Caroline L. Waller. Capitalizing on Collections and the Crowd: What needs to be considered before embarking on a digital crowdsourcing project? A Master's Paper Proposal for the M.S. in L.S degree. May, 2020. 46 pages. Advisor: Helen Tibbo

This study provides insights into the decision-making process archivists and librarians undergo when considering incorporating crowdsourcing elements into a digital project. The research was conducted via semi-structured interviews with eight American archivists and librarians that have conducted a crowdsourcing project. The study found that these professionals chose crowdsourcing as a tool for a variety of reasons and, while they discussed many challenges associated with crowdsourcing, they believed their projects a success. The study establishes four suggestions that will help library and archives managers be better prepared to manage and access the viability of potentially incorporating crowdsourcing elements into their digital projects.

Headings:

Digitization of museum collection

Crowdsourcing

Archives--Social aspects

CAPITLIZING ON COLLECTIONS AND THE CROWD: WHAT NEEDS TO BE CONSIDERED BEFORE EMBARKING ON A DIGITAL CROWDSOURCING PROEJCT?

by
Caroline L. Waller

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

May 2020

Approved by

_____

Helen Tibbo

Table of Contents

Introduction

Journalist Jeff Howe coined the term *crowdsourcing* in his 2006 *Wired* magazine article. Howe presented the concept as a spin on outsourcing--instead of companies contracting work to an outside company, they "tap the latent talent of the crowd" at a lower cost than paying traditional employees (p. 2). Saxton, Oh, and Kishore (2013) refined the definition of crowdsourcing as not only an enterprise that incorporates "crowds" and small-scale outsourcing, but also one that utilizes social web technologies (p. 2). Estellés-Arolas and González-Ladrón-de-Guevara (2012) analyzed the existing definitions of crowdsourcing to extract and establish the common elements and basic characteristics of crowdsourcing initiative, developing the following definition:

> Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken. (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p. 197)

Examples of early crowdsourcing initiatives across industries include Wikipedia, a community-made online encyclopedia; YouTube; Delicious; Amazon Mechanical Turk, an online marketplace for businesses to hire remotely located workers; Flickr; iStockphoto; the ModCloth Be the Buyer Program, which allow users to vote on products to push to production; and Fiat Mio.

During the first two decades of twenty-first century archives and libraries conducted a wide range of projects with crowdsourcing elements. These projects have demonstrated, to varying degrees of success, the benefits of crowdsourcing initiatives in archives. Among these benefits include enhancing archival description, metadata, and usability; creating publicity for a project and the larger institution; obtaining financial backing for archival digitization projects; and decentering the power of the archivist in describing and stewarding cultural heritage materials. However, crowdsourcing initiatives also have many potential detriments, which archivists need to weigh against these benefits. These projects are often time consuming, labor intensive, expensive, and may not yield high-quality results. This study aims to explore the use of crowdsourcing in archives and special collections digitization projects with these many costs and benefits in mind. Of particular interest is why librarians and archivists choose to incorporate crowdsourcing in their digitization projects. Data will be gathered from semi-structured interviews with archivists and librarians who are currently conducting digital crowdsourcing projects, or have completed such a project in the past. The following four questions will be examined in this exploratory study:

- Does crowdsourcing generate benefits beyond that of simply adding metadata or content to digital collections?
- What are the challenges of crowdsourcing projects?
- Are the results of crowdsourcing worth the time and effort to plan, implement, and assess the results?

- What factors and/or circumstances make a digital collection an ideal candidate for a crowdsourcing project? How do archivists and librarians weigh the pros and cons of such a project?

The goal of this study is to gain insights into the decision-making process archivists and librarians undergo when considering incorporating crowdsourcing elements into a digital project. The interviews conducted for this study explore questions relevant to archivists and librarians who are considering conducting crowdsourcing. Given the small sample size, and the unique circumstances of every archival institution and their digital collections, the study cannot make generalizations about crowdsourcing that will apply to all such projects. However, the experiences of the participants do allow us to glean suggestions about how and when to approach crowdsourcing projects. As a result, the study will help library and archives managers assess if their digital project is a good candidate for crowdsourcing and establish suggestions for helping them implement such a project based off of the practical insights of other professionals.

Literature Review

In the decade and a half since Howe coined the term, crowdsourcing initiatives have proliferated in libraries and archives. In 2007, the National Library of Australia launched the Australian Newspapers Digitisation Program, and by 2009 began recruiting volunteers to correct Optical Character Recognition (OCR) transcription mistakes in newspaper text (Ayres, 2013). Five European national libraries—the Moravian Library (Brno), the Nationaal Archief (The Hague), the National Library of Scotland (Edinburgh), the British Library (London), and the Institut Cartografic de Catalunya (Barcelona)—launched a crowdsourcing project to georeferencing, or adding spatial coordinates, to historical maps (Fleet et al, 2012). There have been several manuscript transcription projects, including the University College London organized the *Transcribe Bentham* project, the University of Oklahoma's *Transcribing the Past: Civil War Manuscripts* project, the Smithsonian Institution Transcription Center's *Field Book Project*, and Amazon's Mechanical Turk *Written Rummage* project (Causer et al, 2012; Lang & Rio-Ross, 2011; Reese, 2016; Parilla and Ferriter, 2016). Additionally, the New York Public Library (NYPL) launched several web-based crowdsourcing projects, including *What's on the Menu?*, which utilized volunteers to transcribe the Library's historical restaurant menus, while the Archives and Special Collections at the University of Louisville used CONTENTdm to crowdsourcing the transcription of the *Louisville Leader*, an historic African American newspaper (Lascarides & Vershbow, 2014; Daniels, Holtze, Kuehn, and Kuehn, 2014). The Library of Congress launched utilized

the popular photo sharing website Flickr to invite the public to help describe photographs through tags or comments (Library of Congress, 2008). Mary Flanagan and Peter Carini at Dartmouth College developed Metadata Games, a software system that uses computer games to collect metadata about archival images (2012). The University of Tennessee Libraries experimented with crowdsourcing, utilizing the public to add citations and subject tags to a new online bibliography, Database of the Smokies (DOTS: dots.lib.utk.edu) (Baggett, Bridges, Wise, Tanner, & Mezick, 2014).

The desire to **enhance access and develop robust archival description and metadata** drives many of libraries and archives crowdsourcing initiatives. Several institutions report that their crowdsourcing projects add significant content and value to their digital collections. The *Transcribe Bentham* project allowed wider access to Jeremy Bentham's nearly illegible manuscripts (Causer et al, 2012, p. 133). Daniels, Holtze, Kuehn, and Kuehn (2014) report that the *Louisville Leader* project also provided greater access to the collection for researchers (p. 47). The University of Tennessee Libraries DOTS experiment resulted in a significant amount of crowdsourced submissions, but many of the citations contributed by the public were obscure or difficult-to-locate publications that the DOTS team likely would not have identified (Baggett et al, 2014, p. 256). The DOTS team also found that while user-submitted tags may be useful, they did not contribute substantially to the taxonomy, and did not substantially modify or change the taxonomy (Baggett et al, 2014, p. 257).

