Jessica M. Sedgwick. Let Me Tell You About My Grandpa: A Content Analysis of User Annotations to Online Archival Collections. A Master's Paper for the M.S. in L.S degree. July, 2008. 42 pages. Advisor: Christopher Lee

Archivists have a long history of striving for objectivity in their descriptions of archival materials for their users. However, archivists are only capable of providing limited contextual knowledge about a collection due to their own subjective point of view and limitations on time and resources in processing. One proposed solution is the implementation of user-contributed annotations to online archival content, though many institutions have been slow to adopt this feature into their own digital initiatives. This study provides a content analysis of user annotations from three online archival collections: The Polar Bear Expedition Digital Collections, Beyond Brown Paper, and the Keweenaw Digital Archives. Findings showed that users most often engage in information-providing communication behaviors, with some variances between the three sites. In analyzing existing annotations, this study seeks to inform the role of user-contributed content in archival description.

Headings:

Archives

Archives -- United States -- Information Services

Libraries -- Information Technology

Libraries and the Internet

LET ME TELL YOU ABOUT MY GRANDPA: A CONTENT ANALYSIS OF USER ANNOTATIONS TO ONLINE ARCHIVAL COLLECTIONS

by
Jessica M. Sedgwick

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

July 2008

Approved by

_____

Christopher Lee

Table of Contents

Introduction

Archivists are continually searching for ways to make their arrangement and description practices more efficient and more effective for both users and themselves. In their 2005 study, Dennis Meissner and Mark Greene found that unprocessed backlogs are a major problem at most archival institutions, meaning that archives are collecting materials more quickly than they can process them and make them available for research. This renders a large amount of their holdings inaccessible to researchers, and puts a resource strain on processing archivists. Meanwhile, archival description must attempt to keep up with evolving user needs while avoiding the pitfalls of subjectivity in archival description as has been highlighted by postmodern thinkers, such as Terry Cook and Joseph Deodato. Much of the literature in archival description addresses these issues, and increasingly researchers are pointing to social software (such as annotations, tagging, and bookmarking) as a partial remedy. In their 2007 study, Magia Ghetu Krause and Elizabeth Yakel posited, “By allowing researchers to contribute descriptive notes and other information to archival collections and items, these materials will become more intellectually accessible to a wider variety of users” (288). Despite the promise of increased accessibility, archivists have been slow to adopt this new technology in their own digital initiatives. This study will focus on one such form of social software: user-created annotations, as they have been applied to three online archival collections: Beyond Brown Paper (http://beyondbrownpaper.plymouth.edu/), the Keweenaw Digital

Archives (http://digarch.lib.mtu.edu/), and the Polar Bear Expeditions Digital Collections (http://polarbears.si.umich.edu/).

According to Tom Hyry and Michelle Light's 2002 study, annotations would serve as a means for users and archivists to create and share their thoughts, insights, and other commentary about a given item or collection. User annotations would create a space in archival description -- separate from the archivists' own relatively static, standardized metadata -- where researchers, genealogists, reference archivists, archivists at other institutions, and any other imaginable users can contribute further information to the existing document. Annotations could add value that is sustainable and yet dynamic in that it can be continually augmented over time. This study seeks to explore the utility of user annotations for researchers and archivists by addressing the question: When allowed to contribute to archival description through online annotations, what kinds of content do users provide?

Background

Materials housed in archival repositories may be characterized as "messy," to say the least. While print materials in traditional libraries may be effectively cataloged in an Online Public Access Catalog, archival collections must be described at a more aggregate level, as item-level description are rarely possible. Effective archival description is crucial in facilitating use and navigation of archival collections for several reasons. First, most archival collections do not circulate because they are made up of rare or unique materials, requiring patrons to visit the on-site repository in order to use them. So ideally, a user should be able to determine the usefulness and appropriateness of a collection from its description before expending the resources to visit the repository. Once there, the user

must additionally have some way of navigating archival collections, since some collections consist of a hundred linear feet of materials or more. The ability to navigate is also important for the public services staff of the archive; if patrons are able to locate approximately what materials in a large collection are going best serve their needs, it will save the public services staff much time and effort. Currently, information regarding the physical location and arrangement of materials is represented in archival description.

Traditionally, archival description is presented in a finding aid, a document that provides description and metadata crucial to understanding the nature, contents, and context of a collection, such as: administrative information, processing and acquisition details, historical and biographical context, descriptions of the materials themselves, and subject headings to provide access points. Finding aids are also typically designed to reflect the structure of a collection at various levels by providing collection-, series-, and often file-level descriptions.

Archival description of a given collection is typically created by a single archivist, and therefore can present only a singular view or interpretation of a collection. Since it would require an impossible amount of resources to describe all collections at the item level, processing archivists must assess what parts are of greatest research value, because those require the most descriptive attention. This process includes value judgments that are susceptible to biases. Try as they might, processing archivists cannot totally avoid their own world-views and biases, which may influence their descriptions of the collection and of the biographical and historical context of the materials. For example, women and minorities were often left out of early archival description due to the biases inherent in the dominant social paradigm of the time. This is problematic since archival

description serves as an access point to collections, and helps users to understand not only the content of the materials, but also the context and importance. Users often make decisions about whether and how to use materials based on what they can glean from the archival description. If adequate access points are not provided, users may misjudge the relevance of given materials or never discover the materials at all. In short, the limited resources in archives cultivates an environment in which single archivists are often the sole describers of collections, thus creating potential limitations in access and use for patrons.

Purpose of study

Hyry and Light suggested that allowing users to annotate archival collections would help eliminate the singularity and biases that plague archival description. However, incorporating and maintaining this kind of interactive function could demand more resources from the archivist than it is worth for researchers and archivists alike. Archivists currently have little empirical data inform their expectations about what it would be like to allow user annotations. Annotations could potentially create more work for archivists by demanding constant monitoring, or by creating a new reference vehicle, burdening public service archivists. If archivists have more knowledge of how users annotate materials in online archival collections, they will be better able to plan their resources accordingly, and could feel more comfortable in using this new type of affordance. This paper addresses the following research question: When allowed to contribute to archival description through online annotations, what kinds of content do users provide?

Research efforts in many fields can suffer if the limitations of archival description prevent researchers from using a collection, misrepresent a collection in such a way that the collection appears more appropriate to a researcher's needs than it actually is, or hinders a researcher's understanding of the context of materials. The better the archival description is able to act as an access point for a collection, the more positive users' experiences will be, lending to the success of the repositories. Annotations can not only potentially give voice to users and researchers, but also can also provide a space for reference archivists to maintain insights gleaned from working with the collections in a sustainable and accessible manner, perhaps even lightening their work load to a degree, as they will not have to answer some of the same questions repeatedly. Annotations may also open a new line of communication across the boundaries of archival institutions, as the annotation space may be used to link between related materials. The purpose of this study is to understand the content of annotations currently being provided by users of online archival collections.

