ASFE and PROSO
conditions and models for interoperability

Lyons, March 2014

Jacopo Zingoni
Introduction (1)

• **ASFE: *Amore scientiae facti sunt exules***
  – A project focused on the time frame between 1500 and 1800, with three main targets:
    1. Conduct a census of the student population of the University (Studium) of Bologna (*Onomasticon Studii Bononiensis*)
    2. Survey the graduates of Italian Universities (*Italici Doctores*)
    3. Conduct a census of the foreigners undertaking academic travels (*peregrinatio academica*) in Italy (*Iter Italicum*)

• **CRR-MM University of Bologna**
  – As a multi-disciplinary support center, we are cooperating with the authors of ASFE to provide consulting for improving and expanding their information science solutions for their research data and multimedia content
• **PROSO** [1]
  
  – A conceptual *reference model* for storage, use, publication and sharing of *prosopographic records*, presented at Heloïse 2013
  
  – Aims to provide researchers with a common vocabulary and a conceptual organization to use while collecting, debating, sharing and modeling the data
  
  – Has a fully documented XML serialization of the model, RNG schema included

Contents

• Contents of this presentation
  – Goals and advantage of interoperability
  – PROSO conceptual model proposal
Interoperability: Advantages

• Interoperability != Loss of freedom
• Interoperability allows:
  – Effective use of existing, parallel efforts
  – Emergence of good practice and standards
  – Reduced costs, ambiguity and complexity
  – Makes cooperation much easier
  – Enables new discoveries on the shared data, facilitating the extension of group knowledge
Desirable Goals (1)

• *Datasets should become interoperable by adopting an agreed-upon exchange format*

• Federated search and data visualization

• Shared repository for nomenclatures of common interest
  – E.g.: an authority list or a concordance table for names of places and political entities
Desirable Goals (2)

• Provenance tracking for every record, and possibly every data fact
  – This is in addition to source tracking
• *Common and extensible vocabulary for describing concepts and properties*
• Importing, exporting and crosslinking records
• Ability to generate data mashups
Final, ambitious goals

• **Create a web of linked open data (LOD), expressing our data in RDF triples.**
  – With SPARQL endpoints (single-federated or multiple)
  – And an user friendly interface that could query the data shared, tracking provenance.
  – Results presented in several interactive ways.

• **Enchant the public and allow domain experts to make new and significant discoveries more easily.**
Features of PROSO

- Handles all of the common concepts
- Event driven, factoid based approach
- Implementation agnostic
  - XML serialization, ready for RDF & SemWeb
- Lightweight: ease of implementation
- Modular and extensible architecture
- Flexibility to represent peculiarities of individual projects: highly customizable
- Reuses and references already existing models
- Versatile and expressive time specification
Main Concepts of PROSO

- **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
  - Both are parts of an extensible class hierarchy
(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.

• All entities should be uniquely identifiable (URI).
Factoids

• They are the means used to assert that:
  – the source S believes that the 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:
  – A source claims that fact F involves this entity
  – Factoids can be contradictory with each other!
  – A degree of reliability can be associated to factoids
Event driven approach

- Prosopographical and biographical info modeled as events “changing” the subject

For instance, let’s suppose we want to state the following:

This individual has held a title of “DomHerr” (Canon) from 1545-01-19 in Augsburg, and was born someday between the 20 and the 30 of January 1521. (fictional facts)
Example (XML)

```xml
<person xml:id="EXAMPLE.001">
    [...]
    <!-- Being promoted to office is a change of Social Relation -->
    <changeOfSocialRelation type="officeCommission">
        <office><value>Domherr</value></office>
        <moment>1545-01-19</moment>
        <place href="Augusburg"></place>
        <source "...fiction…"/>
    </officeCommission>
    <!-- Let’s show some Bio Data: birth is a Change of Health -->
    <changeOfHealth>
        <birth>
            <interval>
                <begins><moment>1521-01-20</moment></begins>
                <ends><moment>1521-01-30</moment></ends>
            </interval>
        </birth>
        <source "...fiction…"/>
    </changeOfHealth>
```
Same Example (RDF: Turtle)

```
ex:00001 a proso:Person ;
# Factoids are rendered with the anonymous nodes like _:bn-title
# The details of the changeOfSocialRelation factoid
_:bn-title rdfs:label "Domherr"@de , “Canon"@en ;
proso:moment "1541-12-07"^^xsd:date ;
proso:takesPlaceIn rag:places_Augsburg ;
proso:office "Domherr Mag. Art." ;
proso:source “FICTION” .

# Some information about his birth (subclass of Change of Health)
_:bn-health rdfs:label “birth”@en ;
proso:interval _:bn-interval .
proso:source “FICTION” .
_:bn-interval proso:beginsMoment "1521-01-20"^^xsd:date ;
proso:endsMoment "1521-01-31"^^xsd:date .
```
Jacopo Zingoni

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

http://www.unibo.it/it/ricerca/progetti-e-iniziative/centro-di-risorse-per-la-ricerca-multimiddiale-crr-mm