

CCO and CDWA Lite: Complementary Data Content and Data Format Standards for Art and Material Culture Information

Murtha Baca, Getty Research Institute

Introduction

CCO and the CDWA Lite XML schema,¹ which is based on CCO's rules and guidelines, were developed in response to a very specific need for certain types of data standards in the art and material culture information communities.

To set the stage, I will begin by briefly 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),² and the Dublin Core Metadata Element Set (DCMES)³ are examples.
- 2. Data value standards:** controlled vocabularies, thesauri, subject headings, etc. The *Library of Congress Subject Headings* (LCSH),⁴ *Thesaurus for Graphic Materials* (TGM),⁵ *Medical Subject Headings* (MeSH),⁶ and the *Art & Architecture Thesaurus* (AAT)⁷ are examples. Data values are used to populate data structures.
- 3. Data content standards:** cataloging rules. *Anglo-American Cataloging Rules* (AACR), *Describing Archives: A Content Standard* (DACS), *Descriptive Cataloging of Rare Books* (DCRB),⁸ and *Regole italiane di catalogazione per autori* (RICA)⁹ are examples.
- 4. Data format/technical interchange standards:** the expression of data structure standards in machine-readable form. MARC21, the MARC XML schema, the VRA Core 4.0 XML schema, MODS, and the Dublin Core XML schema are examples. As is evident, in some cases type number 1 and type number 4 are identical, and/or type 4 may be a technical "translation" of a set of elements expressed in type 1 (e.g., Dublin Core and Dublin Core expressed as an XML schema).

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 outlined above had been developed for the art and material culture communities: these were data value and data structure standards (types 1 and 2 above) specifically created for art and material culture collections.

The data value standards were the first to emerge. As early as 1980, the Getty Information Institute (now the Getty Research Institute) had begun to develop the *Art & Architecture Thesaurus* (AAT), 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)¹⁰ and the *Getty Thesaurus of Geographic Names* (TGN)¹¹ were added to the controlled vocabularies published by the Getty. The Library of Congress' *Thesaurus for Graphic Materials* (TGM) was first published in 1987.¹² Other vocabulary and classification tools for describing works of art, including ICONCLASS,¹³ have been in use not only in the Anglophone world but also in several European countries for several decades.¹⁴

Data structure standards for art and material culture came next. Of course, MARC is a data structure standard that can be used to describe visual works. There was even a version of this standard, MARC VIM, intended specifically for the description of visual works; like all the other special formats, such as those developed for archives and manuscripts, musical scores, computer files, and so on, MARC VIM was integrated into MARC in the 1990s. But essentially MARC was designed and is optimal for describing bibliographic works, especially works in the FRBR sense.¹⁵

Realizing that there was a need in the art documentation and museum communities for a data structure standard specifically designed for describing unique works of art, architecture, and material culture, 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).¹⁶ Two other data structure standards, based on subsets of the CDWA elements, are the VRA Core Categories¹⁷ and Object ID.¹⁸ I have written in some detail about these standards elsewhere.¹⁹

What was still missing were types 3 and 4 above: a data content (i.e., cataloging) standard specifically for 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 was the response to the former, and CDWA Lite and VRA Core 4.0 are responses to the latter need.

CCO and other Data Content Standards

Although 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, there was no "AACR for art objects"—that

is, there was no cataloging or data content standard designed specifically to deal with unique items of art, architecture, and material culture. The response to this lack was CCO, which is amply explicated and analyzed in several of the other essays in this issue. It is our hope 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)²⁰ and DACS²¹ (the primary data content standard for archives in the United States, the successor to *Archives, Personal Papers and Manuscripts*)²² as an essential tool for describing and providing end-user access to information and digital surrogates of works of art, architecture, and material culture.

In fact, an important, relatively recent trend in metadata is what we might call “cross-cultural” metadata—that is, the use of several different metadata standards together, as appropriate for the specific materials in hand. For example, MARC (data structure/data format) and AACR (data content) could be used to describe a rare book, while CDWA Lite or VRA Core (data structure/data format) could be used to describe an engraving that is contained in that book. A variety of data value standards, including LCSH, AAT, ICONCLASS, TGM, and others could be used to populate both the “parent” MARC record for the book and its linked “child” record for the engraving (Figure 1). Similarly, DACS (data content) and EAD (data structure/data format) could be used to describe an archival collection, while CCO (data content) and CDWA Lite (data structure/data format) could be used to describe a photograph, print, or other object that is part of that collection.

