April 2008

Parallel Systems: The Coexistence of Subject Cataloging and Folksonomy

Elaine Peterson
Montana State University, elainep@montana.edu

Parallel Systems: The Coexistence of Subject Cataloging and Folksonomy

Elaine Peterson
Associate Professor, Information Resources Specialist
Montana State University
Bozeman, Montana USA

Introduction

Catalogers have always had to balance adherence to cataloging rules and authority files with creating cataloging that is current and relevant to users. That dilemma has been complicated in new ways because of user demands in the world of Web 2.0. Standardized cataloging is crucial for communication between computer systems, but patrons now have an expectation of social interaction on the Internet, as evidenced by the popularity of folksonomy. After a description of traditional subject cataloging and folksonomy, this article discusses several institutions where subject cataloging is still used, but where patron interaction is also encouraged. User-generated tags can coexist with controlled vocabulary such as subject headings.

Two Types of Subject Analysis

Subject cataloging has certain features that distinguish it from folksonomy. First, subject cataloging is a top-down approach, where the library professional determines the topical scope of the item being analyzed and assigns subjects to the bibliographic record. While recognizing user and library needs, the cataloger is the final arbiter of the subject headings. Second, the subject headings assigned are either prescribed or are derived according to rules that have been articulated. An obvious example is the Library of Congress Subject Headings (LCSH) with its accompanying manuals and interpretations.

Traditional subject cataloging is also hierarchical. Similar to taxonomy, subject groupings are generated in a tree-like, hierarchical structure. A cataloger working with a book on horses would apply the specific subject heading *Horses* even though a subset of the horses in the book are Arabians and some of them pictured are white. Moreover, if one did use the term *White horses*, it would be incorrect to assign the additional term *Black horses.* Contraries matter in a traditional classification system, and there are right and wrong subject headings.

In sum, library subject cataloging is based on established principles and rules. It is restrictive rather than inclusive, because choices are made by an information specialist who assigns a limited number of relevant subject headings. In the classical application of subject terms, the underlying assumption is that there are particular, relevant heading(s) to apply and that the author's intent is of primary importance, even though the needs of the library patron are given consideration. The cataloger aspires to be the unbiased professional, making cataloging decisions and without the interference of personal interpretations.

In contrast to subject cataloging, folksonomy is a reflection of personal preference. Folksonomy, or social bookmarking, emerged as a trend in 2005 as an alternative to traditional hierarchical cataloging.

As defined by Wikipedia, "Folksonomy is a collaboratively generated, open-ended labeling system that enables Internet users to categorize content such as Web pages, online photographs, and Web links." The article continues by explaining that "the freely chosen labels, called tags, help to improve search engine's effectiveness because content is categorized using a familiar, accessible, and shared vocabulary" ("Folksonomy"). It is appropriate to define folksonomy by using Wikipedia, which is a very good example of a work that is maintained by a social network of individual contributors. Hammond and others have preferred to call folksonomies by the name social bookmarking, emphasizing the social networking that may be achieved by the collaborative effort of assigning subjects and tagging an online piece (Hammond, 2005).

In her article on the high-stakes game of tag now being played by companies, Dye (2006) differentiates two types of folksonomies, broad and narrow. The first type characterizes an application whose users assign tags to the same content. One of the fastest growing folksonomy communities is del.icio.us, a well-known and representative social bookmarking site. A user who wishes to apply a tag chooses one, applies it, and then a bookmark is stored in a personal folder. The site states that tags are just like bookmarks, and one can invent them as needed. The site emphasizes the social aspect of folksonomy, allowing one user to tag another's items, to monitor the top tags used by the group, or to review what others are bookmarking.

