Zachery P Whitaker. Moving With the Crowd: Gauging the Prevalence and Execution of Crowdsourcing in Archives. A Master's Paper for the M.S. in L.S. degree. July, 2014. 50 pages. Advisor: Denise Anthony

This study describes an online survey sent to the Society of American Archivists' Archives and Archivists listserv to illicit responses from archivists regarding the use of crowdsourcing in archives. The survey was conducted to determine the prevalence of crowdsourcing in the archives profession. In addition, the survey sought to gain insight into the level of interest in crowdsourcing among archivists as well as the methods by which archivists have executed crowdsourcing projects.

Slightly over half of archivists surveyed indicated they have either engaged in, are currently engaging in, or have plans to engage in crowdsourcing, while seventy-nine percent of archivists who identified as having no plans to engage in crowdsourcing indicated they were slightly to very interested in the subject. In addition, the survey found that archivists execute crowdsourcing in myriad ways and often employ several technologies and competencies simultaneously.

Headings:

Crowdsourcing

Archives -- Public relations

Archives -- Social aspects

Archives users

MOVING WITH THE CROWD: GAUGING THE PREVELANCE AND EXECUTION OF CROWDSOURCING IN ARCHIVES

by
Zachery P Whitaker

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 2014

Approved by

_____

Denise Anthony

Table of Contents

Introduction

In his *Manual of Archive Administration*, famed British archivist Hilary Jenkinson (1922) declared that in addition to providing for the safeguarding of the Archives in his/her custody, an archivist's job was to "provide to the best of his ability for the needs of historians and other research workers" (p. 15). While Jenkinson may have viewed the process by which archivists ensure access to archival materials as secondary to ensuring their preservation, there can be little argument that accessibility and usability of archival material has assumed a much greater role and emphasis in the archival profession since Jenkinson published his foundational book in 1922.

There have been many societal changes that have led to the increased impetus on accessibility for archival material, the first, and perhaps most fundamental, being the rise of the personal computer. Further, the invention of the World Wide Web has fundamentally changed the way people interact with information. The Institute for Prospective Technological Studies asserts that affordable, widespread broadband access to the World Wide Web has turned the personal computer (and subsequently the PDA, telephone, and the mp3 into the "ultimate collaborative device" (Huijboom, van den Broek, Frissen, Kool, Kotterink, Nielsen & Millard, 2009, p. 30).

It is this environment that has given rise to social computing. Social computing is a term used to describe online communities where users interact with each other by creating and

sharing information. Social computing has changed the way in which people expect to gain access to information. It has fostered a collaborative environment where people expect instant access to information and also that their voices be heard in regards to the creation of new information. Studies have shown that "a new social structure is emerging in which technology puts power in communities, not institutions" (Huijboom et al., 2009, p. 30). Online communities are having an effect on the missions of institutions because social computing technologies are resulting in a shift in the balance of power between the two, essentially taking the power away from formal institutions and putting it in the hands of online communities (Huijboom et al., 2009, p. 31). The notion of citizen empowerment that has arisen from ready access to information and social computing has had a tremendous effect on the way people view the centrality of the individual and the need for increased access to information, transparency of information and decision making, and the ability of individuals to communicate directly with the disseminators of information (Huijboom et al., 2009, p. 22).

Postmodern archival theory reflects this fundamental change brought on by social computing in the way archivists view traditional top-down authority structures. Postmodern archival theorists claim that traditional archival description is inadequate, and that archivists should explore the ways in which user contributions can enhance archival description. For example, the late Terry Cook asserted that the archival record is not static, but rather a mediated, ever-evolving construction influenced by its use, interpretation, and reinterpretation (Cook, 2001). Postmodern archival theory recognizes that information has traditionally flowed outward from the archivist and institution to the researcher and suggests that archivists and institutions should seek to understand the

ways in which researchers interact with material in order to enhance the meaning of their collections (Krause & Yakel, 2007, p. 289). Wendy Duff and Verne Harris (2002) argue that archivists and institutions need to "create holes that allow in the voices of… users" (p. 279). By extension, postmodernist archival theory rejects the notion that digitization is a means to an end in terms of providing access to material. Rather, digitization is a first step in the process of remediation and user interaction that has the potential to create new meanings and data related to archival collections in the digital format. Therefore, user interaction with digital materials has the potential to actually create new data and new experiences for the users of archives (Vershbow, 2013, p. 81).

In recent years, there has been an emphasis on innovation in the archives community to take advantage of new digital possibilities. Archival institutions are under ever-increasing pressure to develop processes more effectively and efficiently and to ensure a return on investments (Huijboom et al., 2009, p. 24). This drive in innovation has often occurred in tandem with shrinking budgets for many archival institutions. Archivists have had to consider new ways in which to connect with today's researchers, while simultaneously adding value to, and demonstrating the value of, their collections. Many archivists argue that social computing technologies hold great potential to provide opportunities for future services, and that the rise in cheap computers and high-speed Internet access has created opportunities to break down and distribute complex work processes in order to add value to, and increase engagement with, archival collections (Huijboom et al., 2009, p. 49; Chrons & Sundell, 2011, p. 20). In addition, the burden has increasingly fallen upon archives to draw attention to collections, in order to prove their viability. Consequently,

over the past several years archives have experimented with crowdsourcing as a way to simultaneously provide increased access

and value to their collections, while promoting the increased use of their collections.

Many have celebrated crowdsourcing's ability to democratize the process of archival description and knowledge production while lowering the costs of such production and increasing awareness of archival collections. However, detractors of crowdsourcing point to institutions exploiting contributors for free labor, while others point to concerns over the lack of archival authority and commodification of culture (Woods, 2009). Commodification of culture refers to the notion of placing extreme value on the immediate access of materials, as opposed to future accessibility through authority control (Van Hooland, Mendez Rodriguez & Boydens, 2011, p. 709). Museum exhibition designer Nina Simon notes that “Many museums fear losing control… but there’s a difference between having power and having expertise… museums will always have the expertise, but they may have to be willing to share the power" (Wright, 2010). This quote suggests there may be archivists who bristle at the notion of sharing their power with the crowd to provide information for collections.

In spite of the concerns of a few archivists, the literature regarding crowdsourcing is overwhelmingly positive. There are many stories of successful crowdsourcing projects undertaken by archives all over the world. Often, these stories extoll the virtues of crowdsourcing, claiming that these initiatives have resulted in added value to collections by accomplishing tasks that institutions would not have the resources to complete otherwise. In addition, many articles and blogs praise crowdsourcing efforts for creating

connections with archival users, increasing the use of collections and notoriety for the institutions taking part in the various projects. But when we examine the archival literature on crowdsourcing are we only seeing the success stories? Do the real-life experiences of archivists who have engaged in crowdsourcing projects match the overwhelmingly celebratory literature? This study also seeks to shed light on the extent to which crowdsourcing has penetrated the archives profession. Do crowdsourcing projects undertaken by archives institutions signify an emerging practice with staying power or just a passing trend? To provide answers to these questions, I conducted a study to gather information about archivists' experiences with crowdsourcing, the design and results of which are discussed in this paper.

Background

The term "crowdsourcing" has its origins in a 2006 article written by Jeff Howe, titled The Rise of Crowdsourcing (Howe, 2006). The Oxford English Dictionary defines crowdsourcing as "The practice of obtaining information or services by soliciting input from a large number of people, typically via the Internet and often without offering compensation" (as cited in Ellis, 2014, p. 1). Rose Holley, of the National Library of Australia and an expert in the field of crowdsourcing in the library community, makes a distinction between social engagement and crowdsourcing. She defines social engagement as giving the public the ability to communicate with the professional and each other, and includes photo tagging, commenting, and giving ratings to resources in this category. Holley's definition of crowdsourcing entails using social engagement techniques to facilitate a group of people in achieving a shared and usually large goal; it

entails a greater level of time and intellectual input from an individual than simply socially engaging (Holley, 2010, p. 2). Brabham (2012) defines crowdsourcing as "an online, distributed problem solving and production model whereby an organization leverages the collective intelligence of an online community for a specific purpose" (p. 395). Simply put, all crowdsourcing involves an organization issuing a task to an online community that participates in a task for the benefit of the organization.

Crowdsourcing Transcription

Crowdsourcing efforts in archives have generally taken on a couple forms, one of them being transcription. Many institutions have undertaken efforts to harness the power of the crowd in order to provide transcription services for digitized manuscripts or newspapers. The primary motivation behind many projects calling on crowds to transcribe digitized documents lies in the fact that Optical Character Recognition software (OCR), while adequate for modern typesetting, does not typically fare well with handwritten documents or older documents with odd fonts or for poorly scanned documents (Chrons & Sundell, 2011, p. 20). Consequently, there have been several instances of institutions engaging in processes designed to augment OCR with human computation, including the National Library of Australia's Trove project, the Improving Access to Text (IMPACT) project, and the reCAPTCHA project (Chrons & Sundell, 2011, p. 20).

Trove is a multi-dimensional platform that is used by the National Library of Australia as a searchable digital repository, a metadata aggregator, and a website developer, among other things. In addition, Trove is widely used for correcting OCR text of digitized newspapers. For example, on July 10, 2014, users made 82,571 corrections ("Trove," n.d.).

IMPACT, a project funded by the European Commission, relies on user contributions to correct text that has undergone the OCR process as part of its program to develop OCR tools to enhance transcriptions of digitized versions of texts created during the period from the advent of the Gutenberg press to the advent of standardized industrial printing processes ("IMPACT," n.d.).

Google's reCAPTCHA project took the approach of correcting faulty OCR results of digitized text in a word-by-word manner by creating a method in which users of various websites are provided with a word as a CAPTCHA. A CAPTCHA is verification process where a word that has been distorted to prevent spam is presented to a user. The user then deciphers the word and retypes it in a text box. While this verification process allows administrators to protect websites from automated software, the words copied by users come from actual digitized text in need of transcription ("Google reCAPTCHA," n.d.).

Similar to reCAPTCHA is a project from the National Library of Finland called Digitalkoot. This project relies on the use of CAPTCHA technology and simple games to solicit user contributions for the correction of OCR text results on a word-by-word basis for digitized newspapers from the late 19th century using the Fraktur typeface. The use of games for transcription, or gamification, is a method in which a mundane task can be enhanced to provide motivation to participants. Gamification works best when users are provided with scores, instant feedback, and a sense that what they are doing contributes to the common good (Chrons & Sundell, 2011, p. 20). In the case of Digitalkoot, users are presented with words which they type back. This process is augmented with games such as Whack-a-Mole, where for each word a user gets correct, a mole is hit with a mallet and the user's score increases. In order to verify the accuracy of tasks, the same

words in Digitalkoot are presented to players simultaneously and the system compares the results. The accuracy verification process used by Digitalkoot works best when a significant number of players participate simultaneously, which is due to latency derived from the act of comparing transcription results. This latency grows when not enough results are available for comparison. To remedy this problem, the developers of Digitalkoot sought to design a system which automatically adjusts to varying levels of participation (Chrons & Sundell, 2011, p. 21). This method has proven to be successful for the Digitalkoot project, as evidenced by a 99% accuracy rate (Chrons & Sundell, 2011, p. 23).

