

Harmonization of CIDOC CRM ontology in the context of archives, libraries and museums



Laís Barbudo Carrasco
Universidade Estadual Paulista
laiscarrasco@hotmail.com



Silvana Ap. Borsetti Gregório Vidotti
Universidade Estadual Paulista
vidotti@marilia.unesp.br



Phil Manfred Thaller
Universität zu Köln
manfred.thaller@uni-koeln.de

Introduction

By the fact that Digital Cultural Heritage Repositories, such as Libraries, Archives and Museums, use different metadata standards to describe their information resources, the metadata harmonization from the cultural heritage field is a challenge, because the data models are more designed on the community requirements than on requirements of cross-community interoperability.

In order to integrate information from heterogeneous sources, ontologies as semantic technologies, are already being used and CIDOC Conceptual Reference Model (CRM) is a very prominent ontology used for such purposes.

The **CIDOC CRM** is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be

mapped to. [...] In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. (CIDOC CRM)

Metadata are forms used to describe, manage, catalog, and classify documents (information objects). In Libraries, they are traditionally used to describe books, in other words, bibliographic data. In addition, in museums, metadata describes artifacts, paintings, and sculptures. On other hand, in Archive, the metadata are used to describe finding aids, charters, and official documents. Therewith, this treated information can be effectively retrieval. The cultural memory institutions aim to treat their information in order to obtain access and retrieval.

A way to make metadata interoperate is mapping them into an ontology. CIDOC Conceptual Reference Model (CIDOC CRM) is an ontology from the cultural heritage domain, which has been created as a tool for information integration.

Cultural heritage institutions such as museums, archives, or libraries are confronting a crescent necessity to integrate their system. For this reason, to make cultural resources accessible, it is necessary the use of rich metadata structures, capable to cover the variety of material held in these memory institutions.

In this context, ontologies are used as an important tool for achieving information integration, in other words, metadata can be semantically mapped and integrated into an ontology, which has the competence not only to conceptualize specific domains, but also to express their semantics.

The CIDOC CRM arised from the CIDOC Documentation Standards Group in the International Committee for Documentation of the International Council of Museums and CIDOC CRM was accepted as the ISO 21127 in 2006. The purpose of the CRM model is to provide a common language for heterogeneous information systems, and to permit their integration, despite possible semantic and structural incompatibilities. In that way, cultural heritage information can be exchanged and retrieved; and cultural heritage institutions can make their information systems interoperable without having to compromise their specific needs or the current level of precision of their data. (CIDOC CRM)

In summary, CIDOC CRM is a high-level ontology to permit information integration for cultural heritage data and their correlation with museum, library and archive information, which is easily converted to other machine readable formats such

as RDF and XML. Possibly the most ambitious application of the CRM is in the development of integrated query tools, mediation systems and data warehouses. At present, much of the information stored in library catalogues, archival finding aids and museum collection management systems remains isolated. Different information resources normally need to be queried individually, and cross-system links are rare. The ability to combine and integrate information from multiple sources has the potential to add significant value to existing data - facilitating research and enhancing the quality of the user's experience. (ARTUR; CROFTS; LE BOEUF; 2002).

When a user makes a search in a data base he or she doesn't look for a library resource or an archival resource or a museum resource, the user searches for information. In this way, when the user insert a query into the search engine the user does not take into consideration the nature of the resource, that's why it's very important to develop a harmonization of CIDOC CRM ontology in the context of archives, libraries and museums, because it will be able to understand how the CRM entities can be interpreted in the context of archives, libraries and museums and then it will be possible to potentialize the information retrieval.

State of art

Library, Archive, and Museums are repositories of collective memory and they have some functions, for example, the maintenance and conservation of collections; the exposure; and provision of the means for the objects or information retrieval.

Storing, cataloging, classifying and exhibiting objects are part of the relationship process of the man with the world, and they are operations that do not arise with the museum, nor with library, and nor with the archive, but with collections.

Library, Archive and Museum – all the three institutions preserve the memory and organize the access to information. On other hand, according to the institution, the role assigned to the document is different. Traditionally, libraries saves book, archives saves legal documents and museum masterpieces. Thus, each typological material has a way to be treated. In this way, it's needed to differentiate what a term refers to each institution. According to the author Alvares (s.d.) archive, library and museum have the following meanings:

- **Archive:** Ordained storage of documents, created by a person or institution in the course of its activity, and preserved to achieve their goals, aiming the utility that

they may offer in the future, therefore, the archival information is information accumulated by institution to prove or inform, as the institution pursues its purposes/ends, with whom it relates, which rights and duties it represents.

