

Tagging: An Organization Scheme for the Internet

Marijke A. Visser

How should the information on the Internet be organized? This question and the possible solutions spark debates among people concerned with how we identify, classify, and retrieve Internet content. This paper discusses the benefits and the controversies of using a tagging system to organize Internet resources. Tagging refers to a classification system where individual Internet users apply labels, or tags, to digital resources. Tagging increased in popularity with the advent of Web 2.0 applications that encourage interaction among users. As more information is available digitally, the challenge to find an organizational system scalable to the Internet will continue to require forward thinking. Trained to ensure access to a range of informational resources, librarians need to be concerned with access to Internet content. Librarians can play a pivotal role by advocating for a system that supports the user at the moment of need. Tagging may just be the necessary system.

Who will organize the information available on the Internet? How will it be organized? Does it need an organizational scheme at all? In 1998, Thomas and Griffin asked a similar question, "Who will create the metadata for the Internet?" in their article with the same name.¹ Ten years later, this question has grown beyond simply supplying metadata to assuring that at the moment of need, someone can retrieve the information necessary to answer their query. Given new classification tools available on the Internet, the time is right to reassess traditional models, such as controlled vocabularies and taxonomies, and contrast them with folksonomies to understand which approach is best suited for the future. This paper gives particular attention to Delicious, a social networking tool for generating folksonomies.

The amount of information available to anyone with an Internet connection has increased in part because of the Internet's participatory nature. Users add content in a variety of formats and through a variety of applications to personalize their Web experience, thus making Internet

content transitory in nature and challenging to lock into place. The continual influx of new information is causing a rapid cultural shift, more rapid than many people are able to keep up with or anticipate. Conversations on a range of topics that take place using Web technologies happen in real time. Unless you are a participant in these conversations and debates using Web-based communication tools, changes are passing you by. Internet users in general have barely grasped the concept of Web 2.0 and already the advanced "Internet cognoscenti" write about Web 3.0.²

Regarding the organization and availability of Internet content, librarians need to be ahead of the crowd as the voice who will assure content will be readily accessible to those that seek it. Internet users actively participating in and shaping the online communities are, perhaps unintentionally, influencing how those who access information via the Internet expect to be able to receive and use digital resources. Librarians understand that the way information is organized is critical to its accessibility. They also understand the communities in which they operate. Today, librarians need to be able to work seamlessly among the online communities, the resources they create, and the end user. As Internet use evolves, librarians as information stakeholders should stay abreast of Web 2.0 developments. By positioning themselves to lead the future of information organization, librarians will be able to select the best emerging Web-based tools and applications, become familiar with their strengths, and leverage their usefulness to guide users in organizing Internet content.

Shirky argues that the Internet has allowed new communities to form. Primarily online, these communities of Internet users are capable of dramatically changing society both on- and offline. Shirky contends that because of the Internet, "group action just got easier."³ According to Shirky, we are now at the critical point where Internet use, while dependent on technology, is actually no longer about the technology at all. The Web today (Web 2.0) is about participation. "This [the Internet] is a medium that is going to change society."⁴ Lessig points out that content creators are "writing in the socially, culturally relevant sense for the 21st century and to be able to engage in this writing is a measure of your literacy in the 21st century."⁵ It is significant that creating content is no longer reserved for the Internet cognoscenti. Internet users with a variety of technological skills are participating in Web 2.0 communities.

Information architects, Web designers, librarians, business representatives, and any stakeholder dependent on accessing resources on the Internet have a vested interest in how Internet information is organized. Not only does the architecture of participation inherent in the Internet encourage completely new creative endeavors, it serves as a platform for individual voices as demonstrated in

Marijke A. Visser (marijkea@gmail.com) is a Library and Information Science graduate student at Indiana University, Indianapolis, and will be graduating May 2010. She is currently working for ALA's Office for Information and Technology Policy as an Information Technology Policy Analyst, where her area of focus includes telecommunications policy and how it affects access to information.

personal and organizationally sponsored blogs: Lessig 2.0, Boing Boing, Open Access News, and others. These Internet conversations contribute diverse viewpoints on a stage where, theoretically, anyone can access them. Web 2.0 technologies challenge our understanding of what constitutes information and push policy makers to negotiate equitable Internet-use policies for the public, the content creators, corporate interests, and the service providers. To maintain an open Internet that serves the needs of all the players, those involved must embrace the opportunity for cultural growth the social Web represents.

