

Draft: Crowdsourcing in cultural heritage: a practical guide to designing and running successful projects

[Please note: the final version of this chapter is available in the [Routledge International Handbook of Research Methods in Digital Humanities](#), edited by Kristen Schuster and Stuart Dunn.]

Mia Ridge, British Library

Have you ever wanted to recruit hundreds of members of the public to assist with the task of making cultural heritage collections findable online? Or to connect with passionate volunteers who'll share their discoveries with you?

Crowdsourcing in cultural heritage is a broad term for projects that ask the public to help with tasks that contribute to a shared, significant goal or research interest related to cultural heritage collections or knowledge.¹ As participants receive no financial reward, the activities and/or goals should be inherently rewarding for those volunteering their time. This definition is partly descriptive and partly proscriptive, and this chapter is *largely concerned with explaining/describing* how to meet the standards it implies.

One of the key challenges that projects face is creating interfaces that turn a series of tasks that create and validate usable outpoints, whether transcribing, describing or collecting source collections, into an enjoyable experience. As crowdsourcing is inherently productive in intent, each activity should contribute to a meaningful, collective goal. This chapter will help cultural heritage practitioners and digital humanists plan and design rewarding crowdsourcing projects with tasks and 'behind the scenes' processes that contribute to a meaningful wider outcome. Understanding the motivations of cultural heritage organisations and the backstage work that goes into building a crowdsourcing project should also help academics and others seeking to collaborate with or study crowdsourcing projects and cultural heritage institutions.

It introduces key principles and stages in developing crowdsourcing projects and designing interfaces and communications that link to participant motivations. Based on the author's extensive practical experience and theoretical engagement with the field, it discusses topics including: choosing appropriate measures of success for evaluating projects; finding the right balance between productivity and engagement; validating and integrating the results of crowdsourced tasks into core collections

¹ Mia Ridge, 'From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing', *Curator: The Museum Journal*, 56.4 (2013).

systems; organisational and volunteer motivations for participation; and the organisational and personal impact of crowdsourcing.

An introduction to crowdsourcing in cultural heritage

Crowdsourcing as we know it has been transformed by technology, but cultural heritage, scientific and other knowledge-based projects have a long history of asking people to voluntarily collect information and objects.² From the 1850s, Joseph Henry's meteorological observation project at the Smithsonian asked volunteers to submit weather observations via the still-new telegraph network.³ Later, through correspondence with the Smithsonian's second Secretary, Spencer F. Baird, participants' own research and contact with the wider scientific community was facilitated, so that each group benefitted in ways that were meaningful to them.⁴

Some aspects of crowdsourcing - particularly 'citizen science' and 'citizen history' - also draw on a more recent history of public participation in scientific research.⁵ Citizen science projects involve members of the public assisting professional scientists with research,⁶ most commonly through data processing tasks like image classification but potentially also through fieldwork or observation tasks, data analysis or research design.⁷ Humanities scholars interested in public participation in scholarly research

² See, for example, the special edition of *Science in Context* (2011) on 'Lay Participation in the History of Scientific Observation', Anne Secord, 'Corresponding Interests: Artisans and Gentlemen in Nineteenth-Century Natural History', *The British Journal for the History of Science*, 27.04 (1994), 383-408

<<https://doi.org/10.1017/S0007087400032416>>. and Jonathan Silvertown, 'A New Dawn for Citizen Science', *Trends in Ecology & Evolution*, 24.9 (2009), 467-71 <<https://doi.org/10.1016/j.tree.2009.03.017>>. Jeremy Vetter, 'Introduction: Lay Participation in the History of Scientific Observation', *Science in Context*, 24.02 (2011), 127-141 <<https://doi.org/10.1017/S0269889711000032>>.

³ Smithsonian Institution Archives, 'Meteorology', *Smithsonian Institution Archives*, 2012 <<https://siarchives.si.edu/history/featured-topics/henry/meteorology>> [accessed 25 November 2017].

⁴ Daniel Goldstein, "'Yours for Science": The Smithsonian Institution's Correspondents and the Shape of Scientific Community in Nineteenth-Century America', *Isis*, 85.4 (1994), 573-599.

⁵ Rick Bonney and others, *Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report* (Washington D.C.: Center for Advancement of Informal Science Education (CAISE), July 2009), pp. 1-58

<<http://caise.insci.org/uploads/docs/PPSR%20report%20FINAL.pdf>>.

⁶ M. Jordan Raddick and others, 'Galaxy Zoo: Exploring the Motivations of Citizen Science Volunteers', *Astronomy Education Review*, 9.1 (2010), 18.

⁷ Bonney and others.

may find the significant body of prior work on this topic by citizen science researchers particularly valuable.

An example from the humanities also neatly encapsulates many aspects of crowdsourcing. The editors of the Oxford English Dictionary (OED) appealed for examples and definitions of words from the public in the 1850s and in 1879.⁸ Indexing, storing and managing the slips of paper subsequently received was a considerable task, as was coordinating and targeting requests for information about specific words. The OED continues to appeal to the public for help defining or providing examples of words in the present day.

These early projects sought to gather data at a geographic and quantitative scale not possible for individuals acting alone by adding documentation and communication tasks to existing leisure activities like observing wildlife or reading historical books. However, the manual work of compiling the information received was time-consuming, and projects could easily fall behind in processing and analysing the incoming data.

The availability of the web as a platform has transformed crowdsourcing.⁹ Data can be easily entered via websites or applications, automatically validated against set criteria and aggregated with other data. Sites can acknowledge and thank contributors immediately, and if tasks are carefully designed, they can even provide instant feedback on the quality of contributions. For institutions that previously relied on volunteers having physical access to collections or records, remote contributions based on digital images relieves physical conservation requirements and pressures on venue space and hours.

Reaching potential participants is also easier online. Social media and specialist email lists or discussion boards can reach broad or niche audiences to advertise a project, according to the skills or numbers needed to complete tasks. In addition to traditional scholarly publications resulting from projects, email newsletters and blog posts can provide more timely updates on progress and developments. Unlike volunteer projects that require attendance at specific locations and times, crowdsourcing volunteers can

⁸ Peter Gilliver, “‘Your Dictionary Needs You’: A Brief History of the OED’s Appeals to the Public”, *Oxford English Dictionary*, 2012

<<https://public.oed.com/history/history-of-the-appeals/>>.

⁹ And of course, these old methods on new platforms led Jeff Howe and Mark Robinson to coin a term to describe the act of taking work once performed within an organisation and outsourcing it to the general public through an open call for participants Jeff Howe, ‘The Rise of Crowdsourcing’, *Wired*, June 2006

<http://www.wired.com/wired/archive/14.06/crowds_pr.html> [accessed 6 January 2013].

contribute from anywhere in the world at any time of day or night, choosing tasks that match their interests and the time they have available.

Key conceptual and research frameworks

As evident in the paragraphs above, research on volunteer work in cultural heritage organisations and open source software has been particularly useful for thinking about fundamental aspects of participation in crowdsourcing. To understand participant motivations, I referenced research from related fields including citizen science, cultural heritage volunteering,¹⁰ 'serious' leisure,¹¹ commercial crowdsourcing, contributions to open source software and *Wikipedia*, and the emerging literature on cultural heritage crowdsourcing. While some aspects of volunteering enabled and mediated by online tools are novel,¹² volunteers are still looking for a meaningful leisure activity that fits into their life - some just want casual activities they can pick up whenever suits them, others want an opportunity to develop deeper skills and interests, or to socialise with other people with similar interests. Many of the skills needed to work with in-person volunteer or community programmes are similar to the community engagement and management skills needed for online projects.

As with traditional volunteering, crowdsourcing is not merely an opportunity to get work done - it is also an opportunity to engage the public with collections, encouraging curiosity and learning as participants pay close attention to collection items. However, the need to justify the resources required to run projects designed to enhance collections can put pressure on projects to focus on productivity at the expense of participant enjoyment. This chapter will discuss some of the tensions between designing for productivity - the number of items processed and rate of data production - and for public engagement over the life of a project.

When thinking about the impact of interface, task and workflow design I drew on research from the fields of human-computer interaction, user experience design and usability. My views on the role of design in enabling curiosity and learning, and the potential of crowdsourcing for deeper engagement were informed by research on

¹⁰ Kirsten Holmes, 'Volunteers in the Heritage Sector: A Neglected Audience?', *International Journal of Heritage Studies*, 9.4 (2003), 341-355 <<https://doi.org/10.1080/1352725022000155072>>.

¹¹ Robert A. Stebbins, 'Casual Leisure: A Conceptual Statement', *Leisure Studies*, 16.1 (1997), 17-25 <<https://doi.org/10.1080/026143697375485>>.

¹² See for example Clay Shirky, *Cognitive Surplus: Creativity and Generosity in a Connected Age*. (London, U.K.: Penguin, 2011).

communities of practice,¹³ learning through legitimate peripheral participation,¹⁴ instructional design¹⁵ and museum studies.