Literature Review

Most archival collections are processed and described by a single archivist; therefore all of the description, historical and biographical context, and access points provided for that collection are subject to the limitations of the archivist's own knowledge and personal judgments. Postmodern thinkers have highlighted this problem for some time now, but little action has been taken to remedy the situation. Terry Cook claims that, "For archivists, the paradigm shift requires moving away from identifying themselves as passive guardians of an inherited legacy to celebrating their role in actively

shaping collective (or social) memory" (4). The chief problem in archival description, as revealed by Cook and other postmodern thinkers, is the singularity and narrow perspective through which archival collections are arranged, described, and interpreted. Hyry and Light argue that "Both the absence of documentation of under-represented groups and efforts to correct the situation belie the extent to which values of individual archivists and the profession as a whole influence the historical record" (219). In 1970, Howard Zinn criticized the perception of archivists as impartial keepers of cultural memory by highlighting the lack of archival materials available by women, minorities, the lower class, etc: "The archivist, in subtle ways, tends to perpetrate the political and economic status quo simply by going about his ordinary business" (Zinn, qtd. in Hyry and Light 218).

Archivists are beginning to step out from the guise of impartiality, recognizing their role as active agent in shaping the historical record. Many researchers agree that a dialogue should be opened between processing archivists and users, and descriptive features made inclusive of not merely the authoritative voice of the archivist, but of the general user community (including the underrepresented) who bring additional insight, experience, and value to archival description (Duff and Harris; Yakel, Shaw, and Reynolds). However, the issue of embracing user insights in archival description becomes especially messy and complex when one considers the contradictory importance of descriptive standards and authority in archival description.

Several issues arise with the prospect of allowing users to alter, or even contribute to, the archivists' very standardized, carefully-crafted description. Most archival description and metadata conforms to some established semantic or syntactic standard,

such as Describing Archives, A Content Standard, Encoded Archival Description, Metadata Encoding and Transmission Standard, Metadata Object Description Standard, and Dublin Core, among others. The purpose of these standards is partly to make the work of archival description easier for the archivist, taking out much of the guess work when deciding how to approach archival description. The standards can also benefit users, because they require that a basic set of elements, typically necessary elements for understanding an archival collection, be included in the description. Standards also help users and archivists by making archival description across different institutions more recognizable; if several repositories conform to a particular standard, the user will not have to re-learn how to navigate the descriptions of collections at each institution, because the descriptions will look similar. Archivists put much time and effort into their descriptions, and so it is not surprising that many could be hesitant to allow users, who know little or nothing about their standardized practices, to have any hand in contributing descriptive content.

These issues of authority are acknowledged by Peter Van Garderen, an archivist and blogger:

> I assume, of course, that professional archivists will have issues with blurring the lines between institutionally managed archival materials and descriptions and those contributed, enhanced or re-used by patrons … Another legitimate concern would be to protect the authenticity of archival materials and the context of their original creation and use (8 May 2006).

For instance, imagine an archivist who spends several months processing a collection, writing the finding aid according to the standards outlined in *Describing Archives: A*

Content Standard (Society of American Archivists, 2004), and encodes the finding aid to be put online using Encoded Archival Description (Society of American Archivists, 2002). Much care and effort has been put into ensuring the integrity of the information the archivist has provided; how should she feel about allowing users to add their own potentially misguided tags or comments? However, Van Garderen goes on to envision a situation in which user-created content might live in harmony with archivist-created content:

> I therefore see the introduction of community-managed collections, descriptions, exhibits and discussions as something that happens in parallel to the authoritative archives access systems that are managed by archival institutions and their professional staff … archival institutions stand to benefit from taking a leadership role in encouraging new and innovative use of their collections and being the benefactor and host of new, online communities (8 May 2006).

Thus, Van Garderner recommends that the perceived threat to authority posed by user annotations could be handled by smart metadata management, ensuring that user-created content and archivist-created content live and display in separate spaces.

A 2007 article by Michael Middleton and Julie Lee further addressed this issue, acknowledging that archivists "Have … deep-seated concerns about authority once user content is brought into the mix. The concept of external parties editing the content of an institutional site is problematic from both a brand and a 'trusted organisation' perspective" (19). This perceived threat to descriptive authority will likely remain a very real fear for many archivists, even as some repositories are now opening up and

experimentation with this type of technology. However, like Van Garderen, Middleton and Lee offer suggestions on ways for cultural institutions to successfully incorporate social software into their digital initiatives, such as implementing an interface design that, through graphics, clearly differentiates between archivist- and user-generated content (Middleton and Lee 20).

Duff and Harris advocate descriptive standards that have enough wiggle room to allow for a plurality of interpretation and representation of collections:

> We need to create descriptive systems that are more permeable. In doing so archivists will have to relinquish some of their power to control access to, and interpretation of, their records with which the current descriptive approaches invest them … We need to create holes that allow in the voices of our users. We need descriptive architectures that allow our users to speak to and in them. Architectures, for instance, which invite genealogists, historians, students, and other users to annotate the finding aids or to add their own descriptions would encourage the leaking of power (Duff and Harris, 279).

Though this threat to the authority and control of the archivist remains one of the primary drawbacks to implementing user annotations, Duff and Harris hold that the benefits of presenting a more complete historical record outweigh the costs. They further emphasize that the archivist, who represents the mainstream, should not attempt to assume the voice of the marginalized, but must rather make room in archival description for the marginalized to speak for themselves. One of their recommendations is to allow user-created annotations in archival description, providing an opportunity for the user

community to enter a dialogue with and alongside the "authoritative," standardized, archivist-provided descriptive content. In the postmodern world, the archival literature demonstrates the importance of finding a way to balance the integrity and authority of the archivist, while allowing for open dialogue with users and alternative voices in description and interpretation.

Some archivists may also be concerned that social software, while potentially useful for the tech-savy, may not appeal as directly to their predominant user base of historians, genealogists, and other researchers. There may also be some concern that social software is a trend that will eventually fall out of vogue, and will leave archivists with quite a mess to clean up. Archivist Jesse Brown comments on this in his blog: "I'm not sure if such functionality [social software] will be immediately welcomed by our user community. Most of our users, particularly the more (ahem) 'experienced' researchers, are unlikely to be familiar with using Web 2.0 technologies" (7 August 2006). It is true that the typical user base for archives may not be as ready and willing to adopt social software into their research patterns as are the users of, say, public libraries. However, this fact is changing; current research shows that users of archives are becoming more and more tech-savy, and are even beginning to expect more online utility in their archival research. Helen Tibbo's 2003 study of online primary source research behavior of American History professors shows that the amount of primary source investigation and information-seeking that is conducted online is growing across all levels of academic research. Tibbo supports the concept of archivists sharing power over tools for description and discovery, claiming that, "It is time to make the electronic finding aid and archival databases historians' tools" (29).

Amanda Hill's 2004 study demonstrates that physical users make up an increasingly small percentage of archives' user base, while "invisible" online users are increasing in number. Her study also reveals that since online users are operating though a special medium, the web, this medium should be fully exploited to offer the best service possible to these users in the absence of face-to-face reference assistance. Hill's study found that online users require more detailed information in archival description, such as item-level description, in order to satisfy their informational needs, as they are less likely than visiting patrons to seek out the help of a reference archivist.

Barbara L. Craig argues that, if an archive becomes technologically stagnant, refusing to take advantage of increased functionality offered by new web technologies, it will be left behind in the growing market of electronic information. Some archivists may fear that becoming "too accessible," i.e. attracting new, accidental, uninformed users, may result in a pressing demand on resources. However, Craig claims that "If we manage their visits well, we may convert them into regular visitors and we may realize our larger goal of education" (125). Krause and Yakel envision user annotations as powerful tools that can transform and empower archival description in the digital age: "Annotations could take the form of additions or amendments to existing descriptions, information about use of the materials, and references to other collections, transforming fairly static finding aids into dynamic documents and creating a more open descriptive system" (291). Thus, it is archivists' responsibility to best serve the needs of their user community, needs which these studies show to include enhanced web resources and more detailed descriptive content, both of which Hyry and Light suggest can be satisfied by allowing annotations in archival description.

