Pamela H. Mayer. Like a Box of Chocolates: A Case Study of User-Contributed Content at Footnote. A Master's Paper for the M.S. in L.S. degree. November, 2011. 78 pages.
Advisor: Christopher A. Lee

User-contributed content has been suggested as a means of narrowing the gap between the level of description that resource-constrained repositories are able to provide and the level that users need or have come to expect. An examination of the experience of entities with large-scale online collections of cultural heritage materials that allow users to contribute content can provide valuable insight to other institutions that are considering doing so. This case study examines 183 users and 1,495 instances of user-contributed content at Footnote. The study identifies individuals with family connections as the largest group of contributors while annotations are the most common type of contribution. The data suggests that users' are predominately interested in information about individuals. Additionally, the study reveals the existence of users who contribute a disproportionate number of annotations. The findings of the study also indicate issues of consistency, authenticity, and context with regard to user-contributed content.

Headings:

Annotations – Evaluation

Archives – Internet Resources

Footnote (Web site)

Use Studies – Information Systems

Web 2.0 – Case Studies

LIKE A BOX OF CHOCOLATES: A CASE STUDY OF USER-CONTRIBUTED
CONTENT AT FOOTNOTE

by
Pamela H. Mayer

A Master's paper submitted to the faculty
of the School of Information and Library Science
of the University of North Carolina at Chapel Hill
in partial fulfillment of the requirements
for the degree of Master of Science in
Library Science.

Chapel Hill, North Carolina

November, 2011

Approved by:

Christopher A. Lee

Table of Contents

Introduction.....	2
Literature Review.....	4
Methodology	20
Users	29
Annotations.....	33
Comments	41
Connections.....	44
Spotlights	46
Uploads	47
Footnote Pages.....	50
Footnote Page Contributions.....	52
Limitations	54
Conclusion	55
Works Cited	59
Appendix A.....	65
Appendix B.....	66
Appendix C	68
Appendix D.....	70
Appendix E	71
Appendix F.....	73
Appendix G.....	75
Appendix H.....	76

Introduction

Archives are more than warehouses. Richard Pearce-Moses stated in his Presidential Address at the 2006 annual meeting of the Society of American Archivists, “I believe that society entrusts archivists with preserving the cultural record and our documentary heritage” (2007, p. 16). Elsie Freeman Finch extends the *raison d’être* by saying, “Use is our reason for being” (1994, p. 1). Before the advent of the Internet and online collections, users were required to physically visit the repository to use cultural heritage materials. The materials were located using a finding aid, a guide to the contents of a collection. Rarely do these finding aids describe individual items in the collection; they usually provide information at a higher level. With digitization of archival material, 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. There is no way for users to easily get from the finding aid directly to the item they wish to view. Some fishing around is required. This is at odds with the needs and expectations of users who have become accustomed to discovering information with ease through search engines, such as Google. Thus, while online digital collections provide ease of access in that users are not required to visit the physical location of the collection, ease of access in terms of being able to access a specific item or piece of information contained in an online collection is still problematic. Additionally, online finding aids in the traditional format do not allow online collections to take advantage of

the additional affordances of being digital. Unlike paper materials, document and photos in digital form can be linked to provide additional context and can be arranged in completely new ways while also maintaining the original order.

To provide the additional item-level description, linkages, and arrangement would require additional resources from already financially overstretched archival institutions. There has historically been a tension between the amount of material that needs to be processed and the level of resources available for processing. In many instances, the materials that are selected for online collections have already been processed once in the traditional manner. While these are usually the most significant or popular collections, should additional resources be spent on them in the digital realm? Significant resources are required to create a boutique collection that has additional contextual information and additional descriptive metadata to provide additional access points for users. Those same resources could be used to post a large-scale collection online with minimal additional information and descriptive metadata.

One way to mitigate the lack of resources available for additional descriptive information of online content is to shift the burden of description to the user. Once information contained in an item is discovered, it could be marked or annotated by the user so that a future user will find the information more easily. Now imagine that a user has additional information about an item. Allowing him to record this information with the item would have the potential to enrich the archival experience for other users. Providing this type of opportunity for users to contribute information to an online collection can not only increase the level of information available about items in the

collection, but also has the potential to create a higher level of commitment and support for the institution from the users.

Yet archivists may have reservations about opening up the collections to this type of activity. If resources are spent to create a means for users to contribute information, what will the users contribute? One way to explore this question is to look at existing online collections with user contribution functionality. Who is contributing and what are they contributing?

Literature Review

Backlogs and Item-level Description

As professionals charged with collecting and preserving materials of enduring value, archivists have been faced with a seemingly endless quantity of materials. Gilliland-Swetland (1991) notes the profound effect of this abundance on the evolution of the fundamental principles and practices of the archival profession: principles such as provenance and practices such as series-level description.

Today this deluge of material 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 materials that have been accessioned but not yet arranged or described, also known as a backlog. 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” (2005, p. 209). 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 2005, Greene and Meissner wrote their widely cited article in which they proposed a strategy to expedite the processing of collections known as More Product, Less Process (MPLP). They present the problem of backlogs and examine the metrics associated with processing. They also discuss their own survey of archivists across the United States. They found that for sixty percent of respondents' repositories at least one third of the collections were unprocessed; for another thirty-four percent at least one half of the collections were unprocessed; and forty-six percent did not permit researcher access to unprocessed collections. They conclude that all collections should receive a basic level of processing with increased attention given to collections only when warranted. Additionally, they recommend that "arrangement, preservation, and description work should all occur in harmony, at a common level of detail" (Greene & Meissner, 2005, p. 240).

Also in 2005, the National Archives and Records Administration (NARA) began the development of a processing initiative to address a backlog of one million cubic feet (Bucciferro, 2008). The new processing approach incorporated the specialization of labor with the creation of teams within a designated processing staff. It also included adjustments to the processing activities. In MPLP fashion, basic processing was considered to be adequate for most collections while those in higher demand would receive more attention.

Weideman (2006) notes the issue of backlogs at Yale University Manuscripts and Archives. She attributes the backlog not only to a greater volume of materials but also to an ever widening array of responsibilities. In her words:

We need to accomplish more processing in less time because those archivists who once spent all of their time processing are now also involved in EAD

reconversion projects, planning and carrying out digitization projects, designing infrastructure to handle electronic records, applying for grants to preserve audio and video collections, or designing new storage areas and moving materials into them (2006, p. 276).

She found that the implementation of what she termed “minimum standards” was not enough to reduce the backlog and keep up with incoming accessions. She described a case study in which processing was pared down further to reflect the recommendations of Greene and Meissner. She reports significant gains in the 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” (Weideman, 2006, p. 276).

In 2010, Crowe and Spilman published the results of their study to assess the impact of the MPLP as proposed by Greene & Meissner (2005). This study consisted of an electronic survey of the archives community in an attempt to gain insight as to the effect on a broad range of repositories. Their survey requested information about both the processing and the reference functions. They found that MPLP has been widely adopted by the archival community. Additionally, they report that the “responses indicated a positive effect of MPLP procedures on processing backlogs, researcher access, and the ability of reference staff to more easily assist researchers in their quests” (Crowe & Spilman, 2010, p. 122). However, they note that further study could focus specifically on the effect of MPLP on descriptive practices.

Not all archivists have embraced MPLP without reservation. Cox shares Greene and Meissner’s “concern for cost efficiency and fitting the level of treatment to the needs of the collection” (Cox, 2010, p. 143). However, he notes that while there may be cost savings in the area of processing, the application of MPLP could also result in additional

costs elsewhere. “The additional costs imposed on appraisal, reference, retrieval, and preservation may individually be small, but when tabulated across the entire minimally processed corpus and extended across the decades in which we curate those collections, it becomes more difficult to dismiss the costs as ‘probably’ not excessive” (Cox, 2010, p. 139). He proposes a maximal processing model which, while compatible with MPLP, emphasizes the function of description. In this model “archivists should seek to maximize both physical and intellectual care of collections (and here is the important part) within the real and practical limits of the resources at their disposal” (Cox, 2010, p. 143). The process is a continuous one of pre-description, description, and post-description.

While the previous articles focus on backlogs and processing of analog materials, they do so within the larger framework of user access. The same issues arise when considering user access to digitized versions of analog materials. Erway (2008) shares ideas from an OCLC Research Library Group (RGL) event that focused on increasing the scale of digitization of special collections with the ultimate goal of increased access. She suggests that special collections librarians “stop thinking about item-level description. Learn from archivists and let go of the obsession to describe items” (Erway, 2008, p. 327). Even with the minimal description, material that is available online is more accessible than that which is available only to those able to travel to the repository. The article includes brief descriptions of a number of repositories with successful digitization projects. One such project at the University of Wisconsin used a streamlined model of scanning in bulk without item-level metadata which resulted in a significant reduction in cost.

