

Modeling CDWA Lite as an OWL Ontology

Richard J. Urban
Graduate School of Library & Information Science
University of Illinois at Urbana-Champaign
rjurban@illinois.edu

Abstract

The Categories for the Description of Works of Art (CDWA) is a foundational framework for describing cultural heritage resources. This poster presents a model of the CDWA Lite core categories as an OWL ontology.

1. Introduction

The *Categories for the Description of Works of Art* “describes the content of art databases by articulating a conceptual framework for describing and accessing information about works of art, architecture, other material culture, groups and collections of works, and related images.” Developed by the Art Information Task Force (AITF) in the early 1990s, CDWA has served as a foundational framework for the description of cultural heritage materials. The CDWA has been influential in the development of the Visual Resource Associations VRACore XML schema and the related content standard, *Cataloging Cultural Objects* (CCO). In order to facilitate sharing CDWA compliant records through the Open Archives Initiative – Protocol for Metadata Harvesting (OAI-PMH)¹, the Getty Research Institute created the CDWA Lite XML schema² based on CDWA core categories in 2006. However, at this time a representation of CDWA’s categories does not exist in a format that could be used by Semantic Web applications that rely on specifications encoded in the Web Ontology Language (OWL).

2. The CDWA “Data Model”

The domain that CDWA encompasses – “distinct intellectual or artistic creations limited primarily to objects and structures made by humans, including built works, visual art works and cultural artifacts...” – is extraordinarily complex. CDWA consists of more than five hundred categories and subcategories, of which approximately thirty-six are considered “core” categories. The CDWA Lite XML schema contains twenty-two XML elements based on the “core” categories found in the CDWA documentation.

Both CDWA and the CDWA-Lite assert that they “describe the content of art databases” and a “format for core records for works of art and material culture.” They *are not* intended to be a model of the larger domain of art and material culture objects themselves, expressly avoiding thorny questions about the “ontology of art.” The loose “data model” that lies within CDWA is frequently represented by the following problematic “E-R Diagram.”

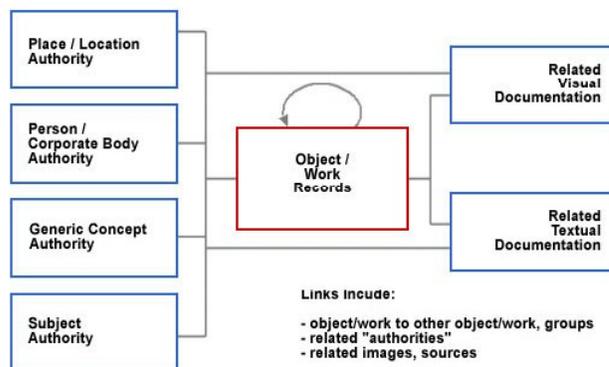


Figure 1 The problematic CDWA “E-R Diagram”³

While this diagram points the way toward the initial set of high-level entities for CDWA, it lacks important features of functional E-R diagrams, such as clearly named relationships between entities or indications of relationship cardinality. Is *Related Visual Documentation* related to *Place/Location Authorities*? Or vice versa? Are “groups” groups of records? Or are they groups of object/work “things?” Are *Visual* and *Textual Documentation* both sources? What is the relationship of the included entities to other kinds of things, like collections, that are identified in prose descriptions? Does this model actually reflect what is stated in those descriptions?

Recent work by the MuseumDAT project uncovered other confusions in the CDWA Lite XML model. ⁵ Using the CIDOC Conceptual Reference Model (CIDOC-CRM) as an evaluation tool, the project created a more “event-based” XML schema based on the original CDWA Lite elements. MuseumDAT also disambiguated confusions between “Repositories” (corporate bodies) and “Locations” (topographical features or political entities) that were present in both CDWA and CDWA Lite. While MuseumDAT’s efforts demonstrate the value of CIDOC-CRM for evaluating and refining existing XML schemas, it does not represent a move towards a CDWA ontology that might be employed by emerging semantic web environments.

3. Towards a CDWA Ontology

Ontologies are one of the important approaches to expressing the structures in how we represent knowledge. Ontology languages, such as the Web Ontology Language (OWL) and ontology-based serializations in RDF are core technologies that undergird the emerging Semantic Web. While XML schemas, such as CDWA Lite, represent an important move towards interoperable *syntaxes*, they still rely heavily on human intervention (and there fore need to be human readable) to reliably interpret their meanings. The

¹ <http://www.openarchives.org>

² http://www.getty.edu/research/conducting_research/standards/cdwa/cdwalite.html

³ http://www.getty.edu/research/conducting_research/standards/cdwa/entity.html

Semantic Web, on the other hand, aims to increase the ability for machines to understand and process the meaning of data without the need for this intervention. To be used to their fullest, these languages heavily on logical inferences; and the kinds of ambiguities present in the existing CDWA models could lead to unintended results. Developing an OWL representation of the CDWA, based on the core categories defined in CDWA Lite, would allow CDWA compliant descriptions to be used in emerging Semantic Web environments.

