Chapter 5 Supplement
Controlled Vocabularies for Improved Resource Discovery

Questions for Review, Study, or Discussion

1. In what ways do controlled vocabularies improve resource discovery for users? What is one important difference between the search and browse methods of information seeking? What happens when metadata creators enter inconsistent values into the same field in different metadata records? What is a "split file?"
2. How do controlled vocabularies help deal with *ambiguity* and *synonymy* in natural human language? What is *disambiguation*?
3. In what ways may controlled vocabularies express the three primary semantic relationships: *equivalence*, *hierarchical*, and *associative*?
4. What six types of controlled vocabularies are covered in Section 5.2 of the text? What are some of the most important characteristics of each? What are some specific examples of each? (with the exception of synonym rings, which do not entail established and shared vocabularies).
5. Why is the assertion made that the *list* is probably the most commonly used form of controlled vocabulary in a majority of cultural heritage digital collections? In what sense is this the case?
6. What controlled vocabularies commonly recommended for use with DC, MODS, and VRA are listed on the sidebar on page 140?
7. What are some primary similarities and differences between the Library of Congress *Thesaurus for Graphic Materials* and the Getty *Art and Architecture Thesaurus*?
8. In what cases might an organization choose to create one or more of its own controlled vocabularies?
9. What is the difference between the so-called *top-down* and *bottom-up* methods of controlled vocabulary creation, as those terms are used in this chapter? What are typical steps involved in the bottom-up process of creating a controlled vocabulary?

Recommended Readings and Resources for Reference or Further Study

- Broughton, Vanda. 2006. Essential Thesaurus Construction. London: Facet.
- Leise, Fred, Karl Fast, and Mike Steckel. 2002. "What Is a Controlled Vocabulary?" *Boxes and Arrows*. Last modified April 7. http://www.boxesandarrows.com/view/what_is_a_controlled_vocabulary_. See also linked articles: "Creating a Controlled Vocabulary" and "Synonym Rings and Authority Files."
- NISO (National Information Standards Organization). 2005. *ANSI/NISO Z39.19-2005 - Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Bethesda, MD: NISO Press. http://www.niso.org/kst/reports/standards?step=2&gid=&project_key=7cc9b583cb5a62e8c15d3099e0bb46bbae9cf38a
- The Getty Research Institute. *Art & Architecture Thesaurus Online*. The J. Paul Getty Trust. http://www.getty.edu/research/tools/vocabularies/aat/index.html.
- Library of Congress Prints and Photographs Division. Thesaurus for Graphic Materials I: Subject Terms (TGM I). http://www.loc.gov/rr/print/tgm1/.
- Hedden, Heather. 2010. *The Accidental Taxonomist*. Medford, NJ: Information Today.

- Hodge, Gail. 2000. *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files*. The Digital Library Federation. http://www.diglib.org/pubs/dlf090/dlf090.pdf.

Exercises

Recommended Exercises

1. Apply controlled vocabularies to the Dublin Core records created for the Chapters 3 and 4 exercises. Although this more often than not done in practice, for the sake of the exercise, use two different subject vocabularies for expressing the same concept. If applicable to your resources and application scenario, search for and enter terms for the same concept from both the TGM and AAT vocabularies.

2. Different controlled vocabularies sometimes use identical or near identical terms to represent the same or similar concept, type of resource, genre, person, language, date, and the like. In this sense, although differences in capitalization may differ, we can still regard the term as essentially the same. But sometimes the actual terms differ in more than just lexicography. There are various reasons for this. Some vocabularies, for example, include a larger number of terms and therefore have a greater degree of granularity or specificity for expressing subject concepts more specifically than others. The following examples have been selected because they have differ among vocabularies. How would you represent each of the following using the controlled vocabularies indicated? You will need to search these vocabularies online and find the form of term, name, or code used in each.
    A. Subject concept: *statues*; in TGM, AAT, and LCSH. Also notice what semantic relationships (equivalent, broader, narrower and related terms) are given for the term in each of the following these three vocabularies.
    B. Form/material type term for *etchings* in the sense of prints made from an etched printing plate in TGM and AAT.
    C. The name for the Greek painter El Greco (Domenico Theotokopoulos) in LCNAF, ULAN, and the VIAF (http://viaf.org/viaf/). Note that the VIAF includes three different forms used in various different national library authority files; identify which countries/states use which forms by holding your mouse cursor over the flag next to each. What advantages might the VIAF have over the other two name vocabularies?
    D. Type of resource term for *interactive multimedia* in MODS typeOfResource and DCMIType vocabularies (see http://www.loc.gov/standards/mods/mods-dcsimple.html).

3. Imagine a scenario in which you might need to create your own controlled vocabulary for a digital collection. Describe the scenario and why an established vocabulary would not suffice. What kinds of terms would you include? Give some specific examples. Would you incorporate any semantic relationships?

Suggestions for Instructors

- Specify which vocabularies students are to use for which metadata elements/fields in the exercises associated with Chapters 3 and 4 by including them in an application profile document.
- Be prepared with your own answers to exercise 2 above.

Additional Exercise Ideas

1. Find some examples of controlled vocabularies used in existing records in one or more existing digital collections. Can you tell if they are widely-established or locally-developed vocabularies? Find examples of flat lists. Try also finding examples of taxonomies and of full-fledged thesauri implemented in a digital collection content. How are the semantic relationships handled for users?