Crowdsourcing Transcriptions of Archival Materials

Aaron G. Noll
UMB History Department
HIST 676
Fall 2012

Introduction

In recent years, users of archives have increasingly demanded greater accessibility of historical records through digitization. The capability of accessing information and viewing documents online is becoming a norm in 21st century society, rather than an exception. This increased demand is evident even as funding for projects that increase the accessibility of historical documents through digitization and transcription is falling. In the United States, a key funding source for such projects is the National Historical Publications and Records Commission (NHPRC), a branch of the National Archives and Records Administration (NARA). From 2010 to 2012, NHPRC's budgets fell from $13 million to $5 million. For FY2013, the Obama Administration has proposed cutting the NHPRC budget to $3 million.[1] The NHPRC provides grants to archival institutions attempting to preserve and increase access to documents significant to the history of the United States.[2] Each budget cut therefore impacts the ability of U.S. archival institutions to acquire the technology and retain staff necessary for increasing the accessibility of their holdings through digitization and transcription.

While the recent economic recession is partially responsible for falling NHPRC budgets, another cause may be a general lack of public understanding about the content and uses of archives. Archivists have responded to this problem during the past two decades by boosting their public outreach efforts. These efforts have ranged from establishing an annual Archives Week to presentations at professional conventions. John J. Grabowski argues that the best solution lies in creating more users of archives "on the assumption that someone who uses the

---

[1] National Humanities Alliance, "National Historical Publications and Records Commission FY2013," National Humanities Alliance. http://www.nhalliance.org/advocacy/funding-priorities/national-historical-publications-records-commissio-2.shtml (accessed February 21, 2013) Fiscal Year 2013 began in October 2012 and will end in September 2013. At this time, the U.S. Congress has not yet passed a budget for FY2013.

[2] National Historical Publications and Records Commission, "Eligibility," National Archives and Records Administration, http://www.archives.gov/nhprc/apply/eligibility.html (accessed February 21, 2013).

product is more likely to value the industry that produces it."[3] He contends that more archives users can be created by publicizing and tailoring archive services to genealogists and minorities who have often been missing from dominant historical narratives. Grabowski also supports efforts by archivists to engage high school and college students with primary source materials.

Some archival institutions have proposed crowdsourcing transcriptions of historical records and manuscripts as a solution to both the problem of constrained funds and the need for greater public engagement with archives. According to Wikipedia, crowdsourcing is

> a process that involves outsourcing tasks to a distributed group of people. This process can occur both online and offline. The difference between crowdsourcing and ordinary outsourcing is that a task or problem is outsourced to an undefined public rather than a specific body, such as paid employees.[4]

In this paper, "crowdsourcing" will refer exclusively to the online aspect of the above definition. Online crowdsourcing utilizes some form of Web 2.0 technology to allow interaction between archivists and volunteers. In the context of archival work, crowdsourcing has been used for a variety of processing tasks that either describe archival materials at varying levels of detail or extract information from archives for database input.

Several archives have crowdsourced the creation of metadata that describes historical documents. "Folksonomy" refers to the practice of crowdsourcing the jobs of tagging and indexing both archival documents such as photographs and library items.[5] The New York Public

---

[3] John J. Grabowski, "Keepers, Users, and Funders: Building an Awareness of Archival Value," *The American Archivist* 55 (Summer 1992), http://archivists.metapress.com/content/r9gmpj0rj8530811/?p=45a04f86615f42dd974acdd9239b28d0&pi=0 (accessed February 21, 2013)

[4] Wikipedia, "Crowdsourcing." Wikimedia Foundation, Inc. http://en.wikipedia.org/wiki/Crowdsourcing (accessed November 17, 2012).

[5] The National Archives' Citizen Archivist Dashboard and the University of Iowa's DIY History website are two examples of archival institutions that use folksonomy. For more on folksonomy, see Eric Willey. "A cautious partnership: The growing acceptance of folksonomy as a complement to indexing digital images and catalogs."*Library Student Journal* 15. *Library, Information Science & Technology Abstracts*, EBSCO*host* (accessed February 21, 2013).

Library's (NYPL) Map Division and NYPL Labs have created NYPL Map Warper, a

crowdsourcing website that "warps" or "overlays" digitized historical maps on to online

contemporary maps.[6] Volunteers orient historical maps by matching historical locations with

contemporary locations on current maps. This work makes it possible to locate historical

landmarks that no longer exist and to see how a landscape has changed over time.[7]

Other archives have crowdsourced document transcriptions. Transcribed information is

used either as a simple representation of the document itself or as selected data that is loaded into

a database. There are several robust projects in the field of scientific archives that crowdsource

data transcription for the creation of large databases. For example, the Old Weather project has

its volunteers transcribe weather data from ship logs dating to the 18th century. The data is

entered into a database that allows scientists to study historical weather conditions.[8] Old Weather

is hosted by Zooniverse, a major crowdsourcing metasite maintained by the Citizen Scientist

Alliance.[9] However, none of these scientific projects have the objective of providing access to

original records. Users of this transcribed information are scientists who work with the databases

that are produced.

Many crowdsourced transcription projects are created in order to generate descriptive

information that provides access to original historical documents. A wide spectrum of languages,

historical period documents, and geographic areas are represented by this type of project. The

---

[6] New York Public Library. "NYPL Map Warper." New York Public Library. http://maps.nypl.org/warper/
(accessed February 21, 2013).

[7] For more information on crowdsourcing metadata of historical maps, see Kimberly C. Kowal and Petr Pridal,
"Online Georeferencing for Libraries: The British Library Implementation of Georeferencer for Spatial Metadata
Enhancement and Public Engagement," *Journal of Map & Geography Libraries* 8 (September 2012): 276-89.

[8] Citizen Science Alliance, "Old Weather: Our Weather's Past, the Climate's Future," Zooniverse.
http://www.oldweather.org/ (accessed February 21, 2013).

[9] Citizen Science Alliance, "Zooniverse: Real Science Online," Zooniverse, https://www.zooniverse.org/ (accessed
February 21, 2013).

Leuven City (Belgium) Archive plans to crowdsource the transcription of more than 950,000

Dutch-language register pages from the Leuven court of aldermen during the years 1362 to

1795.[10] The Ancient Lives[11] and Carolingian Canon Law[12] projects seek online volunteers to

transcribe ancient Greek papyri and medieval Latin-language law manuscripts, respectively.

 Some crowdsourced transcription projects use optical character recognition (OCR) tools

to assist online volunteers in transcribing typed documents.[13] The National Library of Australia

is leading the Australian Newspapers Digitisation Program (ANDP) with OCR and

crowdsourcing. The ANDP's goal is for the public to help transcribe forty million newspaper

articles that cover 4.4 million pages.[14] The OCR tool first transcribes the article from a digital

image, then a volunteer needs to correct the mistakes.