Archivists are recognizing the need to keep up with their evolving user community and have responded in a variety of ways, including the notion of "light processing." Greene and Meissner's surveys have shown that processing backlogs are prevalent problems in the archival profession, and much of this is due to time and resources spent on tasks that are not of absolute necessity. Greene and Meissner also demonstrated that the backlog problem can negatively affect relationships with donors as well as accessibility of materials for researchers. They propose a paradigm shift in processing prerogatives: that archivists make materials usable for researchers more quickly, that they ensure *adequate* arrangement for user needs, perform *minimal* preservation procedures necessary, and provide description which is *sufficient* to promote use. This suggests that, where appropriate, *less* descriptive detail should be provided in archival description. Once more common, it is now rare in the archival field to process and describe a collection at the item level. Many archival institutions have adopted approaches similar to Meissner and Greene's light processing model.

However, considering the finding that online users of archives require *more* detailed description, archivists must find a way to make archival description as descriptive and detailed as possible without over-dedicating their limited human resources, neglecting an entirely inaccessible backlog of collections.

Greene and Meissner's piece has several implications for the role and potential effects of implementing annotations in archival description. First, implementing a platform to support user annotations in web-based archival description -- as well as developing and maintaining some system of screening, monitoring, managing, and responding to those annotations -- creates additional work that archivists, be they

processors, reference, or other staff members may not have to spare. This will be a real concern for many archivists, and stands as one of the major obstacles to implementing annotations in archival description. Spam can also be a problem when allowing user annotations, though measures can be taken to prevent and manage spamming as Hammond et. al's 2005 article described.

On the other hand, as has been noted, implementing user annotations also has the potential to assist the archivist in her work by helping to fill in descriptive gaps left by the varied levels of description applied to collections, especially lighter treatments. For instance, it may not be possible (due to a lack of time, resources, or personal knowledge) for an archivist to identify and list every person pictured in a large photograph collection, however, users may be able to offer further information, identification, and access points, due to their personal knowledge or experience in working with the collection. In this case, if user annotations are a possibility, the user will have the opportunity to contribute this information alongside the archivist's descriptive content, thereby filling in the gaps. An example of this type of user behavior can be seen in Beyond Brown Paper, a collaborative digital collection launched in October 2006 at Plymouth State University. The collection is made up of digitized photographs that document the history of the Brown Paper Company of Berlin, New Hampshire. At the time of data collection for this project, there were roughly 17,000 objects on the site. The example below (Figure 1) shows an image from this site of a group of people, and a user's annotation to the image, identifying many of those pictured:

Mar 13, 2007 at 1:32 pm

What a wonderful surprise. I have in my archives, a glossy copy of this group photo dating from 1949. Taken in Sanmaur. Upper right, is Marguerite Bourassa,, my first teacher (1950) (grade 1 to 7) . Among the students : Jean-Pierre Ricard, Louis Lacasse, Pauline and Robert Bouchard, children from Brown's white collars in Sanmaur, north of La Tuque. Also Houle, Pelletier, Elie and many other boys and girls from contractors for the Brown and the CIP. I have this original photo, saved from the dump in 1965, in La Tuque. I am trying to identify the other teacher and students.

Fig. 1. Example annotation from Beyond Brown Paper, "Item #860" <http://beyondbrownpaper.plymouth.edu/item/973>

The only archivist-created description for this image is the plate number, year, month, two subjects ("Sanmaur Depot" and "School children"), and a tag ("Exterior"). The user's annotation not only adds additional descriptive information, but is also one of eleven such annotations on this individual item that does so. As this example demonstrates, in a world where archival description is becoming necessarily less-detailed, archivists and users alike may find it incredibly useful and worthwhile to have a space wherein users, reference archivists, and archivists from other institutions can contribute further detail and insight to archival description, augmenting what authoritative, though "light," descriptive work has already been done.

Some of the archival literature points to prototypes and platforms which can be used or built on to see how annotations work, and how they might really work for archivists. One such study was done by Roscheisen et. al, which discusses the development of “ComMentor,” an architecture that supports third-party annotations to existing documents. Roscheisen’s article explains how annotation platforms may live outside of arbitrary web documents, such as EAD finding aids and digital object displays, for easy management. Roscheisen also notes that, though most web platforms only allow whole-document annotations, the development of “in-place” annotations, replicating what is done by hand in books, is becoming a more realistic possibility, as well as the ability to overlay and hide layers of annotations. This is particularly important because a study by Catherine Marshall’s demonstrated that, in digital environments, users prefer the same flexibility they have when making paper annotations.

The literature on annotations in the online environment also suggests that making annotations searchable is a crucial way of exploiting their utility. For instance, Thiel Ulrich et al.’s 2004 study determined that including annotations in search and discovery tools allows for a novel and more advanced and context-based search, assuming that the annotations provide contextual information about the digital objects they are related to (11). This utility can be seen at play in the Keweenaw Digital Archives, a collection of digitized photographs documenting Michigan's historic copper mining district, launched in early 2006. The example below (Figure 2) shows an image of an ice sculpture of Disney character Scrooge McDuck, which has been incorrectly identified in the metadata as Donald Duck. Because this particular site searches across user annotations as well as archivist-created metadata, this image now comes up as a result for a search on “Scrooge

McDuck."

User Reviews

Hide Comments (1)

User	Time	Comment
Anonymous	1/30/2008 5:46:47 PM	I believe this is Scrooge McDuck, the rich uncle of Donald. The monacle and money bags are dead giveaways.

Fig. 2. Example annotation on Keweenaw Digital Archives, "Winter Carnival," < http://digarch.lib.mtu.edu/showbib.aspx?bib_id=629978#>

Another online archival collection that has been experimenting with user annotations and since 2006 is the Polar Bear Expedition Digital Collections, a project of the Bentley Historical Library and the School of Information at the University of Michigan. This collection includes materials, representing a wide variety of formats, on a World War I American intervention to Russia. Many users of this collection have personal memories, first-hand knowledge, and family stories that can augment the archival description. The example below (Figure 3) shows an informative user annotation on an item in this collection that had received no descriptive metadata from the archivist. Note that in this case, the user also links to a source of further information. Since, like the

Keweenaw Digital Archives, this site includes the annotation text itself in its search functions, user annotations can serve as added access points. For example, if one searches the Polar Bear site for the term “Solovetski Island,” she will get the item in Fig. 3 as a result, because of the added information in the annotation. Before that annotation was contributed by the user, the search would have turned up zero hits.

7/24/2007 at 8:16 pm

This is the monastery on Solovetski Island in the White Sea. In the summer of 1919, submarine chaser SC354 became the first USN vessel to visit the island. For a view of the monastery taken during that expedition, see: http://www.subchaser.org/north-04

Fig. 3. Example annotation from Polar Bear Expedition Digital Collections, “Simpson 4-9.1” <http://polarbears.si.umich.edu/index.pl?node_id=15472&lastnode_id=18065>

The Polar Bear Expedition Digital Collections platform also caters to a need identified in the online annotation literature by Robert Wilensky. His 2000 study found that users value the ability to annotate at various levels of collections, from the collection level to the item level, helping them get as specific as needed and accurately associate the annotation with the related object or set of objects. The Polar Bear Expedition Digital Collections offers this utility because the pages are actually structured like an EAD-encoded finding aid that allows users to annotate at the various levels presented in finding aids.

