Museum Implementation of Encoded Archival Description

by Layna White, Grunwald Center for the Graphic Arts, University of California, Los Angeles

Museums have long provided descriptions of objects selected for particular exhibitions and publications. However, the history of sharing such information more directly with researchers, teachers, and students—often our targeted audience—in a freely accessible, online environment is relatively short.

Like many museums, the Grunwald Center for the Graphic Arts and UCLA Hammer Museum plan to widen public access to collection information. The Museum found a way to ease delivery and exercise its strength—knowledge of museum objects and their uses—by joining the Museums and the Online Archive of California (MOAC), a cross-community consortium motivated by the common goal of standards-based online public access to cultural heritage information. MOAC is the museum component of the Online Archive of California (OAC), a digital access project managed by the California Digital Library (CDL), which is a portal for sharing information and resources with the public online.

Museum professionals from the University of California, Berkeley Art Museum/Pacific Film Archive lead and manage the MOAC project.1

The OAC is a searchable union database with a range of contributions from variously sized and specialized California museums, archives, and libraries. Participants contribute finding aids to the project which become part of OAC's permanent collection of online descriptions and digital images.2 Finding aids are formal documentations of the context and contents of acquisitions. They can include collection indexes, registers, inventories, and guides. MOAC partners are, for the most part, contributing collection guides to OAC. At this writing, the OAC has a searchable database of over 5,000 online representations of California collections. Adhering to agreed-upon standards keeps content from these different collections co-existing peacefully in the union database.

Much of MOAC's work is supported by the University of California and funds from the Institute of Museum and Library Services (IMLS). Partners received an IMLS National Leadership award in 1999. One reason for applying to IMLS and for participating in MOAC was to provide online public access to information about geographically scattered resources. MOAC's choice in delivering public access is experimental and standards-based.

EAD As The Core Standard

MOAC partners use Encoded Archival Description (EAD) as one data structure standard for collection and object descriptions in the Online Archive of California union database. EAD is a relatively young standard developed by and for the archival community. The Society of American Archivists and the Library of Congress's Network Development/MARC Standards Office are its sponsors and promoters. The Library of Congress hosts the official EAD Web site.3

Encoded Archival Description is an application of Standard Generalized Markup Language (SGML), with a Document Type Definition (DTD), written by the archival community. EAD, then, comprises a particular set of data elements (a.k.a. units of information or data fields) selected by the archival community to represent finding aids. EAD is also a set of technical definitions for structuring and communicating these data elements, and was developed principally to move the finding aid model into the Web environment.4

Archivists describe acquisitions, typically of numerous materials related to a single source, at the collection level (for example, with biographies of collectors or agencies and notes describing the arrangement and extent of the collection). Acquisitions are further described at the series level or by thematic sections (for example, with explanations of context between objects). Lastly, archivists might provide item- or object-level descriptions. Object descriptions are clustered below series, in container lists, and range from just enough information for unique identification to rather full descriptive and subject cataloging for the most treasured objects. Arrangements of archival material often reflect the physical order constructed by the previous owner. Finding aids provide archivist and user with a means of getting a handle on or making sense of the items within a collection. EAD further assists by packaging finding aids (collection, series, and object-levels) into a single, easily manipulated machine readable text file.

Packaging overall collection descriptions and detailed object information in a single EAD text file and placing it in an online environment like the Online Archive of California can support different ways to access information. Users can enter finding aids and collection guides at any point to retrieve single, known objects, or users might view selected sections or series singly, in self-contained blocks or groups. OAC users might also advance from the top of a guide down, studying the arrangement of a collection as a whole, an approach similar to following the
presumed path through an exhibition. Currently, OAC users cannot gather 'like' content from the different descriptions in an ad hoc manner, as group display features allow. The finding aids and guides are collected and accessible in the OAC union database but they must be viewed individually. Future iterations of the OAC will surely accommodate such retrieval and display functionality.

Implementing EAD

Conceptually, the hierarchical design supported by EAD may aid in communicating context about collections and objects. Usability of the machine readable text files, however, improves when descriptive standards are applied.

The Society of American Archivists and Library of Congress maintain an EAD metadata tag library which contains data elements with recognizable labels like Personal Name, Title, and Genre/Physical Characteristic. Many data elements in this tag library are defined rather imprecisely, leaving room for elements to be used with diverse collection and object types. While there are application guidelines, EAD does not dictate how to fill data elements exactly (e.g., it does not prescribe how to express Personal Name).