The Missing Piece of the Puzzle: CDWA Lite

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.²³

The CDWA Lite schema is a distillation of the very ample, exhaustive set of elements and sub-elements (more than three hundred in total) in *Categories for the Description of Works of Art*, with rules for populating those elements that were devised with CCO as their guiding force. 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).²⁴ The VRA Core 4.0 XML schema also holds great promise as a way to express and make harvestable records for visual resources.

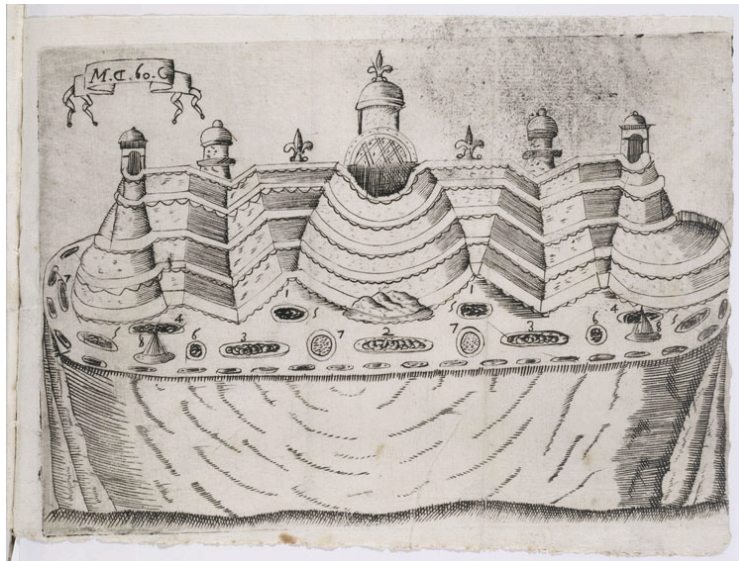
The metadata harvesting model has several advantages. Among the most important are the fact that the metadata comes from (or should come from) the institution that owns the corresponding objects or items, and is therefore accurate and authoritative. It is much better, for example, to have metadata from a museum that owns a particular work of art, than from a visual resources collection that has simply cataloged a work it does not own, using whatever information it can find. Also, if both the institutions that offer metadata for harvesting and those that harvest and aggregate metadata records act responsibly, the harvesting model can offer the most up-to-date information about an object. As we all know, in the field of art and material culture, new information is often discovered about works, and attributions may change. A work once attributed to Luciano Laurana may later be attributed to the Master of the Barberini panels. The sitter in a sixteenth-century portrait once identified as Cosimo de' Medici may later be identified as Francesco Guardi. A work once believed to be an ancient Greek kouros may later be identified as a skillful twentieth-century forgery. Also, works are constantly accessioned and de-accessioned, which is another strong argument for getting metadata from the holding institution, and refreshing it at regular intervals.

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 (Figure 2). 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 describes the metadata schema that has been so extensively talked about, used, and sometimes misused in recent years).²⁵ Since the Dublin Core XML schema is the standard format for OAI harvesting, records expressed in other schemas must include a sort of Dublin Core header or wrapper in order for the protocol to work. CDWA Lite is included among the recommended formats for OAI harvesting on the OAI Best Practices Web pages.²⁶

The goal of the CDWA Lite schema is to provide *core* descriptive metadata about cultural works—i.e., an essential metadata record that can be easily shared and contributed to union resources and that provides enough information to enable users to understand what the work is and what institution owns it. However, it is up to data providers (in OAI parlance, individuals or organizations that expose metadata records for harvesting by service providers) to decide how “light” or “heavy” their records will be. For example, in the CDWA Lite pilot project that the Getty did with ARTstor, the Getty Museum chose to expose fairly minimal metadata records, assuming that users could follow the link back to the objects on their own Web site for richer contextual information, additional images, and so on (Figure 3).

The Getty Research Institute Photo Study Collection, instead, chose to map as much of the data as possible from their local records into the CDWA Lite schema, seeing this as an opportunity to both expose their records to a wider audience and to make as much information available as possible to that audience (Figure 4). A key factor that contributed to these different approaches to mapping metadata from local information systems to a standard schema like CDWA Lite is the fact that the Getty

Figure 1: Example of the use of CDWA Lite (data structure and data format) and CCO (data content) for an item-level record describing a visual work (engraving), and MARC (data structure & data format) and AACR (data content) for a "parent" record describing the rare book that contains the engraving.



