2010-01-01

Moving Authority Control from Manual to Automated

Shannon Hoffman

MOVING AUTHORITY CONTROL FROM

MANUAL TO AUTOMATED

Shannon Hoffman

Brigham Young University

Thesis Statement: Authority control is very expensive; however, a controlled vocabulary is important for the users of a library catalog. Ways must be found to do authority control faster, cheaper, and better if we are to maintain a high-quality library catalog.

Moving authority control from manual to automated

Is the high cost of authority control overshadowing the needs of the patron? Has Cutter's rules out lived their purpose and been replaced by full text and computer searching? No our patrons still need to find a book by author, title or subject. They still need to find what the library has access to by a given author, or about a given subject or literature. The only thing that has really been changed is the number of places and way that information can be accessed. At present, authority control is a vast and expensive undertaking whether done by the library or by a vendor. New names and subjects are added or changed every day in the authority files, databases, and thesauri created and maintained by the Library of Congress, other libraries, and organizations. The maintaining of these authority databases require an in-depth knowledge and understanding of the standards used. Working as an authority control librarian for nineteen years has given me a broad understanding of authority control as it applies to The Library of Congress name and subject authority file and the MARC Standards for bibliographic and authority record; Anglo-American Cataloguing Rules, Second Edition (AACR2); Library of Congress Rule Interpretations (LCRI) and the Library of Congress Subject Cataloging Manual (SCM).

In the process of investigation the problem of how to have a more accurate, less expensive, timely process of authority control, I volunteered as a consultant for Backstage Library Works (our authority control vendor), who were in the process of developing a new authority control system. As a library consultant, my job was to help the developers understand and correctly apply the cataloging and authority standards. My hope was that a better understanding of the vendor's system would allow for streamlining of our workflow, which would permit us to use our time, employees, and the vendor more proficiently.

While consulting with the programmers, I was introduced to a different paradigm for authority control, opening my eyes to the road blocks created by the cataloging standards and rules. If we are to use controlled vocabulary in our libraries, we need to be able to catalog materials once and then let automation handle the maintenance of the files. Unfortunately,

because of the present standards and rules many changes to headings may require human intervention.

Working at Backstage, it became clear that many of the cataloging rules originated from constrains of the space on the 3 x 5 card. Space constrains have now been lifted by technology, however, in many cases the rules and standards have not changed, and do not work well with the automated processes.

Cataloging requires human knowledge, education and subject expertise but once the cataloger has cataloged the material and created the necessary bibliographic and authority records, an automation process should be able to do the required maintenances. The maintaining of the library catalog is affected by our changing world. As our language changes new words are added, new meaning is applied to old words, words become politically incorrect; authors change their names; new authors may have the same name as another author; and sometimes mistakes need to be corrected. However, once the records are entered, changes should be handled by an automated process.

As a library community, we need to look at places where our standards and process do not allow computers to update our bibliographic and authority records, and then ask how the rule or process can be updated. We need to work with our systems, vendors, and other libraries to make necessary changes to the standards and processes; to leave behind the antiqued 3x5 card and other old rules that impede our embracing the automated world. We must take a close look at the standards in MARC21, AACR2, SCM, and RDA to see what changes are necessary to facilitate the use of technology and maintenance of the catalog.

Space was a big problem when it came to placing all the needed information on a 3 X 5 card. When we first went to computers the cost of space was still a problem. However, the cost of space is no longer a problem. So why are we still dealing with rules that were made because of space constraints? Example of rules that are entirely about space is the rule of three and the rule of four.
Rule of three is if the material is about more than three subtopics assign a broader headings for the topic. Rule of four is if the topic is a very broad topic then you can assign four headings. (see SCM, H180)

If you are cataloging a book about dachshunds, coonhounds, basset hounds, and foxhounds the rule of three would apply to this book and the cataloger would assign the

broader subject term "hounds" to the book. The rule of four would be applied to a book that is about dachshunds, foxhounds, golden retrievers, coonhounds, and poodles to which the heading "dogs" would be assigned. However, if a patron wants information on dachshunds they would not find all the information the library owned by looking under the heading "dachshunds". To find all the material, they would be required to look under "dachshunds," "Hounds", and "Dogs. However, the patrons is inclined to believe that he will find all the information the library has under the first headings he looks at that has material under it, and so would not continue to go from "dachshunds", to "hounds" then on the "dogs. Do these rules help our user or do they just save space?