Another notable transcription effort that harnesses the power of the crowd is University College of London's Transcribe Bentham Project. The Transcribe Bentham project presents contributors with digitized images of Jeremy Bentham's unpublished manuscripts which they then transcribe using a markup tool known as Transcription Desk. The final transcriptions are then published in the *Collected Works* of Jeremy Bentham, an ongoing publication of Bentham's manuscripts begun in 1958 (Causer, Tonra & Wallace, 2012, p. 120). Transcription Desk converts transcribed manuscripts into XML markup text. These transcribed texts then undergo a rigorous editing process before they are removed from the site and included in the *Collected Works*. The Transcribe Bentham project began in September, 2010, and is still ongoing (Causer et al., 2012, p. 120). As of June 12, 2014, users have contributed a total of 8,185 complete and verified transcriptions ("Transcribe Bentham," n.d.).

Transcription crowdsourcing projects do not need to rely heavily on computer programming skills and programs developed in-house in order to be executed, however.

The University of Louisville proved this with their project aimed at transcribing the *Louisville Leader*, an African-American newspaper that was produced from 1917 to 1950 (Daniels, Holtze, Howard & Kuehn, 2014, p. 38). Staff of the University Archives and Special Collections at the University of Louisville Libraries used Scripto, an out-of-the-box transcription tool, Omeka to build the digital exhibit, and CONTENTdm to collect and store transcribed newspapers. The latter two tools are considered Content Management Systems (CMS). Omeka was used as a functional interface for the front-end of the exhibit, while CONTENTdm was used for the back-end task of storing the data. Project staff decided not to expend time editing the transcribed newspapers and was pleased with the results of their project; only one transcription was found to be grossly inaccurate (Daniels et al., 2014, p. 46).

Crowdsourcing Photographs

Another popular task that archival institutions have used crowdsourcing for is the identification, or tagging, of photographs. Tagging involves users creating subject terms for images that are used for indexing and increased online discovery. Tags may be related to the subject matter of the image or to the people or places found in the image. Tagging of images requires relatively little effort on the part of contributors when compared to transcription.

There are numerous examples of crowdsourcing projects related to photographic image tagging and commenting in archives, one in particular being the Library of Congress' joint project with Flickr launched on January 16, 2008. Flickr is a social media site where users can upload their photos for viewing, commenting, and tagging by others. The Library of Congress posted non-copyright images in their collections to the Flickr site in

order that users may engage with the photos by commenting on them. The aim of the Library of Congress' project is to ensure better access to collections and to ensure that the best possible information about collections will be collected and preserved (Raymond, 2008). During the first nine months of the project there were 10.4 million views of the photos posted on Flickr; 7,166 comments were left on 2,873 images by 2,562 users; 67,176 tags were created by 2,518 users; and less than 25 instances of inappropriate content were removed from the site (Springer, Dulabahn, Michel, Natanson, Reser, Woodward & Zinkham, 2008, p. 4).

Several institutions have followed the Library of Congress' lead by spearheading crowdsourcing efforts through the use of Flickr in the hopes of providing more robust information for their photographic collections. One example is Virginia Commonwealth University's Freedom Now Project. The Freedom Now Project involves images taken by a police photographer during protests revolving around Prince Edward County, Virginia's decision to close public schools rather than integrate in 1963. The project has used crowdsourcing as a method to gather more information about the people involved in the protests as well as the protests themselves. The target group for this project has been residents of Prince Edward County over the age of 65. Perhaps because people over the age of 65 are less likely to be familiar with social media than their younger peers, many of the project's participants felt more comfortable providing information directly to the archivist than through the Flickr site (McNeill, 2014). This occurrence shows that the targeted audience of a particular project may influence crowd participation.

Participation

The percentage of contributors on a crowdsourcing project depends largely on the aim of the project, the targeted community, and the required level of skills needed for contributors to complete the task (Huijboom et al., 2009, p. 36). The predominate method of user participation in the Freedom Now Project confirms that a project's targeted community has an effect on the amount of online work generated by users. User skill and difficulty of the crowdsourcing task also plays a role in the amount of online participation a project receives. The Transcribe Bentham project compiled a user survey and found that many people who created user profiles found the transcription tasks rather complicated due either to extensive instructions, problems identifying untranscribed material, or difficulties using the Transcription Desk interface. The survey found that these barriers often kept people from contributing to the project (Causer et al., 2012, p. 127).

Crowdsourcing projects may also suffer from a problem of critical mass. When there are not enough people participating in a crowdsourcing project there may be difficulty in seeing any effects of the presence of others (Krause & Yakel, 2007, p. 294). "According to Technorati, 'all large-scale, multi-user communities and online social networks that rely on users to contribute content or build services share one property: most users don’t participate very much'" (Huijboom et al., 2009, p. 35). Studies have found that many users simply lurk in the background while a small minority does the majority of the work. This is evidenced by the "90-9-1 Rule." According to Nielsen, 90% of all users lurk in the background, while 9% of users contribute sporadically and may have other priorities. The final 1% of participants account for the majority of contributions to crowdsourcing projects (as cited in Huijboom et al., 2009, p. 35).

A critique of the "90-9-1 Rule" is provided by Mijke Slot at the COST Conference in Copenhagen, Denmark in 2009. Slot found that while passive activities, or those that simply consume information, exist online, there is evidence that a large number of users actively create content. Through the use of a survey, Slot found that 38% of respondents reported having a website, 27% reported they created a weblog, over 15% reported to writing news messages, and 3.5% reported to uploading a podcast at least once a year. While there may be a possibility that Slot's study attracted a preponderance of respondents who are active online, it raises important questions regarding the absoluteness of the "90-9-1 Rule" (Slot, 2009).

The statistics of participation among contributors of the Transcribe Bentham project seem to be in line with the "90-9-1 Rule." Among all users who created an account between September 8, 2010 and March 8, 2011, only 21% did any transcription. Of these users, around two-thirds worked on a single manuscript, whereas over a quarter transcribed between two and five manuscripts. The seven most active volunteers, or 0.6% of all registered users, produced a total of 709 transcripts. This small minority of users accounted for 70% of all transcribed manuscripts (Causer et al., 2012, p. 126).

Marketing

Participation in crowdsourcing projects often depends on how aggressively projects are marketed. In order for people to participate they must know the project exists in the first place. To quote the popular movie *Field of Dreams*, the "If you build it, they will come" philosophy does not typically lead to successful crowdsourcing projects in archives (Frankish, B. & Robinson, P., 1989). The idea that crowdsourcing is simply a free source

of labor is erroneous. Institutions must take the time to publicize their crowdsourcing projects, which usually entails a considerable effort.

Critical mass is an essential element in the success of any crowdsourcing project and marketing plays a key role in achieving it. Projects such as Wikipedia, Distributed Proofreaders, FamilySearchIndexing, and the Australian Newspapers Digitisation Program all launched with little fanfare. Each project had fewer than 4,000 volunteers in their first year. Through word of mouth and viral marketing, such as blogs, forums, and email, participation in these projects rose dramatically in subsequent years (Holley, 2010, p. 8). In contrast, the University of Michigan's Polar Bear Expedition project, while marketed locally through the library's blog and by adding links to Wikipedia articles about the expedition and to various articles related to World War I, experienced a lack of critical mass. Successful projects are those that have had long-range, aggressive marketing goals (Krause & Yakel, 2007, p. 287). Several successful projects have also received attention from the press on a national or international scale. While it may be unclear in some cases whether a project's success is due to national or international press exposure or if that exposure is simply a byproduct of success, the Transcribe Bentham project attributes much of their participation to such exposure. The marketing strategies behind the University of Louisville's *Louisville Leader* transcription project, University College London's Transcribe Bentham project, and the National Library of Finland's Digitalkoot project provide examples that illustrate the ways in which marketing can positively affect the success of crowdsourcing projects in archives.

University of Louisville and the Louisville Leader

The University of Louisville took an aggressive, long-term approach in marketing their project aimed at transcribing issues of the *Louisville Leader*. Before the project was launched, the Director of Archives and Special Collections met with a liaison in the University of Louisville's Office of Communication and Marketing to discuss methods of marketing the project and decided to create a press release and run an article in the school's newspaper, *U of L Today*. The project group then linked the project to the collection's CONTENTdm homepage, as it was thought that this would provide the best access to the public and because it also described the historical relevance of the newspaper. February 12, 2013, marked the publishing of the article in *U of L Today* and the post to the University of Louisville Libraries blog. Press releases were also sent out to local media outlets. In addition, the project group created social media posts for both the libraries' Facebook and Twitter pages. A local radio station interviewed the Director of Archives and Special Collections on the day of the launch. This aggressive marketing strategy netted the first uptake of transcription statistics, with 37 article sections being transcribed in the first two days. Participation in the project was boosted once more when a local television station aired a piece on the project along with an accompanying web page story on February 19, 2013. The day of the story and the day that followed, participants transcribed an additional 68 article sections (Daniels et al., 2014, p. 44).

The project team did not stop with the aforementioned marketing efforts. Several potential interest groups were identified and project announcements were sent to these groups over a period of time. For the remainder of February an average of 16 transcriptions were completed per day. The project continued to experience jumps in

completed transcriptions as new releases were sent out. On March 1, 2013, an email was sent to the university's History Department listserv and a post ran on the local interest blog, *Consuming Louisville*. This new coverage resulted in an additional 49 transcriptions on March 1 alone. Marketing efforts for the project continued and announcements were sent to blogs, community groups, professional organizations, and academic departments. In addition to local media coverage, the project gained national coverage when it was picked up by the *Journal of Blacks in Higher Education* on February 22, 2013. This aggressive marketing strategy was instrumental to the success of the project. Four months into the project transcribers had transcribed 1,648 article segments, an average of 14.5 transcriptions per day (Daniels et al., 2014, p. 45).

The marketing efforts that the project team put into the project had the effect of providing the University of Louisville Libraries and the Archives and Special Collections Department with positive publicity. The project team found that though they intended to increase participation, the aggressive marketing campaign also had the effect of advertising the other work that was being done in the libraries and reinforced the libraries' "commitment to providing resources of interest and use to the community" (Daniels et al., 2014, p. 46).

Transcribe Bentham

The Transcribe Bentham project experienced a jump in participation similar to the University of Louisville's project due to increased coverage in the media. The project planners' marketing plan was unclear early on but timely coverage in the press boosted participation in the project considerably. On December 27, 2010 a feature article on the *Transcribe Bentham* project ran in the *New York Times*. Participation in the project can

thus be measured in two separate periods, the first being the day of the project's launch (September 8, 2010) to December 26, 2010. The second period covers the time from when the *New York Times* article appeared to March 8, 2011 (Causer et al., 2012, p. 125).

In the first period, 1,207 people registered an account with Transcribe Bentham. Of these, only 259, or 21%, did any transcription. With the publishing of the article in the *New York Times*, the project experienced a dramatic increase in user participation. In the period from December 27, 2010 to January 7, 2011, the total of transcribed manuscripts numbered at 187. Though achieved in only about a week and a half, this total constituted an increase of 41% over the total of completed transcriptions in the entire first period before the *New York Times* article. The total number of completed transcriptions per week jumped from about 22 in the first period to roughly 56 completed transcriptions per week during the second period (Causer et al., 2012, p. 127).

