

3-2015

Crowdsourcing Digital Public History

Jason A. Heppler

University of Nebraska at Omaha, jheppler@unomaha.edu

Gabriel K. Wolfenstein

Follow this and additional works at: <https://digitalcommons.unomaha.edu/crisslibfacpub>



Part of the [Library and Information Science Commons](#), [Public History Commons](#), and the [United States History Commons](#)

Please take our feedback survey at: https://unomaha.az1.qualtrics.com/jfe/form/SV_8cchtFmpDyGfBLE

Recommended Citation

Heppler, Jason A. and Wolfenstein, Gabriel K., "Crowdsourcing Digital Public History" (2015). *Criss Library Faculty Publications*. 27.

<https://digitalcommons.unomaha.edu/crisslibfacpub/27>

This Article is brought to you for free and open access by the Dr. C.C. and Mabel L. Criss Library at DigitalCommons@UNO. It has been accepted for inclusion in Criss Library Faculty Publications by an authorized administrator of DigitalCommons@UNO. For more information, please contact unodigitalcommons@unomaha.edu.

Crowdsourcing Digital Public History

by Jason A. Heppler and Gabriel K. Wolfenstein

The generation of communal knowledge is not a new phenomenon. In the late nineteenth century, the *Oxford English Dictionary* solicited volunteers to submit words and their usage for inclusion in the dictionary (1). Carl Becker, writing in 1932 on what was already an old discussion in the historical profession, noted that "if the essence of history is the memory of things said and done, then it is obvious that every normal person, Mr. Everyman, knows some history" (2). The historian Jo Guldi's work on participatory mapping shows that urban planners in the middle of the twentieth century attempted to learn from and listen to members of a community.

There is plenty of precedent, then, for harnessing participatory knowledge. Today, the digital turn has offered new technologies to engage with communities and significantly widened the number of possible participants. The success of recent digital crowdsourcing projects, including Flickr Commons, the National Archive's Citizen Archivist Dashboard, History Harvest, and Transcribe Bentham have demonstrated the degree of success that crowdsourcing offers to cultural heritage and public digital history. Like any research, a crowdsourcing project requires careful planning and an understanding of what is meant by crowdsourcing in a specific project. In this essay we discuss the importance of these definitions, describe a few successful and well-known crowdsourced projects, and discuss one of the projects we are working on here at Stanford's Center for Spatial and Textual Analysis (CESTA).

Why does the definition of crowdsourcing matter? The short answer is that if you aren't clear on what you mean by crowdsourcing, you will have a hard time designing your project and, importantly, engaging with your targeted community (3). In his book *Crowdsourcing* (2013), Daren C. Brabham defines crowdsourcing as an "online, distributed problem-solving and production model that leverages the collective intelligence of online communities to serve specific organizational goals. Engaged volunteers are given the opportunity to respond to crowdsourcing activities promoted by the organization, and they are motivated to respond for a variety of reasons" (4). Here, the important thing to note is that a crowd is neither left to its own devices nor is it used only as a provider of information or mechanical work. Rather, there must be an almost symbiotic relationship between the organization (or researcher) and the crowd. In such projects, you need to identify your community from the start and figure out how to make it possible for them to contribute effectively and become your partner in the research.

This, however, is not the only kind of crowdsourcing. Some projects are driven by the community. These would include Wikipedia and open-source software. Here, the project leaders are providing the space, but it is the community which defines both scope and content. There is also a third kind of crowdsourcing: something along the lines of an online poll to find a new M&M's color by the candy's manufacturer. Here, control is held completely by the company—as opposed to jointly, as in Brabham's definition, or by the community, as in the case of Wikipedia.

Depending on the kind of data you want to collect, and the kind of community you want to engage with, you need to be very clear on what you mean by crowdsourcing. We see one commonality that links all crowdsourcing projects: for your project to be successful, you have to think about why the crowd would want to participate. Crowdsourcing's greatest strength is two-fold, at least for humanities and social science researchers: first, it fosters engagement with new publics, and second, it opens up data sets and skills that were formerly difficult, if not impossible, to access. But if you don't consider the community's motivations, you will handicap your project from the start.

When done well, crowdsourcing can result in widespread interest from the public and engaged volunteers. Wikipedia illustrated what collaborative knowledge creation could look like on a large scale. Wikipedia's rapid growth signals a significant shift in crowdsourcing as something that is feasible on a large scale. Other large-scale crowdsourcing projects have likewise produced favorable results: the Galaxy Zoo's volunteers have identified more than ten million images of galaxies, while volunteers working on the National Library of Australia's historic newspaper digitization project have corrected 36.5 million lines of OCR text. Participants are motivated by a number of factors, including a passion for the subject, an interest in giving back to the community, contributing to a collective goal, or wanting to play a role in a professional field (5). For historians, this means leveraging the knowledge of engaged volunteers who are eager to share their accumulated knowledge and passion for the past.