Krause and Yakel's study analyzed a set of annotations from the Polar Bear Expedition Digital Collection site. They found that the two major behaviors users engaged in when submitting annotations were error correction (spelling, death date of individuals, etc.) and information sharing, such as in Fig. 3. They determined that user annotations offered two essential benefits to users: that through participation, users "Gain a sense of ownership and a vested interest in the site's continuation and improvement to address their information needs," and that the "Corrections benefit future users of the materials," because the annotation "stands for later visitors to consider" (299). Krause and Yakel also speculate about the potential utility of annotations for reference archivists, suggesting that they allow archivists to document the discoveries they make during the course of their work in a sustainable way that is accessible to them as well as users (300). This study seeks to expand upon Krause and Yakel's preliminary study by analyzing a larger set of annotations from the Polar Bear Expedition Digital Collections, the Keweenaw Digital Archives, and the Beyond Brown Paper project.

The literature that supports user annotations in archival description emphasizes how user-contributed content can fill in the limitations and gaps in archival description. If archivists cannot escape their role as mediators, they can at least allow other voices to contribute, thereby building a more complete understanding of a collection and its context, adding increasing detail in description, and cross-referencing related material, collectively illustrating and illuminating the historical record.

Methodology

Content analysis

This study investigated the types of communications happening in the form of annotations in current online archival collections. For the purpose of this study, a "user" is defined as anyone engaged in the use of a collection of archival materials. An "online archival collection" has been defined as a collection of archival materials that have been digitized and made accessible online by an archival institution. The unit of analysis for this study is one complete annotation contributed to any image or page on a digital collection site. An "annotation" has been defined as any user-contributed, publicly-viewable content added to an online archival collection at any level (collection, series, or item). Annotations were treated the same regardless of the author (the data were anonymized) and regardless of whether or not the annotation was part of a longer thread. A content analysis was conducted on a sample of existing, publicly contributed user annotations to online archival collections, considering user communication behaviors as the variable. This method was chosen as an unobtrusive research method that would yield valid analysis of communication behaviors (Babbie, 324). This method was chosen for its high degree of reliability given that the communication behaviors examined represent manifest content that, for the most part, can be clearly assigned to the various categories.

The set of categories used were initially developed based on the two types of communication behavior observed by Krause and Yakel, sharing further information and error correction, along with several other anticipated categories, which included questions, answers, establishing personal connections, linking to or referencing related resources, and general comments (this category includes opinionative and evaluative

statements; praise and thanks for the website; and other commentary which does not provide further information about the subject). Through the process of open-coding, several more categories were added iteratively as new patterns became apparent. These include translations of other non-English annotations, donation offers, requests for copies of materials, and annotations from archivists noting recent changes to metadata based on user contributions. See *Table 1* for an elaboration of the coding categories with examples.

Sample

For this study three online archival collections were selected from the total known population of collections that allow user annotations. As there is no source or list of online archival collections which allow user annotations, such collections were discovered through a combination of web searching, examining the relevant literature, and speaking with colleagues. Selection criteria for these collections included number, depth, and variety of annotations accrued, how long the digital collections have been up and running, and documentation surrounding the development and goals of the collections. The collections selected should also represent different institutions, in order to include distinct approaches and local user populations. After surveying the small identifiable number of online archival collections which allow user annotations, the following three collections were chosen: Polar Bear Expeditions Digital Collections, Beyond Brown Paper, and the Keweenaw Digital Archives.

As many annotations were collected from each site as possible. For the Keweenaw Digial Archives site, a spreadsheet was provided by the institution containing annotations, user ids, item titles, and links to the digital objects. For the Beyond Brown

Paper (BBP) project, a WGET function was utilized to extract a list of all items with annotations from the site, which were then hand-entered into a spreadsheet with fields that matched those provided by the Keweenaw Digital Archives (KEW). This particular site contained a number of annotations in French. These were not included in the final data set, however, see Table 1 for the total number of such annotations.

Table 1

Total Annotations, Annotations by Archivists, and Annotations in French

Site	Total Annotations Identified	Annotations by Archivists	Annotations in French
KEW	177	1	0
BBP	279	78	34
POL	112	28	0
Total	**568**	**107**	**34**

For the Polar Bear Expedition Digital Collections (POL), there was no discernable way to access the complete set of annotations on the site, and an advanced search for no character (leaving the search box blank) yielded 84 annotations as results. From there, further threads of annotations were followed to collect all additional identifiable annotations. See Table 1 for a breakdown of how many comments were extracted from each site, as well as the number of annotations from archivists and annotations in French. Archivist-contributed annotations were only counted as such when self-identified in the content of the annotation or in the user name provided, such as: “Susan, The Archivist.”. Across all three sites, there were a total of 107 comments from 6 distinct individuals identified as archivists working on their respective collections. Figure 4 below shows the percentage of annotations from each site in the total sample. It should be noted that about half of the annotations collected were from the Beyond Brown Paper digital collection.

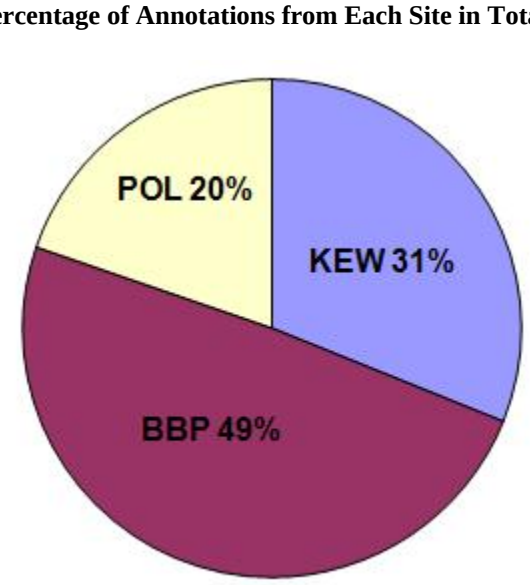

Fig. 4. Percentage of Annotations From Each Site in Total Sample

Codebook

The following codebook (Table 2) was developed to define the categories assigned to each of the annotations. The categories were developed based on types of communication that were anticipated, and were iteratively revised to reflect any unanticipated categories as they arose. The final codebook is displayed below.