The continued success of the Transcribe Bentham project appears to be due largely to the coverage the project received in the New York Times. The project has since gathered momentum, winning the 2011 Prix Ars Electronica Award of Distinction and garnering second place in the 2012 Knetworks competition ("Transcribe Bentham," n.d.). In all, at least 184 press articles, journal articles, blogs, and announcements have been published about the Transcribe Bentham project ("Transcribe Bentham," n.d.).

Digitalkoot

Perhaps due in part to the Digitalkoot project's innovative design, which employs games to encourage participants to transcribe old newspapers word-by-word, the project has garnered considerable media attention. On February 8, 2011, the National Library of

Finland launched the Digitalkoot project. Exactly one week after launch, Digitalkoot was featured in a national radio broadcast and appeared in several newspapers. These press releases resulted in almost 200 new users. By March 15, the project appeared in local business newspapers and was featured in a Wired.com article on March 17. *The New York Times* published a piece about Digitalkoot on March 23. By the end of March, 2011, over 30 articles had appeared in the press. These articles raised public interest in Digitalkoot and resulted in increased users taking part in the project (Chrons & Sundell, 2011, p. 24).

Social media appears to have played a vital role in obtaining participants for Digitalkoot. In addition to the marketing power that social media can hold, the 2007 Oxford Internet Survey found that social networking sites may enhance social capital. (Huijboom et al., 2009, p. 38). Digitalkoot may have taken advantage of this phenomenon by allowing users to login to the project using their Facebook accounts. During the first week of the project 1,756 users had friends who also joined; this number amounted to more than a third of all registered users that logged in with Facebook. The first week was by far the most active when it came to friends sharing information about Digitalkoot with each other. After the first week, only 341 users had friends who also joined the project. While the number of users skyrocketed around the time of the project's release and subsequent media coverage, the number of Digitalkoot contributors leveled off to about 300 per week during the period under inspection (Chrons & Sundell, 2011, p. 24).

Accuracy/Trustworthiness of the Crowd

When considering whether or not to engage in a crowdsourcing project, archivists must be assured that participants can be trusted to provide reliable, accurate information. The

idea that participants are amateurs is deeply tied into the concept of crowdsourcing. Brabham (2012) conducted critical discourse analysis of more than 1,300 articles related to crowdsourcing and found that the word "amateur" was used over 100 times (p. 399). He argues that as crowdsourcing matures, it is imperative to remain critical of how crowdsourcing is discussed, including whether or not participants in crowdsourcing projects can definitively be labeled as amateurs (Brabham, 2012, p. 395). In his book, *Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business*, Jeff Howe notes that the majority of participants are products of liberal arts educations. He suggests that many people may feel stifled by the hyper-specialization of modern capitalism and may engage in crowdsourcing efforts as a way to be part of a task that better utilizes their creative talents (as cited in Brabham, 2012, p. 396).

Surveys of participants in various crowdsourcing projects have shown that many participants are not really amateurs at all. InnoCentive, a crowdsourcing project that allows enthusiasts to attempt to solve scientific problems that the scientific establishment has yet been able to solve, conducted a survey and found that based on 320 respondents, 65% held PhD degrees, while 19.1% held advances degrees, and that the majority of degrees were in the sciences (Brabham, 2012, p. 400). Next Stop Design, a crowdsourcing project related to transit planning where users were asked to design bus stop shelters, conducted a survey and found results similar to those found in the InnoCentive survey. Of the 23 users who responded to the Next Stop Design survey, 18 reported to being either architects, architecture teachers, or intern architects seeking licensure. Those who reported other professions included an electrical engineer, a

surveyor, two graphic designers, and a computer programmer, some of these who reported having studied architecture in college (Brabham, 2012, p. 400).

The majority of users who participate in crowdsourcing projects do so out of a genuine interest in the project. A user survey of the Transcribe Bentham project found that most respondents participated in the project because of an interest in Jeremy Bentham, a general interest in history or philosophy, or an interest in crowdsourcing projects. Some reported to having an interest in contributing to the common good or making Bentham's manuscripts available to others, while others reported that they thought transcription was fun (Causer et al., 2012, p. 127). In a survey conducted by the University of Louisville aimed at participants in the *Louisville Leader* transcription project, one contributor noted, "I am enjoying this. . . . I know I am making a contribution, and in the process I am getting a good look at history from a different perspective. Because I have generally transcribed in a consecutive timeline, I feel that I have known some of these people, their clubs and church work, etc., as well as some of the issues that had meaning for them" (Daniels et al., 2014, p. 47). When people have a sincere interest in a crowdsourcing project and personally identify with the subject matter, it only makes sense that they can be trusted to perform to the best of their ability.

While participants may approach a crowdsourcing project with pure intentions, people are not perfect and are bound to make errors. Human error must certainly be taken into account when examining the effectiveness of any crowdsourcing project. Despite the propensity of humans to make errors, it must be conceded that most crowdsourcing projects involve tasks that computers are unable to do, or do accurately. In many cases, archives institutions must make the choice whether to get their information out to the

world, imperfections and all, or have a "dark" archive that few have access to (Ellis, 2014, p. 5). Results of crowdsourcing projects may prove that human error may be an overblown fear among archivists. The project team of Digitalkoot found that while OCR software only achieved roughly 85% accuracy when transcribing Finnish newspapers with archaic type font, contributors were able to achieve 99% accuracy while doing the same task (Chrons & Sundell, 2011, p. 23). In addition, staff at the *Louisville Leader* project reported that only one transcription they received was grossly inaccurate (Daniels et al., 2014, p. 45).

The everyday context in which a crowdsourcing project occurs may also have an effect on the quality of information provided by users. As stated previously, some authors warn against the commodification of culture, that information provided by members of the crowd, while sufficient and quick for the needs of today, may not be suitable to meet the information needs of future users, as opposed to authority control, which has the future needs of users in mind (Van Hooland et al., 2011, p. 709).

Spammers are another concern that many crowdsourcing projects face. Some people may take the opportunity to promote themselves or businesses rather than contribute meaningfully to a project. Others may provide false information simply out of malice. Crowdsourcing projects have struggled with ways to keep spammers from skewing project results or creating unnecessary work for project staff. For example, Digitalkoot project staff designed verification tasks that all users must complete satisfactorily before they are given any real tasks. A series of verification words are presented to users when they first create an account. These words are already known to the computer and are used as a measurement to determine whether a user is benevolent in their intentions. Users

never know whether the words they are transcribing are real or are used for verification. Once a user correctly transcribes the verification words, these words appear less frequently until the system can be sure the user is sincere at which point they are given a steady diet of untranscribed words (Chrons & Sundell, 2011, p. 22). The solution employed by Digitalkoot relies on simple CAPTCHA technology yet requires a significant level of programming expertise to implement. Other less technologically-driven projects may have to rely on a member of the project staff to check for spam and delete it, and so they should weigh the risk of spam and the implications it may have on the time team members can devote to the project.

Staff

Crowdsourcing presents challenges to archives staff and institutions. Though the work being done by participants may be free of charge, crowdsourcing projects require time, money, and careful planning if they are to be executed properly. Money will have to be devoted to a project manager. This person should be chosen carefully, as a project manager with vision, knowledge, leadership, and a strong work ethic is instrumental to the success of any crowdsourcing endeavor (Ellis, 2014, p. 5). Obviously, a project manager employed to work only for the project and nothing else will require substantial funding to pay their salary. The project manager may already be employed by the institution and must devote a portion of their time to the project, which means that the institution loses resources in terms of time that could be devoted to other collections. Time on the clock must also be spent to properly plan the project. An institution must weigh these considerations carefully before embarking on a crowdsourcing project or they are likely to be unpleasantly surprised (Ellis, 2014, p. 5).

The aspect of quality control must be addressed if a crowdsourcing effort is to be successful in providing accurate information of high quality. Depending on the nature of the project, this may involve employing a person or multiple people as editors responsible for ensuring the crowd's contributions are sufficient. Various projects have taken different approaches to quality control in terms of the number of staff and staff time devoted to it. These approaches have ranged from no additional staff contributing little time, to multiple staff members devoting a considerable amount of time to quality control. The University of Louisville's *Louisville Leader* transcription project did not employ any additional staff and very little time, if any, was spent on quality control (Daniels et al., 2014, p. 45). The *Louisville Leader* project seems to be more of an exception than the rule. Most other projects have devoted considerable resources to quality control. In the United Kingdom, the National Maritime Museum's Old Weather project required that submissions be checked three times before being included (Romeo & Blaser, 2011). The Transcribe Bentham project also expended a great deal of resources on quality control. Transcribe Bentham hired two full-time Research Associates to coordinate the various aspects of the project. One of their main priorities concerned the moderation of transcripts submitted by participants (Causer et al., 2012, p. 128). Since Bentham's transcribed manuscripts are to be published as part of his *Collected Works*, additional quality control measures are taken. Manuscripts to be included in the *Collected Works* are seen by at least four people at minimum, including one or more transcribers, the moderator, the *Collected Works*' editor, and the edition's general editor. Sometimes manuscripts are passed back and forth by multiple editors before being included in the *Collected Works* (Causer et al., 2012, p. 128). The University of Louisville employed the

crowd to transcribe manuscripts to be included locally on the Archives and Special Collections' CONTENTdm site for the *Louisville Leader* collection, while Old Weather created a national resource to promote scientific discovery and Transcribe Bentham created a product for publication in an ongoing print series. Due to their nature, the latter two projects needed to pass strict quality control measurements. The aim and scope of the crowdsourcing project certainly affects the amount of staff and staff time needed for it to be successful.

Assessment

A seemingly important aspect of any crowdsourcing project, and one that does not get a lot of attention in the literature, is assessment. Projects have used various methods to determine participants' thought and feelings, or levels of project use and exposure. Surprisingly, no article reviewed during the course of this study mentioned setting goals in the beginning of a project to compare against results in order to measure its success.

Surveys predominate as the preferred method of assessment among crowdsourcing project staffs. The Transcribe Bentham project created a web-based survey aimed at understanding the motivations users had for contributing, or not contributing, to the project. The survey was posted on the project blog, as well as on both the project's Facebook and Twitter pages. The survey was also posted on the project's Transcription Desk tool and was also sent to each user's profile. This approach resulted in about 8% of all account holders responding to the survey (Causer et al., 2012, p. 127).

The staff of the University of Michigan's Polar Bear Expedition, an early example of crowdsourcing in archives, put more emphasis on Web analytics than surveys to assess

their project. Transaction logs, user statistics, and search term analysis were part of the Web analytics regime. Content analysis, an online survey, and three semi-structured interviews rounded out the project's assessment methods. In truth, the project team used a multimethodological approach to assessment but the surveys and interviews did not yield much data. The survey only garnered six responses; coupled with only three interviews, the project team's ability to collect quantitative data via Web analytics outstripped their ability to collect qualitative data through surveys and interviews (Krause & Yakel, 2007).

