Semantic interoperability via ontology mapping

Andrea D'Andrea[1], Franco Niccolucci[2]
EPOCH
[1] Università di Napoli "L'Orientale" – [2] PIN, Prato

1. Introduction.

According to a well-known definition [Staab, Studer 2004], an ontology is "*a formal explicit specification of a shared conceptualization for a domain of interest*" – in other words, it is related to a community of users. As long as such users share the same concept organization, namely the same ontology, there is no issue and interoperability is guaranteed. Problems arise, on the contrary, when different domains and the related communities of interest (and practice) try to co-operate and aim at interoperable archives without any common ontology. In such cases, one might believe that they generate a more general (perhaps interdisciplinary) domain, and a more general community, and that there exists some overarching ontology, of which the original domain ontologies are just specializations. So, for example, if one deals with Renaissance paintings and another one with 19th century stamps, a 2D image ontology could be expected to guarantee interoperability among them, since both domains are subsets of the set of 2D images. It is a tempting approach, particularly for those neither involved with paintings nor with stamps, who can claim to solve the problems of both in this way. In many cases, however, such a generalizing approach leads to build a very complex theoretical structure based on a bare-bone ontology (e.g. Dublin Core), unfortunately of rather little utility for the individual domains. In fact, there are very few (if any) applications where this approach has proved to be successful in practice.
The opposite, bottom-up approach is often based on ad-hoc solutions, strongly relying on peculiar features of the domain, which loose significance when another domain is considered. This approach is sometimes strongly supported by heritage professionals, perhaps scared of loosing the control on their discipline if it is contaminated with concepts coming from elsewhere. The usual argument is that interoperability is unnecessary in cultural applications, and the unavoidable loss of specialization that accompanies it has no gain in improved knowledge. Paradoxically, the ineffectiveness of the top-down approach brings support to this point.
This argument has become weaker with the request for trans-national interoperability and the extensive use for cultural purposes of non-traditional data, as images, 3D models and movies. This has led to the need of incorporating such data into cultural databases, previously generally limited to text and reference to pictures, in general stored in external files or archived as blobs. Nowadays digital cultural objects have a complex nature and the organization and management of digital cultural archives (the so-called digital libraries) must reflect this complexity. A "librarian" approach is unsatisfactory, because it tends to ignore the peculiarity of each domain, as explained above for the top-down approach. In conclusion, digital libraries are no librarian's business at all.

To avoid this dilemma we have adopted a third approach, focusing on the following aspects:

- Determine which is the most effective way of storing the many facets of digital cultural objects; find the best "container" format and give it a sound theoretical basis.
- Establish guidelines for mapping existing data structures (and ontologies) on some established standard.
- Accompany each step with real examples based on extensive datasets, and provide tools for their management.

2. The features of digital cultural objects

Work on this issue is still in progress. At present, a preliminary list of features has been established and for each of them a standard has been chosen. The overall container will probably be MPEG-7 or METS.
Reconciliation of the different ontologies involved has been analyzed in some special, but rather general, cases, for example as far as 3D models are concerned [Niccolucci, D'Andrea 2006], using X3D as standard for the 3D part. Whether the geometry must be considered as a feature of a cultural object, or, vice versa, cultural information is to be considered as a set of attributes of a physical artefact, is in fact irrelevant, as both approaches have been shown to be viable [Niccolucci, D'Andrea 2006, Niccolucci 2007].