Testing EAD required partners to flesh out the EAD data elements with instructions for museum implementation. These instructions are based on the OAC's supplement to the EAD application guidelines (because MOAC's implementation of EAD must be compatible with other OAC components) and standards that more adequately represent the specialized descriptive practices and anticipated use of museum information. Because partners expect to export available descriptions from local information systems, they took into consideration the practices and guidelines at their own institutions. Representing different disciplines, partners first found common ground for exchange by identifying core elements likely to recur across simple descriptions of paintings, photographs, masks, diaries, and other object types.

Data Structure Standards

The Dublin Core (DC) data structure standard identifies fifteen widely understood and easily procured data elements, which can serve as access points for discovering metadata of cultural heritage organizations, businesses, and governments when mixed on the Web. DC can help users find and gather 'like' content: for instance, students interested in Vincent van Gogh might want text written by and about van Gogh (resources described thoroughly in a MARC record). Students may also want information about works of art made by van Gogh and by his contemporaries (resources often described in museum information systems).

In 1999 the Consortium for the Computer Interchange of Museum Information (CIMI) and representatives from the museum community, such as the Walker Art Center and Solomon R. Guggenheim Museum, developed and tested a 'Guide to Best Practice' for implementing Dublin Core. The test project confirmed that given DC's intentional simplicity, the data elements could be derived and exported from museum information systems without much difficulty, a prelude to interoperability between different local systems. However, with its simplified structure, DC offers limited value to users beyond discovery and integration. Clearly, MOAC partners needed to pay attention to DC but could not limit themselves to it.

MOAC partners, therefore, looked to the REACH data element set to extend Dublin Core's simple base. REACH (Record Export for Art and Cultural Heritage) was developed by the Research Libraries Group (RLG) and informed by the Museum Education Site License Project (MESL), a test project conducted in the 1990s by the Getty Information Institute. With museum records as its reference pool, REACH lists attributes regularly required by researchers for basic object identification or access (e.g., Creator/Maker, Title, and Medium/Materials). REACH picks up more of the data routinely recorded in museum information systems. This makes it attractive because, again, MOAC partners planned to move descriptions from local systems to EAD.

REACH project managers planned to evaluate the element set, especially when integrated with other museum and library resources and delivered through a single interface, but the project was short-lived and the results of any formal evaluation remain unpublished. MOAC partners, however, found much of REACH acceptable and adopted a dozen REACH elements, and their definitions, as the MOAC data element set. Partners mapped this set, in turn, to related EAD data elements; for example, the REACH Creator/Maker data element maps to EAD's Personal Name. Even if the REACH project is defunct, what is the likelihood that any new, common, recurring categories for museum objects will be discovered or demanded by OAC users? The California Digital Library is targeting a broad audience, cutting a wide swath, from the K-12 community to experienced academic researchers.

To refine the MOAC target set as desired, partners like the University of California Bancroft Library looked to the much more in-depth Categories for the Description of Works of Art (CDWA). Produced in the 1990s by a task force comprised of information providers and seekers specializing in art and material culture, CDWA continues to evolve in response to community interests and practices through the efforts of the Getty Standards Program. Information professionals might refer to the CDWA when planning description and documentation of single objects, complex objects, collections, and their visual surrogates. While CDWA deals with hierarchical arrangement for archival and similar museum collections, MOAC partners have not discussed collective use of CDWA for describing such arrangements.

As a data structure standard, CDWA defines broad description types (e.g., Creator Identification or, nearly everything users might want to know about Creators); it also defines several more exacting subcategories (e.g., Creator Identification Name) related to objects, their makers, owners, commentators, history, and to related objects, documents, and surrogates. A select number of categories are identified as core or primary access points. These categories, thought necessary for identifying and locating objects, are common to data standards such as the Visual Resources Association (VRA) Core Categories.

CDWA is sophisticated. The standard makes a distinction between gathering descriptions from objects and creating authority or meta-records for recurring concepts, personages, and places. Maintaining separate authorities, when combined
with controlled vocabulary, can improve access because ‘like’ or related information is grouped rather than scattered throughout descriptions. CDWA’s lengthy explanations help information providers think about the type of information collected in an object’s life-cycle and desired by expert users. While the categories can be simplified for non-academic use, they are very much a wish list for researchers.

