Proceed with Caution: Deepening Practitioner Concerns about Social Tagging within Digital Collections

Edward Benoit III and Amanda L. Munson

abstract: Based on a comparison of two nearly identical surveys in 2010 and 2016, this study examines the deepening practitioner concern over the use of social tagging within digital collections during the past six years. While participants reported a greater level of experience and increased use of social tagging, concerns regarding control, consistency, and potential abuse strengthened over the six years of the study. The authors discuss several potential causes for this increase and suggest future directions for social tagging.

Introduction

The number of digital collections has quickly grown over the past decade. As the participatory movement took hold between about 2005 and 2010, some institutions began to allow social tagging within their digital collections—that is, they employed keywords generated by users rather than specialists to classify and describe their online content. Early in the development of social media, the research community began exploring social tagging, its potential benefits, and its limitations. The research soon shifted to include tagging within traditional information retrieval systems such as databases, online public access catalogs (OPACs), and digital libraries. The studies continued expanding through explorations of tagging development, organization, and the taggers themselves. While several studies highlight problems with tagging consistency and tagging abuse, few considered the practitioner’s perception of social tagging itself.
Considering the opinions and concerns of digital librarians and archivists serves a vital role in the development of new digital platforms that better integrate tagging and quality assurance tools. A 2010 study of practitioners' views regarding social tagging in digital collections found a generally positive perception, with limited concerns regarding control and consistency. With the further implementation of tagging in digital collections during the past six years and the increased use of social media, one might expect a continued favorable view of social tagging. However, by definition, early adopters view new technology positively, while the early and late majority adopters are more skeptical. Likewise, the participants in the 2010 study had little practical experience with tagging in their collections. Their opinions might have shifted as theoretical opportunities encountered experiential realities.

The following article compares the findings of the original 2010 study with one conducted in 2016 and addresses the following research question: How have digital librarians' and archivists' perceptions regarding the integration of social tagging within digital collections changed between 2010 and 2016? What concerns, if any, intensified?

**Literature Review**

There have been several studies on the use of social tagging in the digital collections of libraries, archives, and museums. Many investigations focus on measuring the benefits and limitations of tagging or on ways to improve it. Most studies center on one of several themes: comparing tags with controlled vocabularies, tagging as a method to expand traditional metadata (or fill in gaps), the use of tags to increase collection access, and concerns about the use of tags. The following section highlights studies related to each of these themes.

**Testing Uncontrolled Vocabulary**

One of the largest areas of tagging research explores how user-generated tags compare with more traditional controlled vocabularies, such as Library of Congress Subject Headings (LCSH). Harry van Vliet and Erik Hekman conducted one such study to determine if tags generated by nonprofessionals differed in quality from those created by experts. They found that experts and nonprofessionals used different vocabulary. Museum personnel ranked the expert tags as more informative and the nonprofessional tags as more effective search terms. The museum personnel may have had a slight bias, however. When the researchers compared the tags on their own, they found the tags created by both groups equally informative. Van Vliet and Hekman were not the only researchers to question whether professional index terms differed from user-generated terms. Christine DeZelar-Tiedman compared user-generated tags in academic library catalogs with LCSH to determine the benefits of using tags to catalog items. The records she examined had six times as many tags as subject headings, and 53.9 percent of the tags contained concepts that the LCSH did not cover. In Athens, Greece, Constantia Kakali used data from Panteion University Library and its social OPAC as well as the results of a national survey of 81 cataloging librarians to study social tags. She found that 57.82 percent of the tags in the study
matched LCSH terms. She also determined that catalogers frequently agreed with the
tagger’s choice. Unlike many other researchers, who declared controlled vocabularies
better than social tags or vice versa, Kakali concluded that new and improved controlled
vocabularies could be created from tags.

Praveenkumar Vaidya and N. S. Harinarayana conducted a study like those by
DeZelar-Tiedman and Kakali. They compared the tags assigned to 100 titles on Libr-

aryThing, a social tagging Web application, with the LCSH terms that appeared on the
machine-readable cataloging (MARC) records for the same titles. They found that 78.5
percent of the tags had no equivalent LCSH term, 46.9 percent of the LCSH terms were
also used by the taggers, and 88 percent of the titles had at least one tag that matched
an LCSH term used in the MARC record. The tags Vaidya and Harinarayana observed
contained bibliographic information, personal references, and opinions about the books.
The tags that did not overlap with LCSH terms, they said, revealed “much information about the sources” and acted “as a bridge between the professional terms.” They
named “homographs, synonyms, and polysemy” as the main limitations of the tags and
an increase in access points as the main benefit.

In 2006, Krystyna Matusiak tested the ability of user-created tags to make images in
digital collections more accessible through user searches. Describing image collections
has long been a difficult task due to the lack of text and the many possible interpretations
of an image. Matusiak believed that a tagging system would allow users to describe im-
ages from their own unique perspective, which would create more searchable metadata
and more access points. She compared the description of images in a digital library to
tags applied to images on the photo-sharing website Flickr and found that tags alone
failed to provide adequate descriptions but could supplement metadata and include
more diverse points of view.