Prochaska (2009) reports on the Association of Research Libraries (ARL) Special Collections Working Group's consideration of issues related to digitization of special collections. One of their major areas of concern was "ensuring discoverability and access" (Prochaska, 2009, p. 18). While noting the increased demand for digitized material, she states that "one critically important recommendation of the ARL Special Collections Working Group is that there should be no digitization without metadata" (Prochaska, 2009, p. 23). She also points to the Council on Library and Information Resources (CLIR) that is seeking "models of good practice for the lean and effective description of collections" as a means of promoting access and discoverability (Prochaska, 2009, p. 23).

In 2010, Mark Greene, one of the original proponents of MPLP, expanded the application of the strategy beyond the function of processing. He discusses ways in which MPLP might also be applied in terms of appraisal, reference, preservation, electronic records, and digitization. He cites 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" (Greene, 2010, p. 189). He posits that archivists are stymied in their efforts to make digital content available, because "we have been inculcated with the belief that we must provide item-level metadata, which is exorbitantly expensive, for collections of any size" (Greene, 2010, p. 194). He rejects the notion that item-level description must be provided. He, instead, favors file or series-level description. He states, "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'" (Greene, 2010, p. 194).

User Needs and Expectations

This movement by archivists toward minimal descriptive metadata risks running counter to user needs and expectations. Pearce-Moses notes, “Today we live in an Amazoogole world, where people expect comprehensive information, 24/7, offering immediate gratification, and customized to the customer” (2007, p. 18). As early as 1998, Craig noted the effect of computer technology on users’ expectations for access to materials. She refers to the notion held by some users that everything is available on the Internet as the “myth surrounding information technology” (Craig, 1998, p. 120). She ponders the question, given the growing reality of digitization, of “how best to deliver our archival goods to all users?” so that archives do not become marginalized players in the information marketplace (Craig, 1998, p. 122). She writes of the need to determine what users want and need from archives.

In 2000, Duff and Cherry reported on their usage study of a digital library collection. Users were asked to compare original paper, microfiche, and a web-accessible digital format. Users preferred the digital format primarily because of its accessibility. However, their expectation for digital material was that they should be able to quickly locate relevant material. Graham (2003) mentions user expectations related to the American Memory collection of digital materials available at the Library of Congress website. She notes that users underestimated the complexity of making content available online and assumed “that everything is or will be there soon” (Graham, 2003). Additionally, users expected to be able to use the site as they would an encyclopedia.

Sweet and Thomas (2000) report their experience at the Public Records Office of the UK National Archives. They note that collection-level archival description that is

rooted in the concept of provenance is “of enormous value to some researchers, particularly those who are conducting research into the history of individuals or institutions” (Sweet & Thomas, 2000). However, they also note “that they [collection-level descriptions] are of limited value on their own to a broad range of researchers” (Sweet & Thomas, 2000). In actual practice many archive users require “clear, accurate and searchable descriptions of individual files” (Sweet & Thomas, 2000). “They then ‘move bottom upwards’ to see the context in which the documents were created and used” (Sweet & Thomas, 2000).

Gilliland-Swetland argues that finding aids in a digital environment should not merely be digitized versions of the paper format, but rather, should exploit the Encoded Archival Description (EAD) metadata infrastructure to enhance information discovery “in order to meet the diverse range of information needs and associative information-seeking behaviors and practices” of online users (2001, p. 200). She provides an overview of the diverse information seeking practices of several user groups. She cites genealogists as the most prevalent type of avocational users of archives. This user group can benefit from item-level search and retrieval. She suggests enhancing EAD so that users will be able to conduct name and natural language geographic keyword searches. She also suggests that dates be associated with the name or geographic metadata to provide context.

Duff and Johnson (2003) performed an in-depth study of the information-seeking behavior of ten genealogists. They reported on the stages of genealogical research as well as the methods for searching for information. In one instance they determined that the ideal search engine “would support a search by name, geographic area, and a range of

dates. It would also contain digitized images of the original documents” (Duff & Johnson, 2003, p. 87). They also note that “provenance-based finding aids provide significant challenges to novice genealogical researchers” (Duff & Johnson, 2003, p. 91). They recommend that traditional and EAD finding aids be redesigned to meet the needs of genealogists.

Tibbo (2003) and Anderson (2004) collaborated in separate studies of the information seeking-behavior of another important archival user group: historians. Tibbo found that American historians used both print and electronic resources to locate original material. By implication, archivists need to maintain traditional finding aids while enhancing the electronic ones. Ultimately, she calls for the electronic finding aid and archival databases to be transformed into “historians’ tools” (Tibbo, 2003, p. 29). Anderson found that academic historians in the UK “desire to see more online finding-aids with greater levels of detail” (2004, p. 114). He posits 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” (Anderson, 2004, p. 83).

Elena, Katifori, Vassilakis, Lepouras, and Halatsis (2010) studied historians with a focus on query formulation and the search process in the context of both print and digital materials. Based on their findings, they identify requirements to facilitate discovery of information by users. One of the requirements is that contents of digital repositories should have metadata for items that include information typically used in researchers’ queries. They recommend that such metadata should be structured to allow

“the researcher to gain more control over the search procedure and get more relevant results” (Elena et al., 2010, p. 34).

Yakel and Torres (2003) did an in-depth study of archive users that explores the concept of archival intelligence. They determined that researchers need a “basic conceptual knowledge and the development of a general framework of archival management, representation and descriptive practices, and search query formulation” in order to become expert archives users (Yakel & Torres, 2003, p. 54). This is an important consideration for user search and discovery in the unmediated online digital collection environment.

A study by Duff and Stoyanova (2004) found that users are often unfamiliar with the terminology within finding aids and, thus, find it to be confusing. One study participant referred to the finding aid as a “crazy quilt” that made sense to the archivist but not to the user (Duff & Stoyanova, 2004, p. 60). For a researcher in the repository, this confusion can be mitigated by the mediation of an archivist. In an online setting, this type of user instruction becomes more problematic.

In her article about the characteristics of online users, also referred to as the “invisible researcher”, Hill (2004) cites two surveys as well as the experience of the National Archives (UK) that point to the fact that a large proportion of such users are leisure users engaged in searching for family information. Because leisure users seek information about specific individuals and places, detailed item-level description is critical to serving their needs. In her view, serving the needs of remote users should be a priority.

Adams (2007) conducted a study that focused on the use of electronic records at the National Archives and Records Administration (NARA) by historians and genealogists. She found that, while they may use the same or similar sources, users who sought facts or specific information had different needs and expectations than users who were analyzing bodies of information with the goal of producing “new knowledge”. She also cites NARA statistics that indicate that in terms of electronic records there has, indeed, been “a shift from an archivist-user interpersonal exchange to a user self-service mode” (Adams, 2007, p. 31).

Daines and Nimer (2009) describe Web 2.0 and possible applications within an archival environment. They also discuss user expectations:

Our patrons have come to expect that our content will be available in digital form, and that they will be able to interact with that content and obtain research help if necessary while they are engaged with our content – all of which they expect to occur virtually (Daines & Nimer, 2009).

User-Contributed Content

User-contributed content has been suggested as a means of describing digital collections or of enhancing such descriptions (Anderson, 2004; Erway & Schaffner, 2007; Prochaska, 2009; Greene, 2010; Elena et al., 2010). It has also been suggested as a means by which to engage users and enrich their experience with respect to online collections (Pearce-Moses, 2007; Horava, 2009).

Light and Hyry (2002) discuss user-contributed annotation of finding aids. They posit 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. However, they note some of the drawbacks as

well, not the least of which is the potential to “threaten archival professionalism” (Light & Hyry, 2002, p.228).

In 2006 Yakel discussed current and future use of Web 2.0 features for archival access. In her review of several websites that had employed such features, she found that “engaging the researcher and eliciting their knowledge base can strengthen metadata about collections as well as the collections themselves” (Yakel, 2006, p. 160).

Evans (2007) introduces 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. 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 (Evans, 2007, p. 397). However, he cautions 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.

Anderson and Allen (2009) also propose a peer-based model which would expand user involvement in the description of online collections in an open, interactive archival commons. In their view:

The archival commons would be an interactive, hands-on environment offering users the ability to interact with [digital] objects by contributing linking between objects, (re)arrangement, tagging, naming, annotation, and narration, with provenance tracking, recommendations, and collaborative, associative, and visualization opportunities throughout (Anderson & Allen, 2009, p. 400).

DEBORA (Digital Access to Books of the Renaissance) is an early example of user-contributed information functionality. The online archival system contained an interface that was designed to support scholarly collaboration through the use of

annotations and highlighting. The system also allowed error correction and user-supplied metadata in the form of additional keywords. Nicols et al. (2000) noted that image collections often lacked detailed metadata. They wrote, “It is at this level, of detailed page-specific metadata, that collaborative contributions by the users of image collections could prove most valuable” (Nichols et al., 2000, p. 243).