Narrow folksonomy has more of a focus on the individual user. Flickr, a photograph-sharing website, is an example of narrow folksonomy. In Flickr, the focus is on the individual's personal tags and finding one's own photographs, although tags can be shared. Narrow folksonomy lacks the social discussion and cohesion of the broader applications. One can add photographs to Flickr and assign tags that have personal meaning. One can choose a tag like *Firenze,* instead of *Florence*, or even one that is meaningless to the rest of the community, such as *Uncle.*

Whether tags are created individually or collaboratively, what can folksonomy achieve? The Wikipedia article explains that,

> In contrast to professionally developed controlled vocabularies (also called taxonomies), folksonomies are unsystematic and, from an information scientist's point of view, unsophisticated; however, for Internet users, they dramatically lower content categorization costs because there is no complicated, hierarchically organized nomenclature to learn. One simply creates and applies tags on the fly ("Folksonomy").

Guy (2006) and others recognize the overall importance of folksonomies, even while acknowledging their problems, including typographical errors or spelling variations. The usefulness of folksonomies is not called into question, just their refinement without losing the openness that makes them so popular.

Problem

Both traditional cataloging and folksonomy can be useful, and both systems are being refined to relieve some of their inherent shortcomings. WIth the availability of online cataloging resources and the ease of updating online catalogs, the vocabulary of traditional subject cataloging has become more current. Folksonomy by its nature will always produce some inconsistency within a database (Peterson, 2006). Nevertheless, other problems with folksonomies such as variant terms, duplication, and typos are being resolved.

A database that truly integrates the two systems does not seem feasible at this time. Since traditional cataloging is rule-bound and limiting, while folksonomy is open-ended and relative to each user, the two remain distinct, self-contained systems of subject analysis. Each is useful on the Internet, and it is not surprising that some institutions are beginning to employ both systems to exploit the benefits

of each. Examining some of those institutions could be a key to understanding the future of subject analysis.

Examples That Allow for Both Systems

Outside the library world, patrons encounter many Web 2.0 applications. Amazon.com is a premier example. Amazon sells the same products that libraries provide, and libraries actually receive these same product descriptions, book cover images, and editorial content from their vendors. Amazon is the epitome of user engagement, a hallmark of Web 2.0. Amazon has user reviews and invitations to participate in various ways on virtually every page. Amazon even uses online customer activity to influence search results. Most online catalog vendors now offer features for patron interaction whether it is managing one's own library account or commenting on books. One author (Spiteri, 2007) has wondered whether libraries are really addressing patron needs and whether they should consider adding collaborative tagging to the online catalog. Spiteri points to folksonomies as organizers of personal information spaces, but also broaches the subject of using those tags to supplement existing controlled vocabularies. Speaking about public libraries, she suggests taking advantage of customizable online catalog features now available to incorporate end-user metadata in the OPAC.

One academic library has begun to use both LCSH headings and alternative tagging in the online catalog--PennTags at the University of Pennsylvania. PennTags are intended for use by the University of Pennsylvania community to tag events and news items. However, individuals can also use PennTags on records in the online catalog. Folksonomy terms exist side by side with the LCSH headings (PennTags, 2007), although they are not yet prevalent in the catalog.

Another example of a combination approach is Montana State University 's treatment of ETDs (Electronic Theses and Dissertations). Theses and dissertations are entered into the online catalog using standard LCSH headings. In addition to the OPAC, dissertations since 2003 are also entered in a separate relational database (Montana State University ETDs, 2007). The same LCSH subject headings appear in the ETD database that appear in the online catalog; however, within this database, users are given the opportunity to assign their own tags. For example, the ETD title *Other spaces, other voices* was assigned three LCSH headings

- Boundaries in literature
- Dystopias in literature
- Islands in literature.

Users have taken the opportunity to supplement the LCSH headings by adding the tags:

- Socio-spatial regime
- Foucault
- Dominant voices
- Island narrative

The ETD reader(s) who assigned those tags obviously thought it would be useful to other readers to provide those specific terms. Additionally, because the ETD database is a relational database, there is now a link from that thesis to other titles with those tags.

