

# LOD2014 LINKED OPEN DATA: WHERE ARE WE?

Roma, 20<sup>th</sup> - 21<sup>st</sup> Feb 2014

Archivio Centrale dello Stato, Roma

Organized by W3C Italy

## METHODS AND EXPERIENCES IN CULTURAL HERITAGE ENHANCEMENT

Francesca Tomasi	Fabio Ciotti	Maurizio Lana	Diego Magro	Silvio Peroni	Fabio Vitali
<i>University of Bologna</i>	<i>University of Roma Tor Vergata</i>	<i>University of Piemonte Orientale</i>	<i>University of Torino</i>	<i>University of Bologna</i>	<i>University of Bologna</i>

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE



<http://www.umanisticadigitale.it>

# THE PROJECT

- ❖ **CH and LOD**
- ❖ **our approach: conversion, extraction, creation**
  - ❖ database **conversion** into LOD (ontology reuse);
  - ❖ **extraction** of LOD from XML/TEI texts;
  - ❖ **creation** of new ontologies to produce LOD.
- ❖ the CH domain: people and roles, ancient and modern places, books and archival documents
- ❖ the aim: best practices in LOD production and dissemination in the CH domain
- ❖ **common strategy:**
  - ❖ ontologies creation and reuse
  - ❖ standoff markup and Open Annotation data Model

# THE CASE STUDIES

## ❖ relational database

- ❖ Zeri Photo Archive database

<http://www.fondazionezeri.unibo.it/catalogo>

## ❖ digital edition

- ❖ Vespasiano da Bisticci Letters  
doi:10.6092/unibo/vespasianodabisticciletters

## ❖ geographical ontology

- ❖ Geolat project

<http://www.geolat.it>

## ❖ archival ontology

- ❖ Proles ontology

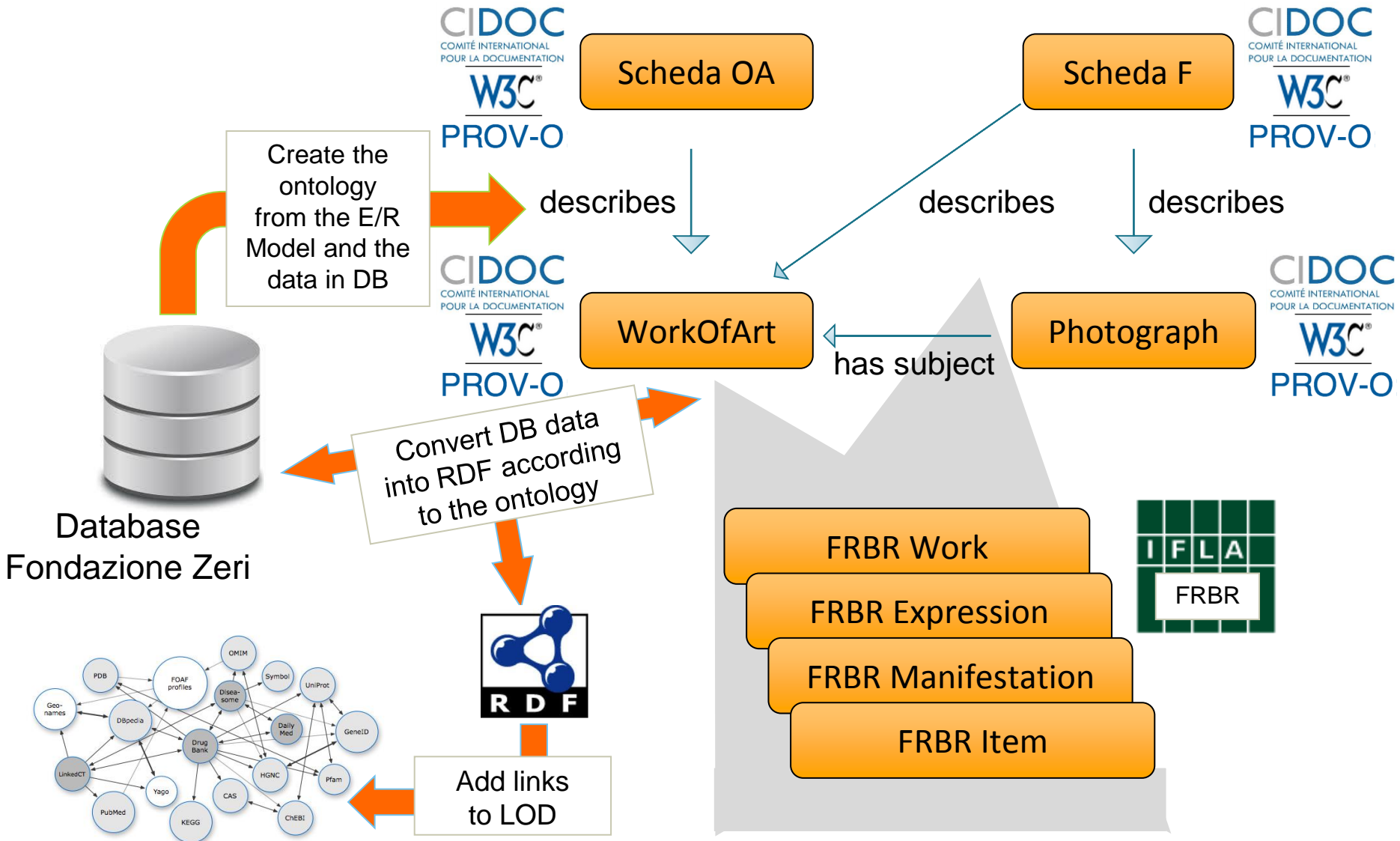
<http://www.essepuntato.it/2013/10/politicalroles>

