# Social Tagging in the Web 2.0 Environment: Author vs. User Tagging

Heather Lea Moulaison

Social tagging increasingly attracts the attention of information scientists. Much research has focused on large-scale assessment of tag sets in systems. The reasons for tagging remain uncertain. To date, the LIS field has not explicitly explored the differences between author-supplied metadata (endo-tagging) and user-supplied metadata (exo-tagging). This paper reviews a number of studies on social tagging and recommends further user-centered research into motivation; proposed research could be carried out in conjunction with studies focusing on the different reasons for endo- and exo-tagging.

**Keywords:** Social tagging, Web 2.0, User motives

Author Bio: Heather Lea Moulaison is a doctoral student at Rutgers University in the School of Communication, Information and Library Studies. She has been named a Fulbright Scholar for the 2008-2009 school year to teach at the Ecole des Sciences de l'Information in Rabat, Morocco. Ms. Moulaison received both an MA in French and an MSLIS in Library and Information Science from the University of Illinois at Urbana-Champaign. Previously, Ms. Moulasion worked in academic libraries, primarily as a cataloging librarian. She has taught Cataloging & Classification and other LIS courses through ALA-accredited library programs. Ms. Moulaison's scholarly interests include the Library 2.0 use of new technologies, digital libraries, and how both can affect users in American and Francophone countries' libraries. This paper was written under the guidance of Dr. Sherry Vellucci; to whom a debt of gratitude is wholeheartedly acknowledged. Edward Corrado's insightful comments on the manuscript are also acknowledged and appreciated.

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

Social tagging in the Web 2.0 environment is a relatively new aspect of personal organization of online information. Although social tagging has attracted the attention of information scientists and others who study such online phenomena, relatively little published literature exists in Library and Information Science (LIS) to inform the field's perceptions and understandings. This paper argues that a yet unstudied aspect of tagging research involves the different reasons for which users supply tags based on the kind of indexing taking place. Communities of users will tag resources differently depending on whether resources have been bookmarked for personal finding purposes (exo-tagging) or whether the resource itself is a creation of the tagger and the tag is a means of advertising that content to the community (endo-tagging). User-centered studies are a necessary part of the process of testing current assumptions and of advancing this field of research.

In the online environment, users tag multiple types of content for various reasons using a variety of Web 2.0 social networking systems. Little is known about why users tag materials; however, it is possible to study the tags themselves as artifacts of the personal classification process and to make inferences about the way certain terms are being used by individuals. It is also possible to look at the tags from the point of view of an entire Web 2.0 social tagging site, as commonalities will be present due to the structure and purpose of the system. Lastly, it is possible to compare users with each other based on the similarity of the tags and the works of interest to which they assign these keywords. All of these elements, along with others, may play a part in the decision to use tags and are viable areas of study in LIS research.

Although the mechanics of applying tags is similar in disparate online environments, it is reasonable to expect users to tag resources differently according to the characteristics of the resource itself and the function of the system in use. Some social tagging sites like del.icio.us and

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

Connotea allow users to create exo-tags, metadata for available Web content, that they intend to use or otherwise revisit. These sets of exo-tags or "bookmarks" tend to pertain to content that was created by others but that has some value to the tagger. By exo-tagging the resource, the user will be able to find more easily the resource at a later time, and will be able to classify it using terms that are personally meaningful.

Other sites such as YouTube and Flickr allow users to endo-tag content that they have created as they make that content available to others in the online community. Endo-tagging pertains to content that has been created by the tagger him- or herself and uploaded expressly for consumption by other users of the system. Although these creator-supplied endo-tags help anchor the resource in the online world, the tags are primarily to help others in the community in their own discovery of online resources within the system and serve an advertising function.

Library and Information Science literature on the topic of tagging is still somewhat undeveloped due to the relatively recent appearance of the phenomenon. A review of the literature reveals that, initially, explorations of social tagging took place in unpublished venues such as the blogosphere or thorough coursework at the academy. Citations appearing in works prior to 2005 draw largely from unpublished online sources (Mathes 2004; Walker 2005). Indeed, the 2006 publication by Marlow, Naaman, boyd, and Davis notes that "despite a considerable amount of attention in academic circles, as represented in various blog posts, little academic research work has been invested in tagging systems to date (Marlow, Naaman, boyd, and Davis 2007, p. 32).