For users who access, create, and distribute digital content, information is anything but static; nor is using it the solitary endeavor of reading a book. Its digital format makes it especially easy for people to manipulate it and shape it to create new works. People are sharing these new works via social technologies for others to then remix into yet more distinct creative work. Communication is fundamentally altered by the ability to share content on the Internet. Today's Internet requires a reevaluation of how we define and organize information. The manner in which digital information is classified directly affects each user's ability to access needed information to fully participate in twenty-first-century culture. New paradigms for talking about and classifying information that reflect the participatory Internet are essential.

Background

The controversy over organizing Web-based information can be summed up comparing two perspectives represented by Shirky and Peterson. Both authors address how information on the Web can be most effectively organized. In her introduction, Peterson states, "Items that are different or strange can become a barrier to networking."⁶ Shirky maintains, "As the Web has shown us, you can extract a surprising amount of value from big messy data sets."⁷ Briefly, in this instance ontology refers to the idea of defining where digital information can and should be located (virtually). Folksonomy describes an organizational system where individuals determine the placement and categorization of digital information. Both terms are discussed in detail below. Although any organizational system necessitates talking about the relationship(s) among the materials being organized, the relationships can be classified in multiple ways.

To organize a given set of entities, it is necessary to establish in what general domain they belong and in what ways they are related. Applying an ontological, or hierarchical, classification system to digital information raises several points to consider. First, there are no physical space restrictions on the Internet, so relationships among digital resources do not need to be strictly identified.

Second, after recognizing that Internet resources do not need the same classification standards as print material, librarians can begin to isolate the strengths of current nondigital systems that could be adapted to a system for the Internet. Third, librarians must be ready to eliminate current systems entirely if they fail to serve the needs of Internet users.

Traditional systems for organizing information were developed prior to the information explosion on the Internet. The Internet's unique platform for creating, storing, and disseminating information challenges pre-digital-age models. Designing an organizational system for the Internet that supports creative innovation and succeeds in providing access to the innovative work is paramount to moving the twenty-first-century culture forward.

Assessing alternative models

Controversy encourages scrutiny of alternative models. In understanding the options for organizing digital information, it is important to understand traditional classification models. Smith discusses controlled vocabularies, taxonomies, and facets as three traditional methods for applying metadata to a resource. According to Smith, a controlled vocabulary is an unambiguous system for managing the meanings of words. It links synonyms, allowing a search to retrieve information on the basis of the relationship between synonyms.⁸ Taxonomies are hierarchical, controlled vocabularies that establish parent-child relationships between terms. A faceted classification system categorizes information using the distinct properties of that information.⁹ In such a system, information can exist in more than one place at a time. A faceted classification system is a precursor to the bottom-up system represented by folksonomic tagging. Folksonomy, a term coined in 2004 by Thomas Vander Wal, refers to a "user-created categorical structure development with an emergent thesaurus."¹⁰ Vander Wal further separates the definition into two types: a narrow and a broad folksonomy.¹¹ In a broad folksonomy, many people tag the same object with numerous tags or a combination of their own and others' tags. In a narrow folksonomy, one or few people tag an object with primarily singular terms.

Internet searching represents a unique challenge to people wanting to organize its available information. Search engines like Yahoo! and Google approach the chaotic mass of information using two different techniques. Yahoo! created a directory similar to the file folder system with a set of predetermined categories that were intended to be universally useful. In so doing, the Yahoo! developers made assumptions about how the general public would categorize and access information. The categories

and subsequent subcategories were not necessarily logically linked in the eyes of the general public. The Yahoo! directory expanded as Internet content grew, but the digital folder system, like a taxonomy, required an expert to maintain. Shirky notes the Yahoo! model could not scale to the Internet. There are too many possible links to be able to successfully stay within the confines of a hierarchical classification system. Additionally, on the Internet, the links are sufficient for access because if two items are linked at least once, the user has an entry point to retrieve either one or both items.¹² A hierarchical system does not assure a successful Internet search and it requires a user to comprehend the links determined by the managing expert. In the Google approach, developers acknowledged that the user with the query best understood the unique reasoning behind her search. The user therefore could best evaluate the information retrieved. According to Shirky, the Google model let go of the hierarchical file system because developers recognized effective searching cannot predetermine what the user wants. Unlike Yahoo!, Google makes the links between the query and the resources after the user types in the search terms.¹³ Trusting in the link system led Google to understand and profit from letting the user filter the search results.