Another driving motivation of crowdsourcing projects is **publicity**. The *Transcribe Bentham* project resulted in significant publicity for Bentham studies, history, and philosophy, and for crowdsourcing (Causer et al, 2012, p. 132). The project

reportedly engaged people from ninety-one countries and has been utilized by educators at Queen's University Belfast, Bloomsburg University, the University of Virginia, and King's College London (Causer et al, 2012, p. 133). The *Louisville Leader* project not only publicized the digitized newspaper, the crowdsourcing project advertised the positive work of the entire University of Louisville Libraries as a whole (Daniels, Holtze, Kuehn, and Kuehn, 2014, p. 46). The DOTS project facilitated opportunities for collaboration between research communities by creating a user base that would not only contribute but also spread the word of the existence and value of the database. (Baggett et al, 2014, p. 257).

Obtaining **financial backing** for digital projects may require libraries and archives to incorporate crowdsourcing in elements into their projects in efforts to increase publicity, encourage collaboration, and to demonstrate innovation. Crowdsourcing the *Transcribe Bentham* project was not particularly cost-effective, but the project team argued "this point is somewhat moot for one main reason: no funding body would ever provide a grant for mere transcription alone" (Causer et al, 2012, p. 131). New York Public Library (NYPL) Labs resorted to crowdsourcing in various projects because they had limited funding and staffing to devote to "cool-but-not-core" projects, and volunteer labor mitigated such limitations (Schwartz, 2012).

Crowdsourcing projects not only works well as a tool for generating transcribed data or metadata, it also **generates public engagement in these materials and the archive** (Parilla and Ferriter, 2016, p. 438). The Library of Congress Flickr pilot project resulted in 1.1 million total views of the account within the first 24 hours after its launch, caused the archival images to become visible for new audiences, and increased traffic to

the Library of Congress' website (2008, p. 14, 17). Parilla and Ferriter (2016) argue that part of the Smithsonian Institution Transcription Center's *Field Book* project was successful, in part, because of the well-designed program of public engagement (p. 448). This project explores the varying levels of engagement in crowdsourcing projects. Typically, crowdsourced transcription projects are highly and carefully structured with detailed templates, so Parilla and Ferriter ask, "[b]y fashioning a highly structured experience, are we fully engaging volunteers to interact with the materials?" (p. 438, 2016). Ultimately, they concluded that the increased communication between volunteers and staff, reinvigorated the online community by bringing in new volunteers, acted as an informal introduction to the Web interface, materials, and online community, and intensified interactions and output (p. 457).

When crowdsourcing projects encourage public engagement with the archive, this also contributes to a larger project of **decentralizing the power of the archive.** Shirky (2005) argues that a collaborative vocabulary created from the bottom up by contributors is more valid than a controlled vocabulary imposed from the top down by professionals. On the other hand, Adam Mathes (2004) argues that folksonomies actually create a controlled vocabulary of its own, reproducing a more traditional top-down controlled vocabulary. Furthermore, professional librarians and archivists are still required to moderate most crowdsourcing projects. Rafferty and Hidderly (2007) contend that most information professionals doubt that self-organizing folksonomies can work without there being some element of control and some form of "representative authority" (p. 376). Folksonomies are produced for specific audience and context and, therefore, may not effectively meet the needs of future researchers. The DOTS project demonstrates the

continued need of having professional librarians and archivists verify and correct citations and review copyright restrictions (Baggett et al, 2014, p. 256). Stvilia and Jorgensen (2009) suggest that these professionally moderated collections with folksonomy tend to exhibit a top-down mentality.

Handing over the work of creating metadata or transcribing archival materials to the public also has several potential downsides. Several crowdsourcing projects, including the *Transcribe Bentham* and the NYPL's *What's on the Menu?* projects, demonstrate the need for **quality control** of data generated by the public Causer et al, 2012; Saylor & Wolfe, 2011). The *Metadata Games* project found that image metadata and controlled vocabularies were applied inconsistently across collections. Flanagan and Carini acknowledge that lack of consistency does prevent crowdsourcing from resulting in a broader range of metadata from widely varying perspectives (2012, p. 438). Other projects found that crowdsourcing resulted in high quality transcriptions. Munyaradzi and Suleman (2014) propose a crowdsourcing method to transcribe handwritten manuscripts from the Bleek and Lloyd Collection. Their initial experiments demonstrate that volunteers were able to produce high quality, reliable transcriptions, with inter-transcriber agreement of 80% for iXam text transcripts and 95% for English text; when compared to ideal quality transcriptions, the average accuracy was around 65% (p. 117). Overall, they determined that these transcriptions are reliable and consistent (p. 125).

Crowdsourcing projects can be **labor-intensive.** The *Transcribe Bentham* team estimates that during the project's six-month testing period two staff members spent the equivalent of a month's full-time labor on moderating submissions and the associate upkeep of the Transcription Desk (Causer et al, 2012, p. 130). The *Papers of Abraham*

Lincoln editors found that when experimenting with non-academic transcribers they spent more time correcting errors than they would have had they carried out the transcription themselves (Cohen, 2010). Daniels, Holtze, Kuehn, and Kuehn (2014) found that transcribed articles that came in from public transcribers often required staff members to edit the same page multiple times, whereas they could more efficiently upload transcriptions to CONTENTdm if a staff member transcribe the entire page at once (p. 46).

There are several other potential challenges to crowdsourcing projects. First, crowdsourcing takes **time and money**, although Anastasiou and Grupta (2011) found the long-term maintenance costs are lower than outsourcing (p. 640-41). Second, crowdsourcing projects are often only funded for a limited amount of time, thus **lacking long term sustainability**. Several projects found that the motivation of participants is particularly difficult to sustain. Projects were most successful when institutions were able to motivate volunteers by acknowledging their contributions (Causer et al, 2011; Anastasiou & Grupta, 2011; California Digital Newspaper Collection; Zooniverse Old Weather Project). Finally, there are many **issues concerning privacy, ownership, intellectual property, and anonymity**. The Flickr pilot project team shared their concerns over putting images into the public domain when they are unsure or have no knowledge of its copyright restrictions (Library of Congress, 2008, p. 5). Anastasiou and Grupta (2011) state "although the Web 2.0 with social media and Web 3.0 with semantic web and metadata has opened many doors for open access and sharing through open standards supported by open tools and services, very often the way is open to commit antisocial, unethical, and even illegal activities (p. 642). There have also been

controversies in regards to certain transcription projects. A petition for the group *Translators for Ethical Business Practices* declare that crowdsourcing translation is an unethical business practice because "translation and interpretation are professions that require years of training, extensive general culture, and excellent command of both the target and source language." The group argues this form of crowdsourcing will lead to translations by unqualified people or the exploitation of qualified translators who should be compensated for their training and labor (Anastasiou & Grupta, 2011, p. 642).