By 2005 and 2006, more formal work on tagging came into being, including Clay Shirky's Internet writings (2005) and the description of the phenomenon in a French-language library publication (Le Deuff 2006). Around this time, conference proceedings appeared more

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

regularly with social tagging as a focal point of study, and studies themselves were being published (i.e. Guy & Tonkin 2006; Golder & Huberman 2006; Marlow et al. 2006; Lin et al. 2006). In 2007, articles have appeared in LIS-related bulletins for professional communities such as the *Bulletin of the American Society for Information Science and Technology* and their special issue on social tagging that appeared in October/November 2007 (i.e. Morrison 2007), and a book chapter on social tagging appeared in a book exploring Library 2.0 (Kroski 2007). Despite the recent emphasis on social tagging and the quality research being done in the field, authoritative background literature is still not available in abundance.

## *FOLKSONOMIES AND RESOURCE-SHARING INTENTIONS*

To study social tagging, it is necessary first to define what is meant by tags. Guy and Tonkin describe tags as “any word that defines a relationship between the online resource and a concept in the user's mind” (2006). It stands to reason, then, that if the content has a different meaning to the user based on the relationship to the content, then the tags employed by different taggers for different tasks will also be different. When referring to del.icio.us, Golder and Huberman note the inherent inconsistency of having personalized bookmarks in an open, online environment. "These two features – storage of personal bookmarks and the public nature of those bookmarks – are somewhat at odds with one another" (2006, p. 201). Although tags are not part of a controlled vocabulary, when viewed together, observations that are of interest to LIS research can be explored.

Social tags may be discovered through the use of tag clouds generated by the system. Tag clouds are a common way of visualizing the tags that are present, either in the whole of the

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

system or for a given user. In tag clouds, tags are listed alphabetically, with larger, bolder, and more prominent fonts for more popular tags. Tag clouds may be of interest on a system-by-system basis, and serve as a good way to gauge the popularity of tags at a glance. In her chapter on tagging, Kroski asserts that tag clouds are a glimpse into the "Zeitgeist, or what is currently in the public favor" as well as, in certain systems, tags for a particular user (2007, p. 94). The way tags are used may also be changing over time. Russell notes that, in the past, users supplied plural common nouns that were similar to indexing terms. More recently, users have been supplying singular nouns that seems to function more as labels than indexing terms (Russell 2008).

Broadly speaking, groups of tags are of interest to information scientists who want to study the phenomenon of social tagging. The user-centered and community-centered collection of tags can be seen to form its own sort of organizational scheme or structure that is akin to an informal taxonomy. Although the term "folksonomy" is disputed by different groups of researchers, the basic underlying concepts bear to be acknowledged. One current way of looking at folksonomies is the following: "Folksonomies are generally used to organize information and support information retrieval (IR)" (Morrison 2007, p. 12). This ontology-like function occurs in the online environment when a critical mass of user-supplied tags exists. Social tags can be considered to be wild forces that do not require taming (Walker 2005) because these tags can function as a way of establishing order in the Web 2.0 environment (Shirky 2005). When a critical mass of community-generated tags is reached, users from within that community are able to navigate a set of intuitive tags to find meaningful and relevant resources.

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

## *TAGGING STUDIES*

Much of the LIS body of literature devoted to social tagging focuses on the tags themselves; studies shedding light on the social tagging phenomenon were gaining attention by 2005 and 2006. *D-Lib Magazine* has published studies on social tagging, and in January 2006 it published one by Guy and Tonkin. In this article, the authors study large sets of tags in del.icio.us and Flickr. The methodology of this study involved the accumulation of large numbers of tags and the distillation of trends based on the tags themselves. By virtue of their work, Guy and Tonkin assert that tags in these systems are "ambiguous, overly personal, and inexact" (2006). Despite the personal nature of tags, Guy and Tonkin find that only 10-15% of tags in Flickr and del.icio.us are single-use tags.