Methodology

There are many stories of successful crowdsourcing projects undertaken by archives all over the world that extoll the virtues of crowdsourcing, claiming that these initiatives have resulted in added value to collections by accomplishing tasks that institutions would not have the resources to complete otherwise. In addition, many articles and blogs praise crowdsourcing efforts for creating connections with archival users, increasing the use of collections and notoriety for the institutions taking part in the various projects. But is the literature providing only one side of the story? Do the real-life experiences of archivists who have engaged in crowdsourcing projects match the overwhelmingly celebratory literature? In addition, this study also seeks to shed light on the extent to which crowdsourcing has penetrated the archives profession. Do crowdsourcing projects undertaken by archives institutions signify an emerging practice with staying power or just a passing trend?

To answer these questions, I created a survey using the Qualtrics survey tool that was sent via email to the Society of American Archivists' Archives and Archivists listserv to

solicit responses among archivists in the field seeking their opinions and experience regarding crowdsourcing. The results of this survey were examined quantitatively to determine the extent to which archivists in the field are taking part in crowdsourcing projects. Additionally, qualitative analysis of participants' responses uncovered the methods archivists used to implement crowdsourcing projects, the tools they employed, and the perceived success of their projects. Results tables from this survey are included to provide a visual representation of these findings. This paper concludes with a discussion of the limitations of this study and a discussion of the survey results and what they mean for the archives profession going forward, with particular attention paid to crowdsourcing as an archival practice.

Results

The first question that all respondents were given was: "Have you ever engaged in, or are currently engaging in, crowdsourcing to enhance archival description of collections or to engage users with collections?" Respondents were given the option to answer either "Yes" or "No." As shown in Figure 1, a majority of respondents answered "No" to this question. This question received 54 total responses. Of the 54 respondents who answered this question, 29 indicated having not participated in crowdsourcing either in the past or present; this totaled 56% of all respondents. 25 respondents reported to having either engaged with crowdsourcing either in the past or the present, which accounted for 46% of respondents.

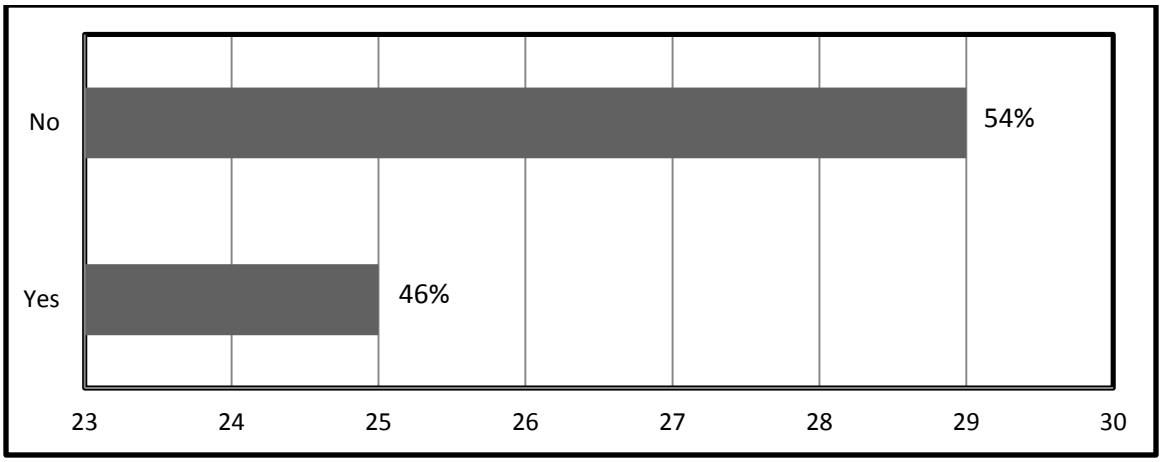

Figure 1: Number of respondents who have engaged or are currently engaged in crowdsourcing versus respondents who have not.

Respondents to the first question who answered that they have not engaged in crowdsourcing either in the past or present were given the follow-up question: "Do you have plans to engage in a crowdsourcing project?" Of the 29 responses to this question, 5 respondents (17%) indicated they have plans to engage in a future crowdsourcing project. Figure 2 illustrates the data gathered from this question. The data from this follow-up question, coupled with data from the first question of the survey indicates that of the respondents who answered both questions, over half of all respondents 51% (30 out of 59) reported to having either engaged in crowdsourcing, are currently engaging in crowdsourcing, or have plans to engage in crowdsourcing in the future, as shown in Figure 3.

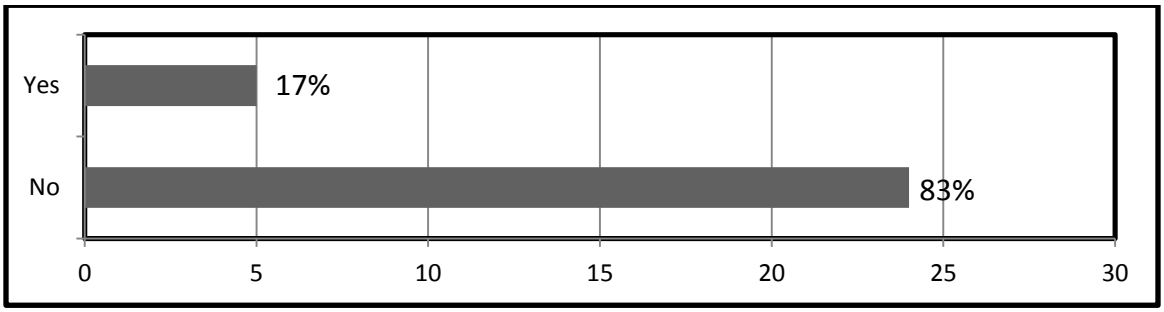

Figure 2: Number of respondents who have not engaged in crowdsourcing who responded they plan to engage in crowdsourcing in the future

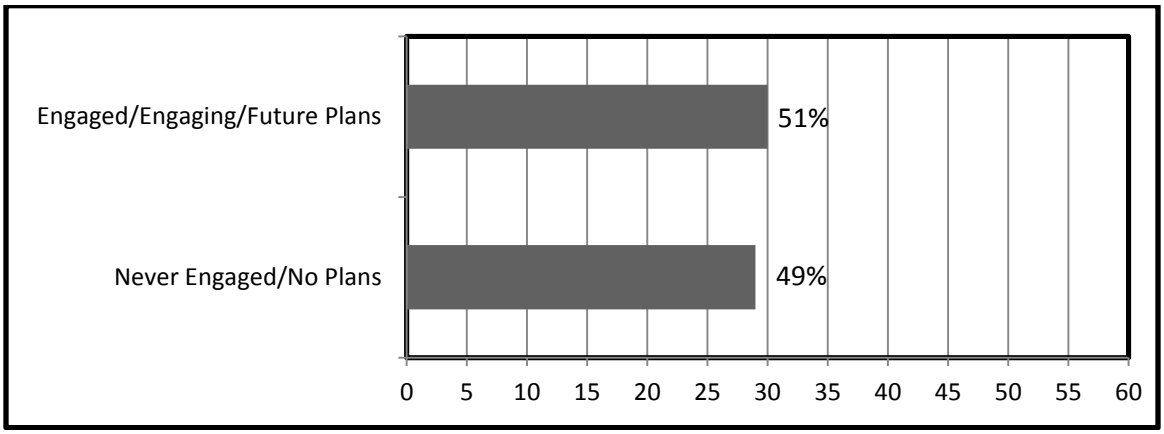

Figure 3: Total respondents who reported to having either engaged in, are currently engaging in, or have future plans to engage in crowdsourcing versus those who have not engaged in crowdsourcing and have no future plans to do so.

Finally, I asked respondents who indicated that they have not engaged in crowdsourcing and have no future plans to do so, to rate their level of interest in crowdsourcing based on a Likert Scale ranging from "Very Interested" to "Not Interested." Of the 24 respondents, 19 (79%) indicated being at least slightly to very interested in crowdsourcing, while only 2 respondents reported being completely uninterested, as shown in Figure 4.

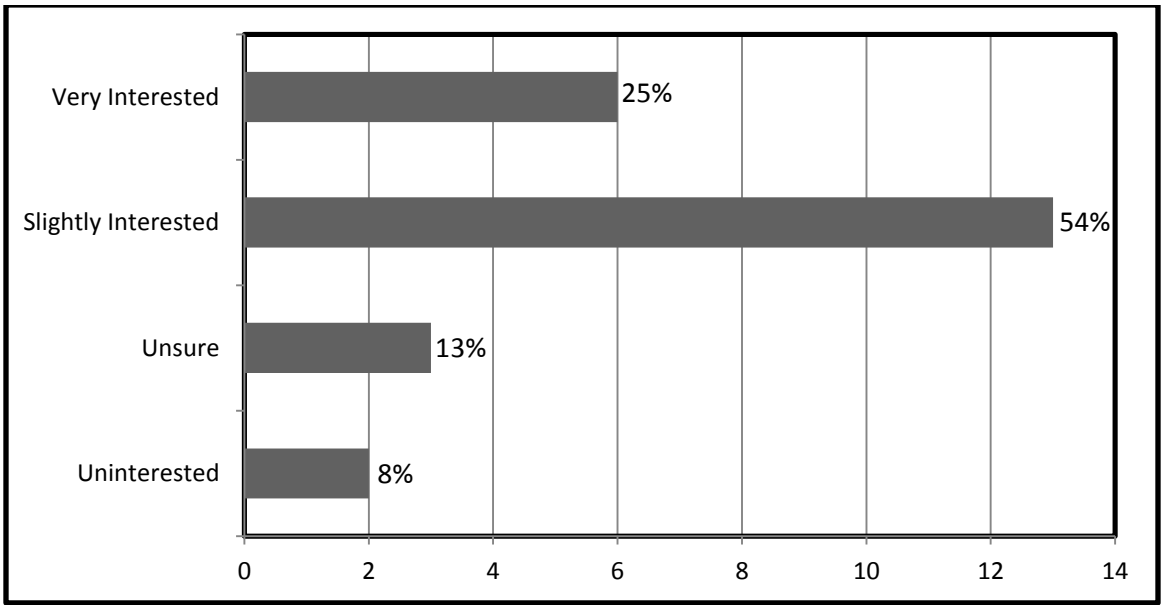

Figure 4: Level of interest in crowdsourcing among respondents who do not have plans to engage in crowdsourcing in the future.

Respondents who indicated in the first question that they have either engaged in or are currently engaging in crowdsourcing were then asked to: "Describe the nature of the archival materials engaged with, or being engaged with, in your crowdsourcing project." They were given the choices of photographs, manuscripts, maps, publications, or other. In the case of "other," respondents were given the opportunity to indicate the materials they worked with in writing. Figure 5 indicates that out of 24 respondents, 11 (46%) reported to having worked solely with photographs in their crowdsourcing project. Photographs constituted the single most cited media being crowdsourced by respondents. Seven respondents (29%) reported to having solely used manuscripts in their crowdsourcing projects, while one respondent reported as solely using publications. An additional five respondents reported to using materials not available in the choices they were given, or to using not just one, but a combination of media formats in their crowdsourcing projects. Of these five respondents, two worked with materials not found in the initial set of choices. These two respondents indicated they had either worked with video or audio. Among the remaining three respondents who utilized crowdsourcing for multiple types of materials, the first respondent reported to having worked with "slides, photographs, films, [and] manuscripts." The second respondent to report using multiple types of media in their crowdsourcing project indicated that they used "photos, anything posted online, documents, maps, [and] publications," while the third such respondent reported using "photos, manuscripts, scientific field notes, diaries, publications and illustrations."

