ICHIM
PARIS 21-23 SEPT. 05
www.ichim.org

Digital Culture & Heritage
Patrimoine & Culture Numérique

Bibliothèque nationale de France, PARIS
Sept. 21st - 23rd, 2005
21 - 23 septembre 2005

INTEGRATING MUSEUM & BIBLIOGRAPHIC INFORMATION : THE SCULPTEUR PROJECT

**Patrick Le Bœuf, Christian Lahanier,
Geneviève Aitken,
Patrick Sinclair, Paul Lewis, Kirk Martinez**

**Published with the sponsorship of the
French Ministry of Culture and Communication**

Actes publiés avec le soutien de la Mission de la Recherche et de la Technologie du Ministère de la Culture et de la Communication, France

Interprétation simultanée du colloque et traduction des actes réalisées avec le soutien de l'Agence Intergouvernementale de la Francophonie

Abstract (EN)

"Infodiversity", that is different ways of structuring information and of defining the content of information, exists among museums, and between museums and libraries. It is neither desirable nor feasible to standardise the practices of all kinds of cultural heritage institutions, as each of them has specific objectives and needs for specific applications. Standardisation and information integration can be achieved at a higher level, through mediation systems that interpret the semantics of each distinct data structure. This process is made possible by semantic reference models, such as the CIDOC CRM in the case of the European-funded SCULPTEUR project, to which the various data structures are mapped. A demonstration of one such mediation tool, the Concept Browser, developed as part of the SCULPTEUR project, follows.

Résumé (FR)

L'« infodiversité » (c'est-à-dire, l'existence de diverses manières de structurer l'information et de définir le contenu de l'information) existe au sein des musées, ainsi qu'entre musées et bibliothèques. Il n'est ni souhaitable ni faisable de normaliser les pratiques d'institutions de mémoire de toutes sortes, dans la mesure où chacune d'entre elles a des objectifs et des besoins spécifiques, pour des applications spécifiques. La normalisation et l'intégration de l'information peuvent être atteintes à un niveau supérieur, par le biais de systèmes de médiation qui interprètent la sémantique de chacune des structures de données distinctes. Ce processus est rendu possible grâce à des modèles de référence sémantique, tel, dans le cas du projet à financement européen SCULPTEUR, le CIDOC CRM, vers lequel est établi un mapping pour chacune des diverses structures de données. Suit une démonstration du Concept Browser, qui fait partie de ces outils de médiation et qui a été élaboré dans le cadre du projet SCULPTEUR.

I. Introduction

Information is structured in a variety of ways in different museums. There is no "one size fits all" standard or data schema covering at the same time fine arts museums, archaeological museums, natural history museums, sites and monuments, conservation-restoration, analytical laboratories, preservation, etc., although the *International Guidelines for Museum Object Information: the CIDOC Information Categories*[1] represent an effort toward standardisation of the *content*, as opposed to the *structure*, of museum information.

This "infodiversity" becomes even more significant when the intention is to integrate museum information and bibliographic information. Bibliographic information is produced by libraries, and both the content and structure of bibliographic information have been highly standardised for a long time. However, "infodiversity" remains an issue due to the fact that not all libraries use the same MARC format nor apply the same cataloguing policy.

Still, the benefit of integrating museum information and bibliographic information would be huge for end-users. As the result of a single query, they could retrieve information related to the location, the date or the iconography of a painting, studies about that painting, accessible reproductions and digitisations of that painting, and working tools that include a reference to that painting. They could navigate seamlessly from an ornithologist's biographical details to information about the species he discovered, specimens of those species, the works he wrote, and the sound recordings he made. They could easily find documentation about the historical context in which an archaeological object came into existence.

Infodiversity is not a bad thing in itself; on the contrary, it can be argued that "infodiversity is good."[2] Since it is at the same time impossible and undesirable to compel different institutions, with different objectives and different needs, to use the same standards, information integration has to be reached through indirect means, via standardisation at a higher level than the level of implementation within each single institution. Information integration has to make use of mediation tools that are able to extract a common semantic value from variously structured information systems.

---

[1] <http://www.willpowerinfo.myby.co.uk/cidoc/guide/guide.htm>.

