Alexandra Lightle
LIS 510 Assignment 5: Issue Paper
May 10, 2017

Crowdsourcing in the Humanities

In the modern global economy, the concept of outsourcing has become very familiar, and in the past ten years a new digital version of this has emerged; crowdsourcing. A succinct definition of crowdsourcing appears in *A New Companion to Digital Humanities*: "[T]he practice of using contributions from a large online community to undertake a specific task, create content, or gather ideas[.]"[1] Digital libraries specialist, Rose Holley, provides figures for the best known and most used example of internet crowdsourcing; Wikipedia. The site was launched in January of 2001, and by December of 2008 they had a total of 10 million articles.[2] At the time of writing, there are over 40 million articles in 293 languages.[3] This paper seeks to provide an oversight of the more specialized definitions of crowdsourcing in a library setting, its various uses, the communities created, tools to make it more effective, as well as address a few criticisms of the practice.

Jeff Howe, in a 2006 article for *Wired* magazine, coined the term crowdsourcing, and it quickly entered the common vernacular. Howe provides two definitions for what it is on his website:

> **The White Paper Version:** Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call. **The Soundbyte Version:** The application of Open Source principles to fields outside of software.[4] [emphasis mine]

In an article for *Wired* magazine, Howe highlights the financial aspects of corporate crowdsourcing. One example he describes is a VH1 show called *Web Junk 20,* which is a curated

---

[1] Schreibman, S., Siemens, R., & Unsworth, J, J. "Crowdsourcing in the Digital Humanities." In *A New Companion to Digital Humanities*, Wiley-Blackwell, 2016. pp. 421
[2] Holley, Rose. "Crowdsourcing, How and Why Should Libraries Do It?" *D-Lib Magazine*, Vol. 16, No. 3/4. Section 4, Figure 5.
[3] Wikipedia.com
[4] Howe, Jeff. *Crowdsourcing*. n.d. http://crowdsourcing.typepad.com/

presentation of user-generated videos from the internet for a fraction of production costs that

scripted shows require. Other industries followed suite and began mining this source of free digital

labor, one of the more notorious being Amazon;

> Amazon Mechanical Turk is a Web-based marketplace that helps companies find people to perform tasks computers are generally lousy at – identifying items in a photograph, skimming real estate documents to find identifying information, writing short product descriptions, transcribing podcasts. Amazon calls the tasks HITs (human intelligence tasks); they're designed to require very little time, and consequently they offer very little compensation – most from a few cents to a few dollars.[5]

This sort of depersonalized labor worked well for the private sector, however, when under-funded

humanities fields wished to also incorporate crowdsourcing into their practices, they had to design

and implement their projects in a different way.

Trevor Owens elaborates upon the kind of interactions that would ideally shape the role of

crowdsourcing in the Humanities in his article, "The Crowd and The Library." Rather than aiming

to reach as large a crowd as possible, these kinds of projects should seek input from a smaller pool

of volunteers. This would increase the likelihood of obtaining higher quality work from a dedicated

group— people who are inspired by personal interest rather than money. Owens insists that this

method is the only ethical way to use crowdsourcing. The goal is not to exploit, but to engage:

> In cultural heritage, we have clear values and missions and we are in an opportune position to invite the public to participate. However, when we do so we should not treat them as a crowd, and we should not attempt to source labor from them. When we invite the public we should do so under a different set of terms. A set of terms that is focused on providing meaningful ways for the public to interact with, explore, understand the past.[6]

---

[5] Howe, Jeff. "The Rise of Crowdsourcing." *Wired*, June 01, 2006. Section 4. https://www.wired.com/2006/06/crowds/

[6] Owens, Trevor. "The Crowd and The Library" [Web log post]. May 20, 2012. http://www.trevorowens.org/2012/05/the-crowd-and-the-library/

Rose Holley examines how the interactive aspect of crowdsourcing can be an extremely useful tool

to libraries. She argues that crowdsourcing is a kind of social interaction, where the objective is not

personal profit, but a personal contribution. She writes:

> Libraries are already proficient in the first step in crowdsourcing: social engagement with
> individuals, but we need to get proficient in the second step: defining and working toward
> group goals.[...] When libraries first started delivering digital resources all these social
> interactions were taken away from users [...] It has taken libraries awhile to realise that
> users still want more than a simple information transaction, and they want the same and
> more social interactions than they had in 'pre-digital' days.[7]

By recreating the sociable environment of the library on a digital platform, libraries can provide

more accessible material that is of interest to a larger group of people who may wish to examine

and contribute, creating an online community in the process.

