

1. Controlled Vocabularies in Context

A controlled vocabulary is an information tool that contains standardized words and phrases used to refer to ideas, physical characteristics, people, places, events, subject matter, and many other concepts. Controlled vocabularies allow for the categorization, indexing, and retrieval of information. This book deals specifically with controlled vocabularies related to cultural works—products of human creativity that have visual aesthetic expression. Such vocabularies are employed with the ultimate goal of allowing cultural works, images of cultural works, and information about them to be discovered, brought together, and compared for study and appreciation.

The intended audience of this book includes students, academics, and professionals in art museums, art libraries, archives, visual resources collections, and other institutions that catalog the visual arts, architecture, and other cultural objects. The audience may include systems developers who support these communities as well as consortia or other groups attempting to compile or use vocabularies for cultural materials. The topics discussed here may also be applicable to disciplines outside the visual arts.

The art and cultural heritage communities have increasingly made use of vocabularies and other standards as they seek to provide access to information that was previously held in paper files or isolated in local systems. Inspired by the power of online databases and the World Wide Web, professionals in the various art and cultural heritage communities now see the value of efficiently exchanging information with each other. Practical concerns and limited resources have proven the value of shared cataloging. In addition, the mission of many cultural heritage institutions has changed over the years to include dissemination of information to the public and to other institutions. Institutions are gradually becoming adept at utilizing appropriate information standards, such as *Categories for the Description of Works of Art (CDWA)* and *Cataloging Cultural Objects (CCO)*, as well as controlled vocabularies that are used or designed specifically for art and architecture, including the *Library of Congress Subject Headings (LCSH)* and the *Library of Congress Authorities*, the *Art & Architecture Thesaurus® (AAT)*, the *Union List of Artist Names® (ULAN)*, the *Getty Thesaurus of Geographic Names® (TGN)*,

Robert Chenhall's *Revised Nomenclature for Museum Cataloging*, the *Thesaurus for Graphic Materials (TGM)*, and the *Iconclass* system, among others.

Such data standards and controlled vocabularies take into account the unique nature of cultural information, which is characterized by conflicting opinions, changing interpretations, and information that must be expressed with nuance and indications of ambiguity and uncertainty. For example, scholars may disagree about the purpose of a given object or its date of creation, or a work may have been attributed to one artist in 1958 and then to a different artist in 2008 based on new analysis. Biographical information about an artist may be amended because of new research; the usage of a generic term such as *naïve art* might differ over time. The history of these changing opinions is valuable in itself, so earlier opinions and original information must be preserved.

1.1. What Are Cultural Works?

In order to understand the context of the vocabularies discussed here, it is first necessary to define the types of materials for which the terminology is created.

In this book, objects representing visual arts and material culture are called *works*. *Material culture* refers to art, architecture, and also more broadly to the aggregate of physical objects produced by a society or culturally cohesive group. *Cultural works* are the physical artifacts of *cultural heritage*, which encompasses broadly the belief systems, values, philosophical systems, knowledge, behaviors, customs, arts, history, experience, languages, social relationships, institutions, and material goods and creations belonging to a group of people and transmitted from one generation to another. The group of people or society may be bound together by race, age, ethnicity, language, national origin, religion, or other social categories or groupings. The works discussed in this book are cultural works, but they are limited to fine arts, architecture, and other visual art as described below.

1.1.1. Fine Arts

Fine arts include physical objects—such as drawings, paintings, and sculpture—that are meant to be perceived primarily through the sense of sight, were created by the use of refined skill and imagination, and possess an aesthetic that is valued and of a quality and type that would be collected by art museums or private collectors. In this book, conceptual art and performance art are included in the visual arts, but the performing arts and literature are not.

1.1.2. Architecture

Architecture includes structures or parts of structures that are made by human beings. Generally, it refers only to structures that are large enough for human beings to enter, are of practical use, and are relatively stable and permanent. Works of architecture are often limited to the built environment that is generally considered to have aesthetic value, is designed by an architect, and is constructed with skilled labor.

1.1.3. Other Visual Arts

In addition to fine arts and architecture, cultural works may include crafts, decorative arts, textiles, clothing, ceramics, needlework, wood-working, furniture, metalwork, decorative documents, vehicles, and other works noted for their design or embellishments and used as utilitarian items or for decorative purposes.