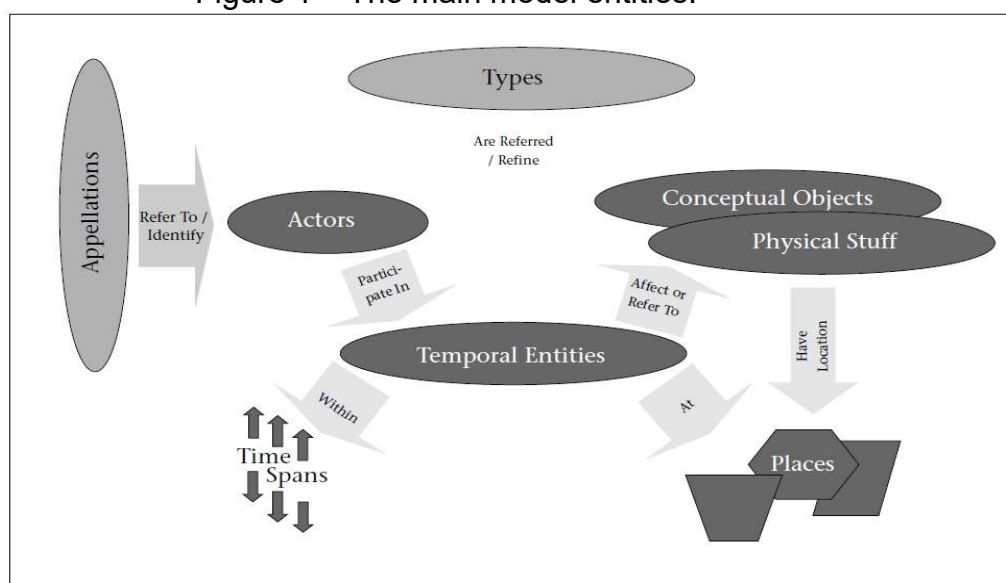
- **Library:** Collection of books, periodicals, audiovisual documents, among others, selected based on their usefulness and arranged to provide access to materials and information, therefore, the organization of the documents in the library reflects the organization given to the area of knowledge (or the subject specifically covered by the documentation center).

- **Museum:** It is a permanent institution, nonprofit organization, in service of society and of its development, open to the public which acquires, conserves, researches, disseminates and exposes the material evidence of people and their environment, for education and delight of society; therefore, the museum, as the instance of the memory representation and social space in which information is considered cultural input, is a fertile ground for the development of studies and actions related to information.

Result

The CRM scope can be defined as all the necessary information for the scientific documentation of cultural heritage collections, in order to enable a broad exchange of information from the area and the integration of heterogeneous sources.

Figure 1 – The main model entities.



Source: Cidoc CRM (DOERR, 2003).

CIDOC CRM contains classes and logical groups of properties. These groups have to do with the notions of participation, structure, location, evaluation and identification, purpose, motivation, usage, and so on. These properties have put temporal entities and with them, the events in a central location (LIMA, 2008).

Table 1 – CIDOC CRM Harmonization in the context of Libraries, Archives and Museums.

CIDOC CRM Entity	Library	Archive	Museum
E1 CRM Entity	CRM Entity about bibliographic material .	CRM Entity about archival material .	CRM Entity about museological material .
E18 Physical Stuff	Literature ("Book") about the Physical Stuff.	Documentation ("Charter") about the Physical Stuff.	Museum object
E28 Conceptual Object	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of bibliographic material.	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of archival material.	Texts, maps, photos, music, sounds, fairy tales, signs, patterns, symbols, plans, rights, and rules, paper, electronic signals, marks, audio media, paintings, photos, human memory, etc; about circumstances of creation and historical implications of museum material.
E39 Actor	Author.	Institution.	Artist.
E41 Appellation	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context referred to the Library and the bibliographic material.	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context referred to the Archive and the archival material.	All names in the proper sense. Codes or words, meaningless or meaningful, in the script of some group or encoding of an electronic system, used solely to identify a specific instance of some category within a certain context referred to the Museum and museological material.
E50 Date	Date of publication.	Date of fund.	Date of creation.
E52 Time-Span	Time-Span of the "Book, Collection".	Time-Span of the "Charter, Fund".	Time-Span of the "Object, Collection".

E53 Place	Place of publication or storage (Library).	Place of provenance or storage (Archive).	Place of creation or storage (Museum).
E55 Type	Typological distinctions of "Book, Collection".	Typological distinctions of "Charter, Fund".	Typological distinctions of "Object, Collection".

Source: Authors.

In this section, it was built from the CIDOC CRM vocabulary a correspondence of terms handled into the cultural heritage institutions – Library, Archive and Museum; because CIDOC CRM entities can have different meanings facing the diverse types of material handled in the three sisters institutions.