Expanding Metadata

The item-level description used by most digital collections requires significant labor
from metadata specialists. Likewise, not every item’s content can be identified or de-
scribed within a limited metadata framework. Social tags can help expand traditional
metadata, identify the unidentified, and fill in gaps. The Library of Congress (LOC)
conducted its own experiment with tagging and user comments to see if it could improve
the visibility and accessibility of its image collections. Michelle Springer and her LOC
colleagues uploaded images from two of

the library’s historical photographs collections to Flickr. Within less than a year, users
created more than 67,000 tags and left over 7,000 comments, enhancing the records of
more than 500 items in the LOC’s Prints and Photographs Online Catalog.

Increased visibility of digital resources is far from the only recorded benefit of tag-
gging. Marek Sroka reported on the use of social tagging by Jewish cultural institutes to
identify Holocaust victims in photographs taken before and during World War II. He
noted two institutions that used this method to great success: The first was the Center
for Jewish History in New York City, which placed its photos on Flickr for users to tag with information. The second was the Virtual Shtetl—an online, Polish-Jewish history archive that allowed users to add descriptions to photos. Tagging has also benefited music cataloging. Paul Lamere explained that music catalogers have difficulties in creating genre hierarchy to use for taxonomic purposes. He stated that tagging systems allow users to tag pieces of music with multiple genres, thereby creating their own hierarchy. Jean-Yves Delort also noted that social tags could make automatic indexing systems more effective and accurate by revealing which terms users employ more frequently to describe certain items.

Finally, tagging can help complete description for items with minimal metadata. Edward Benoit III, one of the authors of this article, examined the use of domain-expert tags as supplemental metadata for minimally processed digital archives. The study participants created tags for a set of photographs and letters based on their assessed prior domain knowledge. Although Benoit concluded that domain knowledge could not serve as a controlling mechanism, the generated tags could expand the existing metadata.

**Increasing Access and Use**

Since the adaptation of social tagging in libraries, archives, and museums, its advocates promote their ability to increase access and use of collections. Not surprisingly, several studies explore how users leverage tags for increased access of digitized materials using natural language rather than traditional controlled vocabularies. Sebastian Chan stated that serendipitous discovery in libraries, archives, and museums was an important aspect of user searches. He noted that tagging systems have the benefit of re-creating serendipitous discovery when integrated into library online public access catalogs. Linda Zajac studied the Philadelphia Museum of Art’s website, which included a user tagging system and a list of randomly selected tags for users to browse. Though Zajac did not mention serendipitous discovery in her analysis, the browsing feature she described was possibly meant to re-create the sense of serendipity in a digital environment.

Jen Pecoskie, Louise Spiteri, and Laurel Tarulli studied the effect of user-generated content on readers’ advisory services in Canadian libraries. They found that titles had more tags than subject headings assigned to them overall, but just under one-third of the bibliographic records in the study had no tags at all. The tags included single categories, compound categories, and atypical compound categories. Some tags consisted of administrative notes or personal notes about the titles but lacked information that was standard in MARC records, such as authors’ names. Pecoskie and her colleagues stated that tags improved the ability to search for titles based on subject, protagonists, awards won, and the tone or mood of the story. The tags also tended to focus on the emotional aspects of the stories, whereas subject headings concentrated more on objective facts. Pecoskie, Spiteri, and Tarulli explained that the inclusion of the emotional information...
serves as a reader advisory because users often want that type of information before deciding what to read. They also stated that tagging and other user-generated content create a sense of community for the readers.23

Lucy Clements and Chern Li Liew studied the effects of social tagging in the Auckland Libraries system in New Zealand. They specifically looked at the ways in which library staff used the tags in their work and any problems they experienced. Clements and Liew interviewed 12 staff members ranging from library assistants to specialized librarians. The interviewees stated that they used the tags for keyword searching and browsing. One said that it made searching for works by a certain author easier because titles were tagged with the author’s name. Another said that performing several quick keyword searches for tags was easier than looking up the proper LCSH and searching for materials cataloged with that heading. Other librarians used the tagging system to add tags to titles when they felt that no subject heading fit the exact genre of a work. Still others mentioned that the tags allowed staff and users to apply local slang and spellings to the resources, which made them more culturally relevant to New Zealanders than the American-created LCSH. Some interviewees disliked the lack of structure in the tags and preferred to use LCSH instead, but others found that sorting features built into the tagging system reduced these issues. Clements and Liew stated that most of the interviewees believed the tags benefited the library system.24

Tagging Issues

The final theme of tagging research focuses on perceived (or realized) issues of tags and proposed solutions to address these issues. Although some archivists and librarians complain about the inconsistencies in user-created tags, Margaret Kipp and D. Grant Campbell argue that these inconsistencies, when made by multiple users, indicated relationships between the tags and shared index terms. In other words, users tried to tag items with the same term or taxonomy, but due to common misspellings, grammatical errors, and the like, they created separate tags instead.25 In a solo study, Kipp also found that tags often represented emotions, tasks to be completed, and time.26 Seth van Hooland, Eva Méndez Rodríguez, and Isabelle Boydens observed similar results in their study of the steve.museum project, a social tagging effort to improve public access to art in United States museum collections. Van Hooland and his coauthors reached a different conclusion than Kipp did, however. They argued that social tagging did not benefit libraries, archives, and museums because the user-generated metadata were too contemporary. Tags that reflect the perspectives and search
methods of one generation might not be useful to future generations, who would likely have different perspectives and search methods. Van Hooland, Méndez Rodríguez, and Boydens also found that many tags contained terms too general to be useful in a search and that many users created tags in only one session, showing that users were less engaged with the collections than steve.museum researchers liked to believe. Van Hooland’s team instead recommended user comments, stating that comments allowed users to provide information that could correct erroneous metadata or tell a story about an item in the collection, thus leading to a higher level of engagement.