 This paper will examine the crowdsourced transcription of typed and handwritten

historical documents without the use of OCR. The focus will be on English-language projects of

archives located primarily in the United States with the exception of Transcribe Bentham, a

project out of University College London. These projects will be analyzed on the basis of three

major principles or areas of archival work: 1) the transcription phase of processing archives, 2)

accessibility, and 3) outreach. Preservation, another major archival principle, looms large in this

[10] Leuven City Archive, "Itinera Nova," Leuven City Archive, http://www.itineranova.be/in/home (accessed February 21, 2013). For information on another European crowdsourcing transcription project, see Cimtech, "Crowdsourcing Project Helps Finnish Library to Digitise Historical Documents," *Information Management & Technology (IM@T.Online),* (March 2011), http://www.imat.cimtech.co.uk.ezproxy.lib.umb.edu/Pages/IM@T_Online/2011/February-March/IMAT_0211_NewsTrack_11.htm (accessed February 21, 2013).

[11] University of Oxford, "Ancient Lives: Help Us to Transcribe Papyri," Zooniverse, http://ancientlives.org/ (accessed February 21, 2013).

[12] University of Kentucky Department of History, "Carolingian Canon Law Project," University of Kentucky, http://ccl.rch.uky.edu/ (accessed February 21, 2013).

[13] Tobias Blanke, Michael Bryant, and Mark Hedges, "Open Source Optical Character Recognition for Historical Research," *Journal of Documentation* 68 (September 2012): 659-83.

[14] Holley, Rose. *Many Hands make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers.* Canberra: National Library of Australia, 2009. http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/1406 (accessed February 21, 2013).

discussion despite the fact that little mention has been made of it in the nascent literature on crowdsourcing archival work. The importance of crowdsourced transcriptions to preservation will be briefly noted. This paper is not intended to offer best practices or advice on how to organize a successful crowdsourced transcription project, although some of its content bears on these subjects. Rather, it is an analysis of this field of archival activity through the categories of major archival principles.

Overview of Projects to be Discussed

The following projects will be discussed in this paper. They will each be introduced shortly here, then further described and analyzed through each of the major archival principles cited above. Each project was chosen for its unique characteristics that contribute to a rich discussion of the topic of crowdsourced transcriptions.

1) The World Archives Project[15] was created by the genealogical research website Ancestry.com in August 2008. A wide spectrum of documents of genealogical importance are placed on the Project website at any given time. A recent selection included prison records, World War II-era Jewish ID Card Applications from Poland, and Mexican War muster rolls of the U.S. Army. During October 2012 alone, 2.7 million records were transcribed.[16] It should be noted here (and will be explained below in more detail) that this project does not demand full transcriptions of documents, only transcriptions of key details that are then entered into an indexing database.

---

[15] Ancestry.com, "World Archives Project," Ancestry.com, http://community.ancestry.com/awap (accessed February 21, 2013). A similar "indexing" project is hosted by FamilySearch.org, another major genealogical research website. FamilySearch.org Indexing approaches their task in much the same way as Ancestry.com; they require volunteers to download their indexing software.

[16] Anna Fechter. "October was A Great Month!," Ancestry World Archives Project Blog, entry posted November 6, 2012, http://blogs.ancestry.com/worldarchivesproject/?p=1584 (accessed February 21, 2013).

2)   The Edgerton Digital Collections Project[17] is sponsored by the Edgerton Center at the

Massachusetts Institute of Technology. Dr. Harold Edgerton was a popular MIT professor who

left behind a handwritten notebook of 8,400 pages when he passed away in 1990. The Edgerton

Center digitized the entire notebook and, beginning in 2009, began crowdsourcing transcriptions

of it.

3)   The What's on the Menu?[18] project, created by the New York Public Library's Menu

Collection archivists and NYPL Labs, seeks the help of volunteers in transcribing 40,000 historic

New York restaurant menus. The collection contains both printed and handwritten menus. From

the project's launching in 2011 to February 2013, 16,489 menus were transcribed.

4)   The Frederick Douglass Diary: Written Rummage Project[19] was organized by Andrew

S.I.D. Lang and Joshua Rio-Ross, both of Oral Roberts University. They used Amazon.com's

Mechanical Turk website to hire transcribers for a 72-page diary of Douglass's that is held by the

Library of Congress. The project was begun and completed in 2011.

5)   The Genealogy Vertical Files Transcription Project[20] is being run by the North Carolina

Government and Heritage Library's genealogy research room. For years, they have collected

copies of the research of genealogists using the Library. In 2011, they began crowdsourcing the

transcription of these copies.

---

[17] The Edgerton Center. "The Edgerton Digital Collections Project." Massachusetts Institute of Technology. http://edgerton-digital-collections.org/ (accessed February 21, 2013).

[18] New York Public Library Labs, "What's on the Menu?," The New York Public Library, menus.nypl.org (accessed February 20, 2013).

[19] Andrew S.I.D. Lang and Joshua Rio-Ross, "Frederick Douglass Diary: A Written Rummage Project," Wikispaces, http://frederickdouglassdiary.wikispaces.com/ (accessed February 21, 2013).

[20] Government & Heritage Library, State Library of North Carolina, "Genealogy Vertical File Transcription Project," State Library and State Archives of North Carolina, http://statelibrary.ncdcr.gov/digital/ncfamilyrecords/verticalfiles.html (accessed February 21, 2013).

6)   The University of Iowa's DIY History[21] website is a revamped version of the previous Civil War Diaries and Letters Transcription Project which, at more than 15,000 manuscript pages transcribed by online volunteers, is nearly complete. The latter project was begun in the spring of 2011. Beginning in October 2012, the new website has crowdsourced the tagging of historical photographs and the transcription of a 20,000 (and growing) item collection of handwritten primary sources in culinary history.

7)   The Papers of the War Department[22] project was created by the Center for History and New Media at George Mason University. A fire on November 8, 1800 destroyed all the records of the U.S. War Department dated from 1784 to 1800. The Center for History and New Media has digitized 55,000 of the documents that were originally part of the collection from duplicates that are held in more than 200 archives around the world. As of January 2013, 1,559 of the documents have been transcribed by 1,205 online volunteers.[23]

8)   Transcribe Bentham[24] was created by the Bentham Project at University College London (UCL) in September 2010. Since 1958, the Bentham Project has been working to transcribe and publish *The Collected Works of Jeremy Bentham*, a planned seventy-volume set that will contain all 60,000 pages of Bentham's writings now held in the UCL archives. 20,000 pages have been

---

[21] University of Iowa Libraries, "DIY History," University of Iowa, http://diyhistory.lib.uiowa.edu/ (accessed February 21, 2013).

[22] Center for History and New Media, "Papers of the War Department," Center for History and New Media, http://wardepartmentpapers.org/ (accessed February 21, 2013).

