

Curating Crowds: A Review of *Crowdsourcing Our Cultural Heritage* (Ashgate, 2014)

Victoria Van Hyning <victoria_at_zooniverse_dot_org>, Library of Congress

Abstract

Through case studies and theoretical reflections, Mia Ridge's edited volume *Crowdsourcing Our Cultural Heritage* makes a comprehensive addition to crowdsourcing research and practice. Authors discuss how issues of project roles, public and volunteer engagement, data use and user choice reshape institutional presence.

Crowdsourcing Our Cultural Heritage is an important collection for anyone working in cultural heritage or academia who is interested in the pros and cons of implementing crowdsourcing projects, whether for manuscript transcription, image or video tagging or crowd-curated exhibitions. Most essays in the collection, starting with the introduction by editor Mia Ridge, offer a definition of crowdsourcing, engage with some of the theoretical material pertaining to the topic such as James Surowiecki's *The Wisdom of Crowds* [Surowiecki 2004], and give an overview of the challenges that crowdsourcing can help GLAMs and academics overcome. Ridge is one of the most cogent advocates for, and careful critics of, crowdsourcing in cultural heritage industries, and she gets the volume off to a thought-provoking start with sections on "Key Trends and Issues" and "Looking to the Future of Crowdsourcing in Cultural Heritage". While the essays themselves are somewhat repetitive in terms of definitions and theoretical ground, almost any one of them could be read on its own and provide insight into the broad issues surrounding crowdsourcing. This volume would be valuable as a teaching resource for a range of specialists, including GLAM practitioners such as archivists and curators, as well as educators, audiovisual specialists, art historians, web designers and developers, and sociological and history of science theorists interested in crowdsourcing. Most articles include robust bibliographies with references to grey and formal publications. Although this is a fast-moving area of research and practice, the volume is still remarkably up-to-date over four years on from its publication in 2014.

1

Part I contains eight case studies: seven from UK and USA-based GLAMs, and one case study of a video tagging project from the Netherlands. The titles of many of the former refer to text transcription or metadata extraction projects, in which volunteers are invited to transcribe or add tags to digital images of texts held in the online catalogues of diverse repositories. And while each case study delivers useful insights into the transcription and tagging projects mentioned in their titles, each at least touches on a wider range of public engagement and crowdsourcing activities undertaken by the authors' respective GLAMs and/or universities or broadcasters. Some of these are in-person events such as "roadshows", while the use of surveys by numerous authors helps to surface users' or patrons' voices. There is a good balance between quantitative and qualitative assessment of projects' successes and failures as well as the reach and impact of digital initiatives. Most articles are illustrated with images of the web-based tools under discussion, and many include figures and tables communicating user participation and engagement.

2

Shelley Bernstein's opening essay "Crowdsourcing in Brooklyn", offers insights into the Brooklyn Museum's strategies of digital and in-person engagement over the better part of a decade, including the process of curating and displaying an exhibition with input from members of the public—*Click! A Crowd-Curated Exhibition* (<https://www.brooklynmuseum.org/exhibitions/click>). One of the strengths of Bernstein's piece is her acknowledgement of the design influences and goals of Flickr whose founder, Caterina Flake, said "You should be able to feel the presence of other people on the Internet", a principle Bernstein and her team translated for the GLAM setting: "How