Some researchers have tested methods for correcting the issues van Hooland and his colleagues reported. For instance, Brian Matthews, Catherine Jones, Bartłomiej Puzoń, Jim Moon, Douglas Tudhope, Koraljka Golub, and Marianne Lykke Nielsen tested whether knowledge organization systems such as the Dewey Decimal Classification system could limit the creation of general or misspelled tags by controlling the vocabulary with preset terms. They found that participants liked the freedom and ease of use of the uncontrolled tagging system but felt that the system needed to suggest terms. Participants who used a knowledge organization system liked the consistency of the controlled vocabulary but only appreciated the suggested terms when they believed them useful. Some participants complained that irrelevant suggestions and the number of steps required to tag documents made the enhanced tagging system inefficient. Some believed the vocabulary too restrictive, and others stated that the limits on vocabulary would make the users think more deeply about the subjects of the papers. Overall, the researchers found that most participants liked the consistency of a controlled vocabulary and found tagging in general provided more ways to access items than regular subject indexing did.

Several years before the Matthews team’s study, Marieke Guy and Emma Tonkin argued that system-suggested tags reduced inconsistencies inherent in user-created terms but also encouraged hegemony, thereby reducing the diversity in the tags. Clearly, a need exists for a less biased way of making tags more consistent. In a study of the National Library of Australia’s Australian newspapers collection, Rose Holley revealed that users made a concerted effort to create consistency in tags when no tagging rules existed. The users developed their own guidelines designed to ensure their fellow researchers could find the information they needed. These guidelines included using natural order for names and phrases as opposed to listing the surname or most general term first, putting spaces between words and punctuation marks when necessary, employing hyphens to indicate subject hierarchy (similar to LCSH), using terms that were meaningful to others, adding occupations and dates of birth and death after the names of individuals, and allowing the creation of tags for personal reference. Another proposed solution to the problem of inconsistency is Spiteri’s suggestion that tagging systems adhere to National Information Standards Organization (NISO) guidelines. However, Spiteri also notes that NISO does not allow slang terms or jargon, which would limit the amount of diversity in tags much as Guy and Tonkin claimed system-suggested terms would.

Another issue with social tagging is the potential for user abuse. Georgia Koutrika, Frans Adjie Effendi, Zoltán Gyöngyi, Paul Heymann, and Hector García-Molina studied the impacts of spam and user abuse on tagging systems and possible ways to prevent them. The researchers created an algorithm to calculate to what degree each user’s tagging
activity conformed to the activities of other taggers. They then used this algorithm to locate tags created by malicious users. Koutrika and her team found that a small number of malicious users operating in an extensive digital collection with many non-abusive taggers did not significantly affect search results when they used their algorithm. Spam affected Boolean searches more. However, when the researchers introduced a moderator into the system to vet the tags, the Boolean searches improved. Koutrika and her co-authors also simulated a large-scale, coordinated spamming attack that negatively affected both Boolean search systems and search systems based on the concentration of tags. The algorithm proved inadequate as a countermeasure in that instance, but Koutrika and her team stated that moderators employed to check the tags would likely be an effective means of combating targeted attacks.35

Despite the abundance of research on social tagging’s applications in libraries, archives, and museums and its benefits and limitations, few studies focus on the perceptions and experiences of the library and archival professionals who work with digital collections. Like Clements and Liew’s examination of how general librarians view and use tags, this study focuses on practicing digital librarians and archivists, exploring their perceptions of the use of tagging in digital collections, its benefits, and its limitations.

Methodology

As noted earlier, this article compares the findings from two nearly identical studies conducted in 2010 and 2016. The original study began with the analysis of 10 semi-structured interviews of digital librarians regarding their personal and professional use of social tagging as well as their opinion of its place within digital collections. The interviews elicited mixed opinions, suggesting the need for further data gathering through a more substantial survey.

The 2010 survey included a combination of open- and closed-ended questions (30 total) based on the semi-structured interviews and was hosted on the Qualtrics online survey platform. Participation invitations were posted on eight different e-mail lists popular among digital librarians and archivists. The survey remained open for three weeks and collected data from 112 participants. The second study, in 2016, closely followed the earlier study with some minor adjustments to its 32 questions. For example, open-ended questions regarding the location of participants’ repositories were replaced with dropdown lists. The remaining questions for the 2016 version were based on the original 2010 study, and semi-structured interviews were not conducted. Participant invitations for the 2016 version used five e-mail lists.1 The survey remained open for three weeks and garnered 102 total respondents, with 73 completing the survey in its entirety.

