THEODORE CALVIN PEASE AWARD

Like a Box of Chocolates: A Case Study of User-Contributed Content at Footnote

Pamela H. Mayer

ABSTRACT

User-contributed content has been suggested as a means to narrow the gap between the level of description that resource-constrained repositories are able to provide and the level of description that users need or have come to expect. Research seems to indicate that allowing users to contribute content holds some promise for augmenting traditional description, thus increasing the discoverability of materials. As yet, the practice of allowing user-contributed content has not been widely adopted, especially for large-scale online collections. Because this is not an endeavor to be entered into lightly in terms of required resources or policy considerations, it is important for decision makers to have as much information as possible about who will contribute content and what that content looks like. It is informative to look at the experience of Footnote, an entity with an existing online collection with user contribution functionality. This case study identifies individuals with family connections to a collection as the largest group of contributors, while annotations are the most common type of contribution. The data suggest that users are predominately interested in information about individuals. This study also indicates that there are issues of consistency, authenticity, and context with regard to user-contributed content.

© Pamela H. Mayer. (cc) BY-NC

KEY WORDS

Description, Metadata, Online Collections

Prior to the advent of the Internet and online collections, users were required to visit a repository physically to access cultural heritage materials. With digitization of archival materials, collections of items such as photographs and documents can be made available to users via the Internet. Despite the fact that individual items are now easily accessible, Web-based finding aids still typically present information in the traditional manner—at the collection level. Users are unable to get easily from the finding aid directly to the item they wish to view. This is at odds with the needs and expectations of users who have become accustomed to discovering information through search engines such as Google. Thus, while online digital collections provide ease of access to remote users, access to a specific item or piece of information contained in an online collection is still problematic.

There has historically been a tension between the amount of material to be processed and the level of resources available for processing. Providing the item-level description needed to facilitate user access to online content would require additional resources from already financially overstretched archival institutions. One way to mitigate this lack of resources is to shift the burden of description to the users. Users could mark or annotate information they discover in an item to enable future users to find the same information more easily.

Yet archivists may have reservations about opening up the collections to this type of activity. What will result if resources are allocated to create a means for users to contribute information? One way to explore this question is to look at an existing online collection's experience with user-contribution functionality.

Backlogs and Item-level Description

As professionals charged with collecting and preserving materials of enduring value, archivists are faced with a seemingly endless quantity of materials. Today this deluge of materials continues unabated and has, in fact, grown in magnitude with the addition of born-digital content. In an environment of limited resources, this often results in accessioned materials going without arrangement or description. In the words of Greene and Meissner, "It should be dismaying to realize that our profession has been struggling with backlogs for at least sixty years."[1] The existence of such backlogs has prompted archivists to re-evaluate their processes in an effort to reduce or eliminate backlogs and make materials accessible to users.

In a 2005 article, Mark Greene and Dennis Meissner proposed a strategy to expedite the processing of collections that is known as More Product, Less Process (MPLP).[2] They concluded that all collections should receive a basic level of processing with increased attention given to collections only when warranted.

Additionally, they recommended that "arrangement, preservation, and description work should all occur in harmony, at a common level of detail."[3]

Also in 2005, the National Archives and Records Administration (NARA) began the development of a processing initiative to address a backlog of a million cubic feet. In MPLP fashion, basic processing was considered to be adequate for most collections while those in higher demand would receive more attention.[4] Christine Weideman described a case study in which processing was pared down further to reflect the recommendations of Greene and Meissner.[5] She reported significant reduction of processing time per linear foot of several collections. "Today in [Yale University] Manuscripts and Archives, most processing consists of arrangement at the series level and description in the form of a catalog record and box listing."[6]

Stephanie Crowe and Karen Spilman found that MPLP has been widely adopted by the archival community. However, they noted that further study could focus specifically on the effect of MPLP on descriptive practices.[7]

Not all archivists have embraced MPLP without reservation. While Robert Cox shares Greene and Meissner's "concern for cost efficiency and fitting the level of treatment to the needs of the collection,"[8] he proposed a maximal processing model which, while compatible with MPLP, emphasizes the function of description.

The previously mentioned articles focused on backlogs and processing of analog materials within the larger framework of user access. The same issues arise when considering user access to digitized versions of analog materials. Even with minimal description, materials available online are more accessible than those available only to researchers able to travel to the repository.[9] The University of Wisconsin used a streamlined model of scanning in bulk without item-level metadata which resulted in a significant reduction in cost.[10] Mark Greene, one of the original proponents of MPLP, cited an American Heritage Center survey with findings that suggest that "higher-level description for all collections served researchers as well or better than granular descriptions of a few collections."[11] He rejected the notion that item-level description must be provided. He, instead, favored file- or series-level description. "As MPLP would argue, 'Every dollar spent to make [online] collections perfect is a dollar we're not spending to get another collection online and to a larger potential audience.'"[12]

User Needs and Expectations

This movement by archivists toward minimal descriptive metadata risks running counter to user needs and expectations. "Today we live in an Amazoogle world, where people expect comprehensive information, 24/7, offering immediate gratification, and customized to the customer."[13] As early as 1998, Barbara

Craig noted the effect of computer technology on users' expectations regarding access to materials. She referred to the notion held by some users that everything is available on the Internet as the "myth surrounding information technology."[14]

A usage study of a digital library collection by Wendy Duff and Joan Cherry revealed that users prefer the digital format primarily because of its accessibility. However, users also expect that they should be able to locate relevant material quickly.[15] Lauren Graham mentioned that users of the American Memory collection of digital materials available at the Library of Congress website expect to be able to use the site as they would an encyclopedia.[16]

Meg Sweet and David Thomas, in a report of their experience at the Public Records Office of the U.K. National Archives, noted "that they [collection-level descriptions] are of limited value on their own to a broad range of researchers."[17] In actual practice, many archives users require "clear, accurate and searchable descriptions of individual files."[18]

In an overview of the diverse information-seeking practices of several user groups, Anne Gilliland-Swetland cited genealogists as the most common type of avocational users of archives.[19] This user group can benefit from item-level search and retrieval. She suggested enhancing Encoded Archival Description (EAD) so that users can conduct name and natural language geographic keyword searches. She also suggested that dates be associated with the name or geographic metadata to provide context.

Wendy Duff and Catherine Johnson noted that "provenance-based finding aids provide significant challenges to novice genealogical researchers."[20] Like Gilliland-Swetland, they recommended that EAD finding aids be redesigned to meet the needs of genealogists.[21]

Ian Anderson found that academic historians in the United Kingdom "desire to see more online finding-aids with greater levels of detail."[22] He posited that "surrounded by ever more efficient means of retrieving information, historians, as well as other user groups, are not likely to remain tolerant of archival services that do not perform in a comparable manner."[23] A 2010 study of historians, with a focus on their query formulation and the search process, recommended that contents of digital repositories should have metadata for items that include information typically used in researchers' queries. Such metadata should be structured to allow "the researcher to gain more control over the search procedure and get more relevant results."[24]

Amanda Hill cited two surveys as well as the experience of the National Archives (U.K.) that identified a large proportion of online researchers as leisure users searching for family information.[25] Because leisure users seek information about specific individuals and places, detailed item-level description is critical to serving their needs.