3

could we highlight the visitor's voice in a meaningful way and utilise technology and the web to foster this exchange?" [Ridge 2014, 18]. She also engages with Surowiecki's idea put forward in *The Wisdom of Crowds* #surowiecki2004 that for crowds to be wise they must be diverse and their actors independent. Brooklyn attempted to foster both attributes in their open call for photographers to submit one image each on the theme of "Changing Faces of Brooklyn", to be judged by a crowd for inclusion in a new exhibition. 389 entries were assessed by 3,344 evaluators in a thoughtfully created interface that attempted to minimize outside influence on evaluators. 410,089 evaluations were submitted; the top 20% of images were then displayed as the exhibition *Click!*, drawing 20,000 visitors in six weeks. Bernstein provides a range of other useful statistics about user engagement with the evaluation interface and the exhibition. The remaining case studies include a Tinder-style app through which volunteers assess images quickly (*Split Second: Indian Paintings*, <https://www.brooklynmuseum.org/exhibitions/splitsecond>), and an open-studio tour program spanning 73 square miles and 67 neighbourhoods (Go). The article argues persuasively that GLAMs can work with visitors and volunteers to transform crowds into communities. Regional museums may be better suited to this approach than GLAMs such as the National Library of Wales, which is represented in a later chapter of the volume, whose authors remark on how relative regional isolation and poor public transportation have made online methods of crowdsourcing particularly useful and productive.

Chapter 2, "*Old Weather: Approaching Collections from a Different Angle*", by Lucinda Blaser (Royal Museums Greenwich), provides a valuable overview of a project in which volunteers transcribe historic ships' logs in an effort to extract climatological data that will improve the British Met Office's weather predication and climate models. As members of the *Old Weather* team have reported previously, volunteers' interest in the historic information they encountered along the way has proved an unexpected hit, and has been a major reason for sustained engagement with the project over time (<https://blog.oldweather.org/>). The project (<https://www.oldweather.org/>) is a collaboration between *Zooniverse* (www.zooniverse.org) – an online crowdsourcing research group based at the University of Oxford – the Adler Planetarium in Chicago, the University of Minnesota, the Met Office, National Maritime Museum and Naval-History.net, using original ship log sources held in the National Archives, UK (a number of spinoff projects since 2014 have added further material accessible through the same site). Blaser remarks that although the source material was not held at the National Maritime Museum, the museum held images of the ships and other material that "could engage users further with links to historic photographs that would bring these vessels to life, making this project more than just a two-dimensional transcription project" [Ridge 2014, 51]. Moreover, she asks if this model could be more widely applied, and poses a rhetorical question: "Do we have to be selfish and only think of ourselves in the results of crowdsourcing and citizen science projects, or is the ability to say that as an institution you have helped a large number of users engage with your subject matter in a meaningful way more than enough?" [Ridge 2014, 51].

Blaser discusses some of the challenges and opportunities inherent in crowdsourcing, including the need to foster communities over the long term and incorporate crowdsourced results into content management systems. She is admirably attentive to the experience of volunteers, quoting a number of contributors in her essay, and reflecting on the ways in which volunteers' engagement can result in new learning opportunities and a sense of fulfillment and ownership that can ultimately drive new research questions. She briefly mentions instances of crowd-curation' of exhibitions at Royal Museums Greenwich, such as *Beside the Seaside* and *Astronomy Photographer of the Year* "where crowdsourced images and collection items share the same gallery space" [Ridge 2014, 47]. Blaser argues that "crowdsourced displays will become more common" [Ridge 2014, 47] thus allowing volunteers to work directly with museum staff and feel greater ownership of collections. Her reflections link in well with the previous essay.

Tim Causer and Melissa Terras of University College London discuss the *Transcribe Bentham* project in chapter 3 of the volume. *Transcribe Bentham* invites members of the public to transcribe and apply TEI XML tags to Jeremy Bentham's voluminous archives, which are slowly being edited by a team at UCL. The essay describes the original transcription interface—a customized MediaWiki web application—the initial call for engagement, updates to the interface, funding, staffing, cost-effectiveness, quality control, as well as future collaborations (now underway) to use the vetted Bentham transcripts as training data for Handwriting Recognition Technology. Significantly, the authors acknowledge that the difficulty of the transcription and marking tasks led to a narrowing of participation and reliance on a small cohort of seventeen Super Transcribers (the threshold at which one becomes a Super Transcriber is not specified). They aver