Table 2

Codebook

Category	Description	Example
identification	identification of subject (person, place/building, thing) not already identified in metadata	I believe this is the National Lutheran church on Eighth Street in Calumet. http://digarch.lib.mtu.edu/showbib.aspx?bib_id=603553#
correction	correction to existing archival metadata	The correct spelling is Gregoire. http://digarch.lib.mtu.edu/showbib.aspx?bib_id=595221

further information	further information about subject such as date, explanation of contextual information	This locomotive was later sold to the Quincy & Torch Lake Railroad (Quincy Mining Co.) and it became their #5. http://digarch.lib.mtu.edu/showbib.aspx?bib_id=605334
link to additional resources	links and references to resources for further information, including websites, books, and email addresses of those with personal knowledge to offer. includes citations to original or other sources of material (see example)	I have pictures of the Daughter's at that final convention if anyone is interested I would be happy to share. [email@address.com] http://digarch.lib.mtu.edu/showbib.aspx?bib_id=595319# This is actually West Genesee Street in Iron River. (cf. Superior View website, http://www.viewsofthepast.com, under Images, Michigan Cities and Towns, under "I", "TO-IRNR 05 Genesee Street 1920.") http://digarch.lib.mtu.edu/showbib.aspx?bib_id=606569# The church is identified as the Methodist Church in Donald Chaput's book "The Cliff". http://digarch.lib.mtu.edu/showbib.aspx?bib_id=633345#
answer	answers to questions posed by other users (back-and-forth communication)	Hello Nancy: A listing of graves in the Liminga on the Finnish Genealogy site at http://www.genealogia.fi/haudat/index4e.htm They show a Selma Mastola (1903-1939) buried in plot C 17 http://digarch.lib.mtu.edu/showbib.aspx?bib_id=595306#
comment	general comments (non-informational), opinions, and praise	I particularly like this photograph. The image demonstrates the totality of the 1913 strike in the sense that it involved all elements of the community. http://digarch.lib.mtu.edu/showbib.aspx?bib_id=602307
personal connection	establishing personal connection to subject, such as noting a family member, providing an anecdote, etc.	This actual keg is in my mom's house in Oskar Bay. At the time that the Bosch plant closed, my parents were running Schmidt's Corner Bar. Somehow this keg ended up in their possession, and we still have it. http://digarch.lib.mtu.edu/showbib.aspx?bib_id=621414#
question	asking questions	I would like to know if a Roy C. Overholt was in this National Guard division. If he was, what was his rank, and his duties ???? Thank you Lester Stark http://digarch.lib.mtu.edu/showbib.aspx?bib_id=600471
donation offer	offering to donate further materials	I have about a dozen photos he took while stationed in Russia of many of the men that were stationed with him there. May I somehow send you them by email for your site. http://polarbears.si.umich.edu/index.pl?node_id=1251&lastnode_id=18283
order request	request for copies of materials	My great great-grandfather (J.R. Toothaker) from Rangeley owned over 1800 acres north west of Rangeley and operated a logging camp from about 1870 through 1900. If possible I'd love to have a copy of these Rangeley logging photographs. Is that possible and what would be the cost. http://beyondbrownpaper.plymouth.edu/item/198
translation	translation provided for non-english annotations	Thank you for the above correction! The edit has been added under the "person" category. http://beyondbrownpaper.plymouth.edu/item/7411
edit	edit or addition made to metadata based on user contribution	Thanks! We will add some of those changes to the database. http://polarbears.si.umich.edu/index.pl?node_id=3038&lastnode_id=7260

Because any one annotation may equally represent more than one category, each annotation was counted toward every applicable category in order to most accurately assess which types of communications occur most often. For example, an annotation

might identify an individual in a photograph, mention that the individual was her grandfather, and ask if anyone else knew where to find further information. In this case, the annotation would be counted toward the identification, personal connection, and question categories.

Limitations

This study sought to identify archivist-contributed annotations within the sample; however, as noted earlier, it was not always possible to make a definitive determination of whether a given annotation was contributed by an archivist or an external user. Similarly, it was difficult to accurately record back-and-forth communication, such as in the case of coding questions and answers. This is because data was extracted from each of the sites and recorded in a spreadsheet, removed from the original context of the annotation. Such communications were thus identified through analysis of the annotation content itself, or through the comment title information that was recorded (for instance, a comment title may begin with “Re:” indicating a response to a previous comment).

As described previously, best efforts were made to extract all comments from each of the sites; however, for the Polar Bear Expedition Digital Collections, a complete set of annotations from the date extracted cannot be guaranteed. Finally, no tests were done to check inter-coder reliability, so the reliability of the results depends on the consistency of coding judgments made throughout.

Findings

A total of 568 annotations were collected from the three sites. Non-English comments, all of which were in French, were not included among the other coded data;

however, these were counted separately as displayed previously in Table 1. The total number of annotations from all three sites counted toward each category is shown in Table 3 below.

Table 3

Total Annotations per Communication Behavior Category

Communication Behavior	Total Annotations
Identification	321
Further information	297
Link to further resources	116
Establishing personal connection	84
General/opinionative comment	68
Correction	66
Question	42
Answer	36
Donation offer	9
Translation of non-english annotation	8
Edit to existing metadata	3
Complaint	3
Request for copy of item	2

Generally, the results of the analysis show users engaging primarily in contributory communication behaviors, such as providing further information about the subject, identifying subjects, and providing links to further resources. There was also a high number of instances of users establishing personal connections to the materials. Somewhat less common communication behaviors include general or opinionative statements, correcting existing metadata, and asking or answering questions. Few instances were counted of users offering donations, requesting copies of materials, complaining about the site, or translating other non-English annotations. The relatively high number of annotations that established personal connections with image subjects possibly suggests that many of the participative users among the communities are

genealogists or connected to the subjects and creators in some other way. It is notable that no instances of spam were detected among the comments, which would have been identified by a lack of relevancy in content to the subject or site, and the presence of unrelated external links. While this bodes well for the successful management of an annotation system for archivists, it is not known how extensively the three sites currently manage and monitor annotation creation. It may be the case that quite a bit of maintenance is done to keep spam and vandalism at bay.

Figure 5 on the following page illustrates the total number of instances for each category across sites. It is important to note that the total number of comments collected from each site varies, and so differences in data set sizes from each site may be reflected in the cross-site comparison. Beyond Brown Paper (BBP) shows more extreme spikes in the data than the Polar Bear Expedition Digital Collections (POL), which has a much more even number of annotations representing the various categories. The Keweenaw Digital Archives (KEW) shows a prominent spike in the Identification, Further Information, and Correction categories, with the numbers dwindling across the remaining categories. Very few instances of edits being made to the metadata based on user contributions were identified (only 1 for BBP and 2 for POL); however, this can only account for edits that are acknowledged in annotation form, and cannot tell us how many edits are being made to metadata without an accompanying annotation.

User name data was also collected; however, due to the anonymity allowed on each of these sites, data was inconsistent and widely varied. Still, it is interesting to note that 116 of the 568 annotations (about 20%) collected were from the user name