While the availability of crowdsourcing platforms¹⁶ has reduced the technical overhead of setting up a project and managing the resulting data, they have also increased the competition for participants. Much of this chapter discusses design principles that will help projects attract and retain participants. While the underlying principles may hold, new and refined design techniques are still emerging. The best way to keep up is to try participating in newly launched or refreshed projects; a process which also provides valuable, grounded insights into volunteers' motivations, different types of barriers to participation, and the impact of communications and institutional processes. It also develops your ability to critically assess projects and define a shared language to discuss your own ideas.¹⁷

A brief note on language - throughout this chapter the shorthand 'crowdsourcing' will stand for 'crowdsourcing in cultural heritage and the humanities'. Discomfort with the term 'crowdsourcing' has led some to use terms such as 'community-sourcing',¹⁸

¹³ Etienne Wenger, 'Communities of Practice and Social Learning Systems: The Career of a Concept', in *Social Learning Systems and Communities of Practice* (Springer Verlag and the Open University, 2010).

¹⁴ Gabriel Mugar, Carsten Østerlund, Katie DeVries Hassman, and others, 'Planet Hunters and Seafloor Explorers: Legitimate Peripheral Participation Through Practice Proxies in Online Citizen Science', 2014
<http://crowston.syr.edu/sites/crowston.syr.edu/files/paper_revised%20copy%20to%20post.pdf>.

¹⁵ David Wood, Jerome S. Bruner, and Gail Ross, 'The Role of Tutoring in Problem Solving', *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 17.2 (1976), 89–100; Priya Sharma and Michael J. Hannafin, 'Scaffolding in Technology-Enhanced Learning Environments', *Interactive Learning Environments*, 15.1 (2007), 27–46
<<https://doi.org/10.1080/10494820600996972>>.

¹⁶ Platforms available include Flickr Commons, the Zooniverse Project Builder, FromThePage and Pybossa.

¹⁷ Sample projects to try are listed in links from Mia Ridge, 'British Library Digital Scholarship Course 105: Exercises for Crowdsourcing in Libraries, Museums and Cultural Heritage Institutions', 2018
<<https://docs.google.com/document/d/1tx-qULCDhNdHoJyURqXERoPFzWuCreXAsiwHlUKVa9w/>> [accessed 22 October 2018].

¹⁸ Amy Sample Ward, 'Crowdsourcing vs Community-Sourcing: What's the Difference and the Opportunity?', *Amy Sample Ward's Version of NPTEch*, 2011
<<http://amysampleward.org/2011/05/18/crowdsourcing-vs-community-sourcing-whats-the-difference-and-the-opportunity/>> [accessed 6 January 2013].

'nichesourcing',¹⁹ 'micro-volunteering'²⁰ or 'targeted crowdsourcing',²¹ acknowledging that the 'crowd' is often neither large nor truly anonymous. At times I use the acronym 'GLAMs' (galleries, libraries, archives and museums) as shorthand for 'cultural heritage institutions'. Crowdsourcing projects that seek to engage members of the public may also be described as a form of public history or public digital humanities. By enhancing digitised collection records, crowdsourcing in GLAMs can enable digital humanities projects, but the potential for more integrated projects between these groups is relatively little explored.

A single chapter cannot provide a definitive account of such a large, constantly changing topic. Instead, my aim is to provide a common language for discussing crowdsourcing, outline issues for consideration in planning and running projects, and provide pointers to further information. Many crowdsourcing projects are committed to transparency about their processes and results, contributing to a field rich in formal and informal publications,²² including blog posts, conference presentations and newsletters from project stakeholders and contributors. Conference papers and publications on human computation, collective intelligence and computer-supported cooperative work (CSCW) contain a lot of deep technical expertise from specialist researchers in related fields, and are worth seeking out where relevant to specific design questions. Subjects not covered in this chapter include crowdfunding, user-generated content, the 'wisdom of the crowd', co-production or co-curation, or commercial crowdsourcing on platforms such as Amazon's Mechanical Turk.

¹⁹ Victor de Boer and others, 'Nichesourcing: Harnessing the Power of Crowds of Experts', in *Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2012* (Springer, 2012), pp. 16–20 <http://dx.doi.org/10.1007/978-3-642-33876-2_3> [accessed 12 May 2013].

²⁰ A term used by the New York Public Library. Michael Lascarides and Ben Vershbow, 'What's on the Menu?: Crowdsourcing at the New York Public Library', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014) <<http://www.ashgate.com/isbn/9781472410221>>.

²¹ Lyn Lewis Dafis, Lorna M. Hughes, and Rhian James, 'What's Welsh for "Crowdsourcing"? Citizen Science and Community Engagement at the National Library of Wales', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014) <<http://www.ashgate.com/isbn/9781472410221>>.

²² For example, Rose Holley's extensive publications, the 'Meta' publications about the Zooniverse project (<https://www.zooniverse.org/about/publications>) and individual books and reports including: Crowdsourcing Consortium, *Engaging the Public: Best Practices for Crowdsourcing Across the Disciplines* (University of Maryland, May 2015) <<http://crowdconsortium.org/>>; Nina K. Simon, *The Participatory Museum*, 2010 <<http://www.participatorymuseum.org/chapter4/>> [accessed 11 March 2013]; United States Government, 'Federal Crowdsourcing and Citizen Science Toolkit', *CitizenScience.Gov* <<https://www.citizenscience.gov/toolkit/>> [accessed 9 December 2018].

Fundamental concepts in cultural heritage crowdsourcing

I will introduce some of the fundamental concepts in crowdsourcing by describing tasks commonly found in cultural heritage projects, with examples for each.

When teaching crowdsourcing in GLAMs, I use a simple, informal categorisation of participant actions in conjunction with categories of task size and role.²³ I group tasks into three types, according to their size and role: microtasks, macrotasks, and meta tasks, which I will briefly define before describing participant actions.²⁴

Microtasks are small, rapid, self-contained tasks. For example, the New York Public Library's *Building Inspector*²⁵ has broken the task of checking building shapes and text transcribed from historical fire insurance maps into five extremely focused, tiny microtasks embedded in a specialised interface. Microtasks can be addictively satisfying because several can be completed in a short amount of time. Tasks such as tagging images are popular microtasks, and in some cases, the unpredictability of the items that appear in a task can 'hook' participants.

Macrotasks are longer, and/or more complex tasks that often involve higher order decisions about what to record and how. The text transcription task in *Transcribe Bentham*²⁶ is a macrotask because the handwriting is difficult to decipher, it transcribes page-by-page rather than line-by-line, and because participants can also 'mark-up' transcribed text to highlight insertions, deletions, etc., adding complexity to the task.²⁷

²³ A more academic approach is discussed in Mia Ridge, 'Making Digital History: The Impact of Digitality on Public Participation and Scholarly Practices in Historical Research' (unpublished Ph.D., Open University, 2015) <<http://oro.open.ac.uk/45519/>>.

²⁴ The categories used here are designed to provide an overview of task types for people planning crowdsourcing projects, rather than formalise a typology. The most useful typology will depend on the context in which it is being used. In 2011, I devised 'activity types' related to crowdsourcing games in museums, and in 2012 Dunn and Hedges devised a typology for academic humanities crowdsourcing based on asset type, process type, task type, and output type. Mia Ridge, 'Playing with Difficult Objects – Game Designs to Improve Museum Collections', in *Museums and the Web 2011: Proceedings*, ed. by Jennifer Trant and David Bearman (presented at the Museums and the Web 2011, Toronto, Canada: Archives & Museum Informatics, 2011) <http://www.museumsandtheweb.com/mw2011/papers/playing_with_difficult_objects_game_designs_to> [accessed 7 June 2013]; Stuart Dunn and Mark Hedges, 'Crowd-Sourcing as a Component of Humanities Research Infrastructures', *International Journal of Humanities and Arts Computing*, 7.1–2 (2013), 147–69 <<https://doi.org/10.3366/ijhac.2013.0086>>.

²⁵ <http://buildinginspector.nypl.org/>

²⁶ <http://blogs.ucl.ac.uk/transcribe-bentham/>

²⁷ Causer and Terras report that the requirement to mark-up the text in *Project Bentham* appears to be an added 'aggravation'. Tim Causer and Melissa Terras, "Many Hands Make Light Work. Many Hands Together Make Merry Work": *Transcribe*

Metatasks are activities that relate to the overall project, rather than individual tasks. This includes taking part in project design or analysis, and contributing questions, comments and answers to participant discussion fora. The *Old Weather* forum²⁸ is a justly famous example of the benefits of participant discussion, with a wealth of information shared and topics discussed.

Participant actions can be described according to how much creative freedom they have when completing the task and where it fits into the overall workflow. An informal grouping I have found effective in teaching crowdsourcing is: 'type what you see', 'describe what you see', 'share what you know', 'share what you have', 'validate other inputs'.

