The 8th International Symposium on Virtual Reality, Archaeology and Cultural Heritage
VAST (2007)
D. Arnold, F. Niccolucci, A. Chalmers (Editors)

CIDOC-CRM Spider: Stonehenge as an example of semantic data integration

G. Sugimoto[1] A. Felicetti[1] C. Perlingieri[2] and S. Hermon[1]

[1]Vast-lab, PIN, Servizi Didattici e Scientifici per l'Università di Firenze, Prato, Italy
[2]Università degli Studi di Napoli "l'Orientale", Naples, Italy

Abstract

Since archaeology in particular and humanities in general are interdisciplinary disciplines, there is an imperious need to enhance the accessibility and harmonise data integration, given their varied resource types (books, archives, scientific data, GIS, 3D models etc) and their different conceptual and technological structures and standards. A factor that further reduces accessibility and query performances is related to storing, such as physical location of data, language, and interface. The advent of Semantic Web technology represents an important advance in creating networks of archaeological knowledge based on various resources available on-line. While the valuable use of CIDOC-CRM (Comité International pour la Documentation -Conceptual Reference Model) ontology for specific CH (Cultural Heritage) domains (e.g. museums, libraries, etc) has been partially demonstrated, the interdisciplinary implementations are limited. In this article, we explore the potential of the CIDOC-CRM for a cross-domain implementation of CH data integration. We conducted tha mappings of different on-line resources related to Stonehenge to CIDOC-CRM, and evaluated them in the MAD database, a web-based application that manages natively XML-based datasets. The result ensures the validity of mapping mechanism and the semantic integration of CH data from different sources.

Categories and Subject Descriptors (according to ACM CCS): I.2.4 [Knowledge Representation Formalisms and Methods]: Semantic networks

1. Premise

The most immediate role of digital documentation in archaeology and CH in general, is as an aid to preservation, comprehension and exchange. But, how many on-line documentations produced by archaeologists and CH experts are designed as a proper information system? How many of them are effectively usable? And, how many are actually comparable? The use of different languages, heterogeneous systems and multiple schemas causes the total fragmentation of knowledge and complicates data transfer between institutions and data migration between systems. Besides, searching for information regarding CH on the Web is not an easy task, given its nature, related to the many research fields that CH covers; books describing various aspects of CH, numerical observations and measurements performed on the physical manifestations of CH (the artifacts themselves), art pieces depicting CH objects, or any other information pertaining to CH. As such, conducting a multi-disciplinary research on CH requires access to a wide array of sources organised, semantically and conceptually, in different ways. The need for a conceptual and technical "common ground", and a common data format that can assure clearer and unambiguous languages and tools comes as a consequence. While the need for such a cross-disciplinary query is self evident, its implementation on the Web is still in its infancy. There are as yet few CH information sources that use a global standard for their metadata, each preferring, and quite rightly in many cases, to maintain their own standards. Starting from the premise that no standard is better than anyother, and that there is a major need for data harmonisation, it was decided to experiment with an ontological approach. Although there are several possibilities for the ontological harmonisation approach to CH information, the CIDOC-CRM seems to be the best to "provide the "semantic glue" needed to mediate between different sources of CH information, such as that published by museums, libraries and archives" [CID06].

2. Concept and methodology of CIDOC-CRM-based integration

The CIDOC-CRM, initially developed as a detailed conceptual model for CH by CIDOC's members [CID06], provides a common framework for the exchange of CH information in a much wider perspective [CDG03]. It is a suitable standard for scientific documentation and for the requirements of the complexity, depth and quality of information concerning CH collections. The fact that it became an ISO standard (ISO 21127:2006) in 2006 encouraged us to adopt it as an intermediary format.

It is worth noting that mapping from CIDOC-CRM to other international standards in CH, have been conducted within the official CIDOC project [Doe98]. The standards include DCMES (Dublin Core Metadata Element Set), EAD (Encoded Archival Description), AMICO (Art Museum Image Consortium), MDA SPECTRUM, and OPENGIS (Figure 1). Discussion is currently underway for FRBR (Functional Requirements for Bibliographic Records). In contrast, practical implementations are still in an infant stage. The applications of CIDOC-CRM mostly concentrate on the museum sector [Sug06], given that the CIDOC-CMR was initially developed for museum documentation. This phenomenon is also echoed by the availability of relevant thesauri such as Art and Architecture Thesauri (AAT), Union List of Artist Names (ULAN), and Thesaurus of Geographic Names (TGN). The creation of standardised thesauri is one of the essential factors to implement the CH interoperability in practice. In other domains such as archives and field archaeology, CIDOC-CRM seems to be relatively unknown and the numbers of applications have been limited.

