Dig/talia

Rivista del digitale nei beni culturali

Cataloging Cultural Objects and CDWA Lite: New Data Content and Data Format Standards for Art and Material Culture Information

Murtha Baca
Getty Research Institute

Great strides were made during the 1980s and 1990s in the development of standards for art and cultural heritage information in the anglophone world. But what was still lacking in the first years of the twenty-first century was a data content (i.e., cataloging) standard for cultural heritage information, and a technical format specifically designed for the communication and exchange of cultural heritage data in machine-readable form. This paper will provide a brief overview of cultural heritage data standards tools in the anglophone world, culminating in the development of CCO (Cataloging Cultural Objects), a set of guidelines for describing cultural objects and their visual surrogates. It will also introduce CDWA (Categories for the Description of Works of Art) Lite, a data format/technical interchange standard for expressing and sharing CCO-compatible metadata records.

Introduction

Cataloging Cultural Objects (CCO)[1] and the CDWA (Categories for the Description of Works of Art) Lite XML schema[2] were developed in response to a very specific need for certain types of data standards in the art and material culture information communities.

For clarity, I will start by defining the four types of essential data standards to which I refer in this paper:

1 *Data structure standards.* Metadata element sets. The MARC set of fields, the International Standard for Archival Description (ISAD), and the Dublin Core Metadata Element Set (DCMES) are examples.

2 *Data value standards.* Controlled vocabularies, thesauri, subject headings, etc. The Library of Congress Subject Headings (LCSH), Thesaurus for Graphic Materials (TGM), Medical Subject Headings (MeSH), and the Art & Architecture Thesaurus (AAT) are examples.

[1] http://www.vraweb.org/ccoweb/.
[2] http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/index.html.

3 *Data content standards.* Anglo-American Cataloging Rules (AACR), Describing Archives: A Content Standard (DACS), Descriptive Cataloging of Rare Books (DCRB), and Regole italiane di catalogazione per autori (RICA) are examples.

4 *Data format/technical interchange standards.* MARC21, the MARC XML schema, and the Dublin Core XML schema are examples.

Data Value and Data Content Standards for Art and Material Culture

By the turn of the twentieth century, two of the four types of essential data standards had been developed for the art and material culture communities: there were data value and data structure standards specifically created for art and material culture collections. As early as 1980, the Getty had begun to develop the Art & Architecture Thesaurus (AAT)[3], responding to the need for a controlled vocabulary specifically designed for cataloging works of art, architecture, and material culture; later in that same decade, the Getty Vocabulary Program was formed, and the Union List of Artist Names (ULAN)[4] and the Getty Thesaurus of Geographic Names (TGN)[5] were added to the controlled vocabularies published by the Getty. The Library of Congress' Thesaurus for Graphic Materials (TGM) was first published in 1987[6]. Other vocabulary and classification tools for describing works of art, including ICONCLASS[7], have been in use not only in the anglophone world but also in several European countries for several decades[8].

MARC is a data structure standard that can be used to describe visual works; one version of this standard, MARC VIM, is intended specifically for the description of visual works, but it is rarely used now. Essentially, MARC was designed and is optimal for describing bibliographic works. Realizing that there was a real need in the art documentation and museum communities for a data structure standard specifically designed for describing unique works of art and architecture, in the late 1980s the Getty and the College Art Association of America (CAA) developed an extensive set of metadata elements and guidelines, Categories for the Description of Works of Art (CDWA)[9].

---

3 http://www.getty.edu/research/conducting_research/vocabularies/aat/.
4 http://www.getty.edu/research/conducting_research/vocabularies/ulan/.
5 http://www.getty.edu/research/conducting_research/vocabularies/tgn/.
6 http://www.loc.gov/rr/print/tgm1/.
7 http://www.iconclass.nl/. Also available in Italian at
   http://www.iconclass.nl/libertas/ic?style=index.xsl&taal=it.
8 For more details, see my article *Fear of Authority? Authority Control and Thesaurus Building for Art and Material Culture Information,* «Cataloging & Classification Quarterly», 38:3–4 (2004).
9 http://www.getty.edu/research/conducting_research/standards/cdwa/.

Two other data structure standards, based on subsets of the CDWA elements, are the Visual Resources Association (VRA) Core Categories[10] and Object ID[11]. I have written in some detail about these standards elsewhere[12].