[2] Gill, Tony. "When the rubber hits the road: using the CIDOC CRM in the real world." In: *Sharing the knowledge: International CIDOC CRM Symposium, March 26-27, 2003* [on line]. Heraklion, Greece: FORTH, 2003 [cited 5 August 2005]. Available at: <http://cidoc.ics.forth.gr/docs/symposium_presentations/gill_2003-when-rubber_hits_road.ppt>.

II. Mediation through Semantic Reference Models

Such mediation tools are based on "ontologies" or, more modestly speaking, semantic reference models. A semantic reference model provides a synthetic view of all interrelated "classes" or "entities" the existence of which is recognised in a given domain, and defines the precise meaning of the relationships that exist between and among them.

The CIDOC Conceptual Reference Model (CIDOC CRM) is one such semantic reference model. It was developed specifically for museum information by the ICOM CIDOC (International Council of Museums, International Committee for Documentation[3]). It has been in development since 1996 and it is about to be published as an ISO standard, ISO 21127. The CIDOC CRM can be used as the basis for data exchange between systems, as a reference guide for the design of new cultural heritage information systems, and as the basis for integrated query tools and mediation systems' data schemas.[4]

Although the CIDOC CRM focuses on museum information, it proved possible to use it in the context of bibliographic and museum information integration. Much of the semantics defined in the CIDOC CRM for museum objects is also valid for the description of bibliographic resources. The main difficulty comes from the fact that museum descriptions relate to physical, individual, "unique" objects, whereas bibliographic descriptions focus on the abstract notion of "publications", which are exemplified by the individual items actually held by libraries. An international working group is currently striving to harmonise those two views by expressing the FRBR model, the conceptual model developed by IFLA (International Federation of Library Associations and Institutions) for bibliographic information, and "plugging" it into the CIDOC CRM. However, this objective is far from being achieved, and there is no publicly available documentation about this process for the time being. Since bibliographic and museum information integration was one of the objectives of the European-funded SCULPTEUR project, and this project had a rather tight deadline, it was decided not to wait until FRBR and CIDOC CRM models were harmonised, and to start working using solely the CIDOC CRM model.

[3] <http://www.willpowerinfo.myby.co.uk/cidoc/>.

[4] Crofts Nicholas, Doerr Martin, & Gill Tony. "The CIDOC Conceptual Reference Model: a standard for communicating cultural contents". In: *Cultivate Interactive* [on line]. Issue 9. February 2003. Available from the World Wide Web: <http://www.cultivate-int.org/issue9/chios/>.

III. Mapping

To achieve semantic interoperability through the use of a common conceptual reference model, mappings from the legacy metadata structures to the common model are required.

Each distinct data structure has to be mapped to the common semantic reference model, in the most detailed way possible. In the specific case of the SCULPTEUR project, the museum metadata schemas that were mapped to the CIDOC CRM included: the EROS database format developed by the C2RMF (Centre de recherche et de restauration des musées de France), the format developed by the National Gallery in London, the Victoria and Albert Museum format, and the Galleria degli Uffizi format.

The only bibliographic format that we have taken into consideration so far is UNIMARC, as it is the format employed by the DMF museum libraries network, comprising the C2RMF library. A first, rough mapping of UNIMARC to CIDOC CRM was proposed in 2004 (REFERENCE!!!). This mapping was too quickly done and it was not satisfactory: it is full of mistakes, often too vague, and even inconsistent at times. It is currently being corrected but that is a very slow process.

Mappings express the semantics of source data schemas in the form of triples involving a "domain" class from the CIDOC CRM (the same CIDOC CRM class should be consistently used as the "domain" class within a given data schema), a CIDOC CRM property (i.e., a relationship), and a "range" class from the CIDOC CRM (i.e., the "target" for the relationship). Additionally, some properties defined in CIDOC CRM have in turn properties, the range of which is always the E55 Type class. Complex meanings are expressed in the form of a concatenation of triples:

(Domain) Class — property —> (Range) Class (Domain) — property —> (Range) Class etc.

For instance, the EROS database information element "oeuvre_title", introducing the title under which the original creation that is infixed on a physical carrier is known, maps to:

E84 Information_Carrier P128F carries E73 Information_Object P102F has title E35 Title.

