Controller Vocabularies for Art, Architecture, and Material Culture

Murtha Baca
Getty Research Institute, Los Angeles, California, U.S.A.

Abstract
There are a number of controlled vocabularies that can be used for describing and enhancing access to information resources relating to art, architecture, and material culture. These data value standards are essential for the description of movable art works, built works, and objects of material culture, as well as their visual and digital surrogates. Local, collection-specific authority files for museum and visual resource collections also play a key role in describing collections materials and providing access points for a range of users, both internal and external to the particular organization. A recent trend is reflected in the so-called folksonomies, which are the result of social tagging.

INTRODUCTION

The simplest way to define vocabularies is to say that they are groupings of terms or names; vocabularies are said to be controlled if they are limited to a particular domain or set of concepts, and designate a preferred form for each concept, person, or entity represented. Controlled vocabularies are organized collections of words, phrases, and/or names, structured to show the relationships between terms and concepts. Authority files and thesauri are types of controlled vocabularies, but a simple picklist (a user interface feature that allows users to select terms from a preset list) can also be controlled. An authority file is a type of controlled vocabulary that serves as a source of standardized forms of names, terms, or titles for a particular application, discipline, or domain. Authority files include references or links from variant forms to preferred forms. For example, in the Library of Congress Name Authority File (LCNAF), the name form “Ovile Master” links to the preferred form “Bulgarini, Bartolomeo” for the fourteenth-century Italian artist whose hand has been variously identified with the “Master of the Ovile Madonna,” “Ugolino Lorenzetti,” and “Bartolommeo Bulgarini.” These are all considered to be variants or synonyms for the preferred name form. Authority files, originally designed to regulate usage in library catalogs, also provide additional name forms. Authority files and thesauri are types of controlled vocabularies, but a simple picklist (a user interface feature that allows users to select terms from a preset list) can also be controlled. An authority file is a type of controlled vocabulary that serves as a source of standardized forms of names, terms, or titles for a particular application, discipline, or domain. Authority files include references or links from variant forms to preferred forms.

There are a number of controlled vocabularies that can be used for describing and enhancing access to collections of any kind. They can range from a simple alphabetical list of approved terms to a complex, carefully structured thesaurus, including broader and narrower concepts, related concepts, and any number of equivalent or near-equivalent terms or names denoting the same concept, as shown in the examples above. In information science, controlled vocabularies are often referred to as data value standards or value encoding schemes, since they provide the specific terms, names, and other data values that are used to populate data structures or metadata schemas such as MARC,
MODS, CDWA Lite, Dublin Core, and many others. The cataloging rules that are used to guide the way that data values are entered into data structures are often (and, for many people, confusingly) referred to as data content standards—AACR, CCO, and DACS are examples of data content standards for bibliographic materials, unique cultural objects, and archival collections, respectively. For a chart that clearly outlines the types of data standards, with examples, see Gilliland.[1]

 Elaine Svenonious gives a thorough overview of controlled vocabularies in her entry “Design of Controlled Vocabularies in this encyclopedia.” (See Svenonius, ###.)

CONTROLLED VOCABULARIES FOR ART AND ARCHITECTURE

In the areas of art, architecture, and material culture, there are several vocabulary tools that have been specifically designed for describing these types of collections, such as the AAT. Other vocabularies have been developed for other types of materials, but can also be applied to art and architecture, such as the Library of Congress Subject Headings (LCSH).

Controlled vocabularies have three main uses. They can be used as sources of standard terminology for descriptive cataloging and other forms of documentation; as knowledge bases or lookup tools for discovering the meaning of a particular term or a term for a particular concept or object, or a description of a particular individual or corporate body in the case of a name authority file; and as online searching assistants that can enhance both precision and recall in querying databases that can have multiple data values representing the same concepts, objects, or persons. (See Lanzi[21].) As of this writing, a new publication, Introduction to Controlled Vocabularies: Terminology for Art, Architecture, and Other Cultural Works by Patricia Harpring, is slated for publication in 2010.