What was still missing were a data content standard specifically applied to unique museum and special collections-type objects and built works, and a technical format or data interchange standard for expressing and exchanging metadata records about those kinds of works.

CCO and other Data Content Standards

By the time the twentieth century was about to turn, much progress had been made in the art and material culture communities in the development of two of the fundamental types of data standards –data value and data structure standards. But there was no "AACR for art objects"– that is, there was no cataloging standard designed specifically to address unique items of art, architecture, and material culture. The response to this lack was Cataloging Cultural Objects (CCO), an extensive set of guidelines (to be published by ALA Publications in the summer of 2006) and examples developed by an editorial team under the auspices of the Visual Resources Association. In a 2004 article in the «Art Libraries Journal», Elisa Lanzi, the principal investigator on the CCO project, offered an excellent history and overview of CCO, including cataloging and authority record examples[13]. It is the hope of the cataloging community that CCO will take its place beside AACR (the primary data content standard for libraries in the English-speaking world, now in the process of evolving into RDA[14] (Resource Description and Access) and DACS[15] (the primary data content standard for archives in the USA, the successor to *Archives, Personal Papers and Manuscripts*[16]), as an essential tool for describing and providing end-user access to information and digital surrogates of works of art, architecture, and material culture.

---

[10] http://www.vraweb.org/datastandards/VRA_Core4_Intro.pdf.

[11] http://www.object-id.com/.

[12] *A Picture Is Worth a Thousand Words: Metadata for Art Objects and Their Visual Surrogates*, in: *Cataloging the Web: Metadata, AACR, and MARC 21* (ALCTS Papers on Library Technical Services and Collections, no. 10; Lanham, MD: Scarecrow Press, 2002); and *Practical Issues in Applying Metadata Schemas and Controlled Vocabularies to Cultural Heritage Information*, in: *Electronic Cataloging: AACR2 and Metadata for Serials and Monographs*, «Cataloging & Classification Quarterly», 36:3–4 (2003), 47–55.

[13] Elisa Lanzi, *Cataloging Cultural Objects: New Guidelines for Descriptive Cataloging*, «Art Libraries Journal», 29:4 (2004), 25–31.

[14] http://www.collectionscanada.ca/jsc/rda.html.

[15] http://www.archivists.org/catalog/pubDetail.asp?objectID=1279.

[16] *Archives, Personal Papers, and Manuscripts: A Cataloging Manual for Archival Repositories, Historical Societies, and Manuscript Libraries*, compiled by Steven L. Hensen, second edition, Chicago: Society of American Archivists, 1989.

What do CCO and AACR have in common? They are both data content or cataloging standards. They are both conceptually (but by no means exclusively) associated with particular data structure standards—AACR with MARC, and CCO with CDWA and VRA Core. Implicit in both is the use of certain data value standards—in the case of AACR, mostly (but again, not exclusively) the Library of Congress authorities and TGM, and in the case of CCO, the Getty vocabularies, ICONCLASS, but also the Library of Congress authorities and local authorities, word lists, etc.

What are the essential differences between CCO and AACR? First and foremost, while AACR was originally created and is optimal for bibliographic cataloging, the focus of CCO is the cataloging of works of art and material culture. Unlike AACR, CCO explicitly deals with database/information systems issues, which cannot be ignored in building information systems and in developing rules for cataloging in those sytems. Again unlike AACR, a considerable portion of the CCO manual is devoted to how to build and implement authorities. Unlike AACR, CCO explicitly addresses indexing versus display issues. And perhaps the most fundamental difference between AACR and CCO is the definition of what a "work" is. In the bibliographic world, as clearly stated in the Functional Requirements for Bibliographic Records (FRBR) final report,[17]

> «A work is an abstract entity; there is no single material object one can point to as the work. We recognize the work through individual realizations or expressions of the work, but the work itself exists only in the commonality of content between and among the various expressions of the work. When we speak of Homer's *Iliad* as a work, our point of reference is not a particular recitation or text of the work, but the intellectual creation that lies behind all the various expressions of the work[18]».

Within CCO, instead, a work is a creative product, including architecture, works of art such as paintings, drawings, graphic arts, sculpture, decorative arts, and fine art photographs, and other cultural artifacts. A work may be a single item or it may be made up of many physical parts. In either case, there is «a single material object one can point to as the work[19]». Thus, for CCO, the FRBR model of *work, expression, manifestation, item,* does not fully apply. Rather, the CCO entity-relationship model focuses on the relationship between works, images (or other surrogates), authorities, and sources (see next page).