# ZERI PHOTO ARCHIVE

- ❖ “it is a rich digital catalog, and is today considered one of the most important repertories of Italian art on the web”.
- ❖ our mission is to **convert the database to LOD**:
  - ❖ **reengineer the E/R model** implemented by the database tables, which contain data according to the Scheda F, into **OWL**, to obtain a **first version of an ontology**;
  - ❖ **iteratively enhance the ontology** according to the specifications described by the **Scheda F** and **CIDOC-CRM**, (changing the whole conceptual organisation and entity naming of the existing model as less as possible);
  - ❖ **transform data originally stored in the database into RDF statements** compliant to the OWL ontology developed, by using appropriate scripts ;
  - ❖ apply automatic and semi-automatic mechanisms to generate **links to existing datasets**, such as DBpedia and Europeana.

# ZERI: THE PROCESS

## ONTOLOGY REUSE AND LOD POPULATION



# VESPASIANO, LETTERS A DIGITAL EDITION

- ❖ a digital annotated (XML/TEI) collection of letters from the XV century sent/received to/by the florentine copyist Vespasiano da Bisticci.
- ❖ a web environment that focuses on: **persons mentioned** in the documents; classical latin and greek **manuscripts requested/copied/proposed** to/by Vespasiano da Bisticci's school and their **description**.
- ❖ the purpose is to **identify** persons related to manuscripts in order to expose datasets of **people** related to **manuscripts**, these last described by **technical words**.
- ❖ the XML/TEI annotation (persons, manuscripts and technical terms) has been realized with **embedded markup** (@ref="URI") pointing to **stand-off RDF file** (with assertion) and **controlled form of the names** (VIAF, LCA, Geonames, etc.) for managing attributes values.

# VESPASIANO: THE MODEL

## RDF SUPPORT TO STANDOFF ANNOTATION

SUBJECT	PREDICATES	OBJECT
people.rdf#PdM  <b>URI:</b> <a href="http://vespasianodabisticciletters/people/PdM">http://vespasianodabisticciletters/people/PdM</a>	has_normalized_form	Medici, Piero de': <b>Dbpedia:</b> <a href="http://eu.dbpedia.org/page/Piero_de_Medici">http://eu.dbpedia.org/page/Piero_de_Medici</a> <b>VIAF:</b> <a href="http://viaf.org/viaf/25406033">http://viaf.org/viaf/25406033</a>
	has_variant_forms	Piero, Piero di Cosimo de' Medici, Principe di Firenze
	is_owner_of	manuscripts.rdf#P_SN manuscripts.rdf#L_D_III manuscripts.rdf#L_D_IV_E

SUBJECT	PREDICATES	OBJECT
manuscripts.rdf#P_SN  <b>URI:</b> <a href="http://vespasianodabisticciletters/manuscripts/P_SN">http://vespasianodabisticciletters/manuscripts/P_SN</a>	has_normalized_form	Plinio, <i>Storia naturale</i>
	is_requested_by	people.rdf#PdM
	is_owned_by	people.rdf#PdM
	is_copied_by	people.rdf#PS
	is_illuminated_by	people.rdf#FT