The research question requires a longitudinal analysis of data from both surveys. A comparison of the demographic information of both sample populations indicates a relatively homogeneous cross sample population. Table 1 shows the comparative percentages for participant gender, age, and educational background. While the demographic data from the two surveys are similar, the types of institution where participants are employed differ slightly between 2010 and 2016 (see Table 2). The more recent study included more academic libraries and fewer institutions listed as “other” than in 2010. Several of the 2010 participants indicated working for library consortiums under the “other” category. Perhaps the transition toward more in-house digital collections departments within academic libraries explains the difference between 2010 and 2016.
Table 1.
Demographics of the sample population, 2010 versus 2016

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>29.5%</td>
<td>20.5%</td>
</tr>
<tr>
<td>Female</td>
<td>70.5%</td>
<td>79.6%</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
</tr>
<tr>
<td>18–25</td>
<td>3.6%</td>
<td>4.1%</td>
</tr>
<tr>
<td>26–35</td>
<td>39.3%</td>
<td>37.0%</td>
</tr>
<tr>
<td>36–45</td>
<td>18.6%</td>
<td>17.8%</td>
</tr>
<tr>
<td>46–55</td>
<td>13.4%</td>
<td>12.3%</td>
</tr>
<tr>
<td>56–65</td>
<td>13.4%</td>
<td>12.3%</td>
</tr>
<tr>
<td>65+</td>
<td>0.0%</td>
<td>2.7%</td>
</tr>
<tr>
<td>Education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Some college</td>
<td>0.9%</td>
<td>1.4%</td>
</tr>
<tr>
<td>4-year college degree</td>
<td>9.8%</td>
<td>5.5%</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>79.5%</td>
<td>82.2%</td>
</tr>
<tr>
<td>Doctoral degree</td>
<td>8.0%</td>
<td>9.6%</td>
</tr>
<tr>
<td>Professional degree</td>
<td>1.8%</td>
<td>1.4%</td>
</tr>
</tbody>
</table>

Table 2.
Workplaces of the sample population, 2010 versus 2016

<table>
<thead>
<tr>
<th></th>
<th>2010</th>
<th>2016</th>
</tr>
</thead>
<tbody>
<tr>
<td>Institution type</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Academic library</td>
<td>46.4%</td>
<td>62.4%</td>
</tr>
<tr>
<td>Archives/Historical society</td>
<td>14.3%</td>
<td>12.9%</td>
</tr>
<tr>
<td>Museum</td>
<td>6.3%</td>
<td>2.0%</td>
</tr>
<tr>
<td>Public library</td>
<td>4.5%</td>
<td>5.9%</td>
</tr>
<tr>
<td>Special library</td>
<td>1.8%</td>
<td>4.0%</td>
</tr>
<tr>
<td>Other</td>
<td>26.8%</td>
<td>12.9%</td>
</tr>
<tr>
<td>Institution location</td>
<td></td>
<td></td>
</tr>
<tr>
<td>United States</td>
<td>90.2%</td>
<td>83.2%</td>
</tr>
<tr>
<td>Non-United States</td>
<td>9.8%</td>
<td>16.8%</td>
</tr>
</tbody>
</table>
The remaining data from the two studies were combined into a single data set for further comparative analysis. Several of the variables required recoding with SPSS, a software package used for statistical analysis, due to reversed Likert scales between 2010 and 2016. This was the only data manipulation required during the merging process. Since the surveys used ordinal dependent data (that is, noncontinuous scales such as Likert scales) and the data were not normally distributed, the comparative statistical analysis used Mann-Whitney U tests to determine whether the two studies were significantly different from each other.

**Results**

This article presents the results of the comparative analysis in four sections: use of social tagging, institutional use of social tagging, perceptions of social tagging within digital collections, and views on potential system features.

**Use of Social Tagging**

The surveys asked participants to indicate their knowledge of social tagging as well as their experience using tagging personally, professionally, or both. When asked to rank their knowledge of social tagging (1 = no knowledge; 5 = very knowledgeable), the 2010 participants indicated a significantly higher median (4) than the 2016 participants (3), U = 3722.5, z = –2.625, p = 0.009, where U assesses whether the two studies were significantly different from each other; z measures how far the score diverges from the mean; and p estimates the probability that the result has occurred by statistical accident. A low level of p indicates a high level of statistical significance. Personal use of social tagging increased slightly from a median of once per month to once per week, but the increase was not significant, U = 1956.5, z = –1.842, p = 0.065. The comparison of professional use of social tagging, however, found a statistically significant increase in use from once per six months to once per week, U = 865, z = –3.089, p = 0.002. Participants who indicated they had neither professional nor personal experience using social tagging were asked to indicate their reasons for not using tagging from a list of five possibilities. Significantly more participants agreed with four of the justifications in the earlier survey. Table 3 reports the associated Mann-Whitney U test results for these statements with statistically significant results highlighted in gray.