Still other MOAC partners looked to the VRA Core Categories to augment the MOAC specifications. Developed principally for visual resource professionals, the Core Categories list common, contemporary visual resource-related data elements. The Core Categories data elements recur in standards like Dublin Core, the MESL data dictionary, Museum Documentation Association’s Spectrum, and CDWA. Visual resource professionals might refer to this data structure standard when planning descriptions of images (such as slides, photographs, and digital images) and the works depicted or captured in images. The UCLA Fowler Museum of Cultural History used VRA Core Categories to develop in-house guidelines for describing both 35mm slides and the works they represent (for example, slides capturing events like Pasadena’s Tournament of Roses Parade and images of tattoos on people).

Visual resource professionals are interested in standards for managing and organizing items in their care, and have also considered standards for shared cataloging. The Core Categories, now at Version 3.0, continues to evolve under the guidance of VRA’s Data Standards Committee and the standard continues to attract users. ArtSTOR, the online collection of digital images and related cultural heritage materials in progress under the sponsorship of the Mellon Foundation, plans to develop a common presentation format for ArtSTOR based on the VRA Core Categories.

With interest in integrated access to museum information and resources collected by libraries and archives come efforts in data exporting (as evident in MOAC) and mapping to cross-community standards. The Getty Standards Program, RLG, and other information experts constructed a helpful metadata crosswalk and this record of mappings or translations across data elements is available on the Getty Research Institute’s Web site.

With our content specifications informed by standards and guidelines, MOAC’s implementation of EAD should cover the basic categories people expect to find for museum objects. This assumes that the discovery of new principal or core categories beyond what is already detailed in, for instance, CDWA, is unlikely, and assumes we have provided all the relevant data elements to Online Archive of California users.

Data Content Standards

If one can find common categories for museum objects, the next step is deciding how to answer questions such as “who made the object?” With few exceptions the MOAC specifications do not prescribe exact ways to enter data for the project. Such insistence might have meant rewriting descriptions recorded in local information systems. Recognizing that museums will likely employ idiosyncratic descriptive practices, partners agreed to fill the data elements within an acceptable range of styles. The goal is to retain local practices yet remain within project boundaries. The University of California Bancroft Library and Japanese American National Museum, for instance, prefer to enter Creator/Maker with surname first, while the Grunwald Center elects to enter the surname last.

While the international museum community has not unanimously embraced a single data structure or content standard as its own, museum professionals generally record data in predictable ways. Museum methods reflect practices, written or unwritten, understood by curators, collections managers, and experienced information users. In developing implementation guidelines for EAD, partners took care not to tie decisions too closely to MOAC. Equally, partners take care not to alter descriptions to fit the project. Readying object descriptions for public access prompted the Grunwald Center to review and improve existing descriptions (e.g., adding title translations, refining subjects, and adding birth location information). Center catalogers did not adjust descriptions to conform to MOAC. Rather, catalogers wanted to ensure that data were entered consistently and thoroughly in the local information system.

That standards can be incorporated into EAD is not surprising or exclusive. Data standards can just as easily be incorporated into vendor-based and in-house museum information systems. The Solomon R. Guggenheim Museum and Cornell University’s Image Management and Access Group (IMAG/CU) Software Project are two such examples. Each used CDWA as one source when developing local guidelines for composing descriptions posted to their respective collections information systems, and publication outside those systems. The Guggenheim, Cornell, and others write prescriptive local cataloging rules based on the CDWA’s detailed discussions and, when necessary, augment such standards by writing provisions to handle challenges and special requirements for object documentation.

In October 2001 the Visual Resources Association announced that funding from the Getty Grant Program and the Digital Library Federation will support VRA and an advisory board of cultural heritage representatives in developing a manual for describing and documenting objects and images. Such form rules, especially if museum information professionals contribute to their development, might lessen cataloger and user confusion between data elements and theoretically increase opportunities for sharing.

Data Value Standards

As with content standards, museums have similar pride of ownership when choosing words to represent objects. It follows that MOAC partners discussed controlled vocabulary, a discussion difficult to avoid, given that partners are from both the library and museum professions. MOAC project partners encourage use of controlled vocabulary but do not require it for the MOAC project. Partners apply controlled vocabulary at their discretion or as implemented in their local information systems and evaluate the rewards of controlled vocabulary against the resources required to apply it.

