Maddalena Andreussi – Achille Felicetti

The CIDOC CRM Encoding of the "Fontes ad Topographiam Veteris Urbis Romae Pertinentes" by Giuseppe Lugli

Abstract: This paper presents a project for the creation of an ontology-encoded version of the "Fontes" by Giuseppe Lugli, one of the most important collections of sources for the study of the topography of ancient Rome. Only seven volumes of the work were published by Lugli between 1952 and 1962; the publication of the remaining volumes is still in progress. The goal of the project is the creation of a semantic "ancient sources" GIS, a set of interactive maps of ancient Rome which will provide spatial information along with the descriptions recorded by ancient sources. The model chosen for the encoding is the CIDOC CRM, an international ISO standard developed to describe concepts and relationships used in cultural heritage documentation. The event-based CIDOC CRM ontology seems ideal to describe the topographic and historical layers designed by Lugli, as it will clarify the relations existing among the events recorded by the ancient sources, and the monuments and places to which they refer.

Giuseppe Lugli's "Fontes"

The "*Fontes ad Topographiam Veteris Urbis Romae Pertinentes*" is a collection of ancient sources related to the topography of ancient Rome, edited by the University of Rome, Institute of Ancient Topography. The first volumes were published between 1952 and 1969 under the direction of Giuseppe Lugli, professor of Roman Topography at the University of Rome.

The work is not yet complete. Currently, only the first two volumes containing general information about the history of the city and the ones collecting the sources about the *Regiones Urbis* I, II, V, VI, e VII, the *Fora Imperatorum*, the *Capitolium* and the *Regiones Urbis* X e XI have been published. The editing of the books about *Forum Romanum* (by Maria Giovanna Forni), *Campus Martius* (by Maria Pia Muzzioli), *Aventinus Mons* and *Transtiberim* (by Maddalena Andreussi) is still in progress.

Essentially the aim of the original work was to collect, through a thorough inquiry of the ancient authors, all the textual citations in ancient Greek and Latin sources pertaining to the topography and monuments of Rome, including epigraphical evidence, numismatic evidence and topographical information gained from historical reliefs.

The Internal Structure of the "Fontes"

Each volume of the work has a general organization that is more or less the same. The first part of the book (typically chapters 1 and 2) contains the sources that generally refer to the geography and topography of the place considered, while in the following chapters all the sources that specifically mention buildings and monuments that rose within the borders of that particular Augustan region are recorded. Every quotation in each chapter has a progressive number for easy searching. Such a rigid organization, based mainly on topographic criteria, is very simple to mark up using languages able to describe tree structures like XML.

Lugli enriched the main topographical organization described above by grouping sets of textual citations that refer to the same event and by writing a short description, often extracted from one of the ancient passages, to summarize their content. Thus he introduced an event-based semantic level through the tree structure, creating a complex and detailed set of relations among the places described by the ancient sources and the events they witnessed. Events are usually complex structures in which elements like people, objects, places and time spans are linked together by implicit and explicit relationships. These are very difficult to describe using a simple tree structure; something more complex is required. The use of an ontology allowing the addition of a semantic level to the tree structure

represented by the general (topographical) organization of the volumes seems to be perfect for this purpose.

The CIDOC CRM Ontology

The ontology we have chosen for the semantic encoding of the Lugli's internal organization is the CIDOC CRM. It is an international ISO Standard developed by an interdisciplinary working group of the International Committee for Documentation of the International Council of Museums (CIDOC/ICOM) under the scientific lead of ICS-FORTH. The ontology is composed of 81 classes and 132 properties designed to enable information exchange and integration between heterogeneous sources of cultural heritage information. The CIDOC CRM standard ontology was also chosen for its capability to describe, in a logical and detailed way, concepts and relationships used in cultural heritage documentation. The CIDOC CRM analyses the common conceptualizations behind data and metadata structures to support data transformation, mediation and merging.

Designed to create data sets that can be understood by people and processed by machines, CRM instances can be encoded in many forms: RDBMS, ooDBMS, XML, RDF(S) to make information coming from different contexts and sources (books, museums, excavation records) interoperable and immediately available for implementation of the Semantic Web vision.

The CIDOC CRM classes and properties can be used to encode the summaries provided by Lugli for each group of quotes, while also capturing and encoding the relations among events, people, places, objects and time spans directly referred to by the sources. The ontology is flexible enough to allow a perfect conceptualization of the meaning contained in each sentence and to encode it in a machine-understandable set of documents to be used in different semantic applications.

As the RDF (Resource Description Framework) language is capable of expressing concepts and relationships defined by the ontology in a succint and easily understandable syntax, it will be used to build our set of semantic documents.