Class: Prints
 Work Type: engraving
 Creator: Unknown Spanish
 Title: **Table Setting for Sixty Covers**
 Creation Date: ca. 1747
 Materials/Techniques: engraving on laid paper
 Measurements: plate mark 14.6 x 20 cm (5 3/4 x 7 3/4 inches), on sheet 16 x 21.1 cm (6 3/8 x 8 3/8 inches)
 Subject: table setting; food; decoration; centerpieces; confectionery; garnishes; cookery; desserts; tablecloths; tabletop fountains; food presentation; courts; courtiers
 Description: Table setting for sixty covers described under the entry "Mesa de sesenta cubiertos, larga, y sus esquinas redondas." The sculptural decoration represents a rampart and its fortified towers (no. 1). The table with rounded corners is adorned with platters of glass (no. 2), and vessels for holding sweets, sugar, and caramel figures, compotes, cakes, cheese, and fruit.
 Current Location: Special Collections, Research Library, Getty Research Institute (Los Angeles, California) (1405.324_pl6)
 Related Work:
 Relationship Type: part of
 [link to Related Work:]
 Juan de la Mata, (Spanish, 18th century); *Arte de reposteria*. Madrid: 1747.

Author: [Mata, Juan de la, 18th cent.](#)
 Title: *Arte de reposteria* : en que se contiene todo genero de hacer dulces secos, y en liquido, vizcochos, turrone, y natas: bebidas heladas de todos generos, rosolis, mistelas, &c. : con una breve instruccion para conocer las frutas, y servir las crudas : y diez mesas, con su explicacion / su author Juan de la Mata, repostero en esta Corte ...
 Publication Information: En Madrid : Por Antonio Marin : se hallara en casa de Simon Moreno, mercader de libros ..., año M.DCCXLVII [1747]
 Description: [26], 196 p., [10] leaves of folded plates : ill. (engravings) ; 21 cm. (4to)
 References: Palau y Dulcet (2nd ed.) 157655
 Notes: Signatures: para⁸ 2para⁴, A-M⁸ N².
 Illustrations: comprise arms of the dedicatee (para1 verso) and bound in at the end ten ill. on folded leaves depicting table settings for 10, 20, 30, 40, 50, 60, 70, 80, 90 and 100 pieces of covers. Each table setting is also described (p. 188-196).
 Subjects: [Table setting and decoration --Early works to 1800.](#)
[Dinners and dining --Early works to 1800.](#)
[Confectionery --Early works to 1800.](#)
[Food --Early works to 1800.](#)
 ID number: 1405-324
 Location: SPECIAL COLLECTIONS - CONTACT REFERENCE
 Call Number: [TX783 .M38 1747](#)
 Copy: Copy 1
 Status: ID: 1405-324 ID: 1405-324 c.1 Checked Out