[23] Center for History and New Media, "Community Transcription Update-Twenty One Months On," Papers of the War Department Blog, entry posted January 24, 2013, http://wardepartmentpapers.org/blog/?p=1271 (accessed February 20, 2013)

[24] The Bentham Project, "Transcribe Bentham," University College London, http://blogs.ucl.ac.uk/transcribe-bentham/ (accessed February 21, 2013).

transcribed and published to date, mostly by a professional editorial staff. Since Transcribe Bentham was launched, 4,408 pages have been transcribed by 2,171 online volunteers.[25]

9)   The National Archives Transcription Pilot Project[26] is just one crowdsourcing component of the NARA's Citizen Archivist Dashboard. It contains more than 300 handwritten and typed historical documents ranging from the 18th to 20th centuries. They include "letters to a civil war spy, presidential records, suffrage petitions, and fugitive slave case files."[27]

Transcribing Digitized Archival Materials

A transcription of a document constitutes the greatest level of attention that an archivist can provide to a handwritten document's content. Most archival materials never receive this level of attention. Writing a description of some aggregate of documents at the folder, box, or collection levels is how most materials are processed by archivists. Transcription, particularly of digitized documents, permits a level of access to a handwritten document's content that cannot be improved upon. When a digitized handwritten document is transcribed, its content (now contained by a database) becomes keyword searchable.

But widespread transcription of handwritten documents is a processing standard that cannot be met by archivists. Many archivists have difficulty in processing their materials at even a basic descriptive level due to the large quantities of materials they must handle. This is evident in the substantial processing backlogs that have built up in many institutions. These backlogs have caused some archivists to reassess their approaches to processing. Mark Greene has

---

[25] The Bentham Project, "Transcribe Bentham: About Us," University College London, http://blogs.ucl.ac.uk/transcribe-bentham/about/ (accessed February 21, 2013).

[25] The volunteers total was given by Tim Causer, a Bentham Project Research Associate, in an email to the author on November 18, 2012.

[26] U.S. National Archives and Records Administration, "Citizen Archivist Dashboard," U.S. National Archives and Records Administration, http://www.archives.gov/citizen-archivist/ (accessed February 21, 2013).

[27] Ibid.

proposed a "More Product, Less Process" (MPLP) processing model.[28] He favors a revolutionary rethinking of processing in order to make more archival materials accessible to users. Crowdsourcing archival work is an idea that originates in the same vein of assertive reassessment of traditional archives processing that Greene's ideas can be located in. Both solutions are grounded in the assessment that traditional processing models are not meeting current needs and that bold, new approaches should be considered.

The objectives that an archivist has for how the information derived from transcriptions is to be utilized determines what software or transcription tools will be used and what level of intellectual control is exercised over transcribed content. For example, a genealogist is primarily interested in the capability to search a collection of digitized documents for names of ancestors. In this case, once the desired document image has been located, a genealogist can read the handwriting for information. On the other hand, an academic researcher may wish to keyword search the entire text of a document collection. A high degree of intellectual control is desired in this case since transcriptions may be quoted in research. In this section, I will examine the variety of approaches to transcription that the above listed projects have exemplified. Software and transcription tools will be discussed as well as the major archival processing issues of intellectual control or accuracy, volunteers' knowledge and education, and the cost effectiveness of crowdsourced transcription.

Crowdsourced Transcription Tools

Crowdsourced transcription tools may be free-form, XML or HTML code-based (i.e., "marked up"), or oriented toward filling database fields with information culled from records

---

[28] Mark A. Greene and Dennis Meissner, "More Product, Less Process: Revamping Traditional Archival Processing," *American Archivist* 68, no. 2 (Fall-Winter 2005): http://www.jstor.org/stable/40294290 (accessed February 21, 2013).

(i.e., indexing).[29] Most tools now being used have a free-form format. They position a digital image of a manuscript on the screen with an adjacent text box. The transcription is pure text with only the letters and symbols of a standard computer keyboard available to express the handwritten text. One of the simplest free-form tools is the comments box on a Flickr image page. This is the tool used by the Genealogy Vertical Files Transcription Project.[30] One difficulty with this method for documents relevant to genealogy is that it cannot easily accommodate varying document formats. For example, a family tree or a table form cannot be easily transcribed. This problem could be mediated by the provision of standard procedures for dealing with commonly seen document formats. For example, the project might insist that volunteers transcribe not just the handwriting filling out the form, but also the typed content as well. In fact, since very little transcription guidance has been given in this project, transcribers have taken a variety of approaches with some transcribing only handwriting and others choosing a combination of handwriting and some or all typed content.

Another crowdsourced transcription project that uses a free-form tool is the DIY History project. Since the majority of the documents to be transcribed are completely handwritten, the problems of the Genealogy Vertical Files Transcription Project are avoided. However, other common difficulties of transcription are present. For example, how is the transcriber to handle crossed out writing, marginal notes, and line and paragraph breaks? How should the transcriber note or mark words that are illegible? DIY History provides guidance on these questions. "No need to account for formatting (e.g., spacing, line breaks, alignment); the goal is to provide text

---

[29] This distinction and related terms were made by Ben Brumfield, "What does it Mean to "Support TEI' for Manuscript Transmission?" Collaborative Manuscript Transmission.Blogspot, entry posted November 10, 2012, http://manuscripttranscription.blogspot.com/2012/11/what-does-it-mean-to-support-tei-for.html (accessed February 21, 2013)

[30] Government & Heritage Library, State Library of North Carolina, "Family History Transcription Project," Flickr, http://www.flickr.com/photos/statelibrarync/sets/72157627124710723/ (accessed February 21, 2013).

for searching" appears above the text box on all pages to be transcribed. Further, a link to "transcription tips" is given for additional standard procedures for common issues. These guidelines help to establish consistency within the DIY History collections of transcribed documents for how a particular transcription problem is to be handled. They exert a minimal level of intellectual control in the sense that users of transcribed material gain more accurate knowledge of content by knowing how a particular issue has been consistently dealt with throughout the collection.

A free-form transcription tool might be chosen by a crowdsourced transcription project for at least two reasons. The first is that the simplicity of the tool requires comparatively small start-up costs for the technical aspects of the project. Using the comments box in Flickr as a transcription tool requires only the cost of digitizing documents and uploading them to Flickr. One initial barrier to the Civil War Diaries and Letters Transcription Project at the University of Iowa Libraries was the development by the systems staff of a web-based transcription tool. The systems staff believed that they did not have the sufficient time or expertise to develop such a tool. They eventually decided to begin their project with webpages composed of the digital image to be transcribed and a simple text web form.[31] Secondly, a free form transcription tool is considered to be more user-friendly and less technically demanding to potential volunteers. Free-form transcription is an example of a "What You See Is What You Get" (WYSIWYG) interface; the text that is immediately visible in the transcription box is identical to the text that is displayed in the final

---

[31] Nicole Saylor, "Experimenting with Strategies for Crowdsourcing Manuscript Transcription," *Research Library Issues* 277 (December 2011): 9-14.

