

8. Indexing with Controlled Vocabularies

In the context of this book, *indexing* is the process of evaluating information and designating indexing terms by using a controlled vocabulary that aids in finding and accessing the cultural work record. This indexing is done by human labor, as opposed to indexing resulting from the automatic parsing of data (*automatic indexing*) into a *database index*, which is used by a system to speed up search and retrieval. Indexing as described in this book is a conscious activity performed by knowledgeable catalogers who consider retrieval implications when assigning indexing terms.

8.1. Technical Issues of Indexing

When building a database and in the process of cataloging, it is important to employ the best design and editorial practice possible. However, if a cataloging or retrieval system is less than ideal, it will be necessary to adjust cataloging rules to accommodate the shortcomings of an information system or software, particularly concerning the application of controlled vocabularies and authorities.

As discussed in **Chapter 7: Constructing a Vocabulary or Authority**, it is critical to invest in both the data structure and the data used to populate the data elements in that structure; the data should survive through a succession of computer systems over time. However, in the real world of cataloging, technical concerns may limit or enhance cataloging in various ways. Ideally, the technical environment will not dictate limitations on good cataloging practice, but practice must sometimes be adjusted nonetheless. For example, if it is not possible to link to hierarchical authorities, it may be necessary for catalogers to index both specific terms and their broader contexts in each record to allow access.

8.1.1. Availability of Indexing Terms to the Cataloger

Ensuring successful indexing using a controlled vocabulary is determined in part by how the vocabulary is expressed to the cataloger or indexer. If possible, terminology should be customized for each particular field in the work record. For example, when filling in values for the Materials field, ideally, catalogers should *not* have access to the Styles and Periods

terms from the *AAT*, because excluding access to extraneous terms reduces the possibility for errors in indexing. However, access to terms should not be limited too narrowly. For example, a collage or other such work may be made of other works, so terminology generally reserved for Work Type (e.g., *photograph*) may be considered a Material in a collage.

Methods for applying vocabulary in the cataloging system may range from copying and pasting from online vocabulary sources to a thorough integration of one or more vocabularies in an information system. The copy-and-paste method is easy and typically inexpensive; however, there are caveats associated with it. Most notably, by copying and pasting terms, the link to the original vocabulary record and all of its variant terms and associated information is lost. In addition, it is not possible to automatically update the records in the future as the vocabulary changes over time. Integrating a controlled vocabulary into the editorial or cataloging system is a much more efficient way of incorporating vocabularies, either through the use of local authorities or by including the published controlled vocabularies in their entirety. Incorporating the vocabularies in the software allows access to variant terms and the unique numeric identifiers of the vocabulary, which accommodate updates to the terms in the system when the published controlled vocabularies issue updates.

Ideally, the system should allow the cataloger to use the preferred or any variant term in the same authority record to refer to the concept. In order to facilitate this, unique identifiers may be assigned to the individual terms, in addition to the unique identifier for the overall concept record.

Fig. 58. *AAT* terms in an editorial system illustrating how there are unique numeric identifiers for the overall record (Subject ID) and for each term (Term ID).

Subject ID: 300038715	Review: Finished	Merged: Merge				
Parent: 300142742	Record: Concept	Candidate: Non				
<i>Terms/Names</i>						
#	P	H	T	V	Terms/Names	Term ID D
1	<input checked="" type="checkbox"/>	C	D	U	lits \$02a la duchesse	1000038715 N.
2	<input checked="" type="checkbox"/>	C	AD	U	lit \$02a la duchesse	1000292667 N.
3	<input checked="" type="checkbox"/>	C	UF	U	beds \$02a la duchesse	1000138184 N.
4	<input checked="" type="checkbox"/>	C	UF	U	duchesse beds	1000138182 N.

8.2. Methodologies for Indexing

Institutions should adopt rules and methodologies for indexing work records that are appropriate to their collections and priorities.