To select the best organizational model for the Internet it is critical to understand its emergent nature. A model that does not address the effects of Web 2.0 on Internet use and fails to capture participant-created content and tagging will not be successful. One approach to organizing digital resources has been for users to bookmark websites of personal interest. These bookmarks have been stored on the user's computer, but newer models now combine the participatory Web with saving, or tagging, websites. Social bookmarking typifies the emergent Web and the attraction of online networking.

Innovative and controversial, the folksonomy model brings to light numerous criteria necessary for a robust organizational system. A social bookmarking network, Delicious is a tool for generating folksonomies. It combines a large amount of self-interest with the potential for an equal, if not greater, amount of social value. Delicious users add metadata to resources on the Internet by applying terms, or tags, to URLs. Users save these tagged websites to a personal library hosted on the Delicious website. The default settings on Delicious share a user's library publicly, thus allowing other people—not limited to registered Delicious account holders—to view any library. That the Delicious developers understood how Internet users would react to this type of interactive application is reflected in the popularity of Delicious. Delicious arrived on the scene in 2003, and in 2007 developers introduced a number of features to encourage further user collaboration. With a new look (going from the original *del.icio.us* to its current moniker, Delicious) as well as more ways for users to retrieve and share resources by 2007, Delicious

had 3 million registered users and 100 million unique URLs.¹⁴ The reputation of Delicious has generated interest among people concerned with organizing the information available via the Internet.

How does the folksonomy or Delicious model of open-ended tagging affect searching, information retrieving, and resource sharing? Delicious, whose platform is heavily influenced by its users, operates with no hierarchical control over the vocabulary used as tags. This underscores the organization controversy. Bottom-up tagging gives each person tagging an equal voice in the categorization scheme that develops through the user generated tags. At the same time, it creates a chaotic information-retrieval system when compared to traditional controlled vocabularies, taxonomies, and other methods of applying metadata.¹⁵ A folksonomy follows no hierarchical scheme. Every tag generated supplies personal meaning to the associated URL and is equally weighted. There will be overlap in some of the tags users select, and that will be the point of access for different users. For the unique tags, each Delicious user can choose to adopt or reject them for their personal tagging system. Either way, the additional tags add possible future access points for the rest of the user community. The social usefulness of the tags grows organically in relationship to their adoption by the group.

Can the Internet support an organizational system controlled by user-generated tags? By the very nature of the participatory Web, whose applications often get better with user input, the answer is yes. Delicious and other social tagging systems are proving that their folksonomic approach is robust enough to satisfy the organizational needs of their users. Defined by Vander Wal, a broad folksonomy is a classification system scalable to the Internet.¹⁶

The problem with projecting already-existing search and classification strategies to the Internet is that the Internet is constantly evolving, and classic models are quickly overcome. Even in the nonprint world of the Internet, taxonomies and controlled vocabulary entail a commitment both from the entity wanting to organize the system and the users who will be accessing it. Developing a taxonomy involves an expert, which requires an outlay of capital and, as in the case with Yahoo!, a taxonomy is not necessarily what users are looking for. To be used effectively, taxonomies demand a certain amount of user finesse and complacency. The user must understand the general hierarchy and by default must suspend their own sense of category and subcategory if they do not mesh with the given system. The search model used by Google, where the user does the filtering, has been a significantly more successful search engine. Google recognizes natural language, making it user friendly; however, it remains merely a search engine. It is successful at making links, but it leaves the user stranded without a means to organize search results beyond simple page rank. Traditional

hierarchical systems and search strategies like those of Yahoo! and Google neglect to take into account the tremendous popularity of the participatory Web. Successful Web applications today support user interaction; to disregard this is naive and short-sighted.