'Type what you see' tasks ask participants to type out or correct transcriptions from the item presented to them, and offer very little creative freedom. These tasks may be micro- or macrotasks. Transcription has been called a 'mechanical' task²⁹ but the difficulty varies according to the source material - printed text is easier to decipher than unfamiliar older forms of handwritten text with unorthodox orthography that may require the transcriber to make difficult decisions. The National Library of Australia's Trove platform for newspaper collections³⁰ includes functions to correct errors in automatically-generated text, and has been both hugely influential and productive.³¹ Other examples include the New York Public Library's What's on the Menu³² project. Like Trove, the Menus interface shows the benefits of expert attention during the design process - the front page anticipates and addresses common barriers to participation, provides a range of tasks to suit different preferences, and the task

Bentham and Crowdsourcing Manuscript Collections', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge, Digital Research in the Arts and Humanities (Farnham, Surrey, UK: Ashgate, 2014) <<http://www.ashgate.com/isbn/9781472410221>>. See also: Tim Causer and Valerie Wallace, 'Building A Volunteer Community: Results and Findings from Transcribe Bentham', *Digital Humanities Quarterly*, 6.2 (2012) <<http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>> [accessed 5 August 2013]; Doug Reside, 'Crowdsourcing Performing Arts History with NYPL's ENSEMBLE', 2014 <<http://dharchive.org/paper/DH2014/Paper-131.xml>> [accessed 22 February 2015].

²⁸ <http://forum.oldweather.org/>

²⁹ Stuart Dunn and Mark Hedges, *Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research* (London, U.K.: King's College, 2012), p. 56 <<http://www.humanitiescrowds.org/>>.

³⁰ <https://trove.nla.gov.au/newspaper/>

³¹ Rose Holley, 'Crowdsourcing: How and Why Should Libraries Do It?', *D-Lib Magazine*, 16.3/4 (2010) <<https://doi.org/10.1045/march2010-holley>>; Rose Holley, *Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers* (Canberra: National Library of Australia, 2009).

³² <http://menus.nypl.org/>

itself is tightly focused on the transcription task, with items pre-processed to minimise distractions.

Transcription tasks may require a single contributor to transcribe an entire passage or page of text or audio, or they may break the task into smaller components (e.g. a line of text or a snippet of a recording). The British Library's In the Spotlight ³³ project first asks participants to mark out the titles of plays on historical playbills; marked titles are then transcribed in a separate task. These tasks may not offer much creative freedom, but they can be immensely engaging, and lead to exploration of the collections and related topics outside the task.

'Describe what you see' tasks are designed to annotate items with additional information from formal taxonomies or informal folksonomies,³⁴ and includes identification and classification tasks such as tagging items with descriptive keywords. Image tagging on Flickr Commons³⁵ is perhaps not quite 'crowdsourcing', as the tagging activity can be spontaneous rather than being directly requested by the relevant GLAMs, but it provides a good example of the benefit of user-contributed keywords in aiding discoverability. [image - screenshot of results?] Other early, influential projects include the art tagging projects *steve.museum*,³⁶ Brooklyn Museum's game, *Tag! You're It*,³⁷ and *Waisda?*³⁸ for video tagging. The BBC's World Service Archive prototype used a combination of crowdsourcing and automated tagging on audio files.³⁹ Non-text forms of descriptive annotation include the Klokan

³³ <http://playbills.libcrowds.com/>

³⁴ Thomas Vander Wal, 'Folksonomy', *Vanderwal.Net*, 2007
<<http://vanderwal.net/folksonomy.html>> [accessed 8 December 2018].

³⁵ <https://www.flickr.com/commons/>

³⁶ Archived at <https://web.archive.org/web/sitemap/steve.museum>

³⁷ Shelley Bernstein, 'Crowdsourcing in Brooklyn', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014)
<<http://www.ashgate.com/isbn/9781472410221>>.

³⁸ Johan Oomen, Riste Gligorov, and Michiel Hildebrand, 'Waisda?: Making Videos Findable through Crowdsourced Annotations', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014)
<<http://www.ashgate.com/isbn/9781472410221>>.

³⁹ Yves Raimond, Michael Smethurst, and Tristan Ferne, 'What We Learnt by Crowdsourcing the World Service Archive', *BBC R&D*, 2014
<<http://www.bbc.co.uk/rd/blog/2014/08/data-generated-by-the-world-service-archive-experiment-draft>> [accessed 15 September 2014].

Georeferencer implemented by the British Library⁴⁰ and the Micropasts ‘photomasking’ task that helps generate 3D models from photographs.⁴¹

‘Share what you know’ tasks may collect factual information or personal stories about collections by drawing on existing knowledge, or asking volunteers to conduct research. The Lives of the First World War⁴² project asked participants to commemorate people who served in the war by ‘sharing their stories, find their records and adding known facts’, targeting the enthusiasm and research abilities of family and local historians. The Museum of Design in Plastics 10 Most Wanted project crowdsourced research into their specialist collection,⁴³ and comments on Flickr Commons sometimes record personal research or family stories about people, places, artefacts and events.

‘Share what you have’ projects collect items physically or digitally. RunCoCo’s Community Collection Model⁴⁴ has been adapted by Europeana for their First World War and Migration collecting projects.⁴⁵ The British Library’s UK Soundmap project collected audio recordings over 2010-11.⁴⁶

Tasks to ‘validate other inputs’ can be designed to crowdsource quality control processes for content created in other tasks. They tend to occur within ‘ecosystems’ of tasks,⁴⁷ a design pattern in which task interfaces or applications are combined to process different aspects of the same source materials. *Building Inspector* is an example of this, as each of the five tasks offered contribute to the larger goal of digitising the maps. Validation tasks may be micro-, macro- or meta-tasks, and include checking tags or annotations added by others, or moderating forum discussions. Increasingly,

⁴⁰ <http://www.bl.uk/georeferencer/> See Christopher Fleet, Ki Kowal, and Petr Přidal, ‘Georeferencer: Crowdsourced Georeferencing for Map Library Collections’, *D-Lib Magazine*, 18.11/12 (2012) <<https://doi.org/10.1045/november2012-fleet>>.

⁴¹ Bart Veldhuizen and Adi Keinan-Schoonbaert, ‘MicroPasts: Crowdsourcing Cultural Heritage Research’, *Sketchfab Blog*, 2015 <<https://blog.sketchfab.com/micropasts-crowdsourcing-cultural-heritage-research/>> [accessed 8 December 2018].

⁴² <https://livesofthefirstworldwar.org/>

⁴³ Susan Lambert, Marcus Winter, and Phil Blume, ‘Getting to Where We Are Now’, *10most.Org.Uk*, 2014 <<http://10most.org.uk/content/getting-where-we-are-now>> [accessed 4 March 2015].

⁴⁴ Ylva Berglund Prytz, ‘The Oxford Community Collection Model’, *RunCoCo*, 2013 <<http://blogs.it.ox.ac.uk/runcoco/2013/06/24/the-oxford-community-collection-model/>> [accessed 22 October 2018].

⁴⁵ <https://contribute.europeana.eu/migration>

⁴⁶ <https://sounds.bl.uk/Sound-Maps/UK-Soundmap>

⁴⁷ Also described as ‘suites’ in Ridge, ‘Game Designs to Improve Museum Collections’.

participants are verifying the results of tasks by software, not people, as 'human computation' systems develop.⁴⁸

Why do cultural heritage institutions support crowdsourcing projects?

Understanding why cultural heritage institutions undertake crowdsourcing projects provides important context, not only for measuring their success but also for understanding some of the barriers to success they face. The most obvious reason is that the size of the backlog of collection items needing transcription or description is beyond the scope of 'business as usual' projects. Resources are rarely available to adequately catalogue or describe digitised collection items held by museums, libraries, archives and other institutions. Software designed to transcribe printed or handwritten text rarely runs without some percentage of character- or word-level errors, hindering full-text search. Images, audio and moving image files often similarly lack detailed information about subjects depicted; audio transcription may not be 100% accurate and cannot recognise subtle references to important individuals, events or subjects that a human can. If digital images or media files can be shared on crowdsourcing interfaces, tasks such as those discussed above can be applied to them.

Keywords and phrases suggested by the public can bridge the 'semantic gap' between the language used in catalogues designed for internal or specialist users, and the everyday language used by the public, to make collection items more discoverable.⁴⁹

As deeply specialist roles have been phased out and curatorial or reference teams are asked to cover longer periods or wider regions of specialist collections, it is increasingly likely that the most expert person on a specific item or collection may not work for the institution; crowdsourcing can create opportunities for them to share their expert knowledge with an institution.

A number of projects have shown that crowdsourcing can create meaningful experiences with collections, and provide opportunities for learning and delight.⁵⁰ Well-designed projects can help people discover new interests, communities, or just encourage them to have a brief moment of deeper engagement with cultural heritage.

⁴⁸ Rachel Collings, 'The Art of Computer Recognition', *Art UK*, 2015 <<https://artuk.org/about/blog/the-art-of-computer-recognition>> [accessed 22 October 2018]; E. J. Crowley and A. Zisserman, 'The Art of Detection', 2016 <<https://www.robots.ox.ac.uk/~vgg/publications/2016/Crowley16/crowley16.pdf>> [accessed 22 October 2018].