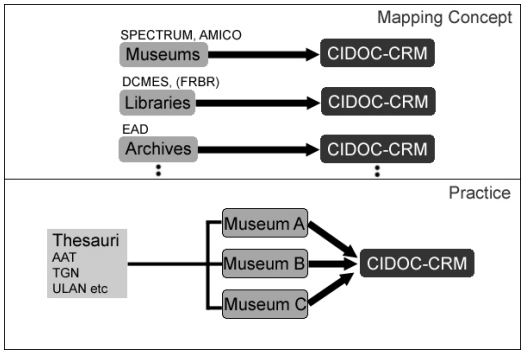

Figure 1: *Mapping concept and current practice*

In order to stimulate current practices of CH documentation, integration experiments are overdue. it was decided to adopt CIDOC-CRM as a core reference point for mapping from various standards of CH data representation. The ontology would maintain the original structure of the data and yet allow the performance of semantic searches. Figure 2 shows the landscape of CH resources and existing standards and the scope of our possible harmonisation. In the centre, there is CIDOC-CRM which connects to specific domains such as library, archives, computer science, museum, field archaeology, science, and sites and monuments. At the bottom, thesauri and classification system support the interoperability of CH data. The integration of CH "web" network is fulfilled by this "spider" mapping concept. This approach of CH data integration is similar to the Information Society project in Greece [BCD07] in the sense that it aims at establishing an integrated information access system for the CH, using CIDOC-CRM as a core ontology. Instead, this project deals with a real example and specialises in the mapping mechanism from various standards to one framework (e.g. CIDOC-CRM) in order to examine the data harmonisation in a semantic application already in operation. Thus, this research not only sheds light upon the semantic data integration in practice, but also assesses the advantages and disadvantages of CIDOC-CRM as a core ontology for CH.

The next chapter will explain MAD system that allows us to explore the potential of semantic database.

3. MAD (Managing Archaeological Data) system

MAD is a database originally designed as a web-based application to natively manage and query XML-based archaeological datasets encoded using CIDOC-CRM ontology through a complete set of web interfaces [Fel06]. The main advantage of the native approach resides in the portability and long term preservation of XML data, since the advanced features of MAD make them simple to share and reuse in different contexts [DFLP07]. MAD has been extended to become a semantic web tool able to fully support RDF for the CIDOC-CRM implementation and for the description of the explicit and implicit concepts and relations underlying the document structure [BL98].

RDF is a W3C standard language which is conceived as a base for the automatic "understanding" process and for the creation of machine-readable information distributed throughout the web. It is commonly used to represent metadata regarding digital artefacts and to provide a complex integration over different sources of information [RDF]. The W3C has also developed an RDF query language called SPARQL to query RDF graphs. It is built on the triple pattern, which consists of a subject, predicate and object. Its syntax is designed to combine information taken from one or more RDF graphs without the need for query authors to explicitly identify the mechanisms by which the graphs are combined. Operations such as joins are implicit rather than explicit in the language's syntax [SPA].

MAD can be considered an ideal environment for the semantic data integration since it is able to store and manage huge sets of RDF documents coming from different sources and to perform semantic queries on them using the powerful SPARQL framework [SW]. MAD is also able to query distributed RDF archives using the namespace mechanism to

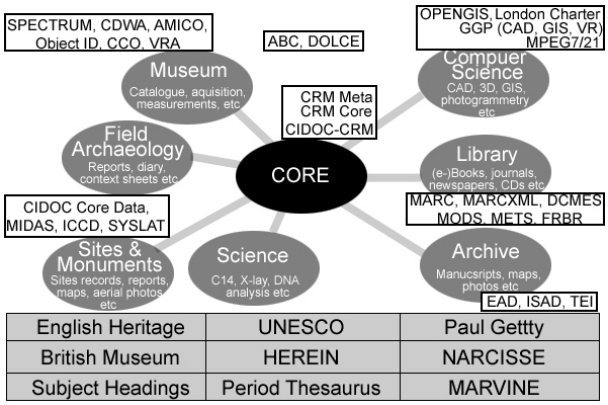

Figure 2: *Concept of CIDOC-CRM spider for CH data*

reach remotely stored sets of triples. The web interfaces also provide all the facilities to browse the complexity of RDF graphs with an advanced set of libraries to extract subgraphs and subelements according to given criteria. This semantic browser is based on the faceted browsing UI paradigm which gives the users the ability to find items based on more than one dimension using the facet, a particular metadata field considered important for the datasets involved in the browsing (Figure 3). Once the facets are selected, the browser extracts a list of relevant results from the model and the number of times each facet value occurs in the dataset. It is possible to add or remove restrictions in order to focus on more specific or more general slices of the model.