8.2.1. Indexing Display Information

Retrieval issues should be considered when assigning terms and values to controlled fields. All important information contained in a free-text display field should be indexed in a controlled field to provide good access to the information. Display fields should generally utilize the preferred terms listed in index fields for consistency, especially if both are visible to end users. Display and indexing issues are defined in **Chapter 2: What Are Controlled Vocabularies?**

Display materials/technique (free-text):

brown ink and brown wash over black chalk underdrawing on white laid paper, with squaring, for an engraving

Indexing fields (repeating, controlled):

Material Names:

ink **Role:** medium

wash **Role:** medium

black chalk **Role:** medium

laid paper **Role:** support

Technique Names:

drawing

squaring

underdrawing

8.2.2. When Fields Do Not Display to End Users

Any field that contains a controlled number (e.g., Start Date), values controlled by pick lists (e.g., Preferred flag), or controlled values linked to authorities are indexing fields. Such indexing fields may or may not display to end users. If an indexing field in a work record will be displayed to end users, values that will not confuse or mislead the user should be used, not guesses or estimates based on incomplete data. For example, if a writing table seems to be constructed of a dark wood that the cataloger guesses may be walnut, the cataloger should not index the material as *walnut* without technical verification from the repository. Instead, the cataloger should index only what he or she knows, perhaps using the broad term *wood*.

Other fields may be used for searching but do not display to end users. For example, dates may be expressed in a free-text display field for end users, and indexed with Start Date and End Date fields, which do not display to end users. If fields do not display to end users but are used behind the scenes for retrieval, indexing may be done more broadly or liberally without fear of confusion. For example, for Start and End Dates, a broad span of time should be estimated, because estimating too narrowly will result in failed retrieval; however, estimating too broadly will result in some false hits in retrieval.

Display Date: ca. 1730–ca. 1750

Start: 1725 **End:** 1755

Display Date: 17th century

Start: 1600 **End:** 1699

Display Date: New Kingdom, 18th dynasty (1404–1365 BCE)

Start:–1404 **End:**–1365

8.2.3. Specificity and Exhaustivity

Applying indexing terms involves consideration of the precision and quantity of terms applied to a particular field in the work record; in cataloging, these characteristics are known as *specificity* and *exhaustivity*. *Specificity* refers to the degree of precision, or *granularity*, used in assigning terms. For example, the cataloger would ideally choose the most specific term to describe a work type, such as *amphora*, rather than the more general term, *storage vessel*. *Exhaustivity* refers to the degree of depth and breadth that the cataloger uses in description, typically expressed by using a larger number of indexing terms.

In order to ensure consistent indexing by catalogers, guidelines should be established regarding the number of terms to be assigned and the method to be used for analyzing a work to determine indexing terms for each field.

Catalog records are more valuable to researchers if they are indexed with a greater level of specificity and exhaustivity. However, practical considerations often limit the ability of cataloging institutions in assigning large numbers of terms to each field of every work record. Is it useful to index every aspect of the work? If not, where do you draw the limit?

8.2.3.1. Specificity Related to the Authority Records

Do specific details of the authority record need to be included in a work record if those topics are already part of the authority record? Generally, those aspects that are apparent, important, unusual, or particular in

the work being cataloged should be indexed, even if they are also in the authority record.

One consideration is whether the particular information system being used will link a specific term to its broader context and synonyms in an authority. One primary purpose of the authority is to reduce the cataloger's labor in linking all variant names and broader contexts for a concept to every work record. However, if the authority does not do this, the broader context and synonyms in the work record should be included.

Assuming that the authority is linked to the work record, there is no need to repeat basic information, such as names. The issue is complicated by the fact that not all aspects of a given authority record will necessarily apply to the work being indexed. Even though the authority record for the subject *Adoration of the Magi* may include the names of the magi, the names of the gifts, the types of animals generally present at the scene, the symbolic significance of the scene, etc., not every depiction of the *Adoration of the Magi* will include all of these topics. Therefore, the indexing of this subject for a particular work should focus on the major aspects of the subject as portrayed in that specific work.

8.2.3.2. General and Specific Terms

In certain fields, it is advantageous to include both general and specific indexing terms, particularly when the general and specific terms are not linked hierarchically in the authorities. For example, with subject indexing, it is useful to label a general subject (e.g., *landscape* or *portrait*) for overall access, in addition to specific terms that name the location or person depicted. For example, an authority record for a geographic place is usually linked to the broader geographic contexts of that place but not to the concept of *landscape*. Without this general designation, the work cannot be retrieved in a search by general subject classification.

