

Crowdsourcing and Community Engagement

In 2011, the New York Public Library (NYPL) released 9,000 digitized restaurant menus with “delicious data” that had been “frozen as pixels,” making the menus difficult to search, index, and discover online. Along with the menus, the NYPL launched an interface that asked the public to help transcribe the thousands of menus and the hundreds of thousands of dishes. In only three months, the menus (and dishes) were fully transcribed.

The success of NYPL’s crowdsourced *What’s on the Menu?* (<http://menus.nypl.org/>) demonstrates how enthusiastically public audiences respond to a well-defined project to which they can contribute through an expertly designed interface. While crowdsourcing has been used in the corporate world as a way to outsource tasks to nonemployees, it is increasingly being used in cultural and academic institutions for projects that seek to harness the energy and brainpower of the masses to complete specific tasks more quickly and inexpensively than would otherwise be possible. Many competing definitions of *crowdsourcing* exist, but perhaps one of the most helpful is offered by Enrique Estellés-Arolas and Fernando Gonzáles-Ladrón-de-Guevara: “A type of participative online activity in which an individual, an institution, a nonprofit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task.”⁷¹

In some instances, academic institutions have taken the lead in developing platforms that facilitate crowdsourcing. *Zooniverse* (<https://www.zooniverse.org/>), owned and operated by a partnership of eight academic, nonprofit, and corporate institutions, currently hosts thirty-three projects that ask participants to carry out a wide range of tasks—from analyzing cancer cells to classifying galaxies to identifying the seasons in photographs of landscapes. The resulting input has led to ninety-four published articles to date.

Crowdsourcing is particularly well suited to simple, repeatable tasks. The challenge often is to find ways for keeping participants engaged. Many classifications of crowdsourcing tasks have been proposed, but for academic and cultural institutions, the tasks may best be organized into four main categories:

- Transcription
- Supplementing Metadata
- Collection Building & Curation
- Identification & Provenance

Transcription. Many crowdsourcing projects and games seek to address the gap that still-lagging optical charac-

ter recognition (OCR) technologies leave in the quality of their digital transcription capabilities. Most OCR software, for instance, is unable to convert handwriting or languages that use a non-Roman alphabet into machine-readable text. *Ancient Lives* (<http://www.ancientlives.org/>), part of *Zooniverse*, teaches users how to identify and then transcribe ancient Greek text that appears on Egyptian papyri. Even more projects use crowdsourced transcription to transform handwritten documents into readable and mineable texts. For example, *DIY History* (<http://diyhistory.lib.uiowa.edu/>), built at the Iowa Digital Library, uses a customized version of Omeka to crowdsource the transcription of several of its collections of letters, diaries, and manuscripts.

Supplementing Metadata. Whereas text-based projects often require transcription, many cultural heritage institutions are increasingly facing the challenges of digitizing images and three-dimensional objects. These objects are virtually undiscoverable on the web without rich metadata. Recognizing this challenge, Dartmouth College’s *Tiltfactor* (<http://www.tiltfactor.org/about/>) has partnered with several cultural institutions, including the British Library, to create a suite of competitive and collaborative games through which users provide metadata via a variety of tagging tasks. For instance, in *Beanstalk* (<http://www.tiltfactor.org/game/beanstalk/>), the user transcribes un-OCR-able words and fragments from digitized botanical texts. As the user provides transcriptions, the beanstalk grows higher and higher through various landscapes. Once the word has been transcribed multiple times in the same way by multiple users, the verified text is then returned to the library that holds the document and is added to its collection. As the game appeals to viewers: “Play *Beanstalk*, save scanned books from digital oblivion.”

Collection Building & Curation. Beyond projects that seek to make existing texts digitally discoverable, crowdsourcing is also being used to build and curate collections online. Crowdsourcing has been especially useful in documenting and providing a space for public and private responses to community tragedies. Projects like the *September 11 Digital Archive* (<http://911digitalarchive.org/>), the *Hurricane Digital Memory Bank* (<http://hurricanearchive.org/>), and *Our Marathon: The Boston Bombing Digital Archive* (<http://northeastern.edu/marathon/>) engage with members of the public to collaboratively build sites of collective cultural memory. In crowdsourcing collection building and curation, cultural institutions retain the responsibilities for preserving and providing access to their collections while transferring the assemblage of the content of those collections to community members.