---

[17] http://www.ifla.org/VII/s13/frbr/frbr.htm.
[18] FRBR final report, 3.2.1.
[19] Of course, the situation is more complex when one encounters works that occur in series or multiples, such as prints or certain decorative arts objects. In any case, in CCO a work is never an "abstract entity" as in FRBR.
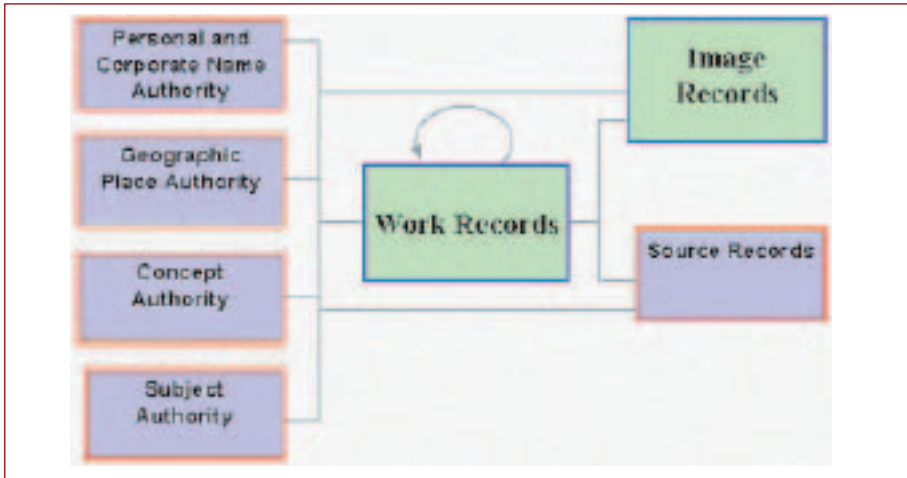
Figure 1. Entity Relationship Diagram for CCO.

As mentioned earlier, AACR is now in the process of evolving into RDA, and it is interesting to note that CCO and RDA have much more in common than do CCO and AACR. Both CCO and RDA explicitly address functions of display versus indexing. Both standards stress the importance of "relationships", although as stated above their entity-relationship diagrams are different (CCO's being much simpler because it considers unique "physical" items). Both unambiguously stress the importance of authorities. Both are output-neutral, albeit implicitly associated with particular data structure standards (CCO with CDWA/VRA Core, RDA with MARC). Both are compatible/combinable with other standards. Both are designed to build the cataloger's judgment: they are principle-based more than rule-based. Both are derived from English-language conventions, but both could be adaptable to different language conventions.

The chief differences between CCO and RDA lie in their definitions of "work" and in their "target" items for cataloging. That is, while RDA could be used for description of Web resources as well as more traditional bibliographic items, CCO's specific focus is on unique works of art and material culture, and their visual and/or digital surrogates; it is not designed to catalog Web sites or other electronic resources. It can, of course, be used to describe "born digital" works of art.

The other data content standard mentioned above is DACS, the U.S. cataloging standard for describing archival materials, approved by the Society of American Archivists. What do CCO and DACS have in common? Like AACR/RDA, both are data content (i.e., cataloging standards); both were developed because AACR did not fully meet the needs for description of cultural works and their images (CCO) and archival collections (DACS); both are output-neutral, albeit implicitly associated with

a particular data structure standard (CCO with CDWA/VRA Core, DACS with EAD, Encoded Archival Description Standard); both are combinable with other standards; both are hierarchical, or can accommodate hierarchical relationships among objects in a collection or group; and both are flexible, to accommodate diverse types of objects and groupings of objects.

What are some of the differences between CCO and DACS? First, DACS is intended for the description of intact archival collections with a common provenance and/or meaningful "collections" of materials. Second, DACS may assume a deeper hierarchy than CCO, but not necessarily, and not always; the CCO focus is on individual objects, but it can also include hierarchical groups and collections. In most cases, the hierarchical relationships may not be as deep or complicated as those expressed in DACS, especially when a cataloging archivist using DACS is working with a large, diverse archival collection.

