Linked Open Data and research in the humanities: revised methodology and case studies

Francesca Tomasi
Department of Classical Philology and Italian Studies
University of Bologna

http://www.unibo.it/docenti/francesca.tomasi
Methodology renewal in the humanities

How to extract meaning from humanistic resources?

The interlinking idea and approach on data:

✓ Knowledge graph
  • Meaning is the network of connections

✓ Disciplines
  • Meaning is found going beyond the barricades

✓ Context
  • Meaning is acquiring knowledge from surroundings

✓ Link
  • Meaning is typed link, i.e. semantic relationships

i.e. give meaning to data
Emerging kws in the (digital) humanities

✓ Collaboration. Not only in the hard sciences
  • working method; sharing skills (and data!); idea interchange; common projects

✓ Tools and environment implementation
  • online collaboration in creating tools

✓ Markup vs annotation or Information vs knowledge
  • information representation: embedded (in-line) and stand-off markup
  • knowledge representation: OWL ontology and TM topic maps

✓ Provenance and context
  • The archival study lesson… Who said what? Where? From which source is the information extracted?

✓ Authorship and authoritativeness
  • authorities for semantic interchange
  • defining quality principles for attribution of statements

✓ Final user approach
  • Web of data; layering: back end annotation (classification) and ‘faceted’ (filtered) navigation; network of semantically related items
How do we work on LOD in the humanities?

- Developing and integrating tools
  - Collaborative dimension
- Elaborating ontologies
  - Reuse, mapping, merging, alignment, creation
- Reflecting on the concept of knowledge sites
  - Production, preservation, managing, dissemination, access
- Creating datasets
  - Formalize CH standards and semantic linking to the LOD cloud

The research group @Unibo

- Marilena Daquino, Silvio Peroni, Francesca Tomasi, Fabio Vitali
How do we work on LOD in the humanities?

✓ Developing and integrating tools
  • Collaborative dimension

✓ Elaborating ontologies
  • Reuse, mapping, merging, alignment, creation

✓ Reflecting on the concept of knowledge sites
  • Production, preservation, managing, dissemination, access

✓ Creating datasets
  • Formalize CH standards and semantic linking to the LOD cloud

The research group @Unibo
✓ Marilena Daquino, Silvio Peroni, Francesca Tomasi, Fabio Vitali
The first layer of the PRoles Ontology: role attribution

An agent declare that - in a source (document) – it is attested that another agent, having a specific role, participated in a certain event.

The second layer of the PRoles Ontology: participation to events

The Political Roles (PRoles) Ontology is an OWL 2 DL ontology that allows one:

- to represent political role attributions;
- their possible links to related events;
- and the provenance declaration;
- by means of particular classes and properties imported, and used, by several models (PRO, n-ary participation pattern and PROV-O).
An agent (3) extracts some contents form a source as RDF statements (7). Such contents are born out, i.e. generated, from an interpretation activity (1), of a certain type (6), made by another agent (e.g. an editor) by using a particular criterion (6). The agent (e.g. editor or cataloguer) makes interpretations (1) in a new work (5) regarding the cultural object (4).

The interpretation act declares the authority on which it’s based (‘cito’) and/or the agreement (or disagreement) with another interpretation act.
How do we work on interlinking in the humanities?

✓ Developing and integrating tools
  • Collaborative dimension

✓ Elaborating ontologies
  • Reuse, mapping, merging, alignment, creation

✓ Reflecting on the concept of knowledge sites
  • Production, preservation, managing, dissemination, access

✓ Creating datasets
  • Formalize CH standards and semantic linking to the LOD cloud

The research group @Unibo

✓ Marilena Daquino, Silvio Peroni, Francesca Tomasi, Fabio Vitali
Vespasiano da Bisticci’s Letter
A knowledge site
http://vespasianodabisticciletters.unibo.it

- Homepage
- A letter
- XML/TEI transcription (first level markup) based on a primary source (archival document/library manuscript);
- Comments as a second level (un-formal) annotation;
- Extracted letters’ metadata;
- Tags/entries (indices);
- Links to LOD cloud (Dbpedia).
- Index for mentioned persons, linked to authorities (VIAF, LCCN, SBN, ISNI, GND)

In Firenze, adì 23 di giugno 1458.
Vespasiano di Filippo [a tergo]

<br>

We notice: **personal names** (Vespasiano di Filippo, Piero di Chosimo de’ Medici, Piero Strozi, Pipo), **places** (Firenze), **dates** (june 23 1458), latin **texts and authors** (Prinio, Decha), **technical lexicon** (minî, legaranno).