MOAC partners who are using controlled vocabulary select terms from a relatively small pool of published sources, both general and subject specific. The University of California Riverside/Museum of Photography, for example, uses the Library of Congress’s Thesaurus for Graphic Materials and the Getty’s Thesaurus of Geographic Names to describe the subject content of stereographs in its collection. The Japanese American
National Museum refers to the Library of Congress Subject Headings, the Getty’s Art and Architecture Thesaurus (AAT), and subject-specific sources like Japanese American History: An A-to-Z Reference From 1868 To The Present. Given the scope of the Grunwald Center collection (works on paper dating from the Renaissance to the present), Center catalogers turn regularly to the Getty’s Union List of Artist Names (ULAN) as a name authority and the AAT for subjects, mediums, and object formats. The advantage of subscribing to a trusted controlled vocabulary source like the AAT is that its developers often invest resources in expanding and maintaining relationships between terms. However comprehensive these publications are, they may not suggest every term necessary to justly describe museum objects. Many MOAC partners import vocabulary from local indexes and authority files, which are generally built over time with in-depth knowledge of a specific collection and its uses.

Archivists often provide subject access at the collection or series level, rather than the object level, because it is an expedient way to describe a group of related objects. Objects, then, inherit subject headings from the upper levels of a hierarchy. Somewhat similarly, museums might gather together objects with the same broad type of subject matter. The J. Paul Getty Museum, for example, constructed subject groupings based on broad conceptual categories like Mythology and Natural World as one way of organizing objects for its public access system.

Differences in term preferences or usage become apparent in a union database like the Online Archive of California. In 1998 MESL evaluators, led by Howard Besser and Robert Yamashita, wrote that “the different metadata vocabulary and general language used by different institutions made the creation of an integrated and consistent database problematic at best.” For example, OAC users might discover that one repository prefers Rembrandt van Rijn, while another is determined to use Rembrandt Harmensz van Rijn.

As with content standards, parameters or guidelines might help establish what is expected of describers and what users can expect from descriptions. MOAC partners, for instance, are encouraged to record sources of data values and, indeed, the EAD tag library makes provisions for recording controlled access sources. Tracking the sources of many local and published terms might lay the groundwork for building a comprehensive, underlying authority file, where synonymous, variant, broader and narrower terms can be disambiguated and linked. There is no current underlying authority control for data values in the Online Archive of California. Such an essential task would be extraordinarily demanding but quite valuable for access.

Can Museums Use EAD?

MOAC is evaluating the effectiveness of EAD to package and present descriptions of collections, other than archival, in a Web-accessible format. Can museums use the standard? While museums certainly populate local information systems with thorough object descriptions, formal collection-level descriptions may not be as readily available. When local systems are found wanting, MOAC partners evaluate and repurpose contextual information published in exhibition catalogs and brochures or recorded in research notes.

If archivists arrange collections into meaningful, hierarchical structures, how do museum professionals organize objects for access, display, and interpretation? Museums are typically selective in tapping objects for public presentation, whether in galleries, exhibition catalogs, or on the Web. Museums often impose arrangements, moving objects in and out of projects based on curatorial interests in, for example, single-artist or thematic exhibitions. This practice lends a transience to arrangements of related museum objects. Is it important to maintain such ephemeral relationships in a collection guide? Of course, many MOAC partners describe self-contained collections, such as the Oakland Museum of California’s finding aid for its excellent collection of Dorothea Lange photographs.

In addition to arrangement, MOAC is exploring EAD’s ability to handle museum object information. It is important to note that partners are not testing EAD for primary object cataloging, documentation, or management. Rather, partners are testing EAD’s ability to represent museum object descriptions in a public access system. It may be, however, that museum object descriptions are so different or special that EAD may not be appropriate. While some information about museum objects may be known and not prone to change, other information is both local and unique because the objects are unique. Creators, titles, values, and related objects are often interpretations or opinions which must be qualified.

Museums place great importance on researching and describing objects and regard changes to the object —its condition, its history of display —as valuable documentation. Museum professionals want to provide users with trusted, consistent descriptions but want to retain the flexibility to write about objects differently at different times and for different purposes. Museum descriptions lead users to objects, helping users find all objects by a given artist, for example. Once found, descriptions might support a working theory (title and subject results might suggest that Henri Matisse drafted more reclining nudes than seated nudes), or museum data might inform users with facts (Edgar Degas was born ten years earlier than Mary Cassatt), or descriptions might encourage users to look for and consider additional data (the “aha!” response when discovering something intriguing).

Information recorded, added to, and amended in museum information systems is likely more layered and complicated than that packaged and contributed to the Online Archive of California. Partners are providing public access to derivatives of our master images and derivatives of our complex, fuller master description records; that is, partners move a subset of data from a larger, more substantial source to a smaller publication format.

Evaluation

Early in the project, MOAC partners defined content and image specifications based on an understanding of community data standards and EAD application guidelines. Partners might now examine their consistency in implementing EAD.

