Subject Indexing and Cataloging: Controlled Vocabularies

Melissa Foss

University of South Florida

LIS 6711: Organization of Knowledge I

Fall Semester 2013
Subject Indexing and Cataloging: Controlled Vocabularies

Authority control is “[d]ecisions about which terms will be used to refer to authors, titles, or subjects” (Rubin, 2010, p. 132). Authority control uses controlled vocabularies, especially in the area of subject terms or headings (Rubin, 2010, p. 132). When dealing with subject access, there are typically three steps: “(1) determining the overall subject content of the item being cataloged, (2) identifying multiple subjects and/or subject aspects and interrelationships, and (3) representing both in the language of the controlled vocabulary at hand” (Chan, 2007, p. 208).

Controlled vocabularies or controlled indexing languages are systems “in which both the terms that are used to represent subjects, and the process whereby terms are assigned to particular documents, are controlled or executed by a person” (Rowley, 1994, p. 109). The authority for the terms that are to be assigned to index a particular document is derived from subject heading lists or a thesaurus (Rowley, 1994, p. 109–110). Both the person indexing the document, and thereby assigning the terms to the document, and the searcher are expected to use the terms reflected in the lists (Rowley, 1994, p. 110).

Controlled vocabularies “are vital for effective searching and for the collocation function of the library catalog” (Rubin, 2010, p. 132). They are used to address synonymy, hierarchical relationships, associative relationships, and homographs (Rubin, 2010, p. 132–133). Many words mean the same thing (synonyms) and when organizing information, it is important to be consistent in the terms that are being used to make retrieval of information more effective (Rubin, 2010, p. 133). Controlled vocabularies assist with this issue by choosing a term and applying that term consistently when indexing materials on a particular topic. Additionally, the individual selecting the particular terms can also create “see” references in order to provide additional access points for a synonym (Rubin, 2010, p. 133). Controlled vocabularies can provide the hierarchy of related terms by “reveal[ing] when a topic identified by a particular term
Subject Indexing and Cataloging: Controlled Vocabularies

is part of a larger concept or when it can be narrowed further” (Rubin, 2010, p. 133). Similar to the assistance controlled vocabularies can render in terms of hierarchical relationships, controlled vocabularies can also identify related concepts and terms to help broaden the search (Rubin, 2010, p. 133). Additionally, some words that are spelled the same way are actually two completely different concepts; controlled vocabularies help “reveal this ambiguity and refer information seekers to the appropriate terms” (Rubin, 2010, p. 133). If there were not controlled vocabularies, any number of terms could be used to index a document. It would be necessary for a researcher to try to think of all the possible terms for a subject in order to make sure that all of the relevant resources are being returned in the search (Rubin, 2010, p. 133).

Types of controlled vocabularies include thesauri, subject heading lists, and ontologies (Taylor & Joudrey, 2009, p. 334). There are similarities and differences among the various types of controlled vocabularies. Subject heading lists and thesauri are both used to provide access to information by using “consistent and reliable rather than uncontrolled and unpredictable” terms (Taylor & Joudrey, 2009, p. 335). Subject heading lists and thesauri also use hierarchical structures and give the preferred terms as well as providing context for related, broader, and narrower terms (Taylor & Joudrey, 2009, p. 335). However, there are differences between the two; thesauri have single terms or bound terms, which cannot be separated. Thesauri often only have single terms for broadening or narrowing rather than phrases and multiple terms (Taylor & Joudrey, 2009, p. 335). Thesauri tend to relate to a specific area rather in comparison to the broad and general scope of subject heading lists.

Ontologies, on the other hand, while similar in reflecting multiple terms to express a concept and reflecting references and relations to other terms, are different because they do not have an authorized or preferred term (Taylor & Joudrey, 2009, p. 335). There does exist
argument that ontologies do not actually meet the definition of controlled vocabulary, unless the
definition is expanded to include situations where terms are brought together to reflect concepts
and relationships, but does not include preferred terms (Taylor & Joudrey, 2009, p. 336).

Examples of thesauri include the Art & Architecture Thesaurus (AAT), which aids in
accessing (by specified terms) cultural heritage information, and the Thesaurus of Educational
Resources Information Center (ERIC), which provides access to education-related works (Taylor
manuals, computer files, and many, many other resources (Taylor & Joudrey, 2009, p. 256).
Additionally, MeSH can be considered to be a thesaurus. WordNet and the Unified Medical
Lanugage System (UMLS) are examples of ontologies. WordNet provides nouns, verbs,
adjectives, and adverbs relative to certain concepts in groups and does not provide a particular
preferred term (Taylor & Joudrey, 2009, p. 256). The UMLS does not consider itself and
ontology, but its set up and characteristics make it fit the concept (Taylor & Joudrey, 2009, p.
359).

Relative to subject heading lists, the subject heading list that is the most prevalent,
perhaps the most widely accepted and used, is the Library of Congress Subject Headings (LCSH)
(LOC, last updated June 28, 2013). Part of the reason for the wide use of LCSH is that it is
general and is not relegated to one particular discipline (Rubin, 2010, p. 135). LCSH provides
subject listings that will provide items regardless of the particular discipline; however, a search
using subject headings will also provide the classification information for that subject because of
134–135). Although widely used, LCSH is not without its critics; critics of LCSH have voiced
concern that the terms in LCSH are culturally biased, and that some of the subject headings were
Subject Indexing and Cataloging: Controlled Vocabularies

too formal and would not be helpful to people using popular search terms (Rubin, 2010, p. 135). The LOC reviews proposals submitted for additions and changes to the LCSH regularly to help address issues that arise in the subject headings (LOC, last updated June 28, 2013), this also includes work to update LCSH headings that are affected by RDA (LOC, 2013). To assist individuals in need of information regarding LCSH, the Library of Congress provides PDF versions of their subject heading lists and subdivisions (LOC, last updated Aug. 1, 2013).