Subject:

landscape

poetry

Longqiu Waterfall, Yandang Mountain (Zhejiang province,
China)

waterfalls

pool

human figures

mountains

clouds

pine trees

literati (Chinese scholars-artists)

8.2.3.3. Preferred or Variant Terms

The term that best fits the characteristic being indexed should be used. Ideally, system constraints do not require the use of only the preferred term or descriptor for indexing. This is particularly important when end users can see the terms. In some cases, a singular term may be appropriate, while in others the plural makes more sense. In other cases, the cataloger may wish to index with a *used for* term, a historical term, or a descriptor in another language. So long as all of these terms are linked to the same vocabulary concept record, the cataloger should be able to use any that fits the situation at hand.

8.2.3.4. How Many Terms

Rules regarding the number of terms to assign and the method of analysis that is most appropriate to local needs should be established. Strategies should be devised that allow catalogers to be thorough, without expending more time than necessary, so that production quotas can be met.

To ensure that the entire work is indexed evenly and consistently with other works in the collection, guidelines should be set for the catalogers to treat the work systematically. Catalogers should index by whatever is most appropriate to a given field in the work record, whether that be by moving front to back, top to bottom, most important to least important, or chronologically. For instance, they could index materials according to the level of importance of the materials or the order in which the media were applied. For example, for a table, the mahogany used as the primary material would be more important than the brass fittings on the feet; for a design drawing, the squaring in pencil would be applied before chalk outlines of figures, with the white highlighting applied last. For the subject of the work, assigning indexing terms according to the following three levels of subject analysis is appropriate: description of the generic subject, identification of the specific subject, and interpretation of symbolic meaning contained in the subject. See *CDWA* and *CCO* for further suggestions regarding how to index specific fields in the work record.

8.2.3.4.1. How to Establish Core Elements

How much information should a catalog record contain? Standards such as *CCO*, *CDWA*, and *VRA Core 4.0* can provide guidance for core data. Not every field in the work record needs be filled with the maximum number of indexing terms. The focus of cataloging should be twofold: promoting good access to the works, and providing clear, accurate descriptions that users will understand. This can be achieved with either a full or a minimal cataloging record, so long as the cataloger follows standards

and the descriptive cataloging and indexing is consistent from one record to another.

8.2.3.4.2. *Minimal Records*

Minimal records contain the minimum amount of information in the minimum set of elements, as defined by the cataloging institution. What comprises a minimal work record for the institution must be decided; this includes which fields are required, which are required if known, and which are optional. All required fields must be included for every record. Even when it appears that two fields overlap, if they are both required, values should be included in both. For example, if the Subject of a utilitarian work is the same as the Work Type or Title, the term should be repeated in all required fields. Noting the values in fields or metadata elements dedicated specifically to certain content elements ensures that the data is consistently recorded and indexed in the same place, using the same conventions for all works in the database.

8.2.3.4.3. *Missing Information*

What should the cataloger do if core information is limited or unavailable? Occasionally, data for any element may be missing during the cataloging process. It is up to the cataloging institution to determine how to deal with missing data. Default values should be established to index unavailable but required fields, so that it is apparent to users that the data is unavailable for a particular record (as opposed to the field having simply been skipped).

Possibilities for dealing with missing data include the following:

- (1) using a value such as *unavailable*, *unknown*, *not applicable*, *destroyed*;
- (2) making the value NULL on the database side; or (3) leaving the field blank entirely and supplying data for missing values at the public access end (e.g., if the creator is unknown, rather than filling in the value *unknown Celtic* in the Creator field, it could be left blank in the local database but filled with the value *Celtic* from the Culture field in displays). How these defaults are implemented is a local decision that may vary from institution to institution. See also the discussion in 8.2.3.5.5. **Expertise of Catalogers and Indexers.**

Descriptive Note: Location unknown; formerly at Aghia Triadha (Iraklion department, Crete, Greece)