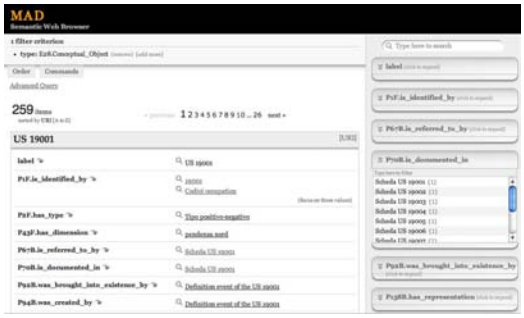

Figure 3: *Facet browsing interface in MAD*

4. Case Study: Stonehenge

4.1. Resource background

Stonehenge was chosen as a case study. Since the monument has been a centre of archaeological attention throughout history, it was presumed that various types of resources were created. It is, in fact, not that difficult to find online resources related to Stonehenge. However the limitation of HTML-based services prevents web users from collecting data without visiting every relevant Stonehenge website. This is the reality to be conquered. Although one particular resource provider from each domain (e.g. the Library of Congress from the library domain) had to be selected, it was the intension of this study to use standardised data (such as MARC and EAD) of each sector in order for the experimentation to be as applicable as possible in the near future. It is also obvious that this approach technically helps our experimentation.

4.2. Library data

British-based libraries were initially considered for data acquisition, but it was decided to use the digital collection of the Library of Congress in the United States [Lib07], because it offers flexible ways to present data. One can download a metadata of Stonehenge resource in simple text or MARC (Machine-Readable Cataloging) format. MARC is an internationally recognised standard (based on ISO 2709 bibliographic description) in the domain of library, thus, to pave a mapping path from MARC to CIDOC-CRM is a significant task. There are several paths to map MARC to CIDOC-CRM. In particular, MARCXML, developed by the Library of Congress, can be adopted as a pivotal format to other formats [Lib06b]. For example, it can be converted to DCMES (Dublin Core Metadata Element Set) and MODS (Metadata Object Description Standard). Figure 4 illustrates the existing paths from various standard formats to CIDOC-CRM with indications of some conversion tools. As far as conceptual mapping is concerned, the formal mapping from DCMES to CIODC-CRM was published in the website [Doe00]. The XML schemas of DCMES are also available from [Dub07a].

The online catalogue returns 356 hits for the query of Stonehenge as a keyword, which are mainly books and cartographic resources, but also digital photographs of Stone-

henge accessible on the web. In this case study, MARC was converted to MARCXML with MARCXML toolkit [Lib06b], and then an XSLT was applied to transform it to the RDF of DCMES, followed by another transformation to the RDF of CIDOC-CRM. The experimentation path seems to be redundant, but this is one of the few ways to obtain actual data in preferred formats. Although depending on one path, this conversion methodology is promising for the future integration of any bibliographical formats.

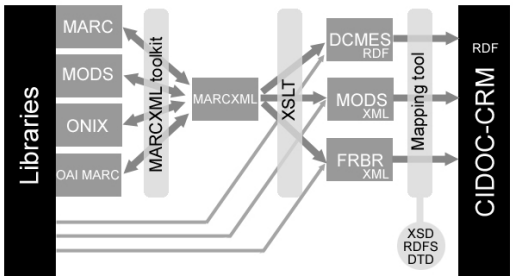

Figure 4: *Mapping path from MARC to DCMES and to CIDOC-CRM*

Since the formal mapping from DCMES to CIDOC-CRM was carried out in 2001, using the old version of CIDOC-CRM (version 2.2), it was necessary to consult its latest version for updates. [Doe00] pointed out that it was not possible to create a complete conceptual mapping due to the implicitness of DCMES definition, persuading the practitioners to discuss the valid mapping. Indeed, the primary goal of this experimentation is not to create a perfect mapping. Having said so, it may be interesting how different ways of conceptual mapping affect the data integration at the end. The agenda for the "official" mapping will be dealt with in the debates among specialists, therefore, it was decided to follow Doerr's philosophy. It was also possible to execute the experiment with FRBR (with a toolkit [Net04]), however the harmonisation of FRBR and CIDOC-CRM was not thoroughly completed (FRBRoo version 0.8.1 [Int07]), letting us set it aside for a while.