We understand that: 1. Piero de’ Medici requested a Pliny and the Decades from Vespasiano laboratory and that 2. the copy has been made before june 1458. We know also that 3. Piero Strozzi finished the Epitomi and 4. delivered the volume to Pipo, which is identifiable as Filippo Torelli, 5. in order to do the illumination and then 6. to realize the binding.
**The embedded XML/TEI mark-up**

ns:metadata + rdf files#XML:ID + string + ns:metadata

**Personal names**
<tei:persname ref="people.rdf#PS">Piero Strozi</tei:persname>
<tei:persname ref="people.rdf#FT">Pipo</tei:persname>
<tei:persname ref="people.rdf#VdB">Vespasiano di Filippo</tei:persname>
<tei:persname ref="people.rdf#PdM">Piero di Chosimo de’ Medici</tei:persname>

**Manuscripts**
<tei:bibl ref="manuscripts.rdf#P_SN">  
   <tei:author>Prinio</tei:author>  
</tei:bibl>
<tei:bibl ref="manuscripts.rdf#L_D_III">  
   <tei:title>Decha</tei:title>  
</tei:bibl>
<tei:bibl ref="manuscripts.rdf#L_D_IV_E">  
   <tei:title>abreviationi</tei:title>  
</tei:bibl>

**Lexicon**
<tei:term type="binding" ref="lexicon.rdf#leg">legaranno</tei:term>
<tei:term type="illumination" ref="lexicon.rdf#min">minî</tei:term>
RDF for internal relationships. Towards the LOD

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>PREDICATES</th>
<th>OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>has_variant_forms</td>
<td>Piero, Piero di Cosimo de’ Medici, Principe di Firenze</td>
</tr>
<tr>
<td></td>
<td>is_owner_of</td>
<td>manuscripts.rdf#P_SN manuscripts.rdf#L_D_III manuscripts.rdf#L_D_IV_E</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>PREDICATES</th>
<th>OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>[manuscripts.rdf#P_SN]</td>
<td>has_normalized_form</td>
<td>Plinio, <em>Storia naturale</em></td>
</tr>
<tr>
<td></td>
<td>is_requested_by</td>
<td>people.rdf#PdM</td>
</tr>
<tr>
<td></td>
<td>is_owned_by</td>
<td>people.rdf#PdM</td>
</tr>
<tr>
<td></td>
<td>is_copied_by</td>
<td>people.rdf#PS</td>
</tr>
<tr>
<td></td>
<td>is_illuminated_by</td>
<td>people.rdf#FT</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>SUBJECT</th>
<th>PREDICATES</th>
<th>OBJECT</th>
</tr>
</thead>
<tbody>
<tr>
<td>[lexicon.rdf#min]</td>
<td>has_normalized_form</td>
<td>miniare, miniatura, miniato</td>
</tr>
<tr>
<td></td>
<td>is_referred_to</td>
<td>manuscripts.rdf#L_D_IV_E</td>
</tr>
</tbody>
</table>
The knowledge base

Ontologies: CIDOC-CRM, SKOS, DC, FRBR, EDM, PROV-O, FOAF....

<persName xml:id="LdM">Lorenzo de’ Medici</persName>

- xml:id=LdM
- Access_key: Medici, Lorenzo de’
- Dbpedia URI: http://it.dbpedia.org/page/Lorenzo_de%27_Medici
- VIAF permalink: http://viaf.org/viaf/54169908
- Father-of: Piero de’ Medici (URI)
- Born-in: Firenze (URI)
- Died-in: Firenze (URI)
- Went-in: Roma (URI) - function: marriage
- Trip-to: Venezia (URI) - function: military mission
- Born-when: 1449 (xsd:integer)
- Died-when: 1492 (xsd:integer)
- Iconography: http://www.google.com/search?hl=en&q=lorenzo+de%27+medici&um=1&ie=UTF-8&tbm=isch&source=og&sa=N&tab=wi&ei=_lvcUOEAJqeI4ASikoCYBQ&biw=1146&bih=709&sei=AVzCUNnjIOXV4gTV1oHQCQ
- Biography: https://it.wikipedia.org/wiki/Lorenzo_de%27_Medici
- Author-of: https://it.wikisource.org/wiki/Canti_carnascialeschi_%28Lorenzo_de%27_Medici%29; https://it.wikisource.org/wiki/Novella_di_Giacoppo
- Attested-in: http://it.wikiquote.org/wiki/Maria
- Owner-of: Plinio, Storia Naturale (URI manuscript)
In the first note of the edition (first letter of the collection), Francesca Tomasi asserts that Donato Acciaiuoli held the role of copyist instead of Vespasiano. This assertion is corroborated by the researcher personal reading and by another critical edition of the same letter (Greco). The researcher based her interpretation on a manuscript preserved in BML in Florence (Plut. 90 sup. 30).
Vespasiano’s environment

Letters
XML/TEI documents

RDF Extraction
about People | Manuscripts | Lexicon
RDF data
specified according to
Domain Ontologies

Computer-aided annotation
Text annotation
specified according to
Open Annotation Data Model (OA)

Mapping and linking
The linked open data cloud
Other RDF datasets
specified according to
Existant Ontologies
How do we work on interlinking in the humanities?