Content production for the project is decentralized; time is spent working independently from fellow collaborators when packaging content. Partners are more likely to be concerned first with how their own collection guides look. Since integrated access to museum, library, and archive descriptions is promoted in OAC, however, partners should examine how their various descriptions work in concert with other partners, and, in particular, whether partners use EAD data elements the same way and
for the same purpose. Is there confusion, for instance, over use of scope notes in EAD? Are partners putting typically granular museum data in appropriate EAD data elements or forcing museum data into EAD? Granularity (or, breaking data elements down into smaller, distinct components) is typically desired in searchable museum object records and suggested by the exacting subcategories in the Categories for the Description of Works of Art.

Without a popular data structure standard for museums to call their own, are MOAC partners attracted to Encoded Archival Description because it is ready and appears well positioned for persistent availability, as an SGML (and eXtended Markup Language (XML) compliant) DTD? With the Society of American Archivists and Library of Congress as its maintenance agencies, EAD seems to offer an enduring, stable mechanism for content placement—even as complementary or more advanced schemas develop. What if, after evaluating EAD, partners find that museum practices and needs are incompatible with EAD? Some suggest museums need a museum standard, one more reflective of museum practices. A data structure standard serving the many types of museums and their different purposes, and one compatible with other SGML/XML DTDs, is needed if content integration is desired. In fact, EAD is but one example of an XML DTD developed within the cultural heritage community. CIMI, in collaboration with the Museum Documentation Association (MDA), is conducting an XML DTD test-bed project within the CIMI membership using MDA’s Spectrum as the underlying data structure standard.

Conclusion

Museums and the Online Archive of California (MOAC) complements other public access projects engaged in by partners. Portals like the California Digital Library are useful for publicizing collections and helping users locate and evaluate the appropriateness of objects for their needs. Museum information systems are useful for serving a more familiar, local constituency.

Collaborations such as MOAC can help object-rich but resource-poor cultural heritage institutions sustain larger public access projects. While partners are still exploring the intellectual issues around museum use of Encoded Archival Description, participation in the project has helped museums like the Grunwald Center and UCLA Hammer Museum share collections information with the public in an open, online environment.

Notes

1. MOAC partners include: Berkeley Art Museum/Pacific Film Archive; Japanese American National Museum; Oakland Museum of California; Phoebe Hearst Museum of Anthropology; Iris & B. Cantor Center for Visual Arts at Stanford University; UC Berkeley Bancroft Library; UC Berkeley Museum of Paleontology; UCLA Grunwald Center for the Graphic Arts; UCLA Fowler Museum of Cultural History; UCR/California Museum of Photography.


3. The official EAD Web site is located at http://lcweb.loc.gov/ead.

4. This paper discusses aspects of EAD related directly to collection and content description, or, frontmatter and (archival) description. EAD also makes provisions for documenting the finding aid or collection guide itself (e.g., the EAD header might include data about when and by whom the finding aid was created). Several relevant articles on EAD can be found in a special issue of Archives and Museum Informatics [12, nos. 3/4 (1998)], edited by Richard Rinehart and Anne J. Gilliland-Swetland. For an introduction to EAD implementation, see also Jackie M. Dooley, ed. Encoded Archival Description: Context, Theory, and Case Studies (Chicago, IL: Society of American Archivists, 1998).


8. The CDWA is available online at http://www.getty.edu/research/institute/standards/cdwa/.

9. The VRA Core Categories (and helpful cataloging examples) can be found at http://php.indiana.edu/~fryp/vracore3.htm.


11. Museum Documentation Association’s Spectrum is a widely used standard in the United Kingdom and, as noted, the VRA Core Categories has a popular base as well.

12. The works referred to are as follows:


14. The challenges of describing and documenting museum objects are suggested in CDWA's extensive discussions.

15. While partners are responsible for readying content for the OAC, the Electronic Text Unit of UC Berkeley offers MOAC partners exceptional technical support and advice. Berkeley's ETU plays a major role in reviewing and evaluating partner EAD text files and implementing style-sheets for the project.

The author acknowledges the generosity of Murtha Baca, Rob Cox, Richard Dolen, Anne J. Gilliland-Swatland, Richard Rinehart, and Günter Waibel.

---

**Springtime in the Rockies**

[Advert for the ARLIS Canada / NW Chapter Regional Conference]

ARLIS Canada / NW Chapter Regional Conference
Banff, Alberta, Canada
Friday May 31 - Sunday June 2, 2002