SUBJECT	PREDICATES	OBJECT
lexicon.rdf#min  <b>URI:</b> <a href="http://vespasianodabisticciletters/lexicon/min">http://vespasianodabisticciletters/lexicon/min</a>	has_normalized_form	miniare, miniatura, miniato
	is_referred_to	manuscripts.rdf#L_D_IV_E

❖ geolat-geography for latin literature, is a research project now funded by Fondazione Compagnia di SanPaolo

## ❖ **main aims:**

❖ increasing the value of geographic references in latin texts

❖ enabling innovative access to latin works (e.g. through geography)

❖ contributing to the LOD cloud

## ❖ **work in progress**

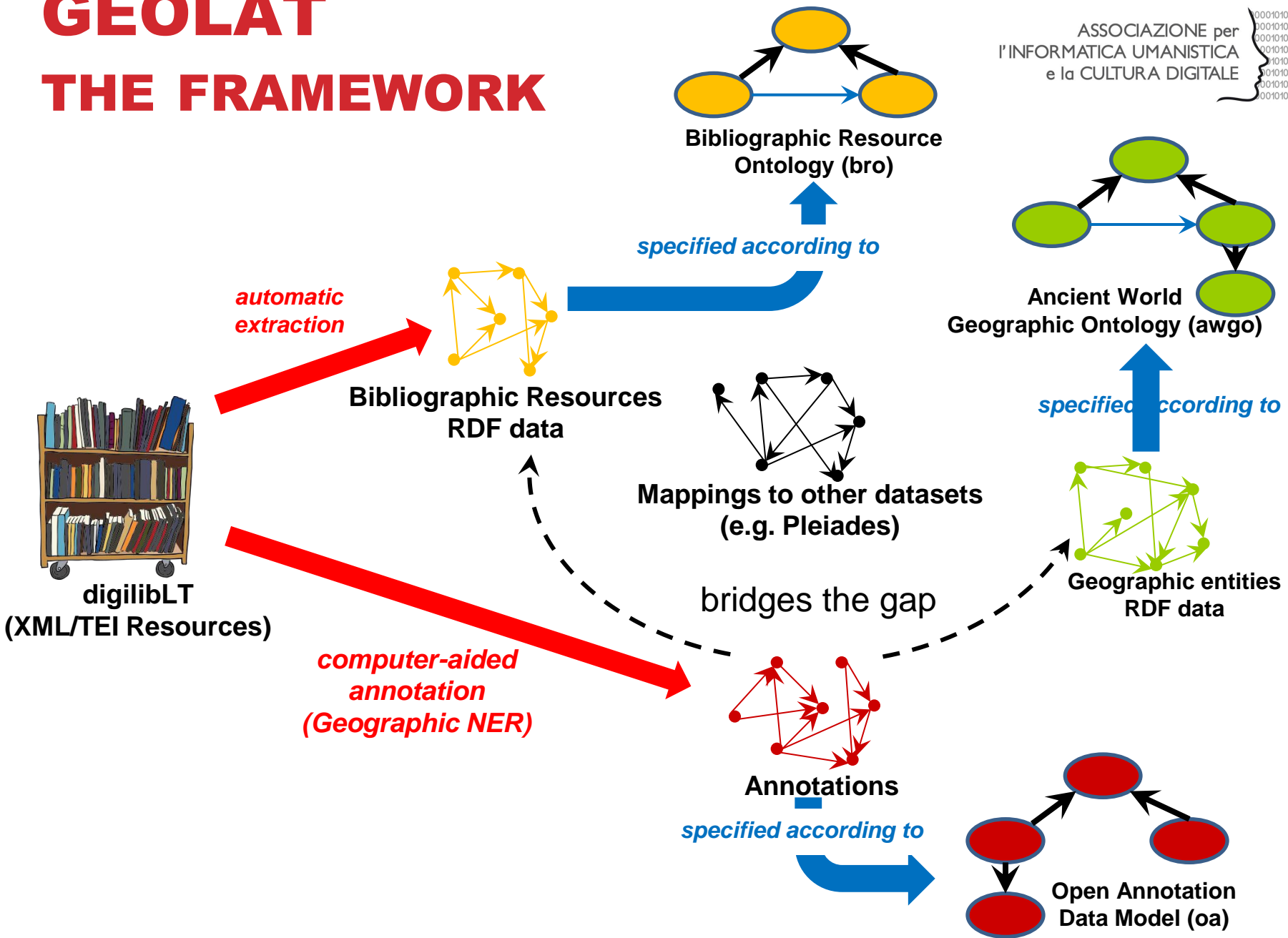


# GEOLAT

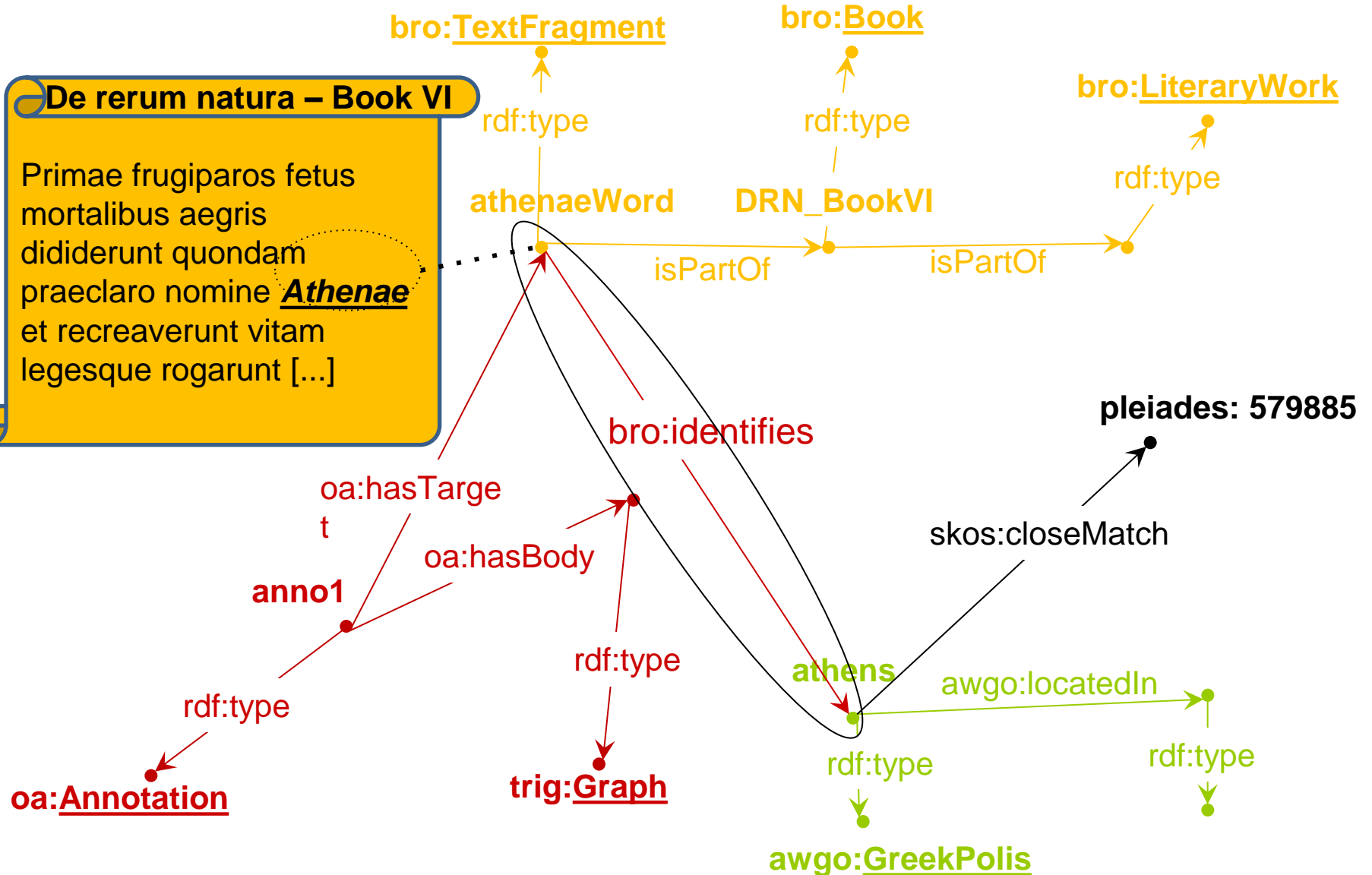
## THE FRAMEWORK

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE

000101010  
0001010101010  
000101010  
0001010101010  
00010101010  
0001010101010  
000101011  
00010101010