Other 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. Van der Sluijs and Houben (2008) designed a system for creating structured metadata through the use of software that correlates user-generated descriptive tags with existing metadata ontologies. Their subsequent user study of the system revealed that while users were enthusiastic about the improvement in the discoverability of items, they “struggled with large results sets within a specific facet” (van der Sluijs & Houben, 2008, p. 91). However, van der Sluijs and Houben also mention the benefits of user involvement. They note such involvement allows their institution to connect with and potentially expand the user base, provides a new channel of communication with users, and reduces the cost of metadata creation.

Couasnon, Camillerapp, and Leplumey (2004) incorporated users in their system to provide access to content in handwritten documents. They devised a system that employed user annotations in conjunction with automated document image analysis for documents that were not suitable for traditional optical character recognition (OCR) methods. In a follow up article, Couasnon (2006) assesses the implementation of the system. He reports on “unexpected difficulties” associated with the automated

recognition of 165,000 nineteenth century military forms (Couasnon, 2006, p. 174). He notes problems due to damage of the original documents and as well as the conditions of digitization. While admittedly the article focused on the automated portion of the system, it should be noted that he did not report any difficulties associated with user annotations.

Doumat, Egyed-Zsigmond, and Pinon (2011) also devised a system for creating access to content in handwritten documents, in this case an online archive of ancient manuscripts. Their method integrated user annotations with semi-automatic annotation. They also included automated user activity tracing and recommendation functions. While they did assess the tracing function, they did not perform an assessment of user annotations. In the future, they plan to investigate further means to facilitate discovery of documents through the connection of images or collections based on semantic analysis of annotations.

Several projects have incorporated user-contributed keywords to enhance discoverability of items within collections. This is often known as tagging. Clayton, Morris, Venkatesha, and Whitton (2008) report on user tagging of online cultural heritage items in Australia. They found that the benefits derived from the presence of tagging functionality included improved discoverability of items, an increase in users' sense of engagement with the institution and collections, and users' positive feelings associated with contributing to the public good. However, they also note some of the difficulties associated with allowing users to engage in tagging of content. Such issues included limited resources available for investment in technological tools and platforms, resistance from repository staff, and concern about the quality of user assigned tags. They caution, however, that "by not adopting user tagging, cultural institutions may miss the

opportunity to connect with the online community and make their collections more accessible as Web 2.0 technology becomes more prevalent in the online environment” (Clayton, Morris, Venkatesha, & Whitton, 2008, p. 24).

Oomen et al. (2010) write about the tagging of a large-scale video project using an online game called *Waisda?*. The purpose of the tagging was to provide a means of locating short clips of very specific content as is often need by media professionals. Such tagging serves as an alternative to video discovery aids such as speech-to-text or visual concept detection. A qualitative analysis of the user-generated tags revealed that users focused on what was seen or heard, while professionals assigned descriptors based on topical subjects. They note that the user tags differed from the controlled vocabulary of professionals which may indicate “that tags contribute to bridging the semantic” gap (Oomen et al., 2010). Oomen et al. conclude that tagging is beneficial and “a lot can be gained for heritage institutions and the public alike” (2010).

Leason (2009) reports on qualitative research into the motivations and experiences of those who participated in tagging digital surrogates of museum objects as part of the *steve* project. Survey responses revealed a positive reaction from participants who expressed altruistic motives and a desire to be involved in the work of the museums. The project team deliberately chose to use “a set of simple interfaces” and did not explore social media functionality (Leason, 2009).

In January 2008, the Library of Congress (LOC) was the first to participate in the Flickr Commons in partnership with the photo-sharing website Flickr. Flickr Commons is a site on which cultural heritage institutions can showcase collections of digital images for the purposes of making them accessible to the public and obtaining information about

the images from the public. The public is able to contribute tags, comments, annotations, and engage in virtual discussions. Zarro and Allen studied the annotations added to the LOC images on the site in order to understand how user contributions of information enhanced the library's holdings. "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" (Zarro & Allen, 2010, p.46). They found that the comments and notes appear to have been contributed by users for the purpose of adding context to an image. In their view, personal remembrances are desirable because they "give the researcher access to 'hidden facts' about a resource" (Zarro & Allen, 2010, p. 51). They note that "the public has shown they are willing and able to provide detailed and valuable annotations, corrections, and translations for the Library [of Congress]" (Zarro & Allen, 2010, p. 53). However, they also cite "trust and authority of the work" as issues that need to be addressed (Zarro & Allen, 2010, p. 53).

Graham (2003) reports 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 (Graham, 2003).

Clearly users had a desire to contribute information even though the site lacked the specific functionality to do so.

The Polar Bear Expedition Digital Collections, however, does provide the functionality for user contribution of information to the finding aid by incorporating features such as bookmarking, commenting, and linking. Krause and Yakel (2007)

conducted a preliminary assessment of the site. They analyzed the first six months of data associated with site visitors and their use of the site. 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 note that use of the social interaction features was limited (Krause & Yakel, 2007, p. 312). In their conclusion, they ponder whether other tools, such as finding aid annotation, tagging, and explicit ranking might be more effective.

Nimer and Daines (2008) were also interested in the redesign of the traditional finding aid to include Web 2.0 features in order to better meet the needs of users. Their usability studies “reinforced the need to examine social navigation tools and selectively integrate them with our finding aids” (Nimer & Daines, 2008, p. 229). Additionally, they “were surprised to find that archivists need social navigation tools almost as much as our patrons”, citing these as means to inform users about collections (Nimer & Daines, 2008, p. 230).

Sedgewick (2008) studied user annotations from three online archival collections. She concludes that “users most often contribute informational content, such as identification, further contextual information, and links to related resources” (Sedgewick, 2008, p. 37).

Summary

It is an unfortunate coincidence that archives are moving toward minimal description at a time when many users need and expect more. Research seems to indicate that allowing users to contribute content holds some promise for augmenting traditional description, thus increasing the discoverability of materials, and for engaging users in an

online environment. 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 other entities that have tried this with digitized cultural heritage materials on a large scale.

Methodology

The purpose of this study is to examine the types of users who contribute content to a large-scale online collection of cultural heritage material as well as what such users contribute. I have built on the research of Krause and Yakel (2007), Sedgewick (2008), and Zarro and Allen (2010). For this study, online collections of cultural heritage material are defined 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 collections are 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, large-scale online collections of cultural heritage material are limited to those collections available at Footnote (www.footnote.com).

About Footnote

Launched in January 2007, Footnote¹ is a subsidiary of iArchives², a provider of digitization services. Footnote has entered into partnerships with entities, both public and private, that seek to have collections of analog material digitized and made available to users online. Currently, Footnote makes more than 70 million original documents and images available through their 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. The site allows two levels of access: (1) subscription membership with complete access and interaction with all content on the site and (2) free membership which is more limited and requires only registration with an email address. The more restrictive free level of access allows the user to search and browse all collections with full viewing and interactive features only for the free collections or for documents and images contributed by other members. Additionally, all National Archives content is available at no charge in the National Archives research rooms in Washington D.C., regional facilities, and Presidential libraries.

I selected this website because it met several criteria. First, although Footnote itself is not a repository for cultural heritage materials, the content of the website is comparable 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.

¹ On 18 August, 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.

² On 21 October, 2010, iArchives was acquired by Ancestry.com.

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 is predominately comprised of 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 collections are digitized and presented in a manner that is designed to be appropriate for online archival collections. The agreement with NARA requires the creation of technical and functional metadata that will “enable retrieval of the material at the fundamental level of archival control” in accordance with NARA standards (*NARA - IArchives*). 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 archive, 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.

Figure 1. Footnote image viewer with Filmstrip feature <http://www.footnote.com/image/#1|271943614>

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 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 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. The Member Discoveries page, that reflects many of the user contribution categories, is updated in real time. It is not unusual to visit the page and find that the last contribution was made just seconds ago. In fact, that particular page has a pause feature so that it remains static for better readability. Thus, there is a substantial amount of user-contributed content available for analysis.

Figure 2. Footnote User-Contributed Content composition

Content Analysis

For this study, users are limited to members of Footnote who have engaged in activity that results in content being contributed to the website. No distinction is made between members with basic membership and those who are paid subscribers.

Information about each user is limited to the information available on the user's profile

page. User-contributed content is the addition of information about a collection or an item within the collection or the addition of new materials to a collection on a voluntary basis. For purposes of this study, user-contributed content is limited to information that is freely and publicly available on the Footnote site as a “Member Discovery.” The unit of analysis is each instance of the contribution by a member of an annotation, comment, connection, the creation of or the 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. This method is unobtrusive and considers only “the visible, surface content” (Babbie, 2007, p. 338). Because each category of user-contributed content is different and the available information associated with each category is varied, I developed a coding scheme that is customized to each category. The codebook for each category is contained in Appendices A through H.

