Towards a common description format for prosopographical data

Bologna, 05 June 2013

Introduction & Contents

- CRR-MM University of Bologna
  - We are a multi-disciplinary group tasked to assist researchers in improving and expanding their information science solutions for their research data and multimedia content
- Contents of this presentation
  - Goals and Importance of Interoperability
  - Contents of our proposal
  - Details of our proposal and Examples

Current Status

- Several important projects exist and are brought together by the Heloïse initiative
- At the moment, each project stands on its own, and no automatic interoperability frameworks or mechanisms are in place
- However, the efforts of these projects are all in the same field of scholarly knowledge even if with different data subsets

Goals

- Investigate on the existing projects, obtaining an overview of their structure, purpose, scope and conceptual models
- Assess shared issues and needs
- **Formalize a first proposal for a shared metadata model**
- Recommend some solutions and encourage a discussion on future developments

Importance of Interoperability

- Semantic and technical Interoperability brings important advantages to the participants and the recipients:
  - Enables a more effective use of the results produced by existing, parallel efforts
  - Encourages the emergence of standards
  - Reduces costs, ambiguity and complexity
  - Makes cooperation much easier
- **Interoperability != Loss of freedom**

Advantages of Interoperability

- In short, data from these research efforts becomes machine readable, allowing automatic programs for data interchange to efficiently leverage existing scholarly efforts, empowering researchers and users
- Services enabled can vary, including:
  - Federated search and visualization platforms
  - Exporting, importing and crosslinking records
  - Publishing results as Linked Open Data

Questionnaire Assessment

- In addition to our investigation of available online repositories, we sent out a questionnaire to interested stakeholders
- Interesting results from the answers:
  - At the moment there are just very small space and time concordances in the projects' scope
  - However, there are relevant affinities in the research subjects, both as conceptual entities and on their related properties.
  - There is definitely an interest in the potential for interoperability and data interchange

Aims of our proposal

- We hope to achieve:
  - Satisfying handling of the common concepts
  - Lightweight approach: ease of implementation
  - Modular and modern architecture
  - Flexibility to represent individual peculiarities
  - Expansibility and customizability
  - Reusing and referencing already existing and accepted models, wherever possible

Contents of our proposal

- Given the aims and the results of our assessment , our proposal hinges on:
  - An explanation on the metadata concepts by an implementation through XML, providing:
    - A formal RelaxNG schema (both syntaxes)
    - Documentation inside the schema itself
    - Some usage examples
  - The metadata proposal architecture is structured in an appropriate way for the concepts to be translated to ontologies

Main Concepts

Elements in our metadata proposal mostly belong to one of two macro categories:

- Entities (*Person, Place, Studium*…)
  - Are the concepts handled by the data supplier
- Factoids (*Name, Affiliation, Kinship*…)
  - Are sourced assertions about 1 or more entities
- Relationship to RDF Triples (S-P-O):
  - Entities are akin to subject & object resources
  - Factoids are akin to the predicates

(Conceptual) Entities

- Are the elements modeling the main concepts handled by the projects we examined, like a *Person*, an *Office* or a *Study Subject*.
- They are either the subject or the objects of the information provided by the data suppliers and described by the means of Factoids.
- They can appear either at a high level (as the subject), containing one or more *Collection Of Information*, or at lower level, referenced (or directly inserted in the markup) as the object of assertion made by factoids

Factoids

- They are elements used to assert that:
  - from the source S, some fact F can be stated about subject entity E.
  - This can be coupled with a set of time information T
  - Or express a relationship between subject entity E and other object entities O1, O2, etc.
- **They are always backed by 1+ sources**
- **A factoid is not an absolute assertion:**
  - It records that a selected source claims that a fact involves this entity.
  - Factoids can be contradictory with each other!