⁴⁹ Jennifer Trant, *Tagging, Folksonomy and Art Museums: Results of Steve.Museum's Research* (Archives & Museum Informatics, 7 January 2009), p. 197 <<https://web.archive.org/web/20100210192354/http://conference.archimuse.com/files/trantSteveResearchReport2008.pdf>>.

⁵⁰ Ridge, 'From Tagging to Theorizing'; Ridge, 'Making Digital History: The Impact of Digitality on Public Participation and Scholarly Practices in Historical Research'.

This makes crowdsourcing a good fit for institutions whose missions encourage access, creativity, engagement or learning through their collections and knowledge.

Why do people contribute to crowdsourcing projects?

Understanding participant motivations is vital for designing successful projects that can attract and retain participants. Research on traditional volunteering, citizen science and GLAM crowdsourcing projects has provided insights into why people donate their time. Research into volunteering by psychologists Clary et al found six groups of motivations for volunteers: values ('altruistic and humanitarian concerns for others'), understanding (new learning experiences and the chance to practice knowledge, skills and abilities), social 'relationships with others', career-related benefits, ego-protective ('eliminating negative aspects surrounding the ego'), and enhancement (positive strivings for growth and development).⁵¹ Research with museum volunteers found that 'doing something enjoyable', an interest in the subject, meeting people and 'making friends' were the main reasons for volunteering.⁵²

Zooniverse projects have made a substantial contribution to research on motivations in citizen science. In one study, nearly 40% of *Galaxy Zoo* participants selected 'I am excited to contribute to original scientific research' as their main motivation, with the next most common primary motivation being: 'I am interested in astronomy'.⁵³ Alam and Campbell have investigated how motivations change over time.⁵⁴ A common thread across other projects is an interest in the subject,⁵⁵ with participants self-fashioning roles within a project to suit their interest.⁵⁶

⁵¹ E. Gil Clary and others, 'Understanding and Assessing the Motivations of Volunteers: A Functional Approach', *Journal of Personality and Social Psychology*, 74.6 (1998), 1516–30.

⁵² Deborah Edwards and Margaret Graham, 'Museum Volunteers and Heritage Sectors', *Australian Journal on Volunteering*, 11.1 (2006), 19–27.

⁵³ Raddick and others.

⁵⁴ Sultana Lubna Alam and John Campbell, 'Temporal Motivations of Volunteers to Participate in Cultural Crowdsourcing Work', *Information Systems Research*, 2017 <<https://doi.org/10.1287/isre.2017.0719>>.

⁵⁵ Sharon M. Leon, 'Build, Analyse and Generalise: Community Transcription of the Papers of the War Department and the Development of Scripto', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014) <<http://www.ashgate.com/isbn/9781472410221>>; Kathryn Eccles and Andrew Greg, 'Your Paintings Tagger: Crowdsourcing Descriptive Metadata for a National Virtual Collection', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Farnham, Surrey, UK: Ashgate, 2014) <<http://www.ashgate.com/isbn/9781472410221>>.

⁵⁶ V Das Gupta, N Rooney, and S Schreibman, 'Notes from the Transcription Desk: Modes of Engagement between the Community and the Resource of the Letters of 1916', in *Digital Humanities 2016: Conference Abstracts* (presented at the Digital Humanities 2016, Kraków: Jagiellonian University & Pedagogical University) <<http://dh2016.adho.org/abstracts/228>> [accessed 22 October 2018].

When thinking about motivations in practical terms, I find grouping motivations relevant to heritage crowdsourcing into extrinsic, intrinsic and altruistic motivations is useful. Very few cultural heritage crowdsourcing projects support extrinsic motivations, such as tangible rewards.⁵⁷ Intrinsic motivations including fun, an interest in the subject and socialising are inherently rewarding and come into effect when an activity is worth doing for its own sake, regardless of external rewards.⁵⁸ Altruistic motivations include those related to the 'collective' or greater good, 'the importance attributed to the project's goals',⁵⁹ and ideological values or principles. Jane McGonigal summarises much of the literature in her memorable overview of 'what humans crave' and 'what museums give us': '1. satisfying work to do 2. the experience of being good at something 3. time spent with people we like 4. the chance to be a part of something bigger'.⁶⁰

Turning crowdsourcing ideas into reality

My analysis of non-commercial crowdsourcing projects⁶¹ found that successful crowdsourcing projects have several features in common, including good publicity (whether through luck or design), well-designed task interfaces and processes, and presenting the project as a contribution to a shared, significant goal that links to participant motivations. Key challenges include recruiting and maintaining volunteer participation over time and integrating the results of crowdsourced tasks back into core catalogues, repositories or IT systems within the institution.

This section discusses important points in the process of planning, implementing and running crowdsourcing projects. Defining 'success' for your project will influence design decisions, as will the choice of source material and your desired outcomes. The

⁵⁷ For exceptions to this, see Luis von Ahn and Laura Dabbish, 'Designing Games with a Purpose', *Communications of the ACM*, 51.8 (2008), 57 <<https://doi.org/10.1145/1378704.1378719>>; 'WieWasWie Project Informatie', *VeleHanden* <http://velehanden.nl/projecten/bekijk/details/project/wiewaswie_bvr> [accessed 1 August 2014]; National Archives of Australia, 'ArchIVE – Homepage', *ArchIVE* <<http://transcribe.naa.gov.au/>> [accessed 18 June 2014].

⁵⁸ Mihaly Csikszentmihalyi and Kim Hermanson, 'Intrinsic Motivation in Museums: Why Does One Want to Learn?', in *Public Institutions for Personal Learning: Establishing a Research Agenda*, ed. by John Falk and Lynn D. Dierking (Washington D.C.: American Association of Museums, 1995), pp. 66–77.

⁵⁹ Oded Nov, Ofer Arazy, and David Anderson, 'Technology-Mediated Citizen Science Participation: A Motivational Model', in *Proceedings of the AAAI International Conference on Weblogs and Social Media* (Barcelona, Spain, 2011).

⁶⁰ Jane McGonigal, 'Gaming the Future of Museums' <<http://www.slideshare.net/avantgame/gaming-the-future-of-museums-a-lecture-by-jane-mcgonigal-presentation#text-version>> [accessed 14 February 2010].

⁶¹ Ridge, 'Making Digital History: The Impact of Digitality on Public Participation and Scholarly Practices in Historical Research'.

exact order of decisions will vary according to the specific project, but you should expect that some decisions will be revisited as more information is gathered and allow for this when allocating resources. Designing iteratively also allows you to fine-tune the prioritisation of efficiency and engagement, adjust workflow and quality controls measures as necessary, improve usability, and update text and tasks for specialist or generalist audiences as you learn from showing your project to potential participants.

Just as interfaces need to be carefully designed to maximise productivity and engagement, projects need to be carefully designed to ensure long-term success. Project design considerations include how the organisation sets up and resources a project, its coordination with other staff and work, how it evaluates and responds to results.

Decisions made in the planning phase will affect the implementation and running phases, so some points to consider for these later stages are discussed under the heading of planning.

Planning crowdsourcing projects

Key stages in the planning process include defining success for your project, managing any impact on the organisation, choosing source material and determining desired outputs, workflows and data re-use, communications and participant recruitment, and applying practical and ethical 'reality checks'.

Understanding the impact of logistical issues such as workflow, quality control and target systems for information collected through crowdsourcing by cultural heritage organisations should also help digital humanities researchers and practitioners interested in collaborating with GLAMs.

Defining 'success' for your project

Potential quantitative metrics for measuring the success of heritage crowdsourcing projects include: the number of hours participants have spent on a project; initial and sustained participation rates; participant retention; the extent and types of use of community discussion platforms; the number of tasks completed; and the percentage of tasks validated against required quality standards. Efficiency can be measured as the number of tasks accurately completed per volunteer minute. Valuable but less easily measured outcomes include the extent to which participants gain related skills and knowledge, or the number of new research questions or discoveries that emerge during a project. Qualitative measures include the extent to which participants expressed support or appreciation for the project, the number of participants who pursue activities related to their new interest, or some wider impact on participants' behaviour or attitudes.

Three definitions of success seem to have the most utility for project stakeholders: productivity, reach and engagement. However, two of these metrics are inherently

opposed: time spent posting on discussion platforms or learning about collection items means less time is available to spend on the core task.⁶² However, there is also an argument that both engagement and contributions are needed for citizen science projects to count as a success.⁶³ Accordingly, measurements of success should be judged and weighted according to the overall goals of an individual project.

Productivity is the simplest to define and to measure, and the easiest metric to design for. How many tasks have been completed to the standards required? Figures for prominent projects can be impressive, with Trove and Zooniverse contributions numbering in the hundreds of millions.⁶⁴

Reach measures the number or type of people contributing to projects. This might be the 1.7 million (at the time of writing) volunteers contributing to Zooniverse or a small group of volunteers drawn to a highly specialist project. Reach can extend beyond individual participants to include those who access research that results from projects, or who are more easily able to find cultural heritage collections online.