We'd like to share a few examples of recent crowdsourced projects that offer a detailed glimpse of how crowdsourcing can work in digital history. For the first project, we need to go all the way back to the turn of the eighteenth century. On the night of November 8, 1800, a disastrous fire swept through the U.S. War Department, destroying a collection of papers, books, and records relating to the military, veterans, Indians, social welfare, internal security, and frontier diplomacy. In short, the loss not only destroyed key documents of the military history of the early Republic, but a wealth of information about the country's early history more broadly. Not until the 1990s did scholars begin reconstructing the lost archive. Under the work of Ted Crackel, a military historian at East Stroudsburg University, the project sought to rebuild the archive with copies. The project then moved to the Center for History and New Media at George Mason University in 2006, which initiated the digitization of forty-five

thousand documents. In 2010, the Papers of the War Department began a crowdsourcing project to transcribe the documents in the archive. The project generated widespread interest: after thirty-three months, the War Department Papers' 1,685 volunteer users have contributed 12,071 saves to the project and another 565 edits, initiated roughly five hundred conversations using the "talk" feature of the transcription tool (5).

We can also share details about a crowdsourcing digital history project that both of us have worked on. [Living with the Railroads](#), under the direction of Richard White in Stanford's Spatial History Project, conforms closely to the crowdsourcing definition offered by Brabham. Following White's work in his book *Railroaded: The Transcontinentals and the Making of Modern America* (2011), this project seeks to use crowdsourcing to learn more about the social, cultural, and environmental impact of the development and expansion of the railroads in the American West. The goal was to get railroad enthusiasts to upload their data, identify images and documents whose provenance is unknown, and help facilitate the growth of connections between historians and enthusiast communities. Rather than a top-down research project, *Living with the Railroads* is understood as a partnership between the wider expert community and us. To accomplish that, we had to go out and engage the community, figure out what they wanted and what they needed, and then work with them to generate both site and content. In a real sense, they are the experts, and they cannot just be mined for knowledge. By interacting with the community as partners and considering their own goals, we help increase the chances for the project's success.

We would like to conclude with a very important point about crowdsourcing, or community sourcing. We have realized, over the course of the crowdsourcing projects we have worked on, that if you are looking to have a successful project, you must take the community and their wants and needs into account as one of the primary questions. Having a research question is obviously the key starting point. But if you approach your crowd like you would approach an archival source or a data set, you will most likely fail to get engagement, or get a poor return on your investment. Whether your crowd is getting paid or participating because you are helping to build a useful tool (as with *Living with the Railroads*), you need to understand your crowd. That means that before you start your project, you need to do research on your crowd, and if you are working with people directly, most likely you will need to have someone conduct community outreach. Crowdsourcing is not merely an easy way to gather data. Rather, it is a rewarding research process, and it can be cost- and time-effective, but only if you lay the groundwork and take the needs of your crowd or community into account.

Notes

- (1) Nate Lanxon, "How the Oxford English Dictionary Started out like Wikipedia," *Wired UK*, <http://www.wired.co.uk/news/archive/2011-01/13/the-oxford-english-wiktionary>.
- (2) Carl Becker, "Everyman His Own Historian," *American Historical Review*, 37 (1932), 223.
- (3) We use crowd and community interchangeably in this article, as we think that both terms get at the same idea of engaging with a potentially large group of people.
- (4) Daren C. Brabham, *Crowdsourcing* (2013), p. xix.
- (5) "Community Transcription—Thirty-Three Months," Papers of the War Department, Jan. 29, 2014, <http://wardepartmentpapers.org/blog/?p=1449>.

Selected Crowdsourcing Projects

- <http://storycorps.org>
- <http://ancientlives.org>
- <https://www.wir-waren-so-frei.de>
- <http://www.bl.uk/maps/>
- <https://transcription.si.edu/>
- <http://www.archives.gov/citizen-archivist/>
- <http://braceroarchive.org/>
- <http://wardepartmentpapers.org/>
- <http://historyharvest.unl.edu/>
- <http://www.historypin.com/project/22-yearofthebay/>
- <http://www.historypin.com/project/42-railroads/>

The Authors

Jason A. Heppler is the Academic Technology Specialist in the Department of History at Stanford University, Affiliated Staff in the Spatial History Lab in the Center for Spatial and Textual Analysis, and a Ph.D. Candidate in History at the University of Nebraska-Lincoln. Prior to joining Stanford, he served as the project manager for the William F. Cody Archive (www.codyarchive.org) at the Center for Digital Research in the Humanities at the University of Nebraska-Lincoln. He can be found online at www.jasonheppler.org or on Twitter @jaheppler.

Gabriel K. Wolfenstein is a historian by training, whose work is in Victorian Britain, with specific focus on the history of statistics and the census. He is particularly interested in how the making and taking of such surveys impacts the way people think about themselves and the world around them. Other interests include the rise of bureaucracies and the popularization of science. Formerly a post-doctoral Fellow in the Humanities in Stanford's Introduction to the Humanities program, he has lately become interested in questions of the utility of crowd-sourcing for humanities research, and crowd-sourcing in general. He is currently the Project Manager for CESTA's Mellon Grant supported research into these very questions. He earned his B.A. from the University of California, Berkeley, an M.A. from The New School for Social Research, and his Ph.D. from the University of California, Los Angeles.