User-Contributed Content

User-contributed content has been suggested as a means of describing digital collections or of enhancing such descriptions.[26] In their discussion of user-contributed annotation of finding aids, Michelle Light and Tom Hyry posited that such contributions would be beneficial in terms of (1) allowing for the expression of perspectives other than those of the creators of the finding aids, (2) increasing the amount of detail within the finding aid, and (3) promoting discovery through additional descriptive language.[27] However, they noted some of the drawbacks as well, not the least of which is the potential to "threaten archival professionalism."[28] Elizabeth Yakel addressed current and future use of Web 2.0 features for archival access.[29] In her review of several websites employing such features, she found that "engaging the researcher and eliciting their knowledge base can strengthen metadata about collections as well as the collections themselves."[30]

Max Evans introduced the commons-based peer-production model to address the difficulty that financially constrained repositories have in providing costly item-level description in an age of large-scale digitization of collections for online access.[31] In addition to recruiting and managing individual volunteers to create descriptive metadata, his model would use "the eyeballs and intellect of thousands of volunteers" in an electronic environment.[32] However, he cautioned that such a project should be managed carefully in terms of determining or protecting the rights to added-value information as it pertains to user contributions. Scott Anderson and Robert Allen also proposed a peer-based model that would expand user involvement in the description of online collections in an open, interactive archival commons.[33]

Projects have emerged that allow for user-contributed content for various purposes with varying levels of interactivity. In some instances, user-generated information is used in conjunction with automated tools to promote discovery.[34] Other projects incorporate user-contributed keywords to enhance discoverability of items within collections.[35] Michael Zarro and Robert Allen studied the annotations added to the Library of Congress images on the Flickr Commons site to understand how user contributions of information enhance the library's holdings.[36] The public is able to contribute tags, comments, and annotations, and to engage in virtual discussions. "The resulting set of user-contributed metadata is a valuable source of descriptive information that may be utilized for information retrieval, resource identification, and outreach."[37] In their view, personal remembrances are desirable because they "give the researcher access to 'hidden facts' about a resource."[38] They noted that "the public has shown they are willing and able to provide detailed and valuable annotations, corrections,

and translations for the Library [of Congress]."[39] However, they also cited "trust and authority of the work" as issues that need to be addressed.[40]

Lauren Graham reported on the experience with users of the Library of Congress American Memory project, an online collection of historical resources. "Naturally, we see American Memory as a flow of historical collections from the Library of Congress to American citizens everywhere. Unanticipated is the flow of content and information back to the Library of Congress from people who have local history, genealogy, or other specialized information to offer for correcting and enhancing descriptions of items in the institution's collections."[41] Users provided comments, corrections and bibliographical information via email communications. They appear to have a desire to contribute information even though the site lacks the specific functionality to do so.

The Polar Bear Expedition Digital Collections, however, do provide the functionality for user contribution of information to the finding aid by incorporating features such as bookmarking, commenting, and linking. Magia Krause and Elizabeth Yakel conducted a preliminary assessment of the site.[42] While they found evidence to suggest "that archivists can employ social interaction tools productively in finding aids to add to the depth and accuracy of descriptions," they noted that use of the social interaction features is limited.[43] In their conclusion, they pondered whether other tools, such as finding aid annotation, tagging, and explicit ranking might be more effective. In a study of the Polar Bear Expedition Digital Collections as well as two other online archival collections, Jessica Sedgwick concluded that "users most often contribute informational content, such as identification, further contextual information, and links to related resources."[44]

Research seems to indicate that allowing users to contribute content holds some promise for augmenting traditional description, thus increasing the discoverability of materials. Because this is not an endeavor to be entered into lightly in terms of required resources or policy considerations, it is important for decision makers to have as much information as possible about who will contribute content and what that content looks like. It is informative to look at the experience of an entity that allows user-contributed content to digitized cultural heritage materials on a large scale.

Methodology

Building on the research of Krause and Yakel,[45] Sedgwick,[46] and Zarro and Allen,[47] I examined the types of users who contributed content to a large-scale online collection of cultural heritage materials as well as what such users contributed. For this study, I define *online collections of cultural heritage materials* as analog materials of cultural or historic value that have been digitized and made

accessible via a Web-based interface by institutions including, but not limited to, archives and special libraries. Such a collection is termed "large scale" because a large number of items, often entire analog collections, are made available with minimal descriptive metadata. This is in contrast to smaller boutique online collections with extensive description and additional information. For the purpose of this study, I limited the large-scale online collections of cultural heritage materials to those available at Footnote.[48]

Launched in January 2007, Footnote is a subsidiary of iArchives,[49] a provider of digitization services. Footnote has entered into partnerships with entities, both public and private, that seek to digitize collections of analog materials and make them available to users online. Currently, Footnote makes more than seventy million original documents and images available through its website. In addition to putting the content online, Footnote provides the means for users to contribute information about the collections as well as to interact with each other through a website interface.

I selected this website for the case study because it meets several criteria. First, although Footnote itself is not a repository for cultural heritage materials, the content of the website compares to that of large-scale online collections of archives and special libraries. Indeed, the content originates from such respected repositories as the National Archives, the Library of Congress, the South Carolina Department of Archives and History, and others. Second, the content at Footnote has broad appeal for a variety of users, including historians and leisure users. The content consists of surrogates of original sources and includes military records, *Amistad* court records, Brady Civil War photos, and the Pentagon Papers. Third, the content predominately comprises textual collections. Other studies have focused on user-contributed content for collections that were entirely or predominately photographic in nature. Footnote has large collections of handwritten documents that do not readily lend themselves to accurate optical character recognition (OCR). The website has collections of typescript documents as well. Fourth, the site observes conventions similar to those of archival arrangement and description. The agreement with the National Archives and Records Administration requires the creation of technical and functional metadata that will "enable retrieval of the material at the fundamental level of archival control."[50] Footnote also provides a feature resembling a filmstrip which displays thumbnail images, allowing users not only to view an individual item, but also to view the item in the context of other items in the series (see Figure 1). As with a traditional archives, information is provided at the collection level. Users are often provided with a link to the original repository's finding aid for the collection to which the item belongs as well. Fifth, the site provides a variety of ways for users to add content. Users can contribute content to collections with features such as annotation and comment. They

FIGURE 1. All Footnote images can be viewed using the filmstrip feature. This example of the Footnote image viewer featuring a filmstrip can be found at http://www.footnote.com/image/##246|4032653.

are also able to augment and add context to collections by uploading content, creating and adding information to Footnote pages, and making connections between items within and among collections. They can highlight interesting items for other users by creating historical spotlights and communicate with others by contributing entries to the Footnote blog. Additionally, Footnote has a presence on the social networking sites Twitter and Facebook. Users are able to create their own collections from the Footnote collections as well as create watches for notification about additions to collections of interest. Sixth, the website provides for the contribution of content with minimal instruction or intermediation by Footnote staff. Analysis of the user-contributed content under