**Institutional Use of Social Tagging**

Participants answered several questions regarding their institution’s use of social tagging within and outside of digital collections. Figure 1 shows the comparative percentage of affirmative responses for each question. Between 2010 and 2016, participants reported a statistically significant increase in the number of institutions allowing social tagging or user commenting, U = 3314, z = –2.244, p = 0.025. The increase in the percentage of institutions using social tagging within digital collections was also statistically significant, U = 3319.5, z = –2.431, p = 0.015. Although the inclusion of user commenting within and outside of digital collections also rose between 2010 and 2016, the increase was not statistically significant.
Table 3.
Reasons for not tagging, 2010 versus 2016

<table>
<thead>
<tr>
<th>Reason for not tagging</th>
<th>U*</th>
<th>z†</th>
<th>p‡</th>
<th>2010 median</th>
<th>2016 median</th>
<th>2010 mean</th>
<th>2016 mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>I haven’t had the opportunity.</td>
<td>7</td>
<td>−3.651</td>
<td>0.001</td>
<td>2</td>
<td>1</td>
<td>1.86</td>
<td>1</td>
</tr>
<tr>
<td>I am not familiar with tagging.</td>
<td>15</td>
<td>−2.581</td>
<td>0.026</td>
<td>2</td>
<td>1</td>
<td>1.64</td>
<td>1</td>
</tr>
<tr>
<td>It doesn’t benefit my life or work.</td>
<td>15</td>
<td>−2.581</td>
<td>0.026</td>
<td>2</td>
<td>1</td>
<td>1.64</td>
<td>1</td>
</tr>
<tr>
<td>I don’t use systems that allow tagging.</td>
<td>22</td>
<td>−3.162</td>
<td>0.005</td>
<td>2</td>
<td>1</td>
<td>1.71</td>
<td>1</td>
</tr>
<tr>
<td>I have privacy concerns.</td>
<td>12</td>
<td>−1.368</td>
<td>0.3</td>
<td>1</td>
<td>1</td>
<td>1.43</td>
<td>1</td>
</tr>
</tbody>
</table>

Statistically significant results are highlighted in gray.

*U assesses whether the two studies were significantly different from each other.
†The z-score is a measure of how far the score diverges from the most probable result, the mean.
‡The p-value is an estimate of the probability that the result has occurred by statistical accident; a low level of p indicates a high level of statistical significance.
Perceptions of Social Tagging

The surveys asked participants to indicate their agreement or disagreement with 17 statements focused on the benefits or limitations of social tagging within digital collections. The mean responses for all statements increased, indicating a higher level of agreement in 2016. Comparative analysis of their responses found significant differences between 2010 and 2016 for five statements (see Table 4). The benefit statements account for only one of the significant increases, “Social tags identify the previously unidentifiable,” U = 3595, z = 2.151, p = 0.031. Participants were more concerned with potential issues and abuses of tagging systems in 2016, as shown by the significant increase in agreement with statements regarding uncontrollability, troublesome users, consistency issues, and sensitive topics.

The open-ended responses echo several of these findings. Both surveys asked participants for their general opinions regarding the use of social tagging within digital collections. A comparative analysis of the responses identified key differences between 2010 and 2016. The initial survey participants in 2010 framed their responses in terms of the possibilities of social tagging, with many stating that they lacked direct experience...
## Table 4.
Perceptions of social tagging, 2010 versus 2016