The CIDOC CRM Encoding

The encoding process is usually composed of four steps:

1) Finding the relevant elements taking part in the event referred to by the ancient source.
2) Finding the relations that link the relevant elements in order to recreate the referred events in a schematic way.
3) Describing the elements and the relations by using the CIDOC CRM classes and properties.
4) Encoding the CIDOC CRM conceptual descriptions (Crofts et al. 2005) in a formal language (RDF).

As an example of how the various steps can be accomplished, we will attempt to apply them to a fragment taken from Book 19 of *"Fontes"* (*LIBER XIX – Mons Palatinus*). The fragment (XIX, 4, B, 263) is a quote from Suetonius's "Life of Nero" (*De Vita Caesarum, Nero*), recorded by Lugli in the section regarding the House of Caesars on the Palatine (XIX, 4, B – *"Domus Palatinae Caesarum"*) because of the information it gives on the various works Nero carried out on it during his reign. Suetonius says:

…domum a Palatio Esquilias usque fecit, quam primo Transitoriam, mox incendio bsumptam restitutamque Auream nominavit (Suetonius, Nero 31, 1)

(...he made a palace extending all the way from the Palatine to the Esquiline, which at first he called the House of Passage, but when it was burned shortly after its completion and rebuilt, the Golden House.)

The structure of the sentence expresses certain relationships between items, which may or may not be globally identified. Other relations are hidden, but can be guessed at from the context or could be recovered from secondary sources or from background knowledge. In the present case, for instance, neither the name of Nero, nor the date of the events recorded by Suetonius are present in the sentence. However, we can extrapolate them from the context and from the additional indications provided by Lugli, who took pains to make explicit every element that was not immediately recognizable in the source's texts. Sometimes sentences are full of simplifications and hidden constants. One of the main problems is understanding how to recover this information during data integration.

Considering the previous observations, the list of the main elements that can be extracted as fol-

lows (in brackets are the corresponding CIDOC CRM classes):

Actors:
Nero (*E39 Actor*)
Events:
Palace burning (*E6 Destruction Event*)
Activities:
Palace extension (*E79 Part Addition*)
Palace rebuilding (*E11 Modification*)
Name assignments (*E13 Attribute Assignment*)
Objects:
The palace (*E24 Physical Man-Made Thing*)
Appellations:
House of Passage (*E41 Appellation*)
Golden House (*E41 Appellation*)
Places:
Palatine (*E53 Place*)
Esquiline (*E53 Place*)
Dates:
64 AD (*E52 Time Span)*

When the list of relevant elements is defined, it is possible to establish the relationships among the classes by using the CIDOC CRM properties (predicates). The whole sentence can now be conceptually represented as a set of assertions (triples) composed by a subject, a predicate and an object:

Nero (*E39 Actor*) **P14B Performed** Palace extension (*E79 Part Addition*)
Palace extension (*E79 Part Addition*) **P7F Took Place At** Esquiline (*E53 Place*)
Palace extension (*E79 Part Addition*) **P110F Augmented** The palace (*E24 Physical Thing*)
Nero (*E39 Actor*) **P14B Performed** Name assignment (*E13 Attribute Assignment*)
Name assignment (*E13 Attribute Assignment*) **P141F Assigned** "House of Passage" (*E41 Appellation*)
Palace burning (*E6 Destruction*) **P13F Destroyed** The palace (*E24 Physical Thing*)
Palace burning (*E6 Destruction*) **P4F Has Time Span** 64 AD (*E52 Time Span*)
Nero (*E39 Actor*) **P14B Performed** Palace rebuilding (*E11 Modification*)
Palace rebuilding (*E11 Modification*) **P7F Took Place** At Palatine (*E53 Place*)
Palace rebuilding (*E11 Modification*) **P31F Has Modified** The palace (*E24 Physical Thing*)
Nero (*E39 Actor*) **P14B Performed** Name assignment (*E13 Attribute Assignment*)
Name assignment (*E13 Attribute Assignment*) **P141F Assigned** "Golden House" (*E41 Appellation*)

Once we have semantically expressed the sentence through the CIDOC CRM set of triples, we can finally create its digital version by encoding each triple in an RDF triple. The RDF metadata model is based upon the same idea of making statements about resources in the form of subject-predicate-object expressions in RDF terminology. The subject denotes the resource, and the predicate denotes traits or aspects of the resource and expresses a relationship between the subject and the object.