FIGURE 2. This Footnote user-contributed content composition analysis does not include Footnote pages, blog, member galleries and watches, Twitter, or Facebook.
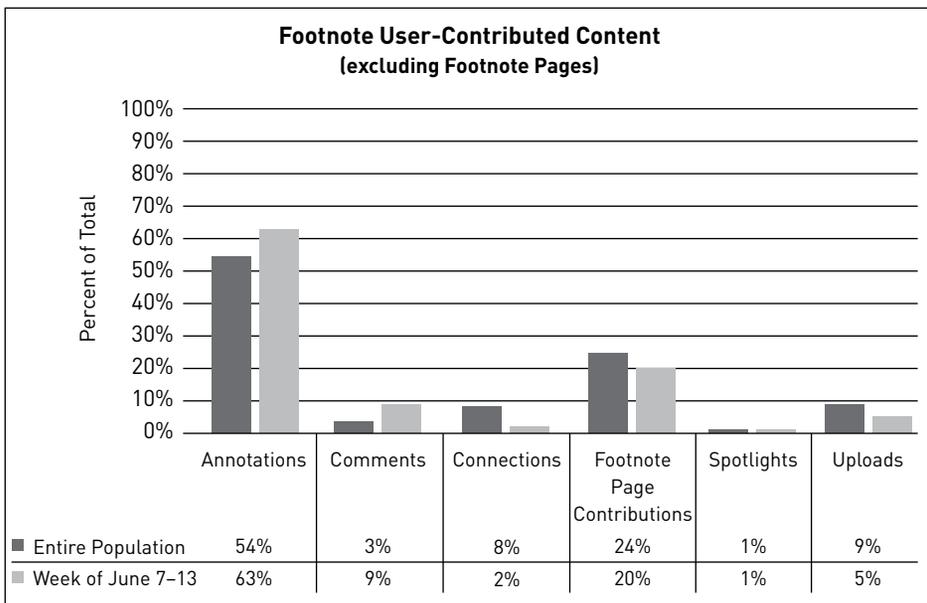
these conditions provides information about the content obtained when using a hands-off approach. Finally, and perhaps most importantly, users contribute content to the Footnote site at an astonishing rate. Figure 2 shows the number of user contributions for various categories. Thus, there is a substantial amount of user-contributed content available for analysis.

Content Analysis

For this study, I limited users to members of Footnote who have contributed content to the website. Information about each user is limited to the information available on the user's profile page. *User-contributed content* is defined as the voluntary addition of information about a collection or an item within the collection or the addition of new materials to a collection. For purposes of this study, I limited user-contributed content to information that is freely and publicly available on the Footnote site as a "member discovery." The unit of analysis is each contribution by a member of an annotation, comment, or connection; the creation or addition of information to a Footnote page; the designation of a spotlight; or the uploading of new content. In examining the information associated with users and the content they contributed, I used manifest content analysis.

I originally defined the sample for analysis as all user-contributed content and the related users for the one-week period of June 7–13, 2011. To obtain the sample, I created PDF (Portable Document Format) images of member discoveries

	Annotations	Comments	Connections	Footnote Page Contributions	Spotlights	Uploads
■ Entire Population	54%	3%	8%	24%	1%	9%
▨ Week of June 7–13	63%	9%	2%	20%	1%	5%

FIGURE 3. Footnote user-contributed content composition of the sample for the week of June 7–13, 2011, compared to the total contributions.

for the week for all categories except Footnote pages. This resulted in a sample size of 8,654 items. The composition of the sample appears to approximate the overall composition of the entire population of 1.7 million instances of user-contributed content, excluding Footnote pages, as shown in Figure 3.[51] However, content analysis of such a large sample was unfeasible. Therefore, I elected to analyze 15% of each category with a minimum of 50 items. This narrowing of the scope of the study resulted in an adjusted sample size of 1,398 items, excluding Footnote pages. The sample of 97 Footnote pages consisted of all member-created pages within the last 5,000 Footnote pages for the designated period.[52]

The addition of the sample of 97 Footnote pages to the samples for the other categories increased the total user-contributed content for analysis to 1,495 items as shown in Table 1. The sample of 183 users was composed of those who contributed the content that was included in the study.

Table 1. Sample for Analysis

Contribution Category	Number of Items
Annotations	847
Comments	58
Connections	122
Footnote Page Contributions	261
Spotlights	50
Uploads	60
Footnote Pages	97
Total	1,495*

*Contributed by 183 Users

Findings and Discussion

Users

I assigned each of the 183 users to one of four types: (1) individual with family connections, (2) organization, (3) researcher, or (4) other. A user is not assigned to more than one type. I determined user types from information on the user's Member Profile Page. I identified a user as *an individual with family connections* if family relationship descriptors appeared on his or her profile page. Such descriptors include, but are not limited to, mother, father, uncle, and grandfather. I identified a user as an *organization* if the profile page contained information about the user's association with an organization. The profile page of a *researcher* mentions nonfamily related research. *Other* users have profile

pages that do not contain information identifying them as one of the other three types.

More than two-thirds of users who contributed content were identified as individuals with family connections to the collections (see Table 2). I also identified two organizations and a single researcher. One-third of the users did not provide identifying information on their profile pages and are classified as "other." It is likely that some of these users are researchers or individuals with family connections, but in the absence of more information it is not possible to assign them to a user category. Users with family connections contributed twice as much content as those in the "other" category (see Table 3). The findings of other studies also reflect the high level of engagement of individuals with family connections.[53] This finding is also compatible with Evans's description of the tradition within the genealogical community of sharing information.[54] The engagement of these users may also reflect their information needs. Margaret Adams described this type of user as "fact seeking."[55] The contribution of content may support the contributing user's fact-seeking behavior as well as that of other users.

Table 2. Types of Users Who Contributed Content

User Type	Number of Users	Percentage of Sample
Individual with Family Connections	119	65%
Organization	2	1%
Researcher	1	>1%
Other	61	34%
Total	183	100%

Table 3. Number of Contributions by User Type

| User Type	Number of Contributions	Percentage of Sample
Individual with Family Connections	648	44%
Organization	526	35%
Researcher	1	>1%
Other	320	21%
Total	1,495	100%

The study revealed that while organizations constituted just 1% of users, they contributed 35% of the content. The analysis also shows that 71% of the

content was contributed by just 8% of the users. These figures are similar to those cited by Rose Holley in her description of "super volunteers" and library crowdsourcing.[56] In the *Waisda?* tagging project, Oomen et al. also noted "the exceptional effort put into the game by a small number of users" whom they labeled "super taggers."[57] They recommended finding a way to target these users specifically. This study of Footnote suggests that this strategy might also be valid for online cultural heritage collections.