Noy & McGuinness suggest beginning ontology development by adopting existing ontologies.⁵ Currently the CIDOC Conceptual Reference Model (CIDOC-CRM) provides the richest domain ontology for the cultural heritage sector.⁴ As a model for data interchange and harmonization, the CRM has been used to model many of the metadata standards used by the cultural heritage sector, including MARC, Dublin Core, Encoded Archival Description and others and could also serve as a model for a CDWA ontology.⁵ However, CIDOC-CRM can be daunting in its complexity and may contain assumptions about works of art that are not shared by CDWA. Instead of attempting to fit CDWA Lite within CIDOC-CRM, this exercise allows it to speak for itself. Stanford's *Protégé* ontology editor was used to define classes and properties based on "core" CDWA categories and CDWA Lite elements.⁵ In some cases, specific CDWA elements were included as sub-classes of a more general, but un-defined class. In other cases, such as the *Materials/Techniques* element set, two separate disjoint classes were defined to acknowledge that physical materials are different from temporally rooted processes.

Frequently the decision about how to model a CDWA category/element required some reading between the lines. Although CDWA asserts that it does not deal with questions about the ontology of art, modeling CDWA as an ontology immediately uncovered implicit conceptualizations of the relationships between *Object/Work* "Things" and the *Object/Work Records* that describe them. For example, a core CDWA Lite element is *Record Type*; defined as a "Term establishing whether the record represents an individual item or a collection, series, or group of works." Since "things" such as *collections*, *series*, and *groups of works* are not included in the broader model, we are left to wonder whether or not they should be considered as a member of the *Object/Works* class or as disjoint class. The description of *Record Type* also suggests that the values recorded for this element modify the meaning of other elements — even when alternative meanings are not explicitly stated in the definition of these other elements. For example, the required *Creator* element set is defined as identifying the creators of *works of art*, however if the *Record Type* value is "collection" it must be assumed that the element identifies the creator of a collection. As these subtle changes of meanings are not accessible for computation, such assertions might lead a system to conclude that the "Getty Museum" was itself a creator of works of art.

The CDWA Lite Schema also provides a mechanism to link *Object/Work Records* with *Resources* that are surrogates or provide documentation. However, *Resources* may possess many of the same intrinsic features of the *Object/Work* they are related to. For example, a photograph may be taken to depict the work. Like the *Object/Work*, the photograph has a creator; a size; is

composed of certain materials; and may have been assigned a *Title* and an *Identifier*. In a CDWA ontology, it is possible to model these as single classes of attributes. However doing so raises questions about how to distinguish *works of art* from other types of "surrogate" works. In addition, defining shared attributes of both works and their surrogates will require re-thinking how definitions are written for these features.

4.Future Directions

For the reasons noted above, this work represents an attempt to create an CDWA ontology that stood on its own merits, based on the available documentation. The documentation however often presented further puzzles that needed to be solved, rather than clarifying how CDWA should be modeled. An alternative approach would be to map CDWA concepts into the CIDOC-CRM classes and properties. The advantage of this approach is that CIDOC-CRM is already well tested as a robust ontology capable of modeling the cultural heritage domain. A CDWA model informed by CIDOC-CRM could still retain essential characteristics, but would at the same time strengthen it by refining and clarifying muddled class and property concepts. This may however, also require rethinking how definitions of CDWA are written and how to resolve places where CDWA concepts do not fit in the CIDOC-CRM.

The challenges presented here also suggest the usefulness of ontology modeling for improving existing metadata standards, even if they are ultimately used as XML encoded data. While not all institutions are prepared to fully engage in Semantic Web environments, ontology modeling can improve the quality of existing metadata by highlighting problematic areas in standards.

5.References

- [1] J. Paul Getty Trust and College Art Association. 2006. *Categories for the Description of Works of Art*. http://www.getty.edu/research/conducting_research/standard_s/cdwa/index.html
- [2] Stein, R. and Coburn, E. 2008. CDWA and MuseumDAT: New Developments in Metadata Standards for Cultural Heritage. The annual conference of the International Documentation Committee of the International Council of Museums. Athens, 15-18 September 2008. <http://www.cidoc2008.gr/cidoc/Documents/papers/drfile.2008-06-17.0283568160>
- [3] Noy, N.F. and McGuinness, D.L. 2001. Ontology Development 101: A Guide to Creating Your First Ontology <http://ksl.stanford.edu/people/dlm/papers/ontology101/ontology101-noy-mcguinness.html>
- [4] Horridge, M., Knublauch, H, Rector, A., Stevens, R., and Wroe, C. 2004. A Practical Guide to Building OWL Ontologies Using the Protégé-OWL Plugin and CO-ODE Tools. <http://www.co-ode.org/resources/tutorials/ProtegeOWLTutorial.pdf>

⁴ <http://cidoc.ics.forth.gr/>

⁵ http://cidoc.ics.forth.gr/crm_mappings.html