Intercoder Reliability

All of the coding for the study was performed by a single coder. However, a limited test of intercoder reliability was conducted for the purpose of testing the reliability of the code design as well as to reveal potential bias on the part of the primary coder. The primary coder and a secondary coder each coded the information for a sample of user profile pages and user-contributed content. A sample of five instances was drawn from the most recent for each type, for a total of 55 items. The two individuals were in complete agreement in coding with two exceptions as shown in Table 1. The differences in both cases did not result from a different understanding of the coding criteria, but rather, from a differing level of knowledge about the digitized item or the associated

metadata arrangement. In the first instance, the user submitted a comment with a correction to a World War II Missing Air Crew Report number. This was abbreviated in the comment as MACR. The primary coder coded the comment as a correction while the secondary coder, who was unfamiliar with the context and terminology, coded the comment as “other.” In the second instance, the user created a spotlight for “lewis burling” in the Revolutionary War Rolls collection. The primary coder coded the spotlight as “person”. The secondary coder did not recognize “lewis burling” as the name of an individual because the user did not capitalize the name in his/her note. The name appears with proper capitalization in the image of the item, but the secondary coder was not adept at reading 18th century script. Thus, the secondary coder coded the spotlight as “undetermined.” This highlights the importance of having an understanding of the items within the collections as well as the user contributions in the coding process.

Table 1. Results of Intercode Reliability Testing

Coding Category	Agreement/ Total Tested
Users	5/5
Annotations: People	5/5
Annotations: Place	5/5
Annotations: Date	5/5
Annotations: Other	5/5
Comments	4/5
Connections	5/5
Footnote Page Contributions	5/5
Spotlights	4/5
Uploads	5/5
Footnote Pages	5/5
Total	53/55

Sample

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 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 3³. However, content analysis of such a large sample was unfeasible. Therefore, I elected to analyze 15% of each category with a minimum of 50

Figure 3. Footnote User-Contributed Content composition for the sample for the week of June 7-13, 2011 compared to the total contributions.

³ Statistical analysis of the data was not performed.

items. This narrowing of the scope of the study resulted in an adjusted sample size of 1,398 items excluding Footnote Pages.

Determining a sample for Footnote Page analysis proved more problematic. I captured Member Discovery information for 5,000 Footnote Pages and still had not captured the entire week of data. The high volume of page creation is due to the fact that, in addition to users, Footnote staff members are also actively engaged in the creation of Footnote Pages. They created 4,903 out of the 5,000 Pages. I elected to consider the 97 member-created Footnote Pages as the sample for analysis for this category. The addition of the sample of 97 Footnote Pages to the samples for the other categories brings the total user-contributed content for analysis to 1,495 items as shown in Table 2. The sample of users to be examined is composed of the 183 users for whom contributed content of all categories was included in the study.

Table 2. Sample for Analysis (Adjusted)

Contribution Category	Number of Items
Annotations	847
Comments	50
Connections	122
Footnote Page Contributions	261
Spotlights	50
Uploads	60
Footnote Pages	97
Total	1,495*
*Contributed by 183 Users	

Users

Methodology

One of the facets of this study is to gain more insight into who is contributing content to Footnote. I propose to build upon the research of Krause and Yakel (2007). Based on data from user profiles and a survey, they found “that genealogists are the more engaged visitors to the Polar Bear Expedition site” (p. 297). With this in mind, I designed the coding scheme to attempt to identify the genealogist/family historian user of Footnote. Additionally, based on a cursory review of contribution activity by users, I attempted to identify users who are affiliated with organizations. Finally, because the literature has identified historians and academic researchers as users of cultural heritage collections, I attempted to identify those as well. I assigned each user to one of four types: (1) individual with family connections, (2) organization, (3) researcher, or (4) other. A user may not be assigned to more than one type. User types were determined from information on the user’s Member Profile Page. A user was identified as an *individual with family connections* if family relationship descriptors appear 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 contains information about the user’s association with an organization. A user was categorized as a *researcher* if the profile page mentions non-family related research. A user is categorized as *other* if the profile page does not contain information identifying him or her as one of the other three types. The codebook is contained in Appendix A.

Findings & Discussion

An analysis of the 183 users who contributed content shows that more than two-thirds could be identified as individuals with family connections to the collections (see Table 3). 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 as shown in Table 4.

Table 3. Types of Users who Contributed Content

User Type	Number of Users	Percent of Sample
Individual with Family Connections	119	65%
Organization	2	1%
Researcher	1	>1%
Other	62	34%
Total	183	100%

Table 4. Number of Contributions by User Types

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

The high level of engagement of individuals with family connections is also reflected in the findings of other studies (Krause & Yakel (2007), Sedgewick (2008)). This finding is compatible with Evans' (2007) comment describing the tradition within the genealogical community as one of sharing information as well. The engagement of these users may also be a reflection of their information needs. Adams (2007) describes this type of user as fact seeking. The contribution of content may be a means to support their own fact-seeking behavior as well as that of other users.

The study reveals that while organizations comprised 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 Holley (2010) in her description of “super volunteers” and library crowdsourcing. She writes, “Although there may be a lot of volunteers the majority of the work (up to 80% in some cases) is done by 10% of the users.” In the *Waisda?* tagging project, Oomen, et al. also note “the exceptional effort put into the game by a small number of users” whom they labeled “super taggers” (2010). They recommend finding a way to specifically target these users. This study of Footnote suggests that this strategy might also be valid for online cultural heritage collections.

As shown in Figure 4, users appear to demonstrate a preference for different contribution types⁴. 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

⁴ Statistical analysis of the data was not performed.

contributions of individuals without family connections most often took the form of Footnote Page contributions.

Figure 4. Number of Contributions by User Type

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 5). 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% was associated with collections that require subscription. This bodes well for repositories with open access to their digital collections.

Figure 5. Number of Annotations, Comments, and Spotlights Contributed to Free Collections & Subscription Collections

Annotations

Methodology

The Footnote annotation feature provides a means for users to add labels or transcriptions to an item. This study seeks to extend the research done by Yakel and Krause (2007) and Sedgewick (2008). Sedgewick comments, “It would be particularly interesting to examine user annotations in an online archival collection community with only very high-level aggregate description” (Sedgewick, 2008, p. 36). Footnote includes collections of that nature.

The Footnote site provides minimal user instruction about the content and structure of annotations. Users are instructed to type exactly what they see. The study examined whether user annotations are an accurate reflection of text within a document, especially in terms of word order and date format.

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. The study examines whether users selected the appropriate designation for the content they contributed.

Additionally, I considered the format of the item that was the subject of the annotation. Because Couasnon (2006) and Doumat, et al. (2011) comment about the difficulty of using OCR technology for handwritten documents, I was especially interested in whether users annotated items in the handwritten format within the collection. The issue of format is also raised by Sedgewick when she poses the question of whether “digitized images invoke more response than written documents” (2008, p. 36).

In the process of my coding, it became apparent that users use the annotation feature to add identifying information to photographs. While this does not appear to be the purpose of the annotation feature as defined by Footnote, the users have adapted this practice to suit their purpose. I adjusted the codebook to reflect this unexpected use. The codebook is contained in Appendix B.

Sample

I elected to analyze a stratified sample of 15% of the annotations that were created during the week of June 7 – 13, 2011. A stratified sample of annotations was selected for the purpose of analyzing information associated with each type (people, place, date, other) in proportion to the amount contributed. I analyzed the most recent annotations for each type within the specified time period. The distribution of the total sample of 847 annotations is shown in Table 5.

Table 5. Stratified Sample of Annotations

Annotation Type	Number of Annotations	Percent of Sample
People	629	74%
Place	67	8%
Date	(minimum) 50	6%
Other	101	12%
Total	847	100%

Findings and Discussion

Table 6 shows that organizational members were much more engaged in contributing this kind of content than the other groups. They contributed 62% of all

Table 6. Annotations Contributed by User Type

User Type	Number of Annotations	Percent of Sample
Individual with Family Connections	238	28%
Organization	525	62%
Researcher	1	>1%
Other	84	10%
Total	847	100%

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 in the annotation category. Additionally, individuals with family connections contributed nearly three times as many annotations as other individual users.

With regard to the minimal instruction for users to “type exactly what you see” when contributing annotations, they followed the recommendation 64% of the time. However, the remaining 36% of the annotations, as shown in Figure 6, point to some

Figure 6. Composition of the 303 Annotations that are not "What You See"?

interesting issues. The largest group of annotations (104), for which text is not recorded as it appears, combines two pieces of information within the item instead of being a literal transcription of text in the item. All of these annotations are associated with a record in which a slave is mentioned. Paterson (2001) discusses the indexing problems faced by archivists when trying to create access to recorded information about slaves. The user chose to combine the slave’s name with the designation “slave”. Figure 7 is an example of this type of annotation. While this method does not coincide with Paterson’s recommendation of incorporating the slave owner’s name in the reference point, the user

is consistent in the application of the method he or she has devised. (Because this user is identified as an organization, it is not possible to identify whether this is one or more individuals.) Additionally, where there is a mention of a slave buyer or seller, the name of that individual is also annotated along the role he or she had in the transaction. Thus, while perhaps not in accordance with accepted indexing principles, this form of annotation provides a form of access to information that is typically not discoverable through traditional search methods.