1.2. Creators of Art Information

In addition to the complexity inherent in art information itself, the issues surrounding the development and maintenance of such information are further complicated by the diverse spectrum of information creators, including museum professionals, librarians, archivists, visual resources specialists, art and architectural historians, archaeologists, and conservators. Users of the information may include all of these groups as well as the general public. While these communities share a vast overlap of required information about works, they also have various requirements and different cataloging and indexing traditions, as described below.

1.2.1. Museums

Traditional museums house collections of works of art, antiquities, or other artifacts that are displayed for public benefit. Art museum professionals may include registrars, curators, conservators, and other scholars in the fields of art and architectural history and archaeology. These are the people who acquire, catalog, care for, and research the history and significance of the works in their collection. They are accustomed to dealing with unique objects, unlike librarians, who typically catalog an item in hand as a nonunique representation of an intellectual work.

Unlike the library and archival communities, museums have historically recorded information about works using long-established local practices rather than a shared, standard set of rules. Even so, there was always a certain amount of consistency in the way that museums recorded information because it was based on common practice in

art-historical literature. However, consistency was uneven and unreliable, so the advent of data standards like *CDWA* and *CCO* provided much-needed written guidelines based on generally familiar practice in this community.

The standards and vocabularies required by the cultural heritage community must take into account the fact that the people who document the works typically derive much of the information directly from the objects themselves, rather than relying on other sources, as visual resources professionals must do. Therefore, rules must include instructions on, for example, not only how to record the dimensions of an object but also how to actually measure it. Unlike librarians, museum professionals usually deal with works that do not have the vital information printed or inscribed on the work itself. For instance, there is generally no title page or inscribed creator name on a museum object. It may be necessary for a museum to devise a title for the work, to establish the identity of the creator, or to estimate the date of creation through research and stylistic analysis.

Also in contrast to other communities, a museum actually houses and cares for valuable and unique works, requiring a great deal of administrative information, such as conservation and treatment history, exhibition history, provenance, and information concerning the specific circumstances of the excavation of an artifact. This community, as compared to librarians or visual resources specialists, requires more areas of the record in which to document detailed scholarly research, such as how a work fits into the evolution of an artist's style or details regarding why a work is dated to a particular year. Controlled vocabularies must provide names and terms to support these needs.

1.2.2. Visual Resources Collections

Visual resources collections maintain images that are typically collected to support the teaching and research requirements of universities, museums, or research institutions. Visual resources professionals are involved in the cataloging, classification, and indexing of images. They generally deal with slides, photographic prints, and digital images depicting art, architecture, or other subjects. They routinely catalog, manage, and store large numbers of images, often in the hundreds of thousands or even millions. Their work includes cataloging single items as well as sets of images.

Because their users will need to retrieve images based on the works depicted in them, the visual resources professional must catalog both the item in hand (slide, photograph, or digital image) as well as the art work or other cultural object depicted in it.



Fig. 1. Oil paintings such as this one are collected by art museums.

Vincent van Gogh (Dutch, 1853–1890); *Iris*; 1889; oil on canvas; 71.1 × 93 cm (28 × 36½ inches); J. Paul Getty Museum (Los Angeles, California); 90.PA.20.

Visual resources professionals were formerly called *slide librarians*. While the images they now deal with are of many media, these professionals are still often trained as librarians and may work in an image collection that is affiliated with or even located in a library. In addition, they are generally familiar with traditional museum cataloging. They have long been accustomed to using library standards and have been active in developing new standards and vocabularies that accommodate the unique requirements of image cataloging.

1.2.3. Libraries

Libraries are collections of documents or records that are made available for reference or borrowing. Librarians are professionals schooled in the cataloging and classification of books, journals, and other published textual materials. Since libraries may also collect rare books, prints, and

art, librarians are often called upon to catalog these items as well. They are guided by principles and practices originating from national institutions, such as the United States Library of Congress and the British Library. Their approach is based primarily on the concept that the item in hand is one of many of the same thing, not a unique item in itself. For this reason, data sharing among libraries has long been seen as economically advantageous, because copy cataloging is more economical than original cataloging.

The librarian's model of the world is codified in the *Functional Requirements for Bibliographic Records (FRBR)* model. In *FRBR*, a *work* is defined as an abstract notion of an artistic or intellectual creation (not analogous to the art community's *work*). The *FRBR expression* is the intellectual or artistic realization of a work; a work may have many expressions, such as in different languages. The *FRBR manifestation* is the physical embodiment of an expression of a work, such as a particular print run of a book. The *FRBR item* is a single exemplar of the manifestation, such as a specific book in hand, which is a physical object that has paper pages and binding (comparable to a unique *work* in art standards, but considered by *FRBR* to be only one of many identical items). The corresponding model for authority information is found in the *Functional Requirements for Authority Data (FRAD)*.