We emphasised the process of mapping and available tools in order to prove that the automation of data conversion is feasible. Thus, it is expected that server technology, in the future, will implement the conversion and mapping on the fly (Figure 5).

4.3. Archive data

Archives Hub offers a variety of services for archival data in the UK. It publishes archive guidelines and an online database with 19,827 descriptions of archives held in more than 150 UK universities and colleges [Arc7b]. It encompasses, for instance, a scrapbook relating to Arthur Hadrian Allcroft, created around 1953, consisting of several articles on Stonehenge. The Llewellyn Treacher Photograph

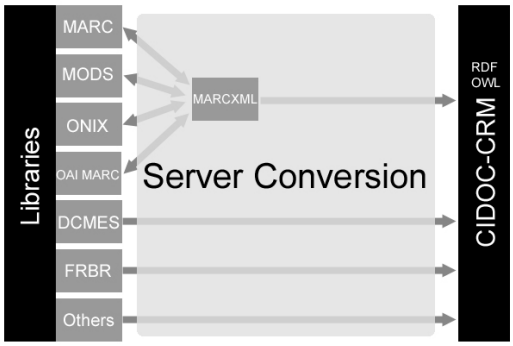

Figure 5: *Future scope of library path*

Collection of 1890s - 1930s contains archaeological subjects from across England and, among others, Stonehenge. The database demonstrates a good visualization of datasets, based on the EAD standard. Despite the value of the famous A2A Database [A2A07], it was not used for this project for the same reason as British libraries were discarded.

EAD is an international standard for encoding archival finding aids using XML [EAD07]. An advantage of the use of EAD is the accessibility to some useful tools on the web. For example, Archives Hub maintains EAD 2002 Online Template [Arc7a]. In addition, EAD Schemas are available in DTD and XMLSchema [EAD07], as well as tools for EAD version update (1.0 to 2002). Needless to say, [Doe01] conducted the conceptual mapping from EAD to CIDOC-CRM.

Stonehenge-related data from the Archives Hub database was inserted into the template, creating an EAD encoded XML. Consequently, the XML, together with the XMLSchema, was mapped to CIDOC-CRM RDF. Again, the problem lay in the obsolescence of versions. In this case, Doerr used older versiosn of both EAD (1.0) and CIDOC-CRM (3.0). Although the gap does not significantly affect the mapping, it is necessary for the CIDOC-CRM community to renew the mapping publications. Because of time constraints and the lack of official mapping, other standards such as ISAD(G) (General International Standard Archival Description) [Int00] and TEI's (Text Encoding Initiative) Guidelines for Electronic Text Encoding and Interchange (TEI P4) [Tex07] were not explored. As the need for CH integration grows, CH records await the conceptualisation of these standards (accordingly, the mappings to CIDOC-CRM).

Unlike the library sector, the mapping path for archives is simpler due to the limited amounts of standards (Figure 6). Metadata crosswalk in libraries and archives has been scholarly discussed by [Day02], [Net01], and [Sat05], thus this kind of work will resolve the complex picture of mapping mechanisms.

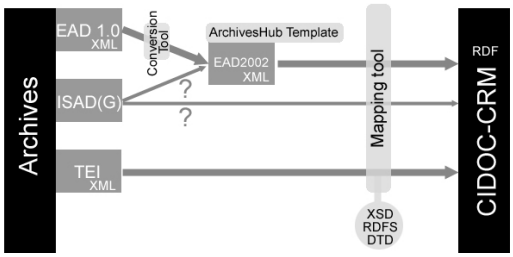

Figure 6: *Mapping path from EAD to CIDOC-CRM*

4.4. Sites and monuments data

Contrary to the sphere of library and archives, archaeological standards are often not accepted in a wider community. However some useful standards are recognised in England. MIDAS, the Monument Inventory Data Standard [Eng98], is an agreed statement of best practice for the compilation of inventories of monuments. The wordlists and thesauri recommended to support MIDAS are brought together as INSCRIPTION [FIS07]. This is a set of standard wordlists that define appropriate terms recommended by FISH for use in compiling inventories. The MIDAS Data Standard is divided into two parts: discussion on the information scheme, which is a cluster of facts constituting the information required to record a particular subject; and definition of each unit of information which are the fundamental facts of interest.