Illustration 1: This image shows the DIY History project's free-form transcription tool. The digitized image is at the top. Below it are some brief guidelines for volunteer transcribers. The text box for the transcription is at the bottom.[32]

product utilized by the archival institution's database. Although Transcribe Bentham's academic objectives require marked-up transcriptions, project organizers have found through surveys of their volunteers that the use of code within the transcriptions' text (as opposed to "hiding" codes in a WYSIWYG interface) intimidates and discourages some potential volunteers.[33]

---

[32] University of Iowa Libraries, "Bean Family Letters, 1862-1863," http://diyhistory.lib.uiowa.edu/transcribe/scripto/transcribe/99/5452 (accessed November 23, 2012).

[33] Tim Causer and Valerie Wallace, "Building A Volunteer Community: Results and Findings from *Transcribe Bentham*," *Digital Humanities Quarterly* 6, no.2 (2012), http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html (accessed February 21, 2013).

A marked-up transcription produces a more powerful product for scholars to research. For this reason, Transcribe Bentham decided to use a more complex tool for its crowdsourced transcriptions. Transcribe Bentham intends to not only make transcriptions available in a database for scholars, but also to publish them in new volumes of *The Collected Works of Jeremy Bentham*. The project's Transcription Desk tool, developed by the UCL Computer Centre, features code solutions to "line breaks, page breaks, headings, and paragraphs; linguistic features like notes, unusual spellings, and foreign-language text; compositional features such as additions and deletions; and interpretive decisions about questionable readings and illegible text."[34] These codes are in compliance with the Text Encoding Initiative, an increasingly accepted set of standards for online scholarly texts.[35] However, volunteers do not need to know or memorize the appropriate codes for each of these text forms. The codes can be inserted by highlighting some piece of transcription text, then pressing the appropriate button on the toolbar. For example, if the word "tables" was written, but then crossed out, the correct transcription of it is: ~~tables~~. "Tables" can be typed into the transcription box and highlighted; then the "deletions" toolbar button can be pushed to produce the marked-up transcription above.

Transcribe Bentham has lessened the technical impact of marked-up transcriptions on volunteers in several ways. The first is the code toolbar that appears at the top of every transcription text box. Secondly, the mark-ups are only slightly modified from those used in MediaWiki, the tool used by volunteer editors of Wikipedia. The similarities between MediaWiki and the Transcription Desk may encourage Wikipedia editors to also volunteer for Transcribe Bentham. Transcribe Bentham orients its volunteers with tips for transcribing

---

[34] Tim Causer, Justin Tonra, and Valerie Wallace, "Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *the Collected Works of Jeremy Bentham*," *Literary and Linguistic Computing* 27, no. 2 (2012): 119-137.

[35] Text Encoding Initiative Consortium, "Text Encoding Initiative," Text Encoding Initiative, http://www.tei-c.org/index.xml (accessed February 21, 2013).

Bentham's difficult handwriting and a simplified list of basic mark-up codes for the most frequently seen text features (a longer list of rarely seen features and corresponding codes is also available). Finally, a discussion board provides a forum for volunteers and project moderators to address unusual transcription difficulties.

Not all crowdsourced transcription projects require whole page or full-text transcriptions, as with the above projects. Some archivists are concerned with making only specific types of a document's data searchable by users; whole document images are viewable, but they can only be accessed or discovered through the data categories that have been transcribed or indexed. Ancestry.com's World Archives Project asks volunteers to transcribe only information such as names, dates, and locations. Transcriptions are entered into fields pre-set for data types that vary with type of document being indexed. This is the only project examined in this paper that does not use a web-based transcription tool. Volunteers need to download the "Keying Tool", a highly polished and efficient tool that Ancestry.com clearly spent significant time and funds to develop.

Ancestry.com's Keying Tool has many features that assist volunteers in producing accurate transcriptions at an efficient pace. The scanned document image covers the top half of the screen. On the bottom half is a data table in which transcriptions are entered. The data fields change depending on the collection the user has chosen to transcribe in and on the form of the document to be transcribed. Before opening a new collection for volunteers to transcribe, Ancestry.com selects twenty volunteers to participate in a Contributor Testing Program for that collection. These groups of volunteers enable project administrators to determine the patterns with which the wider volunteer community will interact with the types of records in that collection. In this way, when the Keying Tool is opened for a collection, advice that is collection, form, and field-specific will appear in the bottom left corner of the screen when a volunteer

clicks on a data field to enter a transcription there. If the field is for a document date, the advice

may indicate a particular area of the document where a date can be found; or if there are several

dates on the document, the advice box will indicate which date should be transcribed.

Ancestry.com supports its volunteers in many other ways. Transcription tips are easily

accessible from the toolbar in the Keying Tool. If a volunteer needs it, a box with four samples

of various ways in which each letter has been written in the past will appear above the writing to

be transcribed. The box contains a further link to more advice in reading antiquated handwriting.

Once a volunteer begins typing a transcription into a data field, a list of word possibilities

appears that is based on other words beginning with the same entered letters that are frequently

transcribed in that particular collection. Document images can be manipulated in multiple ways

to assist in reading handwriting. Images can be rotated, have their contrast adjusted, and

magnified; the Keying Tool permits more image manipulation than any other transcription tools

examined in this paper. Finally, an active message board of volunteers assists in answering

remaining questions.[36]

The New York Public Library's What's On The Menu? project also uses a database

field-based transcription tool. This project's objective for volunteers is the extraction by

transcription of dishes on historical restaurant menus. The unique contribution of this project to

this paper's discussion is its provision of marked-up digital images to specify what content is to

be transcribed. The text of each dish on the menu is marked as in need of transcription; other

text, such as prices and dates, are not marked. Once a single dish's transcription is complete,

NYPL Lab's transcription tool marks it with a checkmark. Although this particular tool was

---

[36] Ancestry.com, "World Archives Project."

developed by NYPL Labs and is not open source, marking the digital image is one technique

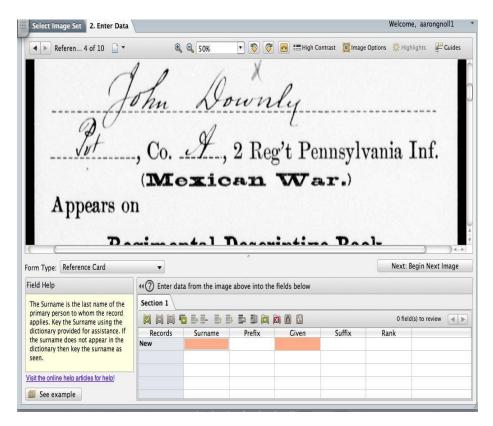available to archivists to ensure that the correct text is transcribed.

Illustration 2: Ancestry.com's Keying Tool used for its World Archives Project.