that *Transcribe Bentham* might be better described as “crowd-sifting”, beginning with the traditional open call of crowdsourcing but resulting in the retention of a small group of highly dedicated individuals. Although the interface was tweaked to ease participation, the authors argue that it is worth attempting to attract more Super Transcribers than casual or short-term users. Other research teams, including Zooniverse, have attempted to lower barriers to participation by developing more granular approaches to text transcription. *Shakespeare’s World* and *AnnoTate*, both launched in late 2015, are transcription projects built with GLAM partners on the Zooniverse platform, for which I served as project lead. These allow participants to transcribe as little as a word or line on a page and have resulted in higher levels of participation from a broader base. For example, as of May 2017, volunteers who worked on fewer than nine pages contributed 20% of *Shakespeare’s World* transcriptions overall, a significant contribution.

Case study 4, “Build, Analyse and Generalise: Community Transcription of the *Papers of the War Department* and the Development of *Scripto*” by Sharon M. Leon, describes how the creation of a particular project resulted in the release of *Scripto* (www.scripto.org) “a customisable software library [built on WikiMedia] connecting a repository to an editing interface, and as extensions for three popular web-based content management systems” [Ridge 2014, 97] including Omeka, Drupal, and WordPress. *Papers of the War Department* (PWD) digitally assembles nearly 45,000 documents from archives in the US, Canada, Britain and France pertaining to the period 1784–1800. The papers had long been believed to be lost due to a fire in the War Office in 1800, which destroyed the central repository. Through the efforts of scholar Ted Crackel in the 1980s and 1990s, the whereabouts of copies and examples of the original correspondence were located and imaged, originally for the purposes of a printed edition, then a CD-ROM, and finally, in 2008, for *PWD*, which invites members of the public to transcribe the sources. The sources were lightly catalogued by experts by 2010, but as Leon points out this only opened the corpus to researchers who knew precisely what they were looking for, while those “with less concrete demands” [Ridge 2014, 92] found the early index less useful. Project funding was used to add more detailed metadata to a third of the collection, but could not stretch far enough to cover the whole. At this point, in 2013, *PWD* staff analyzed their site traffic and concluded they had a ready-made group of users who might be willing to contribute their own transcriptions and expertise back into the collection.

Before describing *Scripto* Leon gives an overview of some of the theoretical work and existing transcription tools and crowdsourcing platforms that inspired staff at the Roy Rosenzweig Center of History and New Media (RRCHNM) to engage with the public. She cites Max Evans’s “2007 call for commons-based peer production as a way to create ‘Archives of the People, by the People, for the People’” [Ridge 2014, 92]: Wikipedia, Flickr Commons, Zooniverse and *Transcribe Bentham*. Like many organisations that have harnessed crowdsourced transcription, RRCHNM realized that “public contributions [could] provide transcriptions where there once were none, and where there likely would be none in the future” [Ridge 2014, 96] due to budgetary constraints and the sheer scale of the job. Moreover “public contributions” where volunteers choose what to transcribe “can serve as a barometer of the most interesting materials within a particular collection” [Ridge 2014, 96] and perhaps have a bearing on editorial choices for print or digital editions in the future. Surely many publishers would be swayed by concrete evidence of this kind.

User testing of the early site led the team to implement a series of innovations to the standard MediaWiki transcription interface, for example showing the manuscript document at the top of the page and the transcription pane beneath. Login accounts are required and new users can wait a business day to be approved. I tested this on a working day, and was confirmed for a new account in less than twenty-four hours. The project team felt approved login was necessary to reduce vandalism and spam, but I would argue it probably acts as a deterrent to users who might feel motivated to engage, but are unwilling or unable to return to the project in future. The remainder of the case study traces the support, development, and editorial time devoted to DWP and the release and uptake of *Scripto*, which has been particularly popular amongst university libraries [Ridge 2014, 108].