Table 4. Number of Contributions by User Type[58]

User Type	Annotations	Comments	Connections	Footnote Page Contributions	Spotlights	Footnote Pages	Uploads
Individual with Family Connections	238	37	117	137	32	46	41
Organization	525	0	0	0	1	0	0
Researcher	1	0	0	0	0	0	0
Other	83	21	5	124	17	51	19
Total	847	58	122	261	50	97	60

As shown in Table 4, users appeared to demonstrate a preference for different contribution types. Organizations contributed annotations exclusively with the exception of a single spotlight. Individuals with family connections also contributed annotations more than any other type. However, they also engaged in higher levels of contribution in the form of connections and Footnote page contributions than the other groups. The contributions of individuals without family connections most often took the form of Footnote page contributions.

Annotations, Comments, and Spotlights
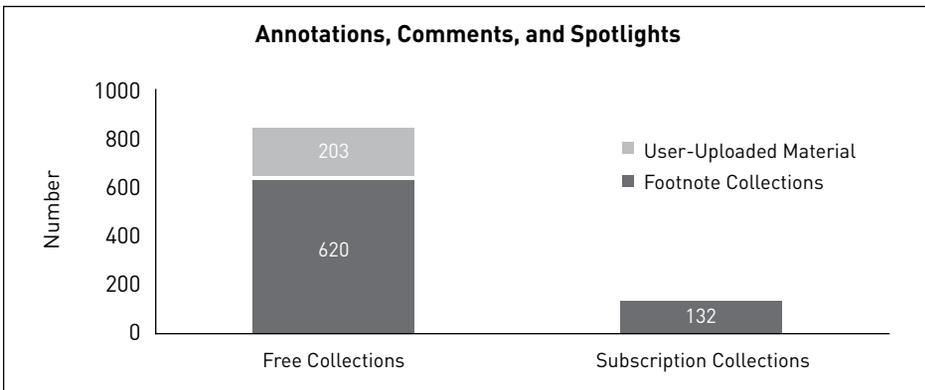
FIGURE 4. This graph compares the number of annotations, comments, and spotlights contributed to free collections and subscription collections.

Users also demonstrated a preference for contributing annotations, comments, and spotlights to items that were freely available over those that required a paid subscription (see Figure 4). Sixty-five percent of those user contributions were made to free collections, while an additional 21% were made to user-uploaded material. The remaining 14% were associated with collections that require subscription. This bodes well for repositories with open access to their digital collections.
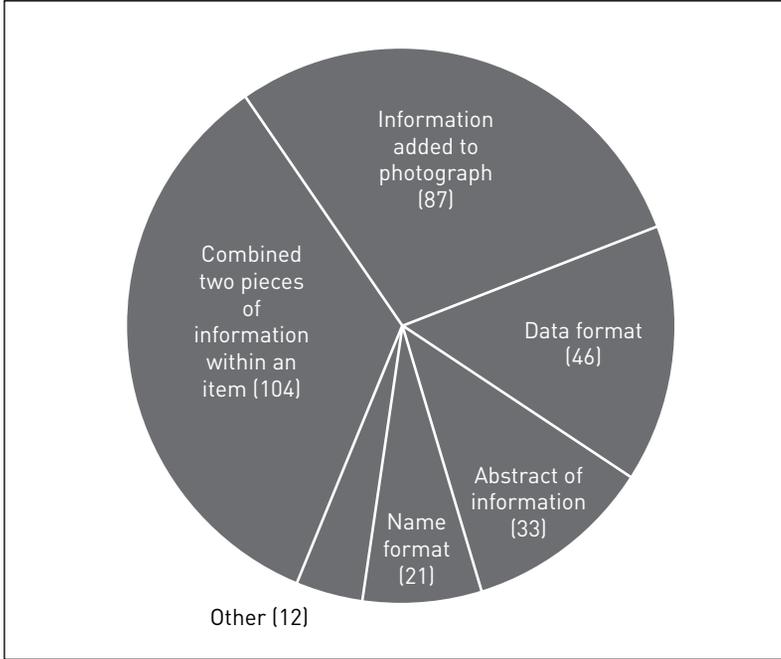
Annotations

The Footnote annotation feature allows users to add labels or transcriptions to an item. Organizational members were much more engaged in contributing this kind of content than were the other groups (see Table 5). They contributed 62% of all annotations. In fact, a single organization was responsible for 54% of the total annotations. It is notable that, with the exception of one instance of contributed content, all organizational contributions were made in the annotation category. Additionally, individuals with family connections contributed nearly three times as many annotations as did other individual users.

Table 5. Annotations Contributed[59] by User Type

User Type	Number of Annotations	Percentage of Sample
Individual with Family Connections	238	28%
Organization	525	62%
Researcher	1	>1%
Other	83	10%
Total	847	100%

The Footnote site provides minimal user instruction about the content and structure of annotations. When labeling or transcribing, users are instructed to type exactly what they see; they followed this recommendation 64% of the time. However, the remaining 36% of the annotations, as shown in Figure 5, point to some interesting issues.

In the largest group of annotations (104) not transcribed verbatim, the user combined two pieces of information contained in an item. All of these annotations are associated with records that mention slaves. The user combined the slaves' names with the designation "slave." Figure 6 is an example of this type of annotation. David Paterson discussed the indexing problems archivists face when trying to create access to recorded information about slaves.[60] He recommended incorporating the slave owner's name in the reference point. While

FIGURE 5. Composition of the 303 annotations that are not transcribed verbatim.

the Footnote user's method does not follow Paterson's recommendation, the user consistently applied the method he or she devised. (Because this user is identified as an organization, it is not possible to know whether one or more individuals are contributing.) Additionally, where a slave buyer or seller is mentioned, the user also annotated the name of that individual along with the role he or she played in the transaction. Thus, while perhaps not in accordance with accepted indexing principles, this form of annotation provides access to information typically not discoverable through traditional search methods.

The second largest group of annotations (87) not transcribed verbatim pertains to photographs, to which users added identifying information. While
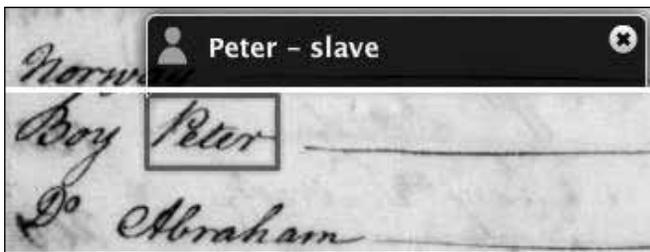
FIGURE 6. This is an example of a Footnote annotation of a slave record from South Carolina Estate Inventories and Bills of Sale, 1732–1872, http://www.footnote.com/image/#266967750.

this does not appear to be the purpose of the annotation feature as defined by Footnote, users have adopted this practice to associate identifying information directly with photographs. In 49 instances, individuals with family connections contributed the annotations and provided information about photographs uploaded by members. Another 33 annotations contained information abstracted from an item adjacent to the photograph within the same collection.

In another group of annotations, users abstracted information from documents rather than transcribing it verbatim. While these 33 annotations do not follow the recommendation, they can enhance discoverability of the item.

In 46 instances, the date format in the annotation did not reflect the date as it appeared in the item. In addition, in 21 instances the name order varied. Clearly, the formats of names and dates will vary between items within and among collections. The findings of the study reflect that users will introduce an additional level of variability in the structure of the information. Perhaps they are trying to guess at proper format, since it is not clearly defined. A controlled data-entry format would perhaps help users and add consistency in names and dates.