Illustration 3: An image of a menu with completed transcriptions in the What's On The Menu? project at the New York Public Library.[37]

Intellectual Control

A significant issue inherent in crowdsourcing the transcription of archival materials is the accuracy of the work. How do archivists maintain intellectual control over transcription content when, by definition, transcribers in this type of project are not required to have any sort of expertise, background knowledge, or experience in either transcription or the subject of the materials? This problem threatens to undermine crowdsourcing as a method for transcriptions. An initial response to the problem was given by the University of Iowa Libraries staff's consensus opinion (in the early stages of planning for the Civil War Diaries and Letters Transcription Project) that "some imperfect access was better than none, and any staff-side inefficiencies would be worth the trade-off in public outreach benefits."[38] In other words, the benefits of increased accessibility to documents offered by transcription and the public outreach intrinsic to crowdsourcing outweigh the negative prospect of some inaccurate transcriptions and project oversight needed by staff to correct these.

One corrective measure taken by project administrators to raise accuracy levels is supporting volunteers through transcription help offered within the tools themselves. Many of these have been described in the above section on transcription tools. Some crowdsourced transcription projects such as the World Archives Project offer well-used discussion forums for volunteers to communicate their difficulties and thoughts with one another. These are all methods for reducing transcription inaccuracies from the start. The remainder of this section will

---

[37] New York Public Library Labs, "What's on the Menu?."
[38] Saylor, "Experimenting with Strategies."

discuss possibilities for correcting inaccuracies after an initial transcription has already been accomplished.

The effort and resources that are devoted by project administrators to correcting inaccurate transcriptions should correspond to the objectives of the project. For Transcribe Bentham, whose objective is scholarly publication of their transcriptions, expert moderation of crowdsourced transcriptions is a necessity.[39] The Genealogy Vertical Files Transcription Project, on the other hand, is transcribing secondary genealogical sources. The transcriptions will remain in an online database for genealogists to keyword search. The nature of the documents, their accessibility, and their intended users all play a role in assessing what level of attention should be given to inaccurate transcriptions.

Another variable in the effort to have volunteers produce accurate transcriptions is whether the transcriptions will be "locked" at some planned point in time. When Transcribe Bentham moderators determine that "no appreciable improvements could be made to the transcript through further crowdsourcing", the transcription is locked.[40] It is then saved in a searchable database of Bentham's writing and sent to the editors of *The Collected Works of Jeremy Bentham.* In contrast, the administrators of the Papers of the War Department project have no intention of publishing their transcriptions. Rather, they will indefinitely remain in a searchable online database for scholars to access. The transcriptions will remain unlocked; i.e., if a future user of the database discovers an error, the text box containing the transcription will

---

[39] Transcribe Bentham retains two full-time Research Associates. One of their main priorities is "moderation of the transcripts submitted by volunteers." Tim Causer, et al. "Transcription Maximized."
[40] Ibid.

remain editable.[41] While transcription errors may exist, it is hoped that as time passes and transcriptions remain editable, they will trend toward the reduction of inaccuracies.

Some project administrators keep their transcriptions unlocked even after the document has been transcribed. To increase accuracy, they intend to have the transcription "double-keyed" or "triple-keyed." Using this terminology, if a document has only had an initial transcription, then it has been "single keyed." The other terms refer to a document being checked for accuracy by volunteers other than the original transcriber. Some project administrators that plan to eventually lock their transcriptions have a policy specifying a minimum number of times a transcription should be checked by volunteers before locking it. The World Archives Project, Transcribe Bentham, and the National Archives Transcription Pilot Project all use double and triple-keying to improve accuracy before locking a document transcription.[42]

If a low level of staff oversight for the insurance of accuracy exists, then it is critical that the original scanned document image remain available to users along with the transcription. The unavailability of original images would ensure that erroneous transcriptions become semi-permanent. One could easily imagine erroneous quotes being cited online or in published material, progressively achieving a kind of immortality as it would become increasingly difficult to dislodge and replace an incorrect quote with a correct quote. Fortunately, in all of the projects reviewed in this paper, it is possible to view original images. However, a related but more troublesome question can be raised: How can a user access a document in which key terms have

---

[41] Email to the author from Christopher Hamner, Principal Investigator & Editor-in-Chief of *The Papers of the War Department*, November 16, 2012.

[42] Crista Cowan,"Keying during the Holidays," Ancestry World Archives Project Blog, entry posted December 26, 2009, http://blogs.ancestry.com/ancestry/2009/12/22/ancestry-world-archives-project-keying-during-the-holidays/ (accessed February 21, 2013); Tim Causer et al. "Transcription Maximized"; Email to the author from Meredith Stewart, Management and Program Analyst, Office of Innovation, U.S. National Archives and Records Administration, November 13, 2012.

42

been erroneously transcribed? In other words, a user who correctly spells a search term would be unable to discover a document for which the related transcription contains a misspelled version of the term. It is not difficult to imagine documents remaining in prolonged obscurity due to a single transcription error.

Volunteers' education, subject knowledge, and transcription experience all impact the accuracy of transcriptions. Critics of crowdsourced transcription as an archive processing method fear that historical documents will be transcribed by inexperienced volunteers with poor historical knowledge. One critic is Edward G. Lengel, editor-in-chief of the Papers of George Washington, a traditional transcription/editing project with the objective of publishing an eighty-eight volume set of Washington's writings; sixty-four volumes have been completed to date.

> Mr. Lengel's staff comes to the work with deep training in how to recognize different types of handwriting, determine the dates of documents, and situate the materials in context. They know which record to use, for example, when confronted with multiple copies of a document, like a draft, a letterbook copy (people used to copy their letters into bound books of blank pages), and a receiver's copy.[43]

The skills of the "crowds" pale in comparison to those of the professionals described above. It can be granted that some of the above skills, such as distinguishing between multiple copies of a document, cannot be left to volunteers in a crowdsourcing project. But there is perhaps some misapprehension among critics of precisely what kinds of people volunteer for crowdsourced transcription projects. The gap in transcription capability between Mr. Lengel's staff and crowdsourcing volunteers may not be quite as wide as is popularly believed.

The majority of the work in crowdsourced transcription projects is completed by about 10% of the volunteers. The people of this minority are sometimes referred to as "super-

---

[43] Marc Parry, "Historians Ask the Public to Help Organize the Past; but is the Crowd Up to it?" *The Chronicle of Higher Education* 59, no. 2 (September 3, 2012), http://go.galegroup.com.ezproxy.lib.umb.edu/ps/i.do?id=GALE%7CA302048583&v=2.1&u=mlin_b_umass&it=r&p=AONE&sw=w (accessed February 21, 2013).

volunteers." Rose Holley of the National Library of Australia reports from multiple project surveys that this volunteer minority tend to "be a mix of retired people and young dynamic high achieving professionals with full-time jobs."[44] Transcribe Bentham administrators have found that 97% of its volunteers have at least an undergraduate degree, with 24% holding a doctorate. However, only a small minority have experience in working with manuscripts or reading handwriting.[45] A Papers of the War Department blogger notes that "our transcribers truly represent a cross-section of life: we have high school teachers, librarians, demographers, doctoral candidates, journalists, historical re-enactors, CEOs, and many other kinds of folks transcribing."[46] While volunteers may not have the experience and subject expertise of professional transcribers, they do appear to be capable of producing generally reliable transcriptions.