EXAMPLES OF USES OF CONTROLLED VOCABULARIES

As sources of terminology for catalog records and other forms of documentation. In a catalog record for a painting in a museum’s collection or an image in a photographic archive, the following database fields or metadata elements should be populated with values from appropriate controlled vocabularies: object type; creator name; materials/techniques; creation location; subject matter; and so on. See the cataloging examples provided on the Categories for the Description of Works of Art (CDWA) Web site: http://www.getty.edu/research/conducting_research/standards/cdwa/examples.html), which indicate which fields in a database record should be populated with values from a controlled vocabulary.

As knowledge bases or lookup tools. A user wishes to know what “trompe-l’oeil” means; the user does a search in the AAT on that term (or on a right-truncated version of the term, in the event of being unsure of the exact spelling), and retrieves a record with the scope note that begins “French term meaning ‘deceive the eye,’ applied to images so realistic that they may fool the viewer into thinking that the represented objects, scenes, textures, or points of view are real rather than images.” If a controlled vocabulary has carefully constructed hierarchical relationships and scope notes, users can even find terms for concepts for which they do not know the word. In her article “Indexing and Access for Digital Libraries and the Internet: Human, Database, and Domain Factors,”[3] Marcia J. Bates discusses the issues associated with the fact that many users are essentially trying to describe something they don’t know or for which they have knowledge gaps, while indexes know what they are describing and use specific terminology in the process of purportedly providing access points for specific materials. The use of carefully structured thesauri and other vocabularies that link both specific and more generic terms, and even “right” and “wrong” (but frequently used) terms, can go a long way toward bridging the gap between indexes and end-users. For example, in the AAT users can navigate down the Visual Works hierarchy from “sculpture” to “funerary sculpture” to “ushabti,” or they can enter a keyword search for “Egyptian and tombs” on the scope notes and also find the record for “ushabti” (with eight variant spellings), the small sculptures that the ancient Egyptians buried with their dead to accompany them in the afterlife. Or a user may have come across the name “Beggarstaff Brothers” and wants to know who these artists were. A search in the Union List of Artist Names (ULAN) reveals that this was the pseudonym adopted by brothers-in-law William Nicholson and James Pryde (each a well-known artist in his own right) while they were collaborating in the 1890s as poster artists. The ULAN record also has links to the individual records for each artist.

As online searching assistants. An Italian searcher wishes to know if the Getty Museum has any works by the artist designated in the scholarly literature and exhibition catalogs in the Italian language as “Gherardo delle Notti.” The searcher enters this name in the Getty search engine, and retrieves works by “Gerrit van Honthorst,” the name used in non-Italian-language publications and institutions to designate this Dutch artist who was active in Italy during the first half of the seventeenth century. This is because the ULAN, with all the variant names for this artist clustered together in a single record, is being used as an intermediary between the searcher and the resources being searched. Similarly, a user can
Controlled Vocabularies for Art, Architecture, and Material Culture

search on “armchair” and retrieve a “fauteuil” (a type of armchair) in the Getty’s collection, even if the word “armchair” does not appear on the Web page for that object. This is because the power of the AAT hierarchy is being used to assist the searcher who is using a more generic, nonexpert term for a very specific object. Marcia J. Bates discusses the power of faceted searching (made possible by faceted thesauri like the AAT) in her article “The Cascade of Interactions in the Digital Library Interface.”[4]

Vocabularies in the realm of art, architecture, and material culture may contain personal names (e.g., Leonardo da Vinci), corporate names (Richard Meier & Partners), object names (squat lekythoi), iconographic subjects (Hercules and the Nemean Lion), geographic names (Luxor, Qina governorate, Upper Egypt), proper names of built works (Hagia Sophia), and so on. They can take a variety of forms. Alphabetical lists of approved terms or names are often used in local applications, to ensure consistency and accuracy. In a controlled list, all terms should be equal in specificity, and should not have overlapping meanings. Subject headings are terms and phrases, usually arranged in alphabetical order, that can be combined to express compound concepts (e.g., Triptychs—I taly—Florence—Congress). A taxonomy is an orderly classification system for a defined domain, with hierarchical relationships (e.g., Flagellation of Christ, Passion of Christ, New Testament, Bible, in ascending hierarchical order). A thesaurus is a structured vocabulary that explicitly includes the relationships between and among terms and concepts. As indicated above, the types of relationships in a thesaurus are: equivalence (still life = natura morta); hierarchical (a gisant is a type of effigy, which is a type of funerary sculpture); and associative (a charterhouse or chartreuse is a type of building associated with the Carthusian monastic order). Any one of these types of controlled vocabularies can be used as an authority file—that is, a vocabulary that is used to enforce consistency by providing standardized forms of names, terms, titles, and so on. Authority files regulate usage, but also provide additional access points for users who may be searching on non-preferred forms of names or terms. For example, in the ULAN, the Italian name “Schiavone, Andrea” is the preferred name form for a Dalmatian artist active in Italy during the sixteenth century, but searchers can also enter queries on variant forms as diverse as “Medulić, Andrij” and “Andrea Meldolla.”