Finally, you can consider how many participants become more engaged with the subject of the collections or disciplines (such as history or science) related to them. Engagement might appear as learning, attitude change, or other changes in behaviour linked to feelings or knowledge gained.⁶⁵ Once you have decided the most appropriate

⁶² In a telling example, the first post on an *Old Weather* thread called 'Signs of OW addiction' said one of the 'Top Ten' signs of addiction might be that 'You spend more time on the forum than you do transcribing'. Forum posters, 'Signs of OW Addiction ...', *Old Weather Forum* » *Shore Leave* » *Dockside Cafe*, 2010
<<http://forum.oldweather.org/index.php?topic=1432.0>> [accessed 11 April 2014].

⁶³ Brooke Simmons, 'Measuring Success in Citizen Science Projects, Part 2: Results', *Zooniverse*, 2015
<<https://blog.zooniverse.org/2015/08/24/measuring-success-in-citizen-science-project-s-part-2-results/>> [accessed 28 August 2015].

⁶⁴ Current figures are available from <https://www.zooniverse.org/> and <http://trove.nla.gov.au/system/stats?env=prod#corrNewspapers>.

⁶⁵ Stephen Bitgood, *An Attention-Value Model of Museum Visitors* (Center for the Advancement of Informal Science Education, 2010), pp. 1–29
<http://caise.insci.org/uploads/docs/VSA_Bitgood.pdf>; The Culture and Sport Evidence (CASE) programme, *Evidence of What Works: Evaluated Projects to Drive up Engagement* (London: Culture and Sport Evidence (CASE) programme, January 2011), p. 19 <http://www.culture.gov.uk/images/research/evidence_of_what_works.pdf>; Museums, Libraries and Archives Council, 'Generic Learning Outcomes', *Inspiring Learning*, 2008
<<http://www.inspiringlearningforall.gov.uk/toolstemplates/genericlearning/>> [accessed 8 September 2014].

mix of success metrics, you can determine how you'll measure and evaluate progress against them.⁶⁶

Managing organisational impact

The Zooniverse guide to 'building a great project' begins '[k]now that you are making a commitment!'.⁶⁷ Crowdsourcing projects require ongoing attention from staff, and assessing whether you can provide resources for the life of the project is an important step in assessing the feasibility of a project. Staff can be supported by volunteers for some tasks, such as answering questions from other participants, but they must also have time to report on progress to internal and external stakeholders, and prepare newsletters and social media updates for outreach and marketing purposes.⁶⁸ Staff might also need support in gaining new skills such as community management or workflow integration.

Crowdsourcing projects can have an impact on the workload and outputs of departments across the organisation. For example, they can lead to increased attention to collections, and requests for new or reprioritised digitisation to keep items flowing into the platform. If your project is to generate metadata, annotations or other information about collection items, talking to the teams that manage the information systems that already store information about collections is vital. They can specify import formats and help you determine what information will be most useful to collect to improve catalogue or discovery systems. If information collected does not fit into existing interfaces (for example, your collections management system has no capacity to store user-generated content), where will it be kept? Ensure technical documentation is shared with relevant staff even if the platform is developed externally. The work of preparing material for ingest into the platform, and of reviewing and packaging task results for ingest into internal systems should also be included in overall resource plans.

Finally, an internal communications strategy, however informal, will help the rest of the organisation feel involved in the success of the project. You can share updates via internal presentations and emails, and invite staff to test interfaces, brainstorm ideas

⁶⁶ The Europeana Impact Playbook provides some useful headings for planning to report impact on various factors (but the overall process can be opaque). Harry Verwayen and others, *Impact Playbook for Museums, Libraries and Archives* (Europeana Foundation, 2017).

⁶⁷ 'Part I: Building a Great Project', *Zooniverse Help* <<https://help.zooniverse.org/best-practices/1-great-project/>> [accessed 9 December 2018].

⁶⁸ The time required will vary according to your specific goals, material, volunteers, etc., but as a rough guide, I may spend up to an hour and a half on *In the Spotlight* each week.

for outreach methods to reach potential participants, and plan publicity in physical venues.

Choosing source collections

Crowdsourcing relies on the availability of digitised collections. Digitisation can be expensive and time-consuming, so get estimates for delivery dates before planning other milestones of new digitisation. Source collections may also be determined by the goals of the project - research projects on a particular topic may choose items in a range of formats, while a project aimed at increasing discoverability might work through one format at a time.

Some content has a wider immediate appeal, and consequently makes the work of recruiting participants easier. Lascarides and Vershbow said it's best to 'choose your parents wisely' when describing the choice of material for what become *What's on the Menu* - 'it is much easier to get patrons excited about participation in a project if they are already excited about the source material'.⁶⁹ *DIY History* selects handwritten, historically significant, 'interesting' and extensive materials.⁷⁰ They also note a preference for material is that is 'old enough' to avoid copyright and privacy issues. While collections that appeal to both casual viewers and scholars make attracting interest and justifying participation much easier, it is possible to create compelling stories about more obscure collections or to invite specialist communities to become involved.

Planning workflows and data re-use

The source material - text, digitised images, audio-video, etc. - and goals of the project determine the types of tasks that will be crowdsourced. Planning the workflows necessary to make data usable is part of the process of assessing the feasibility of a project - there is no point asking people to help create data or knowledge that you cannot use as intended. Creating a workflow plan is part of managing the organisational impact and integration of a project, and should ensure that you can move digitised source material into your crowdsourcing platform, then move validated data to the system (which might be a collections management system, web publishing system etc.) in which it can be used. Collections management teams can advise on the most useful data for discoverability or help work out how to publish research datasets. Any changes required are likely to take time to implement so begin conversations with relevant departments as early as possible.

Source items might need pre-processing before they are presented in tasks. For example, some projects categorise manuscript items by how easy or hard they are to

⁶⁹ Lascarides and Vershbow.

⁷⁰ Michelle DiMeo, 'First Monday Library Chat: University of Iowa's DIY History', *The Recipes Project*, 2014 <<http://recipes.hypotheses.org/3216>> [accessed 6 September 2014].

transcribe. Some pre-processing tasks can be built into the task ecosystem, such as Fossil Finder, which asked participants whether a photo was ‘good enough to study’, instructing them, ‘If it is too blurry, dark, noisy, or bushy then bin it!’.⁷¹ Workflow also includes task validation and quality control processes, although these are usually built into the crowdsourcing platform. Data might also need post-processing to convert it into formats⁷² suitable for ingest and sharing with project contributors.

Workflows should be tested as early as possible to allow time to manage any logistical, technical, legal or institutional issues that arise. This work is not visible to contributors but it ensures that new items can be easily added to the platform, and that data created is put to work as soon as possible, helping demonstrate the value of volunteer contributions.

Workflow systems should be designed for modularity to allow for changes in other platforms over time. Collections management systems are refreshed, new tasks may require different export formats, or you may start to integrate machine learning processes to create human computation systems. Finally, in order to re-use content created by volunteers, you should ensure that you have put in place terms and conditions that give you the right to use the data.

Planning communications and participant recruitment

There is someone, somewhere, interested in every single thing collected by a cultural heritage institution. The hard part is finding them and reaching them with a compelling invitation to join your shared endeavour.

Without participants, there is no project, so invest time in planning your communication strategy. Publicity material, including posts on social media, text on project sites, must clearly explain the project's goals and tasks without jargon or assumptions about what recipients already know. For example, in testing *In the Spotlight* I found that not everyone will be familiar with terms like ‘transcription’ or understand why it cannot be done automatically. As you develop prototype interfaces, test and revise messages until they effectively motivate target participants to complete their first task.

Marketing and outreach may not come naturally. It can help to find out (or remind yourself) what people already love about the relevant collections - what stories do they share with front-of-house or social media staff about them, or why do they already value them? Similarly, you can involve existing communities of interest in the process of designing the project (bearing in mind that they will not be able to represent the

⁷¹ <http://www.fossilfinder.org/>

⁷² Each platform produces differently formatted outputs - text transcribed via FromThePage will look different to that transcribed via the Zooniverse Project Builder. Technical resources to convert JSON and XML-formats might be required.

needs of novices with those collections). You might find them on listservs or discussion forums, social media hashtags or via academic or community contacts.

Clary et al found that messages that resonate with recipient motivations have enhanced 'persuasive impact' and help volunteers find more enjoyable and satisfying roles that match their motivations.⁷³ Furthermore, volunteers whose experiences matched their motivations were more satisfied and more likely to intend to continue volunteering,⁷⁴ suggesting that the text used to market and describe projects could be as important as interface and task design. Favourite examples of 'straplines' that encapsulate the goals and attitudes of projects include: 'With a few keystrokes, you could bring a family together';⁷⁵ 'We know the names of these children; can you help us tell their stories?';⁷⁶ 'Kill Time. Make History.'; ⁷⁷ 'Historians need your help!'.⁷⁸

Final considerations: practical and ethical 'reality checks'

Having defined success for your project, talked to teams across your organisation, chosen your goals and source material, and considered workflow, a final 'feasibility check' can be useful. Will anyone have the necessary skills and knowledge to undertake the task you propose, and can you motivate them to take part?