Additionally, I considered the format of items that were the subjects of annotations. Because other studies referred to the difficulty of using OCR technology for handwritten documents,[61] I was especially interested in whether users annotated handwritten items in the collection. In terms of format, 74% of the items annotated by users were handwritten documents (see Table 6). This indicates that users are, indeed, willing to annotate the items for which OCR is problematic. Sedgwick also raised the issue of format when she posed the question of whether "digitized images invoke more response than written documents."[62] My analysis seems to suggest, at least for these collections, that this is not the case.

Table 6. Annotation Item Formats

Item Format	Number of Items	Percentage of Sample
Handwritten	625	74%
Typescript	130	15%
Photo	88	10%
Other	4	> 1%
Total	847	100%

It appears that user groups have preferred formats (see Table 7). Individuals with family connections engaged in annotating handwritten items 67% of the time, while other individuals annotated typescript items 84% of the time.

Additionally, although it may appear that the organizational users preferred to annotate handwritten documents, annotations can be further broken down by organization. One organization was responsible for 455 of the 456 annotations for handwritten items. Another organizational member annotated all of the typescript items and photographs attributed to organizations. This appearance of format preference may actually indicate a preference for the content contained within the format rather than for the format itself.

Table 7. Types of Formats That User Types Annotate

Number of Items					
User Type	Handwritten	Typescript	Photograph	Other	Total
Individual with Family Connections	159	26	53	0	238
Organization	456	34	35	0	525
Researcher	1	0	0	0	1
Other	9	70	0	4	83
Total	625	130	88	4	847

In making an annotation, users are required to designate whether it refers to people, place, date, or other. These designations are mutually exclusive. Annotations designated as "other" are textual in nature, with the text derived from within the document and, thus, differ from the comments form of contributed content in which the information comes from the user. Ninety-nine percent of the time, users correctly associated the annotation designation of people, place, date, or other with the type of content they annotated. In an example of an incorrect association, a user annotated a date but selected the place annotation type. Users also appear to have a preference with regard to the type of annotation they contribute (see Table 8). All user types, except the

Table 8. Types of Annotations by User Type

Number of Items					
User Type	People	Place	Date	Other	Total
Individual with Family Connections	216	7	3	12	238
Organization	350	58	46	71	525
Researcher	0	1	0	0	1
Other	63	1	1	18	83
Total	629	67	50	101	847

researcher, contributed more people annotations than any other designation. A single organization contributed more than half of the people annotations. Individuals with family connections contributed more than three times as many people annotations as did other individuals.

My analysis of annotations suggests that while users will follow suggestions for input, they will also adapt the feature to suit their information needs. Additionally, organizational users overwhelmingly contributed this type of content. It appears that this type of user prefers to annotate handwritten items. However, this likely reflects an interest in the content rather than a format preference. The high level of engagement with the creation of annotations related to people may indicate a need for item-level information about individuals.

Comments

The Footnote comment feature provides a means for users to augment the descriptive information of items as well as to note corrections to it. Because this feature allows for the typing of free text without formatting restrictions, users can contribute information of any kind. Comments can be judged to belong to more than one category, but multiple expressions within a comment of information in the same category are counted as a single instance.

My analysis of the comments shows that users primarily provide additional or corrected information (see Table 9). This finding mirrors that of the Krause and Yakel study[63] as well as the Sedgwick study.[64] Additionally, the contributed information is most often associated with individuals rather than with places or other things. Among comments providing additional information, 21 out of 25 were associated with individuals. As for corrections, 8 out of 9 related to individuals. To a lesser extent, users described relationships either between themselves and others or relationships between other individuals. Almost half of the comments were contributed to Footnote pages rather than to items in the

Table 9. Categories of Comments Contributed by Users[65]

Comment Category	Related to Individuals	Percentage of Sample
Additional Information	25	41%
Correcting Information	9	16%
Link	1	2%
Relationship—Self	5	9%
Relationship—Other	7	12%
Other	11	19%
Total	58	100%

collections (see Table 10). Thus, the additional information does not directly augment the descriptions of items within a collection, but rather relates to items peripheral to it.

Table 10. Format of Items on which Users Commented

Item Format	Number of Comments	Percentage of Sample
Handwritten	16	28%
Typescript	9	15%
Photograph	7	12%
Footnote Page	26	45%
Total	58	100%

Individuals with family connections contributed more comments than any other group (see Table 11). Organizations did not contribute any comments. This contrasts sharply with the substantial number of annotations they contributed, suggesting that organizational members are more interested in increasing discoverability of items rather than augmenting information or engaging with the collections.

Table 11. Comments by User Type

User Type	Number of Comments	Percentage of Sample
Individual with Family Connections	37	64%
Organization	0	0%
Researcher	0	0%
Other	21	36%
Total	58	100%

Connections

The Footnote connection feature is designed to allow users to create links and to describe relationships between themselves and, thus, engage with the collections by creating a personal connection. Users can also use this feature to create links between items, such as photographs and documents, thereby building additional context within or between collections. Additionally, users can describe connections involving items peripheral to the collections such as user-uploaded content and Footnote pages. I limited the information analyzed for the connections feature to that which could be readily obtained by viewing entries on the member discoveries page.

Users largely engaged in demonstrating a connection between themselves and other individuals (see Table 12). Only 4% of the connections created by users demonstrated a connection between two documents. In 90% of the connections, the users themselves are part of the relationship. Individuals with family connections created 109 out of 110 connections of this type. In fact, individuals with family connections engaged in creating connections to a much higher degree (96%) than did any other group (see Table 13). Organizations and the researcher did not create connections.

Table 12. Types of Connections Made by Users[66]

User Type	Number of Connections	Percentage of Sample
Individual with Family Connections	110	90%
Organization	7	6%
Researcher	5	4%
Other	122	100%
Total	58	100%

Table 13. Number of Connections Contributed by User Type

User Type	Number of Annotations	Percentage of Sample
Individual with Family Connections	117	96%
Organization	0	0%
Researcher	0	0%
Other	5	4%
Total	122	100%

It appears that the connection feature functions more as a tool for building community around collections through user engagement than as one that builds context within and among collections.

Spotlights

The Footnote spotlight feature allows users to highlight interesting items they discover in the collections so other users may view them. Other users may comment on spotlights as well. My content analysis focused on the textual descriptions that users submitted when creating spotlights.

Users created spotlights most often to highlight people (see Table 14). Of the 50 spotlights examined, 78% were associated with people. Additionally, as

with comments and connections, individuals with family connections demonstrated the highest level of engagement. This group contributed 64% of the spotlights, nearly double the number contributed by other individuals (see Table 15). No other users commented on the spotlights, but this may be a function of the recent nature of the submissions rather than a lack of interest.