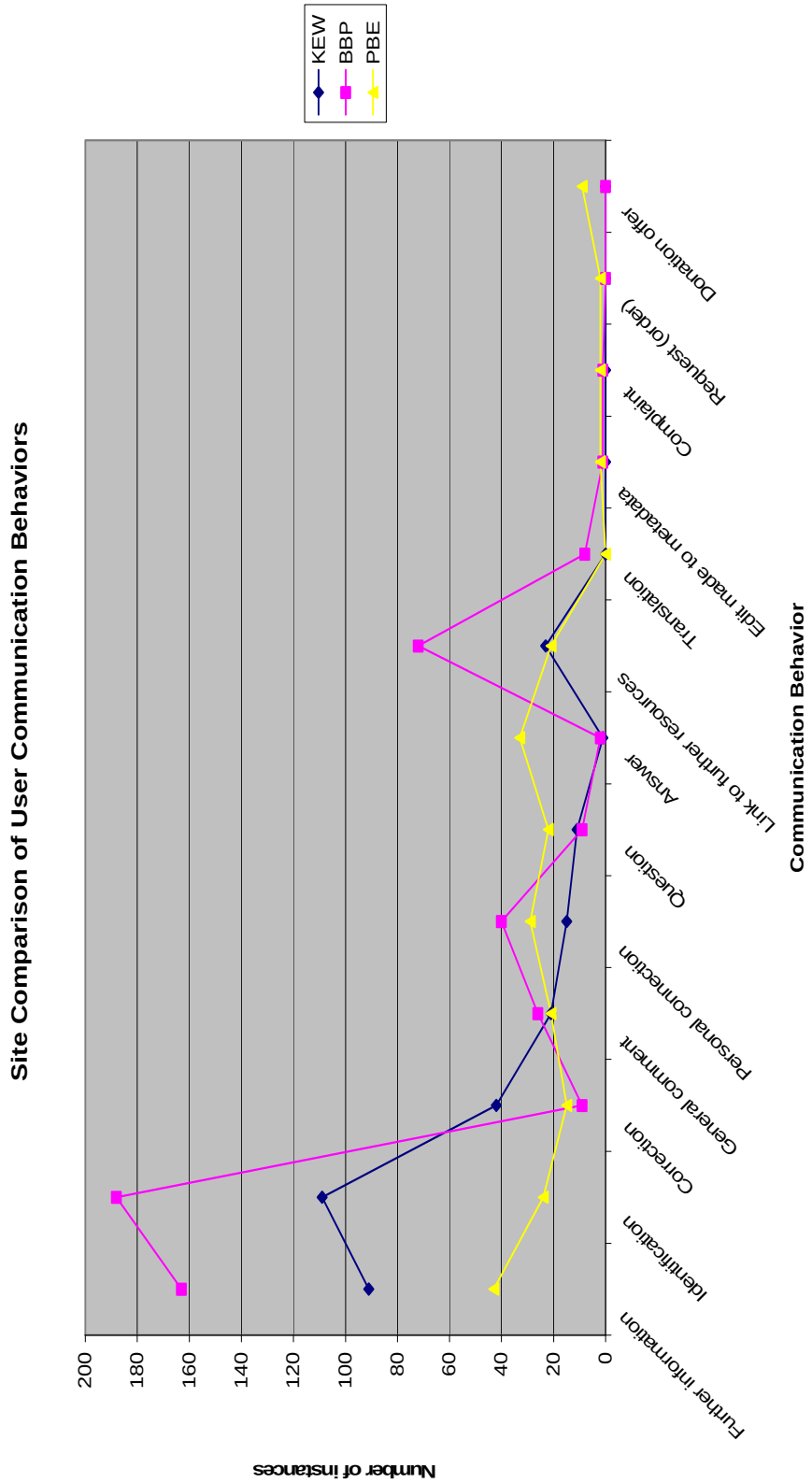

Fig. 5. Site Comparison of User Communication Behaviors

"anonymous," while several frequent annotators were identified. The highest numbers of annotations from single users were 32, 25, and 12.

Keweenaw Digital Archives (KEW) Results

Results for KEW show a strong tendency toward identification, correction, and providing further information about subjects in user communication behaviors. Other notable behaviors include providing links to related resources, general comments, and establishing personal connections, with only a few questions and one answer. This site had very little archivist presence, with only one comment from an identified archivist. This could be the reason for the comparatively low number of questions-and-answer behaviors on the site.

Table 5

Total Annotations for Each Category on KEW

Communication Behavior	Counts	Communication Behavior	Counts
Further information	91	Link to further resources	23
Identification	109	Translation	0
Correction	42	Edit made to metadata	0
General comment	21	Complaint	0
Personal connection	15	Request (order)	0
Question	11	Donation offer	0
Answer	1		

KEW by far had the most archivist-created description of the three sites, providing metadata for 19 distinct fields for every object. However, this site also had the highest number of corrections among the three sites, suggesting that perhaps the more metadata is provided, the more opportunity there is for users to correct that information.

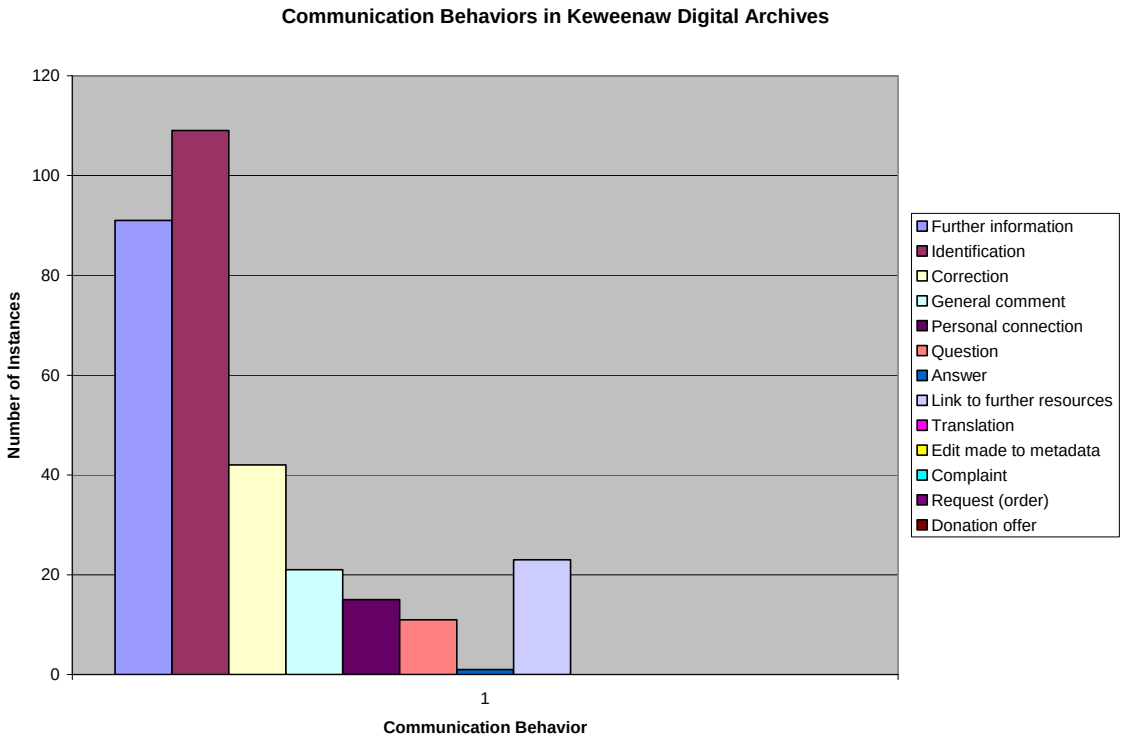

Fig. 6. Communication Behaviors in Keweenaw Digital Archives

Beyond Brown Paper (BBP) Results

The results for the BBP site show an even stronger tendency toward identification and providing further information than KEW, with similar distributions for links to further resources, personal connections, and general comments. This was by far the largest dataset; the annotations from BBP made up 49% of the total sample, showing a very active user base. Furthermore, two of the three most active annotators (based on user ids) were users of BBP.

Table 6

Total Annotations for Each Category on BBP

Communication Behavior	Counts	Communication Behavior	Counts
Further information	163	Link to further resources	72
Identification	188	Translation	8
Correction	9	Edit made to metadata	1
General comment	26	Complaint	1
Personal connection	40	Request (order)	0
Question	9	Donation offer	0
Answer	2		

One surprising finding in this result set is the relatively low number of metadata corrections, especially considering the high frequency of identification, which on might assume is a closely related behavior. This is probably because BBP offers considerably less archivist-provided metadata for objects, especially compared to KEW. So, it appears that the user response tends toward identification of subjects more often when subjects are not already identified in the metadata. Another interesting finding is the use of the annotation space to provide translations of non-English annotations. On BBP, 34 of the 279 annotations are in French, and it appears that archivists related to the project have utilized the annotations as a space for translating these annotations, which frequently provided identification and further information about subjects. Of the 34 French annotations on the site, 8 were translated by a user identified as an archivist through her user id.