A third approach is to encourage patrons to take the data provided by the institution and reuse the records. This provides interaction with patrons and also extends the use of the data. A key feature of Web 2.0 is creating distinctive value for users. An example of this approach is the Minneapolis Institute of Art (ArtsConnectEd, 2007). The Institute has adopted an Amazon-style shopping site with its ArtsConnectEd Project. Users are encouraged to take museum images, drop them into personal collection spaces, and write their own descriptions. While the Institute still creates subject headings and

groups images for their museum inventory on the rest of the website, patrons are encouraged to build their own picture gallery and create their own tags.

A final example demonstrates catalogers moving outside of the library and into the patron's collection and tags--the Picture Australia project (Picture Australia , 2007). Membership in the project includes universities, city libraries, and museums. Although the Picture Australia site uses some traditional cataloging to organize the collections, the images are also loaded into Flickr where folksonomy tagging is promoted. Additionally, through their "click and flick" project, private individuals are encouraged to add both images and tags to the image database. Images of current news items, such as Cyclone Larry or the Commonwealth Games, are particularly encouraged. Picture Australia recognizes the importance of user interaction and is enlarging its digital collection at the same time.

A related example is work that is underway at the University of Washington, where the library is linking articles from Wikipedia back to the library's digital collections. One can see, for example, in the Wikipedia article on *salmon* that there are images and links supplied by the University of Washington Libraries Digital Collections-Salmon Collection ("Salmon").

Conclusion

Are these hybrid efforts successful? Since most of the examples are recent undertakings, many in place for less than a year, it is truly too soon to tell. One benchmark for the future will clearly be use by patrons. As mentioned, the use PennTags do not yet constitute high numbers. In the case of the Montana State University database, 560 ETD titles are included and each title averages four LCSH headings, but users have assigned only twenty-three tags for the entire database. Applying folksonomy tags has the potential to be very popular, but the numbers are not there yet for their use in library catalogs and similar settings. They may increase over time as the database is used. On the other hand, the measure of success may be that patrons are being provided with something useful, however small the number, since folksonomy tagging has little actual cost for an institution. The question of whether library users look for the social interaction features of Web 2.0 in the databases where they conduct library research remains. The assumption that it is beneficial to blend Web 2.0 features into library databases may not be correct.

These questions must be answered with further research and user studies. A single system of subject access might be the ideal, but given the incompatibility of some characteristics of subject cataloging and of folksonomy, it does not seem possible to integrate the two. Instead, subject cataloging and user-generated tags will probably coexist until their cost or usefulness is no justified.

References

ArtsConnectEd. Retrieved October 10, 2007, from http://www.artsmia.org

Dye, J. (2006). Folksonomy: a game of high-tech (and high-stakes) tag. *EContent* 29, 38-43.

Folksonomy (2008, March 28). In *Wikipedia, The Free Encyclopedia* . Retrieved April 4, 2008, from http://en.wikipedia.org/w/index.php?title=Folksonomy&oldid=201598719

Guy, M., & Tonkin, E. (2006). Folksonomies, tidying up the tags? *D-Lib Magazine* 12. Retrieved October 8, 2007, from http://dlib.org/dlib/january06/guy/01guy.html

Hammond, T., et al. (2005). Social bookmarking tools (I), a general review. *D-Lib Magazine* 11. Retrieved October 8, 2007, from http://www.dlib.org/dlib/april05/hammond/04hammond.html

Montana State University ETDs (2007). Retrieved October 10, 2007, from http://etd.lib.montana.edu/etd/view

PennTags (2007). Retrieved October 10, 2007, from http://tags.library.upenn.edu

Peterson, E. (2006). Beneath the metadata. *D-Lib Magazine* 12. Retrieved October 8, 2007, from
http://dlib.org/dlib/november06/peterson/11peterson.html

Picture Australia (2007). Retrieved October 10, 2007, from http://pictureaustralia.org

Salmon (2008, April 4). In *Wikipedia, The Free Encyclopedia* . Retrieved April 4, 2008, from
http://en.wikipedia.org/w/index.php?title=Salmon&oldid=203192489

Spiteri, L. F. (2006). The use of folksonomies in Public Library catalogues. *Serials Librarian* 51, 75-89.