Several crowdsourcing projects have resulted in the **creation of guidelines** that increase the success of future project. The *What's on the Menu?* project team's key suggestion is to make sure the source materials in your collection that you want to use for your crowdsourcing project are already exciting for participants and have public appeal (Lascarides & Vershbow, 2014, p. 135). Once this base is covered, Lascarides and Vershbow offer the following suggestions for designing the project: engage the user on an emotional level; appeal to the user's better nature, or frame the call for participation in terms of helping the institution; be transparent about the purpose of the project; make the task as small as possible; encourage continuation by immediately thanking the participant for their contribution, then asking them to contribute a bit more; show results of their input immediately; lower the barriers to participation; encourage a feeling of shared ownership; place the project in context by linking to the website to other reference sources; reward effort; report results of the project; make the product publicly available; and build an online community by utilizing social networks, blogs, and/or an online forum (2014, p. 132-35). The *Transcribing the Past: Civil War Manuscripts* project team focused on the lessons they learned in regards to crafting an effective funding proposal

for a crowdsourcing project, then suggest using a wiki-style transcription tool, allowing volunteers to edit each other, rather than requiring parallel transcription and triangulation of transcripts into a final version, which was time intensive and required the fulltime attention of a staff member (Reese, 2016, p. 70-1).

Methodology

Designing the Study

The purpose of this research study is to explore the reasons why librarians and archivists choose to include crowdsourcing components in their digitization projects. Crowdsourcing is a broad activity and has been incorporated in a wide variety of ways in library and archive projects, large and small. Crowdsourcing activities include transcription, tagging, quality control, bibliography creation, and metadata creation. However, crowdsourcing is expansive and not limited to these activities. Any participative online activity in which a library or archive utilizes the labor, knowledge, and/or experience of a crowd in a project that entails mutual benefit, may be considered crowdsourcing (Estellés-Arolas & González-Ladrón-de-Guevara, 2012, p. 197). For this reason, this study will rely on qualitative research methods. It will focus on the practical experiences of librarians or archivists who have and/or are currently conducting a crowdsourcing project. Although certainly not exhaustive, these experiences will contribute to recent scholarship that examines experimental crowdsourcing projects in digital archiving.

As a student researcher, I am interested in exploring the practicalities of crowdsourcing projects in archives. I approached this topic with no previous experience with these crowdsourcing projects, but rather with the hope that I will be prepared to weigh the costs and benefits of such projects as I progress in my career. While the study employed non-probability sampling methods and, thus, does not fully cover the wide

variety of crowdsourcing projects completed by many different types of institutions, the results are still transferable to other contexts. The goal of the study is to communicate how particular institutional contexts and circumstances are likely to yield a successful, or unsuccessful, crowdsourcing project in a library or archive. In addition, as a student with little experience with crowdsourcing projects, I entered this study free of professional biases and assumptions when it comes to decision-making surrounding these projects. While neutrality is impossible with a study of this nature, my positionality as well as the transparency of my research methods increase the trustworthiness of the study. This study received IRB approval as it involved human subjects.

Study Participants

The population this study investigates is American librarians and archivists who have conducted in the past, and/or are currently conducting, a digital archives project that has a crowdsourcing component. There is no comprehensive sampling frame for such a population, so the participants were selected using a mixture of snowball sampling, in which participants were asked to identify other potential participants, and convenience sampling (Wildemuth, 2017, p. 128). Potential participants were identified based on knowledge of their past participation in crowdsourcing projects, or the strength and reputation of institution's digital archiving programs and initiatives. In total, seventy-five institutions from forty-four states were contacted by email (Appendix A). Eight archivists and librarians from different institutions participated in the study. The participants were affiliated with either a university/college (7) and a public library (1). Out of the seven participants affiliated with a university/college, six participants were from a public

university/college and one from a private university/college. Out of the seven participants affiliated with a university/college, one was from a small-sized institution (less than 5,000 students), three from a medium-sized institution (5,000-15,000 students), and three from a large-sized institution (more than 15,000 students). Institutional support, resources, budget, and project goals for these crowdsourcing projects varied across these institutions.

Project Participants		
Participant #	**Institution Type**	**Project(s) Type**
Participant 1	university/college public university/college medium-sized institution	• Transcription • Transcription
Participant 2	university/college public university/college medium-sized institution	• Image solicitation, identification and/or cataloguing
Participant 3	public library	• Image solicitation, identification and/or cataloguing • Transcription
Participant 4	university/college public university/college large-sized institution	• Transcription • Compilation of a catalogue of scholarly sources
Participant 5	university/college public university/college large-sized institution	• Transcription
Participant 6	university/college public university/college large-sized institution	• Transcription
Participant 7	university/college private university/college small-sized institution	• Transcription
Participant 8	university/college public university/college medium-sized institution	• Image solicitation, identification and/or cataloguing

The eight participants discussed eleven different crowdsourcing projects varying in scope and scale. Seven of the projects involved transcription of handwritten documents

or newsprint; three of the projects involved image solicitation, identification and/or cataloguing; and one of the projects involved compiling a catalogue of scholarly sources. Several types of software or digital tools were used to facilitate the online crowdsourcing aspects of each project, including Flickr, an online photo sharing and management application; Omeka with the Scripto plugin, which allows the public to crowdsource transcription of the content management system's content; FromThePage, a free software for transcribing documents; Zooniverse, a collaborative research platform; WordPress, a free content management system, and Dropbox, a cloud-based file storage software. One project did not utilize a specific crowdsourcing software, rather volunteers communicated with the project manager via email. The software or digital platforms used for three of the projects was unspecified.

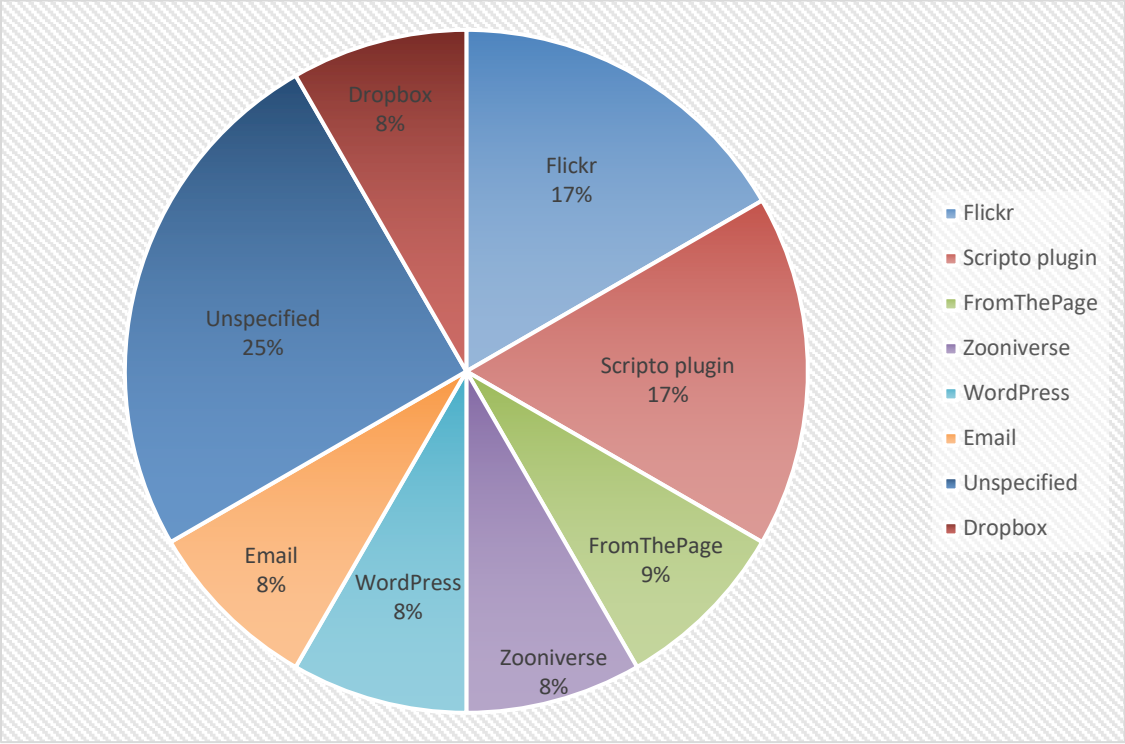

Figure 1. Types of software or digital tools used in participants' crowdsourcing projects