One of the early achievements of crowdsourcing was done by the Australian Newspapers

Digitisation Program, which in August of 2008 uploaded millions of articles and asked volunteers

to search for corrections that needed to be made. In August of 2009 5,437,254 lines had been

corrected. As of May 2017, almost nine years on, there have been 229,347,323 lines corrected.[8]

There are other notable achievements of smaller scale projects. These projects soon become social

spaces for the volunteers:

> [T]he New York Public Library's menu transcription project, *What's on the Menu?*, invites
> members of the public to help transcribe the names and costs of menu items from digitized
> copies of menus from New York Restaurants. Anyone who wants to can visit the project
> website and start transcribing menus. However, in practice it is a dedicated community of
> foodies, New York history buffs, chefs, and otherwise self-motivated individuals who are
> excited about offering their time and energy to help contribute, as volunteers, to improving
> the public library's resource for others to use.[9]

---

[7] Holley, Section 2.
[8] http://trove.nla.gov.au/
[9] Owens.

Projects like this are a gravitational center where people of various professions and backgrounds—each with their own special niche of knowledge— merge together on a topic of mutual interest, while simultaneously contributing to a larger whole.

An ongoing challenge of how to recruit volunteers persists in crowdsourcing work. Holley provides many tips for having a successful project. One of the first tasks of crowdsourcing is for the project to state a clear and concise goal, as well as what results they anticipate from the project, so the volunteers understand exactly what they are working towards. The progress of the project should be completely transparent, so that volunteers can calculate how much work needs to be done to meet the final goal. The matter of usability is an important factor as well; it should be easy to interact with, using quick and intuitive software. Many volunteers are older adults who may not be technologically literate. Therefore, the software must be easy to operate, and site mangers should provide detailed instructions. The material must be interesting, and volunteers should have options on a variety of tasks to do (i.e. transcribing, proofreading, or tagging). Having some kind of non-monetary recognition, such as lists of the top contributors, or public credit given to them for work are effective ways of building loyalty between the institution and the volunteers.

Owens stresses that the managers of a crowdsourcing project should drop their negative connotations of the word amateur, because these nonprofessionals possess a myriad of knowledge. "I suggest we think of crowdsourcing not as extracting labor from a crowd, but of a way for us to invite the participation of amateurs (in the non-derogatory sense of the word) in the creation, development, and further refinement of public goods."[10] Marc Parry reports on the involvement of volunteers outside the so-called academic "ivory towers" who have the extra time, the desire for knowledge, and an eagerness to participate.

[10] Owens.

Crowdsourcing advocates see that kind of [user] excitement as a powerful force to improve access to material, build an engaged audience for collections, and perhaps save money. They speak of democratizing the publication of historical documents, allowing people to produce an online archive about any subject, be it a World War II regiment or a small-town mayor.[11]

Oftentimes the topics of digital archiving branch out to a huge variety of disciplines and interests, and the more diverse the volunteer force is, the more knowledge you can extract from source material.

Of course, this method is not without its disadvantages. One immediate problem that professionals identified in this method is the potential for both unintentional and mischievous mistakes. As for the latter, Matt Enis writes: "Experts interviewed [...] say that trolling or vandalism tends to be a nonissue for most library crowdsourcing projects, particularly those that require volunteers to register with a verified email address in order to participate."[12] As for mistakes, such as typos or misidentification, many programs implement options for volunteers to double check other people's work, so that it can go through multiple editing processes. One employee working on the NYPL's *What's on the Menu?* project says; "We just have a basic honor system and a couple of safeguards built in to detect malicious behavior or hacking. Basically, it's just, 'Here's the menus, transcribe them; when you think it's done, move it along, and it will go into a review category,' until another volunteer confirms that it is correct."[13]

Another concern is how much time paid professionals would need to put into managing volunteers. Many findings suggest that these duties make up a very small percentage of their time. Parry writes that editors of a crowdsourcing project at George Mason University "spent about 30