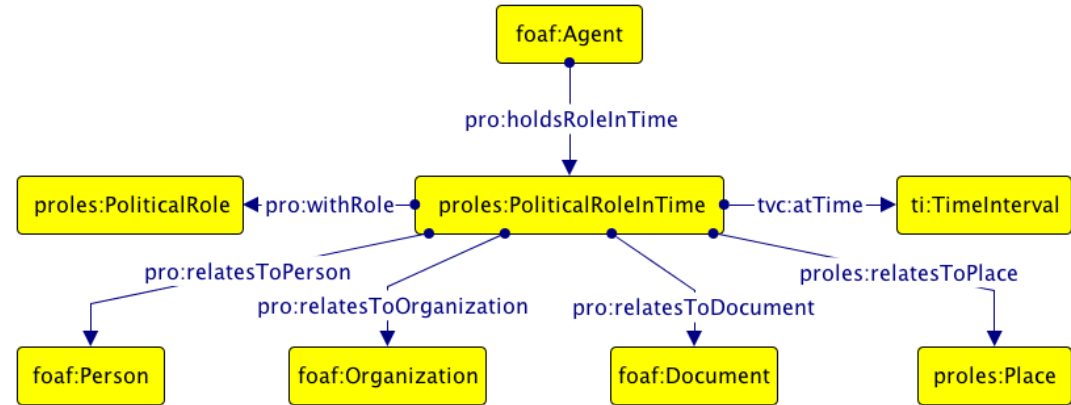
# GEOLAT THE MODEL



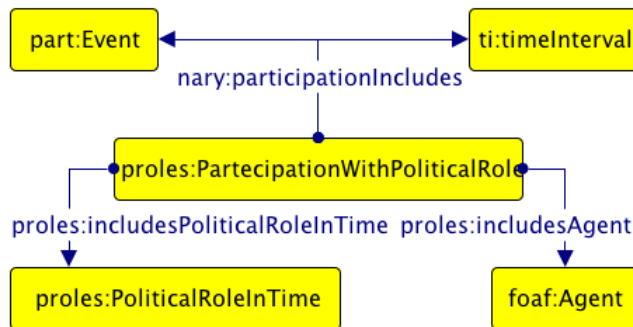
# AN ARCHIVAL ONTOLOGY: PROLES

- ❖ the *Political Roles (PROles) Ontology* is an OWL 2 DL ontology that allows one to represent **political role attributions** and their possible **links to related events** by means of particular classes and properties imported and used by several concepts from **PRO**, **n-ary participation pattern** and **PROV-O**.
- ❖ we are now managing an experiment on Andrea Costa fonds, by exploiting the related authority record (<http://archivi.ibc.regione.emilia-romagna.it/eac-cpf/IT-ER-IBC-SP00001-0000264>), in collaboration with IBC, Soprintendenza per i Beni librari e documentari.

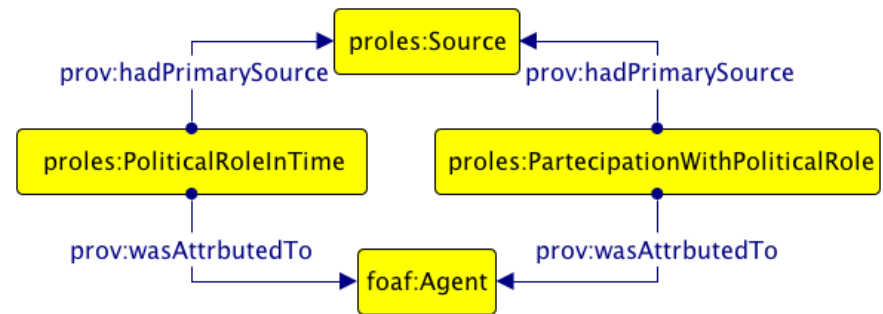
# PROLES: THE MODEL



The **first** layer of the PROLES Ontology: **role attribution**



The **second** layer of the PROLES Ontology: **participation to events**



The **third** layer of the PROLES Ontology: **provenance information**

# FINAL REMARKS

## ❖ the shared method:

- ❖ ontology reuse;
- ❖ definition of new classes and predicates;
- ❖ ontology as the basis for LOD characterization;
- ❖ stand-off markup and OA data model;
- ❖ LOD cloud population;
- ❖ mapping to other datasets

ASSOCIAZIONE per  
l'INFORMATICA UMANISTICA  
e la CULTURA DIGITALE



**THANK YOU!**

FRANCESCA, FABIO C., MAURIZIO, DIEGO, SILVIO, FABIO V.