Data Collection

Data was gathered through semi-structured interviews with eight participants, ranging from 15 to 25 minutes in length. The bulk of the interviews were conducted over the phone since participants were located across the United States. One of the interviews was conducted via email at the request of the participant. The semi-structured interviews allowed for the flexibility that is important for probing into each participants' perspectives and experiences with crowdsourcing, which varied drastically from project to project. This mode of investigation provided the space for discussions beyond the answers to the predetermined questions and allowed for serendipitous responses. During the interview the participants were asked to describe their involvement with crowdsourcing projects, then discuss the motivations behind the projects. The complete interview guide can be found in Appendix B.

Each semi-structured interview was transcribed. Next, qualitative analysis of the interviews was conducted and the interview transcripts were coded according to major patterns, themes, and categories. The major themes that developed during the analysis process address the research questions posed at the beginning of the study, ultimately informing the development of general guidelines and suggestions for librarians and archivists when assessing the viability of embarking on a crowdsourcing project.

Results

Introduction

Through qualitative analysis of these eight semi-structured interviews, several themes begin to emerge regarding the participants' thoughts about their crowdsourcing projects. This chapter organizes these themes around the three broader questions they address:

- Why choose crowdsourcing as a tool?
- What factors did participants need to consider when conducting their project?
- What lessons were learned from these projects?

These results provide insights into the decision-making process archivists and librarians undergo when considering incorporating crowdsourcing elements into a digital project. These results helped establish suggestions for assisting librarians and archivists hoping to implement their own crowdsourcing project based on the practical insights of other professionals.

Why choose crowdsourcing as a tool?

Participants reported choosing to utilize crowdsourcing in their projects for seven general reasons. First, most participants reported the desire to use crowdsourcing to encourage community and/or scholarly engagement with the archival materials or archival institution. Second, all participants noted the importance of creating digital content through crowdsourcing. Associated with the first two reasons, participants chose

crowdsourcing as a tool to increase access to their collections. Other acknowledged reasons behind crowdsourcing projects were lack of funding for traditional archival work, the need for specialized knowledge and/or help, pressure from library or institutional leaders, and the desire to be a trailblazer in their field.

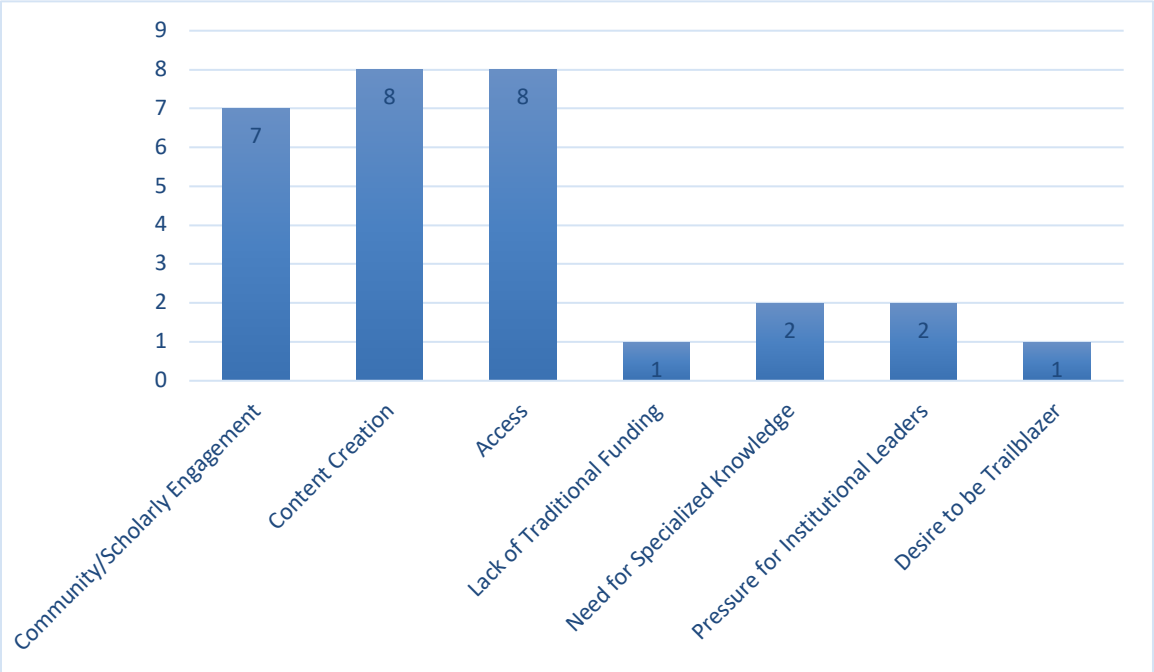

Figure 2. Reasons why participants chose to use crowdsourcing as a tool

The majority of the study's participants acknowledged that one of the major reasons they chose to incorporate crowdsourcing into their digital projects was to encourage **community and/or scholarly engagement** with the archival materials or the institutional repository. The interviews revealed that the participants sought varying types of engagement that was often unique to their collections and institutional culture. One participant articulated a desire to encourage engagement with the content within their collection through crowdsourcing projects as a way to complement other engagement initiatives: "… we do a lot of publications and projects that involve the community, but this is a way to get people engaged at a different level they are providing content"

(Participant 2). Another participant expressed that his institution was focused on more than mere community engagement with the archival materials, but also with transforming the public's perceptions of and relationship with the library itself. The reason his institution chose to utilize crowdsourcing was because it served as a tool to make their users and community feel involved in library affairs, thus "dissolving the 'this and them' mentality" and encouraging users to see the archival collections as belonging to the public (Participant 3). One participant stated that he hoped volunteers' engagement with his repository's collections might encourage donations to his institution: "One of the selfish reasons for creating the user text correction is to get people in and have a sense of ownership for the project and, eventually, hope they would donate funding. I think it has been mildly successful" (Participant 4). Engagement sought through crowdsourcing projects could be narrow in scope, such as a particular academic audience's increased use of a single archival collection, or broad in scope, such as encouraging wider public participation with a library and its resources.