Case study 5 returns us to New York, with an engaging piece titled “*What’s on the Menu?: Crowdsourcing at the New York Public Library*”, by Michael Lascarides and Ben Vershbow. The authors combine a detailed case study of a menus transcription and metadata extraction project launched in 2011 with up-to-date (and still relevant) analysis and insight into user motivation, usability, sustainability, and data ownership. Lascarides and Vershbow argue that “it needs to be made very clear at the outset that your library entirely owns the newly created data to do with whatever it wants, and that the participant willingly relinquishes any ability to restrict those rights” and that “usually, you will want to share the

[resulting] content that results from their labours as broadly as you can” [Ridge 2014, 122]. NYPL have perhaps been more explicit than others about the status of the data they collect. While most GLAMs want their data to be reusable and searchable through a web interface, and clearly state this in their mission statement and other materials geared towards potential volunteers, project owners could do more to highlight that any data produced through their interface will become the property of the institution.

After providing a clear overview of the site functionality, supported by images of the interface, the authors also highlight the depth of user engagement with *What's on the Menu* (<http://menus.nypl.org/>) during its first sixteen months: 163,690 visits, four million page views, and an average of 6.36 minutes on the site, compared to just 2.38 on nypl.org [Ridge 2014, 126], suggesting that transcription and other crowdsourcing interfaces offer patrons ways of engaging with and exploring collections that more traditional GLAM interfaces do not. In a “What’s Next?” section the authors advocate for “crowdsourcing at scale” in which “a new generation of reusable tools that require less maintenance and serve a wider variety of purposes” are deployed across most if not all NYPL domains, as opposed to building and attempting to maintain single stand-alone apps. Lascarides and Vershbow conclude with reflections on the gamification debate, arguing persuasively that participants are often motivated by the collections themselves and do not need an additional layer of play to become or remained involved in crowdsourcing. They cite a range of authorities including Trevor Owens (Library of Congress):

11

When done well, crowdsourcing offers us an opportunity to provide meaningful ways for individuals to engage with and contribute to public memory. Far from being an instrument which enables us to ultimately better deliver content to end users, crowdsourcing is the best way to actually engage our users in the fundamental reason that these digital collections exist in the first place. [Owens, 2012, cited [Ridge 2014], 131] [Owens 2012]

Like a number of other contributors to the volume, Lascarides and Vershbow emphasize the experimental and iterative nature of crowdsourcing projects at their institution (including others beyond *What's on the Menu?*, such as *GeoTagger*, a geo-referencing project). They write about paying down their “technical debt” after a successful beta test of *What's on the Menu*, by overhauling the original application code, implementing new visual design and elements of the user interface, with better search and browsing features, and perhaps most significantly, a new public API “to provide other application developers or digital researchers real-time data from the project” [Ridge 2014, 126].

Chapter 6, “What’s Welsh for ‘Crowdsourcing’? Citizen Science and Community Engagement at the National Library of Wales”, by Lyn Lewis Dafis, Lorna M. Hughes and Rhian James, reports on two crowdsourcing projects undertaken at the National Library of Wales (NLW): *The Welsh Experience of the First World War* (<http://cymru1914.org/>) collecting project and *Cymru1900Wales* (<http://www.cymru1900wales.org/>), a place name gathering project in partnership between NLW, the University of Wales, People’s Collections Wales, the Royal Commission on the Ancient and Historical Monuments Wales and Zooniverse. They also describe the digitized collection of *Welsh Wills Online*, a project with potential for adding crowdsourced transcription. The authors remark that the relatively remote location of the library has led the institution to focus on “mass digitisation of core collections to support access, preservation, research and education” [Ridge 2014, 139], as well as the provision of all tools and web services in both Welsh and English. Moreover, they argue that crowdsourcing “can [...] be seen as the logical development of a long tradition of research and engagement based on the Library’s collections” [Ridge 2014, 144].