Factoids Advantages

- An approach tested successfully by other important prosopographical DBs (e.g: PASE)
- Models the knowledge of subjects of the discourse as a set of sourced data
- Perfect for integration with DB records
- ***Event-driven*****:** Adaptable and simple
- We are providing a set of premade Factoids, but a flexible and easy to use mechanism for extensibility is in place

Other features and elements

- Versatile and Powerful Time Specification
  - Modeled after LKIF:Time
- Flexibility of the content model
  - Entities and factoid can have structured data
  - Data can also be supplied unstructured by using the *Value* tag
  - Easy to reference outside resources (href)
- Ability to express different degree of certainty (reliability) about factoids

Extensibility

- Extensibility and ease of customization
  - Aside from pre-supplied ones, the factoid element itself has a "*type*" attribute that makes extension and sub-typing very easy
  - A "*rel*" attr. for factoids and entities allows to further specify the meaning of a relationship
  - "*Class*" attr. for entities for concept subsets
  - *Note* factoid element to mix text and other existing factoids

Example 1 – Simplicity

```
<person xml:id="atelier.eu.examples.asfe.jsepulveda">
   <infoCollection xml:id="collections.asfe.jsepulveda"
        src="http://asfe.unibo.it/it/persona/LL1012">
        <fallbackSources>
                <source href="atelier.eu/sources/asfe"/>
        </fallbackSources>
   <name normalized="true">
        <firstName>Juan</firstName>
        <surname>Sepúlveda, de</surname>
        <source useFallback="true"/>
   </name>
   <name>
        <firstName>Iohannes</firstName>
        <source href="atelier.eu/sources/asfe/guerrini.1003"/>
   </name>
   <name>
        <surname>Sepulveda</surname>
        <source href="atelier.eu/sources/asfe/guerrini.1003"/>
   </name>
[…]
```

Example 2 – Structured Data

```
<person xml:id="atelier.eu.examples.asfe.wolfgang.kastner">
[…]
 <!-- Enrollment in a nation -->
 <affiliation rel="matriculatus nationis">
   <studium rel="iuristarum"><value>Padova</value></studium>
   <personGroup rel="natio"><value>Germanica</value></personGroup>
   <moment>1571-04-18</moment>
   <source href="atelier.eu/sources/asfe/df.1.2164"/>
 </affiliation>
 <!– Another one, where entities are referred from the outside -->
 <affiliation rel="matriculatus nationis">
   <studium href="atelier.eu/studium/Bologna"/>
   <personGroup class=“natio” href="atelier.eu/groups/Germanica"/>
   <!-- We know it happened BEFORE this date -->
   <before unit="days">
      <moment calendar=“AD”>1573-04-13</moment>
   </before>
   <source href="atelier.eu/sources/asfe/siena.1.123"/>
 </affiliation>
[…]
```

Example 3 – Personal Data

```
<person xml:id="atelier.eu.examples.rag.felix.fredrich.hohenzollern">
[…]
      <!-- Let's show some Bio Data -->
      <changeOfHealth>
        <death>
          <interval>
            <begins><moment>1550-01-20</moment></begins>
            <ends><moment>1550-01-30</moment></ends>
          </interval>
        </death>
        <source useFallback="true"/>
      </changeOfHealth>
      <!-- Being a noble -->
      <officeCommission rel="herkunftSozial">
          <office><value>Graf</value></office>
          <moment>1542-08-30</moment>
          <source useFallback="true"/>
      </officeCommission>
[…]
```

Example 4 – Extensibility

```
<person xml:id="atelier.eu.examples.asfe.wolfgang.kastner">
[…]
<factoid type="liberAmicorum">
   <person rel="owner" href="http://asfe.unibo.it/persona/NG0288"/>
   <moment>1575-09-29</moment>
   <source href="atelier.eu/sources/asfe/lib.2"/>
</factoid>
<!-- Another possible solution -->
<changeOfPersonalRelation><friendship rel="liberAmicorum">
   <person rel=“owner”>
        <name><value>Onophrius Perbinger</value>
        <source href="atelier.eu/sources/asfe.lib.2"/>
   </name></person>
   <!-- A note about this could have also been added -->
   <note>Liber Amicorum – example of note
         <source href="atelier.eu/sources/asfe"/></note>
   </friendship> <source href="atelier.eu/sources/asfe.lib.2"/>
   </changeOfPersonalRelation>
[…]
```

Future Opportunities

- Immediate Future:
  - Shared conceptual metadata model ready to be ported to an ontological one, re-using standards (LKIF:time, Biography Light, SPAR)
  - Agreed-upon standard for exporting records
  - Publishing results as Linked Open Data
- A long-term approach:
  - A network of prosopographical databases providing common services of data search, retrieval and visualization

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Jacopo Zingoni

Gioele Barabucci

Area Sistemi Dipartimentali e Documentali – ASDD:
CRR-MM: Centro di Risorse per la Ricerca Multimediale
mail: asdd.crrmm@unibo.it

www.unibo.it