2. Mapping

The mapping process has been investigated for the case of archaeological data [D'Andrea et al 2006]. This is only apparently a simplification, both for the complexity of such data and the very large amount of legacy

archives, currently in use for research, management and other purposes. The institutions in charge of maintaining such archives, usually at a national or regional level, are reluctant to convert to a different system for several reasons, including national regulations that have not yet been superseded by some European norm. Mapping to some intermediate international standard appears therefore to be the only possibility to guarantee interoperability and maintain the semantic richness of the archives. Our choice for the common standard has been CIDOC-CRM.
Semantic interoperability may be really useful in the archaeological domain. For example, information concerning prehistoric "cultures" spanning over vast areas is usually spread through the archives of several modern states, and is stored not only in different languages, but also according to different methods reflecting different national regulations. Multilingualism appears also to be a key issue, and work is now addressing multilingual thesauri. As it is well known, problems here come not only from the translation, but also from the diverse history of Europe. For example, the term "Iron Age" has a different time span in European countries. As a consequence, the year 600 AD would belong to Middle Ages in Poland, to Iron Age in Norway, to Early Medieval in UK, and to (very late) Classical period for Romania. It would be Byzantine in Greece, Asia Minor, the Levant, and other parts of the Mediterranean region, but not everywhere (e.g. Spain). In Italy, it would depend from the region. So it is becoming clear that simple concepts as *who*, *when*, *where* are in fact all time and space dependent – and there is no such thing as a universal calendar or gazetteer.

3. Tools and applications

As yet, two tools have been provided, both still as prototypes.
The first one, AMA, is a help to create the mapping. It accepts as input the description of two ontologies e.g. in RDF (other ways are possible) and provides a graphical interface for establishing the correspondence. The mapping is then saved as a "template", i.e. an XML description. The tool creates also an XSL for the automatic conversion of the data, which need to be XML encoded to be processed in this way. An additional advantage of AMA is the possibility of editing an existing template in order to define a mapping which differs from an existing one only for some details. The AMA tool is being tested on an extensive number of datasets, provided by several national agencies in charge of archaeological data management. The AMA team is also developing a tool to manage poorly structured or unstructured text documents.
For more details on AMA, including the composition of the research team, visit the EPOCH web site www.epoch.eu: AMA is described in the chapter on NEWTONS, accessible via the tag "Research".
The second tool, MAD, is a data management system based on an XML native DBMS [Felicetti 2006]. The system can work on separate collections, stored on different servers, regardless of their structure, which is nonetheless very important to retrieve significant results. MAD can accept the output of the AMA conversion and thus offers a solution for practical cases when data conversion is performed.. Queries in MAD are based on XQUERY. Semantic (i.e. RDF-based) queries are presently being experimented. MAD has been used for a number of archaeological archives, and could be used for any archive where records consist of XML documents.
Both systems are distributed as Open Source and work on a number of platforms (Windows, Linux, Mac OS). They are available for download (or will be in short) from the above mentioned EPOCH web site.

Acknowledgement

The present research has been partially funded through the project EPOCH by the European Commission, under the Community's Sixth Framework Programme (contract no. 507382). However, this paper reflects only the authors' views and the European Community is not liable for any use that may be made of the information contained herein.

References

Staab S., Studer R., 2004. (eds.) *Handbook on Ontologies*. Berlin, Springer, vi.

Niccolucci F., D'Andrea A., 2006. An Ontology for 3D Cultural Objects. In M. Ioannides, D. Arnold, F. Niccolucci, K. Mania (eds.) *Proceedings of VAST 2006*. Aire-La-Ville, Eurographics, 203 – 210.

Niccolucci, F. 2007. Standards, credibility and philology of virtual models in archaeology. In B. Frisher (ed.) *Beyond Illustration*, in press.

[D'Andrea A., Marchese G., Zoppi T., 2006. Ontological Modelling for Archaeological Data. In M. Ioannides, D. Arnold, F. Niccolucci, K. Mania (eds.) *Proceedings of VAST 2006*. Aire-La-Ville, Eurographics, 211 – 218.

Felicetti A., 2006. MAD – Management of Archaeological Data. In M. Ioannides, D. Arnold, F. Niccolucci, K. Mania (eds.) *The e-volution of Information Communication Technology in Cultural Heritage – Project papers*. Budapest, Archaeolingua, 124 – 131.