Some participants were less concerned with encouraging engagement, and more focused on the **content** created with the help of volunteers. One participant expressed that the hope of increased community and academic engagement had little to do with why he chose to utilize crowdsourcing for his project. When tasked with an extremely time intensive transcription project, he first envisioned using work study students to complete the project, before attending a regional archives conference where crowdsourcing was discussed: "When I went to that conference session, a light bulb went on automatically. Like I have a project for this, it's perfect" (Participant 7). Another participant noted that she needed more images to contribute to her library's photograph project, "and

logistically [crowdsourcing] is a way to get a whole lot of imagery… images that we have not had before with relatively little effort on our part" (Participant 2). Several participants reported that they turned to crowdsourcing when optical character recognition (OCR) conversions of their textual materials failed in some way. Crowdsourcing was used as a tool to correct computer generated text or and improve access to the content.

Tied in with both the desire of increased engagement and creating digital content for a repository, all participants expressed the desire to increase **access** to their collections. Several participants noted that handwritten archival materials are becoming less and less accessible to archival users. One participant attributed this barrier to the new generation of scholars conducting research at her institution: "I think the main driving point is that this material is handwritten and the handwriting is hard to read, period. And [we are] finding that this is coming more and more of a barrier over time because our scholars are younger and younger and they are less able to read the handwriting" (Participant 6). Another participant noted that due to the changes to the research process because of scholar's increased use of computers and the internet, "the collection is a lot less useful if you can't search and read the content in a machine-readable environment" (Participant 3). She viewed crowdsourcing as a way to bridge the gap between the format of her collections and the evolving access needs of scholars. One participant emphasized the importance of producing transcriptions to aid in the development of digital humanities and scholarship associated with her institutions' collections (Participant 6). Participants hoped for increased access for the evolving needs of scholars and types of scholarship.

There were several other reasons for initiating a crowdsourcing project that participants reported. One participant noted the **lack of funding for traditional archival work** associated with his bibliographic data project. He reported that the impetus for his project was "the declining funding for traditional cataloguing work," causing his institution to ask researchers to volunteer to take on the task (Participant 4). Another participant noted a **need for specialized knowledge** when it came to his digital project, as his student assistants were unable to keep up with the demands and subject matter of a photograph cataloguing project, resulting in a backlog of images that needed further processing before they could become accessible. To address this problem, he solicited the help of librarians, historians, or interested community members familiar with the subject matter through Facebook (Participant 8). Two participants explicitly noted **pressure from library and institutional leaders** to conduct their crowdsourcing project. One participant discussed how his decision to transcribe a diary in his collection hinged on the university president asking him to do so (Participant 7). Another participant noted that her university's deans and heads of department were interested in increasing community engagement, to which crowdsourcing was viewed as a tool to encourage such engagement (Participant 1).

One participant became a relatively early adopter of crowdsourcing around 2012. She acknowledged that she became interested in utilizing crowdsourcing because she hoped that her library and herself would be **viewed as a trailblazer** in archives:

> I think it was just the timing of it. If there hadn't been other people already doing crowdsourcing and publicizing their crowdsourcing efforts, it wouldn't have been something to have occurred to us... It was exciting that other people were getting all of this attention and that the entire library and archives world was getting excited about the possibilities with crowdsourcing. So, we really did jump on the

> bandwagon, trying to be trailblazers in a way, since there were a couple of people doing crowdsourcing but it wasn't that many. We were still early adopters. We thought we would be able to get done what we wanted to get done and also be able to do presentations and papers and cast a good light on our operation here by jumping in early and experimenting with this methodology. It was also a way to publicize our collection itself. If we got a lot of publicity for the methodology of crowdsourcing, then it would also publicize the content, you let people know how we got this stuff and what was available online. (Participant 5)

As demonstrated by all eight participants, crowdsourcing was utilized for more than just creating digital content for each institution. While they often expressed similar motivations behind each of their crowdsourcing projects, each participant's decision process was informed by conditions unique to their collections and institutions.

Factors considered before crowdsourcing

Once participants decided they might want to use crowdsourcing as a tool, they had to consider the many factors that would inform the feasibility, scope, and nature of such a project. The most prevalent factors reported related to the resources available to each project team. These resources took the form of project personnel, funding, and software and/or other tools. Other factors considered include the ability to solicit high-quality volunteers, the need for comprehensive technical plans, and university or institutional rules and regulations regarding volunteer labor. Most participants discussed a combination of these factors, which they either considered before their projects, or realized in hindsight that they should have considered them.

One of the most important factors informing the participants' projects was the crowdsourcing **tools and software** available to the project team. One participant noted the need to acquire a scanner that was capable of producing high-quality scans for her transcription project (Participant 1). Most of the participants focused on the factors that

informed the type of software they were able to use for the project. While several participants were able to shop around to select their ideal software, others were stuck with legacy software systems that informed how they were able to logistically complete the project. One project team already had Omeka installed on their library server, so they selected the Scripto plug-in to build off of the existing software (Participant 5). Several participants noted that their project team did not necessarily choose the best software available, rather the cheapest. One participant explained his choice to use Flickr in similar terms: "I've seen some software used in crowdsourcing and they are very nice and the problem with Flickr is your transcript is underneath your original, so scrolling up and down all the time… the better way to have it is side by side on one screen. We didn't have the luxury" (Participant 7). Most participants expressed their need to work with the best possible option given the resources they were provided.

When it came to selecting software, most project participants noted the lack of development personnel, which restricted their options. One participant noted that her project team landed on FromThePage after looking at other open source options and finding they did not have the development staff to support those other options (Participant 6). Another participant noted that, while they could afford their ideal software, they would not be able to get the technical support to customize it for their particular institution. The participant stated: "The computing services people were not interested in taking on another project, and they would see my project as another addition to their workload. So that wasn't going well. That's why when I was offered the suggestion of using something free, it really clicked with me" (Participant 7). Another

participant discussed his need for a development personnel to engineer his institution's commercial software so to facilitate crowdsourcing:

> For the newspaper project the user text correction (UTC) is layered up on top of commercial software that we purchased and we pay a yearly maintenance fee for that software. So, there was development work we had to pay for to create the tool that allows users to go in and correct the text. From a standards or technology perspective, the generated text is not applied to the XML itself and so if you were to try to import the data across systems, user text correction is not in a standardized format. (Participant 4)

Choosing the appropriate software for each crowdsourcing project was one of the most prevalent themes the participants discussed.

The participants also had to consider other types of **project personnel** before embarking on their projects. One participant discussed how she was able to capitalize on the consolidation of information technology staff across her institution's campus and a round of retirements, which left several open positions and allowed her to build her own unit. This unit included a web developer, programmer, and digitization staff that could facilitate a large-scale crowdsourcing project (Participant 1). Another participant noted that, while his institution did not hire new staff or student workers to assist with a crowdsourcing project, he found that he had to take on new duties himself (Participant 3). Yet another participant elaborated on the importance of not only having sufficient personnel to complete the project, but also the need to develop an understanding of how these personnel will maintain and administer the crowdsourcing project (Participant 4).

A significant part of the success of a crowdsourcing project, as well a typically time-consuming aspect for project personnel, is publicity and engagement with volunteers. Several participants expressed their surprise over the amount of time spent participating in such activities in comparison with the time spent completing the more

technical aspects of the projects. One participant admitted that she was blindsided by the how much outreach and volunteer engagement her team's crowdsourcing project required:

> What I should have known or what we should have explored when we started that is how much outreach or engagement with the transcription or volunteer community it takes to get a good product when it comes to transcription... To get solid and continued engagement it definitely takes a lot of effort on our part, to actively engage with them through email, or through setting up transcription challenges, or social media output. (Participant 1)

Another participant noted the difficulty of balancing outreach activities with the other aspects of her repository's crowdsourcing project. During the project, her team constantly were concerned with if they were doing enough on the publicity side of the project (Participant 5). Participants found that such outreach activities need to be considered when assessing the time and personnel required to make a crowdsourcing project a success.