Significant content and guidelines related to monuments and sites are found in the Archaeology Data Service (ADS). The ADS Collections Management provides a standard framework for the administration, description and storage of CH digital data. With no available MIDAS datasets, MIDAS-style data related to the Stonehenge area was chosen from the ADS Online Catalogue [Arc07]. Although MIDAS schemas [FIS07] and the official mappings are available [CID07], which can be applied for ADS data to some extent, the fundamental difference between archaeological standards and datasets (and its contexts) required for heterogeneous data to be mapped directly to the CIDOC-CRM, taking MIDAS into accout.

4.5. Finds (museum oriented) data

MDA is a leading organisation on documentation and information management for museums collections in the UK [MDA], funded by the MLA (Museums, Libraries and Archives Council). It develops and maintains professional standards in trust on behalf of the museums sector. SPECTRUM is an industry standard for documentation practice developed by MDA. It consists of 21 documentation procedures, 8 of which are classed as 'primary'. These primary procedures are embedded into the Museum Accreditation Scheme as a national minimum standard for managing information about collections.

Noteworthy is the ADS contribution to the finds and collections documentation practices. The Stonehenge 20th Century Excavations Databases is based on the Wessex Archaeology Stonehenge Archive [Arc5b]. The archive contains written material (primary, secondary, original, and copies), drawings, photographs, and finds. These relate to a large number of excavations and non-archaeological interventions undertaken in and around the Stonehenge Triangle between 1901 and 1994. Two tables of Finds and Graphics were selected from the archive for the mapping, because they represent common datasets used in typical archaeology. Similar to the case of monuments, the direct mapping was conducted according to [Cro99].

4.6. Chemical analysis data (Radio carbon)

In the domain of scientific resources ADS provides a useful database of radiocarbon analysis in Britain and Ireland. It originated in a printed index compiled by the Council for British Archaeology in 1971. With further elaboration and struggle, it became available on the web as a database holding about 9000 records [Arc5a]. Without a doubt, some datasets of Stonehenge can be related to other CH data. For instance, artifacts and bibliographical resources used for date determination will be considerably valuable.

There are almost no standards for the description of chemical analysis in archaeology, therefore, we simply mapped the database fields to the CIDOC-CRM. The mapping was manually undertaken (Figure 7). [Arc5a] argued that terminological control/thesaurus was not applied for the database, therefore, the maintenance and interpretation of data may be difficult. Although there is a short documentation of database fields, it was not easy to create a semantic model with CIDOC-CRM. For example, the semantics of some data fields are not explicit. Another problem occurred when data fields include multiple values (e.g coordinate field has northing and easting together). Databases are often not well designed for future re-use. These problems are universal for most of the conventional database mappings. Nevertheless, as far as a CIDOC-CRM model was created and data is transferred into the MAD database, the completeness of the mapping is not very important for the experiment. In the future, it is expected that software such as the AMA tool [EPO] will be able to semi-automatically create a mapping with user-friendly interface.

5. Conclusions

5.1. Current state and problems

All data were successfully mapped and imported into the MAD database. Thanks to RDF, it is now possible for the users to explore the semantic network of various Stonehenge data encoded in CIDOC-CRM. They can "jump" across the data from different domains without knowing the resources themselves. This seamless data integration is ex-

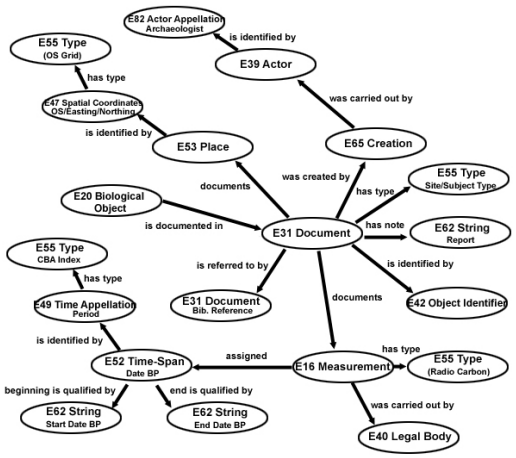

Figure 7: *Radiocarbon Database Mapping to CIDOC-CRM*

tremely powerful and overcome some problems of the current practice of CH data on the web. The facet browser allows users to effectively navigate in the complex data network according to their criteria. Figure 8 shows the current picture and a blueprint of our CH data harmonisation.