Golder and Huberman also performed a study on social tagging and their work had a considerable impact on the LIS community (Marlow et al. 2006, p. 32); the Golder and Huberman study was likewise published in 2006 and the authors also focus on tagger activity as evidenced in tags. The Golder and Huberman study focuses on user tags in del.icio.us. One very noteworthy contribution of their study is the system they devised for categorizing bookmarks according to the assessed function of the tags. The seven mutually exclusive semantic categories that Golder and Huberman apply are the following: 1. Identifying what (or who) it is about; 2. Identifying what it is; 3. Identifying who owns it; 4. Refining categories; 5. Identifying qualities or characteristics; 6. Self reference; and 7. Task organizing (2006, p. 203). By categorizing large sets of tags, the authors are able to assess trends in bookmarking via usage patterns over time.

The fact that Lin, Beaudoin, Bui, and Desai combined their three studies into one paper that was published in the 2006 ASIS&T proceedings is of interest. The authors carry out three

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

very different studies on social tagging and report them under the umbrella of "social classification" (2006). One study involves the comparison of tags in Connotea with Medical Subject Heading (MESH) terms and automated indexing. This study actually reports on user-assigned tags, intermediary-assigned tags, and machine-assigned tags without labeling them as such; the amount of overlap between the systems is judged to be small. A second very different study compares the categories of tags in Flickr and the ease with which coders can or cannot agree on a category. The nature of this study can be seen as calling into question the type of work done by Golder and Huberman in their article. If coders have difficulty agreeing on the objective meaning of tags, presumably because of the highly personal nature of tag usage on social bookmarking Web sites, then categorizing tags may not be a reliable way of implying user intention. The last study focused on user-created bookmarks on del.icio.us, pulling the tags that were listed as "most popular" on the site and their related tags. The authors tried to assess whether tags followed a power law distribution. Their results were less strong than anticipated.

Another study from 2006 was carried out by four researchers affiliated with Yahoo!: Marlow, Naaman, boyd, and Davis. The emphasis of the study is on the taxonomic functions of the endo-tags in Flickr where the creator is also the tagger. In suggesting the interplay at work on social tagging sites, Marlow et al. advance a model that links users and resources, but that also links resources to resources and users to users (2006, p. 32). The user motivation that they cite is divided into two categories, *organizational* and *social*. They include several categories of supporting incentives for tagging, but list as the first three 1. Future retrieval; 2. Contribution and sharing; and 3. Attract attention (p. 35) as user incentives for social tagging. These observations, however astute, are based on incentives that can be "outwardly observed in current social tagging systems" (p. 35) and not on user-centered studies.

A more recent study of tagging by Ames and Naaman investigated geo-tagging in Yahoo!'s Flickr (2007). Geo-tagging is the process of including geographic metadata in tags. In their study, Ames and Naaman looked at tags created by early adopters of the ZoneTag mobile application. ZoneTag provides a limited number of automated tags at upload, including location. Users can also supply endo-tags. Ames and Naaman interviewed a group of thirteen early adopters of ZoneTag in an attempt to document motivations. Although the results are inconclusive, this early study of motivation is an important step in the body of literature on social tagging and users.

Studies that will be of particular interest to the organization of information community compare the current use of social tags to known phenomena associated with indexing. Like the others, Margaret Kipp also carries out a study of tags that are available on publicly available social tagging sites, but in doing so, keeps in mind the different roles of the traditional indexer (professional intermediary) and the user/tagger. In her 2007 article, Kipp takes user tags from CiteULike as well as author keywords and descriptors (MESH terms) from PubMed for scholarly articles in the field of biology. In this way, Kipp's research was similar to the Lin et al. case study using Connotea terms and MESH terms along with automatically generated terms. Kipp's conclusions are that users prefer to follow associative trails rather than to search using controlled vocabulary terms.

It is worth noting that none of the studies examined in the course of this review of the literature considers the inherent differences between endo- and exo-tagging. Reference to an article in *D-Lib Magazine* from 2005 by the authors of Connotea does make mention of this distinction, and the unpublished article from a computer-meditated communication standpoint by Mathes (2004) does as well. However, these works, although influential for information

scientists, have not had an impact on the articles surveyed for this research and have not influenced the course of study of articles and proceedings in the major LIS venues. Instead, work presented at the ASIS&T conferences and published in scholarly LIS journals does not yet attempt to study the user per se. By focusing on the tags that the user leaves and inferring a rationale for their application, LIS is missing out on an important element of the tagging process: the motivation.