<table>
<thead>
<tr>
<th></th>
<th>U*</th>
<th>z†</th>
<th>p‡</th>
<th>2010 median</th>
<th>2016 median</th>
<th>2010 mean</th>
<th>2016 mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social tagging engages users with the collection(s).</td>
<td>3169</td>
<td>0.498</td>
<td>0.618</td>
<td>4</td>
<td>4</td>
<td>3.95</td>
<td>4.01</td>
</tr>
<tr>
<td>Social tags identify the previously unidentifiable.</td>
<td>3595</td>
<td>2.151</td>
<td>0.031</td>
<td>4</td>
<td>4</td>
<td>3.73</td>
<td>4.00</td>
</tr>
<tr>
<td>Social tags expand the metadata for a given item.</td>
<td>3415</td>
<td>1.49</td>
<td>0.136</td>
<td>4</td>
<td>4</td>
<td>3.94</td>
<td>4.12</td>
</tr>
<tr>
<td>Social tags complete the metadata for a given item.</td>
<td>3373.5</td>
<td>1.877</td>
<td>0.06</td>
<td>3</td>
<td>3</td>
<td>2.68</td>
<td>2.96</td>
</tr>
<tr>
<td>Social tags allow inter-user conversation related to the collection/items.</td>
<td>3031.5</td>
<td>-0.038</td>
<td>0.969</td>
<td>4</td>
<td>4</td>
<td>3.75</td>
<td>4.00</td>
</tr>
<tr>
<td>Social tags should be integrated with metadata.</td>
<td>2704</td>
<td>0.625</td>
<td>0.532</td>
<td>3</td>
<td>3</td>
<td>3.24</td>
<td>3.34</td>
</tr>
<tr>
<td>Social tags cannot be trusted.</td>
<td>3302.5</td>
<td>1.299</td>
<td>0.194</td>
<td>3</td>
<td>3</td>
<td>2.73</td>
<td>2.91</td>
</tr>
<tr>
<td>Social tagging is too new of a technology to spend time implementing within collections.</td>
<td>3247.5</td>
<td>0.393</td>
<td>0.693</td>
<td>2</td>
<td>2</td>
<td>2.30</td>
<td>2.42</td>
</tr>
<tr>
<td>Social tags are too uncontrolled to be used as access points for items.</td>
<td>3650.5</td>
<td>2.27</td>
<td>0.023</td>
<td>2</td>
<td>3</td>
<td>2.58</td>
<td>2.96</td>
</tr>
<tr>
<td>Allowing social tagging will overload the collection with information.</td>
<td>3326.5</td>
<td>1.074</td>
<td>0.283</td>
<td>2</td>
<td>2</td>
<td>2.29</td>
<td>2.44</td>
</tr>
<tr>
<td>I am concerned over possible troublesome users.</td>
<td>3513.5</td>
<td>2.093</td>
<td>0.036</td>
<td>3</td>
<td>4</td>
<td>3.14</td>
<td>3.43</td>
</tr>
<tr>
<td>Users will input inappropriate terms (i.e., expletives).</td>
<td>3430</td>
<td>1.818</td>
<td>0.069</td>
<td>3</td>
<td>3</td>
<td>2.99</td>
<td>3.21</td>
</tr>
<tr>
<td>I am concerned over the consistency of tags.</td>
<td>3484</td>
<td>1.993</td>
<td>0.046</td>
<td>4</td>
<td>4</td>
<td>3.34</td>
<td>3.65</td>
</tr>
<tr>
<td>Tagging within sensitive topic collections should be limited.</td>
<td>3719</td>
<td>2.882</td>
<td>0.004</td>
<td>3</td>
<td>3</td>
<td>3.00</td>
<td>3.41</td>
</tr>
<tr>
<td>Older users will not understand a tagging system.</td>
<td>3155.5</td>
<td>0.74</td>
<td>0.459</td>
<td>3</td>
<td>3</td>
<td>2.80</td>
<td>2.92</td>
</tr>
<tr>
<td>I am concerned about the migration of tags to future systems.</td>
<td>3281</td>
<td>1.232</td>
<td>0.218</td>
<td>4</td>
<td>4</td>
<td>3.47</td>
<td>3.69</td>
</tr>
<tr>
<td>Social tags should remain separated from the official metadata of a collection.</td>
<td>3124</td>
<td>0.621</td>
<td>0.534</td>
<td>3</td>
<td>4</td>
<td>3.37</td>
<td>3.48</td>
</tr>
</tbody>
</table>

Statistically significant results are highlighted in gray. *U assesses whether the two studies were significantly different from each other; †The z-score is a measure of how far the score diverges from the most probable result, the mean; ‡The p-value is an estimate of the probability that the result has occurred by statistical accident; a low level of p indicates a high level of statistical significance.
but viewed the potential benefits positively. For example, one participant said, “Social tagging is potentially useful for metadata augmentation, comments are definitely a useful metadata feedback mechanism if the resources are available to respond and implement change.” Another 2010 participant stated, “I think there is potential, but no one is sure how to make it work yet.” In contrast, more of the 2016 participants reflected on their own experiences with tagging. As one noted, “I am very interested in social tagging as a crowd-sourcing tool, and we undertook an outreach project to try and get alumni to comment and tag university photos through our digital collections, but response has been very limited.” Another stated, “I think it is immensely valuable to digital collections” (emphasis added).

Reflecting the Likert-based questions, more participants in 2016 (34.6 percent) than in 2010 (23.4 percent) noted concerns regarding quality control and consistency. These issues mainly focused on perceptions rather than experience. As one 2016 participant stated, “I think it would be wonderfully useful and encourage the cross-pollination of ideas across disciplines. Unfortunately, there is a fear that we would lose control of cataloging at our institution if we opened it up to the ‘masses’ despite my assurances to the contrary.” Others focused on how tagging “can sometimes get really messy, with misspellings or mis-identification of people” and complained, “Vetting that info can be problematic, as is getting user participation.” A participant in the 2016 study said, “My thoughts are somewhat mixed. While we have received some really useful information from users . . . we’ve also received relatively useless information that still requires staff time to review/ve/manage [sic].”

The comparison of open-ended responses also found a decrease in the number of participants with system-based issues preventing them from implementing social tagging from 23.4 percent in 2010 to 15.4 percent in 2016. This drop reflects technological development since 2010, when major digital collection content management systems, such as CONTENTdm®, did not support social tagging as they do now. Other issues arose in the responses in low numbers, such as calls for additional research, lack of user response, and the need for longitudinal development of tagging within systems.