The Missing Piece: CDWA Lite

As we have seen, the development of CCO furnished the third of the four essential types of data standards for the art and material culture communities. What was still lacking was a data format standard, a technical "container" for expressing, exchanging, delivering, and harvesting records in machine-readable form. The response to this need was the CDWA Lite XML schema[20]. This data format standard was developed by the Getty and ARTstor both as a domain-appropriate, CCO-compliant standard schema for expressing core metadata records for art, architecture, and material culture, and as a way to make metadata records and digital surrogates easily "harvestable" and hence more broadly shareable via the Open Archives Initiative Protocol for Metadata Harvesting (OAI/PMH)[21].

The CDWA Lite XML schema has a total of twenty-two high-level elements (which "wrap" around the relevant sub-elements); only nine are required. It is OAI harvestable, relatively simple, and much more appropriate for expressing metadata records for art and material culture than «the wretched Dublin Core» (as Michael Gorman aptly describes the metadata schema that has been so extensively used and misused in recent years)[22].

---

[20] http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/index.html.

[21] For more details on the history and development of CDWA Lite, see Karim B. Boughida, *CDWA Lite for Cataloging Cultural Objects (CCO): A New XML Schema for the Cultural Heritage Community*, in: *Humanities, Computers and Cultural Heritage: Proceedings of the XVI International Conference of the Association for History and Computing, 14-17 September 2005*, Amsterdam: Royal Netherlands Academy of Arts and Sciences, 2005, p. 49–56. Available on line at http://www.knaw.nl/publicaties/pdf/20051064.pdf. On the OAI protocol, see http://www.openarchives.org/OAI/openarchivesprotocol.html.

[22] http://www.sba.unifi.it/ac/relazioni/gorman_eng.pdf.

CDWA Lite XML Schema: Required and Not Required High-Level Elements	
Required Elements (9)	**Not Required Elements (13)**
Descriptive Metadata	
1. Work Type 2. Title 3. Display Creator 4. Indexing Creator 7. Display Materials/Techniques 12. Display Creation Date 13. Indexing Dates 14. Location/Repository	5. Display Measurements 6. Indexing Measurements 8. Indexing Materials/Techniques 9. State/Edition 10. Style 11. Culture 15. Subject 16. Class 17. Description/Descriptive Note 18. Inscriptions 19. Related Works
Administrative Metadata	
21. Record ID and Type	20. Rights for Work 22. Resources *(e.g., images of works, including metadata for the rights to those images)*

Practical Applications: Integrating Diverse Data Standards to Enhance Access and Understanding

The data structure standards I have been describing here should not be seen as "competitors" with one another; rather, they are potential "partners" that can complement one another and enhance the accessibility and usability of descriptive metadata and related digital objects. A simple and quite obvious example would be the use of EAD as a data structure and data format standard, and DACS as the data content standard, at the collection level for an intact collection of personal papers, ephemera, and objects, combined with the use of CDWA as the data structure standard, the CDWA Lite XML schema as the data format standard, and CCO as the data content standard for items within that same collection. The following is an example from the special collections of the Research Library at the Getty Research Institute.

Figure 2. EAD (data structure/data format) and DACS (data content) used at the collection level for an archival collection with a common provenance that includes individual art objects.
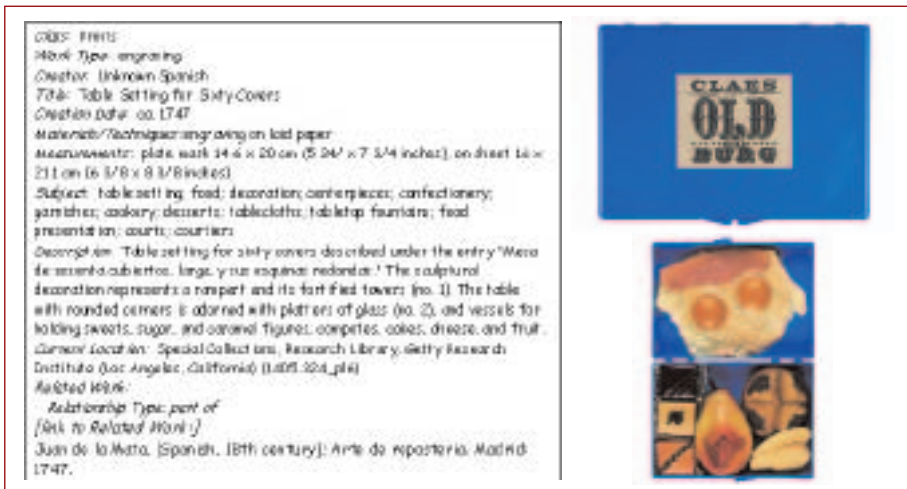

Figure 3. CDWA (data structure), CDWA Lite (data format), and CCO (data content) used at the item level for an individual work within the collection described in Figure 2.