IV. Concept Browser

Once the mappings were available, they could be integrated in the *Concept Browser* that was developed, as part of the SCULPTEUR project, by the University of Southampton, UK.

The concept browser is able to display the CIDOC CRM ontology in a graphical way, and allows users to navigate to concepts of interest and request to view instances of certain concepts. The graph visualisation interface is based on “TouchGraph”[1], an open source graph layout system that has been extended and adapted to suit our requirements.

Informal user trials involving museum and gallery partners showed that the terminology and complexity of the CIDOC CRM proved too challenging to visualise in an intuitive way. This led to the creation of customisable simplifications based on each museum's legacy metadata structure to increase familiarity for the museum users. These simplifications are more appealing to end-users, as well-known concrete notions such as “material,” “technique” and “place,” are used instead of the awkward CIDOC CRM concept and property labels.

An important aspect of ontological visualisation tools is querying for instances of concepts. The concept browser has based instance visualisation and query on mSpace interfaces [2],[3]. mSpace is an interaction model designed to allow a user to navigate in a meaningful manner the multi-dimensional space that an ontology can provide. mSpace offers potentially useful slices through this space by selection of ontological categories. mSpace interfaces are extremely well suited to present the information from the museum systems in SCULPTEUR when structured according to the CIDOC CRM ontology. mSpace interfaces are based on slices through an ontological space, with each slice represented as a list of values. Typically, mSpace interfaces use a multipanel display, where slices are presented as columns arranged from left to right. Selection in a slice will update the display so that the values displayed in the next slice (i.e. to the right of the current slice) are related to that value.

For example, if there is a slice of artists and the next slice is painting titles, selecting an artist will display only that artist's paintings in the titles slice. Values in each slice are filtered, so that there are always results to view in the next column when a selection is made. When an item is chosen in a slice, details about that item are displayed in a detail panel; if no details are available for that item, examples of related objects are shown. Slices can be freely interchanged, removed and new slices can be added to the mSpace.

The museum metadata being dealt with in SCULPTEUR is large and varied, so there are many possible slices as well as combinations of slices that users may be interested in. The simplified CIDOC CRM, based on TouchGraph, allows users to browse and add the slices in which they are interested into the mSpace browser, where they can be arranged to suit the user's preference. A preview panel displays the current slice arrangement, so that users can view the mSpace slices as they are put together in this interface. Predefined slice layouts can be selected, and users are able to save and load their own arrangements.

An animated demonstration of the Concept Browser is available at: <http://sculpteur.it-innovation.soton.ac.uk/help/help.jsp> (click on “Using the Concept Browser”).

IV. Conclusions

In this paper we have discussed how the issue of “infodiversity” can be overcome through the use of common semantic reference models, such as the CIDOC CRM. We have described the benefits of integrating museum information and bibliographic information, and shown how in the SCULPTEUR project we applied mappings from various museum metadata schemas to the CIDOC CRM. Through the use of a mediation tool, the mappings expose the museum information to our concept browsing tool, that we have described.

Acknowledgements

The authors wish to thank: the European Commission for support under grant IST-2001-35372; our collaborators on the project, including Francis Schmitt and Tony Tung of ENST, Paris, Rachel Coates of the V&A museum, Joseph Padfield of the National Gallery, Adrian Pillinger and Dan Prideaux from IT Innovation. M Chapman, P Dibdin, R Sadotra, S Samangooei, A Smithson and

T Wirdyanto who, as Master of Engineering students, contributed to the development of SCULPTEUR prototypes; Patrick Le Boeuf of the Biblioteheque Nationale de France for assistance with mapping to the CRM; TouchGraph (www.touchgraph.com) for software used in the concept browser; and Hewlett-Packard's Art & Science programme for the donation of server equipment.

References

[1] TouchGraph (2003), TouchGraph LLC. http://www.touchgraph.com/

[2] mSpace (2005), http://mspace.ecs.soton.ac.uk

[3] schraefel, m. c., Karam, M. and Zhao, S. (2003) mSpace: interaction design for user-determined, adaptable domain exploration in hypermedia. In *Proceedings of AH 2003: Workshop on Adaptive Hypermedia and Adaptive Web Based Systems* , pages pp. 217-235, Nottingham, UK. De Bra, P., Eds.