Table 14. Types of Spotlights Contributed by Users[67]

Spotlight Type	Number of Spotlights	Percentage of Sample
Person	39	78%
Place	2	4%
Event	1	2%
Other	7	14%
Undetermined	1	2%
Total	50	100%

Table 15. Number of Spotlights Contributed by User Type

User Type	Number of Spotlights	Percentage of Sample
Individual with Family Connections	32	64%
Organization	1	2%
Researcher	0	0%
Other	17	34%
Total	50	100%

Uploads

The Footnote uploading feature allows users to contribute and share documents and photographs. Users can contribute files (JPEG, GIF, PNG, TIFF, and JPEG-2000) without any mediation from Footnote staff.

Table 16 shows that users predominantly contributed photographs. It is especially important to assign descriptive metadata to photographs because such information is not usually contained within the image. Of the 36 photographs depicting people, only 4 were identified with both the first name and last name of the subject, as well as the date or time period and location of the photograph. Table 17 provides additional data about the metadata for photographs of people. Users uploaded 7 of the photographs of people and all 6 of the photographs of places without descriptive metadata. Files containing images of documents were uploaded to a lesser degree. The descriptive metadata assigned

by users included a title for 80% of the documents, but in no instance did users provide source citation information.

Table 16. Content Types of User-Uploaded Files[68]

| Uploaded Content Type	Number of Uploads	Percentage of Sample
Photograph	42	70%
Document	15	25%
Other	3	5%
Total	60	100%

Table 17. Metadata for 36 User-Uploaded Photographs of People

Metadata	Number of Photographs	Percentage of Sample
Both First and Last Name	22	61%
Date/Time Period	6	17%
First Name (no last name)	4	11%
Location	4	11%
Additional Information	9	25%

Ninety percent of the materials were uploaded by individuals with family connections. This group was by far the most engaged in this activity. Other individuals contributed 10% of the uploaded materials, while organizational members and the researcher uploaded no materials. It is ironic that while item-level description is most useful for individuals with family connections and, in fact, they work toward that end with the other types of contributions (especially annotations), they often are not supplying consistently complete descriptions when they upload their materials. Perhaps this is because the assignment of descriptive metadata is separate from the uploading process and requires an additional step. Additionally, users may have adapted the annotation feature to serve this purpose.

Footnote Pages

The Footnote page feature allows users to create pages within the site on which to collect and share information about people, topics, events, places, or organizations. Figure 7 is a sample page for a person. After the initial creation of a page, additional information can be added in the form of Footnote page contributions (see the following section). The user who creates the Footnote page has the option of allowing others to contribute to the page.
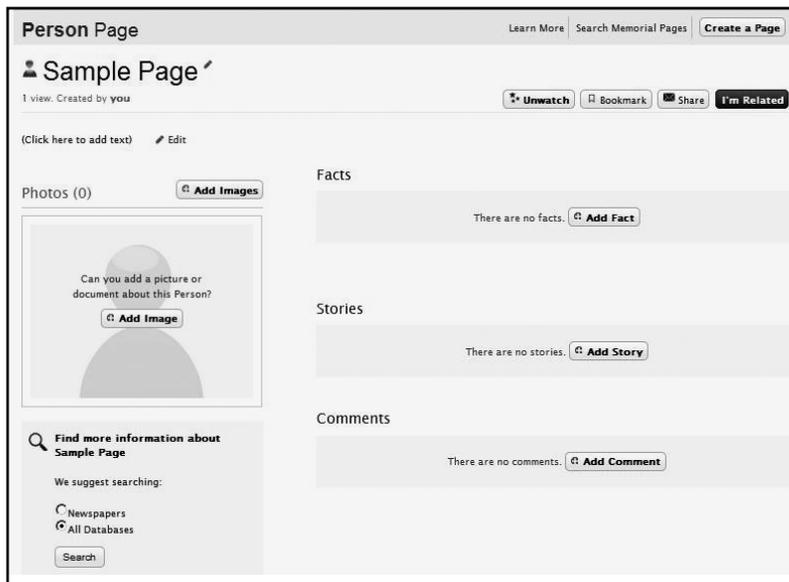
FIGURE 7. Example of a Footnote page.

Footnote page creation is the only contribution activity in which family-connected individuals engaged less than other individuals. Of the 97 Footnote pages analyzed,[69] other individuals created 51 pages (53%), while those with family connections created 46 pages (47%). Organizational members and the researcher did not engage in Footnote page creation.

Person pages comprised 95% of the pages created by users. The remaining pages were topic (4%) and event pages (1%). No place or organization pages were created. In reviewing the Footnote pages, I determined that 3 of the person pages were not related to a person, but to a commercial product. These pages contained links to sites for the commercial product.

Individuals with family connections were much more willing to allow contributions by others to Footnote pages that they created. Out of 45 pages created by this group, others were allowed to contribute in all but one instance. In contrast, individuals without family connections allowed others to contribute to pages they created 56% of the time. This seems to indicate that those with family connections use their pages as collaborative tools, while those without family connections view them as one-way sharing of information.

Footnote Page Contributions

Users can add facts, images, stories, or links to Footnote pages they have created as well as to those created by other members provided the creator of the page allows it. Footnote page contributions were evenly divided between

members with family connections (52%) and those without such connections (48%). These groups were largely engaged in contributing facts (92%) with other types of content contributed to a much lesser degree (see Table 18). As was the case with creating Footnote pages, organizations and the lone researcher did not engage in contributing information to them.

Table 18. Footnote Page Contributions by Type of Contribution[70]

Type	Footnote Page Contributions	Percentage of Sample
Fact	240	92%
Image	12	5%
Link	4	2%
Story	5	2%
Total	261	100%

In reviewing the fact section of the Footnote pages for source citation information, I did not evaluate the format or adequacy of the information. I only noted the presence or absence of the information to see, in effect, if the attempt to add it was made (see Table 19). It is notable that, for 95% of the information contributed to the fact section of a Footnote page, source citation information was absent. Noncited information was submitted in nearly equal amounts by both individuals with family connections (52%) and other individuals (48%). However, individuals with family connections contributed all 12 instances for which citation information was present.

Table 19. Citation of Footnote Page User-Contributed Facts

User Type	Facts with Citation	Facts without Citation	Total
Individual with Family Connections	12	115	127
Organization*	—	—	—
Researcher	—	—	—
Other	0	113	113
Total	12	228	240

*Did not contribute facts to Footnote pages

Conclusion

While the results of this case study of Footnote user-contributed content may not be generalizable to other online collections, the information

is instructive in terms of who might contribute and the content they might contribute. The study shed some light on users and their primary interests. However, it also brought to light issues of consistency, authenticity, and context with regard to user-contributed content.

The study revealed a predominate interest in information about individuals. Users contributed more annotations related to people than any other type. Additionally, comments and spotlights were also most often associated with people. It is not surprising that individuals with family connections focused on people named in the collections. This contrasts with traditional finding aids that do not typically identify an individual at the item level. Yet, this information appears to be the focus of users. For repositories with collections that include "hidden" individuals, providing annotation functionality to users would likely enhance the discoverability of those people.