Figure 7. Example of an annotation from a slave record. From South Carolina Estate Inventories and Bills of Sale, 1732 – 1872 <http://www.footnote.com/image/#266967750>

The second largest group of annotations (87) that is not recorded as they appear in the item is annotations of photographs. In 49 instances, the annotations were contributed by individuals with family connections and provided information about photographs that had been uploaded by members. Another 33 annotations are comprised of information abstracted from an item that is adjacent to the photograph within the same collection. In both situations, it appears that the users are adapting the annotation feature to suit their purpose of associating identifying information directly with a photograph.

Another group of annotations that do not reflect information exactly as it appears in the document are abstracts of information rather than verbatim transcriptions. While

these 33 annotations do not follow the recommendation, they can enhance discoverability of the item.

There were 46 instances in which the date format in the annotation did not reflect the date as it appeared in the item. In addition, there were 21 instances in which 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 serve as an aid to users and would add consistency for names and dates.

In terms of format, 74% of the items that were annotated by users were handwritten documents as shown in Table 7. This is an indication that users are, indeed, willing to annotate these items for which OCR is problematic. The analysis also seems to suggest, at least for these collections, that images do not “invoke more response than written documents” (Sedgewick, 2008, p. 36). It appears that user groups have preferred

Table 7. Annotation Item Formats

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

Table 8. Types of Formats That User Types Annotate

User Type	Number of Items				Total Sample
	Handwritten	Typescript	Photograph	Other	
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

formats as shown in Table 8. 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 type of user 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 was responsible for all of the typescript items and photograph annotations that were attributed to organizations. This appearance of format preference may actually indicate a preference for the content contained within the format rather than an actual format preference. Thus, this finding may instead be an indication that the user is not averse to a particular format.

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. An example of an incorrect association would be an instance in which the user annotated a date but selected the *place* annotation type when creating the annotation. The information in Table 9 shows that they also appear to have a preference with regard to the type of

annotation they contribute. All user types contributed more *people* annotations than any other designation. (The exception to this is the lone researcher who made a single place annotation.) More than half of the *people* annotations were contributed by a single organization. Individuals with family connections contributed more than 3 times as many *people* annotations as other individuals.

Table 9. Types of Annotations by User Type

User Type	Number of Items				Total Sample
	People	Place	Date	Other	
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

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

Comments

Methodology

The Footnote comment feature provides a means for users to augment the descriptive information of items as well as to note corrections to the information. Because this feature allows for the typing of free text without formatting restrictions, users can contribute information of any kind.

User comments have been studied by Krause and Yakel (2007), Sedgewick (2008), and Zarro and Allen (2010). In developing the codebook for the analysis of comments, the categories were adapted from those used in prior studies. The Krause and Yakel study and the Sedgewick study categorize comments as providing additional information and information correction. Graham (2003) also notes these types of contributions as recurrent themes. While I used these categories, I increased the granularity by further analyzing a comment in terms of its association with a person or place. Additionally, Sedgewick used categories that are not applicable for this study. Those categories include donation offer, order request, and edit. As with the Zarro and Allen study and the Sedgewick study, comments can be judged to belong to more than one category, but multiple expressions within a comment of information in the same category is counted as a single instance. As with annotations, I considered the format of the item for which a comment was contributed. The codebook is contained in Appendix C.

Sample

The sample for analysis consisted of the 50 most recent comments that were identified for the week of June 7 – 13, 2011. Because several of the comments included information for more than one category, the number of categorized comments was 58.

Findings and Discussion

The analysis of the comments shows that users primarily provide additional information or correcting information as shown in Table 10. This finding mirrors those of the Krause and Yakel (2007) and Sedgewick (2008) studies. Additionally, the contributed information is most often associated with individuals rather than places or other things. For comments providing additional information, 21 out of 25 were

Table 10. Categories of Comments Contributed by Users

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

associated with individuals. As for corrections, 8 out of 9 related to individuals. To a lesser extent users engage in the description of 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 collections (see Table 11). Thus, the additional information does not directly augment the description of items within the collection, but rather is related to items that are peripheral to the collections.

Table 11. Format of Items on which Users Commented

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

The information in Table 12 shows that individuals with family connections contributed more comments than any other group. Organizations did not contribute any comments. This is in sharp contrast to the substantial number of annotations they contributed. This suggests that organizational members are more interested in increasing discoverability of items rather than augmenting information or engaging with the collections.

Table 12. Comments by User Type

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

Connections

Methodology

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 user-uploaded content and Footnote Pages.

The information analyzed for the connections feature was limited to that which could be readily obtained by viewing entries on the Member Discoveries page. I developed a codebook to use to analyze the types of connections users make (see Appendix D).

Sample

The sample size was equal to 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 ones within that time period.

Findings and Discussion

Users are largely engaged in demonstrating a connection between themselves and other individuals. Table 13 shows that only 4% of the connections created by users show connections between two documents. For 90% of the connections the users themselves are part of the relationship. Of this type of connection, 109 out of 110 connections were created by individuals with family connections. In fact, individuals with family

connections engage in creating connections to a much higher degree (96%) than any other group (see Table 14.) Organizations and the researcher did not create connections.

Table 13. Types of Connections Made by Users

Connection Type	Number of Connections	Percent of Sample
Footnote Member	110	90%
Footnote Pages	7	6%
Connected Images	5	4%
Total	122	100%

Table 14. Number of Connections Contributed by User Type

User Type	Number of Connections	Percent 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 as more of a tool for building community around collections through user engagement rather than one that builds context within and among collections.

Spotlights

Methodology

The Footnote spotlight feature allows users to highlight interesting items they have discovered in the collections so the items may be viewed by other users. Spotlights may be commented on by other users as well. This study examines the types of items that users spotlight. As with comments, I devised a coding system that took into account whether the spotlight highlighted a person or a place with an additional category for event. The content analysis focused on the textual description that the user submitted when creating the spotlight. The codebook for spotlights appears in Appendix E.

Sample

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 a minimum of 50 items. These items are the 50 most recent spotlights that were identified for the time period.

Findings and Discussion

Users created spotlights most often for the purpose of highlighting people. Of the 50 spotlights examined, 78% were associated with people as shown in Table 15. As with annotations and comments, the focus is on individuals. Additionally, as with comments and connections, individuals with family connections demonstrated the highest level of engagement. This group contributed 64% of the spotlights, which was nearly double the number contributed by other individuals (see Table 16). None of the spotlights were

commented on by other users, but this may be a function of recent nature of the submissions rather than a lack of interest.

Table 15. Types of Spotlights Contributed by Users

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

Table 16. Number of Spotlights Contributed by User Type

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

Uploads

Methodology

The Footnote uploading feature allows users to contribute and share documents and photographs. Graham (2003) as well as Krause and Yakel (2007) remark about the users desire to donate these types of materials. At Footnote, users can contribute files

(JPEG, GIF, PNG, TIFF and JPEG-2000) without any form of mediation from Footnote staff. This study examines the types of content that are contributed by users in this fashion. Additionally, the study examines the user-assigned metadata for the uploaded files. I developed a codebook to be used in determining the type of content as well as the presence of a minimum level of descriptive metadata (see Appendix E).

Sample

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.

Findings and Discussion

Table 17 reveals a predominance of instances in which users contribute photographs. In the case of photographs, it is especially important to assign descriptive metadata because it is usually not contained within the depicted image. Of the 36 photographs depicting people, only four were identified with both the first name and last name of the subject, as well as the date or time period, and the location of the

Table 17. Content Types of User-Uploaded Files

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

photograph. Table 18 provides additional data about the metadata for photographs of people. Seven of the photographs of people and all six of the photographs of places were uploaded without descriptive metadata. Files containing images of documents were also 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 18. Metadata for 36 User-Uploaded Photographs of People

Metadata	Number of Photographs	Percent of Sample
Both First & 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 uploaded material was uploaded by individuals with family connections. They were by far the group that was most engaged in this activity. Other individuals contributed 10% of the uploaded material, while organizational members and the researcher uploaded no material. 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 description themselves at the time the material is uploaded. Perhaps this is attributable to the fact that 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

Methodology

The Footnote Page feature allows users to create a page within the site on which to collect and share information about a person, topic, event, place, or organization.

Figure 8 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. The study examines the types of pages most often created by users as well as whether they opted to allow others to contribute to their pages. I devised a codebook with these questions in mind (see Appendix G).