SPECIFIC VOCABULARY TOOLS

Following is a list of some of the major published vocabulary tools that can be used to describe and provide access to materials relating to the visual arts, architecture, and material culture:

Art & Architecture Thesaurus® (AAT)

www.getty.edu/research/conducting_research/vocabularies/aat/

A structured vocabulary that contains terminology and other information related to art, architecture, and related disciplines. Updated monthly.

Getty Thesaurus of Geographic Names® (TGN)

www.getty.edu/research/conducting_research/vocabularies/tgn/

A structured vocabulary containing names and other information about geographic places, focusing on but not limited to places important to art, architecture, and related disciplines. Updated monthly.

ICONCLASS

http://www.iconclass.nl

A classification system for iconographic research and documentation of images. ICONCLASS is a collection of definitions of objects, persons, events, situations, and abstract ideas that can be the represented in the visual arts. The system is organized in 10 broad divisions within which subjects are ordered hierarchically. Although ICONCLASS is primarily a classification system, the textual correlates of the alphanumeric classification codes can be used to describe and/or index visual materials, as can the associated keywords. Also available in print:


Library of Congress Name Authority

http://authorities.loc.gov/

A controlled vocabulary consisting of records for names and titles of works established by the Library of Congress and cooperating libraries under the National Coordinated Cataloging Operations (NACO) program. Updated daily.

Library of Congress Subject Headings

http://authorities.loc.gov/

A structured vocabulary designed to represent the subject and form of the books, serials, and other materials in the Library of Congress collections, with the purpose of providing subject access points to the bibliographic records contained in the Library of Congress catalogs.
More broadly, LCSH is used as a tool for subject indexing of library catalogs and other materials (including visual materials). Updated daily.

**National Monuments Record Thesauri**

http://thesaurus.english-heritage.org.uk/

A set of thesauri developed by the National Monuments Record Centre, English Heritage Project (officially known as Historic Buildings and Monuments Commission for England). The thesauri include Monument Types, Archaeological Objects, Building Materials, Defence of Britain, Components, Maritime Place Names, Maritime Craft Type, Maritime Cargo, Evidence Thesaurus, Archaeological Sciences, Thred Thesaurus (terminology for the description of archive type and format), and a Historic Aircraft Type thesaurus.

**Subject Index for the Visual Arts**


A structured vocabulary for subject access to materials in the Print Room of the Victoria & Albert Museum, London.

**Thesaurus for Graphic Materials (TGM)**

www.loc.gov/rr/print/tgm1/
lcweb.loc.gov/rr/print/tgm2/


A structured vocabulary containing terms that describe both the subjects and the object/work types of graphic materials. Available in print and online versions. First issued in print as two separate works, *LC Thesaurus for Graphic Materials* (1987) and *Descriptive Terms for Graphic Materials* (1986). The online version continues to be separated into two parts, one containing subject terms and one containing object/work type terms.

**Union List of Artist Names® (ULAN)**

www.getty.edu/research/conducting_research/vocabularies/ulan/

A structured vocabulary containing names and other information on artists, architects, other individual and corporate creators of art and architecture, and repositories of art. Updated monthly.