Funding for the crowdsourcing projects was also a frequent theme during the interviews. One participant noted that her team was afforded only $80 a month for four months to complete the project. The limited finances also limited the scope of the project: "... like I said we have really been limiting our time… about how much time we spend in the system. This may come back to be a big mistake, we might decide that we really should have spent more time in the system and we should have jumped in and corrected people more or encouraged people that were doing great" (Participant 6). Another participant discussed how her institution's crowdsourcing project was made possible by an outside grant (Participant 2). While not all participants explicitly discussed how their projects were funded, they did discuss how finances informed decisions regarding tools, software, and project personnel.

Several participants contended that **high quality volunteers** were vital to the eventual success of the entire project. One participant noted that several volunteers "ghosted" him shortly after they began working on the project, but one good volunteer remained active and contributed significantly to the project (Participant 8). Another participant admitted that the project was a success due to the help of one volunteer, who actually transcribed all of the documents that were a part of his diary transcription project: "So, our crowdsourcing wasn't so much a crowd as one. But we had to put it out there and it took a while. He found it and contacted me. The fact that it was available widely through the Internet made it findable, and that helped a lot" (Participant 7). Several participants solicited volunteers from community groups or within participants subject-based interest groups. Knowledgeable and enthusiastic volunteers proved to contribute to several projects' overall success, and many participants strategized how to reach such volunteers during the planning stages of their projects.

Several participants emphasized the importance of having a **comprehensive technical plan** at the onset of the project. Some emphasized the importance of clear workflows for the project team, or detailed guidelines for volunteers to follow. One participant noted that she had to clearly articulate which members of the project team would be accountable for each aspect of the crowdsourcing project after the project had already began, something she encouraged others to implement at the very beginning: "We had to bring structure to it, so my unit manages the platform and does the digitization and loads it in, and they have people working on the frontside, doing the engagement and the outreach. So, we developed that, but it's something you should think about at the very start" (Participant 1). According to these participants, all aspects of the

project, no matter how small, should be tested and planned prior to implementing the project and soliciting volunteer labor.

University or institutional rules and regulations are also important to consider before beginning a crowdsourcing project, although only one of the eight participants mentioned any related issues in their interview. He noted that he had to take into account his university's rules concerning volunteer labor, which required volunteers to fill out paperwork. However, he found that completing volunteer paperwork was "far too intrusive for what these folks were doing," as the university wanted more information than he thought was necessary and worried the paperwork would discourage volunteers. Fortunately, he was able to work with his university's human resources department and they were able to make an exception for the crowdsourcing project (Participant 8).

Lessons learned from crowdsourcing projects

The eight participants shared diverse, and sometimes contradictory, lessons learned from their experiences conducting a crowdsourcing project. Participants discussed several topics as a response to the questions pertaining to the lessons they learned during their projects, including project planning, expectations for volunteer labor, selections of crowdsourcing platforms and/or software, and collaboration with other institutions. These lessons can serve as advice for other librarians and archivists hoping to conduct their own similar project.

Several participants emphasized the importance of intensive **project planning** prior to beginning the project. One participant admitted that he jumped impulsively into his crowdsourcing project, and wished he had spent more time on creating workflows and

writing instructions for volunteers (Participant 8). Another participant suggested writing a project charter, in which project expectations are defined in advance (Participant 6). Another discussed the importance of articulating a clear goal for the project at its inception. She contended that you should not embark on a crowdsourcing project merely for the purpose of doing a crowdsourcing project. She stated:

> I would say… have, or decide on, the one or two desires, either reasons or outcomes, why you want to do it. So, if engagement is what it is, that's great, just figure out all the other aspects, so the technical aspects, the running of it, your standards, that type of thing. If it's more of a multi-prong, so for engagement but also you're trying to get some sort of product, like transcription, then really think about how the engagement and the product and the technical requirements, including not only the technology but in your transcription guidelines and the kind of product you want at the end, to be talked about in the beginning, or at least acknowledged. There is stuff you figure out along the way, but understanding across the project and then seeing how all of that works together. Just know that you have to have one with the other… and it is a lot of work. (Participant 1)

In addition to defining a project's expectations during the planning stages, another participant encouraged others to start with a small project, as he found that his larger crowdsourcing project took much longer than originally anticipated (Participant 7).

Several participants discussed their interactions with or expectations of volunteers. A couple advised that sustained engagement with volunteers is key to conducting a successful crowdsourcing project. Participants discussed the importance of constantly monitoring volunteer's contributions, which includes answering their questions in a timely manner. Sustaining such levels of engagement with volunteers was found to be incredibly time consuming. One volunteer emphasized the importance of offering a high-level of guidance for volunteers from the very beginning of a transcription project:

> I think there could be on our part there could be a little more guidance on that correction, like telling people what kind of papers to correct or somehow how to

> correct it. That is actually something we are working on right now. For example, if a word is misspelled in the original, you tell them to stay with that misspelling and if not, how do they encode both the misspelled and the correctly spelled word. (Participant 4)

Another participant warned that, regardless of guidance provided for volunteers, archivists and librarians should remain open-minded with their expectations when it came to crowdsourcing. He encourages others to "expect a little less of what [volunteers] are immediately capable of," so to prevent disappointment in the results of volunteer labor (Participant 8).

Two participants explicitly mentioned their **selections of crowdsourcing platforms and/or software** when posed with questions concerning the lessons they learned from their projects. They offered somewhat contradictory advice when it came to choosing a software or web platform. One participant expressed that she wished she had chosen another platform for her project's website. She found that the platform chosen and created by their web designer was unnecessarily complicated, and she also wished they had selected an open source program (Participant 2). However, another participant encouraged others to not worry too much about selecting the ideal software or platform for their project: "Don't be afraid to try something that may not be ideal. Had I waited to try and get the good stuff that the National Archives are using for something, I would still be waiting maybe. Going with something that will work and that is free was a good solution for us, and it got our project done" (Participant 7). This contradictory advice is perhaps due to the differences between their respective projects and their project goals.

One participant, who was involved with a larger scale crowdsourcing program, encouraged others to **collaborate with other institutions**. He discussed how collaboration with other institutions led to the success of one of his crowdsourcing

projects. Such a collaboration allows institutions to share technology, expertise, and information, while increasing the reach of your digital project. He also discussed his desire to collaborate in the future on his institution's work on the curation of linked data. Collaboration would allow him to share costs with other institutions, while being able to "explore what is working and what isn't collectively rather than having to do it on our own" (Participant 4). While other participants did not explicitly encourage others to collaborate with other institutions, several of the projects discussed were conducted in such a manner.