Figure 8. Example of a Footnote Page

Sample

The selection of the sample of 97 Footnote Pages for analysis was described previously in the overall Methodology section.

Findings and Discussion

Footnote Page creation is the only contribution activity in which family-connected individuals engaged less than other individuals. Of the 97 Footnote Pages analyzed, 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 that were 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 three 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

Methodology

Users may contribute content to Footnote Pages that they have created. They may also contribute to other Footnote Pages provided the creator allows it. Users can add facts, images, stories or links to the page. The study examines the types of information users contribute to Footnote Pages. The study considers the degree to which users cite the source of the information they contribute to the page as well. As stated by the Organisation for Economic Co-operation and Development, “The availability of large amounts of information (some accurate and some not) shifts the responsibility to users to correctly assess information found on UCC [user-contributed content] sites” (OECD, 2007, p. 90-91). 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 was made. I devised the codebook for Footnote Page contributions accordingly (see Appendix H).

Sample

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

Findings and Discussion

Footnote Page contributions were relatively evenly divided between those with family connections (52%) and those without such connections (48%). These groups were largely engaged in contributing facts (92%). Other types of content were contributed to a

much lesser degree as shown in Table 19. It is notable that, for 95% of the instances of information contribution to the Fact section of a Footnote Page, source citation information was not present. The instances of non-cited information were 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 is present (see Table 20). As was the case with Footnote Page creation, organizations and the lone researcher did not engage in contributing information to Footnote Pages.

Table 19. Footnote Page Contributions by Type of Contribution

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

Table 20. 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			

Limitations

There are several limitations to this study. One limitation is that it is limited to a single website, Footnote.com. The study would have been improved by including several comparable sites. However, to my knowledge, the Footnote site is unique in the manner in which it allows users to contribute content. Other online collection sites are lacking in one or more of the user contribution functions. Although the case study examines a single site, that issue may be mitigated by the fact that the digital content within the Footnote site originates from a number of other repositories and institutions. A second limitation is that full access to a portion of the Footnote site requires a subscription. Thus, the findings of the case study may not be generalized to other repositories without financial restrictions to access. A third limitation of the study is the 34% of users that could not be identified in terms of the types set forth in the codebook. This was due to a lack of information on the users' profile pages. It is likely that the many of the users who were categorized as "Other" could have been attributed to one of the other categories. This information could have been gathered by contacting the users, but this was beyond the scope of the study. Because of the number of users in the "Other" category, I did not analyze the data for statistical significance. A fourth limitation of the study is that, as in any study, there is the potential for bias in design and implementation. In this case, I view user-contributed content in a positive light and believe that it shows promise in terms of supplementing the description of digital collections. I attempted to neutralize this bias by using manifest content analysis, designing a detailed codebook, and performing intercoder reliability testing.

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 sheds some light on users and their primary interests. However, it also brings 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 is in contrast to the fact that traditional finding aids do not typically include the level of detail that identifies 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 a repository to cultivate a relationship with an organization with a shared interest in the materials in their collection. 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 their collections, the study suggests that people with family connections to collections are the most engaged. A repository 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 do not always “type what they see” 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, it is important to monitor how the features are being used or adapted by users to suit their purposes in order to improve those features to better serve users’ needs.

The study demonstrated that users often do not contribute content in a way that allows other users to identify or assess it. They rarely provide complete descriptive metadata when uploading content. They also rarely provide source citation information when contributing facts. This speaks to the trustworthiness of the information. In the world of repositories, “trust remains an important asset” (Horava, 2010, p. 145). In Horava’s words, “trust saves the user’s time, keeps the user’s attention, and provides an implicit stamp of quality” (2010, p. 145). In fact, Yakel posits whether archivists’ slow adoption of Web 2.0 interactive features stems in part from “a desire to maintain authoritative metadata about collections” (2006, p.160). These reservations are not

lessened by what Hazen (2010) describes as “‘Authority 2.0’ as users, singly or in cohorts, participate in an electronic free-for-all” (2010, p. 120). 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. Perhaps user-instruction might mitigate the situation to some degree. Additionally, Anderson and Allen suggest, “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” (2009, p. 395). Thus, this study suggests that a clear delineation of the original content and 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 to the collections. In the case of connections, only 4% of the connections were between items in collections. Additionally, when users uploaded content, such as photographs that could potentially augment the collections, they were not in the habit of supplying complete metadata. Thus, the building of context was not as extensive as it might have been.

There are additional issues to consider that are outside the scope of this study. For example, how does the presence or absence of the ability to interact with other users affect who engages with the collection?

It would also be useful to explore user instruction in the context of contribution of content to online collections. Is it possible to improve data input consistency, descriptive metadata, and source citation without detracting from the user's experience and enjoyment?

In the case of Footnote, some users appear to be less interested in creating connections between items and more interested in creating connections between themselves and the collections. It would be useful to explore the value of that behavior in terms of users' support for a repository. Having made a personal connection to the online collection are they more inclined to be concerned about the long-term preservation of the materials? Are they equally committed to analog and digital materials? Are they more likely to lobby on behalf of the repository?

Because limited funding is typically an issue with repositories, it would also be useful to explore the effect of this form of user engagement on funding bodies. Do they perceive a high degree of user activity as a positive thing? Do they, perhaps, view it as a negative in terms of diluting the authority and trustworthiness of the institution with the "electronic free-for-all" of information?

Despite these issues and questions, 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." I hope that this case study of Footnote has provided some insight into what a repository might get.

Works Cited

- Adams, M. O. (2007, March). Analyzing archives and finding facts: Use and users of digital data records. *Archival Science*, 7(1), 21-36.
- Anderson, I. G. (2004, Fall). Are you being served? Historians and the search for primary sources. *Archivaria*, 58, 81-129.
- Anderson, S. R., & Allen, R. B. (2009, Fall/Winter). Envisioning the archival commons. *American Archivist*, 72(2), 383-400.
- Babbie, E. (2007). *The practice of social research*. Belmont, CA: Wadsworth Cengage.
- Bucciferro, A. (2008, Summer). Attacking the backlog: NARA archivists mobilize to make unprocessed records available to the public. *Prologue Magazine*, 40(2). Retrieved from <http://www.archives.gov/publications/prologue/2008/summer/backlog.html>
- Clayton, S., Morris, S., Venkatesha, A., & Whitton, H. (2008). *User tagging of online cultural heritage items* (Report). Retrieved from <http://hdl.handle.net/10760/12775>
- Couasnon, B. (2006). What can we learn from the processing of 165,000 forms from the 19th century? *Proceedings of the Second International Workshop on Document Image Analysis for Libraries*, 174-179. doi:10.1109/DIAL.2006.47

- Couasnon, B., Camillerapp, J., & Leplumey, I. (2004). 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*, 270-277. doi:10.1109/DIAL.2004.1263255
- Cox, R. S. (2010). Maximal processing, or, archivist on a pale horse. *Journal of Archival Organization*, 8(2), 134-148.
- Craig, B. L. (1998, Spring). Old myths in new clothes: Expectations of archives users. *Archivaria*, 45, 118-126.
- Crowe, S. H., & Spilman, K. (2010). MPLP @ 5: More access, less backlog? *Journal of Archival Information*, 8(2), 110-133.
- Daines, J. G. & Nimer, C. L. (Eds.). (2009, May 18). Web 2.0 and archives [Special section]. *The Interactive Archivist: Case Studies in Utilizing Web 2.0 to Improve the Archival Experience*. Retrieved from <http://interactivearchivist.archivists.org/>
- Doumat, R., Egyed-Zsigmond, E., & Pinon, J.-M. (2011, August 9). A web 2.0 archive to access, annotate and retrieve manuscripts. *Multimedia Tools and Applications*. doi:10.1007/s11042-011-0851-9
- Duff, W., & Stoyanova, P. (2004, Fall). Transforming the crazy quilt: Archival displays from a users' point of view. *Archivaria*, 58, 44-79.
- Duff, W. M., & Cherry, J. M. (2000, October). Use of historical documents in a digital world: Comparisons with original materials and microfiche. *Information Research*, 6(1).