Final considerations

We are facing an explosion of varieties of cultural heritage metadata schemas. Cultural Heritage - as the safeguard of the memory - has heterogeneous materials, and for this reason they are safeguarded, protected and conserved in the Libraries, Archives and Museum. In order to treat these digital objects, metadata standards are used. On other hand, how can these three institutions exchange data?

According to the literature, there are many XML metadata mapping to the CIDOC CRM ontology efforts, since this ontology is considered one of the most appropriate models in integration architectures.

In this way, metadata can be mapped into an ontology to provide interoperability of its data and also to achieve information integration. When the different kind of metadata are mapped into an ontology the system can interoperate and the information access is higher as well as their information retrieval.

“An ontology is a specification of a conceptualization” (GRUBER, 1993). More specially, the CIDOC CRM ontology is the specification of the Cultural Heritage conceptualization. CIDOC CRM has an abstract hierarchy.

In summary, the metadata from the three sisters institutions – Library, Archive and Museum – can be mapped into the ontology CIDOC CRM to provide interoperability and information integration of these heterogeneous data from the digital cultural heritage systems.

References

ALVARES, Lillian. (s.d.) **Informação em Arquivos, Bibliotecas e Museus**. (Slides) - Faculdade de Ciência da Informação - Universidade de Brasília. Available at: <http://www.alvarestech.com/lillian/Analise/Modulo1/Aula13ABMII.pdf>. Access: May, 2012.

ARTUR, O.; CROFTS, N.; LE BOEUF, P. Elag presentation ontologies. Semantic web and libraries. In: LIBRARY SYSTEMS SEMINAR, 26., 2002, Rome. **Proceedings**... Rome: European Library Automation Group, 2002. Disponível em: <<http://www.ifnet.it/elag2002/papers/pap9.html>>. Acesso em: 18 abr. 2012.

CIDOC CRM. Disponível em:<<http://www.cidoc-crm.org/>>. Acesso em: 20 fev. 2014.

CROFTS et al. (Ed.). **Definition of the Cidoc Conceptual Reference Model**. Cidoc, Icom, 2005. Disponível em: <http://www.cidoc-crm.org/docs/cidoc_crm_version_4.2.pdf>. Acesso em: 28 jan. 2015

CROFTS, N. **Museum informatics: the challenge of integration**. 2004. Tese (Doutorado em Ciências Econômicas e Sociais) – Faculté des Sciences Économiques et sociales, Universidade de Genebra, Genebra, 2004.

DOERR M., The Cidoc CRM: an ontological approach to semantic interoperability of metadata. **AI Magazine**, v. 24, n. 3 p. 75-92, 2003.

DOERR, M.; STEAD, S. **The Cidoc CRM: standard for the integration of cultural information**. Glasgow, Jan. 29, 2008. Disponível em: <http://www.uniurb.it/sbc/ist_bal/seminario/carlo%20meghini/CIDOC%20CRM%20introduction.pdf>. Acesso em: 20 maio 2012.

DOERR et al. The Europeana Data Model (EDM). In: IFLA GENERAL CONFERENCE AND ASSEMBLY, 76., 2010, Gothenburg. **Proceedings**... Gothenburg: Ifla, 2010. Disponível em: <<http://conference.ifla.org/past-wlic/2010/149-doerr-en.pdf>>. Acesso em: 18 fev. 2012.

GRUBER, T. R. **What is an ontology?** Stanford: Stanford University, 2001. Disponível em: <<http://www-ksl.stanford.edu/kst/what-is-an-ontology.html>>. Acesso em: 27 fev. 2012.

GRUBER, T. R. Towards principles of the design of ontologies used for knowledge sharing. In: INTERNATIONAL JOURNAL HUMAN-COMPUTER STUDIES, SUBSTANTIAL REVISION OF PAPER PRESENTED AT THE INTERNATIONAL WORKSHOP ON FORMAL ONTOLOGY, Padova, 1993. **Proceedings**... Padova: Standford University, 1993. Disponível em: <<http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.89.5775>>. Acesso em: 20 jan. 2014.

LE BOEUF, P., Using an ontology-driven system to integrate museum information and library information. In: SYMPOSIUM ON DIGITAL SEMANTIC CONTENT ACROSS CULTURES, 2006, Paris. **Proceedings**... Paris: Louvre, 2006.

LIMA, J. A. O. **Modelo genérico de relacionamentos na organização da informação legislativa e jurídica**. 2008. Tese (Doutorado em Ciência da Informação) – Universidade de Brasília, Brasília, DF, 2008.