Moreover, with a carefully planned publicization strategy, it seems possible to influence volunteer demographics for a project. The Edgerton Digital Collections Project publicized its project primarily through two outlets: 1) a display at the MIT Museum, and 2) an article in the MIT alumni magazine. This strategy likely produced a scientifically knowledgeable group of volunteers that was capable of interpreting Dr. Harold Edgerton's notebooks at (at least) an elementary level. In fact, Suzana Lisanti, Edgerton Digital Collections Project Coordinator, has noted that several volunteer transcribers for her project are Dr. Edgerton's former students.[47]

---

[44] Rose Holley, "Crowdsourcing: How and Why should Libraries do it?" *D-Lib Magazine* 16, no. 3 (March 2010), http://www.dlib.org/dlib/march10/holley/03holley.html (accessed February 21, 2013).

[45] Causer, "Building a Volunteer Community."

[46] Center for History and New Media, "Community Transcriptions-18 Months and 1000 Transcribers," Papers of the War Department Blog, entry posted on October 25, 2012, http://wardepartmentpapers.org/blog/?p=1208 (accessed February 21, 2013).

[47] Suzana Lisanti, "Innovation in the Archives." (Panel Discussion at the Fall Conference of New England Archivists, Simmons College, Boston, MA, November 3, 2012).

In contrast, a poor method of recruiting volunteers for transcription would be publicizing the project among people who are likely to have no interest in the archival materials to be transcribed. This was the chosen method for publicizing the Frederick Douglass Diary: Written Rummage Project, which used Amazon.com's Mechanical Turk website to pay transcribers a small rate for each page completed. There is no particular reason that those who are interested in Frederick Douglass or care about transcription of archival materials would visit the Mechanical Turk website. This appears to be the best explanation for the poor transcription of Douglass's diary produced by the project. The first ten pages of transcriptions contained an average of seven errors per page with one error significantly changing the meaning of a sentence.[48] Therefore, a careful consideration of what kinds of people would make good volunteers for the project should precede its publicization. Publicization should be directed toward groups of people who are most likely to produce accurate transcriptions. Volunteers' attributes can ultimately affect the quality and accuracy of transcriptions.

Critics of crowdsourced transcriptions have suggested that "transcribers' political beliefs [could] skew their work on documents related to sensitive historical topics."[49] This concern could be expanded to a consideration of how transcribers' beliefs in all categories, including religious and social, could affect the accuracy of transcriptions. There has been no mention of any actual cases of transcriber bias in the literature on crowdsourced transcriptions; the problem has so far remained at a hypothetical level. Sharon Leon, director of public projects at the Center

---

[48] This examination of transcriptions was done by the author. Of the first ten pages, two transcription pages are completely missing from the final project page. Of the other eight pages, the average number of errors was seven, with one page having ten errors. As a result of transcription error, a conversation between Douglass and a fellow ship passenger was described as "sharp", not "short", as Douglass had written. See an article written by the project's organizers: Andrew S.I.D. Lang and Joshua Rio-Ross, "Using Amazon Mechanical Turk to Transcribe Historical Handwritten Documents," *Code4Lib Journal* no. 15 (October 31, 2011) http://journal.code4lib.org/articles/6004 (accessed February 21, 2013).

[49] Parry, "Historians Ask the Public"

for History and New Media (creators of the Papers of the War Department project), stated, "Lots of fear about crowdsourcing is, Oh, it's going to be a mess, and people are just going to deface things... We haven't seen a hint of any malicious use."[50] It is a good sign for the method of transcription by crowdsourcing that no projects have yet reported any instances of intentional or unintentional bias. However, this method of transcription is only a few years old. Given the short period during which crowdsourced transcriptions have been tried, it seems too early to disregard the apprehension that some critics have expressed over the potential bias of volunteer transcribers.

A Costs/Benefits Analysis of Crowdsourcing Transcriptions

The major benefit of crowdsourcing transcriptions of archival materials is having a large volunteer workforce to transcribe important historical and genealogical documents. As described in further detail below, these transcriptions result in greater accessibility of archives to researchers and the public. The nature of the projects themselves offer substantial benefits in public outreach and increasing the public's awareness of archives and archival institutions.

But it is important to consider the costs of crowdsourced transcription projects. Some costs relate to the processing that prepares documents for transcription: scanning the original documents, recording minimal metadata on each image, organizing the image files in a database, choosing or designing a transcription tool, and publicizing the project. All of these are fixed in the sense that they must be completed before transcription is to begin. Some costs such as the transcription tool are variable depending on whether a new tool is designed from scratch or if an open-source tool is used.

---

[50] Quoted in Parry, "Historians Ask the Public"

Possibly the most variable cost of all is the staff time required for oversight of transcription accuracy. Some project administrators devote no staff time to checking the accuracy of transcriptions. For example, the Edgerton Digital Collections Project relies solely on volunteers themselves for checking the accuracy of its notebook transcriptions. At the other extreme is Transcribe Bentham, whose administrators calculated that

> had the [Transcribe Bentham] Research Associates been employed purely to transcribe manuscripts on a full-time basis, they could have transcribed about 5,000 manuscripts between them over twelve months, or two-and-a-half times as many as the volunteers would have produced had they continued transcribing at the same rate.[51]

In the latter case and others whose objectives require a high degree of transcription accuracy, one can reasonably ask what factor could persuade project administrators to continue with a plan to crowdsource transcription? Philip Schofield, director of the Bentham Project, notes, "it would be virtually impossible to get significant funding for transcription alone."[52] Schofield is referring here to the fact that the Transcribe Bentham project was initially made possible through a grant from the United Kingdom Arts and Humanities Research Council. The current climate of grant priorities appears to favor archival institutions' engagement with the public through methods such as crowdsourcing. So long as this grant climate persists, it seems possible that the ironic calculation that a crowdsourced transcription project with the high standards of accuracy demanded by Transcribe Bentham is desirable on a costs/benefits analysis despite its inefficiency will continue to be a reasonable one. An assessment of the benefits of accessibility and public outreach that accrue from a crowdsourced transcription project is in the next two sections of this paper.

---

[51] Tim Causer et al. "Transcription Maximized."
[52] Quoted in Parry, "Historians Ask the Public"

In summary, the processing of archival materials can involve transcription for particularly important documents. Crowdsourcing can be an effective method for accomplishing this. The objectives of the project and the materials themselves will influence many facets of the project, such as the choice of a transcription tool, how the project ensures the accuracy of transcriptions, and how the project is publicized for the recruiting of volunteers. Project administrators need to engage in a cost/benefits analysis that accounts for the source of project funding, the increased accessibility offered by accurately transcribed documents, and the increased public awareness of archives that results from the project's public engagement.