12

Like others in the volume, the authors have recourse to the theoretical frameworks put forward by Jeff Howe of Wired magazine in regards to crowdsourcing and business practice. They conclude that crowdsourcing in the cultural heritage domain “seeks to utilise the multiple perspectives of the crowd”, a statement most clearly borne out in *The Welsh Experience of the First World War* which collected and digitized primary material provided by the public in a series of five “roadshows” held in geographically diverse parts of Wales. The roadshow format is not new, as the authors point out, citing the Oxford-based *Great War Archive* project, *Europeana 1914–1918* and the JISC-funded *Welsh Voices of the Great War Online*. But the project is different in that it aimed to digitize materials that would fill particular gaps in existing collections. The authors provide a list of those organizations they contacted and the advertising deployed to recruit participants, and reveal that while the 350 items that were digitized were diverse, they did not succeed in

13

gathering non-documentary or text-based items. They conclude that future marketing of roadshows would need to be much more targeted in order to capture other kinds of media.

Cymru1900Wales, the library's first crowdsourcing project, launched in September 2012 and asks volunteers to add local place name information to digitized Ordnance Survey maps from 1900. A number of research goals are referred to in broad strokes by the authors, for example the hope that the dataset will unlock social and linguistic history. Unlike the roadshows, this project was conducted entirely remotely, with academics and participants communicating via email, project blog, FaceBook and Twitter. This is particularly important for GLAMs that are remote from their patron base. *Welsh Wills Online* consists of 800,000 pages of wills and other legal documents collected in the Welsh ecclesiastical courts between the late-sixteenth century and 1858. Like so many of the text-based collections discussed throughout the volume, the materials here are not yet machine-readable, making manuscript transcription necessary if the contents of the images and original documents are to become word-searchable. At the time of writing, NLW had not yet embarked upon such a project and one does not appear to be under development at present.

14

The authors acknowledge that while crowdsourcing may have great promise, many GLAM and academic end-users, including those they surveyed, are anxious that projects be cost-effective. This manifests in two distinct but related anxieties: 1) that time spent setting up and maintaining projects should not exceed the amount of time it would take staff to do the core tasks associated with the project themselves, i.e. transcription and 2) that end-users, i.e. researchers, be able to make use of the results. Quality control and vetting, the authors argue, should not create a heavier burden than any work offset by the use of crowdsourcing. Like Causer and Terras above (*Transcribe Bentham*), the NLW team concludes that it might be best to attract and retain specialists or, to put it another way, a cohort of Super Transcribers. Again, drawing on my experience of *Shakespeare's World* in which volunteers transcribe a range of early modern English manuscripts that share some of the difficulties of the Welsh wills corpus, a significant proportion of volunteers are able to make meaningful contributions to transcription when given some guidance in the form of handbooks, tutorials, shortcut keys for common abbreviations, and so on. But however much we lower barriers to participation, GLAMs and academics still need time, money and support to deal with both the process and the products of crowdsourcing. In this regard, the authors' emphasis on the potential for crowdsourcing to save money is perhaps misleading, though even within the context of the present volume, it is a widely expressed view; one that may have its roots in crowdsourcing for business purposes. As Trevor Owens argues at the close of the volume, and as Blaser argues in chapter 2, crowdsourcing in GLAM and academic environments may save time on tasks such as transcription and metadata extraction, but ideally should create new roles dedicated to public engagement with collections and tweak existing roles and the ways in which GLAMs conceptualize their duties and interactions to patrons. GLAMs might, for example, spend more time nurturing public engagement projects and ingesting the products of crowdsourcing and other kinds of engagement projects into CMSs (content management systems), rather than having specialists add deeper metadata to a lightly catalogued collection.

15

In "*Waisda?: Making Videos Findable Through Crowdsourced Annotations*" (chapter 7), Johan Oomen, Riste Gligorov and Michiel Hildebrand describe two pilot projects that resulted in the contribution of over one million tags to a corpus of video clips in the Netherlands Institute for Sound and Vision, which holds over 750,000 hours of audiovisual material as of 2014. The primary audience for the archive is not the equivalent of library patrons or museum-goers, but broadcasters and journalists who seek out reusable content. Secondary and tertiary audiences are comprised of researchers and students who use materials in a broad range of disciplines, and "home users" who access the materials for "personal entertainment or a learning experience" [Ridge 2014, 169]. The opening pages of the article give an overview of the challenges of making non-machine readable datasets accessible through crowdsourcing, and describe various approaches to crowdsourcing and motivational factors for both GLAMs and "end-users" or participants. Many of these, such as "increasing connectedness between audiences and the archive" [Ridge 2014, 166] are echoed in contributions throughout this volume.