Librarians are accustomed to doing authority work and using controlled vocabularies. This community has a long tradition of following prescribed rules, striving for consistency, and using well-established standards such as the Machine Readable Cataloging (MARC) format and *Anglo-American Cataloguing Rules (AACR2)*, currently evolving into *Resource Description and Access (RDA)*.

1.2.4. Special Collections

Special collections contain rare or unique materials that are held by libraries or historical repositories but are typically not placed in public stacks. These materials may be available to the public only if special arrangements have been made in advance. The items may include rare books, manuscripts, personal papers, artworks such as prints, and other fragile or sensitive items. The people who work with special collections are often trained as librarians but occasionally as archivists, historians, or art historians.

1.2.5. Archival Collections

Archives are repositories for the noncurrent records of individuals, groups, institutions, and governments that contain information that is rare or of enduring historical value. Archival records are the products of everyday activity that are maintained to enable research. Documents represented in an archive may include administrative records, unpub-

lished letters, diaries, manuscripts, architectural drawings, architectural models, photographs, films, videos, sound recordings, optical disks, computer tapes or digital files, and other items.

The archivist's job involves the arrangement and description of these documents with the goal of maintaining physical and intellectual control of the materials. The work is done in accordance with accepted standards, such as Encoded Archival Description (EAD), and following practices of national institutions such as the U.S. National Archives and Records Administration (NARA). Many archivists have been educated as librarians or historians.

The methodology of the archivist emphasizes the function and provenance of archival materials. The archivist typically documents large groups, subgroups, collections, and series of items rather than individual works, creating finding aids that briefly detail the physical location of groups and individual works in the archive.

1.2.6. Private Collections

Private collections are aggregations of objects gathered by or for one or more people but are not intended to be accessible to the general public. Individual collectors, families, architectural firms, corporations such as banks, or others develop private collections. The expertise of the people who maintain such collections varies widely.

Private collections may include a variety of objects of the types that would otherwise be located in museums, archives, or libraries. Materials from private collections may sometimes be seen in exhibitions at publicly accessible institutions.

1.2.7. Scholars

Art and cultural heritage information may also be created by scholars or academics—often art historians or architectural historians who are associated with teaching institutions or museums, but are not trained as librarians, archivists, visual resources professionals, or museum professionals. The information may be collected during the course of research—for example, for the purpose of teaching or writing books, articles, or other publications. Scholars are now beginning to capture information about art and architecture in electronic form in order to organize or aid in their research.

1.3. Standards for Art Information

There are several types of standards used to record art information. Standards for data values provide the actual values to be entered in fields, including the vocabulary terms and allowable character sets. Controlled

vocabularies are standards for data values. They fit into the broader scheme of standards together with standards for data structure and for data content.

Standards for data structure dictate what constitutes a record. They define the names, length, repeatability, and other characteristics of fields and their relationships to each other. Examples are the MARC format and *CDWA*.

Standards for data content indicate how data should be entered, including cataloging rules and syntax for data. They may refer to standards for data values and standards for data structure. Examples of standards for data content are *AACR2* and *CCO*. For a typology of data standards, see the chapter by Anne Gilliland in *Introduction to Metadata*, edited by Murtha Baca.

1.3.1. Standards for the Creation of Vocabularies

While controlled vocabularies may function as standards for data values and be referenced in standards for data content, they themselves should ideally be constructed according to established standards for vocabulary creation. Institutions should use established vocabularies that are compliant with national and international standards. Furthermore, if a cataloging institution creates its own controlled vocabularies or adapts existing vocabularies to its local needs, it should consult these standards in order to make it easier to integrate its local vocabularies into a shared environment for search and retrieval.

The following standards for the creation of thesauri and other controlled vocabularies provide high-level guidelines regarding how a thesaurus should be structured, what kinds of relationships should be included, and how to identify preferred terms. The standards supplement each other in various areas, but where they overlap directly, they are generally in agreement. Thus, being compliant with one typically means being compliant with the others in most respects. More detailed rules for constructing vocabularies for art information may be found in **Chapter 7: Constructing a Vocabulary or Authority**, in *CCO* and *CDWA*, and in the more detailed rules of the *Editorial Guidelines* for the Getty vocabularies.