**Controlled Vocabularies for Art, Architecture, and Material Culture**

**“UNCONTROLLED” VOCABULARIES FOR VISUAL WORKS**

A relatively recent trend in the description of visual materials is the phenomenon known as *social tagging*, prevalent on image Web sites such as Flickr (http://www.flickr.com/). Social tagging (also referred to as *collaborative tagging*, *social classification*, *social indexing*, *mob indexing*, and *folk categorization*) is the decentralized practice and method by which individuals and groups create, manage, and share terms, names, etc. (called “tags”) to annotate and categorize digital resources in an online “social” environment. A folksonomy, which is simply an assemblage of concepts represented by tags, is the result of social tagging. It should be noted that in spite of the “onomy” at the end of the word, a folksonomy is not a taxonomy, which is an orderly, structured classification that explicitly expresses the relationships, usually hierarchical, between and among the things being classified. At the time of this writing, in the world of art museums, “steve” is the most advanced project using the social tagging model to enhance access, especially but not only subject access, to images of works of art. For a description of this collaborative project for user-generated descriptive metadata for images of works of art, see http://www.steve.museum/.

Social tagging and other forms of user-generated metadata are a trend in online resources that is garnering considerable attention and interest. If this type of nonstandardized, unstructured terminology could be linked to more highly structured, standard thesauri and vocabulary tools, the potential for enhanced end-user access could be significant.

**COLLECTION-SPECIFIC LOCAL VOCABULARIES**

Many museums and other collecting institutions have learned that merely taking terms from a published controlled vocabulary like the Library of Congress authorities, the AAT, or ICONCLASS and using them to populate descriptive records for objects in their collections is not enough. They may find that they need to supplement these sources with scholarly or other specialized terms that are out of scope for discipline-specific vocabularies, or too specific for the level of indexing supported by more general vocabularies. Or they may use nonexpert, generic, or even erroneous terms or misspellings that cannot be found in a published, standard vocabulary tool, to enhance access to collections for a wide range of users. The National Gallery of Art in Washington, D.C. allows users to browse its collections by subject, using a controlled vocabulary in the form of an alphabetical list that is based specifically on categories of themes depicted in the works in its collections (available at http://www.nga.gov/collection/searchsub.
shtm). The Getty Museum uses collection-specific thesauri of both object types and subjects to enable users to familiarize themselves with the works in its collections, both on the Web and in the touch-screen kiosk system in its galleries (available as two picklists at http://www.getty.edu/art/gettyguide/).

In a collection-specific thesaurus aimed at enhancing access to nonexpert users and museum visitors, a term as unscientific as “bottles and pots” can be used to lead users to more specific types of vessels with names that would be known only to experts. While this type of “folk classification” is not legitimate for a published, standard authority like the AAT or LCSH, it is extremely useful in a local thesaurus, where it can be linked to the more accurate expert terminology as well.

**CONCLUSION**

The three main ways by which users attempt to find information online are browsing, directed searching, and following links.²

For online collections of art, architecture, and material culture, all three of these methods can be extremely powerful. Museums can use their collection-specific thesauri and descriptive cataloging records to enable users who are unfamiliar with their collections to browse by object type, creation place, chronological period, subject matter, and so on. As shown in the examples given above, directed searching can be significantly enhanced via both the terms and the hierarchical relationships in controlled vocabularies. Likewise, controlled vocabularies can be used to lead users to other materials of interest via links from one specific term or name to another. Many of the debates that are currently taking place among information professionals about how the library profession can modernize itself, compete with commercial search engines like Google, and ultimately not only survive but recreate itself, center on authority control rather than on item-level description (see, e.g., Marcum⁷). The final report of the Library of Congress Working Group on the Future of Bibliographic Control (January 2008) also stresses the need for increased focus on authority files and controlled vocabularies in a collaborative model among different institutions and communities, and the potential value of mapping values from different controlled vocabularies.

(Controlled vocabularies, combined with less structured data values such as those generated by social tagging, may very well be one of the key elements in the future of access to the constantly growing body of online resources. Since the mission of museums, archives, and other collecting and memory institutions is not merely to collect and preserve, but also to provide access to what they collect, the use of controlled vocabularies to describe and lead users to their collections should be as essential as the use of security systems and climate control to protect and preserve those collections.

**REFERENCES**