A more problematic issue of manual vs. automated authority control is the problem of controlling the headings. Library of Congress authorities are divided into two files: The Name Authority File (NAF) and the Subject Authority File (SAF). The NAF is divided into six categories that are clearly defined; (personal names, corporate names, meeting names, geographic names, uniform titles, and series.) The SAF is divided into 39 categories that are not clearly defined; (animals, art, bodies of water, chemicals, Chinese art, Christian denominations, classes of persons, colonies, corporate bodies, diseases, drama, ethnic groups, individual family names, institutions, industries, Japanese art, Korean art, land vehicles, languages, legal topics, legislative bodies, literary works entered under author, literary works entered under title, literatures, materials, military services, musical compositions, organs and regions of the body, names of places, plants and crops, religious and monastic orders, religions, sacred works, topical headings, types of educational institutions, and wars).

In the SAF, there is no identifier to separate one category type from another. However, the rules require that a category can only be combined with another category in a defined way. For example: 'English literature' is in the 'literature' category, so it can only be followed by a subdivision listed under the 'Literatures' category. The heading 'English literature—History' is not correct because the subdivision 'History' can only follow the categories: 'Classes of Persons', 'Ethnic groups', 'Corporate bodies', 'Places', 'Colonies', 'Languages', 'Military Services', and 'Sacred works'. In this case the correct heading would be "English literature—History and criticism", because 'History and criticism' falls into the categories of 'Literatures', 'Musical' compositions', and 'video recordings'. However, there is not a place in the authority record that tells a computer to which category –'History' or –'History and criticism' can be applied. There is also nothing in the 'English literature' record to inform the computer that 'English literature' is in the 'literatures category'.

Another good example is 'Castles—Great Britain—Travel,' and the heading 'Castles—Great Britain—Description and travel.' 'Travel' as a subdivision means that the main heading (in this case Castles) is traveling. So the heading 'Castles—Great Britain—Travel' means that the castles in Great Britain are traveling from one place to another. The second heading 'Castles—Great Britain-- Description and travel' means the travel of people to castles in Great Britain. 'Travel' as a subdivision can be used with the categories: 'Classes of persons', 'Ethnic groups', 'Corporate bodies', and 'Individual persons'. 'Description and travel' can be used with the categories: 'Place' and 'Colonies'. With the present authority file an automated system has no way of telling which category fits which headings.

Another big problem for automated systems with current subject cataloging rules is that almost every rule has an exception and an automated system has a very hard time dealing with such exceptions. Here again let's return to the subdivision "History." History can be used under the categories: 'Classes of Persons', 'Ethnic groups', 'Corporate bodies', 'Places', 'Colonies', 'Languages', 'Military Services', and 'Sacred works'. However, in the standards we find in many places the statement "Do not use." Again using the example of 'history' we find that history cannot be used on headings that are already historical. An already historical heading would be: 'Renaissance', 'World War, 1914-1918', or a heading with dates. We cannot use history with 'literary','music', 'film', 'Television program', and 'video recording. Instead we must use the subdivision 'History and criticism'. (see SCM H1647) However, this is not the only section where "do not use" is listed. We find many sections with lists of "do not use."

Human intervention is required when exceptions to the rules make it impossible for the computer to determine which headings are correct. Many of these exceptions could be eliminated by a carefull reevaluation of our standards and rules; or by creating authority records for exception to the rule.

Another problem with LCSH is the use of patterned headings. A pattern heading is a main heading where an authority record has been created for most of the authorized main heading-- subdivision combinations for that category. The rule states a subdivision established under a pattern heading is usable, if appropriate, and no conflict exists, under any other heading belonging to its category. Tell me how a computer is going to determine if a heading is appropriate, and whether or not a conflict exists.   For example: 'Corn' is the pattern headings for 'Plants and crops'. So if we have an authority record for 'Corn as food—Contamination', we can use the heading 'Wheat as food—Contamination' without creating an authority record for the heading 'Wheat as food—Contamination.' This creates several problems for an automated system. (1) There is no

way for the system to know that 'Corn' and 'Wheat' belong to the same category. (2) Even though corn and wheat belong to the same category, there are some combinations that would not work for both corn and wheat. For example: 'Corn—Husking' is on the list under 'Corn'. This heading would not be used under 'Wheat'. How is the system to distinguish between what works for one but not the other? (3) Under the list of headings for corn are headings that have nothing to do with corn as a plant. For example: 'Corn laws (Great Britain)—Economic aspects'. How will the system know that a heading for 'Wheat laws (Great Britain)—Economic aspects' is not correct? (4) Some of the processes for handling corn are different than the processes for wheat. For example: Corn is picked while wheat is threshed so while 'Corn picking machinery—Accidents' is correct, but 'Wheat picking machinery—Accidents' would be the wrong heading for wheat. The heading for wheat would be 'Wheat—Threshing—Machinery—Accidents'. Each of the 26 categories presents similar problems for an automated system. In the card catalog and the print world pattern headings were a great way to save space; However, systems do not have a effective automated way to handle these headings.