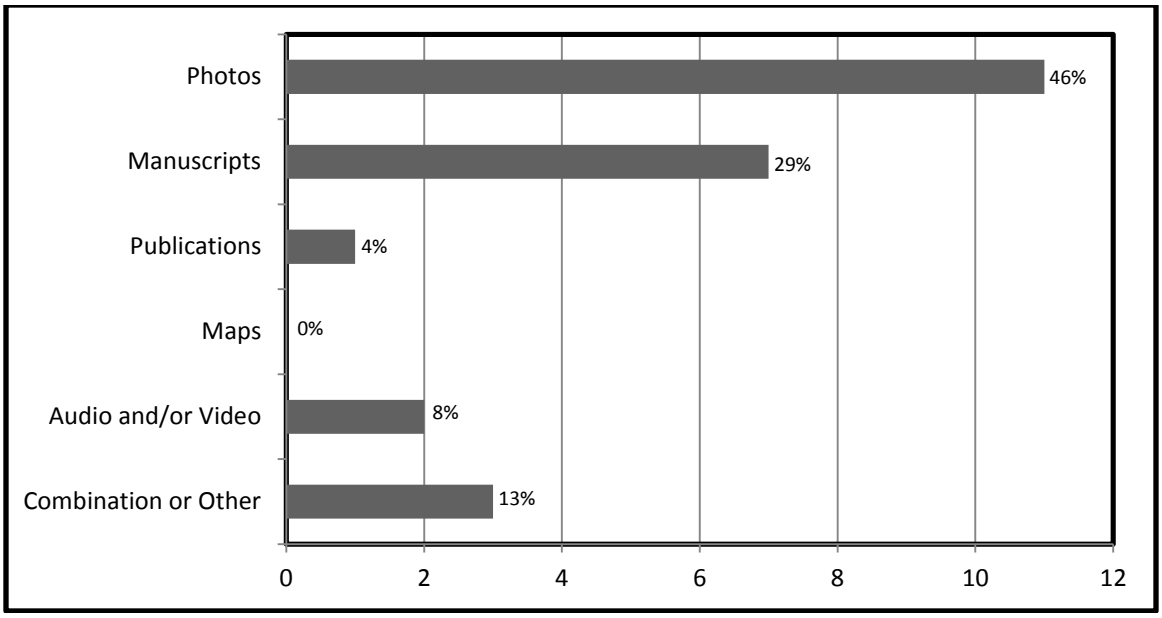

Figure 5: Types of materials engaged with in crowdsourcing projects.

Respondents who reported as having not engaged in nor currently engaging in crowdsourcing but responded to having future plans to engage in crowdsourcing were asked this same question of what materials they would use in their crowdsourcing project. Figure 6 shows that of the five respondents who answered this question, three (60%) reported to having plans to use crowdsourcing for publications, one (20%) reported having plans to use photographs, and one (20%) reported having plans to use audio and video. None of the five respondents reported to having plans to use manuscripts or maps.

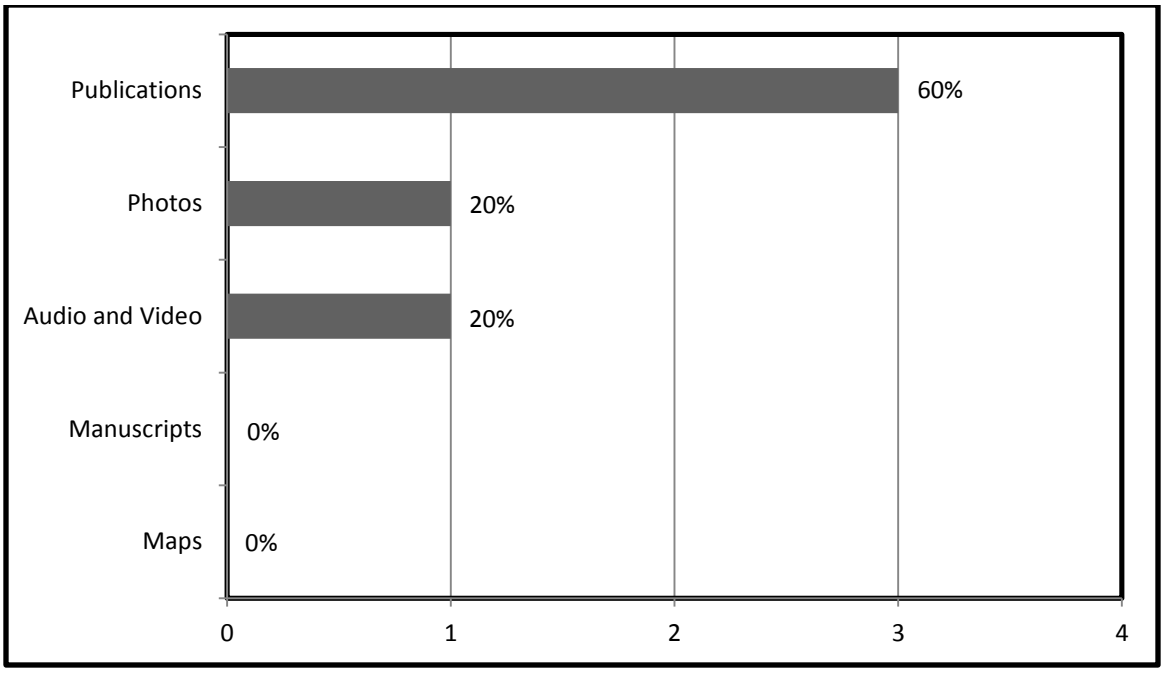

Figure 6: Materials which respondents who have future plans to engage in crowdsourcing plan to use for their project.

All further questions in the survey were responded to by those who indicated in their initial responses that they had either been involved in crowdsourcing projects or were currently involved in crowdsourcing projects.

Next, I asked respondents to "Describe any technology needed to carry out your crowdsourcing project, such as content management systems, blogs, software, social media, special programming knowledge, etc…" I provided respondents with a text box in which they could write their responses in narrative form. From the 20 people who responded came a wide variety of responses. I chose to analyze the content of these responses by classifying the types of technologies and competencies respondents reported utilizing in their crowdsourcing projects. In all, the technologies and competencies mentioned by respondents fell into seven categories. I then proceeded to count the number of times technologies or competencies that fell into these seven categories were

mentioned by respondents. Several responses indicate that respondents used, or are using, various combinations of technologies and/or competencies that fell into multiple categories. I simply tabulated each instance in which a given technology or competency was mentioned. These results are displayed in Figure 7.

The first category, "Blogs and Social Media," includes the blogs Wordpress and Tumblr and the social media applications Facebook and Twitter. Of the 20 respondents, 12 mentioned using blogs, social media, or a combination of both. The second category, "Content Management Systems/Digital Asset Management Systems (CMS/DAMS)," includes CONTENTdm, Omeka, PastPerfect, Archon, Archives Space, and Solr database. Twelve respondents reported using CMS/DAMS in their crowdsourcing projects. The third category, "Image/Video Hosting Sites," includes Flickr and Youtube. Six respondents reported utilizing image/video hosting sites in their crowdsourcing project. The fourth category, "Digital Imaging Equipment," includes cameras, scanners, photocopiers, and outsourced digitization services. Five respondents indicated they relied on digital imaging equipment to carry out their crowdsourcing project. The fifth category, "Image Viewing Equipment," includes slide projectors and computers. Three respondents indicated they used image viewing equipment to carry out their crowdsourcing project. The sixth category, "Special Programming Knowledge," refers to any sort of programming expertise that was needed to carry out a project. Four respondents indicated that specific programming knowledge was needed to carry out their crowdsourcing project. The seventh and final category, "Other Software/Applications," is a catch-all category that includes Microsoft Office applications, Photoshop, Moodle Word Processor, Google Hangouts, Google Maps, email, Amara, and instances of homegrown

applications. Nine respondents reported to utilizing one or more of these software or applications to carry out their crowdsourcing project.

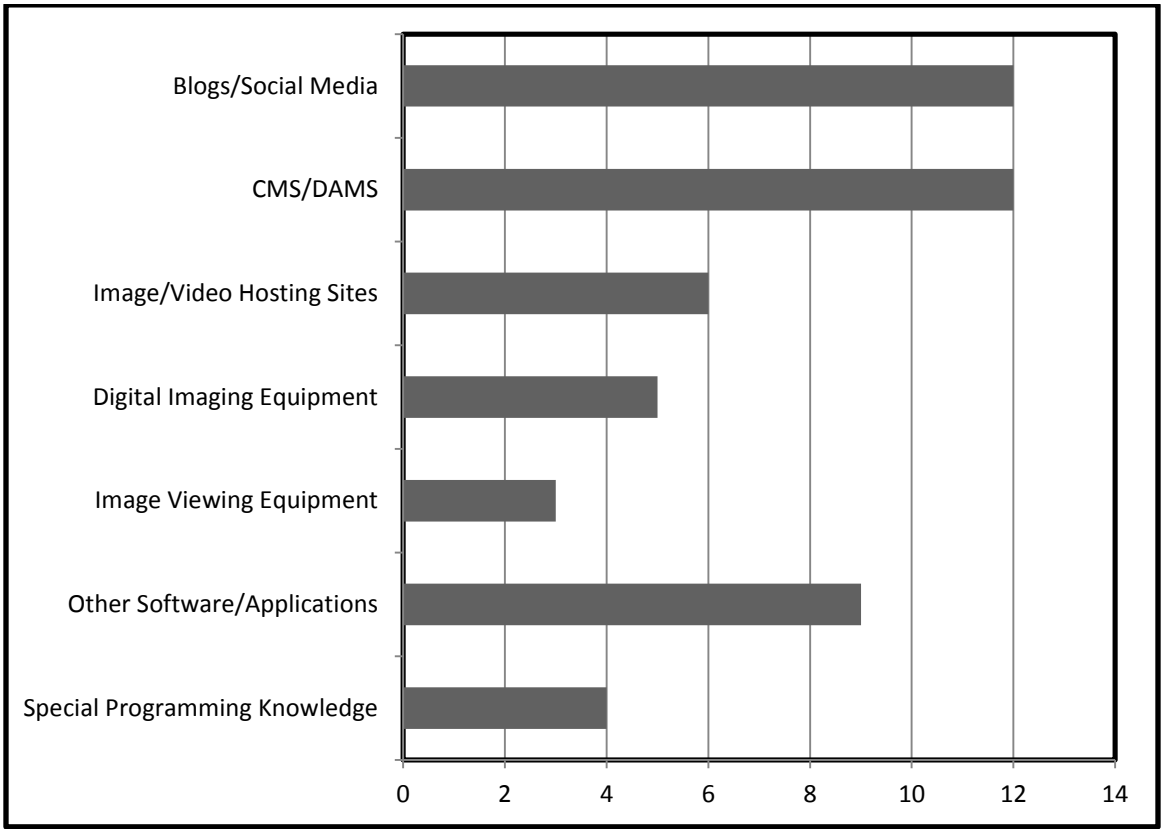

Figure 7: Technologies/competencies mentioned by those who reported engaging in crowdsourcing.