In contrast to a simple page-rank results list or a hierarchical system, Delicious results provide the user with rich, multilayer results. Figure 1 shows four of the first ten results of a Delicious search for the term “folksonomy.” The articles by the four authors in the left column were tagged according to the diagram. Two of the articles are peer-reviewed, and two are cited repeatedly by scholars researching tagging and the Internet. In this example, three unique terms are used to tag those articles, and the other terms provide additional entry points for retrieval. Further information available using Delicious shows that the Guy article was tagged by 1,323 users, the Mathes article by 2,787 users, the Shirky article by 4,383 users, and the Peterson article by 579 users.¹⁷ From the basic Delicious search, the user can combine terms to narrow the query as well as search what other users have tagged with those terms. Similar to the card catalog, where a library patron would often unintentionally find a book title by browsing cards before or after the actual title she originally wanted, a Delicious user can browse other users’ libraries, often finding additional pertinent resources. A user will return a greater number of relevant and automatically filtered results than with an advanced Google search. As an ancillary feature, once a Delicious user finds an attractive tag stream—a series of tags by a particular user—they can opt to follow the user who created the tag stream, thereby increasing their personal resources. Hence Delicious is effective personally and socially. It emulates what Internet users expect to be able to do with digital content: find interesting resources, personalize them, in this case with tags, and put them back out for others to use if they so choose.

Proponents of folksonomy recognize there are benefits to traditional taxonomies and controlled vocabulary systems. Shirky delineates two features of an organizational system and their characteristics, providing an example of when a hierarchical system can be successful (see table 1).¹⁸

These characteristics apply to situations using databases, journal articles, and dissertations as spelled out

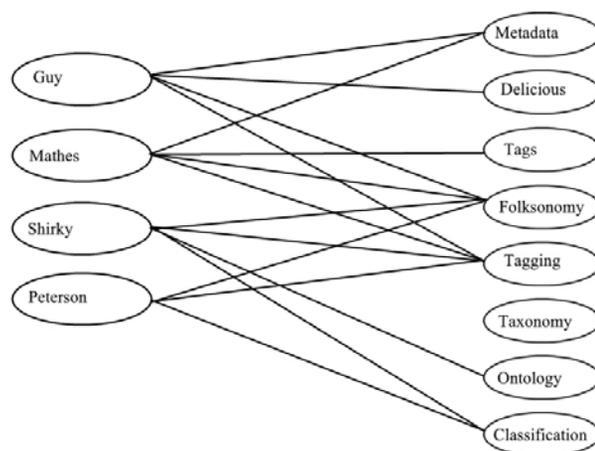


Figure 1. Search results for “folksonomy” using Delicious.

by Peterson, for example.¹⁹ Specific organizations with identifiable common terminology—for example, medical libraries—can also benefit from a traditional classification system. These domains are the antithesis of the domain represented by the Web. The success of controlled vocabularies, taxonomies, and their resulting systems depends on broad user adoption. That, in combination with the cost of creating and implementing a controlled system, raises questions as to their utility and long-term viability for use on the Web.

Though meant for longevity, a taxonomy fulfills a need at one fixed moment in time. A folksonomy is never static. Taxonomies developed by experts have not yet been able to be extended adequately for the breadth and depth of Internet resources. Neither have traditional viewpoints been scaled to accept the challenges encountered in trying to organize the Internet. Folksonomy, like taxonomy, seeks to provide the information critical to the user at the moment of need. Folksonomy, however, relies on users to create the links that will retrieve the desired results. Doctorow puts forward three critiques of a hierarchical metadata system, emphasizing the inadequacies of applying traditional classification schemes to the digital stage:

1. There is not a “correct” way to categorize an idea.
2. Competing interests cannot come to a consensus

Table 1. Domains and their participants

Domain to be Organized	Participants in the Domain
Small corpus	Expert catalogers
Formal categories	Authoritative source of judgment
Restricted entities	Coordinated users
Clear Edges	Expert users

- on a hierarchical vocabulary.
3. There is more than one way to describe something.

Doctorow elaborates: "Requiring everyone to use the same vocabulary to describe their material denudes the cognitive landscape, enforces homogeneity in ideas."²⁰ The Internet raises the level of participation to include innumerable voices. The astonishing thing is that it thrives on this participation.

Guy and Tonkin address the "folksonomic flaw" by saying user-generated tags are by definition imprecise. They can be ambiguous, overly personal, misspelled, and a contrived compound word. Guy and Tonkin suggest the need to improve tagging by educating the users or by improving the systems to encourage more accurate tagging.²¹ This, however, does not acknowledge that successful Web 2.0 applications depend on the emergent wisdom of the user community. The systems permit organic evolution and continual improvement by user participation. A folksonomy evolves much the way a species does. Unique or single-use tags have minimal social import and do not gain recognition. Tags used by more than a few people reinforce their value and emerge as the more robust species.