Another example might be to use MARC as the data format standard and AACR as the data content standard for the "parent" record (in this case, for an eighteenth-century cookbook that includes engravings), and CDWA Lite as the data format standard and CCO as the data content standard for the "child" record for an individual engraving from the larger entity that is represented in the MARC record.

Figure 4. MARC (data structure/data format) and AACR (data content) used for a "parent" item (18th-century book with engravings).

Figure 5. CDWA (data structure), CCO (data content), and CDWA Lite (data format) used at the item level for an individual engraving from the "parent" work represented in the MARC record in Figure 4.

CCO, unlike AACR, allows the cataloger to devise a brief descriptive display title in the language of the catalog record (in this case, English), achieving much more comprehensible results displays and browse displays, and much greater accessibility and usability for end-users. If AACR were used in its strictest interpretation (take the title as inscribed on the item verbatim) for the item-level record, the ti-

tle would be "M. de 60 C". Even if the cataloger chose to spell this out ("Mesa de 60 cubiertos"), the end-user would have to know Spanish and realize what the parent work was constituted by, in order to understand what he or she had retrieved. CCO, instead, instructs the cataloger to devise a brief display title, in this case "Table Setting for Sixty Covers", which is much more user-friendly.

Conclusion

CCO and CDWA Lite are data standards that are still in their infancy. But with the increasing demand for digital images and other media relating to art museum collections, special collections items, works of architecture, and other objects of material culture, it is hoped that they will play a key role in enhancing description and access to art and cultural heritage resources on line. It is also hoped that data content standards like CCO and the emerging RDA, data value standards such as thesauri and other controlled vocabularies, and data format standards like CDWA Lite, the Metadata Encoding and Transmission Standard (METS)[23], and the Metadata Object Description Schema (MODS)[24], combined with protocols like OAI/PMH, will enable greater sharing and dissemination of metadata and associated digital resources for an increasingly broad and more diverse user community.

Nel corso degli anni '80 e '90 il mondo anglofono ha registrato grandi progressi quanto allo sviluppo di standard per la descrizione di opere artistiche e culturali. Purtuttavia, all'inizio del nuovo millennio ancora non si disponeva di uno standard di rappresentazione (ovvero di catalogazione) per i dati culturali, e di un formato tecnico per la trasmissione e lo scambio di dati culturali in forma machine-readable *(ovvero rilevabile meccanicamente). Il presente articolo propone una breve panoramica degli sforzi compiuti nel mondo anglofono sul fronte degli standard per i dati culturali, culminati nello sviluppo delle linee guida CCO (Cataloging Cultural Objects) per la descrizione degli oggetti culturali e dei relativi surrogati visivi. Viene inoltre presentato lo schema CDWA (Categories for the Description of Works of Art) Lite, un formato/standard di interscambio tecnico per la creazione e lo scambio di metadati in forma CCO-compatibile.*

---

[23] An XML schema for "packing" complex digital objects and related metadata: http://www.loc.gov/standards/mets/.
[24] A MARC-derived XML schema: http://www.loc.gov/standards/mods/.

Pendant les années quatre-vingt et quatre-vingt dix le monde anglophone a réalisé de nombreux progrès dans le développement des standards de description d'œuvres artistiques et culturelles. Toutefois, à l'aube de l'an 2000 il ne disposait pas encore d'un standard de représentation (de catalogage) des données culturelles ni d'un format technique de transmission et d'échange des données culturelles sous forme machine-readable *(c'est à dire lisible mécaniquement). Cet article propose un bref aperçu des efforts qui ont été fait dans le monde anglophone pour développer les standards des données culturelles qui ont aboutit dans l'élaboration des lignes directrices CCO (Cataloging Cultural Objects) de description d'objets culturels et des succédanés visuels correspondants. L'article présente de plus le schéma CDWA (Categories for the Description of Works of Arts) Lite, qui est un format/standard d'échange technique de création et d'échange des métadonnées sous forme CCO-compatible.*