Current Location: unknown

Former Location: Aghia Triadha (Iraklion department, Crete, Greece)

Descriptive Note: Destroyed in 1966; formerly Gabinetto Disegni e Stampe (Uffizi, Florence, Italy)

Current Location: destroyed

Former Location: Gabinetto Disegni e Stampe (Uffizi, Florence, Italy)

8.2.3.5. Size and Focus of the Collection

The level of homogeneity of a collection may influence the specificity and exhaustivity of indexing. The more similarity there is among items in the collection, the more specific indexing terms need to be and the more granularity should be used in indexing the vocabulary or vocabularies. For example, to make meaningful distinctions between items in a specialized collection of tapestries, the terminology used to index them should be much more specific than that used for a few tapestries in a more general collection.

The size of the collection may play a role in limiting the levels of specificity and exhaustivity employed by any given institution. An institution that is cataloging a large collection may not have the need or resources to record extensive and specific information for every work. On the other hand, a small institution may be constrained by not having access to specific information; for example, a repository may not have a conservation laboratory to supply accurate analysis of materials.

8.2.3.5.1. Different Works Require Different Indexing

Different levels of specificity and exhaustivity may be dictated by the works themselves. For example, one sculpture may have been cast of a single material, so simply stating the material is sufficient (e.g., *bronze*), while another sculpture may be composed of various materials that should be indexed (e.g., *fiberglass and resin on wire mesh*).

8.2.3.5.2. Cataloging in Phases

Cataloging in phases may influence the way in which terms are assigned. An institution may index a few broad or important elements in minimal records to gain control of a collection and then go back in a second pass to add more specificity and greater numbers of terms.

8.2.3.5.3. Indexing Groups vs. Items

An archival group (or *record group*) is an aggregate of items that share a common provenance. Group-level cataloging focuses on the description of coherent, collective bodies of works. Indexing should emphasize the characteristics of the group as a whole, highlighting the unique and distinctive characteristics of the most important works in the group.

If an institution is cataloging groups of works rather than individual items, an appropriate methodology of assigning indexing terms must be established. The two most common methods are to assign terms that refer to all items in the group, or to assign terms that refer to only the most important items in the group. If the items will eventually be cataloged individually, broad terms or a miscellaneous term applicable to the group, such as *various materials*, should be assigned, and as a second step, narrower terms appropriate to individual items should be assigned in the individual item records.

Title: Group of Points from Bannerstone Site

Work Types:

arrowheads

kirk points

netting

Materials and Techniques: flint, vitric tuff, and rhyolite

Indexing Materials:

flint

tuff

rhyolite

Description: 152 design drawings and models for the East Building project that I. M. Pei & Partners gave to the archives of the National Gallery of Art in 1986.

Work Types:

design drawings

models

Materials and Techniques: various materials

Indexing Materials:

various

8.2.3.5.4. Expertise of End Users

What types of terms will the intended end users be familiar with? A major challenge for catalogers is that indexing terms should accommodate the expectations and knowledge of the intended users of the information system. Many institutions must satisfy a wide range of users, from the scholarly expert to the novice visitor to a museum Web site. Ideally, separate—but related—vocabularies would be used for indexing and retrieval; however, this is not possible for most institutions. If end users will be exposed to the original specialist vocabulary terms, rather than utilizing an intermediary vocabulary designed to bridge the gap between nonexpert and expert users, nonexpert terms should be included along with expert terms in indexing.

A collection may someday be retrieved in a consortial environment, for which the indexing terms may need to be broader or narrower than in a local environment. Indexing terms will need to be specific enough to allow the records to remain meaningful in the context of a larger information repository.

8.2.3.5.5. *Expertise of Catalogers and Indexers*

The indexing and other content of work records necessarily reflect the level of subject expertise of the catalogers. Catalogers may not be experts on the works being cataloged. In general, catalogers of visual resources collections and others who are cataloging works not held in their own institution do not have access to some information about the work.

8.2.4. Indexing Uncertain Information

It is desirable to be specific, so a good general rule is, if you know something, include it. However, an equally important axiom is this: if you do not know, do not guess.