Accessing Transcribed Materials

A primary reason for crowdsourcing transcriptions of archival materials is that professional archives and editorial staff cannot always accomplish the transcriptions themselves in a timely or cost effective manner. If transcriptions are accomplished by volunteers, then the documents' content becomes widely accessible in an online format by researchers and the public. Max J. Evans, former director of the NHPRC, supports crowdsourcing transcriptions for accessibility reasons. He points out that with this approach, documents will not be "held up in these scholarly editing offices for years and years, and not only available to a select group of scholars."[53]

At a fundamental level, documents become accessible through some type of text searching after they are transcribed. Short of transcription, the best accessibility that archivists can offer is keyword searching of document *description* databases. All of the projects examined

---

[53] Quoted in Patricia Cohen. "Humanities 2.0; Scholars Recruit Public for Project." *The New York Times*, December 28, 2010, http://search.proquest.com.ezproxy.lib.umb.edu/docview/821256512 (accessed February 21, 2013)

in this paper offer (or will offer)[54] keyword searching of transcriptions. But there are also more specialized searches of transcriptions that some projects are offering. Transcribe Bentham's mark-ups of transcriptions permit more powerful searching by researchers. "A user may choose to see every instance in which Bentham deleted the word 'panopticon,' or search just within the text of the marginal notes."[55]

Projects that have volunteers transcribe only selected information into database fields offer accessibility to documents only through the mediation of those fields. The benefit of this method of transcription and consequent accessibility is that users can search transcriptions more efficiently. There are generally fewer results from a search by database field than there are from a full-text search. This arrangement is particularly useful for genealogical researchers, as they often search documents by way of select information categories such as surnames, locations, and dates. This is the approach of the World Archives Project. The NYPL's What's On the Menu? project also has volunteers enter transcriptions of select data into database fields. Access to the scanned menu images is offered through searching in the transcribed fields. However, the NYPL has plans to creatively move beyond the original documents by creating a "historical Yelp" that contains "information on theatrical and musical shows and other activities as well as food." There are plans to "mash" the menu data with maps to provide a geographic dimension to historical menus.[56] Some of the NYPL's plans for using their transcriptions are more similar to

---

[54] The National Archives Transcription Pilot Project is not yet offering transcribed documents in a public database. Project administrators eventually intend to offer transcriptions with their descriptions in the National Archives online catalog. This was communicated to the author in an email from Meredith Stewart, Management and Program Analyst, Office of Innovation, National Archives and Records Administration, November 13, 2012.

[55] Tim Causer, et al. "Transcription Maximized."

[56] Meredith Schwartz, "Dicing Data at NYPL Labs," *Library Journal* 137, no. 14 (September 2012): 22-25, *Library, Information Science & Technology Abstracts*, EBSCO*host* (accessed February 21, 2013).

scientific crowdsourcing projects like Old Weather than they are to the document-based transcription projects examined in this paper.

Public Outreach Through Crowdsourcing Transcriptions

Crowdsourced transcription projects are an excellent opportunity for archives to raise public awareness of both their materials and the role of archives in a society. In the process of publicizing transcription projects, public awareness begins to rise--even among people who do not choose to volunteer. But projects have the potential to reach far beyond this initial publicity during their course of development. Ben Brumfield describes a "virtuous circle" for crowdsourced transcription projects: 1) volunteers, 2) deep digitization, 3) findability, and 4) more volunteers.[57] By "deep digitization", Brumfield means transcription of digitized documents as opposed to merely descriptions of digitized documents. By "findability", Brumfield means the increased accessibility consequent upon digitization and transcription. His theory is that the greater access to archival materials offered by crowdsourced transcriptions generates more enthusiasm and support for the project, which in turn produces more volunteers.

While Brumfield's hypothesis is intuitively true, there have not yet been any studies to confirm (or disconfirm) it. One prior condition that should exist is public interest in the archival materials being transcribed. The DIY History project exemplifies this condition. The materials chosen for transcription include a twenty thousand item collection of handwritten recipe books and a collection of the personal papers of Niles Kinnick, a University of Iowa football player and U.S. Navy aviator who died during World War II. The choice of these collections was clearly

---

[57] Ben Brumfield, "Bilateral Digitization at Digital Frontiers 2012." Collaborative Manuscript Trancription, entry posted September 24, 2012 , http://manuscripttranscription.blogspot.com/2012/09/bilateral-digitization-at-digital.html (accessed February 21, 2013)

targeted to produce high public interest. The NYPL's What's On the Menu? project has been popular among New Yorkers for a similar reason.[58]

Due to their online nature, crowdsourced transcription projects are capable of raising public awareness beyond limited geographic confines. After the publication in December 2010 of a New York Times article[59] on Transcribe Bentham, the project's website received hits from people in ninety-one countries around the world. Coverage by a major news outlet was probably due to the novelty of the project. Regardless of the reason, the Transcribe Bentham rapidly raised the public profile of the Bentham Project. All archives that host crowdsourced transcription projects could benefit from publicity by non-scholarly media, whether at the local, regional, or national levels. Benefits include not only more volunteers for transcription, but also the increased support, financial and otherwise, that results from greater public awareness.

Preservation of Archival Materials

When archival materials are digitized and transcribed, it becomes much less necessary for archivists to allow access to the original documents. Researchers' demand to view original materials also falls. When original documents are viewed less often, their preservation for the future is more secure. From the archivist's perspective, this is a major benefit of crowdsourcing transcriptions. As this relatively new method becomes more widespread and there is more experience to analyze, attaching statistics to these commonsense observations will be possible. As yet, the literature on crowdsourced transcriptions has not addressed this matter.

---

[58] Schwartz, "Dicing Data."

[59] Cohen, "Humanities 2.0."

Conclusion

  Crowdsourcing transcriptions is an excellent opportunity for archivists to have materials processed at a much greater rate than is possible with only professional transcribers at work. While some risks of transcriptions' accuracy accompany this approach, errors can be minimized in several ways. The benefits that accompany this approach are substantial; they include increased accessibility and public awareness of archival materials. Moreover, at the present time, grant-based funding for crowdsourced transcriptions may be more available than for traditional, professional transcriptions. Crowdsourcing transcriptions is a processing method that all archivists should be aware of and consider using.

Bibliography

Ancestry.com, "World Archives Project," Ancestry.com, http://community.ancestry.com/awap (accessed February 21, 2013).

Bentham Project, "Transcribe Bentham," University College London, http://blogs.ucl.ac.uk/transcribe-bentham/ (accessed February 21, 2013).