You can talk to potential contributors and undertake usability testing on early paper or digital prototypes of your interface to check whether the tasks proposed make sense and whether they would be motivated to do them. These 'reality check' conversations will also help determine whether you have a compelling 'marketing' story about the difference the project makes that would convince people to donate time to your project, allowing you to refine or abandon an idea.

The final 'reality check' for a planned project is reviewing your plans to ensure that it meets the ethical standards required. For example, organisational policies about volunteering may apply, or there may be local norms about responsibility and fairness. Once you have determined the ethical principles that apply, ensure they are enacted in practices such as task and interface design, communication strategies and data access

⁷³ Clary and others.

See also Paul Fugelstad and others, 'What Makes Users Rate (Share, Tag, Edit...)? Predicting Patterns of Participation in Online Communities', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, 2012), pp. 969–978 <<http://dl.acm.org/citation.cfm?id=2145349>> [accessed 20 December 2013].

⁷⁴ Clary and others.

⁷⁵ Ancestry's World Archives Project

<https://web.archive.org/web/20150905125517/http://landing.ancestry.com/wap/learnmore.aspx>

⁷⁶ *Children of the Lodz Ghetto Research Project*

<https://web.archive.org/web/20180614032124/https://www.ushmm.org/online/lodzchildren/>

⁷⁷ *Building Inspector* <http://buildinginspector.nypl.org/>

⁷⁸ *DIY History* <https://diyhistory.lib.uiowa.edu/>

plans. For example, you may value transparency about process and outcomes, but as this can be challenging for cultural heritage and academic projects, you could initiate internal conversations about publishing more information than usual to meet those goals.

Project teams generally believe that they must honour participants' time and contributions, and honour any commitments they make to them.⁷⁹ (Having seen stakeholders on some early projects disappear after launch, I tend to use a 'party' metaphor - if you have invited people into your space, as a host you are bound to provide for their basic needs, which includes being in the same space as them.) The European Citizen Science Association's Ten Principles of Citizen Science lists some practical ways in which ethical considerations may be operationalised.⁸⁰ Discussion of 'what ethical practice looks like on a daily basis' at an expert workshop on crowdsourcing organised at the Digital Humanities 2016 conference⁸¹ included: timely and responsive communication, defining benefits to each party, communicating expectations and keeping promises to volunteers, participants' right to access and re-use data, data protection (for records related to recent generations), acknowledging and crediting participants; considering participant experience alongside goals and efficiency; and updating ethical practices as necessary. Models for crediting participants could be drawn from traditional volunteer programmes, and might include letters of reference or certificates of participation in addition to credit on academic outputs.

Developing and testing crowdsourcing projects

In this section I will discuss key points and general principles for implementing crowdsourcing projects, including task design, documentation and tutorials, quality control and ensuring that designs work as well as possible through usability testing. It is important to note that basic usability (minimising dissatisfaction) is rarely enough; websites should both offer pleasing features that encourage users to return and

⁷⁹ This is not always as easy as it sounds, as enthusiasm can get ahead of resources. Mia Ridge, 'Citizen History and Its Discontents' (presented at the IHR Digital History Seminar, Institute for Historical Research, London, 2014)

<<https://hcommons.org/deposits/item/hc:17907/>> [accessed 9 December 2018].

⁸⁰ European Citizen Science Association, *10 Principles of Citizen Science* (London, September 2015)

<https://ecsa.citizen-science.net/sites/default/files/ecsa_ten_principles_of_citizen_science.pdf> [accessed 9 December 2018].

⁸¹ DH2016 Expert Workshop, 'DH2016 Crowdsourcing Workshop Session Overview', *DH2016 Expert Workshop: Beyond The Basics: What Next For Crowdsourcing?*, 2016 <https://docs.google.com/document/d/1sTII8P67mOFKWxCaAKd8SeF56PzKcklxG7KDfCRUF-8/edit?usp=drive_open&oid=o&usp=embed_facebook> [accessed 5 October 2018].

minimise annoyances for users. The details of effective task design will depend on your goals and source materials.

Critical points for projects when the quality of 'user experience' (also known as UX, the visible aspects of backend workflow, instructional and marketing text, etc. in addition to interface and interaction design) matters are successfully 'onboarding' a participant so that they can complete their first task, and maintaining participation despite changes over time. As crowdsourcing is a voluntary activity, it is vital to minimise barriers to participation, points of friction and demotivators.

Barriers to participation include compulsory registration,⁸² but some Zooniverse projects have successfully deployed a design pattern called 'lazy registration'.⁸³ Being clear about how data will be used helps. Rose Holley's 2010 summary of research on participation in *Distributed Proofreaders, FamilySearch Indexing, Wikimedia* and *Trove* reported that volunteers 'do not want to feel that their work can be commercially exploited'.⁸⁴ A study of *Old Weather* found that stopping participating is strongly associated with an anxiety about the quality of contribution.⁸⁵ Competitive models like gamification-style leaderboards are an easy way to recognise individuals who have completed more tasks, but they favour those with more free time, and there is some evidence that some participants are deterred by competition.⁸⁶

Usability tests can be conducted throughout the development process, as you can test existing projects, paper prototypes and work-in-progress. Tests can be informal ('guerrilla' usability tests are free apart from the time required to talk to participants) or formal, but the benefits are invaluable. Usability tests allow you to understand and devise creative solutions to problems uncovered. They will help you identify and remove barriers to participation, define rewards appropriate to your goals and community and ensure that the project maximises the return on investment.

⁸² Raluca Budi, 'Login Walls Stop Users in Their Tracks', *Nielsen Norman Group*, 2014 <<http://www.nngroup.com/articles/login-walls/>> [accessed 7 March 2014].

⁸³ 'Lazy Registration Design Pattern', *Http://Ui-Patterns.Com/Patterns/LazyRegistration* <<http://ui-patterns.com/patterns/LazyRegistration>> [accessed 9 December 2018].

⁸⁴ Holley, 'Crowdsourcing: How and Why Should Libraries Do It?'

⁸⁵ Alexandra Eveleigh and others, 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science' (ACM Press, 2014), pp. 2985–94 <<https://doi.org/10.1145/2556288.2557262>>.

⁸⁶ Chris Preist, Elaine Massung, and David Coyle, 'Competing or Aiming to Be Average?: Normification as a Means of Engaging Digital Volunteers', in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing* (presented at the CSCW 2014 - Computer Supported Cooperative Work and Social Computing, Baltimore, MD, USA.: ACM Press, 2014), pp. 1222–33 <<https://doi.org/10.1145/2531602.2531615>>.

Designing the 'onboarding' experience

In user experience design, 'onboarding' refers to orienting people to the features of a site and helping them start to use it.⁸⁷ Ideally, the first page that potential participants see shows (rather than tells) them what the project aims to do, how their help can make a difference, and where to start the task. As discussed earlier, a good communication strategy should include a strong strapline that give a sense of the larger challenge that tasks will contribute to, and ideally connect to probable motivations for action. For example, *What's on the Menu* has manicules (pointing hands) pointing to a button labelled 'Help transcribe'.⁸⁸

The landing page should also include 'social proof' that others have already chosen to participate.⁸⁹ For example, the front page of *What's on the Menu* prominently lists the number of dishes transcribed so far and *Trove* lists the number of corrections already made on a given day, the number of items tagged that week, and the number of comments added that month, showing how updates can be tailored to the frequency of different tasks (a method that supports low-frequency sites).

Some projects feed participants tasks from a queue of material, while others leave the choice of material up to the participant. Providing initial tasks from a queue minimises the number of decisions a participant has to make, which helps reduce cognitive load (the amount of mental effort required to operate a system or learn new information).⁹⁰ This, in turn, leaves more mental resources for learning the task.⁹¹ Feeding the first tasks to participants also allows a project to begin with 'golden tasks', tasks to which the answer is known, so they can assess the participant's performance.⁹²

⁸⁷ Whitney Hess, 'Onboarding: Designing Welcoming First Experiences', *UX Magazine*, 2010 <<http://uxmag.com/articles/onboarding-designing-welcoming-first-experiences>> [accessed 28 July 2014].

⁸⁸ <http://menus.nypl.org/>

⁸⁹ Tanushree Mitra and Eric Gilbert, 'The Language That Gets People to Give: Phrases That Predict Success on Kickstarter' (CSCW, 2014) <<http://comp.social.gatech.edu/papers/cscw14.crowdfunding.mitra.pdf>> [accessed 28 January 2014]. See also Preist, Massung, and Coyle. on the 'normalising' effect of displaying participant activity.

⁹⁰ Kathryn Whitenton, 'Minimize Cognitive Load to Maximize Usability', *Nielsen Norman Group*, 2013 <<http://www.nngroup.com/articles/minimize-cognitive-load/>> [accessed 12 September 2014].

⁹¹ See also Fred Paas, Alexander Renkl, and John Sweller, 'Cognitive Load Theory and Instructional Design: Recent Developments', *Educational Psychologist*, 38.1 (2003), 1–4 <https://doi.org/10.1207/S15326985EP3801_1>. and Van Merriënboer, Kirschner, and Kester, "Taking the Load off a Learner's Mind."