16

Unlike most of the other projects here, *Waisda?* deploys gamification strategies to engage users, and the authors report successful outcomes from what they call "serious game" play [Ridge 2014, 171]. As in the ESP Game, players of *Waisda?* accrue points if their tags match those of other players. *Waisda?* players can see their score relevant to other players and scorekeeping is split into a number of categories including "fastest typers" [Ridge 2014, 173]. Evaluation of

17

the results is provided for the first and second pilot studies, focusing on the overall usefulness of the tags created. As the authors indicate, some of these findings have been published at an earlier date, but an overview is offered here. Ultimately, they conclude that “using only verified user tags (i.e. where there was mutual agreement) for search gives poorer performance than search based on all user tags” [Ridge 2014, 179] and that search functionality improves with the addition of more tags. Like the *Transcribe Bentham* and *PWD* teams, the authors advocate for finding “super-taggers” rather than creating broad appeal, but they do acknowledge that Zooniverse offers an “alternative model”, which they describe as relying on a “sustainable ‘army’ of users” [Ridge 2014, 180]. That said, they do not elaborate how that army might have been engaged in the first place nor how *Waisda?* might emulate Zooniverse to create broader engagement. However, the research team have reuse and sustainability on their agenda, having published their code on GitHub, and connected with the European Film Gateway and *Europeana* [Ridge 2014, 180]. As Lascarides and Vershbow argue in chapter 5, reusability of apps is necessary for long-term sustainability.

Chapter 8, “*Your Paintings Tagger*: Crowdsourcing Descriptive Metadata for a National Virtual Collection” by Kathryn Eccles and Andrew Greg, describes the *Your Paintings* site hosted by the British Broadcasting Corporation (BBC), containing over 200,000 images of paintings in The Public Catalogue Foundation, and a metadata extraction project called *Your Paintings Tagger* (www.tagger.thepcf.org.uk formerly). While many articles in *Crowdsourcing our Cultural Heritage* directly invoke Zooniverse’s *Galaxy Zoo* and other scientific projects, *Your Paintings Tagger* (YPT) was built in partnership with Zooniverse, and *Galaxy Zoo* user motivations have been compared directly with those of YPT participants by researchers at Oxford and the University of Glasgow (the current chapter builds on previous work undertaken by Greg and Eccles). Not all Zooniverse components were deployed in the YPT, for example this project does not have a social forum or other mechanism whereby volunteers and experts can interact. It was only through surveys that the project team learned of volunteers’ desire for a social space. Like Causer and Terras of *Transcribe Bentham*, Greg and Eccles conclude that because most tags are contributed by a small cohort of volunteers, more should be done to engage and retain additional ‘super-taggers’, though the authors do gesture to the prospect that the threshold for agreement among taggers could be lowered and that paintings shown by some logic such as artist or time period might be more engaging than the default Zooniverse mechanism for presenting images, which is random.

