

Book Reviews

Crowdsourcing our Cultural Heritage

MIA RIDGE (ed.)

Farnham, Surrey: Ashgate Publishing, 2014

283p. | ISBN 978 1 4724 1022 1 (hardback) | £70 (website £63)

'Crowdsourcing' has been something of a buzzword in recent years; as such, it's likely to attract in equal measure interest and suspicion. First coined in 2006, the term refers to the idea of organisations accomplishing time-consuming but relatively basic micro-tasks by outsourcing these to a network of online volunteers through an open call on the internet. Initially developed by business, the concept has subsequently been transferred to non-commercial areas, including the GLAM sector (Galleries, Libraries, Archives and Museums). For these resource-constrained public institutions crowdsourcing's appeal lies in the way it promises to provide a cost-effective means of accomplishing organisational tasks *and* an opportunity to invite user engagement with their collections. Yet there remains considerable wariness amongst cultural heritage professionals about this direction. Is crowdsourcing really a promising new tool or does it instead pose a threat to archival standards? Will it prove a major game-changer or just another passing fad?

The recently published anthology, *Crowdsourcing Our Cultural Heritage*, represents a significant contribution to this debate. Edited by British academic Mia Ridge – who gave a well-received keynote address at last year's National Digital Forum in Wellington – this book makes the case for the benefits of crowdsourcing. It does so, however, without the over-hyped, triumphalist language of some advocates, and with a critical eye on the limitations of applying its techniques to the cultural heritage field. To cover both practice and theory, it adopts a two-part approach, with a first section containing eight case studies of crowdsourcing projects written from an implementer's point of view, and the second four essays examining this topic more analytically.

One could describe the eight case studies here as the usual suspects. Anyone interested in crowdsourcing within the GLAM sector will likely have come across references to most in journals, conferences, and blog posts over recent years. That's not really a complaint, though, and the selections are fully justified by the value of making available detailed behind-the-scenes accounts of these pioneering initiatives.

As Ridge explains in her introduction, the case studies are intended as both inspiration and practical how-to guide for GLAM practitioners contemplating crowdsourcing. They provide an opportunity to hear from people behind successful projects 'about their goals, their experiences building and launching crowdsourcing sites, what worked and what did not, how their designs improved over successive iterations, and how these projects changed the host organizations.'

Five of the studies are primarily about collaborative transcription — that is, enlisting the services of the online public to help copy archival texts into digital formats. This emphasis reflects the promise this particular kind of crowdsourcing holds for cultural heritage institutions. Manuscript transcription is often too prohibitively labour-intensive for archives to achieve through their own resources, leaving content less findable, readable, and useable than if fully digitised. Crowdsourcing offers one possible solution.

Of the examples discussed here, the *Old Weather* project involved establishing a website for the public to view early 20th century Royal Navy ship logs and transcribe the weather observations recorded there. The objective was to render this hitherto inaccessible meteorological data computable in a manner that would aid climate change research. With the *Transcribe Bentham* project, University College London sought online volunteers to assist with the task of transcribing their vast collection of manuscripts by 18th century social philosopher Jeremy Bentham. This work was fairly challenging, requiring users to not only grapple with arcane subject matter and the difficulties of 18th century handwriting, but also to encode the results in TEI markup language. The New York Public Library's *What's on the Menu?* project requested public help with the online transcription of their collection of heritage restaurant menus so that this previously under-utilised resource might be available as a database of dishes and prices. It proved so unexpectedly popular that work flows within the Library had to be adjusted to keep up with the demand for digitised menus to transcribe. (It's worth mentioning that one of the principle architects of this project, and a co-writer of this case study, Michael Lascarides, is now with our own National Library). The National Library of Wales's forays into crowdsourcing have included inviting online volunteers to copy out Welsh place names from old Ordnance Survey maps in order to create a georeferenced gazetteer; and preliminary work on a major initiative to crowdsource the transcription of the Library's heritage collection of handwritten wills. Finally, the *Papers of the War Department* project is an initiative to create a virtual

archive of late 18th Century US government documents from content dispersed across numerous physical repositories. To facilitate the online transcription of this archive, the Roy Rosenzweig Center for History and New Media created Scripto, an open source transcription tool developed with the community of researchers already using these papers.

Of the three remaining case studies, two concern social tagging whereby the public is invited to help improve the discoverability of collection items by adding descriptive 'tags'. The BBC's *Your Paintings* website, for example, displays images of all the paintings held in public ownership in the United Kingdom. The original metadata were restricted to basic taxonomic fields like title, artist, date, etc. However, by enabling online users to contribute tags that identify objects and themes in a painting, it becomes more usable through keyword searching. The Netherlands Institute for Sound and Vision has taken this a step further with its social tagging game *Waisda?*, developed to enhance the description of its large audiovisual archives. Drawing on the appeal of computer gaming, *Waisda?* allows online users to compete against one another by tagging the same video simultaneously. Points are awarded for speed and matching other people's tags (the underlying assumption being that tags quickly chosen by more than one are likely to prove the most useful). Finally, the remaining case study looks at Brooklyn Museum's experiments with crowd-curation, where the public can help select, or even contribute, content for an exhibition.