Data should only be indexed when authoritative sources for the information are available. It is important to consider the reliability and idiosyncrasies of the sources and to analyze what is true and what is only possibly or probably true. When important information is described as uncertain by reliable sources, the information may still be recorded, but with an indication of uncertainty or approximation in a Descriptive (Scope) Note or Display Date field (e.g., *ca.* or *probably*).

Materials/Techniques Description: probably soft paste
porcelain

Indexing Material Name: soft paste porcelain

Catalogers should never use a specific term unless they have the research, documentation, or expertise to support that use. A broader but accurate term should be used in place of an incorrect specific term. It is better to be general and correct than specific and incorrect. For example, a cataloger should index the broader material *stone* rather than the specific *banded slate* if he or she is unsure of the specific material. Rules should be established regarding default values for required elements for which no information is available.

Another option is to index multiple values for uncertain information, explaining any ambiguity and nuance in display fields. For example, if scholarly opinion is divided regarding whether a figure represents *Zeus* or *Poseidon*, the names of both gods should be indexed as subjects for retrieval, and the situation should be explained in a note. If sources disagree about whether an artist was French or Flemish, index both nationalities and explain the discrepancy in a note.

Display biography: French or Flemish draftsman, active by 1423, died 1464

Nationalities:

French
Flemish

Descriptive Note: It is uncertain if the work was used as a table or a stool.

Work Type:

furniture
table
stool

Fig. 59. If the cataloger is uncertain of information, such as the composition of the upholstery fabric of a piece of furniture, it is better to choose an indexing term that is broad and correct than to use a specific term that is wrong (e.g., use *cloth* rather than *silk*). For the chair illustrated here, the repository has done analysis, and the specific material is indeed known.

Frames: attributed to François-Honoré-Georges-Jacob-Desmaller (French, 1770–1841); upholstery: Beauvais Tapestry Manufactory (French, founded 1664); *Armchair (Bergère)*; ca. 1810; mahogany and beech, with gilt-bronze mounts, silk and wool tapestry upholstery; J. Paul Getty Museum (Los Angeles, California); 67.DA.6.



8.2.4.1. Knowable vs. Unknowable Information

There is a difference between *knowable* and *unknowable* information: one refers to information that is simply unknown to the cataloger due to lack of expertise or access to research and publications, while the other refers to information that is debated among scholars or unknown despite expert analysis. To maintain high-quality, reliable, and professional catalog records that are in keeping with standard art historical practice, this distinction should be kept clearly in mind during indexing.

8.2.4.1.1. Knowable Information

For information that is *knowable* but simply unknown by the cataloger, a more general term should be used or the information should be omitted. Most catalogers are not experts on all the works they catalog, but information in a catalog record should only be supplied by experts and authoritative sources. When a lack of knowledge is due to ignorance regarding a particular issue, the cataloger's assumptions should not be indexed. In such cases, terms such as *probably* or *perhaps* should not be used, because this would imply that scholars or other experts are uncertain.

For example, if a source describes the material of a Louis XVI chair as *gilded beechwood* but does not identify the material of the upholstery, the upholstery should not be indexed as *silk* or even described as *probably silk*, even if it appears to be so. The fiber content of that upholstery is knowable by technical analysis and perhaps may even be published in other sources. If an end user were to read *probably silk*, he or she should be able to assume that technical analysis was inconclusive or impossible, not that a cataloger was making a guess. In this case, it would be best to index *gilding* and *beechwood* but avoid indexing or describing the upholstery at all, because there is no source of information for the upholstery.

8.2.4.1.2. Debated Information

For information that is *unknowable* because current authoritative sources indicate that scholars disagree, the historical or archaeological information is incomplete, or interpretation of the information differs in reliable sources, multiple possibilities should be indexed with words such as *probably* or *perhaps* in a note explaining the ambiguity or uncertainty of prevailing authoritative sources.

When sources are in disagreement, the preferred information is that which is supported by general scholarly opinion or found in the most recent authoritative sources. If scholarly opinion is evenly split or both sources are equally reliable, neither view can be preferred; the debate should be explained in a note and both possibilities should be indexed.