Conclusion

The benefits of the Internet are accessible to a wide range of users. The rewards of participation are immediate, social, and exponential in scope. User-generated content and associated organization models support the Internet's unique ability to bring together unlikely social relationships that would not necessarily happen in another milieu. To paraphrase Shirky and Lessig, people are participating in a moment of social and technological evolution that is altering traditional ways of thinking about information, thereby creating a break from traditional systems. Folksonomic classification is part of that break. Its utility grows organically as users add tagged content to the system. It is adaptive, and its strengths can be leveraged according to the needs of the group. While there are "folksonomic flaws" inherent in a bottom-up classification system, there is tremendous value in weighting individual voices equally. Following the logic of Web 2.0 technology, folksonomy will improve according to the input of the users. It is an organizational system that reflects the basic tenets of the emergent Internet. It may be the only practical solution in a world of participatory content creation.

Shirky describes the Internet by saying, "There is no shelf in the digital world."²² Classic organizational schemes like the Dewey Decimal System were created to organize

resources prior to the advent of the Internet. A hierarchical system was necessary because there was a physical limitation on where a resource could be located; a book can only exist in one place at one time. In the digital world, the shelf is simply not there. Material can exist in many different places at once and can be retrieved through many avenues. A broad folksonomy supports a vibrant search strategy. It combines individual user input with that of the group. This relationship creates data sets inherently meaningful to the community of users seeking information on any given topic at any given moment. This is why a folksonomic approach to organizing information on the Internet is successful. Users are rewarded for their participation, and the system improves because of it. Folksonomy mirrors and supports the evolution of the Internet.

Librarians, trained to be impartial and ethically bound to assure access to information, are the logical mediators among content creators, the architecture of the Web, corporate interests, and policy makers. Critical conversations are no longer happening only in traditional publications of the print world. They are happening with communication platforms like YouTube, Twitter, Digg, and Delicious. Information organization is one issue on which librarians can be progressive. Dedicated to making information available, librarians are in a unique position to take on challenges raised by the Internet. As the profession experiments with the introduction of Web 3.0, librarians need to position themselves between what is known and what has yet to evolve. Librarians have always leveraged the interests and needs of their users to tailor their services to the individual entry point of every person who enters the library. Because more and more resources are accessed via the Internet, librarians will have to maintain a presence throughout the Web if they are to continue to speak for the informational needs of their users. Part of that presence necessitates an ability to adapt current models to the Internet. More importantly, it requires recognition of when to forgo conventional service methods in favor of more innovative approaches. Working in concert with the early adopters, corporate interests, and general Internet users, librarians can promote a successful system for organizing Internet resources. For the Internet, folksonomic tagging is one solution that will assure users can retrieve information necessary to answer their queries.

References and notes

1. Charles F. Thomas and Linda S. Griffin, "Who Will Create the Metadata for the Internet?" *First Monday* 3, no. 12 (Dec. 1998).
2. Web 2.0 is a fairly recent term, although now ubiquitous among people working in and around Internet technologies. Attributed to a conference held in 2004 between MediaLive

International and O'Reilly Media, Web 2.0 refers to the Web as being a platform for harnessing the collective power of Internet users interested in creating and sharing ideas and information without mediation from corporate, government, or other hierarchical policy influencers or regulators. Web 3.0 is a much more fluid concept as of this writing. There are individuals who use it to refer to a Semantic Web where information is analyzed or processed by software designed specifically for computers to carry out the currently human-mediated activity of assigning meaning to information on a webpage. There are librarians involved with exploring virtual-world librarianship who refer to the 3D environment as Web 3.0. The important point here is that what Internet users now know as Web 2.0 is in the process of being altered by individuals continually experimenting with and improving upon existing Web applications. Web 3.0 is the undefined future of the participatory Internet.

3. Clay Shirky, "Here Comes Everybody: The Power of Organizing Without Organizations" (presentation videocast, Berkman Center for Internet & Society, Harvard University, Cambridge, Mass., 2008), <http://cyber.law.harvard.edu/interactive/events/2008/02/shirky> (accessed Oct. 1, 2008).

4. *Ibid.*

5. Lawrence Lessig, "Early Creative Commons History, My Version," videocast, Aug. 11, 2008, Lessig 2.0, http://lessig.org/blog/2008/08/early_creative_commons_history.html (accessed Aug. 13, 2008).