Useful as the above studies are, it is strange that none deal in any depth with what's probably the most common form of cultural heritage crowdsourcing: inviting users to provide further information about archival items. This may reflect Ridge's concern to differentiate crowdsourcing, as a task-oriented activity, from Web 2.0-style user-generated content projects that simply invite users to post a comment. However, while insisting that crowdsourcing is defined by intent to accomplish specific goals, Ridge adds that this applies "even if that goal is as broadly defined as 'gather information from the public about our collections'." Given this, it's disappointing that there isn't more discussion here about the very kind of crowdsourcing projects New Zealand cultural heritage institutions have most experimented with. Similarly, it would have been nice if one or two case studies had been of projects undertaken by smaller organizations, like community archives, to illustrate the particular crowdsourcing issues for them.

That said, these case studies still serve to identify issues that apply across a broad range of projects. A recurring theme is that not

only do most cultural heritage crowdsourcing initiatives usually attract a comparatively modest number of participants, the vast majority of work is done by an even smaller group of “super contributors”. *Transcribe Bentham*, for example, had around 3000 registered users but most transcriptions were by a core of about ten. Another is the way a crowdsourcing project can be transformed through the influence of its participants. *Old Weather*, for example, was originally conceived as an exercise in climate science but shifted towards maritime history once it was realised this was the interest sustaining its user base. A distinctive feature of cultural heritage crowdsourcing projects is that participants are largely motivated either by a pre-existing interest in the subject matter or by an altruistic desire to contribute to the public good. This suggests the value both of reaching out to already established communities of interest, and of drawing on the public’s desire to help as a way of fostering greater engagement with an institution’s collections.

The concluding four essays all try to better define the parameters and implications of cultural heritage crowdsourcing. Trevor Owens, for example, has two problems with the term ‘crowdsourcing’ from a GLAM perspective: the words ‘crowd’ and ‘sourcing’. He takes issue with the former, not just because participant numbers in cultural heritage projects tend to be small, but also because ‘crowd’ implies an anonymous, undifferentiated mass, and this doesn’t reflect the kind of engaged, self-selected user communities these projects attract. He objects to ‘sourcing’ because it alludes to the business practice of ‘outsourcing’ and consequently carries troubling implications of exploitation inappropriate to a cultural heritage ethos. Rather than abandon the concept altogether, Owens argues the need for crowdsourcing to be shaped and constrained in ways that render it a better fit with cultural heritage values. He positions it as continuous with long-standing traditions of volunteerism within cultural heritage, reinforcing rather than threatening established identities and practices. ‘When done well’, cultural heritage crowdsourcing projects should involve participants in meaningful work as informed and engaged non-professionals, not as a crowd to complete mechanical tasks.

Stuart Dunn and Mark Hedges take a somewhat similar tack, examining how best to frame crowdsourcing for humanities research. Like Owens, they’re anxious about the implications of crowdsourcing’s commercial origins, believing the need for it to be “reconfigured” for application within an academic or cultural heritage setting. They advocate the development of an ‘epistemic model’ of crowdsourcing, which is ultimately about ‘creating knowledge or enabling understanding’.

This contrasts with a 'business model' that is solely seen as a cost-effective means for enhancing digital content. Lori Byrd Phillips looks at the vexed question of how increased user collaboration might impact on the nature of institutional authority within the GLAM sector. While she's more welcoming than the aforementioned essays of crowdsourcing's potential to transform traditional practice, she still sees a need for balance. Her 'open authority' model is based around the metaphor of the Museum as *both* a Temple (a source of authoritative expertise) and a Bazaar (an open collaborative community). It draws on pedagogical theories of child-driven learning to propose a museological theory of practice driven by the interests of visitors.

Finally, Alexandra Everleigh's essay – arguably the most substantial here and certainly that of most direct interest to archivists – endeavours to untangle the somewhat confusing mix of different current approaches to crowdsourcing by archives. Her 'user participation matrix' provides an analytical framework for assessing these initiatives, particularly how they impact on professional practice. The four quadrants of her matrix situate crowdsourcing projects in terms of whether: they adhere to a radically transformative and democratising vision of archival practice (the Archival Commons model); seek to engage online communities on their own terrain, but are constrained by limited institutional objectives (the Outreach and Engagement model); invite collaboration in a manner that allows for the redrawing of boundaries between archives and their users, creating new ways of working (the Collaborative Communities model); or simply use volunteer labour as a cheap and effective means of accomplishing mechanical tasks (the Transcription Machine model).

If the case studies in *Crowdsourcing Our Cultural Heritage* testify to the current flurry around crowdsourcing within the GLAM sector, the analytical essays point to the present state of flux within our professional discourse as to its implications. For a concept and practice not yet a decade old, this uncertainty is not altogether surprising. What is clear, though, is that crowdsourcing is proving a significant development within the cultural heritage domain that all interested in archives should keep abreast of. This anthology provides one convenient and engaging way of doing so.

Reid Perkins
Upper Hutt City Library