Discussion

Introduction

The interviews conducted for this study were intended to explore questions relevant to archivists and librarians that are considering conducting crowdsourcing projects. Given the small sample size, and the unique circumstances of every archival institution and their digital collections, the study cannot make generalizations about crowdsourcing that will apply to all projects. However, the experiences of the participants do allow us to glean suggestions about how and when to approach crowdsourcing projects. Through the experiences of other professionals, librarians and archivists can be better prepared to manage and access the viability of potentially incorporating crowdsourcing elements into their digital projects.

The Research Questions Revisited

Does crowdsourcing generate benefits beyond that of simply adding metadata or content to digital collections?

The most obvious reason for choosing to utilize crowdsourcing in digital archives projects is to create and/or compile content or metadata about the materials. But this study confirms that the benefits of crowdsourcing can be much broader than merely creating transcriptions or metadata. Another very apparent benefit is generating community and scholarly engagement with the archival materials or the archival institution as a whole. The study finds that archivists and librarians may be pressured to

encourage such engagement from library or institutional leaders. Another benefit of crowdsourcing projects was the procurement of specialized knowledge from outside the institution. Given the nature of archives, archivists and librarians are often required to process and describe materials pertaining to subjects outside of their subject expertise. For this reason, the need to solicit this knowledge from the public will persist indefinitely. Obtaining specialized knowledge and generating engagement with the public are indisputable benefits of crowdsourcing. The study also found that some professionals have sought to be trailblazers in their libraries and archives. However, given the abundance of crowdsourcing projects over the past decade, it is unlikely that archivists or librarians will continue to seek crowdsourcing for this reason unless the form of crowdsourcing is particularly innovative.

What are the challenges of crowdsourcing projects?

This study underscores the many ways in which crowdsourcing projects can be a challenge. The archivists and librarians interviewed found that the feasibility and nature of their projects was informed by the resources available to them. The project teams often had to operate their projects with limited budgets, requiring them to opt for less-than-ideal software and tools. The study also finds that crowdsourcing projects require a significant amount of time and labor. Several of the participants discussed how they initially underestimated the amount of time spent engaging with volunteers, conducting outreach or marketing activities, and curating the volunteer created content. In some cases, a crowdsourcing project took up time that could go towards other library projects. While several of the participants found high-quality volunteers, who contributed

invaluable labor and subject expertise to the project, several also found that volunteers tend to produce low-quality transcriptions and/or metadata, or they do not remain invested in the project for long enough to make any helpful contributions.

Are the results of crowdsourcing worth the time and effort to plan, implement, and assess the results?

According to the archivists and librarians interviewed for this study, they believed their crowdsourcing projects were successful. Of course, the perspective of these participants is undoubtedly skewed since they each had already made the choice to incorporate crowdsourcing into their projects. The study does not meaningfully engage with professionals who have had more negative experiences with crowdsourcing. Though the participants assert that their projects were a success, this assertion does not mean the projects were not challenging. Though none of the participants regretted their projects, they will not necessarily choose to complete a similar project in the future. One participant, upon reflecting on her institution's successful newspaper transcription project, admitted that in the future she would spend more money on better tools that would eliminate the need for crowdsourcing. When it came to her new newspaper digitization project, she made sure to spend more money on digitization in order to better capture the full text of the newspapers so as not to need to conduct another crowdsourcing project. Based on the advice provided by these participants, crowdsourcing can be worth the time and challenges, depending on the overall goal of the project. Those who implemented their project with the goal of producing community and/or scholarly engagement succeeded in doing so. When engagement is secondary or

inconsequential to the intended outcomes of a project, the findings suggest that crowdsourcing may be less useful unless other technical avenues fail, given that crowdsourcing is incredibly labor intensive and time consuming. While participants interviewed in this study all believed their projects were successful, these projects still posed challenges, both common across projects or unique to their institutional context. These findings point to the importance of careful and thorough planning, including a solid grasp on the project's key purpose.

What factors and/or circumstances makes a digital collection an ideal candidate for a crowdsourcing project? How do archivists and librarians weigh the pros and cons of such a project?

Based on the recollections of this study's participants, the process of weighing the pros and cons of crowdsourcing projects is unique to each librarian and archivist, as crowdsourcing projects come about in many different manners, for many different reasons. The projects discussed in this study concerned materials that had high research value, whether it engendered popular interest from a local community or interest group or it served a particular type of academic or digital humanities scholarship. Due to the success of the crowdsourcing projects examined by this study, a collection that would be an ideal candidate for crowdsourcing is one that can sustain long term interest with a community or scholar group. Engagement is critical to the success of most crowdsourcing projects, so archivists need to factor in the time, personnel, and resources required to foster this engagement. As several of the participants found that fostering the necessary level of engagement took more time and effort than they initially anticipated,

archivists should overestimate the resources required for the engagement and marketing portion of their projects. Archivists and librarians may also want to reach out to their colleagues at other institutions to help explore crowdsourcing software and strategies, or to possibly facilitate collaboration with these institutions. The suggestions gleaned from this study cannot be generalized to all crowdsourcing projects, but they do offer archivists and librarians a good starting point when weighing the pros and cons of conducting a crowdsourcing project. This study was unable to establish quantifiable factors that make a digital collection an ideal candidate for crowdsourcing, but does suggest that what is an ideal collection for crowdsourcing for one institution, within a particular community, may not be ideal for another.

Suggestions for crowdsourcing

Though this study does not make generalizations about crowdsourcing that apply to all institutions and projects, the experiences of the participants informs the following four suggestions to consider before embarking on a crowdsourcing project. These suggestions are based on the themes gleaned from discussions about successful crowdsourcing projects that varied in size, longevity, and scope.

1. **Define the goal of the project in the early planning stages**.

There is no reason to conduct a crowdsourcing project solely for the purpose of conducting a crowdsourcing project. Goals of a crowdsourcing project may include, but are not limited to, fostering community and/or scholarly engagement, obtaining specialized knowledge, boosting a library's reputation within scholarly or local

community, promoting awareness of collections, and creating transcriptions or metadata that facilitates wider access to collections. Not all projects can achieve all goals. Determining the primary versus the secondary goals of a crowdsourcing project should help inform the project design. The interviews suggest that crowdsourcing may not be optimal for all goals of digital projects. Archivists and librarians should explore all avenues that may allow them to achieve their overall goal. If engagement is the primary goal, is there a better way to foster constructive engagement with the collection or institution? If creating some sort of content for the digital collections is the primary goal, is there a better way to create this content? Does an institution have student workers qualified to work on the project, or are there better tools that may allow you to develop the content without help from the crowd? The ultimate goal of each project should guide the design of the project. Crowdsourcing can be challenging, labor intensive, and time consuming. Project managers should develop an end vision for their project to help them weigh the pros and cons of crowdsourcing.

2. **Planning is key to a smooth project**.

The importance of intensive project planning prior to beginning the project will undergird the viability of a crowdsourcing project. Planning includes selecting and testing the technical tools and workflows, writing thorough instructions for volunteers, creating an outreach or marketing plan, and determining how labor will be divided between members of the project team. Archivists may want to write a project charter, in which project expectations are defined in advance. No matter the level of planning prior to a project, crowdsourcing projects will likely result in unexpected challenges or

serendipitous occurrences. For this reason, one should plan for each aspect of the project to take more time or resources than expected. Careful planning paired with overestimating time and resources will help prepare project managers with any unexpected challenges.