ANSI/NISO Z39.19-2005: Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

The National Information Standards Organization (NISO) is a nonprofit association accredited by the American National Standards Institute (ANSI). This publication discusses how to formulate preferred terms, establish relationships among terms,

and present the information in print and on a computer screen. It also discusses interoperability, methodologies for maintaining a thesaurus, and recommended features for thesaurus management systems.

BS 8723-1:2005, BS 8723-2:2005, BS 8723-3:2007, BS 8723-4:2007: Structured Vocabularies for Information Retrieval

This is a British Standards work published in four parts. Parts 1 and 2 include the basic principles of thesaurus construction, including facet analysis, presentation in electronic and printed media, thesaurus functions in electronic systems, and requirements for thesaurus management software. Parts 3 and 4 cover vocabularies other than thesauri and interoperability between vocabularies.

ISO 2788:1986: Documentation—Guidelines for the Establishment and Development of Monolingual Thesauri

This standard is an International Organization for Standardization (ISO) publication on the construction of monolingual thesauri. It includes guidelines for dealing with descriptors, compound terms, basic relationships, vocabulary control, indexing terms, display, and management of a thesaurus. Updates and additions to this standard are in development at the time of this writing, including *ISO/CD 25964-1: Information and Documentation—Thesauri and Interoperability with Other Vocabularies: Part 1: Thesauri for Information Retrieval*.

ISO 5964:1985: Documentation—Guidelines for the Establishment and Development of Multilingual Thesauri

This standard is intended as an extension of *ISO 2788*, the standard for monolingual thesauri. It includes guidelines for dealing with degrees of term equivalence and nonequivalence, single-to-multiple term equivalence, and thesaural displays.

1.3.2. Issues in Sharing Data

The various types of creators of information described above often wish to share data with each other or in a consortium. There are several steps involved with data sharing, including the extraction of data from a system, mapping data to another system or format, and delivering the data to the new environment.

Data standards and information systems are critical in data sharing. Standards are usually intended to be applicable independently of any particular automated system. However, in practical terms, the ability

of an institution to apply a standard depends in part on the system used to collect and store data. It is easiest to accommodate standards when an institution is building a new system for which requirements of the standard may be planned. Building or implementing a new system allows an institution the opportunity to use the standard as a starting point for incorporating the core fields, planning requirements based on the data model and editorial demands suggested by the standard, and implementing authority files and vocabularies.

However, most institutions must use existing cataloging systems. Sharing data first requires that the different institutions (or multiple departments in one institution) map fields in their existing systems to each other or to a common set of data elements, such as *CDWA*. Data exchange or metadata harvesting standards, such as *Dublin Core* or *CDWA Lite*, may be utilized.

After deciding upon common core (required) fields, the collaborators must agree that, in shared files, there is a range of acceptable ways for different institutions to record display information. This is necessary because it is very unlikely that there will be absolute consensus regarding how to display information. For example, institutions may vary in the way they wish to publish a display date of creation or a creation statement—using different syntax or vocabulary. This is typically acceptable within the parameters of the standards, provided that the information is indexed in a consistent way that allows access across the databases. The distinction between information for display and indexed information is discussed in **Chapter 2: What Are Controlled Vocabularies?**

Following cataloging rules—such as *CDWA* and *CCO*—and indexing using common vocabularies—ideally thesauri that link synonyms—comprise the most efficient course to ultimately achieving good access to the data. The thesauri should also be applied using strategies and interfaces that accommodate the various ways users may try to access the data. The thesaurus should provide end users access via synonyms and relationships between concepts.

In summary: When information providers at a museum or other cultural heritage institution begin the process of making information accessible across departments, between institutions, and for the general public, they must consider the following issues:

- They must decide which data elements are important to share.
- They must identify the audience for the shared information.
- They must use a technical standard for data exchange between systems, such as *Dublin Core*, *CDWA Lite*, or the *Visual Resources Association Core Categories (VRA Core)*.

- They must agree upon guidelines and rules for data content, such as *CCO* and *CDWA*.
- They must agree upon controlled vocabularies for ensuring consistency and coordination of data values.

This book deals primarily with the last issue: it seeks to explain what controlled vocabularies are and how to identify, use, and create controlled vocabularies for ensuring consistency and coordination in data values and for enhancing access for a wide range of users.