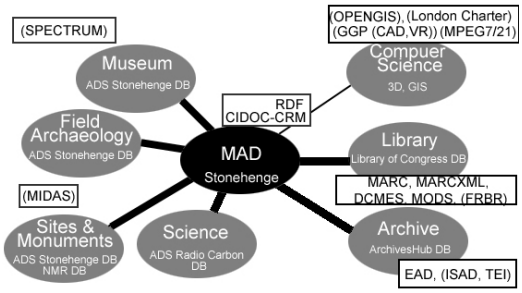

Figure 8: *Current mapping paths to MAD*

There are several problems regarding the mappings. It seems that the Library of Congress does not (currently) intend to provide a tool for loss-less conversion. The XSLT file for MARCXML to DCMES RDF causes the loss of original MARC21 data. This result probably means that DCMES cannot capture all information from MARC21, if so, other metadata schemas and ontologies such as FRBR have to be deployed. Otherwise it is necessary to fully conceptualise MARC for CIDOC-CRM conversion. In addition, the tool created a simple Dublin Core (so-called Hedgehog model), not a qualified one. If more complex semantics are needed, this transformation programme should be re-written based on the guidelines [Dub07b]. Moreover, the Dublin Core description created by XSLT was divided into several dc:description. This is reflected by the MARC21 format which encodes the description data in several paragraphs. This form of conversion is, obviously, not satisfactory for this case study. Thus, properly written XSLT is needed in order to preserve all information of MARC21. The technological infancy also applies to ontology tools in general. Tools such as Protégé [Sta07] and AMA are useful in their own right. However, the lack of functions necessary to the case study led to manual mapping and conversion. For instance, the AMA tool is capable of semi-automatically mapping an XMLSchema (from existing standards schemas or conventional databases) to CIDOC-CRM . The software produces XSLT to convert XML data into simple CIDOC-CRM RDF. However, it is not currently designed to instantiate new entities to represent an implicit data structure. This is problematic because CIDOC-CRM demands the clarification and explication of data fields [DL07], which often requires the creation of event entities.

Although CIDOC-CRM may declare that official mappings to famous existing standards are completed, they have to be updated (most of them are subject to old versions). In addition, domain experts have not discussed fully about the best mapping practice. In particular, most websites of CH standards do not refer to CIDOC-CRM. Similarly, many archaeologists do not know CIDOC-CRM [Sug06], [Sug07a], and [Sug07b]. In contrast, crosswalk is often a burgeoning issue of library and archives standards [Day02]. The research of this paper is, therefore, expected to be a catalyst for its interdisciplinary dissemination.

5.2. Future scope

As described above, the integration has not been executed in some domains; more precisely computing data.

Few have, hitherto, produced academic papers on the practical implementation of 3D application using CIDOC-CRM. In this sense, the integration is challenging yet interesting. As far as we will be able to obtain 3D data of Stonehenge, we are keen to test it by semantic-based approach.

Spatial management functions of MAD will be implemented to integrate spatial information and to create and distribute rich geospatial relationships across the web using the Geographic Markup Language (GML) [GML]. Our system will allow the creation of semantic web meaningful data combining spatial and non-spatial integrated data and using ontologies. It is also noted that the comparison between OPENGIS Abstract Specification and CIDOC-CRM was conceptually completed [Doe], therefore, OPEN GIS compliant data encoded in GML could be mapped to CIDOC-CRM in the future.

In parallel to more integration of CH data, theoritical and technological development will be required.

The preliminary goal of our integration was to extract

CH data and import it in CIDOC-CRM RDF format into a real application (e.g. MAD). Therefore, during the mapping process, some details of CH data were omitted. In other words, the aim of this article was not to discuss how to extend CIDOC-CRM to map reasonably, when the extension is needed. The next step of this research will be to examine how the extension behaves and influences the whole semantic model in reality.

We did not include any thesauri and classification systems in MAD partly due to their inaccessibility and insufficiency. Indeed, there are few internationally recognised thesauri available for archaeology [Sug06](and the interdisciplinary study of Stonehenge). However, depending on the availability of vocabulary system such as the Library of Congress Subject Heading [Lib06a] and thesauri of English Heritage [Eng99], the terminological inclusion is urgently needed.

The final step of the MAD project will be the creation of a web spider able to harvest all the CIDOC-CRM encoded datasets present on the web and to index them in an integrated system. It will hopefully become a semantic web CIDOC-CRM search engine and browser for users seeking cultural heritage information.

In order to avoid data loss, proper XSLT for data conversion should be created. Although conceptual mappings should always be done manually, it is also hoped that server-side conversions will bypass the unnecessary manual process.