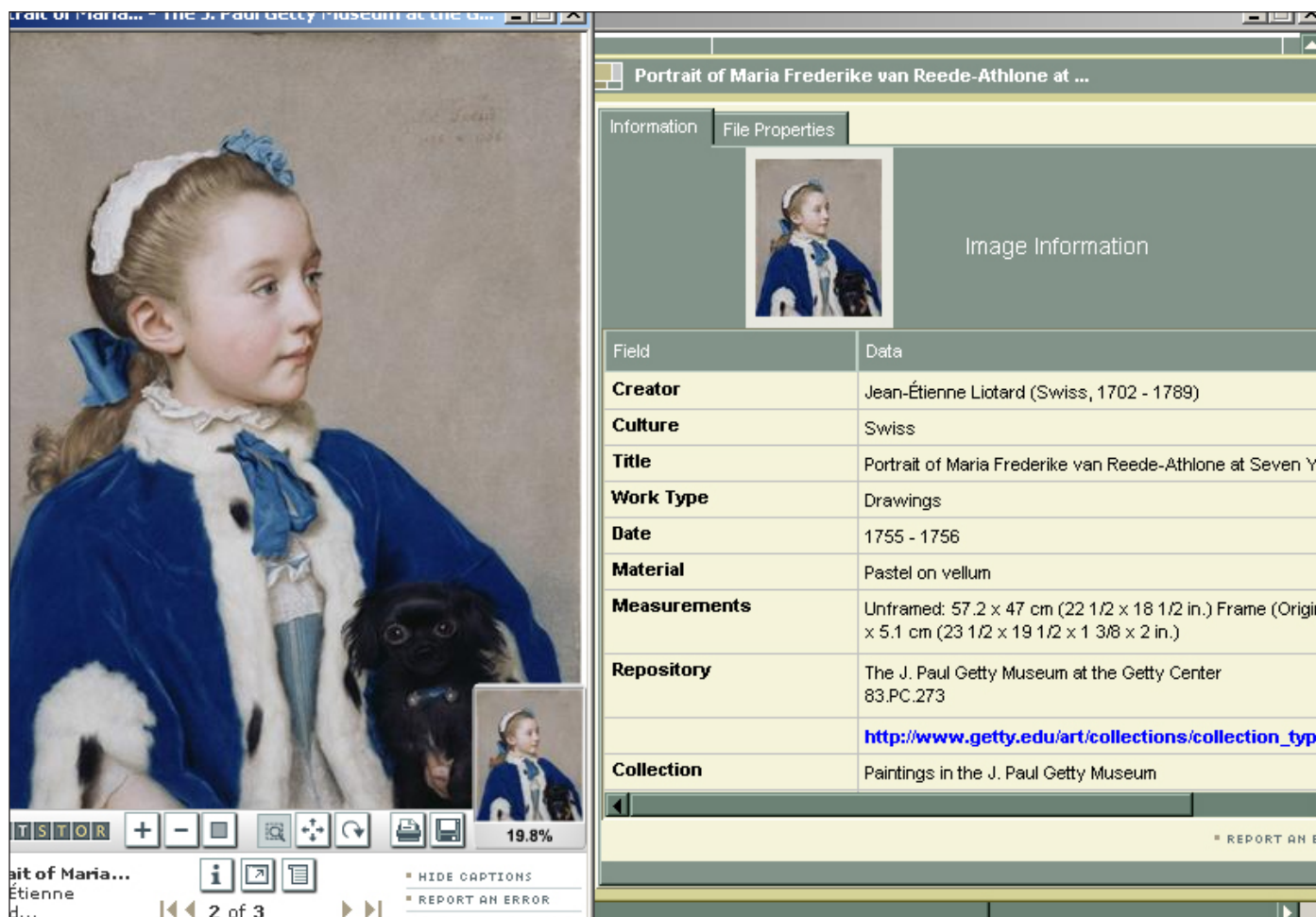
Figure 2: CDWA Lite XML Schema: Required and Not Required Top-Level Elements..

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

Museum’s data is part of the “Visible Web,” i.e., it can be indexed by commercial search engines like Google and thus is easy for users to find, while the Getty Research Institute Photo Study Collection’s records are “Deep Web,” i.e., they are in a proprietary database that is not indexed by commercial search engines and therefore much less likely to be found by users who do not already know about them.

The CDWA Lite schema also contains administrative metadata elements, including a Rights for Work element, as well as a Resources element. In OAI parlance, a “resource” is a digital object (e.g. a digital image, other media file, etc.) that is related to (usually some kind of surrogate of) the work that is described in the metadata record itself. The schema includes a sub-element for an explicit link back to the record for this object in its “native” environment (e.g., on a museum or other institution’s Web site); in some cases, the metadata provided in the “home” environment may be much more detailed and provide greater contextual information than the core metadata record that is expressed in the CDWA Lite schema, as in the case of the Getty Museum records mentioned above. Thus a two-fold goal is achieved—ease of sharing/contributing data, and a link back to the object in its own institutional context. The Resources element also includes a subelement for data about the rights to the resource (for artworks, usually the reproduction rights for images of the work).

Figure 3: Core CDWA Lite harvested metadata record for a painting in the Getty Museum.



Of the eight required high-level descriptive metadata elements in the CDWA Lite schema, Work Type, Title, Display Creator, Display Materials and Techniques, Display Creation Date, and Location/ Repository are those that together make it possible to uniquely identify a work of art or architecture and to distinguish it from other similar works. In keeping with CCO, CDWA Lite clearly distinguishes between data values for display (i.e., for easy comprehension by users) and those for indexing (to facilitate accurate online searching).