## *COMMUNITY ASPECTS AND TAG FUNCTIONS*

A variety of different types of social tagging-enabled environments are now available on the Web, each with a specific purpose and niche. As enumerated above, information science research has examined user tags in several of these, but fails to explain differences in tags as a result of the different sites studied. Researchers have investigated tags in del.icio.us (Golder & Huberman 2006; Lin et al. 2006), Connotea (Lin et al. 2006), CiteULike (Kipp 2007), and Flickr (Guy & Tonkin 2006; Lin et al. 2006). Researchers discuss the purpose of the goals of the sites, but have not been systematically tying the nature of the site in with the tag use results they have found. Instead these studies tend to focus on the number of tags per resource (Golder & Huberman 2006; Lin et al. 2006), the order in which tags appear (Guy & Tonkin 2006), the nature of the tags (Golder & Huberman 2006), and the relationship between user bookmarks that do or do not have specific community-based tags (Corrado 2007). Consistently, effort has been made to understand the tags themselves despite their highly personal nature. The difficulty with which tags may be classified by researchers due to their highly personal nature and to difficulties

with interpreting largely decontextualized polysemous language is evident in the research Lin et al. (2006).

There are instances where, due to the nature of the site, tags can be highly social in their promotion of the community and their attempt at taking part. Whereas Kipp (2007) notes that some social tagging is highly contextual and "short term and highly specific" (p. 4) this may not apply to all instances of tagging, especially if the item is not found but is created and launched for others in the community as a video or other user generated content on a video- or photo-sharing site. In these cases, the tags could be more stable and long-term, as they promote personal content for the community instead of announcing personal intention as toread or todo.

Joseph Tennis presented on social tagging at the 2006 ASIS&T conference. For him, "social tagging systems have grown up around communities that want to share goals" (Tennis, p. 7). However, Tennis also states that "social tagging is done for personal reasons. As such, the purpose and reflection on that process are personal in nature. Likewise, since the act of tagging is for oneself, not someone else" (p. 9). This second statement is perhaps not applicable in all tagging systems based on the previous discussion. Despite the fact that the mechanism for tagging is the same in that it requires users to consider their relationship to the resource, the purpose for the tagging may be significantly different in various Web 2.0 environments and based on the tagger's relationship to the material and generalizations should

## *FUTURE RESEARCH: USER-CENTERED STUDIES*

In order to understand the ways in which different types of tagging may differ in the online environment, studies of tagging need to shift the focus from the tags themselves and to

place the emphasis of the study on the user. Examining the artifacts left by users is indeed a straightforward and enlightening line of study that can be accomplished with relatively limited financial resources. The investments in time and resources that have already been made in the study of social tagging have yielded very interesting results, but do not bring the information scientist significantly closer to understanding the underlying rationale for the tags or the motivation for tagging, either for endo- or exo-tags.

Future studies, therefore, should focus at least in part on the users. By making users the center of study, speculation about motivation could be confirmed through a move toward greater understanding. Research into reasons for tagging can be investigated on several levels. For example, it has been suggested anecdotally that users have more of a feeling of ownership when they actively tag an online resource by applying an exo-tag. The extent to which this is true would be a worthwhile subject of study. An online survey of taggers may be one of the best ways to begin this type of user-centered research. By carrying out face-to-face interviews or in-person observation of taggers as they work, information scientists could also begin to learn more about the motivation and rationale behind tagging.

In the case where continued system-centered study of the tags is preferable, there remain a plethora of options for further study that focus on the community of users. For example, by focusing on the tagging process at the community level across different kinds of social tagging Web sites, information scientists can study the interrelationship of users within the community. The inclusion of other types of complimentary artifacts such as “comments” on uploaded material, “friending” or rating uploaded content, or the simple act of consuming by tagging resources with the community’s tags are also valid points of interest. Studying a group’s tags in situations where taggers are alternately authors and consumers of resources or where taggers are

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

tagging different types of resources would be helpful to understanding the rationale for tagging in each environment and by each group. Sites such as Technorati may be helpful for pulling together tags from a variety of online resources.