---

[11] Parry, Marc. "Historians Ask the Public to Help Organize the Past; But is the crowd up to it?" *The Chronicle of Higher Education*. September 03, 2012. Page 2 of 5.
[12] Enis, Matthew. "Wisdom of the Crowd; As Libraries Turn to Crowdsourcing for Assistance with Large Digitization Projects, a Consensus on Best Practices is Beginning to Emerge." *Library Journal*, Vol. 140, No. 11. pp. 4.
[13] Ibid.

mins a day managing the work of these volunteers– comparing transcriptions to the original

images, creating accounts, and answering questions."[14] Holley argues that:

> The main task of the paid staff in regard to management of volunteers was to create, establish, or endorse guidelines, FAQs, and policies for the digital volunteer process. The site manager may also keep an eye on the forum activity and spot anything which may become an issue and resolve it through FAQ, policy or guidelines. All site managers agreed this was the way to handle large online communities. No attempt should be made to seek paid staff to 'manage' digital volunteers.[15]

This small fraction of time spent supervising the crowdsourcing projects frees up time for paid staff

to work on projects that require more professional training.

Some specialists are skeptical as to the reliability and speed of the work done in

crowdsourcing projects. One of them, Edward G. Lengel, who is editor-in-chief of the Papers of

George Washington at the University of Virginia, argues: "[B]ecause members of the public who

have not been trained in documentary editing are never going to be able to produce complete

editions to the same level of accuracy that trained professionals will do. [...] I just think it can never

be an alternative to documentary editing for a major project."[16] This statement takes a small view of

archival work at large. For such an important collection like the one he works with, trained

professionals *are* going to be able to provide a more in depth study of them. But there are millions

of archives that hold material from much less eminent sources, but still could serve as an important

tool for many researchers, and could easily be handled by competent volunteers.

While digital volunteers will never replace professional analysis, they have the potential to

do the same amount of work it would take numerous paid employees to complete. This free work

is a budget friendly way to continue developing collections, especially in the current political

uncertainty. It helps libraries fulfil one of their main missions; providing a social space for patrons.

[14] Parry, pg. 4
[15] Holley, sec. 5.5
[16] Parry, pg. 3

I believe in the near future crowdsourcing will be one of the first options considered when digitizing new collections, and that participation is often a more stimulating rewarding pastime than TV, games, or social media.

Bibliography

Enis, Matthew. "Wisdom of the Crowd; As Libraries Turn to Crowdsourcing for Assistance with Large Digitization Projects, a Consensus on Best Practices is Beginning to Emerge." *Library Journal*, Vol. 140, No. 11. 5 pp. Retrieved May 05, 2017, http://0-www.lexisnexis.com.liucat.lib.liu.edu/hottopics/lnacademic/?

Holley, Rose. "Crowdsourcing, How and Why Should Libraries Do It?" *D-Lib Magazine*, Vol. 16, No. 3/4. Retrieved May 05, 2017, http://0-resolver.ebscohost.com.liucat.lib.liu.edu/http://0-resolver.ebscohost.com.

Howe, Jeff. *Crowdsourcing*. n.d. http://crowdsourcing.typepad.com/

Howe, Jeff. "The Rise of Crowdsourcing." *Wired*, June 01, 2006. Retrieved May 05, 2017, https://www.wired.com/2006/06/crowds/

National Library of Australia. n.d. [http://trove.nla.gov.au/system/stats?env=prod -corrNewspapers].

Owens, Trevor. "The Crowd and The Library" [Web log post]. May 20, 2012. Retrieved May 05, 2017, http://www.trevorowens.org/2012/05/the-crowd-and-the-library/

Parry, Marc. "Historians Ask the Public to Help Organize the Past; But is the crowd up to it?" *The Chronicle of Higher Education.* September 03, 2012. Retrieved May 05, 2017, http://0-www.lexisnexis.com.liucat.lib.liu.edu/hottopics/lnacademic/?

Schreibman, S., Siemens, R., & Unsworth, J, J. "Crowdsourcing in the Digital Humanities." In *A New Companion to Digital Humanities* Wiley-Blackwell, 2016. pp. 420-439.

Wikipedia. n.d. https://en.wikipedia.org/wiki/Wikipedia:Size_comparisons