- Duff, W. M., & Johnson, C. A. (2003, Spring/Summer). Where is the list with all the names? Information-seeking behavior of genealogists. *American Archivist*, 66(1), 79-95.
- Elena, T., Katifori, A., Vassilakis, C., Lepouras, G., & Halatsis, C. (2010). Historical research in archives: User methodology and supporting tools. *International Journal on Digital Libraries*, 11(1), 25-36.
- Erway, R. (2008, December). Supply and demand: Special collections and digitisation. *Liber Quarterly*, 18(3/4), 324-336.
- Erway, R., & Schaffner, J. (2007). *Shifting gears: Gearing up to get into the flow*. Online Computer Library Center Programs and Research.
- Evans, M. J. (2007, Fall/Winter). Archives of the people, by the people, for the people. *American Archivist*, 70(2), 387-400.
- Finch, E. F., Editor. (1994). *Advocating archives: an introduction to public relations for archivists* (p. 1). Metuchen, N.J.: The Society of American Archivists and Scarecrow Press.
- Gilliland-Swetland, A. J. (2001). Popularizing the finding aid. *Journal of Internet Cataloging*, 4(3/4), 199-225.
- Gilliland-Swetland, L. J. (1991, Spring). The provenance of a profession: The permanence of the public archives and historical manuscripts traditions in American archival history. *American Archivist*, 54(2), 160-175.
- Graham, L. (2003, January). Tales from the vault: A journey over the mountains. *Common-place*, 3(2). Retrieved from <http://common-place.org/vol-03/no-02/tales/>

Greene, M. A. (2010, Spring/Summer). MPLP: It's not just for processing anymore.

American Archivist, 73, 175-203.

Greene, M. A., & Meissner, D. (2005, Fall/Winter). More product, less process:

Revamping traditional archival processing. *American Archivist*, 68, 208-263.

Hazen, D. (2010). Rethinking research library collections: A policy framework for

straitened times, and beyond. *Library Resources & Technical Services*, 54(2), 115-121.

Hill, A. (2004). Serving the invisible researcher: Meeting the needs of online users.

Journal of the Society of Archivists, 25(2), 139-148.

Holley, R. (2010, March/April). Crowdsourcing: How and why should libraries do it? *D-*

Lib Magazine, 16(3/4). doi:10.1045/march2010-holley

Horava, T. (2010). Challenges and possibilities for collection management in a digital

age. *Library Resources & Technical Services*, 54(3), 142-152.

Krause, M. G., & Yakel, E. (2007, Fall/Winter). Interaction in virtual archives: The Polar

Bear Expedition Digital Collections next generation finding aid. *American Archivist*, 70(2), 282-314.

Leason, T. (2009, April). *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, IN.

Retrieved from <http://www.archimuse.com/mw2009/papers/leason/leason.html>

Light, M., & Hyry, T. (2002, Fall/Winter). Colophons and annotations: New directions

for the finding aid. *American Archivist*, 65(2), 261-230.

- Nichols, D. M., Pemberton, D., Dalhoumi, S., Larouk, O., Belisle, C., & Twidale, M. B. (2000). DEBORA: Developing an interface to support collaboration in a digital library. In J. Borbinha & T. Baker (Eds.), *Lecture notes in computer science: Vol. 1923. Research and advanced technology for digital libraries* (pp. 239-248). doi:10.1007/3-540-45268-0_22
- Nimer, C., & Daines, J. G. (2008, December). What do you mean it doesn't make sense? Redesigning finding aids from the user's perspective. *Journal of Archival Organization*, 6(4), 216-232.
- OECD. (2007). *Participative web and user-created content: Web 2.0, wikis and social networking*. doi:10.1787/9789264037472-en
- Oomen, J., Belice Baltussen, L., Limonard, S., van Ees, A., Brinkerink, M., Aroyo, L., et al. (2010). Emerging practices in the cultural heritage domain - social tagging of audiovisual heritage. *Proceedings of the WebSci 10: Extending the Frontiers of Society On-Line*. Retrieved from http://journal.webscience.org/337/2/websci10_submission_23.pdf
- Paterson, D. E. (2001, Spring/Summer). A perspective on indexing slaves' names. *American Archivist*, 64(1), 132-142.
- Pearce-Moses, R. (2007, Spring/Summer). Janus in cyberspace: Archives on the threshold of the digital era. *American Archivist*, 70(1), 13-22.
- Prochaska, A. (2009, Spring). Digital special collections: The big picture. *RBM: A Journal of Rare Books, Manuscripts, and Cultural Heritage*, 10(1), 13-24.

- Sedgewick, J. M. (2008). *Let me tell you about my grandpa: A content analysis of user annotations to online archival collections*. Master's paper, University of North Carolina, Chapel Hill.
- Sweet, M., & Thomas, D. (2000, September). Archives described at collection level. *D-Lib Magazine*, 6(9). Retrieved from <http://www.dlib.org/dlib/september00/sweet/09sweet.html>
- Tibbo, H. R. (2003, Spring/Summer). Primarily history in America: How U.S. historians search for primary materials at the dawn of the digital age. *American Archivist*, 66(1), 9-50.
- van der Sluijs, K., & Houben, G.-J. (2010, November/December). Exploiting user tags to build a semantic cultural heritage portal. *IEEE Intelligent Systems*, 25(6), 84-92. doi:10.1109/MIS.2010.143
- Weideman, C. (2006, Fall/Winter). Accessioning as processing. *American Archivist*, 69(2), 274-283.
- Yakel, E. (2006). Inviting the user into the virtual archives. *OCLC Systems & Services*, 22(3), 159-163. doi:10.1108/10650750610686207
- Yakel, E., & Torres, D. A. (2003, Spring/Summer). AI: Archival intelligence and user expertise. *American Archivist*, 66(1), 51-78.
- Zarro, M. A., & Allen, R. B. (2010). User-contributed descriptive metadata for libraries and cultural institutions. In M. Lalmas, J. Jose, A. Rauber, F. Sebastiani, & I. Frommholz (Eds.), *Lecture notes in computer science: Research and advanced technology for digital libraries* (pp. 46-54). doi:10.1007/978-3-642-15464-5_7

Appendix A

CODEBOOK: Users	
What types of Users are contributing content?	
Criteria	Example
Individual with Family Connections: Member's Profile Page contains family relationship descriptors (e.g. mother, father, uncle, grandfather, etc.).	<div> <p>Related Pages</p> <div> <p>Joe Commander Other He was a brother of my great-grandfather, Edward Forrest Commander.</p> </div> </div> <p>http://www.footnote.com/profile/chambln/</p>
Organization: Member's Profile Page contains information about the user's association with an organization.	<div> <p>About Lowcountry_Africana</p> <p>Lowcountry Africana is a free website dedicated to records of African American genealogy and history in South Carolina, Georgia and Florida, home to the Gullah-Geechee cultural heritage. Sponsored by the Magnolia Plantation Foundation of Charleston.</p> </div> <p>http://www.footnote.com/profile/lowcountryafricana/</p>
Historian/Researcher: Member's Profile Page mentions non-family related research.	<div> <p>About Licameli</p> <p>Researching the New York Campaign of 1776, especially the battle of Pell's Point. Also the 6th Regiment NY Heavy Artillery in the Civil War and other military history of all wars. I teach at West Point.</p> </div> <p>http://www.footnote.com/profile/Verplank1782/</p>
Other: Member's Profile Page does not contain information that would identify the user as one of the above types.	

Appendix B

CODEBOOK: Annotations	
Does the annotation reflect the information exactly as it appears?	
Criteria	Example
Yes: The information in the annotation does not vary from what appears.	<p>The image shows a handwritten note with the name 'Edward Harvey' written in cursive. A yellow rectangular annotation box is placed over the name, and the text 'Edward Harvey' is typed in a sans-serif font within the box. A small 'x' icon is in the top right corner of the box.</p> <p>http://www.footnote.com/image/#269306161</p>
No: The information in the annotation is in the item, but is not recorded as it appears in the item (e.g. word order, date format, abstracting).	<p>The image shows a handwritten note with the name 'Thos Eveleigh' written in cursive. A yellow rectangular annotation box is placed over the name, and the text 'Thos Eveleigh' is typed in a sans-serif font within the box. A small 'x' icon is in the top right corner of the box.</p> <p>http://www.footnote.com/image/#269306161</p>
Other: Provides information that is not already present in the item (e.g. additional information for a photo)	<p>The image shows a handwritten note with the name 'Sgt Leslie H. Schweinebraten' written in cursive. A yellow rectangular annotation box is placed over the name, and the text 'Sgt Leslie H. Schweinebraten (ball turret gunner) 27 missions, killed in action on 24 May 44, mission to Berlin, with Worthley crew, aboard 42-97787' is typed in a sans-serif font within the box. A small 'x' icon is in the top right corner of the box.</p> <p>http://www.footnote.com/image/#161244018</p>
Does the annotation type correspond to the information in the annotation?	
Criteria	Example
Yes: The designated type (<i>people, place, date, other</i>) corresponds to the information in the annotation.	<p>The image shows a handwritten note with the date '30th day of April AD 1833' written in cursive. A yellow rectangular annotation box is placed over the date, and the text '30th day of April AD 1833' is typed in a sans-serif font within the box. A small 'x' icon is in the top right corner of the box.</p> <p>http://www.footnote.com/image/#15537444</p>
No: The designated type (<i>people, place, date, other</i>) does not correspond to the information in the annotation.	<p>The image shows a handwritten note with the date '12/6/1837' written in cursive. A yellow rectangular annotation box is placed over the date, and the text '12/6/1837' is typed in a sans-serif font within the box. A small 'x' icon is in the top right corner of the box.</p> <p>http://www.footnote.com/image/#269307073</p>