The authors spend some early pages of the chapter discussing the complex negotiations between experts at the BBC, participating GLAMs and the University of Glasgow, who tried to pin down a suitable metadata format for *Your Paintings*, before the introduction of a crowdsourced dimension. Other GLAM practitioners may find this account useful when considering the institutional barriers they may need to overcome when trying to make collections more discoverable. It is notable however, that while Greg and Eccles, like others in this volume, suggest that crowdsourcing is more cost effective than traditional metadata improvement projects, time and money are still needed to support communities. Indeed, even without a social forum feature, which generally entails more staff time to maintain than projects without a social dimension, YPT is currently unavailable due to a funding shortage. The project owners are keen to implement changes to the platform and a call for donations (a form of crowdfunding) is prominent on the home page. Rather than conceiving of crowdsourcing as a cheap alternative to metadata extraction, we should focus on other benefits, for example the prospect of engaging people in new ways with collections they might not otherwise encounter; and in the case of tagging developing alternative languages for searching that enable broader access to online collections. Greg and Eccles do in fact report on these benefits throughout their piece, and acknowledge that at the rate of tagging reported in 2013 it would take a long time for the project to come to completion. Project owners will continue to experience the same disappointments over cost effectiveness and speed so long as the (narrow) messaging around the value of crowdsourcing remains the same.

Part II, “Challenges and Opportunities of Cultural Heritage Crowdsourcing”, contains four essays that address different aspects of the relationships between GLAMs and volunteers. Alexandra Eveleigh’s thoughtful and carefully balanced piece, “Crowding Out the Archivist? Locating Crowdsourcing within the Broader Landscape of Participatory Archives”, acknowledges some of the common concerns of archivists and domain specialists in engaging with the crowd—notably concerns about authority and accuracy—while also advocating for careful engagement with online communities. Eveleigh examines the “tension inherent between a custodial instinct to control context and authenticity, and a desire to share access and promote usage” [Ridge 2014, 212] of collections, and suggests that the reality of participatory

18

19

20

archival practices will cause neither the demise of the archivist specialist nor the complete revolution of their role, but rather that the ever-changing landscape of participatory technologies and projects will enable the curator/gatekeeper role to evolve and to place greater emphasis on the perspective of the user/volunteer. Eveleigh's piece brings many of the bubbling concerns from the case studies to the fore, and serves as a strong yet encouraging critique of GLAM practice with regards to crowdsourcing.

Stuart Dunn and Mark Hedges' "How the Crowd Can Surprise Us: Humanities Crowdsourcing and the Creation of Knowledge" is a follow on from their "Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research" [Hedges and Dunne 2012]. In the present essay they offer a series of definitions and typologies of crowdsourcing activities, which may be helpful for researchers interested in terminology and theories of crowdsourcing. They explore distinctions between crowdsourcing for business versus epistemic purposes, arguing that humanities crowdsourcing, while it may draw on "mechanical" micro-tasking approaches common in business crowdsourcing projects, can also provide the circumstances for knowledge co-creation, interpretation, creative responses, editing, investigation, and new research. They conclude that because a small number of people undertake the bulk of tasks in any given crowdsourcing project, "successful uptake of contributor effort in humanities crowdsourcing will be dependent on finding pockets of enthusiasm and expertise for specific areas" [Ridge 2014, 244].

21

The penultimate chapter is "The Role of Open Authority in a Collaborative Web" by Lori Byrd Phillips, which begins by quoting Jane McGonigal's "Gaming the Future of Museums" lecture [McGonigal 2008], and striking a note common to almost all of the other pieces: that there is "pent-up knowledge in museums" and "pent-up expertise" in the public that can be married up for the benefit of all involved. Perhaps more clearly than the other authors in the volume, Byrd Phillips argues that the increase in user-generated content created a "renewed need for authoritative expertise in museums" [Ridge 2014, 247]. This argument essentially turns the more familiar paradigm—that there are collections that cannot be unlocked without volunteer effort—inside out. The piece echoes ideas put forward by Eveleigh and draws on additional theoretical perspectives, including the Reggio Emilia approach to learning, a child-led educational model that emerged in post-WWII Italy. Byrd Phillips argues that this model may be particularly useful for museums wishing to create "opportunities for community learning and collaboration" [Ridge 2014, 259].