Of the non-required elements, Style, Culture, and Inscriptions are particularly relevant for works of art and architecture, and are not elements that will be found in a generic metadata element set like Dublin Core. The CDWA Lite elements for Style and Culture do not correspond to MARC elements, while the State/Edition element, which is meant to accommodate important metadata relating to works that occur in multiples or series, such as prints and some furniture and other decorative arts, for example, does correspond to the 250 Edition field in MARC. Whereas a typical LCSH string of subject headings in a MARC 655 field might be something like "Chalices—Spanish—16th century," in CDWA Lite these are separate (and hence separately searchable) elements: Work Type, Culture, and Display Creation Date. With regard to the Title element, CDWA Lite is informed by the CCO rule that if there is no title *per se* for a work, the value that is used to populate the Work Type element should be repeated. For example, one might have a record with the value "Brewster chair" in both the Title and Work Type elements.

On the other hand, again according to CCO, CDWA Lite clearly distinguishes between the Title and Inscription elements. Unlike bibliographic works, art works and other cultural objects often bear inscriptions that were never intended as titles; hence the AACR rule of "title from item" does not work in these cases. Thus, per CCO, a work might be given the assigned title "Tarot Card Depicting a Peasant Wedding," while it bears the inscription "Alhier auf disen Jäger Wagen läst sich von guten Wildpret sagen"—which, apart from the obvious linguistic difficulties it might pose for end users, simply would not give the user the necessary information to understand what had been retrieved in a results list, for example.

Figure 4: Fuller CDWA Lite harvested metadata record for an image of a tapestry in the Getty Research Institute's Photo Study Collection.

Creator	Reymbouts, Martin II (Netherlandish (before 1600) (workshop, attr.) [weaver]
Culture	Flemish
Title	Scipio upbraiding Masinissa [Massanissa/Massinisa] Story of Scipio Africanus (Major)
Collection	Study photographs of tapestries in the Getty Research Libr.
Work Type	Textiles - Tapestries
Date	c. 1590-1600
Location	Creation Location: Brussels
Repository	Belgium, Brabant, Brussels, Musées royaux d'art et d'histoi http://www.getty.edu/research/conducting_research/digitiz French & Co.
Material	wool (warp: 6-7/cm); silk
Measurements	H 8'3" x W 6'5" Brussels catalog: H 3.30 x W 2.65 m
Description	Scipio dressed in armor, holds saber in right hand & foreground); Masinissa stands before Scipio with arms foreground); 2 soldiers with spears (R, middle ground); ground); group of soldiers (L, background); city within mc background) (L & R BRD) central medallion with man & lion flanked b female figures at top & bottom under trellised arbor; (UF with 2 female archers aiming at wild boar, flanked by grotes lions French & Co. stock sheet missing from archive, J-359 Brussels (Musées Royaux), Tapisseries (1956), 51,52, no. : Cavallo, Boston tapestries v.1 (1967), 109 Cavallo, Boston tapestries v.2 (1967), pl. 30 City mark on lower guard, left [not visible in GCPA: 0240 on right guard, bottom [not visible in photographs]
Related Item	Compositionally similar tapestry, Boston, Museum of Fi Upbraiding Masinissa"
Subject	History
ID Number	Original database: 962 Web database: 304816
Source	Unknown G. L. Hunter Collection

Another very important—and potentially very powerful—element in the CDWA Lite schema that does not correspond to elements in Dublin Core, MARC, or MODS, is the Class or Classification element (see Patricia Harpring's comments on this and other CCO elements elsewhere in this issue), not to be confused with the Work Type element. If judiciously used, the Class element can function as an aid to browsing for users who are unfamiliar with the contents of a particular collection, and it can provide broader and sometimes more user-friendly, less scholarly categories or facets for searching. For example, an object might be designated as a "cartonnier" in the Work Type and Title elements, while the Class element could be populated with values such as "furniture," "chairs," "decorative arts," or whatever else the cataloger or owning institution deems to be meaningful and useful for its users. Of course the Class element can take on different meanings and pose different challenges when records are contributed to a union catalog. In this case, the service provider (in OAI parlance, the individual, institution, or organization that aggregates metadata harvested from diverse contributors and includes value-added services to facilitate access) has the responsibility of seeing that the Class element is used in such a way as to enhance searching and browsing of the aggregated records.