"If appropriate, and no conflict exists" is boldly stated in patterned headings, however in many other places in the SCM it is only implied. For example: the subdivision 'Law and legislation' states "use under topical headings for works about the legal aspects of the topic, or works that contain the text of laws. DO NOT USE: If the topical headings itself is inherently legal, for example: 'Torts', 'Civil Procedure', 'Domestic relations'". How can a computer tell if a heading is inherently legal? Many catalogers have problems with telling if a heading is inherently legal.

Multiple subdivisions are another type of problem for the automated system. A multiple subdivision is a way to allow the use of like subdivision without establishing all headings. Examples: Marriage--Religious aspects--Baptists, [Catholic Church, etc.]" which would allow the use of the heading "Marriage--Religious aspects—(any church)" without establishing an authority record; or the heading "[place]–Foreign public opinion, British, [French, Italian, etc.]" which allows the use of "United States—Foreign public opinion, French." (see SCM H1090) How does a computer figure out all the different churches or places that might be used with the heading? In the card catalog adding a card for every church would have used a lot of space but for the computer it just requires more cataloger intervention.

All our form subdivisions can use either a delimiter x [$x] or a delimiter v [$v] depending on whether the material is or is about the subdivision. If we had a flower catalog, it might have the heading "Flowers $v Catalog" because the book is a flower catalog. However, if the book was about creating a flower catalog the heading would be "Flowers $x Catalog." A computer has no way to look at a bibliographic record and know whether it is a catalog or is about a catalog.

Another problem we run into is with the General International Standard Bibliographic Description (ISBD) proscribing punctuation within a field. When the punctuation is imbedded in the heading but can change from a period to a comma depending on the usage, the computer has a problem knowing what to do.

Conclusion

Taking a closer look at the complex problems of controlled vocabulary, it must be remembered that no one can do this alone. Rules and standards need to change, these changes need to be guided by what technology can offer, and by what information seekers need. Our libraries do not exist in isolation but belong to a vast network of librarians, publishers, system creators, and information seekers. Libraries are constantly changing and to be effective librarians must open their minds and welcome help from this vast group. Librarians must begin to communicate with the world outside of libraries and invite publishers, systems creators, or information seekers to meet with them and take a new look at the rules and standards. They must communicate, not just by stating what they want or what they think is best; but by listening to the ideas of others.

Karen Coyle made this most enlightening statement, in answer to an e-mail from Martha M. Yee on the <u>April 10, 2008 Version of the Statement of International Cataloging Principles.</u>

> "It would be ideal (nay, I should say necessary) for cataloging standards and systems standards to be developed together so that the requirements of both communities can be met. I can see many things in the current RDA drafts that simply WILL NOT DO what they are intended to do when taken off the written page and managed in a machine-readable record. Those of us in systems can only manipulate the data we are given, and we have very little say over its form. A better dialog between catalogers and the folks who will actually bring the catalog into being using computers would only benefit everyone, and the users most of all." ([RDA-L] Comments from Martha M. Yee on April 10, 2008 version of the STATEMENT OF INTERNATIONAL CATALOGUING PRINCIPLES. May 28, 2008 by Karen Coyle)

Cataloging requires both proficiency in cataloging and knowledge of the subject being

cataloged. This expertise is very important when the item is first cataloged, however, once an item has been cataloged by a librarian, no other library should be required to spend the time and money to maintain it. To accomplish this goal it is essential to remove all barriers to automation, permitting material to be cataloged once, then letting the automated system handle future changes and maintenances.

Bibliography

1. LC Subject Cataloging Manual.

2. ([RDA-L] Comments from Martha M. Yee on April 10, 2008 version of the STATEMENT OF INTERNATIONAL CATALOGUING PRINCIPLES. May 28, 2008 by Karen Coyle)