It would also be worthwhile to investigate one community member's tags over time and to assess his or her interaction with other members of the community through the feedback loops like "comments" that are available in many photo or video sharing sites. If the status of a given group member could be ascertained, such as "novice" or "popular", this might inform the perception of the user's place in the group, thereby impacting the tags that he or she assigns in a given system or the comments that the user makes about the content of others. Several video sharing sites allow for users to upload text or explanations along with their materials. In this way, users have the option of making clear their place in the community by contextualizing their resource. The feedback that community members provide for each other on video sharing sites helps situate the relative popularity of a given post within the community. A robust assessment of the tags in context could logically follow accordingly.

## ***CONCLUSION***

Although relatively little is known about social tagging, this popular and timely topic has captured the attention of many information scientists. The challenge now is to refine the understanding of the phenomenon while expanding knowledge of user-centered classification in general. It may be helpful to consider that authors and users may view online resources differently and may be endo- or exo-tagging based on very different motivations. As a way of developing a more robust understanding of user motivation, user-centered studies should be

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

carried out by researchers in LIS. This next step is essential to understanding how taggers view the act of tagging, giving insight into user expectations and motivation in terms of the organization of information in the Web environment.

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

## *REFERENCES*


Ames, M., & Naaman, M. (2007). *Why we tag: Motivations for annotation in mobile and online media*. Paper presented at the CHI 2007, San Jose, CA, April 28-May 3, 2007.

Corrado, E. (2007). *Social tagging: Community tagging or personal tagging in communities? Part of Panel: Tagging and Social Networks: The Impact of Communities on User Centered Tagging*. Panel presented at ASIS&T 2007, Milwaukee, WI, October 23, 2007.

Golder, S. & Huberman, B. A. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science, 32*(2), 198-208.

Guy, M., & Tonkin, E. (2006). Folksonomies: Tidying up tags? *D-Lib Magazine, 12*(1).

Kipp, M. E. I. (2007). *Tagging practices on research oriented social bookmarking sites*. Paper presented at the Canadian Association for Information Science, Montreal, Quebec.

Le Deuff, O. (2006). Folksononomies: Les usagers indexent le web. *Bulletin des Bibliothèques de France, 51*(4), 66-70.

Lin, X., Beaudoin, J. E., Bui, Y., & Desai, K. (2006). Exploring characteristics of social classification. In Jonathan Furner and Joseph T. Tennis (Ed.), *Proceedings of the 17th ASIS&T SIG/CR Classification Research Workshop* (Vol. 17). Austin, TX, November 4, 2006: ASIS&T.

Marlow, C., Naaman, M., boyd, d., & Davis, M. (2006). *HT06, tagging paper, taxonomy, Flickr, academic article, to read*. Paper presented at HT'06, August 22-25, Odense, Denmark.



Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.

Mathes, A. (2004). Folksonomies: Cooperative classification and communication through shared metadata. Graduate School of Library and Information Science, University of Illinois Urbana-Champaign.

Russell, T. (2008). *Watching organizational opinion via social tagging.* Poster presented at the iConference, Los Angeles, CA, February 27, 2008.

Shirky, C. (2005). Ontology is overrated: Categories, links, and tags. *Clay Shirky's writings about the Internet*. http://www.shirky.com/writings/ontology_overrated.html (viewed December 15, 2007)

Tennis, J. T. (2006). *Social tagging and the next steps for indexing*. Paper presented at the ASIS&T 17th SIG/CR Classification Research Workshop, Austin, TX, November 4, 2006.

Walker, J. (2005). *Feral hypertext: when hypertext literature escapes control*. Paper presented at HT'05, Salzburg, Austria, September 6-9, 2005.

Citation: Moulaison, Heather Lea. (2008). Social tagging in the Web 2.0 environment: Author vs. user tagging. *Journal of Library Metadata*, *8*(2), 101-111.