Respondents were then asked "Did you publicize your crowdsourcing project?" Respondents were given a choice between answering either "Yes" or "No." Out of 20 responses to this question, 16 respondents (80%) indicated that they publicized their project, as shown in Figure 8.

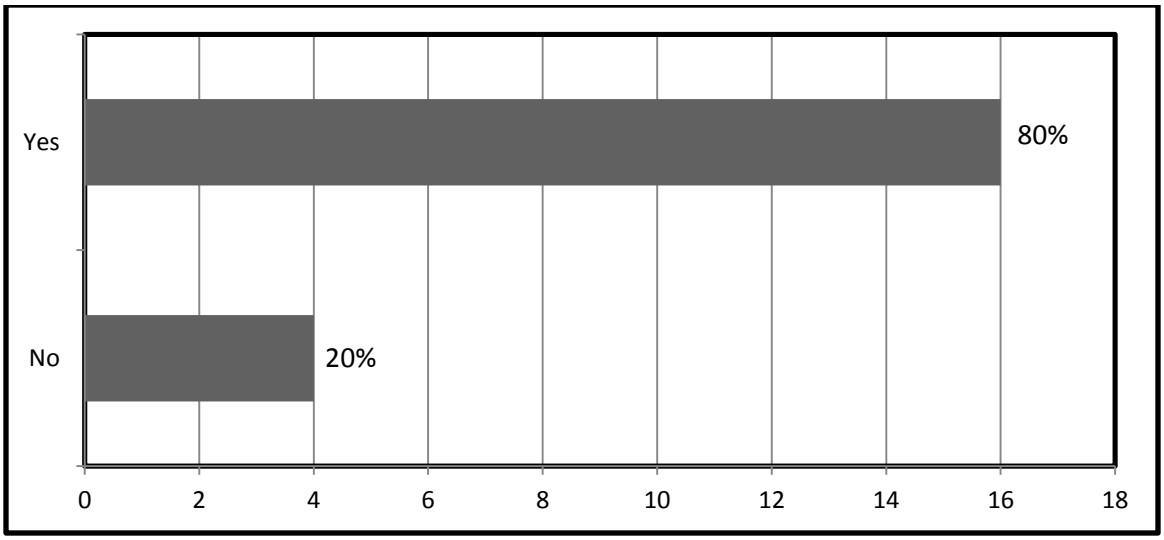

Figure 8: Number of crowdsourcing projects that were publicized versus number that were not.

Respondents who indicated that they had publicized their crowdsourcing project were then asked to: "Describe the methods you used to publicize your project and the communities of users you targeted." Respondents were given a text box and were required to answer the question in narrative form. In all, 15 responses were given to this question. This question can be broken into two parts: Part 1: Methods respondents used to publicize their crowdsourcing projects; and Part 2: User communities respondents targeted in their crowdsourcing projects.

Part 1: The majority of respondents indicated that they used a multi-methodological approach to publicizing their crowdsourcing projects. No two responses were alike, and so I decided to classify like methods together and provide figures for the number of times each method was mentioned. I was able to classify each method respondents employed to publicize their crowdsourcing project into ten categories. The categories are as follows: 1. Social Media; 2. Blogs; 3. Email; 4: In person (respondent to another person or people); 5: Media (radio, news media, press releases); 6: Presentations; 7: Newsletters; 8: Website Announcements; 9: Exhibits; and 10: Word of Mouth (people not executing the project

communicating with each other). Out of 15 responses, the use of social media was mentioned ten times, blogs were mentioned by six respondents, and email and in-person methods were mentioned five times apiece. The media, presentations, and newsletters were mentioned four times apiece. Announcements to websites were mentioned three times, while exhibits and word of mouth were both mentioned once. Figure 8 displays these results.

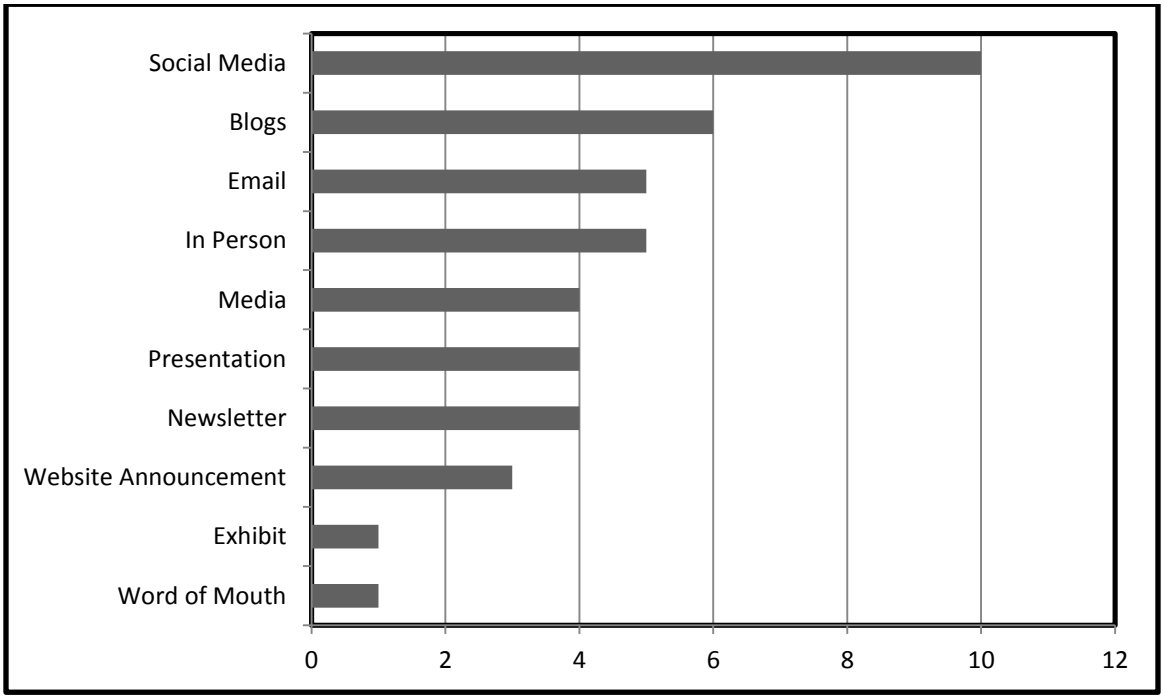

Figure 9: Methods of publicizing collections mentioned by those who reported engaging in crowdsourcing.

Part 2: Several respondents reported targeting more than one community of users. Responses were analyzed and grouped into six categories of user communities. The number of times user communities that fell into these six groups were mentioned was recorded, as shown in Figure 10. The foremost targeted group of users, "Faculty/Scholars," was mentioned five times by respondents. The second targeted user group, "Professional Organizations," was mentioned four times. The third targeted user group, "Alumni," was also mentioned four times. The fourth targeted user group,

"Students," was mentioned twice. The fifth targeted user group, "Local Citizens," was also mentioned twice. Rounding out the results was the sixth group, "Other Archives," which was mentioned once.

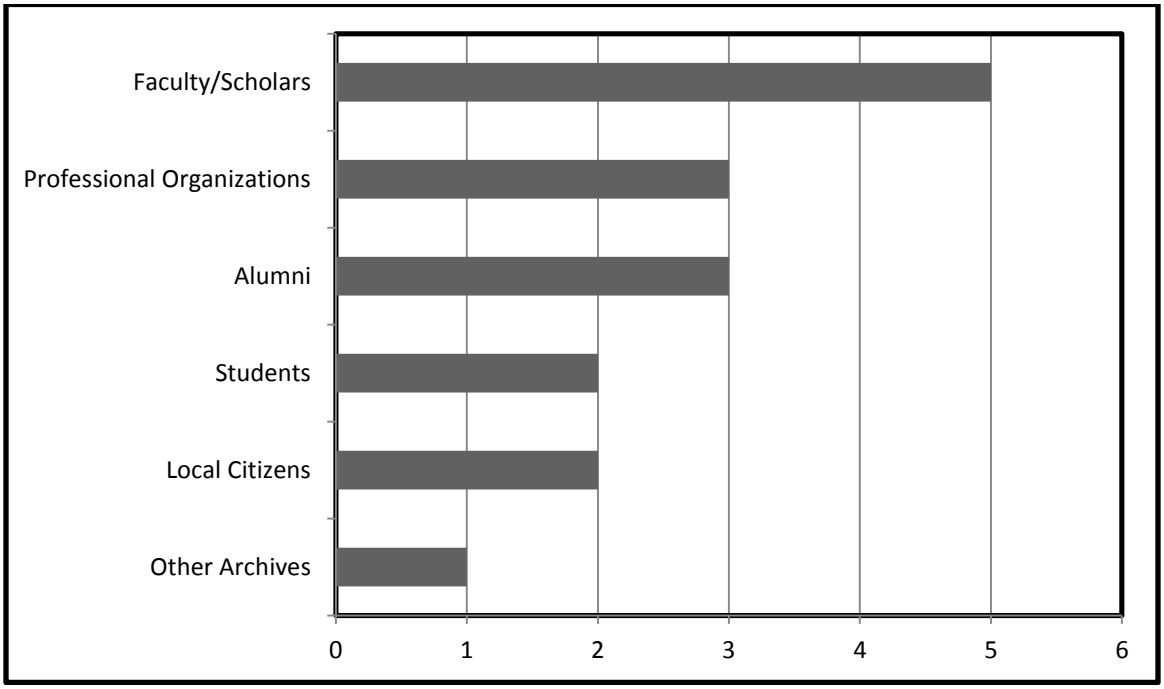

Figure 10: Reported target communities of crowdsourcing projects.

To understand the level of user participation respondents experienced during their crowdsourcing projects, I asked: "Based on your expectations going into your crowdsourcing project regarding user participation, how would you judge participation?" I provided respondents with a Likert Scale that asked them to rank user participation as having "Exceeded Expectations;" "Met Expectations;" being "Unsure;" having "Almost Met Expectations; or "Did Not Meet Expectations." Out of 18 responses, the majority of respondents (67%) indicated that user participation in their crowdsourcing projects met or exceeded expectations. These 12 respondents were split down the middle, with six (33%) indicating that user participation exceeded their expectations and another six (33%) indicating that user participation met their expectations. Three respondents (17%)

reported to being unsure if user participation met their expectations, while one respondent indicated that user participation almost met their expectations. Finally, two respondents indicated that user participation in their crowdsourcing projects did not meet expectations. The results of this question are shown in Figure 11, below.