22

The final essay, Trevor Owens's "Making Crowdsourcing Compatible with the Missions and Values of Cultural Heritage Organisations" closes the volume on a confident and even utopian note, declaring that crowdsourcing should be a core function of the way in which GLAMs serve the public: "crowdsourcing is one of the most valuable experiences we can offer our users" [Ridge 2014, 279]. He argues that crowdsourcing, when done well, can engage users with content in active and meaningful ways – not as mechanical transcribers for instance but as 'authors of our historical record', who contribute their passion and time to tasks that on the one hand open up collections for new kinds of investigation, and which also enable users to encounter primary material more deeply than if they were simply browsing an online catalogue.

23

Owens is a rhetorically skillful proponent of what he calls "ethical crowdsourcing" which is as much focused on the experience of patrons or volunteers as on cultural heritage outcomes. He touches on the work of Surowiecki, the examples of *ReCaptcha*, *BabelZilla* – "an online community for developers and translators of extensions for Firefox web browser" – and *Galaxy Zoo*. Of the latter he argues: "all the work of the scientists and engineers that went into those systems are part of one big scaffold that puts users in a position to contribute to the frontiers of science through their actions on a website, without needing the skills and background of a professional scientist" [Ridge 2014, 276]. His concept of scaffolding is particularly relevant in light of two new platforms, which enable anyone to create their own free project: www.zooniverse.org/lab (*Zooniverse Project Builder*) and <https://crowdcrafting.org/> (*crowdcrafting*), both launched after the publication of Ridge's volume. Finally, as if in answer to some of the contradictory statements about cost-effectiveness and emerging modes of engaging with the public that have been put forward by various case study authors in Part I, Owens argues that "in the process of developing [...] crowdsourcing projects we have stumbled onto something far more exciting than speeding up or lowering the costs of document transcription" [Ridge 2014, 277]. He closes with an example of transcription of Civil War diaries from the University of Iowa Libraries' DIY history site <http://diyhistory.lib.uiowa.edu/>, whose former head of Digital Library Services, Nicole Saylor, sees transcription as a "wonderful by-product" of a process of engaging the public with history. This model is a more realistic image of what

24

GLAMs can hope to achieve by deploying crowdsourcing.

Crowdsourcing our Cultural Heritage has much to offer a range of researchers and GLAM practitioners both in terms of particular examples of projects focused on a diverse range of media, and in terms of the evolving and complex debates about the role of crowdsourcing and public engagement in GLAMs and academia. This is an excellent starting place for anyone interested in studying crowdsourcing or embarking upon or improving existing projects.

25

Acknowledgements

This review was written while I was a British Academy Postdoctoral Fellow at the University of Oxford and Pembroke College, and the Zooniverse Humanities Principal Investigator in 2017, prior to my relocation to the Library of Congress, Washington DC, where I serve as Senior Innovation Specialist and Community Manager for By the People, a new crowdsourcing initiative (crowd.loc.gov).

26

Works Cited

Hedges and Dunne 2012 Hedges, Mark and Stuart Dunne. "Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research." [https://kclpure.kcl.ac.uk/portal/en/publications/crowdsourcing-scoping-study\(abf40e6d-7ece-4d76-94d2-36f91b5707ff\).html](https://kclpure.kcl.ac.uk/portal/en/publications/crowdsourcing-scoping-study(abf40e6d-7ece-4d76-94d2-36f91b5707ff).html) (2012).

McGonigal 2008 McGonigal, Jane. "Gaming the Future Museums." Lecture presented at Newseum, Washington, D.C., hosted by American Alliance of Museums (AAM), <https://www.youtube.com/watch?v=zJ9j7kIZuoQ> (2008, published 2012.)

Owens 2012 Owens, Trevor. "Crowdsourcing Cultural Heritage: The Objectives are Upside Down." <http://www.trevorowens.org/2012/03/crowdsourcing-cultural-heritage-the-objectives-are-upside-down/> (2012).

Ridge 2014 Ridge, Mia (ed). *Crowdsourcing Our Cultural Heritage* Ashgate, Farnham (2014).

Surowiecki 2004 Surowiecki, James. *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Doubleday and Co., New York (2004).