⁹² Tommaso De Benetti, 'The Secrets of Digitalkoot: Lessons Learned Crowdsourcing Data Entry to 50,000 People (for Free)', *Microtask*, 2011 <<http://blog.microtask.com/2011/06/the-secrets-of-digitalkoot-lessons-learned-crowdsourcing-data-entry-to-50000-people-for-free/>> [accessed 9 January 2012].

The *Smithsonian Transcription Center* provides many ways for a participant to find content that they might be interested in, including themes (such as 'Civil War Era' or 'Field Book Project'), source organisations (specific museums or archives), featured projects and those with recent activity. The *Notes from Nature* collection pages list the average time per record (ranging from 3 minutes to 15 minutes) as well as the average 'difficulty' (ranging from 'easy' to 'very hard').

Task design

Nielsen's usability heuristics contain many principles relevant to crowdsourcing projects, including: keeping users informed of the system status through appropriate feedback; speaking the users' language; preventing errors; supporting recovery from error when errors do occur; following platform conventions; minimising memory load by making actions and options visible; and (where necessary) providing concrete instructions that focus on the users' task.⁹³

In design principles specific to crowdsourcing, task 'size' can be measured in terms of the amount of source material to process, the time per task, modularity (whether tasks are independent and asynchronous) and cognitive load (roughly, the amount of mental effort required).⁹⁴ Research has found that microtasks lead to fewer mistakes and an 'easier' experience.⁹⁵ They provide opportunities to learn the skills required for more complex tasks but are easier for novices to complete. If you have to design macro- or more specialist tasks, ensure that motivational text and recruitment are strong enough to match the size or complexity. Finding the sweet spot between tasks likely to attract participants, that provide useful data and are possible within the resources available can require some creativity.

Most crowdsourcing projects report that up to 80-90% of the work is done by 10% of participants and many other participants contribute a small amount each.⁹⁶ Given the

⁹³ I expand on the application of these heuristics in Ridge, 'Making Digital History: The Impact of Digitality on Public Participation and Scholarly Practices in Historical Research'. Jakob Nielsen, '10 Usability Heuristics for User Interface Design', 1995 <<http://www.nngroup.com/articles/ten-usability-heuristics/>> [accessed 29 April 2014].

⁹⁴ Motivation seems to reduce the impact of task size, in that some large, complex tasks (such as those in the *Dickens Journals Online* text correction project or *Children of the Lodz Ghetto*) can still attract participants if the motivation and/or challenge is strong enough. The combination of task size and motivation could be called the task 'weight', but further research is needed to test this model.

⁹⁵ The research compared macrotask and microtask versions of the same overall task. It also found that microtasks took more time combined than the equivalent macrotask. Justin Cheng and others, 'Break It Down: A Comparison of Macro- and Microtasks' (presented at the CHI2015, Seoul, Republic of Korea: ACM Press, 2015), pp. 4061-64 <<https://doi.org/10.1145/2702123.2702146>>.

⁹⁶ For a visual representation of this see: Philip Brohan, 'One Million, Six Hundred Thousand New Observations', *Old Weather Blog*, 2012

role 'super-contributors' play in a projects' productivity, it could be tempting to optimise designs for their need but projects must cater for both casual and super contributors.

Documentation and tutorials

Ideally, interactive tutorials could show new participants how to complete the task successfully while letting them try it, rather than read about it, but the user experience design and technical resources required to do so are rarely available ⁹⁷ and many tutorials appear as modal windows overlaid over the task window. ⁹⁸ However, many users automatically close tutorials without reading or watching them, so it is important to have a visible link for a Help page that includes the tutorial and/or more detailed documentation.

Help text, whether on the task interface or a separate page, should help reassure potential participants by anticipating and answering their questions. It should be clear and unambiguous, and available at the point at which it is needed, ⁹⁹ address 'boundary cases', and ideally provide examples of what is expected. ¹⁰⁰ Balancing the need for simplicity with the need for flexibility is a challenge for projects working with materials that may contain unexpected or inconsistent information. Producing good tutorials and documentation can take several iterations. Including tutorials and help text in usability testing will highlight issues, and test participants may provide more user-friendly alternatives for language used.

Quality control: validation and verification systems

Even the most highly-skilled and well-intentioned volunteer makes occasional mistakes, and crowdsourcing projects carefully check the information they receive.

<<http://blog.oldweather.org/2012/07/23/one-million-six-hundred-thousand-new-observations/>> [accessed 30 October 2012].

⁹⁷ However, in future human computation systems may be able to support participants with feedback tailored to their performance of a task.

⁹⁸ For more on tutorial design see: Aurora Bedford, 'Instructional Overlays and Coach Marks for Mobile Apps', *Nielsen Norman Group*, 2014

<<http://www.nngroup.com/articles/mobile-instructional-overlay/>> [accessed 12 September 2014]; Paas, Renkl, and Sweller. Research on the techniques game designers use for including skills tests and tutorials may be relevant for projects that wish to teach specific skills or knowledge to participants undertaking tasks. See: Mia Ridge, 'Playing with Difficult Objects: Game Designs for Crowdsourcing Museum Metadata' (unpublished MSc Dissertation, City University London, 2011)

<<http://www.miaridge.com/my-msc-dissertation-crowdsourcing-games-for-museums/>>; Richard E. Mayer and Roxana Moreno, 'Nine Ways to Reduce Cognitive Load in Multimedia Learning', *Educational Psychologist*, 38.1 (2003), 43-52.

⁹⁹ Nielsen.

¹⁰⁰ Aniket Kittur and others, 'The Future of Crowd Work', in *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*, 2013, pp. 1301-1318 <<http://dl.acm.org/citation.cfm?id=2441923>> [accessed 7 May 2013].

Most methods involve comparing three or more task results for the same source against each other, with a simple 'majority rules' decision to accept the most common answer. The most appropriate method for reaching consensus will depend on the material, even for 'type what you see tasks', where small differences in punctuation may make transcriptions fail tests drawn from the sciences. Ben Brumfield has provided a useful overview of quality control methods for transcription.¹⁰¹ Verifying tags is difficult to do automatically without excluding potentially valuable unique tags from contributors with specialist knowledge,¹⁰² but verification tasks can help.¹⁰³

Rewards and recognition

Public recognition of volunteer contributions is important, and can be built into many points of the project interface and communications. Some projects name contributors in project updates¹⁰⁴ or list them as co-authors on journal articles.¹⁰⁵ Describing, or even better, showing the impact of contributions towards a project's goals can powerfully link to participant motivations.¹⁰⁶

Metrics for recognition should be chosen carefully. Ben Brumfield has a story that illustrates the dangers of external motivators like leaderboards, where contributors may focus on aspects that are quantified on a leaderboard at the expense of more important but unquantified tasks.¹⁰⁷

¹⁰¹ Ben W. Brumfield, 'Quality Control for Crowdsourced Transcription', *Collaborative Manuscript Transcription*, 2012
<<http://manuscripttranscription.blogspot.co.uk/2012/03/quality-control-for-crowdsourced.html>> [accessed 9 October 2013].

¹⁰² This was difficult back in 2011 (as discussed in Ridge, 'Game Designs to Improve Museum Collections'.) but advances in human computation should make it easier. See also: Richard Grayson, 'A Life in the Trenches? The Use of Operation War Diary and Crowdsourcing Methods to Provide an Understanding of the British Army's Day-to-Day Life on the Western Front', *British Journal for Military History*, 2.2 (2016) <<http://bjmh.org.uk/index.php/bjmh/article/view/96>> [accessed 25 February 2016].

¹⁰³ There is a significant body of literature on this subject. A useful place to start is von Ahn and Dabbish.

¹⁰⁴ For example, Philip Brohan, 'In Search of Lost Weather', *Old Weather Blog*, 2014 <<http://blog.oldweather.org/2014/08/18/in-search-of-lost-weather/>> [accessed 4 September 2014].

¹⁰⁵ For example, the authors whose affiliation is listed as 'Planet Hunter' in 'Planet Hunters. VI. An Independent Characterization of KOI-351 and Several Long' <<http://adsabs.harvard.edu/abs/2013arXiv1310.5912S>> [accessed 4 September 2014].

¹⁰⁶ Dana Rotman and others, 'Dynamic Changes in Motivation in Collaborative Citizen-Science Projects', in *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, 2012), pp. 217–226 <<https://doi.org/10.1145/2145204.2145238>>.

¹⁰⁷ Ben W. Brumfield, 'Crowdsourcing at IMLS WebWise 2012', *Collaborative Manuscript Transcription*, 2012
<<http://manuscripttranscription.blogspot.com.au/2012/03/crowdsourcing-at-impls-web-wise-2012.html>> [accessed 8 September 2014].

Running crowdsourcing projects

The key challenge in running a project is motivating continued participation. In this section I discuss expectations around launching projects, the effect of media stories, consider the role of participant discussion, ongoing communications and maintaining participation, and planning for a 'graceful exit'.