3. **Marketing, outreach, and communication sustains volunteer labor.**

The study finds that crowdsourcing projects require a significant amount of time and labor. Several of the participants discussed how they initially underestimated the amount of time spent engaging with volunteers and conducting outreach or marketing activities. To maintain productive and prolonged engagement, the project teams must consistently communicate with volunteers, answer any questions in a timely manner, and advertise the project through social media or other channels.

4. **Collaborate with other institutions.**

Collaborating with other institutions during a crowdsourcing project or program allows for the share of technology, expertise, and information, while increasing the reach of the project. If archives and libraries have a limited budget for conducting such a project, collaboration may allow a project team to cut back on costs or share costs or resources with another institution. Collaboration may be especially beneficial if the project team hopes to sustain the project for a longer amount of time.

Conclusion and future research

This study confirms that there is no consensus interrogating the viability of a crowdsourcing project. Still, the study provides insights into the decision-making process archivists and librarians undergo when considering incorporating crowdsourcing elements into a digital project. Though the study cannot make generalizations about crowdsourcing that will apply to all such projects, the experiences of the study's participants do allow us to glean suggestions about how and when to approach crowdsourcing projects. As a result, the study will help library and archives managers assess if their digital project is a good candidate for crowdsourcing and establish suggestions for helping them implement a crowdsourcing project based off of the practical insights of others.

Moving forward, more research should be conducted that explores the types of software and digital platforms used in transcription or metadata projects. A complete survey of the digital tools available will help archivists and librarians make the best decision for their projects. In addition, this study does not examine the experiences of archivist who have decided not to conduct a crowdsourcing project. Learning more about the experiences of those who do not have such positive experiences with crowdsourcing will offer a more balanced perspective on the topic.

References

Anastasiou, D., & Gupta, R. (2011). Comparison of crowdsourcing translation with Machine Translation. *Journal of Information Science*, *37*(6), 637-659.

Ayres, M. L. (2013). Singing for their supper': Trove, Australian newspapers, and the crowd.

Baggett, M. P., Bridges, A., Wise, K., Tanner, S., & Mezick, J. (2014). Populating the wilderness: crowdsourcing database of the smokies. *Library Hi Tech*, *32*(2), 249–259. https://doi-org.libproxy.lib.unc.edu/10.1108/LHT-11-2013-0150

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.

Cohen P. (2010). Scholars Recruit Public for Project. *New York Times*. [retrieved November 7, 2010] Available from: http://www.nytimes.com/2010/12/28/books/28transcribe.html?nl=books&pagewanted=all

Daniels, C., Holtze, T. L. . terri. holtze@louisville. ed., Howard, R. I., & Kuehn, R. (2014). Community as Resource: Crowdsourcing Transcription of an Historic Newspaper. *Journal of Electronic Resources Librarianship*, *26*(1), 36–48. https://doi-org.libproxy.lib.unc.edu/10.1080/1941126X.2014.877332

Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science*, *38*(2), 189-200.

Flanagan, M., & Carini, P. (2012). How games can help us access and understand archival images. *The American Archivist*, *75*(2), 514-537.

Fleet, C., Kowal, K. C., & Pridal, P. (2012). Georeferencer: Crowdsourced georeferencing for map library collections. *D-Lib magazine*, *18*(11/12).

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine*, *14*(6), 1-4.

Lascarides, M., & Vershbow, B. (2014). What's on the Menu?: Crowdsourcing at the New York Public Library. *Crowdsourcing Our Cultural Heritage*, 113-1137.

Lang, A. S., & Rio-Ross, J. (2011). Using Amazon Mechanical Turk to transcribe historical handwritten documents. *Code4Lib Journal*, (15).

Library of Congress. Prints and Photographs Division, Springer, M., Dulabahn, B., Michel, P., Natanson, B., Reser, D. W., ... & Woodward, D. (2008, October). For the common good: The Library of Congress Flickr pilot project. Library of Congress, Prints and Photographs Division.

Mathes, A. (2004). Folksonomies-cooperative classification and communication through shared metadata.

Munyaradzi, N., & Suleman, H. (2014). A system for high quality crowdsourced indigenous language transcription. International Journal on Digital Libraries, 14(3-4), 117-125.

Parilla, L., & Ferriter, M. (2016). Social media and crowdsourced transcription of historical materials at the Smithsonian institution: Methods for strengthening community engagement and its tie to transcription output. American Archivist, 79(2), 438-460. doi:10.17723/0360-9081-79.2.438

Rafferty, P., & Hidderley, R. (2007). Flickr and democratic indexing: Dialogic approaches to indexing. Aslib Proceedings, 59(4/5), 397-410. doi:10.1108/00012530710817591

Reese, J. S. (2016). Transcribing the Past: Crowdsourcing Transcription of Civil War Manuscripts. Archival Issues: Journal of the Midwest Archives Conference, 37(2), 59–74. Retrieved from http://search.ebscohost.com.libproxy.lib.unc.edu/login.aspx?direct=true&db=lls&AN=132914026&site=ehost-live&scope=site

Saxton, G. D., Oh, O., & Kishore, R. (2013). Rules of crowdsourcing: Models, issues, and systems of control. *Information Systems Management*, *30*(1), 2-20.

Schwartz, M. (2012). Dicing data at NYPL labs. *Library Journal, 137*(14), 22-n/a. Retrieved from http://libproxy.lib.unc.edu/login?url=https://search-proquest-com.libproxy.lib.unc.edu/docview/1197603530?accountid=14244

Shirky, C. (2005), "Ontology is overrated: categories, links, tags", available at: http://shirky.com/ writings/ontology_overrated.html (accessed 6 November 2019).

Wildemuth, B. M. (2017). Applications of Social Research Methods to Questions in Information and Library Science. Santa Barbara, C.A.: Libraries Unlimited.

Appendix A: Recruitment email

Hello,

I am a student in the MSLS program at the University of North Carolina at Chapel Hill conducting research about online crowdsourcing projects and the reasons why archives and other cultural heritage institutions choose to take on such projects.

If you've helped manage a project that uses volunteers to describe, transcribe, annotate, or curate materials online, I'd appreciate the chance to speak with you. The interview will take approximately 20 to 30 minutes. Participation is voluntary and the results are anonymous. No individual subject or personal identifying information will be shared.

Please feel free to share this message with others who have experience with these types of projects.

Thank you!

Caroline Waller

Appendix B: Interview Guide

1. Can you describe your past and/or current involvement with crowdsourcing projects?
    a. What was your job title at the time of this project? What were your typical duties in this position?
    b. What was the nature of this project(s)?
    c. What was your role in the project?
2. Why did you or the project team choose to embark on a crowdsourcing project?
3. Were there any specific factors or circumstances you or your project team had to consider before adopting such a project? In particular, consider funding, personnel, or other practical concerns.
4. Do you believe your project was a success? If so, what factors do you think helped contribute to its success? If it was not, what factors impeded the project's success?
5. If you could do the project again, would you make the same decisions? Would you do anything differently?
6. Do you have any advice for other archivists or librarians who want to incorporate crowdsourcing elements into their projects?