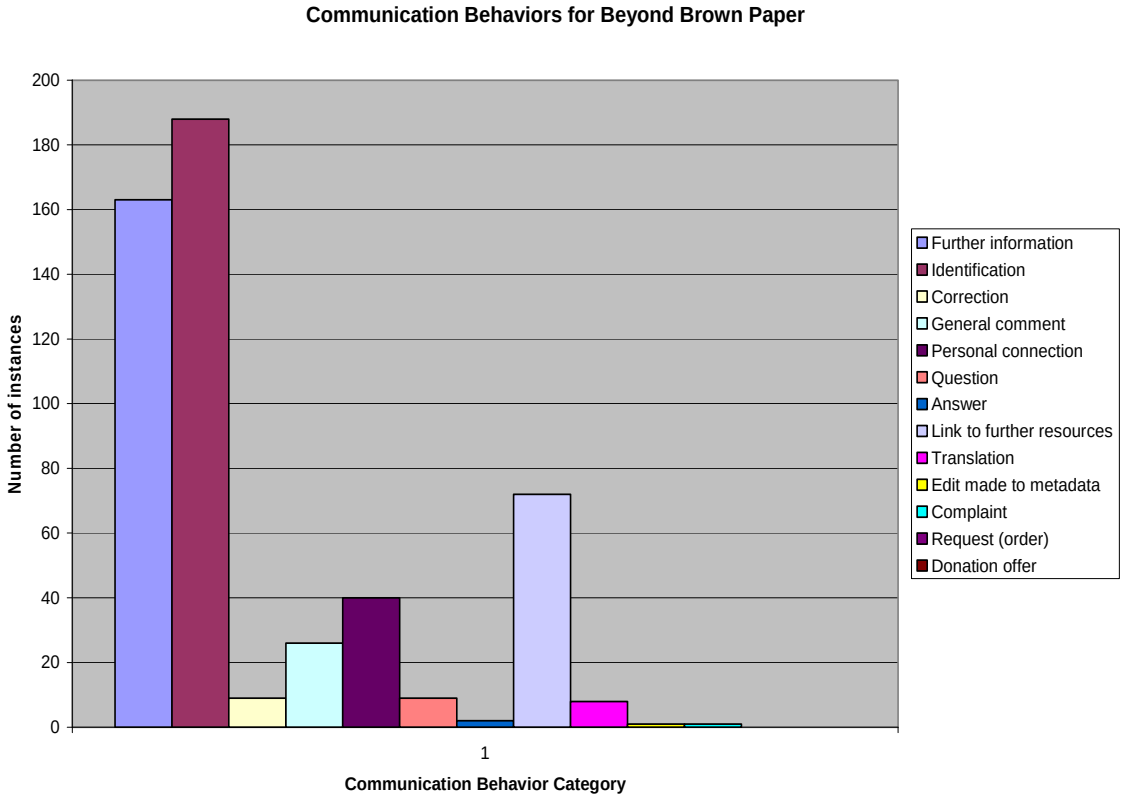

Fig. 7. Communication Behaviors in BBP

Polar Bear Expedition Digital Collections (POL) Results

There are several notable findings in the POL results in comparison to the findings of the other sites. There is a much more even distribution across categories for this site than for the other two. This is likely related to the fact that this project has put considerably more effort into creating a collaborative user community that includes archivists. POL employs the web presence of several archivists who encourage users to ask questions, make corrections, and share personal connections. The archivists respond pretty actively to user questions and requests, as shown in the comparatively high number of Question and Answer behaviors.

Table 7

Total Annotations for Each Category on POL

Communication Behavior	Counts	Communication Behavior	Counts
Further information	43	Link to further resources	21
Identification	24	Translation	0
Correction	15	Edit made to metadata	2
General comment	21	Complaint	2
Personal connection	29	Request (order)	2
Question	22	Donation offer	9
Answer	33		

This site also has annotations for categories that are missing from the other two sites, such as Order Requests and Donation Offers. It appears from the content of the archivists' reply annotations that a system of stock responses has been developed handle to such requests and offers. These stock responses encourage the user's interest in either donating or requesting items, and provide appropriate contact information for further discussion. Similar to the Question and Answer phenomenon, the higher presence of these types of communication on this site could be attributed to the strong archivist presence, which might invite these more varied types of annotations. While this model of involved annotation management would require the most resources from archivists, it appears to create a more varied and enriched style of communication among the user community.

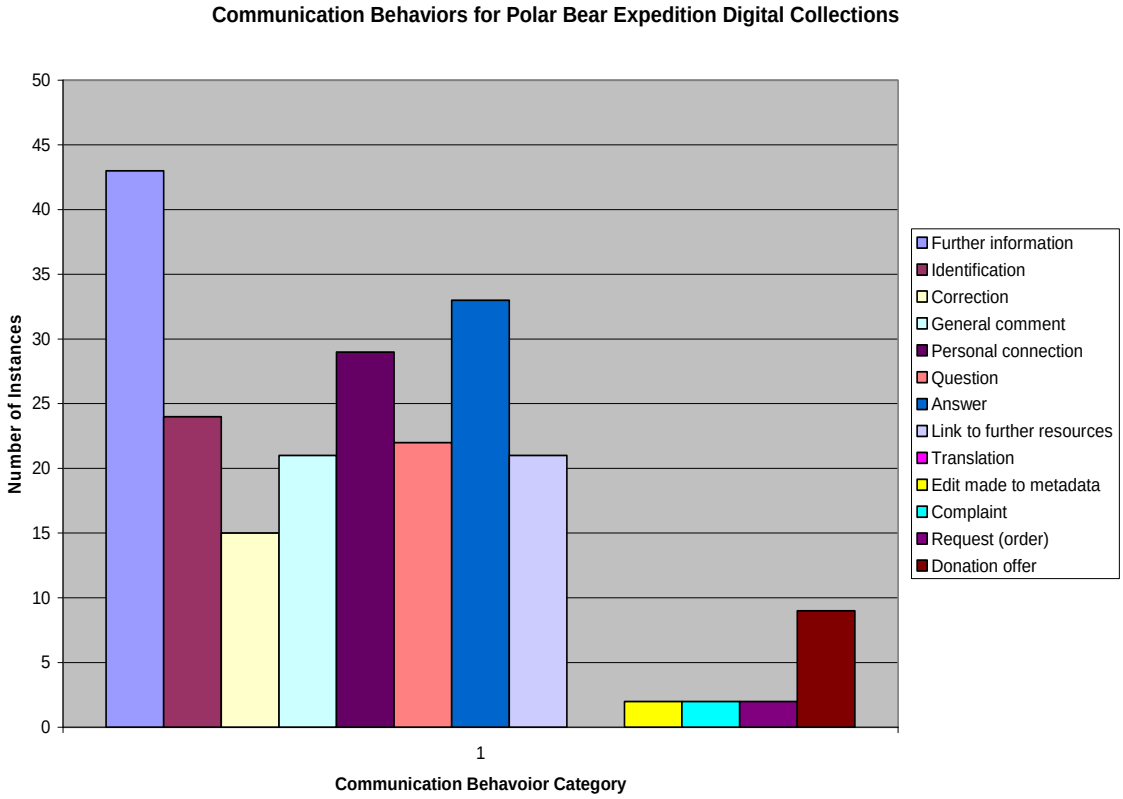

Fig. 8. Communication Behaviors in POL

Discussion

The findings of this study contribute to a better understanding of this topic by specifically exploring what users are doing when given the opportunity to contribute annotations to online archival collections. Since archivists may question the usefulness and manageability of annotations and other social software in archival description, and may be wary to dedicate their valuable resources to developing such experimental features, it is important to have empirical data about how the annotation space is used. This study augments the fledgling literature on the subject, adding to the field's general understanding and acceptance of these new technologies. Hopefully, the findings will allow archivists to become more comfortable with incorporating social software

technologies into their own archival description, using web technologies to their fullest advantage.