**Potential System Features**

Based on the findings of the original pilot project, the surveys presented participants with a range of potential features within a digital collection system that could help ease their concerns. Table 5 lists the presented features with their associated comparative analyses. Participants ranked each feature on a scale from 1 (not important) to 5 (extremely important). Only two of the 14 features received a lower mean score in 2016 than in 2010—a user monitoring system and new tag notification—neither of which was statistically significant. The remaining 12 features were all viewed as more important in 2016 than in 2010, with more than half (seven) experiencing statistically significant increases. Those seven features were...
Table 5.
Perceptions of potential tagging system features, 2010 versus 2016

<table>
<thead>
<tr>
<th>Feature</th>
<th>2010 Median</th>
<th>2016 Median</th>
<th>2010 Mean</th>
<th>2016 Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>A user monitoring system</td>
<td>4</td>
<td>4</td>
<td>3.65</td>
<td>3.59</td>
</tr>
<tr>
<td>Notification every time a new tag is added to the collection</td>
<td>3</td>
<td>3</td>
<td>3.25</td>
<td>3.21</td>
</tr>
<tr>
<td>Approval required for every tag added to the collection</td>
<td>3</td>
<td>3</td>
<td>2.61</td>
<td>2.89</td>
</tr>
<tr>
<td>Ability to block specific users</td>
<td>4</td>
<td>4</td>
<td>3.71</td>
<td>4.21</td>
</tr>
<tr>
<td>Requiring users log in prior to creating tags</td>
<td>4</td>
<td>4</td>
<td>3.23</td>
<td>3.49</td>
</tr>
<tr>
<td>Automatic transfer of tags into searchable metadata</td>
<td>3</td>
<td>3</td>
<td>3.00</td>
<td>3.34</td>
</tr>
<tr>
<td>Requiring indexing approval prior to transfer of tags into searchable metadata</td>
<td>3</td>
<td>4</td>
<td>3.20</td>
<td>3.67</td>
</tr>
<tr>
<td>Independent search systems for official metadata and social tags</td>
<td>2</td>
<td>3</td>
<td>2.37</td>
<td>2.78</td>
</tr>
<tr>
<td>Spell-checking software within the tagging system</td>
<td>3</td>
<td>4</td>
<td>2.56</td>
<td>3.77</td>
</tr>
<tr>
<td>Displaying previously added tags</td>
<td>4</td>
<td>4</td>
<td>3.39</td>
<td>3.78</td>
</tr>
<tr>
<td>A user approval system (such as thumbs up or thumbs down)</td>
<td>3</td>
<td>3</td>
<td>2.80</td>
<td>3.05</td>
</tr>
<tr>
<td>Social tagging tutorials</td>
<td>4</td>
<td>4</td>
<td>2.99</td>
<td>3.59</td>
</tr>
<tr>
<td>Separate systems for digital collection and tagging</td>
<td>3</td>
<td>3</td>
<td>2.30</td>
<td>2.70</td>
</tr>
<tr>
<td>Integrated systems for digital collection and tagging</td>
<td>3</td>
<td>3</td>
<td>2.58</td>
<td>3.34</td>
</tr>
</tbody>
</table>

Statistically significant results are highlighted in gray.

*U assesses whether the two studies were significantly different from each other.
†The z-score is a measure of how far the score diverges from the most probable result, the mean.
‡The p-value is an estimate of the probability that the result has occurred by statistical accident; a low level of p indicates a high level of statistical significance.
1. ability to block specific users;
2. requiring indexing approval prior to transferring tags into searchable metadata;
3. independent search systems for official metadata and social tags;
4. spell-checking software within the tagging system;
5. social tagging tutorials;
6. separate systems for digital collections and tagging; and
7. integrated systems for digital collections and tagging.

The final two features appear contradictory; however, individual participants rarely ranked both statements as important—instead selecting one over the other. Spell checking (+1.21), integrated systems (+0.76), social tagging tutorials (+0.6), and user blocking (+0.5) received the highest average increased rankings. Interestingly, the ability to block specific users received the highest mean ranking for both 2010 (mean = 3.71) and 2016 (mean = 4.21). Likewise, separate systems for digital collections and tagging received the lowest average rankings for both 2010 (mean = 2.3) and 2016 (mean = 2.7).

Discussion

The comparison of how practitioners perceive social tagging in digital collections provided surprising results. The expectation was that participants in the 2016 study would report fewer concerns and exhibit a more positive view of social tagging than those in the 2010 study since people tend to become more comfortable with technology over time. The results do not support this expectation. As noted earlier, the more recent study found a significant increase in concerns regarding abuse, consistency, and descriptive control since 2010. Additionally, participants strengthened their calls for tagging system features that specifically address such concerns. The personal and professional use of social tagging has significantly increased since 2010, and more institutions now allow social tagging or user commenting within digital collections than did so then.

Although a direct explanation for the shift in perception remains unknown, several possibilities exist. One view might be that practitioners are reluctant to relinquish descriptive control over their collections or are fearful of new technologies. This view is not reflected in the open-ended responses, nor does it address the change over time. The consideration of when the original study was conducted might provide more insight. Many of the original participants were likely “early adopters” of new technologies who have a generally more positive view of technological developments and their potential benefits. In this case, as social tagging became popularized through social media development in the past six years, more late adopters with lingering concerns about tagging participated in the 2016 study.
The open-ended responses provide another possible explanation. In the original study, participants reflected on the potential use of tagging and often complained about the lack of tagging support in existing systems. This means most participants spoke not from actual experience but instead described what tagging could theoretically achieve. Such a focus on potential is reflective of the literature at the time. Early tagging studies highlighted success stories such as the Library of Congress Flickr project and Sroka’s positive results from the Center for Jewish History and the Virtual Shtetl projects.39 Even studies that recognized potential issues suggested that the usefulness of tags as supplemental metadata would exceed the risks.40