For our example, the RDF/XML syntax should be used, as shown above:

```
 <cidoc:E53.Place
rdf:about="Esquiline"/>

 <cidoc:E13.Attribute_Assignment
rdf:about="GH_Name_assignment">
 <cidoc:P141F.assigned
rdf:resource="Golden_House"/>
 </cidoc:E13.Attribute_Assignment>

 <cidoc:E41.Appellation
rdf:about="Golden_House"/>

 <cidoc:E13.Attribute_Assignment
rdf:about="HP_Name_assignment">
 <cidoc:P141F.assigned
rdf:resource="House_of_Passage"/>
 </cidoc:E13.Attribute_Assignment>

 <cidoc:E41.Appellation
rdf:about="House_of_Passage"/>

 <cidoc:E24.Physical_Man-Made_Thing
rdf:about="The_Palace"/>

 <cidoc:E52.Time-Span
rdf:about="Year_64 _AD"/>

 <cidoc:E6.Destruction
rdf:about="Palace_Burning">
 <cidoc:P13F.destroyed
rdf:resource="The_Palace"/>
```

```
 <cidoc:P12F.occurred_in_the_pres-
ence_of rdf:resource="The_Palace"/>
 <cidoc:P93F.took_out_of_existence
rdf:resource="The_Palace"/>
 <cidoc:P4F.has_time-span
rdf:resource="Year_64 _AD"/>
 </cidoc:E6.Destruction>
```

In the same way the relations can be expressed in RDF, as demonstrated here:

```
 <cidoc:E39.Actor rdf:about="Nero">
 <cidoc:P14B.performed
rdf:resource="GH_Name_assignment"/>
 <cidoc:P14B.performed
rdf:resource="HP_Name_assignment"/>
 <cidoc:P14B.performed
rdf:resource="Palace_extension"/>
 <cidoc:P14B.performed
rdf:resource="Palace_rebuilding"/>
 </cidoc:E39.Actor>

 <cidoc:E79.Part_Addition
rdf:about="Palace_extension">
 <cidoc:P7F.took_place_at
rdf:resource="Esquiline"/>
 <cidoc:P110F.augmented
rdf:resource="The_Palace"/>
 <cidoc:P31F.has_modified
rdf:resource="The_Palace"/>
 </cidoc:E79.Part_Addition>

 <cidoc:E11.Modification
rdf:about="Palace_rebuilding">
 <cidoc:P7F.took_place_at
rdf:resource="Esquiline"/>
 <cidoc:P7F.took_place_at
rdf:resource="Palatine"/>
 <cidoc:P31F.has_modified
rdf:resource="The_Palace"/>
 </cidoc:E11.Modification>
```

CIDOC CRM also makes it possible to capture the implicit semantic meaning hidden in the sentences in order to extend and enrich the encoding by providing a set of complex spatial relation definition properties ("forms_part_of", "contains", "overlaps_with", "borders_with", "falls_within") and advanced temporal definition predicates ("occurs_before", "occurs_after") to define sequences of events, *termini ante* and *post quem*.

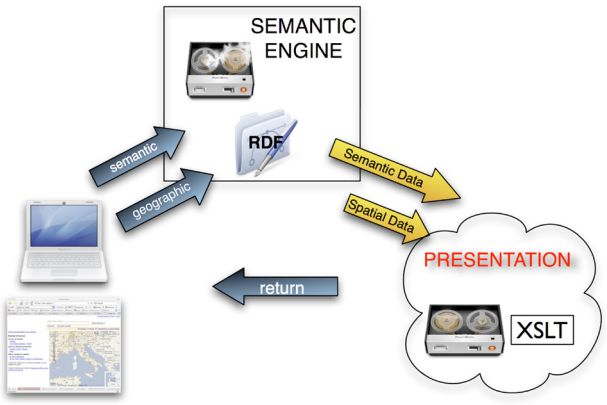

Fig. 1. The semantic engine for the RDF management.

The Encoded Data in Action

The set of RDF documents generated from the encoding process can be stored and managed by powerful semantic engines. These are able to evaluate semantic and geographic queries, implemented through an advanced set of user interfaces, generated according to specific criteria, to return relevant results.

Semantic queries will be serialized to provide a wide range of entry points (people, places, events and other elements described by the ancient sources). This will allow users to extend their search to find all the related items and follow the various relationships among elements.

Spatial queries will be generated directly by using a set of detailed maps of ancient Rome to select spatial areas or specific monuments while obtaining descriptions of events and related elements recorded by the ancient sources.

The query-on-map framework will be implemented using modern and powerful open source GIS, which is able to manage XML formats (such as GML and other standard languages developed by the Open Geo-spatial Consortium) and to generate maps dynamically. This will allow the creation of an "ancient sources" GIS of Rome combining geographic information and event-based semantic data.