Other examples of a subject heading lists are the Sears List of Subject Headings and the Medical Subject Headings (MeSH) (Rubin, 2010, p. 135). The Sears list main users are smaller libraries and schools, and its structure and terms tend to be of a less complex nature when compared to LCSH (Chan, 2007, p. 268). Sears subject headings do not include as much specificity as LCSH (Taylor & Joudrey, 2009, p. 341). The National Library of Medicine created MeSH and it is a very technical list of subject headings that are used in indexing medical documents (Rubin, 2010, p. 135). The National Library of Medicine actually considers MeSH to be a thesaurus in particular based on its strict structure and focus on the medical field (Taylor & Joudrey, 2009, p. 352).

Although the use of controlled vocabularies is often associated with their use in libraries and with traditional library materials, controlled vocabularies are used in many areas. For example, Roszkiewicz (2005) discusses the use of controlled vocabulary in a business. Roszkiewicz discusses the various types of controlled vocabularies, including thesauri and ontologies (2005, p. 13–14). A controlled vocabulary can assist a company when searching and retrieval is an important service (Roszkiewicz, 2005, p. 13). Many companies use a controlled vocabulary without knowing that they do so and implementing a controlled vocabulary can be quite beneficial to the company (Roszkiewicz, 2005, p. 13). There are also sources, such as the
Subject Indexing and Cataloging: Controlled Vocabularies

Taxonomy Warehouse, from which a company can license a controlled vocabulary and receive updates; this is done through licensing (Roszkiewicz, 2005, p. 13). Roszkiewicz indicates that all of the tools needed to develop, implement, and maintain controlled vocabularies are available and that “[d]esktop equivalents of these tools, suited to the rest of us who are not library scientists, are slowing emerging” (2005, p. 16), making it a matter of time before controlled vocabularies become more widely used in various business arenas. Tena, Diez, Diaz, and Aedo (2013) discuss the use of controlled vocabularies in the area of use cases for web tasks. Use cases indicate information about software systems and user needs and it is believed that the use of a controlled vocabulary in this area would improve the standardization of the use cases (Tena et al, 2013, p. 1580). Standardizing the terminology would allow for individuals with an understanding of the problems related to user interface design to more effectively communicate by using accepted terminology (Tena et al, 2013, p. 1580).

Controlled vocabularies are also used in order to assist in issues related to accessing musical works by subject (McKnight, 2012). With musical works, problems arise when determining whether access to the materials is based on what the work is about or what the work is and “[m]ost U.S. libraries still use the Library of Congress Subject Headings (LCSH) to assign subject terms for music materials, even though studies have demonstrated the ineffectiveness of these headings” (McKnight, 2012). Resource Description and Access (RDA) is the new standard for describing music materials and this has opened an opportunity to determine whether there is a better means for providing access to musical works (McKnight, 2012). In order to assist in better search and retrieval results for musical works, the Music Library Association and the Library of Congress Music Genre/Form Project have collaborated in order to develop “a structured vocabulary of terms for genre/form and medium of performance
Subject Indexing and Cataloging: Controlled Vocabularies (McKnight, 2012). The resulting controlled vocabulary from this collaboration is intended to improve access to music (McKnight, 2012).

Although controlled vocabularies are used in order to alleviate some problems that can be involved in the retrieval of information, there are issues that arise with controlled vocabularies. Included in those issues is the expense of using controlled vocabularies because it requires a person to assign the terms to the specific document (Rowley, 1994, p. 110) and criticism over the terms used in the subject headings reflect cultural or racial bias and being too formal for people conducting searches based on popular terms (Rubin, 2010, p. 135). An alternative means of indexing is the use of natural language rather than controlled vocabularies. “Natural indexing languages are not really a separate language, but are the ‘natural language’ or ordinary language of the document being indexed” (Rowley, 1994, p. 110). With natural language, any term in the document can be a potential index term (Rowley, 1994, p. 110). Natural language looks to be able to interpret what a user is seeking even though they are not using particular terms and are free searching (Taylor & Joudrey, 2009, p. 361). A potential advantage of natural language searching is that the indexing can be done by either a person or a computer (Rowley, 1994, p. 110), which means that the expense of human labor in indexing could potentially be avoided or decreased with use of natural language indexing. The fear of removing the use of subject headings or controlled vocabularies is that a researcher using natural language would be more likely to miss results. A 2005 study by Gross and Taylor came to the conclusion aht if subject headings were no longer used, users would miss over one third of the potential hits they would retrieve given the current subject headings (p. 223). The study indicated that users would also lose the other advantages of controlled vocabularies such as showing related terms, synonyms, and cross-referencing (Gross & Taylor, 2005, p. 223). There has been a persistent debate for a
Subject Indexing and Cataloging: Controlled Vocabularies

long time over the use of controlled vocabularies versus the use of natural language in indexing (Rowley, 1994), and that debate continues on today.
Subject Indexing and Cataloging: Controlled Vocabularies

References


Subject Indexing and Cataloging: Controlled Vocabularies