The study revealed the existence of a few users who contributed a disproportionate number of annotations. The most prolific of these super annotators were organizational users. These users contributed annotations only, with the exception of a single spotlight, and did not avail themselves of any other method of contribution. This suggests that it may be useful for repositories to cultivate relationships with organizations that share an interest in the materials in their collections. Additionally, these organizational users did not participate in the social aspects of the site. This suggests that, if the goal is increased discoverability of items, a repository that partners with an organization may not be required to invest in robust social Web 2.0 functionality. However, if the goal of the repository is to increase use of and engagement with its collections, the study suggests that people with family connections to collections are the most engaged. Repositories might consider partnering with genealogical organizations as a strategy for targeting participation of this user group.

The study showed that the consistency of the information contributed by users can be problematic. They did not always "type what they saw" or use the features as they may have been originally conceived. This has implications for the design of the search functions and may indicate a need for controlled data-entry formats. This may also indicate the need for user instruction, if feasible. However, users may also diverge from suggested practices in ways that add meaning and increase the discoverability of items in collections. Thus, to improve features to better serve users' needs, it is important to monitor how users use or adapt them.

The study demonstrated that users often do not contribute content in a way that allows other users to identify or assess it. They rarely provided complete descriptive metadata when uploading content. They also rarely provided source citation information when contributing facts. This speaks to the trustworthiness of the information. In the world of repositories, "trust remains an

important asset."[71] In Horava's words, "trust saves the user's time, keeps the user's attention, and provides an implicit stamp of quality."[72] In fact, Yakel posited whether archivists' slow adoption of Web 2.0 interactive features stems in part from "a desire to maintain authoritative metadata about collections."[73] These reservations are not lessened by what Hazen described as "'Authority 2.0' as users, singly or in cohorts, participate in an electronic free-for-all."[74] That is not to say that the user-contributed content amounts to a free-for-all, just that the information may not meet professional standards. User instruction might mitigate the situation to some degree. Additionally, Anderson and Allen suggested, "Transparency and attribution related to the narrative activity associated with the materials will be critical for preserving the authenticity of the materials themselves versus subsequent additions to them."[75] Thus, this study suggests that clearly delineating original content and its professional description from user-contributed content and description is appropriate and desirable.

Finally, user-contributed content has been proposed as a means of adding context to collections. The Footnote features that support this addition of context are comments and connections. While the comments provided additional information, nearly half of the comments were directed at Footnote pages, which themselves are not part of the collections, but rather, are peripheral. In the case of connections, only 4% were between items in collections. Additionally, when users uploaded content, such as photographs, that could potentially augment the collections, they did not supply complete metadata. Thus, the building of context was not as extensive as it might have been.

Despite these issues, allowing users to contribute content in an online environment shows promise as a means of narrowing the gap between what repositories are able to provide and what users want. The purpose of this study was to look at the experience of an existing online collection in terms of users and user-contributed content. To paraphrase Forrest Gump, user-contributed content "is like a box of chocolates. You never know what you're gonna get." This case study of Footnote provides some insight into what a repository might get.

NOTES

1 Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *The American Archivist* 68 (Fall/Winter 2005): 208–63.

2 Greene and Meissner, "More Product, Less Process," 209.

3 Greene and Meissner, "More Product, Less Process," 240.

4 Ashley Bucciferro, "Attacking the Backlog: NARA Archivists Mobilize to Make Unprocessed Records Available to the Public," *Prologue Magazine* 40 (Summer 2008), http://www.archives.gov/publications/prologue/2008/summer/backlog.html.

5 Christine Weideman, "Accessioning as Processing," *The American Archivist* 69 (Fall/Winter 2006): 274–83.

6 Weideman, "Accessioning as Processing," 276.

7 Stephanie H. Crowe and Karen Spilman, "MPLP @ 5: More Access, Less Backlog?," *Journal of Archival Organization* 8, no. 2 (2010): 110–33.

8 Robert S. Cox, "Maximal Processing, or, Archivist on a Pale Horse," *Journal of Archival Organization* 8, no. 2 (2010): 143.

9 Ricky Erway, "Supply and Demand: Special Collections and Digitisation," *Liber Quarterly* 18 (December 2008): 324–36.

10 Erway, "Supply and Demand," 328–29.

11 Mark A. Greene, "MPLP: It's Not Just for Processing Anymore," *The American Archivist* 73 (Spring/Summer 2010): 189.

12 Greene, "MPLP," 194.

13 Richard Pearce-Moses, "Janus in Cyberspace: Archives on the Threshold of the Digital Era," *The American Archivist* 70 (Spring/Summer 2007): 18.

14 Barbara L. Craig, "Old Myths in New Clothes: Expectations of Archives Users," *Archivaria* 45 (Spring 1998): 120.

15 Wendy M. Duff and Joan M. Cherry, "Use of Historical Documents in a Digital World: Comparisons with Original Materials and Microfiche," *Information Research* 6 (October 2000).

16 Lauren Graham, "Tales from the Vault: A Journey over the Mountains," *Common-Place* 3 (January 2003), http://common-place.org/vol-03/no-02/tales.

17 Meg Sweet and David Thomas, "Archives Described at Collection Level," *D-Lib Magazine* 6 (September 2000), http://www.dlib.org/dlib/september00/sweet/09sweet.html.

18 Sweet and Thomas, "Archives Described at Collection Level."

19 Anne J. Gilliland-Swetland, "Popularizing the Finding Aid," *Journal of Internet Cataloging* 4, nos. 3–4 (2001): 199–25.

20 Wendy M. Duff and Catherine A. Johnson, "Where Is the List with All the Names? Information-Seeking Behavior of Genealogists," *The American Archivist* 66 (Spring/Summer 2003): 91.

21 Duff and Johnson, "Where Is the List," 79–95.

22 Ian G. Anderson, "Are You Being Served? Historians and the Search for Primary Sources," *Archivaria* 58 (Fall 2004): 114.

23 Anderson, "Are You Being Served?," 83.

24 Elena Torou et al., "Historical Research in Archives: User Methodology and Supporting Tools," *International Journal on Digital Libraries* 11, no. 1 (2010): 25–36.

25 Amanda Hill, "Serving the Invisible Researcher: Meeting the Needs of Online Users," *Journal of the Society of Archivists* 25, no. 2 (2004): 139–48.

26 See, for example, Anderson, "Are You Being Served?"; Ricky Erway and Jennifer Schaffner, "Shifting Gears: Gearing Up to Get into the Flow," Online Computer Library Center Programs and Research (2007), http://www.oclc.org/research/publications/library/2007/2007-02.pdf; Alice Prochaska, "Digital Special Collections: The Big Picture," RBM: *A Journal of Rare Books, Manuscripts, and Cultural Heritage* 10 (Spring 2009): 13–24; Greene, "MPLP"; Elena et al., "Historical Research in Archives."

27 Michelle Light and Tom Hyry, "Colophons and Annotations: New Directions for the Finding Aid," *The American Archivist* 65 (Fall/Winter 2002): 261–30.

28 Light and Hyry, "Colophons and Annotations."

29 Elizabeth Yakel, "Inviting the User into the Virtual Archives," *OCLC Systems and Services* 22, no. 3 (2006): 159–63.

30 Yakel, "Inviting the User," 160.

31 Max J. Evans, "Archives of the People, by the People, for the People," *The American Archivist* 70 (Fall/Winter 2007): 387–400.