Blanke, Tobias, Michael Bryant, and Mark Hedges. "Open Source Optical Character Recognition for Historical Research." *Journal of Documentation* 68 (September 2012): 659-83.

Causer, Tim and Valerie Wallace, "Building A Volunteer Community: Results and Findings from *Transcribe Bentham*," *Digital Humanities Quarterly* 6, no. 2 (2012), http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html (accessed February 21, 2013).

Causer, Tim, Justin Tonra, and Valerie Wallace. "Transcription Maximized; Expense Minimized? Crowdsourcing and Editing *the Collected Works of Jeremy Bentham*." *Literary and Linguistic Computing* 27, no. 2 (2012): 119-137.

Center for History and New Media. "Papers of the War Department." Center for History and New Media. http://wardepartmentpapers.org/ (accessed February 21, 2013).

Cimtech. "Crowdsourcing Project Helps Finnish Library to Digitise Historical Documents." *Information Management & Technology (IM@T.Online).* (March 2011), http://www.imat.cimtech.co.uk.ezproxy.lib.umb.edu/Pages/IM@T_Online/2011/February-March/IMAT_0211_NewsTrack_11.htm (accessed February 21, 2013).

Citizen Science Alliance. "Old Weather: Our Weather's Past, the Climate's Future." Zooniverse. http://www.oldweather.org/ (accessed February 21, 2013).

———. "Zooniverse: Real Science Online." Zooniverse. https://www.zooniverse.org/ (accessed February 21, 2013).

Edgerton Center. "The Edgerton Digital Collections Project." Massachusetts Institute of Technology. http://edgerton-digital-collections.org/ (accessed February 21, 2013).

Government & Heritage Library, State Library of North Carolina. "Family History Transcription Project." Flickr. http://www.flickr.com/photos/statelibrarync/sets/72157627124710723/ (accessed February 21, 2013).

———. "Genealogy Vertical File Transcription Project." State Library and State Archives of North Carolina. http://statelibrary.ncdcr.gov/digital/ncfamilyrecords/verticalfiles.html (accessed February 21, 2013).

Grabowski, John J. "Keepers, Users, and Funders: Building an Awareness of Archival Value." *The American Archivist* 55 (Summer 1992).

http://archivists.metapress.com/content/r9gmpj0rj8530811/?p=45a04f86615f42dd974acdd9239b28d0&pi=0 (accessed February 21, 2013)

Greene, Mark A. and Dennis Meissner. "More Product, Less Process: Revamping Traditional Archival Processing." *American Archivist* 68, no. 2 (Fall-Winter 2005): http://www.jstor.org/stable/40294290 (accessed February 21, 2013).

Holley, Rose. *Many Hands make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers*. Canberra: National Library of Australia, 2009. http://www.nla.gov.au/openpublish/index.php/nlasp/article/viewArticle/1406 (accessed February 21, 2013).

———. "Crowdsourcing: How and Why should Libraries do it?" *D-Lib Magazine* 16, no. 3 (March 2010). http://www.dlib.org/dlib/march10/holley/03holley.html (accessed February 21, 2013).

Kowal, Kimberly C. and Petr Pidal, "Online Georeferencing for Libraries: The British Library Implementation of Georeferencer for Spatial Metadata Enhancement and Public Engagement," *Journal of Map & Geography Libraries* 8 (September 2012): 276-89.

Lang, Andrew S.I.D. and Joshua Rio-Ross. "Frederick Douglass Diary: A Written Rummage Project." Wikispaces. http://frederickdouglassdiary.wikispaces.com/ (accessed February 21, 2013).

———. "Using Amazon Mechanical Turk to Transcribe Historical Handwritten Documents." *Code4Lib Journal* no. 15 (October 31, 2011) http://journal.code4lib.org/articles/6004 (accessed February 21, 2013).

Leuven City Archive, "Itinera Nova," Leuven City Archive, http://www.itineranova.be/in/home (accessed February 21, 2013).

Lisanti, Suzana. "Innovation in the Archives." (Panel Discussion at the Fall Conference of New England Archivists, Simmons College, Boston, MA, November 3, 2012).

Moyle, Martin, Justin Tonra, and Valerie Wallace. "Manuscript Transcription by Crowdsourcing: Transcribe Bentham." *Liber Quarterly: The Journal of European Research Libraries* 20, no. 3/4 (2011). http://liber.library.uu.nl.ezproxy.lib.umb.edu/index.php/lq/article/view/7999 (accessed February 21, 2013)

National Humanities Alliance, "National Historical Publications and Records Commission FY2013," National Humanities Alliance. http://www.nhalliance.org/advocacy/funding-priorities/national-historical-publications-records-commissio-2.shtml (accessed February 21, 2013)

New York Public Library. "NYPL Map Warper." New York Public Library. http://maps.nypl.org/warper/ (accessed February 21, 2013).

New York Public Library Labs. "What's on the Menu?" The New York Public Library. menus.nypl.org (accessed February 20, 2013).

Parry, Marc. "Historians Ask the Public to Help Organize the Past; but is the Crowd Up to it?" *The Chronicle of Higher Education* 59, no. 2 (September 3, 2012). http://go.galegroup.com.ezproxy.lib.umb.edu/ps/i.do?id=GALE%7CA302048583&v=2.1&u=mlin_b_umass&it=r&p=AONE&sw=w (accessed February 21, 2013).

Saylor, Nicole. "Experimenting with Strategies for Crowdsourcing Manuscript Transcription." *Research Library Issues* 277 (December 2011): 9-14.

Schwartz, Meredith. "Dicing Data at NYPL Labs." *Library Journal* 137, no. 14 (September 2012): 22-25. *Library, Information Science & Technology Abstracts*, EBSCO*host* (accessed February 21, 2013).

Text Encoding Initiative Consortium. "Text Encoding Initiative." Text Encoding Initiative. http://www.tei-c.org/index.xml (accessed February 21, 2013).

U.S. National Archives and Records Administration. "Citizen Archivist Dashboard." U.S. National Archives and Records Administration. http://www.archives.gov/citizen-archivist/ (accessed February 21, 2013).

University of Iowa Libraries. "DIY History." University of Iowa. http://diyhistory.lib.uiowa.edu/ (accessed February 21, 2013).

University of Kentucky Department of History. "Carolingian Canon Law Project." University of Kentucky. http://ccl.rch.uky.edu/ (accessed February 21, 2013).

University of Oxford. "Ancient Lives: Help Us to Transcribe Papyri." Zooniverse. http://ancientlives.org/ (accessed February 21, 2013).

Wikipedia, "Crowdsourcing." Wikimedia Foundation, Inc. http://en.wikipedia.org/wiki/Crowdsourcing (accessed November 17, 2012).

Willey, Eric. "A cautious partnership: The growing acceptance of folksonomy as a complement to indexing digital images and catalogs."*Library Student Journal* 15. *Library, Information Science & Technology Abstracts*, EBSCO*host* (accessed February 21, 2013).