We believe that our mapping mechanism finds the right way of integrating CH data. Our experimentation has been successful so far, but there are a few domains that we could not explore yet because of the lack of time and data. 3D and GIS data will be evaluated and integrated with other relevant data, resulting in a more comprehensive network of semantically connected Stonehenge information. MAD will be able to encapsulate the networked CH information and to ensure the potential of CIDOC-CRM. Consequently, it is hoped that this paper inspires and evokes the Semantic Web.

Acknowledgement

This research has been conducted under two programs funded by European Commission including EPOCH (The European Research Network of Excellence in Open Cultural Heritage)(IST-2002-507382) and CHIRON (Cultural Heritage Informatics Research Oriented Network) fellowship (MEST-CT-2004-514539). We would like to thank Jacqueline Radebaugh from the Library of Congress, Network Development & MARC Standards Office who gave us a personal support for the use of MARCXML and FRBR display toolkit. In addition, our colleague and a computer programmer of AMA tool, Marco Crescioli was involved for the support of its software. Moreover, we acknowledge the following data providers: ADS (Archaeology Data Service), Wessex Archaeology, Council for British Archaeology, the Library of Congress, and Archives Hub.

References

[A2A07] A2A: Access to Archives. http://www.a2a.org.uk/.

[Arc5a] ARCHAEOLOGY DATA SERVICE: Archaeological Site Index to Radiocarbon Dates from Great Britain and Ireland, Council for British Archaeology 2000. http://ads.ahds.ac.uk/catalogue/resources.html?c14_cba.

[Arc5b] ARCHAEOLOGY DATA SERVICE: Stonehenge 20th Century Excavations Databases, Wessex Archaeology, 2005. http://ads.ahds.ac.uk/catalogue/resources.html?stonehenge_eh_2005.

[Arc7a] ARCHIVES HUBS: EAD 2002 Online Template. http://www.archiveshub.ac.uk/eadform2002.html.

[Arc7b] ARCHIVES HUBS: Introduction to the Archives Hub. http://www.archiveshub.ac.uk/introduction.shtml.

[Arc07] ARCHAEOLOGY DATA SERVICE: Archsearch. http://ads.ahds.ac.uk/catalogue/index.cfm.

[BCD07] BEKIARI C., CONSTANTOPOULOS P., DOERR M.: Information design for cultural documentation. *DELOS Conference 2007 Working Notes* (2007), 221–230.

[BL98] BERNERS-LEE T.: Semantic Web Roadmap, an attempt to give a high-level plan of the architecture of the Semantic WWW, September 1998. http://www.w3.org/DesignIssues/Semantic.html.

[CDG03] CROFTS N., DOERR M., GILL T.: The CIDOC-CRM Conceptual Reference Model: A standard for communicating cultural contents. *Cultivate Interactive 9* (Feb. 2003). http://www.cultivate-int.org/issue9/chios/.

[CID06] The CIDOC-CRM. http://cidoc.ics.forth.gr/.

[CID07] CIDOC CRM SPECIAL INTEREST GROUP: CIDOC CRM Mappings, Specializations and Data Examples. http://cidoc.ics.forth.gr/crm_mappings.html.

[Cro99] CROFTS N.: MDA Spectrum CIDOC CRM mapping. http://cidoc.ics.forth.gr/docs/MDA\%20Spectrum_CIDOC_CRM_mapping.pdf.

[Day02] DAY M.: Metadata Mapping between metadata formats. http://www.ukoln.ac.uk/metadata/interoperability/.

[DFLP07] D'ANDREA A., FELICETTI A., LORENZINI M., PERLINGIERI C.: Spatial and non-spatial archaeological data integration using MAD. *CAA2007* (2007). 2-6 April 2007, Berlin.

[DL07] DOERR M., LEBOEUF P.: Modelling Intellectual Processes: The FRBR - CRM Harmonization. *DELOS Conference 2007 Working Notes* (2007), 133–137.

[Doe] DOERR M.: A comparison of the OpenGIS Abstract Specification with the CIDOC CRM 3.2 Draft.

[Doe98] DOERR M.: Correlation Test Project. http://cidoc.ics.forth.gr/correlation_test.html.

[Doe00] DOERR M.: Mapping of the Dublin Core Metadata Element set to the CIDOC-CRM. http://cidoc.ics.forth.gr/docs/dc_to_crm_mapping.pdf.

[Doe01] DOERR M.: Mapping of the Encoded Archival Description DTD Element Set to the CIDOC-CRM. http://cidoc.ics.forth.gr/docs/ead.pdf.