What is the format of the item to which Users contributed content?	
Criteria	Example
<p>Handwritten:</p> <p>The portion of the item that has been annotated by the user is written by hand. If the item is a printed or typescript form, but the annotation is for the handwritten portion, it is categorized as handwritten.</p>	<div data-bbox="813 405 1057 787"> </div> <div data-bbox="1141 405 1401 787"> </div> <p>http://www.footnote.com/image/#24615629178</p> <p>http://www.footnote.com/image/#269307074</p>
<p>Typescript:</p> <p>The annotated portion of the item is typescript.</p>	<div data-bbox="886 873 1330 1234"> </div> <p>http://www.footnote.com/image/#244483961</p>
<p>Photograph:</p> <p>The annotated item is a photograph.</p>	<div data-bbox="992 1287 1222 1518"> </div> <p>http://www.footnote.com/image/#161244018</p>

Appendix C

CODEBOOK: Comments	
What types of information are in User comments?	
Criteria	Example
Additional Information – Person: Information about a specific person that does not already appear within the item or the related metadata	Carmen died on Normandy Beach during WWII. His mother received a Purple Heart on his behalf for his service. He is buried in Arlington Cemetery with a fellow soldier as their remains couldn't be distinguished separately. http://www.footnote.com/page/86202244_carmen_c_brigandi/#a-comments
Additional Information – Place: Information about a place (building, geographic location) that does not already appear within the item or the related metadata	The building is the Scuola di Guerra Aerea (School of air warfare) in Florence, that still exists, same as then. http://www.footnote.com/image/#161304798
Additional Information – Other: Additional information that does not fall in the above categories	Thse [sic] are Italian pilots of the [sic] Cobelligerant Air Force, of the 51 st Fighter Wing, as can be seen from the crest of the pilot shaking hands. http://www.footnote.com/image/#55064589
Correction – Person: Correcting information about a person that appears within the item or the related metadata	It would appear that this David C. Young served in the 70th OVI rather than the 17th OVI as was listed on the card http://www.footnote.com/image/#5774771
Correction – Other: Correcting information about a place (e.g. building, geographic location) or thing that appears within the item or in the related metadata	It is not an S.81, actually it is an Sm.82 heavy transport. http://www.footnote.com/image/#47312360
Link: Hyperlinks (either linked in or linked out)	see also http://www.footnote.com/page/89946945_robert_t_anderson/ http://www.fold3.com/page/66637488_robert_t_anderson/#a-comments
Relationship – Self: Description of a familial relationship between the user and another person	This is the brother of Henry H. Caneega Sr., my great, great grandfather. http://www.footnote.com/image/#48560072
Relationship – Other: Description of a familial relationship between two people (neither of which is the user)	Their mother was Catherine. http://www.footnote.com/image/#48560072
Other: Comments for which the others categories do not apply	Many Hights 1875 Brooklyn City Directories http://www.footnote.com/image/#228274037

What is the format of the item to which Users contributed content?	
Criteria	Example
<p>Handwritten: The portion of the item that has been annotated by the user is written by hand. If the item is a printed or typescript form, but the annotation is for the handwritten portion, it is categorized as handwritten.</p>	<div data-bbox="813 331 1057 716"> </div> <div data-bbox="1138 331 1398 716"> </div> <p>http://www.footnote.com/image/#24615629178 http://www.footnote.com/image/#269307074</p>
<p>Typescript: The annotated portion of the item is typescript.</p>	<div data-bbox="886 800 1330 1161"> </div> <p>http://www.footnote.com/image/#244483961</p>
<p>Photograph: The annotated item is a photograph.</p>	<div data-bbox="992 1213 1222 1444"> </div> <p>http://www.footnote.com/image/#161244018</p>

Appendix D

CODEBOOK: Connections	
What types of connections are Users making?	
Criteria	Example*
Footnote Member: User is making a connection between himself/herself and another individual.	<p>armistead46 Footnote Member Grandmother added by armistead46 13 Jun 2011</p>
Footnote Pages: User is making a connection between two individuals (neither of which is the user).	<p>Raymond A Zechman Raymond is Dorothy's spouse added by armistead46 13 Jun 2011</p>
Connected Images: User is making a connection between two documents or images.	<p>CONNECTED IMAGES 56 same-person Faustino Jordan.jpg Added: (12 Jun 2011)</p>
	<i>*from dynamic Member Discoveries Page (non-persistent URL)</i>

Appendix E

CODEBOOK: Spotlights	
What type of things are Users highlighting for other Users?	
Criteria	Example
Person: Text description supplied by the user includes a person's name	http://www.fold3.com/spotlight/26709/john_stokes/
Place: Text description supplied by the user includes a location (geographical, building, etc.)	http://www.footnote.com/spotlight/26686/doubleheads_mill/
Event: Text description supplied by the user includes an event or happening	http://www.footnote.com/spotlight/26750/loftusodonnell_wedding/
Other: Text description supplied by the user that does not fall into the category of People, Place, or Event but is still has meaning in relation to the item	http://www.footnote.com/spotlight/26724/ships_cat/
Undetermined: The meaning of the text description supplied by the user cannot be determined in the context of the item	http://www.footnote.com/spotlight/26717/13/

Are Users commenting on spotlights?	
Criteria	Example
Number of comments as shown under the Historical Spotlight	<div>Comments</div> <div>There are no comments.</div>

Appendix F

CODEBOOK: Uploaded Files	
What are Users uploading?	
Criteria	Example
Photograph	<p>http://www.footnote.com/image/#270592336</p>
Document	<p>http://www.footnote.com/image/#270592346</p>
Other	Items for which the photograph and document categories do not apply (e.g. newspaper clipping, portrait)

Are Users assigning descriptive metadata to uploaded materials?	
Criteria	
Photograph: Person	First Name given for at least one person in the Photo
	Last Name given for at least one person in the Photo
	Date or Time Period
	Location
	Other Additional Information
Photograph: Place	Location
	Date or Time Period
	Other Additional Information
Document	Title of the Document
	Source Citation
	Other Additional Information
Other	Reviewed the metadata for the uploaded file for any of the information delineated for Photos and Documents
<div> <div>Example</div> <div> <div>Source information</div> <div>Suggest correction</div> </div> <div> <div>Filename: Gaynor Young, Jersey City, NJ, Easter 4_10_1955.jpg</div> <div>Document Title: Gaynor Eugene Young</div> <div>Document: Gaynor Young, Jersey City, NJ, Easter 4_10_1955</div> <div>Description: Easter 4_10_1955</div> </div> </div>	
http://www.footnote.com/image/#270592325	

Appendix G

CODEBOOK: Footnote Pages	
What types of Footnote Pages are Users creating?	
Criteria	Example
Person: "Person Page" appears at the top of the Footnote Page.	<div> Person Page Dane C. Skuda http://www.footnote.com/page/285045753_dane_c_skuda/ </div>
Topic: "Topic Page" appears at the top of the Footnote Page.	<div> Topic Page Vamosi http://www.footnote.com/page/285045736_vamosi/ </div>
Event: "Event Page" appears at the top of the Footnote Page.	<div> Event Page mt events http://www.footnote.com/page/285045605_mt_events/ </div>
Place: "Place Page" appears at the top of the Footnote Page.	<div> Place Page Treblinka Extermination Camp http://www.footnote.com/page/94255800_treblinka_extermination_camp/ </div>
Organization: "Organization Page" appears at the top of the Footnote Page.	<div> Organization Page Kiwanis Club of Jeffersonville, Indiana http://www.footnote.com/page/111618483_kiwanis_club_of_jeffersonville_indiana/ </div>
Do Users allow others to contribute information to the Footnote Page?	
Criteria	Example*
Yes: Information on Member Discovery entry for the Footnote Page contains the phrase "and others can contribute."	<div> Dane C. Skuda by susanskuda244 and others can contribute. </div>
No: Information on the Member Discovery entry for the Footnote Page does not contain the phrase "and others can contribute".	<div> Fuzi Romavita by b3372521698 </div>
	<i>*from dynamic Member Discoveries Page (non-persistent URL)</i>

Appendix H

CODEBOOK: Footnote Page Contributions	
What types of information are Users contributing to Footnote Pages?	
Criteria Determined by the location of Input(not selected by user)	Example*
Fact
Image
Story
Link
*from dynamic Member Discoveries Page (non-persistent URL)	
Do Users provide source citation information for information they contribute to the Fact Section of a Footnote Page?	
Criteria	Example
Yes: An icon appears in the Source column of the Fact section next to the contributed information.
No: An icon does not appear in the Source column of the Fact section next to the contributed information.