✓ Developing and integrating tools
  • Collaborative dimension

✓ Elaborating ontologies
  • Reuse, mapping, merging, alignment, creation

✓ Reflecting on the concept of knowledge sites
  • Production, preservation, managing, dissemination, access

✓ Creating datasets
  • Formalize CH standards and semantic linking to the LOD cloud

The research group @Unibo
✓ Marilena Daquino, Silvio Peroni, Francesca Tomasi, Fabio Vitali
"Zeri e LODE" project
from DB to LOD

The project is organised in 4 steps:

1. Developing ontologies for Scheda F and Scheda OA
   - Zeri Foundation database
   - Create the ontology from E/R model and data in DB

2. Data conversion into RDF
   - Convert DB data into RDF according to the ontologies
   - Create a responsive website based on a public faceted user interface for browsing RDF data

3. Publishing in LOD through RDF links
   - Add links to LOD
   - Ibis. Reuse existing models

4. Developing smart applications that use such data for real
   - Different devices for browsing
Zeri & LODE dataset

[http://data.fondazionezeri.unibo.it/](http://data.fondazionezeri.unibo.it/) - [http://w3id.org/zericatalog](http://w3id.org/zericatalog)

- About 16,000 works of art and more than 30,000 photographs depicting such works accurately described by means of around 10 million of RDF triples.

- New ontologies: HiCO, FEntry and OAEntry.

- Used Ontologies: CIDOC-CRM, SPAR (Fabio, CiTO and PRO).

- Mapped Ontologies: CIDOC-CRM

- SPARQL endpoint

- Data browsing in LodView RDF browser.

- Linked datasets: AAT, Getty ULAN, VIAF, Geonames, Dbpedia, Getty CONA.

Federico Zeri Foundation | Linked Open Data Explorer
A sample of enrichment: authorship attribution

Source of information: Frick Art Reference Library; Fototeca Zeri; Villa I Tatti, Fototeca Berenson
In black: Information recorded in two or three db

Attributions (and relative SOURCE)
Anonymous Sienese school 15th Century (Collezione Volpi; American Art Association 1916)
Anonymous Florentine school 15th Century (R. Offner, 1925; Frick classification)
Anonymous Sienese school 14th Century (Asta Palazzo D'Anziani 1934)
Neri di Bicci (Christie's 1967)
Bicci di Lorenzo (Biblioteca Berenson classification)
Apollonio di Giovanni (Fototeca Zeri classification)

Previous locations/ collection/ sale, in chronological order
Florence, Collezione Volpi
New York, American Art Association, 1916/11/21
G.K. Stetson
Florence, Palazzo D'Anziani, 1934/06/25
Florence, Galleria Bellini
New York, French & Co.
New Windsor (MD), Thomas S. Hyland collection
Greenwich (CT), Thomas S. Hyland collection
London, Christie's, 1967/06/23 1967
New York, Sotheby's, 1985/01/17

Dating
1400-1499
1435-1465
1400-1452

OA Entry
Interpretation
Act
Work of Art
Creation
pro:RoleInTime
Apollonio di Giovanni
1435-1465

Interpretation
Act
Work of Art
Creation
pro:RoleInTime
Anonymous note - verso of photograph
hico:InterpretationCriterion
weak criterion
Anonymous Sienese

Inferred as preferred attribution

Fondazione Zeri

Authorship attribution
hico:InterpretationType

still the stronger criterion
The Centre ensures creation, management, web enhancement, preservation, accessibility over time and citability of digital collections stored.

1. **Analysis and development of software tools and environments** for the management and use of digital resources and metadata.

2. **Web application development** for dynamic access to resources (data; modeling; design).

3. **Preservation of research data:** conversion and data migration for preservation (storage; hosting; integration; curation; persistent identifiers).
DH@Unibo. The project

- CRR-MM (DH Lab); Digital Library; Education; Research Activity; Work for young people
The added value of Digital Humanities (DH) for LOD

- DH for rethinking **knowledge organization** through the lenses of **interpretation**
- DH and **inter/multi disciplinary** approach (trans/cross domain): knowledge hybridization as a solution, not a obstacle
- DH and **new methodologies** in the traditional disciplines (Publishing; Literature; Arts; Law; Economy;…)
- DH and conceptual **formalization** of cultural heritage standards
- DH as a **new way of thinking** about cultural heritage in a digital dimension: scholarly/scientific vs descriptive/objective approach
- DH and the attention to the **end-user**: improving user experience by ‘friendly’ semantic relationships
Thank you!

Francesca