The incorporation of annotations has the potential to improve archival practices by making allowances for diverse voices and perspectives, but can also improve user satisfaction by making collections more accessible. However, this latter point hinges on whether the annotations are made searchable in the site's system. Annotations to the Keweenaw and Polar Bear digital collections are searchable, while the annotations to the Beyond Brown Paper collection are not. This means that for Beyond Brown Paper, all of the identifications, contextual information, and other types of information provided in annotations cannot function as access points, and are only of use when a patron discovers an item by other means and scrolls down to read the comments. The search model used for the Keweenaw and Polar Bear collections seems to offer far more utility, because those annotations help users retrieve search results that include items for which the archival metadata was lacking or incorrect. This highlights the importance of creating and maintaining distinctive display modes for user-created content and archival description. It should be entirely clear to users what description was created by archivists, and what content was created by users. Each of the three sites examined displayed user annotations at the bottom of the page, below the item display and archival description, in a way that is reflective of blog comment interfaces.

Conclusion

The findings of this study suggest that information-contributing (including identification, providing further information, and linking to additional resources) is the

most common type of user annotation. These findings support and expound upon the findings of Krause and Yakel's 2007 study, though interestingly, results for all three data sets, not just the results for the POL collection, reinforce this finding. Other common user communications include establishing personal connections to the materials and, to some extent, asking and answering questions (though the latter of these seems to depend somewhat on the site's approach to annotation management). Very few instances of donation-offering and copy-requesting were identified, and no spam was detected.

There is still much to learn through future research about user annotations and their place in archival description. One question left open following this study is, how does the amount of archivist-created metadata influence the amount and types of user annotations contributed. To some extent, the findings of this study suggest a correlation between more detailed archivist descriptions (as with KEW) and users offering corrections as annotations, while less detailed description (as with BBP) seems to incite more identification behavior. It would be particularly interesting to examine user annotations in an online archival collection community with only very high-level, aggregate description.

Another question is whether certain types of digitized materials lend themselves more to users participating and contributing with annotations. For example, do digitized images invoke more response than written documents? Do users annotate images of people more frequently than images of buildings, or objects?

It would also be very useful for studies to be done in the areas of usability and platform management, as this type of research would speak to concerns about how best to manage, display, index, and monitor user annotations. Given that the literature

demonstrates a user preference for the ability to annotate at various levels of documents, placing the annotation text as close to the relevant document section as possible, it would be interesting to see research in developing this type of platform. The challenge would be in developing a way to display annotations in context while clearly differentiating them from the original document and the archival description.

Finally, do user annotations have the same utility for archival finding aids as they do for collections of digital objects? It may be of some use for researchers to be able to annotate archival descriptions in online finding aids, but are they as likely to annotate descriptions without the objects right in front of them, as they are in the online environment?

This study demonstrates that, when given a venue of communication, in the case the annotation space, users most often contribute informational content, such as identification, further contextual information, and links to related resources. Further research is needed to help archivists understand how to best implement and manage annotations for their collections.

Works Cited

Babbie, E. *The practice of social research*. 10th ed. CA: Wadsworth, 2004.

Brown, Jesse. “Finding Aids: The Next Generation - Part I (SAA 2006).” Weblog entry. *Anarchists, Architects, Activists…* 7 August 2006. <http://www.novicearchivist.com/> Accessed 3 December 2007.

Cook, Terry. “Archival Science and Postmodernism: New Formulations for Old Concepts.” *Archival Science* 1 (2001), 3-24.

Craig, Barbara L. “Old Myths in New Clothes: Expectations of Archives Users,” *Archivaria* 45 (Spring 1998), p. 119-126.

Deodato, Joseph. “Becoming Responsible Mediators: The Application of Postmodern Perspectives to Archival Arrangement & Description.” *Progressive Librarian*, Summer 2006. 52-63.

Duff W.M. and Harris, V. “Stories and Names: Archival Description as Narrating Records and Constructing Meanings,” *Archival Science* 2/3-4 (2002), 263-285.

Greene, Mark A. and Dennis Meissner. “More Product, Less Process: Revamping33 Traditional Archival Processing.” *American Archivist* 68 (Fall/Winter 2005), 208-263.

Hammond, T., Hannay, T., Lund, B. and Scott, J. “Social Bookmarking Tools (1).” *D-Lib Magazine* April 2005. URL: <doi:10.1045/april2005-hammond>.

Hill, Amanda. "Serving the Invisible Researcher: Meeting the Needs of Online Researchers." *Journal of the Society of Archivists* 25, no. 2 (2004), 139-148.

Hyry, Tom and Michelle Light. “Colophons and Annotations: New Directions for the Finding Aid.” *American Archivist* 65 (Fall/Winter 2002), 216-230.

Krause, Magia Ghetu and Elizabeth Yakel. “Interaction in virtual Archives: The Polar Bear Expedition Digital Collections Next Generation Finding Aid.” American Archivist 70 (Fall/Winter 2007) 282-314.

Marshall, Catherine C. “Toward an ecology of hypertext annotation,” *Proceedings of the ninth ACM conference on Hypertext and hypermedia: links, objects, time and space---structure in hypermedia systems: links, objects, time and space---structure in hypermedia systems*, p.40-49, June 20-24, 1998, Pittsburgh, Pennsylvania, United States

Middleton, Michael and Julie Lee. “Cultural Institutions and Web 2.0.” November 2007. <http://eprints.qut.edu.au/archive/00010808/01/Cultural_Institutions_and_Web_2_0.pdf> Accessed 3 December 2007.

Roscheisen, Martin, Christian Mogensen, Terry Winograd. “Shared Web Annotations as a Platform for Third-Party Value-Added, Information Providers: Architecture, Protocols, and Usage Examples.” Stanford Integrated Digital Library Project, Computer Science Dept., Stanford University, California. 1997.

Tibbo, H.R. (2003). “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.

Ulrich, Thiel. “COLLATE – A collaboratory supporting research on historic European films.” *International Journal on Digital Libraries* 4.1 (2004): 8-12.

Van Garderen, Peter. “Web 2.0 and Archival Institutions.” Weblog entry. Archivematica, 8 May 2006. <http://archivemati.ca/> Accessed 3 December 2007.

Wilensky, Robert. “Digital library resources as a basis for collaborative work.” *Journal of the American Society for Information Science* 51.3 (2000): 228-245.

Yakel, Beth, Seth Shaw, and Polly Reynolds. "Creating the Next Generation of Archival Finding Aids." *D-Lib Magazine* 13, no. 5/6 (2007): http://www.dlib.org/dlib/may07/yakel/05yakel.html.