Conclusions

The strength of the ontological approach we are planning to use for our project lies in the portabil-

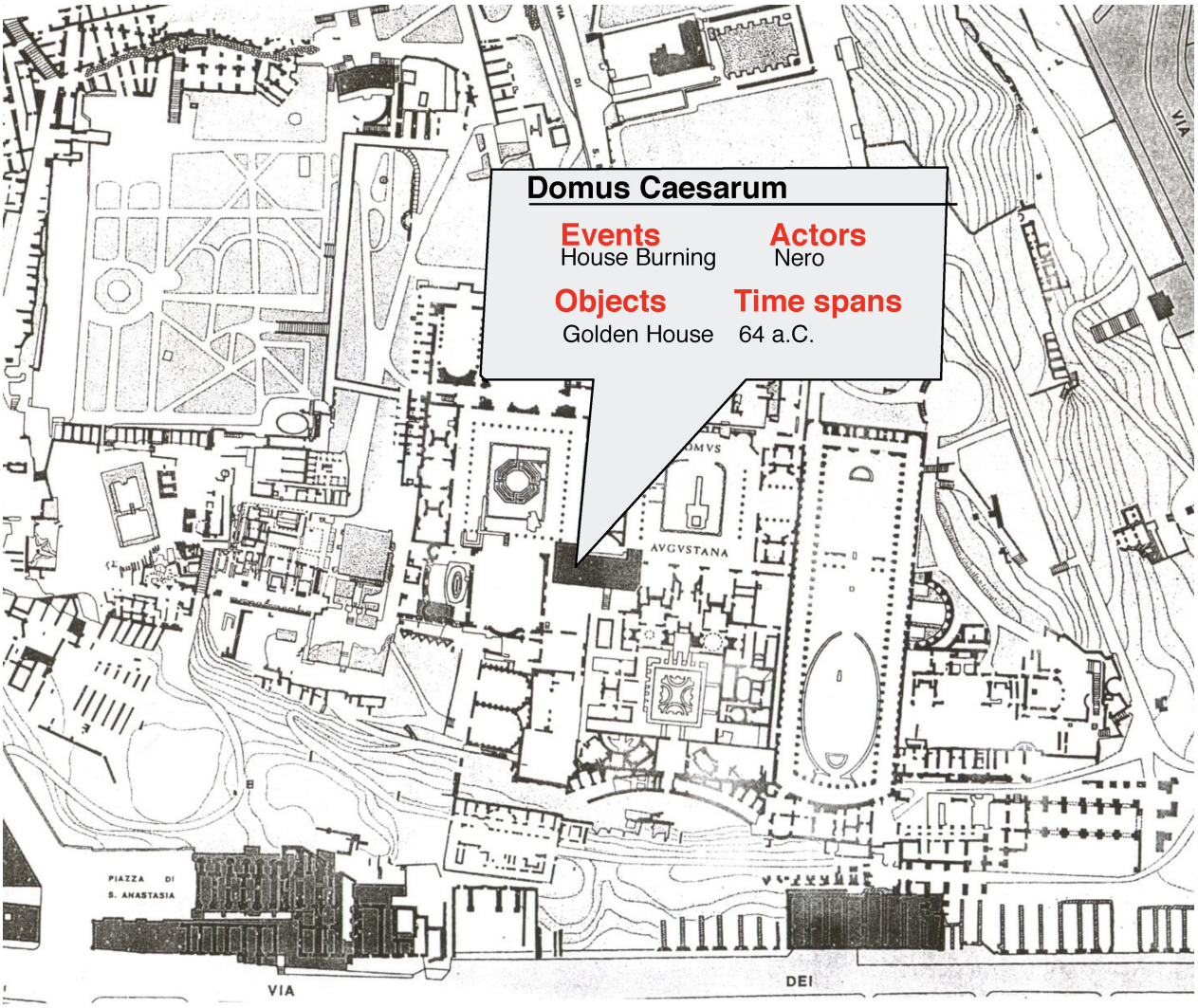

Fig. 2. The query-on-map interface.

ity of data generated using a conceptual model. Standard syntaxes and open formats enhance the archives' reusability and long term preservation. RDF datasets created using the CIDOC CRM formal ontology will be flexible and standard enough to be used in different contexts (not only in GIS frameworks) and integrated with other similar resources for the study of the topography of ancient Rome. Integration among different data sources is merely the first step towards the creation of a Semantic Web where data is linked together in a dynamic and meaningful way. The electronic version of "*Fontes*" will be immediately Semantic Web compliant and ready to be linked with other geographic and bibliographic systems, becoming a new and powerful pillar for the study of the topography of ancient Rome.