6. Elaine Peterson, "Beneath the Metadata: Some Philosophical Problems with Folksonomy," *D-Lib Magazine* 12, no. 11 (2006), <http://www.dlib.org/dlib/november06/peterson/11peterson.html> (accessed Sept. 8, 2008).

7. Clay Shirky, "Ontology is Overrated: Categories, Links, and Tags" online posting, Spring 2005, Clay Shirky's Writings about the Internet, http://www.shirky.com/writings/ontology_overrated.html#mind_reading (accessed Sept. 8, 2008).

8. Gene Smith, *Tagging: People-Powered Metadata for the Social Web* (Berkeley, Calif.: New Riders, 2008): 68.

9. *Ibid.*, 76.

10. Thomas Vander Wal, "Folksonomy," online posting, Feb. 7, 2007, [vanderwal.net, http://www.vanderwal.net/folksonomy.html](http://www.vanderwal.net/folksonomy.html) (accessed Aug. 26, 2008).

11. Thomas Vander Wal, "Explaining and Showing Broad and Narrow Folksonomies," online posting, Feb. 21, 2005, Personal InfoCloud, http://www.personalinfocloud.com/2005/02/explaining_and_.html (accessed Aug. 29, 2008).

12. Shirky, "Ontology is Overrated."

13. *Ibid.*

14. Michael Arrington, "Exclusive: Screen Shots and Feature Overview of Delicious 2.0 Preview," online posting, June 16, 2005, TechCrunch, <http://www.techcrunch.com/2007/09/06/exclusive-screen-shots-and-feature-overview-of-delicious-20-preview/> (accessed Jan. 6, 2010).

15. Smith, *Tagging*, 67–93.

16. Vander Wal, "Explaining and Showing Broad and Narrow Folksonomies."

17. Adam Mathes, "Folksonomies—Cooperative Classification and Communication through Shared Metadata" (graduate paper, University of Illinois Urbana-Champaign, Dec. 2004); Peterson, "Beneath the Metadata"; Shirky, "Ontology is Overrated"; Thomas and Griffin, "Who Will Create the Metadata for the Internet?"

18. Shirky, "Ontology is Overrated."

19. Peterson, "Beneath the Metadata."

20. Cory Doctorow, "Metacrap: Putting the Torch to Seven Straw-Men of the Meta-Utopia," online posting, Aug. 26, 2001, The Well, <http://www.well.com/~doctorow/metacrap.htm> (accessed Sept. 15, 2008).

21. Marieke Guy and Emma Tonkin, "Folksonomies: Tidying up Tags?" *D-Lib Magazine* 12, no. 1 (2006), <http://www.dlib.org/dlib/january06/guy/01guy.html> (accessed Sept. 8, 2008).

22. Shirky, "Ontology is Overrated."

Global Interoperability *continued from page 33*

9. Julie Renee Moore, "RDA: New Cataloging Rules, Coming Soon to a Library Near You!" *Library Hi Tech News* 23, no. 9, (2006): 12.

10. Rick Bennett, Brian F. Lavoie, and Edward T. O'Neill, "The Concept of a Work in WorldCat: An Application of FRBR," *Library Collections, Acquisitions, & Technical Services* 27, no. 1, (2003): 56.

11. Park, "Cross-Lingual Name and Subject Access."

12. *Ibid.*

13. Thomas B. Hickey, "Virtual International Authority File" (Microsoft PowerPoint presentation, ALA Annual Conference, New Orleans, June 2006), <http://www.oclc.org/research/projects/viaf/ala2006c.ppt> (accessed Dec. 9, 2009).

14. LEAF, "LEAF Project Consortium," <http://www.crxnet.com/leaf/index.html> (accessed Dec. 9, 2009).

15. Bennett, Lavoie, and O'Neill, "The Concept of a Work in WorldCat."

16. Alan Danskin, "Mature Consideration: Developing Bibliographic Standards and Maintaining Values," *New Library World* 105, no. 3/4, (2004): 114.

17. *Ibid.*

18. Bennett, Lavoie, and O'Neill, "The Concept of a Work in WorldCat."

19. Moore, "RDA."

20. Danskin, "Mature Consideration," 116.

21. *Ibid.*; Park, "Cross-Lingual Name and Subject Access."