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, METS,²⁷ and MODS,²⁸ combined with protocols like OAI/PMH, will enable greater sharing and dissemination of authoritative, up-to-date metadata and associated digital resources for an increasingly broad and more diverse user community. ♪

Notes

1. CDWA Lite XML schema: http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite/index.html.
2. *International Standard for Archival Description* (ISAD): http://www.ica.org/biblio/cds/isad_g_2e.pdf.
3. *Dublin Core Metadata Element Set* (DCMES): <http://dublincore.org/documents/dces/>.
4. *Library of Congress Subject Headings* (LCSH) and the *Library of Congress Name Authority File* (LCNAF) are available at <http://authorities.loc.gov/>.
5. *Thesaurus for Graphic Materials* (TGM): <http://www.loc.gov/rr/print/tgm1/>.
6. *Medical Subject Headings* (MeSH): <http://www.nlm.nih.gov/mesh/>.
7. *Art & Architecture Thesaurus* (AAT): http://www.getty.edu/research/conducting_research/vocabularies/aat/.
8. *Descriptive Cataloging of Rare Books* (DCRB): <http://www.itsmarc.com/crs/rare0170.htm>.
9. *Regole italiane di catalogazione per autori* (RICA): <http://sbn-ubo.cib.unibo.it/manuali-e-documentazione/docs/corso8/rica/indicrica.html>.
10. *Union List of Artist Names* (ULAN) http://www.getty.edu/research/conducting_research/vocabularies/ulan/.
11. *Getty Thesaurus of Geographic Names* (TGN): http://www.getty.edu/research/conducting_research/vocabularies/tgn/.
12. *Thesaurus for Graphic Materials* (TGM): <http://www.loc.gov/rr/print/tgm1/>.
13. *ICONCLASS*: <http://www.iconclass.nl/>. Also available in Italian at <http://www.iconclass.nl/libertas/ic?style=index.xsl&taal=it>.
14. For more details, see my article "Fear of Authority? Authority Control and Thesaurus Building for Art and Material Culture Information," in *Cataloging & Classification Quarterly* 38:3-4 (2004).
15. However, since CCO is "schema agnostic," it can in fact be used with MARC as well. And for some institutions, which do not have the technical or financial resources to implement multiple information systems, it may make perfect sense to populate the MARC "container" using the CCO rules, when describing cultural works. The Morgan Library's use of CCO as part of a suite of cataloging standards, along with AACR, is a wonderful example. See <http://corsair.morganlibrary.org/>, and Elizabeth O'Keefe's essay in the present issue of the *VRA Bulletin*.
16. *Categories for the Description of Works of Art* (CDWA) http://www.getty.edu/research/conducting_research/standards/cdwa/.
17. *VRA Core Categories*: <http://www.vraweb.org/projects/vracore4/index.html>.
18. *Object ID*, an international standard for describing art, antiques and antiquities: <http://www.object-id.com/>.
19. "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.
20. *RDA*: <http://www.collectionscanada.ca/jsc/rda.html>.
21. *Describing Archives: A Content Standard* (DACS): <http://www.archivists.org/catalog/pubDetail.asp?objectID=1279>.
22. *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.
23. For more details on the history and development of CDWA Lite, see Boughida, Karim B., "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. pp. 49-56. Available on line at <http://www.knaw.nl/publicaties/pdf/20051064.pdf>.
24. *Open Archives Initiative* protocol: <http://www.openarchives.org/OAI/openarchivesprotocol.html>.
25. "Authority Control in the Context of Bibliographic Control in the Electronic Environment," Florence, Italy, February 10-12, 2003: http://www.sba.unifi.it/ac/relazioni/gorman_eng.pdf.
26. *Open Archives Initiative Protocol for Metadata Harvesting* (OAI/PMH): <http://oai-best.comm.nsd.gov/cgi-bin/wiki.pl?MultipleMetadataFormats>.
27. *Metadata Coding and Transmission Standard* (METS) is an XML schema for "packaging together" complex digital objects and related metadata: <http://www.loc.gov/standards/mets/>.
28. *Metadata Object Description Schema* (MODS) is a MARC-derived XML schema: <http://www.loc.gov/standards/mods/>.