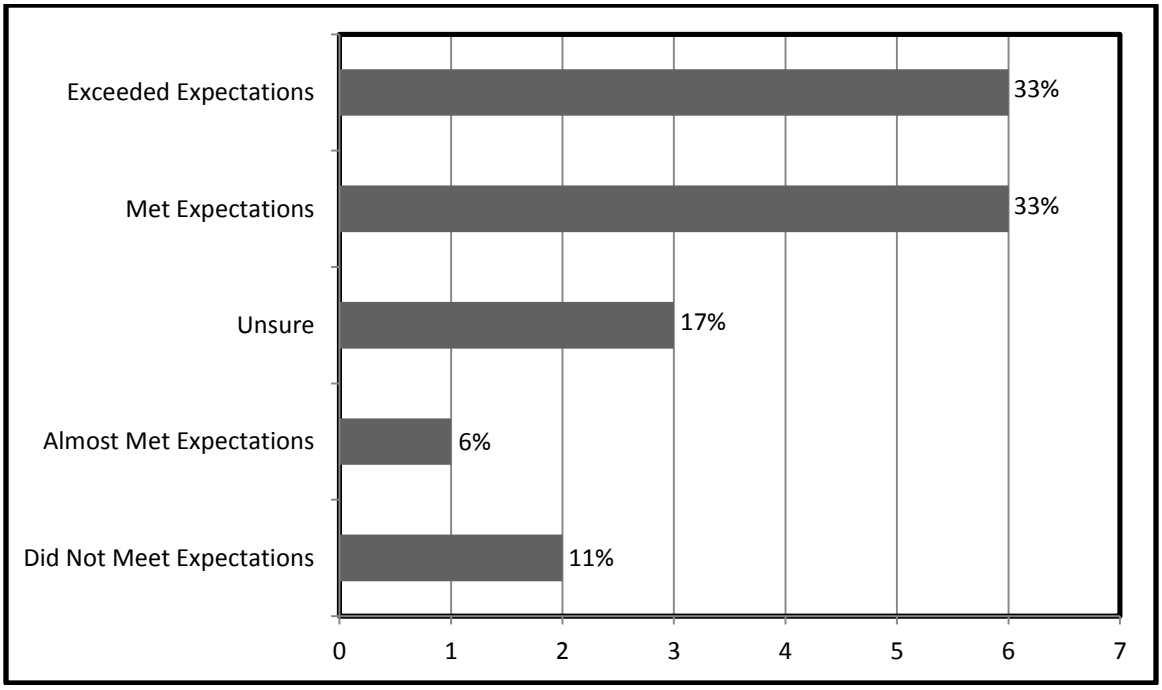

Figure 11: Whether or not user participation met respondents' expectations.

I then asked respondents to assess the overall success of their crowdsourcing project. The choices I offered to respondents represented a Likert Scale, where they could choose "Successful;" "Somewhat Successful;" "Unsure;" "Unsuccessful;" and "Too Early to Judge." Of the 18 respondents who provided their assessment, ten (56%) indicated their crowdsourcing project was successful. Four respondents (22%) indicated their crowdsourcing project was somewhat successful. Nobody responded as being unsure of success. Two (11%) respondents indicated that their crowdsourcing project was unsuccessful, while an additional two (11%) indicated that it was too early to judge the success of their project. The results of this question are shown in Figure 12.

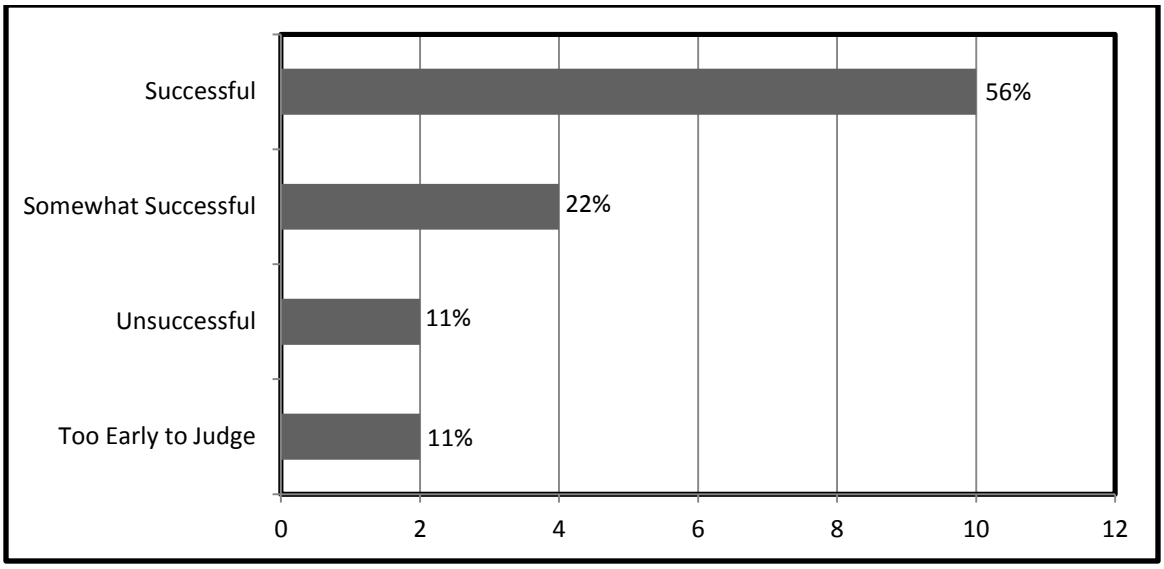

Figure 12: How respondents judged the overall success of their crowdsourcing project.

To determine the criteria by which respondents answered the question above, I asked them: "What made or will make this project a success?" For, this open-ended question I provided respondents with a text box where they could explain their rationale for judging the success of their projects. I received 18 responses to this question. The answers provided made it clear that there was no single way that respondents judged the success of their projects. In lieu of consensus, I realized that despite varying metrics respondents used to gauge the success of their crowdsourcing project, all answers seemed to focus on two aspects: 1. Completion of the work task; and 2. Engagement with the materials. Every response I received indicated that success was either measured by varying levels of completeness of the task being crowdsourced or by varying levels of crowd participation and interaction with the material. Some respondents reported they measured success by how well their projects lived up to expectations in both areas. I decided to count the number of responses that fell into these two, or both, categories. Results are displayed in Figure 13. Seven (39%) respondents reported having expectations for their projects that focused on some level of completion of the crowdsourced task. Eight (44%) respondents

provided answers that indicated they judged the success of their crowdsourcing project based on varying levels of user participation and engagement with the material. Finally, three (17%) respondents measured success by the amount of work they were able to get completed coupled with a certain level of user engagement.

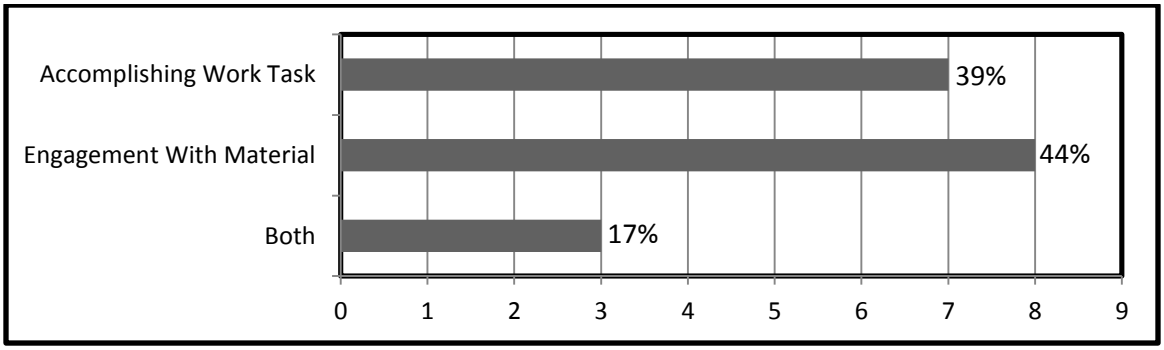

Figure 13: How respondents measured or will measure the success of their crowdsourcing project.

The last question respondents who reported to having been involved with, or are currently involved with, crowdsourcing were asked was: "Would you attempt another crowdsourcing project in the future?" Respondents were provided three choices: "Yes;" "No;" and "Unsure." Out of 18 respondents, 14 (78%) indicated they would attempt a crowdsourcing project again. No respondent indicated they would not attempt a crowdsourcing project again, while 4 (22%) indicated they were unsure whether they would attempt another project. Results of this question are displayed in Figure 14.

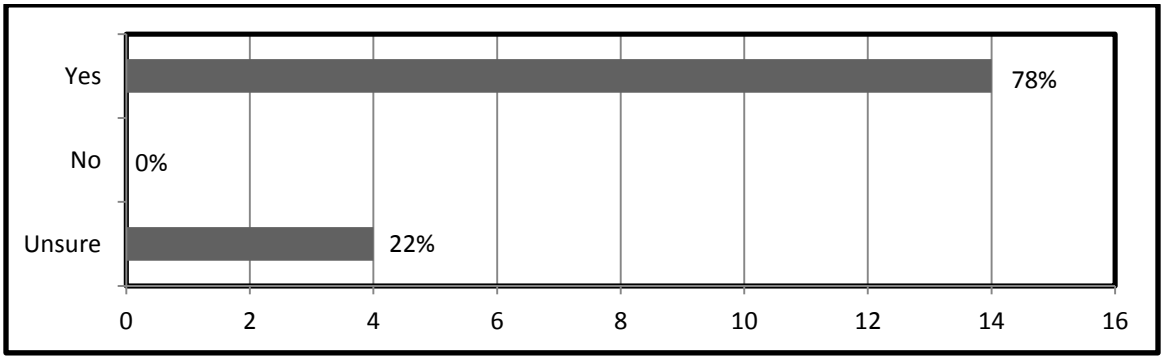

Figure 10: Number of respondents who reported engaging in crowdsourcing who would attempt another crowdsourcing project.

Discussion

Limitations

The impetus to conduct this study was derived from the lack of statistical information regarding the degree to which crowdsourcing is practiced in the archives profession. While there is certainly a substantial corpus of articles and blogs having to do with crowdsourcing, these represent merely a drop in the bucket when juxtaposed with the number of professional archivists in the world. The survey upon which this study is based was sent to a listserv for archives professionals but this does not guarantee everyone who participated in the survey was an archivist. Many professionals practicing in tertiary fields such as librarians and information technologists may have participated in the survey. There also exists a community of people who specialize in crowdsourcing technologies who may have taken the survey. Because of these variables, it seems impossible to definitively say that the results of this study can be applied specifically to archives.

The overwhelmingly positive literature on the subject of crowdsourcing led me to wonder if only the positive stories and viewpoints related to crowdsourcing were being broadcast. This study may not answer this question because I could not eliminate the possibility that people would respond to the survey out of personal interest or personal experience with the subject matter, while those who have no interest in the topic may have decided to not take the survey. Consequently, the results of the survey regarding the percentage of archivists out in the world who have actually conducted crowdsourcing projects may be skewed due to these variables.

Finally, only 54 people responded to the survey. This represents a miniscule fraction of professional archivists. While the results of this study may apply to a minute portion of the profession, it may be irresponsible to extrapolate these results over the rest of the profession. A study involving a much larger cross-section of archivists may be required in order to better answer the questions the present study seeks to answer.

Interpretation

The results of this study show that a significant proportion (46%) of archivists responding to my survey have been involved, or are involved, in crowdsourcing projects. An additional five respondents indicated they were planning a future crowdsourcing project. When these five additional respondents are taken into account, slightly over half (51%) indicated that they have either participated in crowdsourcing, are currently participating in crowdsourcing, or have plans to engage in crowdsourcing in the future; this represents a substantial number of archivists involved in crowdsourcing. While I did not expect to find that roughly half of all respondents indicated involvement in crowdsourcing, this finding seems to be supported by the sheer number of articles and blogs related to the subject. The results of this study suggest that there may be a substantial number of examples of crowdsourcing projects in archives that not everyone gets to hear or read about.