By 2016, many practitioners had moved from viewing tagging’s potential to a more experiential understanding with the growth of digital collections. Therefore, more of the 2016 participants referenced their experience with real collections. People who had a poor experience tended to focus on their individual case rather than consider the broader aspects. Indeed, several of the more recent participants cited a lack of tag development within their digital collections as the justification for their negative views. Not surprisingly, larger institutions such as the Library of Congress have a more extensive user-base to generate tags than smaller, more localized repositories have. Additionally, research on promoting the generation of tags remains lacking. Some more recent studies on gamification and the creation of tagging games show promise, but they are still too limited and lack large-scale testing.41

While some studies such as those of Koraljka Golub and her colleagues and of Guy and Tonkin found the use of system-suggested tags can help standardize tagging and generate more tags through prompting, these types of systems have not been adequately built into digital collection platforms. Likewise, practitioners cannot readily employ potential solutions such as the algorithm of Koutrika and her colleagues—and some may question if the time required to do so would outweigh the benefits of tagging itself. Overall, the deepening concerns of practitioners are not adequately addressed in either the current literature or the development of new tools and digital collection platforms. Although digital collection management systems like CONTENTdm® allow user-generated tags and comments, they remain difficult to moderate and offer few of the features listed in Table 5. Additionally, customizations to propriety software require significant time and expertise. System designers should work more closely with practitioners to identify potential solutions to the issues with tagging as new updates are developed.

Finally, while most statistically significant findings emphasize growing concerns, one should not discount the positive aspects. All the statements of perceived benefits of
tagging, as shown in Table 4, increased from 2010 to 2016. Therefore, it is not necessary
to throw the baby out with the bathwater. While there is room for additional quality
control measures, practitioners still see the potential benefits of tagging.

Conclusion

The use of social tagging in digital collections, as well as other participatory applications,
continues to increase. As a result, digital librarians and archivists increasingly move
toward engaging users with various aspects of their collections. This engagement is
not without issues. While many practitioners initially viewed social tagging as promis-
ing, the past years saw an increase in
their concerns regarding its applica-
tion. A comparison of practitioner
perceptions of social tagging between
2010 and 2016 found stronger calls for
incorporating safeguards within tag-
ging systems, including spell-checking
features, tutorials, the ability to block
users, and requiring that staff review tags before allowing them into the system. Despite
these calls for safeguards, digital librarians and archivists remain optimistic about the
potential benefits of tagging but also wary of uncontrolled and unmediated tags.

The comparative study identified several potential contributing factors for the
change in perception. As more digital collections incorporate user tagging, comment-
ing, or both, system developers need to address the practitioners’ concerns and desire
for additional safeguards within their systems. Likewise, developers should encourage
more direct feedback and utilize methods such as focus groups to elicit more specific
desired features and options.

Academics and researchers must also further engage with practitioners by conduct-
ing more studies in partnership with them and on larger scales. Such studies would
help move discussions from the theoretical realm to practical applications. Additionally,
future research should address different motivating techniques to increase the number
of participating users, as well as the quality and quantity of their tags. These studies
could include reward mechanisms such as point-based systems toward institutional
membership or events, or recognition-based systems.

Moving forward, users will increasingly expect participatory aspects in their en-
gagement of digital content. Therefore, it is imperative to address professional concerns
now, before systems become overly complicated. Additional qualitative and quantitative
studies should explore concerns and solutions at greater depths through combinations
of focus groups, interviews, experiments, and case studies. Only through a mixed meth-
ods approach that brings all stakeholders to the table will true solutions be uncovered.

Edward Benoit III is an assistant professor and coordinator of the Archival Studies program in
the School of Library and Information Science at Louisiana State University in Baton Rouge; he
may be reached by e-mail at: ebenoit@lsu.edu.

Amanda L. Munson received a master’s degree in library and information science in 2017 from
the School of Library and Information Science at Louisiana State University in Baton Rouge.
Notes


10. Ibid.

11. Ibid., 571.


13. Ibid., 41.

14. Ibid.


Proceed with Caution: Deepening Practitioner Concerns about Social Tagging within Digital Collections


25. Kipp and Campbell, “Patterns and Inconsistencies in Collaborative Tagging Systems.”

26. Kipp, “@toread and Cool.”


30. Golub, Moon, Tudhope, Jones, Matthews, Puzoń, and Nielsen, “EnTag.”


32. Guy and Tonkin, “Folksonomies.”


34. Spiteri, “The Structure and Form of Folksonomy Tags.”


36. The 2010 study used the following e-mail lists: DIGLIB (diglib@infoserv.inist.fr), IMGELIB@LISTSERV.ARIZONA.EDU), ContentDM-L (CONTENTDM-L@OCLC.ORG), SAA [Society of American Archivists] Metadata & Digital Object Discussion List (metadata@forums.archivists.org), SAA Visual Materials List (vms@forums.archivists.org), A&A [Archives & Archivists] List (archives@forums.archivists.org), LOOKSEE (listserv@lsu.uky.edu), and Digital Preservation (Digital-Preservation@jiscmail.ac.uk).


38. Rogers, Diffusion of Innovations.