By ALICIA PEAKER



John Lumf/Glow Images © 2015

Identification & Provenance. Crowdsourcing is frequently discussed in terms of the power of a massive and anonymous public. But crowdsourcing techniques are also being applied to smaller, more specialized audiences. For example, the *Provenance Online Project* (<https://provenanceonlineproject.wordpress.com/>), based at the Kislak Center for Special Collections, Rare Books, and Manuscripts at the University of Pennsylvania, relies on a community of experts outside of Pennsylvania to source and correct the provenance of individual books through the image-sharing platform Flickr.

Crowdsourcing remains an important strategy for many emerging digital projects that seek to build community around and through their projects. The best-built crowdsourcing tools make clear what each participant adds to the project and why each participant matters to the project as a whole. Some tools take these strategies a step further, connecting participants to each other through forums, collaborative tasks, and competitive scoreboards.

Perhaps the most-voiced concern about using crowdsourcing for scholarly projects is the quality of nonexperts' contributions. To address this concern, the Maryland Institute for Technology in the Humanities (<http://mith.umd.edu/>) is developing a transcription platform that clearly indicates the quality of available transcriptions from *The Shelley-Godwin Archive* (<http://shelleygodwinarchive.org/>). A red dot indicates an untranscribed document, yellow indicates a transcribed but unvetted document, and green indicates a vetted and approved transcription. As scholars increasingly build and contribute to crowdsourcing software, these technologies will become more reflective of the scholarly practices and concerns we share.

Crowdsourcing, particularly as it is deployed in the private sector, may also become exploitative when unpaid or underpaid labor replaces paid work. On the one hand, crowdsourcing is completely voluntary, at the discretion of the participant. On the other hand, relying on unpaid labor could work to dehumanize participants.² True, many cultural heritage institutions, which are chronically underfunded in the United States, have a long

history of engaging volunteer labor. The argument for such labor is often framed in terms of the “greater good” and engages participants' desires to contribute to an institution whose mission they support.³

Crowdsourcing can open up access to previously inaccessible materials—but often in fragmented ways. Participants see snippets of text or a single letter, image, or document outside of the context of a collection. This fragmentation is frequently a result of trying to balance the need to lower barriers for participation with the overall goals of the project. Project designers often ask themselves questions such as the following: How much contextual information does the participant need to contribute meaningfully to the project? How do the goals of the project relate to the crowdsourcing efforts? How does the technology facilitate or hinder participation? If the crowdsourcing initiative is meant to increase buy-in of the project as a whole (as opposed to simply completing mundane tasks), more contextual information and cross-participant engagement are necessary.

Only in the last few years have advances in technology enabled the large-scale, asynchronous collaboration that can produce a project like the New York Public Library's *What's on the Menu?* And yet the spirit behind such a project reaches back much further. As Mia Ridge reminds us: “Technology has enabled crowdsourcing as we know it, but models for public participation in collection, research and observation pre-date it.”⁴ As crowdsourcing gains purchase in academic and cultural heritage sectors, new crowdsourcing tools and platforms will be transformed by scholars, artists, librarians, and curators who seek to engage with community members in meaningful ways. And these new platforms and modes of engagement will, in turn, transform the very shape of the scholarship and creative works that seek to engage communities. ■

Notes

1. Enrique Estellés-Arolas and Fernando González-Ladrón-de-Guevara, “Towards an Integrated Crowdsourcing Definition,” *Journal of Information Science* 38, no. 2 (April 2012), 197.
2. Aniket Kittur, Jeffrey V. Nickerson, Michael Bernstein, Elizabeth Gerber, Aaron Shaw, John Zimmerman, Matt Lease, and John Horton, “The Future of Crowd Work,” in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (New York: ACM, 2013).
3. Mia Ridge, ed., *Crowdsourcing Our Digital Heritage* (Farnham, UK: Ashgate, 2014), 5.
4. Ibid. The *Oxford English Dictionary's* 1857 appeal to the public for “unregistered” words is one of the most notable examples. See Peter Gilliver, “Your Dictionary Needs You: A Brief History of the OED's Appeals to the Public,” *Oxford English Dictionary*, October 4, 2012, <http://public.oed.com/the-oed-appeals/history-of-the-appeals/>.

Alicia Peaker is the CLIR/DLF Postdoctoral Fellow in the Digital Liberal Arts at Middlebury College. During her doctoral work at Northeastern University, she also worked as the Co-Director of *Our Marathon: The Boston Bombing Digital Archive*, the Project Manager of the *Women Writers Project*, and the Development Editor for *GradHacker*.

© 2015 Alicia Peaker. The text of this article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (<http://creativecommons.org/licenses/by-nc-sa/4.0>).