[Dub07a] DUBLIN CORE METADATA INITIATIVE: DCMI Schemas. http://dublincore.org/schemas/.

[Dub07b] DUBLIN CORE METADATA INITIATIVE: Encoding Guidelines. http://dublincore.org/resources/expressions/.

[EAD07] Encoded Archival Description Version 2002 Official Site. http://www.loc.gov/ead/.

[Eng98] ENGLISH HERITAGE DATA STANDARDS UNIT NATIONAL MONUMENTS RECORD CENTRE: MIDAS A Manual and Data Standard for Monument Inventories. http://www.english-heritage.org.uk/upload/pdf/MIDAS3rdReprint.pdf.

[Eng99] ENGLISH HERITAGE: National Monuments Record Thesauri. http://thesaurus.english-heritage.org.uk/.

[EPO] EPOCH: EUROPEAN RESEARCH NETWORK ON EXCELLENCE IN PROCESSING OPEN CULTURAL HERITAGE: AMA -Archive Mapper for Archaeology. http://www.epoch-net.org/index.php?option=com_content&task=view&id=74&Itemid=120.

[Fel06] FELICETTI A.: MAD: Managing Archaeological Data. *7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST2006)* (2006), 124–131. http://public-repository.epoch-net.org/publications/VAST2006/project1.pdf.

[FIS07] FISH: FISH: Forum on Information Standards in Heritage. http://www.fish-forum.info/.

[GML] GML: the Geography Markup Language. http://www.opengis.net/gml/.

[Int00] INTERNATIONAL COUNCIL ON ARCHIVES: ISAD(G): General International Standard Archival Description, Second edition. http://www.ica.org/en/node/30000.

[Int07] INTERNATIONAL WORKING GROUP ON FRBR AND CIDOC CRM HARMONISATION: FRBR object-oriented definition and mapping to FRBRER (version 0.8.1). http://cidoc.ics.forth.gr/docs/frbr_oo/frbr_docs/FRBR_oo_V0.8.1c.pdf.

[Lib06a] LIBRARY OF CONGRESS: Library of Congress Authorities. http://authorities.loc.gov/.

[Lib06b] LIBRARY OF CONGRESS: MARCXML MARC21 XML Schema. http://www.loc.gov/standards/marcxml/.

[Lib07] LIBRARY OF CONGRESS: The Library of Congress Online Catalogs. http://catalog.loc.gov/.

[MDA] MDA: MDA. http://www.mda.org.uk/.

[Net01] NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE LIBRARY OF CONGRESS: Dublin Core/MARC/GILS Crosswalk. http://www.loc.gov/marc/dccross.html.

[Net04] NETWORK DEVELOPMENT AND MARC STANDARDS OFFICE LIBRARY OF CONGRESS: Functional Analysis of the MARC 21 Bibliographic and Holdings Formats FRBR Display Tool Version 2.0. http://www.loc.gov/marc/marc-functional-analysis/tool.html.

[RDF] W3C Semantic Web Activity, Resource Descpription Framework. http://www.w3.org/RDF/.

[Sat05] SATO Y.: Crosswalk of MARC and metadata. *Current Awareness 283* (2005), 48–62. http://www.dap.ndl.go.jp/ca/modules/ca/item.php?itemid=984&keywords=No.283+2005.03.20.

[SPA] SPARQL Query Language for RDF. http://www.w3.org/TR/rdf-sparql-query/.

[Sta07] STANFORD MEDICAL INFORMATICS: Protege. http://protege.stanford.edu/.

[Sug06] SUGIMOTO G.: Beauty or Beast?: A Review of the CIDOC-CRM Applications and Thesauri in Archaeology. *7th International Symposium on Virtual Reality, Archaeology and Cultural Heritage (VAST2006)* (2006), 203–208. http://public-repository.epoch-net.org/publications/VAST2006/short2.pdf.

[Sug07a] SUGIMOTO G.: A Digital Data Survey for Japanese Archaeology. http://www.chiron-training.org/go_sugimoto/digital_survey/.

[Sug07b] SUGIMOTO G.: Face to Face - A comparative study on the use of digital data in archaeology: UK versus Japan. *Communicating Cultural Heritage in the 21st Century The CHIRON project and its research opportunities* (2007), 95–122. http://public-repository.epoch-net.org/publications/CHIRON/communicating_CH.pdf.

[SW] W3C Semantic Web Activity. http://www.w3.org/2001/sw/.

[Tex07] TEXT ENCODING INITIATIVE: TEI: Yesterday's information tomorrow. http://www.tei-c.org/.