In my review of the literature, I was unable to find articles about crowdsourcing projects that were explicit failures. 11% of respondents in this study indicated that their experience with crowdsourcing was unsuccessful, which may lend credence to the notion that archivists do not readily expose their failures in the literature.

When asked what types of technologies or competencies were needed to carry out their crowdsourcing projects, respondents gave a variety of answers. Answers ranged from simple, low-tech projects such as an in-person event that utilized little more than a scanner and CONTENTdm, to more technologically robust projects such as one reported by a respondent that employed a homegrown tag database and homegrown interface developed to be an open-source program. Examples can be found in the literature that describe projects requiring varying levels of technical proficiency, and the results of this study affirm that there is no one way to create a crowdsourcing project and projects may be carried out with even the most modest technical expertise. The large number of ready-made tools that respondents used in their projects, such as image/video hosting sites, content management systems, and social media, and the fact that only four respondents reported to needing special programming expertise, indicates that archivists do not need to be programming whizzes to successfully carry out crowdsourcing, and that crowdsourcing can be pulled off with minimal additional equipment costs.

This study shows that crowdsourcing generally occurs for two reasons: to get people to engage with archival material, or to accomplish a work task that archivists do not have the resources to do themselves, whether it be from lack of funding, lack of time, or lack of expertise. In referencing *Transcribe Bentham*, Causer, et al., support this claim by noting that, "no funding body would ever provide a grant for mere transcription alone (2012)." When respondents were asked what made or will make their crowdsourcing projects successful, their responses fell into at least one of these two categories, if not both.

While archivists appear to engage in crowdsourcing efforts to complete work they cannot, for whatever reason, do themselves, or to engage people with the materials of a given archives, the criteria they assign to judge the success of their project varies greatly. One respondent reported that success meant "100% identification" of crowdsourced images, while another respondent measured success as the ability to identify "a good number of individuals in photographs." The first respondent was not confident that their project would be successful, while the second respondent seemed pleased with the results of their project. Photographs may be especially difficult to approach with the point of view of the first respondent (100% identification) because of various photochemical processes that degrade the integrity of the image, as evidenced by that same respondent's admission that some images were "not clear." An additional response indicated that the pace of task completion was "slow but steady" and indicated that the amount of work required to engage in crowdsourcing was probably equal to the amount of work it would take to transcribe the material themselves. This respondent still judged the project to be a success because they were able to engage the community with the materials in a substantial manner.

While 100% completion of a work task seems to be hit-or-miss according to the results of this study, community engagement with archival materials seems to be where crowdsourcing can really shine. Respondents reported to building relationships with communities that probably would not have otherwise. One respondent noted that perhaps because the target audience of their project was over the age of 65, they were wary about posting information to the Web. Consequently, the respondent reported engaging in several in-person meetings to gather the information they were after. This interaction and

relationship building is likely to lead to more users of the archives, thus enhancing the institution's viability. Another respondent noted that users were making new connections with the material and with each other. An additional respondent noted that participation was linked to publicity and that their project experienced spikes in participation with each press release. This point was reinforced by several articles in the crowdsourcing literature, and this study reaffirms that marketing plays a major role in achieving the goals set forth in the beginning of the project, whether they be to complete a work task or engage users with archival materials in new ways.

Conclusion

This study was able to answer some key questions related to the use of crowdsourcing in archives. While not something that every archives engages in, crowdsourcing appears to be an emerging practice with staying power in the archives profession, as evidenced by roughly 50% of respondents indicating they have engaged or will engage in the practice eight years after Jeff Howe coined the term.

This study was able to reaffirm the reasons found in the literature as to why crowdsourcing projects are being carried out; that is, to accomplish work tasks that are otherwise prohibitive for archivists to complete and to engage users with archival materials.

This study also shows that photographs constitute the most prevalent materials that archivists are using to engage in crowdsourcing projects. This is not surprising, as photographs often require a greater degree of contextual information to identify.

I consider this study's findings to be important in terms of examining the tools and competencies needed to undertake a crowdsourcing project. By providing evidence that there are a great number of methods in which crowdsourcing is being carried out, this study may work to provide encouragement to other institutions that they may find a crowdsourcing solution that fits them. The results may also work to educate archivists about the tools and competencies needed to carry out crowdsourcing. In a sense, this study may function as an aggregator from which archivists may find examples of crowdsourcing systems without having to track individual archivists down who have experience in the subject to find this information.

While several articles related to crowdsourcing discuss the manner in which institutions have performed assessment of crowdsourcing projects after the fact, often through Web analytics or user surveys, this study sheds light on the ways archivists measure the success of their projects and the expectations the have going into them. I believe that further study needs to be conducted regarding archivists' expectations of their crowdsourcing projects, as the results of this study reveal that archivists' opinions of what constitutes success vary greatly. Gaining a better understanding of how archivists measure success may have the effect of grounding archivists' expectations in some sort of consensus and could possibly lead to a greater number of crowdsourcing projects with realistic goals.

In addition, this study raises important questions as to why archivists are not engaging in crowdsourcing. This study shows that a disproportionate number of archivists who have not engaged in crowdsourcing are interested in the subject compared to the number of those same respondents who indicated they have no future plans to engage in it. Further

investigation of this key finding may result in understanding why archivists are not engaging in crowdsourcing, even though they may have an interest in the subject.

Crowdsourcing is relevant to archives; this study shows that. I would like to extend a thank you to everyone who participated in this study. I believe the answers provided by this study go a long way towards understanding how crowdsourcing is being approached and carried out by archivists today.

Bibliography

Brabham, D. C. (2012). The myth of amateur crowds. *Information, Communication & Society*, 15:3, 394-410. http://dx.doi.org/10.1080/1369118X.2011.641991

Causer, T., Tonra, J. & Wallace, V. (2012). Transcription maximized; expense minimized? Crowdsourcing and editing The Collected Works of Jeremy Bentham. *Literary and Linguistic Computing*, 27:2, 119-137. doi:10.1093/llc/fqs004

Chrons, O. & Sundell, S. (2011). Digitalkoot: Making old archives accessible using crowdsourcing. *Human Computation: Papers from the 2011 AAAI Workshop (WS-11-11)*, 20-25. Retrieved from http://www.aaai.org/ocs/index.php/WS/AAAIW11/paper/download/3813/4246

Cook, T. (2001). Archival science and postmodernism: New formulations for old concepts. *Archival Science*, 1, 3–24. http://dx.doi.org/10.1007/BF02435636%I

Daniels, C., Holtze, T. L., Howard, R. I. & Kuehn, R. (2014). Community as resource: Crowdsourcing transcription of an historic newspaper. *Journal of Electronic Resources Librarianship*, 26:1, 36-48. http://dx.doi.org/10.1080/1941126X.2014.877332

Duff, W. & Harris, V. (2002). Stories and names: Archival description as narrating records and constructing meanings. *Archival Science*, 2:3–4, 263-285. http://dx.doi.org/10.1007/BF02435625%I

Ellis, S. (2014). A history of collaboration, a future in crowdsourcing: Positive impacts of cooperation on British librarianship. *International Journal of Libraries & Information Services*, 64:1, 1-10. doi:10.1515/libri-2014-0001

Frankish, B. (Producer) & Robinson, P. (Director). (1989). Field of dreams. United States: Universal.

Google reCAPTCHA. *Intro*. Retrieved from http://www.google.com/recaptcha/intro/index.html

Holley, R. (2010). Crowdsourcing: How and why libraries should do it? *D-Lib Magazine*, 16:3/4. doi:10.1045/march2010-holley

Howe, J. (2006). The rise of crowdsourcing . *Wired*, 14.06. Retrieved from http://archive.wired.com/wired/archive/14.06/crowds.htm

Huijboom, N., van den Broek, T., Frissen, V., Kool, L., Kotterink, B., Meyerhoff Nielsen, M., & Millard, J. (2009). Public services 2.0: The impact of social computing on public services. *European Commission Joint Research Centre Institute for Prospective Technological Studies*, 1-134. doi:10.2791/31908

IMPACT. *Project architecture*. Retrieved from http://www.impact-project.eu/about-the-project/project-architecture/

Krause, M. G. & Yakel, E. (2007). Interaction in virtual archives: The Polar Bear Expedition digital collections next generation finding aid. *The American Archivist*, 70:2, 282-314. Retrieved from http://archivists.metapress.com/content/LPQ61247881T10KV

Light, M. & Hyry, T. (2002). Colophons and annotations: New directions for the finding aid. *The American Archivist*, 65:2, 216-230. Retrieved from http://archivists.metapress.com/content/L3H27J5X8716586Q

McNeill, B. (2014). Public identifies 60 people in VCU Libraries' exhibit of Virginia civil rights protest photographs. *VCU News*. Retrieved from http://news.vcu.edu/article/Public_identifies_60_people_in_VCU_Libraries_exhibit_of_Virginia

Raymond, M. (2008, January 16). My friend Flickr: A match made in photo heaven. Library of Congress Blog. Retrieved from http://blogs.loc.gov/loc/2008/01/my-friend-flickr-a-match-made-in-photo-heaven/

Romeo, F. & Blaser, L. (2011). Bringing citizen scientists and historians together. *Museums and the Web 2011*. Retrieved from http://www.museumsandtheweb.com/mw2011/papers/bringing_citizen_scientists_and_historians_tog

Slot, M. (2009). Web roles re-examined: Exploring user roles in the online media entertainment domain. *2009 COST Conference*. Copenhagen. Retrieved from https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&cad=rja&uact=8&ved=0CB8QFjAA&url=http%3A%2F%2Fwww.abs-center.si%2Fgbccd%2Fpapers%2FP058.pdf&ei=8dO9U5uIAsvvoATVjILQAg&usg=AFQjCNHv2RogAyGsx07kJ3zUIYJnSD1AUA&bvm=bv.70138588,d.cGU

Springer, M., Dulabahn, B., Michel, P., Natanson, B., Reser, D., Woodward, D., & Zinkham, H. (2008). For the common good: The Library of Congress Flickr pilot project report summary. 1-7. Retrieved from http://www.loc.gov/rr/print/flickr_report_final_summary.pdf

Transcribe Bentham. *Welcome to Transcribe Bentham*. Retrieved from http://blogs.ucl.ac.uk/transcribe-bentham/

Trove. *National Library of Australia*. Retrieved from http://trove.nla.gov.au/

Van Hooland, S., Mendez Rodriguez, E., & Boydens, I. (2011). Between commodification and engagement: On the double-edged impact of user generated metadata within the cultural heritage sector. *Library Trends*, 59:4, 707–720. Retrieved from http://hdl.handle.net/2142/26432

Vershbow, B. (2013). NYPL Labs: Hacking the library. *Journal of Library Administration*, 53:1, 79-96. doi:10.1080/01930826.2013.756701

Woods, D. (2009, September 29). The Myth of Crowdsourcing. *Forbes*. Retrieved from http://www.forbes.com/2009/09/28/crowdsourcing-enterprise-innovation-technology-cio-network-jargonspy.html

Wright, A. (2010, January 19). Online, it's the mouse that runs the museum. *New York Times*. Retrieved from http://www.nytimes.com/2010/01/20/arts/design/20museum.html?pagewanted=all&_r=0