References

D'Andrea et al. 2006

A. D'Andrea / A. Felicetti / S. Hermon / F. Niccolucci / T. Zoppi, I linguaggi del W3C e gli strumenti Open Source per la gestione dei dati archeologici. In: R. Bagnara / G. Macchi Janica (eds.), Atti del I Workshop su Open Source, Free Software e Open Format nei processi di ricerca archeologici, Grosseto, Italy, May 8, 2006 (Grosseto 2006).

Bergamaschi et al. 1998

S. Bergamaschi / S. Castano / S. De Capitani De Vimercati / S. Montanari / M. Vincini, An Intelligent Approach to Information Integration. International Conference on Formal Ontology in Information Systems (Amsterdam 1998).

Berners-Lee 1998

T. Berners-Lee, Semantic Web Roadmap: an attempt to give a high-level plan of the architecture of the Seman-

tic WWW. http://www.w3.org/DesignIssues/Semantic.html [31 Dec 2007].

Crofts et al. 2005
N. Crofts / M. Doerr / T. Gill / S. Stead / M. Stiff, Definition of the CIDOC Conceptual Reference Model 4.2. Heraklion, June 2005. http://cidoc.ics.forth.gr/docs/cidoc_crm_version_4.2.pdf [31 Dec 2007].

Doerr 2001
M. Doerr, CIDOC Conceptual Reference Model: Correlation Test Project Results. http://cidoc.ics.forth.gr/testproject_results.html [31 Dec 2007].

Doerr 2003
M. Doerr, The CIDOC CRM: an Ontological Approach to Semantic Interoperability of Metadata. AI Magazine 24,3, 2003, 75–92.

Doerr et al. 1996
M. Doerr / N. Crofts / I. Dionissiadu / C. Gritzapi, Notes on the transformation of the CIDOC relational data model. http://cidoc.ics.forth.gr/docs/notes_trans_cidoc.rtf [31 Dec 2007].

Erdmann / Studer 1999
M. Erdmann / R. Studer, Ontologies as Conceptual Models for XML Documents (Karlsruhe 1999).

Felicetti 2006
A. Felicetti, MAD: Managing Archaeological Data. In: M. Ioannides / D. Arnold / F. Niccolucci / K. Mania (eds.), The e-volution of Information Communication Technology in Cultural Heritage. Where Hi-Tech Touches the Past: Risks and Challenges for the 21st Century (Budapest 2006) 124–131.

Genereux / Niccolucci 2006
M. Genereux / F. Niccolucci, Extraction and Mapping of CIDOC CRM encodings from texts and other digital formats. In: M. Ioannides / D. Arnold / F. Niccolucci / K. Mania (eds.), The e-volution of Information Communication Technology in Cultural Heritage. Where Hi-Tech Touches the Past: Risks and Challenges for the 21st Century (Budapest 2006).

Guarino 1998
N. Guarino, Formal Ontology and Information Systems. In: N. Guarino (ed.), Formal Ontology in Information Systems. Proceedings of the First International Conference (FOIS'98), June 6-8, Trento, Italy (Amsterdam 1998) 3–15.

Kuhn 1994
W. Kuhn, Defining semantics for spatial data transfers. In: T. C. Waugh/R. G. Healey (eds.), Advances in GIS Research. Proceedings of the Sixth International Symposium on Spatial Data Handling 2 (Edinburgh 1994) 973–987.

Open Geospatial Consortium 2003
Open Geospatial Consortium, Geography Markup Language (GML) Implementation Specification. https://portal.opengeospatial.org/files/?artifact_id=7174 [31 Dec 2007].

World Wide Web Consortium 2004a
World Wide Web Consortium, RDF/XML Syntax Specification (Revised), Recommendation. http://www.w3.org/TR/2004/REC-rdf-syntax-grammar-20040210/ [31 Dec 2007].

World Wide Web Consortium 2004b
World Wide Web Consortium, Resource Description Framework (RDF): Concepts and Abstract Syntax, Recommendation. http://www.w3.org/TR/2004/REC-rdf-concepts-20040210/ [31 Dec 2007].

World Wide Web Consortium 200?
W3C Semantic Web Activity, Resource Description Framework (RDF). http://www.w3.org/RDF/ [31 Dec 2007].

Maddalena Andreussi

Università degli Studi di Roma "La Sapienza"
Scienze storiche archeologiche antropologiche dell"antichità
via Giovanni Pantaleo 5
00152 Roma
Italy
m.andreussi@virgilio.it

Achille Felicetti

PIN, University of Florence
VAST-LAB
Piazza Ciardi, 25
509100 Prato
Italy
achille.felicetti@pin.unifi.it