Participatory projects can be challenging for organisations used to 'launch and forget it' exhibitions and publications. Ideally, iterative design processes can continue after launch. Participants tend to have creative ideas for new tasks,¹⁰⁸ suggest sensible tweaks to existing tasks and text, and report bugs. Over the longer term, an interface that looks amazing in 2019 may look dated in 2022, or you may want to take advantage of emerging technologies. It is important to allow resources for post-launch.

Launching a project

You've planned, designed and tested your project. You've prepared a press release and social media posts. Launch day will (hopefully) be busy. Allow time for answering participant queries and media enquiries, and be prepared to load new material if the first batches are completed.

When your project launches, some of the first visitors will be participants from other projects, and colleagues from academia and cultural heritage institutions. The first group often have a highly sophisticated understanding of crowdsourcing, and will be looking for markers of quality including the importance of the task, the availability of data and how participants are credited or rewarded. The second group will be curious about your project design. Their positive reports may help build your word of mouth marketing.

Pieces in 'traditional' media can be very effective in attracting visitors, some of whom may become participants. It is difficult to disentangle the role of luck in getting media and popular attention but a quirky story or topic, relationships with an existing community, being the first of its type, or an opportunity to access highly-valued content or expertise seem to help. More targeted publicity may reach a smaller number of people, but those reached may be proportionally more likely to participate. *History Harvest* and *Letter in the Attic*¹⁰⁹ found that 'face-to-face contact' at local events and groups is more effective than media attention at gaining contributions.¹¹⁰

¹⁰⁸ E.g. when *In the Spotlight* launched, participants requested the ability to tag playbills with specific topics.

¹⁰⁹ The project collected letters, diaries and items related to Brighton and Hove.

¹¹⁰ Jack Latimer, 'Letter in the Attic: Lessons Learnt from the Project', *My Brighton and Hove*, 2009

<<http://www.mybrightonandhove.org.uk/page/letterintheatticlessons?path=op116p1543>
p> [accessed 17 April 2014]. Face-to-face events might also help reach those not online. One First World War project heard from a potential contributor 'aged 89 and nearly

The role of participant discussion

Some crowdsourcing projects provide ways for participants to communicate with each other via a discussion forum,¹¹¹ on social media, or through comments on specific items.¹¹² Some participants may prefer to comment directly to project staff rather than post in public. Posts might discuss difficulties, help answer queries, collect lessons learnt over time, share stories about interesting finds or potential discoveries, and provide feedback or suggestions for improvement to project stakeholders.

Conversations on forums can have important learning outcomes¹¹³ and provide social opportunities that motivate ongoing participation.¹¹⁴ Participants' expectations about the presence of project staff on discussion forums vary, and projects should be careful about how these forums are described so that their expectations are not disappointed.

Ongoing community engagement

Once a project is up and running, marketing efforts generally need to shift from participant recruitment to participant retention. This is also an opportunity to shift from talking about the project to talking about the impact of the project. You can thank participants individually or collectively, share progress reports and participants' findings and questions, and provide information about how contributions have been used (an important factor in ongoing motivation).¹¹⁵ The *Smithsonian Transcription Center* has been carefully designed to provide multiple opportunities for celebrating success, with small-scale, niche projects within the larger project.¹¹⁶

One of the joys of *In the Spotlight* is the opportunity to amplify the expertise and curiosity of participants (for example, we published a blog post from one participant, Edward Mills, on the British Library's Digital Scholarship blog);¹¹⁷ my hope is that

blind' who had asked neighbour to email the project after hearing about it on the radio. Piers Dillon-Scott, 'How Europeana, Crowdsourcing & Wiki Principles Are Preserving European History', *The Sociable*, 2011
<<http://sociable.co/business/how-europeana-crowdsourcing-wiki-principles-are-preserving-european-history/>> [accessed 15 February 2015].

¹¹¹ Forums were particularly important for early Zooniverse projects, but they have moved to 'Talk' pages that are more closely integrated with task interfaces.

¹¹² E.g. transcribers in the *Smithsonian Transcription Center* can leave notes for other transcribers and reviewers on specific pages or post questions on social media.
<https://transcription.si.edu/>

¹¹³ Gabriel Mugar, Carsten Østerlund, Corey Brian Jackson, and others, 'Being Present in Online Communities: Learning in Citizen Science', in *Proceedings of the 7th International Conference on Communities and Technologies*, C&T '15 (New York, NY, USA: ACM, 2015), pp. 129–138 <<https://doi.org/10.1145/2768545.2768555>>.

¹¹⁴ Holmes.

¹¹⁵ Rotman and others.

¹¹⁶ Discussed further in Ridge, 'Making Digital History: The Impact of Digitality on Public Participation and Scholarly Practices in Historical Research'.

¹¹⁷ Edward Mills, 'The Flich of Bacon: An Unexpected Journey Through the Collections of the British Library', *British Library Digital Scholarship Blog*

posting updates from participants encourages more activity, which I can then share, in a 'virtuous circle'. Even the most ardent fans of a project may forget to revisit it unless it has become a daily habit (even then, it is liable to be interrupted by changes in routine). Regular updates remind participants to come back to a project.

Ongoing communications, whether simple quantitative progress updates, answering questions or liaising with experts to pass on information on the impact of the project, can require significant amounts of time. However, anecdotally, it seems that paying attention to activity on a project reaps rewards in ongoing participation.

Planning a graceful exit

Whether a project finishes because volunteers have completed all the available tasks, key team members move on or funding ends, planning a graceful finish is the best way to honour the work of the project team and volunteers alike. It is important to let volunteers know when the end of a project is in sight, giving them time to complete personal tasks, download data and finish conversations. You should document the final outcomes, deposit any resulting datasets in a repository, and (ideally) submit project URLs to regional or international web archives.

The future of crowdsourcing in cultural heritage

As the increasing success of machine learning-based projects such as *Transkribus*,¹¹⁸ which aims to teach computers to read handwriting, shows, many tasks currently crowdsourced can increasingly be performed by software. Computer vision technologies can increasingly identify even obscure or historical subjects in a picture.¹¹⁹

As computers get better at microtasks such as text transcription and image classification - tasks that many crowdsourcing participants find satisfying, and that may be important first steps in developing new interests - what impact will this have on crowdsourcing projects and participants? 'Human computation' systems that deploy the particular skills of people and machines in order to efficiently complete tasks can help meet the challenges of large-scale collections. However, if they are used in GLAMs, they should be carefully designed to allow for engagement and enjoyment of collections while not unnecessarily duplicating effort that could be better done by software.

<<http://blogs.bl.uk/digital-scholarship/2017/12/the-flitch-of-bacon-an-unexpected-journey-through-the-collections-of-the-british-library.html>> [accessed 17 August 2018].

¹¹⁸ <https://transkribus.eu/Transkribus/>

¹¹⁹ Rachel Collings, 'The Art of Computer Image Recognition', *The Public Catalogue Foundation*, 2014 <http://www.thepcf.org.uk/what_we_do/48/reference/862> [accessed 25 May 2014]. Kyle Willett, 'New Paper: Galaxy Zoo and Machine Learning', *Galaxy Zoo* <<http://blog.galaxyzoo.org/2015/03/31/new-paper-galaxy-zoo-and-machine-learning/>> [accessed 31 March 2015].

To finish on a positive note, these new technologies can also be harnessed to make crowdsourced microtasks even easier.¹²⁰ The success of OCR correction projects like *Trove* shows that providing some pre-processed data might actually make tasks easier, and therefore more enjoyable.¹²¹ Pre-processing items might allow tasks that can be performed on mobile and tablets devices to be created, and machine learning technologies could be used to provide personalised feedback on participant tasks, helping them feel more confident and learn skills more quickly.

Thanks and acknowledgements

I would like to thank participants and supporters of crowdsourcing projects I've created, including *Museum Metadata Games*, *In their own words: collecting experiences of the First World War*, and *In the Spotlight*. I would also like to thank my co-organisers and attendees at the Digital Humanities 2016 Expert Workshop on the future of crowdsourcing. Especial thanks to the participants in courses and workshops on 'crowdsourcing in cultural heritage', including the British Library's Digital Scholarship training programme, the HILT Digital Humanities summer school (once with Ben Brumfield) and scholars at other events where the course was held, whose insights, cynicism and questions have informed my thinking over the years. Finally, thanks to Meghan Ferriter and Victoria Van Hyning for their comments on this manuscript.

¹²⁰ See also research on 'social machines', in which people and computers are part of a larger integrated system. Paul R. Smart, Elena Simperl, and Nigel Shadbolt, 'A Taxonomic Framework for Social Machines', in *Social Collective Intelligence: Combining the Powers of Humans and Machines to Build a Smarter Society*, ed. by Daniele Miorandi and others (Berlin, Germany: Springer, 2014) <<http://eprints.soton.ac.uk/362359/>> [accessed 27 July 2014].

¹²¹ This does not mean that projects should only offer 'easy' microtasks, as the relationship between challenge and enjoyment is complex, but they may allow for a broader range of participants and thereby create more opportunities for deeper engagement.