32 Evans, "Archives of the People," 397.

33 Scott R. Anderson and Robert B. Allen, "Envisioning the Archival Commons," *The American Archivist* 72 (Fall/Winter 2009): 383–400.

34 See, for example, Kees van der Sluijs and Geert-Jan Houben, "Exploiting User Tags to Build a Semantic Cultural Heritage Portal," *IEEE Intelligent Systems* 25 (November/December 2010): 84–92; Bertrand Couasnon, Jean Camillerapp, and Ivan Leplumey, "Making Handwritten Archives Documents Accessible to Public with a Generic System of Document Image Analysis," *Proceedings of the First International Workshop on Document Image Analysis for Libraries* (2004), 270–77; Reim Doumat, Elod Egyed-Zsigmond, and Jean-Marie Pinon, "A Web 2.0 Archive to Access, Annotate and Retrieve Manuscripts," *Multimedia Tools and Applications* (August 9, 2011).

35 See, for example, Sarah Clayton et al., *User Tagging of Online Cultural Heritage Items* (2008), E-lis, http://hdl.handle.net/10760/12775; Johan Oomen et al., "Emerging Practices in the Cultural Heritage Domain—Social Tagging of Audiovisual Heritage," *Proceedings of the WebSci 10: Extending the Frontiers of Society On-Line* (2010), http://journal.webscience.org/337/2/websci10_submission_23.pdf; Tiffany Leason, "Steve: The Art Museum Social Tagging Project: A Report on the Tag Contributor Experience" (paper presented at "Museums and the Web 2009: The International Conference for Culture and Heritage On-line," Indianapolis, Ind., April 2009), Archives and Museums Informatics, http://www.archimuse.com/mw2009/papers/leason/leason.html.

36 Michael A. Zarro and Robert B. Allen, "User-contributed Descriptive Metadata for Libraries and Cultural institutions," in *Lecture Notes in Computer Science: Research and Advanced Technology for Digital Libraries*, ed. Mounia Lalmas et al. (Berlin: Springer, 2010), 46–54.

37 Zarro and Allen, "User-contributed Descriptive Metadata," 46.

38 Zarro and Allen, "User-contributed Descriptive Metadata," 51.

39 Zarro and Allen, "User-contributed Descriptive Metadata," 53.

40 Zarro and Allen, "User-contributed Descriptive Metadata," 53.

41 Graham, "Tales from the Vault."

42 Magia G. Krause and Elizabeth Yakel, "Interaction in Virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid," *The American Archivist* 70 (Fall/Winter 2007): 282–314.

43 Krause and Yakel, "Interaction in Virtual Archives," 312.

44 Jessica M. Sedgwick, "Let Me Tell You about My Grandpa: A Content Analysis of User Annotations to Online Archival Collections" (master's thesis, University of North Carolina, Chapel Hill, 2008), 37.

45 Krause and Yakel, "Interaction in Virtual Archives."

46 Sedgwick, "Let Me Tell You."

47 Zarro and Allen, "User-contributed Descriptive Metadata."

48 On August 18, 2011, the name of the site was changed from Footnote to Fold3 to reflect a new focus on military records. However, the features that afford the contribution of content by users remain the same. The URL for the Footnote website was http://www.footnote.com at the time of this study.

49 On October 21, 2010, Ancestry.com acquired iArchives.

50 "NARA-iArchives Digitization Agreement," U.S. National Archives and Records Administration, http://www.archives.gov/iarchives/iarchives-digitization-agreement.html.

51 Statistical analysis of the data was not performed.

52 Footnote staff members are also actively engaged in the creation of Footnote pages. They created 4,903 out of the 5,000 pages.

53 See, for example, Krause and Yakel, "Interaction in Virtual Archives"; Sedgwick, "Let Me Tell You."

54 Evans, "Archives for the People."

55 Margaret O. Adams, "Analyzing Archives and Finding Facts: Use and Users of Digital Data Records," *Archival Science* 7 (March 2007): 21–36.

56 Rose Holley, "Crowdsourcing: How and Why Should Libraries Do It?," *D-Lib Magazine* 16 (March/April 2010), http://www.dlib.org/dlib/march10/holley/03holley.html.

57 Oomen et al., "Emerging Practices."

58 I did not perform statistical analysis of the data.

59 I elected to analyze a stratified sample of 15% of the annotations that were created during the

week of June 7–13, 2011. I selected a stratified sample of 847 annotations for the purpose of ana-
lyzing information associated with each type in proportion to the amount contributed: 629 People
(74%), 67 Place (8%), 50 Date (minimum), 101 Other (12%). I analyzed the most recent annotations
for each type within the specified time period.

[60] David E. Paterson, "A Perspective on Indexing Slaves' Names," *The American Archivist* 64 (Spring/
Summer 2001): 132–42.

[61] See, for example, Couasnon, Camillerapp, and Leplumey, "Making Handwritten Archives
Documents Accessible"; Doumat, Egyed-Zsigmond, and Pinon, "A Web 2.0 Archive."

[62] Sedgwick, "Let Me Tell You," 36.

[63] Krause and Yakel, "Interaction in Virtual Archives."

[64] Sedgwick, "Let Me Tell You."

[65] The sample for analysis consisted of the 50 most recent comments identified for the week of
June 7–13, 2011. Because several of the comments included information for more than one cat-
egory, 58 were categorized.

[66] The sample size equaled 15% of the connections that that were identified for the week of
June 7–13, 2011. The 122 selected user-contributed connections were the most recent within that
time period.

[67] Within the designated period of June 7–13, 2011, I identified 98 spotlights. Rather than analyze a
sample of 15%, which totaled 15 items, I elected to analyze the 50 most recent spotlights identified
for the time period.

[68] The sample size was equal to 15% of the files uploaded by users that were identified for the week
of June 7–13, 2011. The 60 selected user-contributed uploaded files were the most recent ones
within that time period.

[69] I described the selection of the sample of 97 Footnote pages for analysis in the methodology
section.

[70] The sample size was equal to 15% of the number of Footnote page contributions identified for the
week of June 7–13, 2011. The 261 selected contributions were the most recent within that time
period.

[71] Horava, "Challenges and Possibilities," 145.

[72] Horava, "Challenges and Possibilities," 145.

[73] Yakel, "Inviting the User," 160.

[74] Dan Hazen, "Rethinking Research Library Collections: A Policy Framework for Straitened Times,
and Beyond," *Library Resources and Technical Services* 54, no. 2 (2010): 120.

[75] Anderson and Allen, "Envisioning the Archival Commons," 395.

ABOUT THE AUTHOR

Pamela Mayer earned her MSLS degree from the University of North
Carolina at Chapel Hill. She is interested in the access and use of informa-
tion that documents individuals, events, and culture. She is also interested
in user-contributed content and digital preservation. She currently works as
a visiting archivist at Carolina Friends School in Durham, North Carolina,
where she is guiding the school in establishing an archives.