FOUNDATION PROJECTS

SUPPORTING DOCUMENTATION

Citizen research landscape documentation

Engaging Crowds: citizen research and cultural heritage data at scale

Ashleigh Hawkins, Rebecca Hutcheon, Louise Seaward, Pip Willcox

The National Archives

Contents

Executive summary

This report reflects on existing citizen research activity in the UK heritage sector by summarising both published research and current citizen research activity at Independent Research Organisations (IROs) and other heritage organisations, addressing the following questions:

How do institutions and volunteers experience online citizen research; what are the motivations and rewards?
Were projects successful in what they set out to do?
What challenges have projects faced?
What recommendations can be made based on the experiences of these projects?

In December 2020, we circulated a call for information from cultural heritage practitioners in the UK running citizen research projects online. This report draws on the responses generously sent in answer to our call, focusing primarily on online projects based in the UK. A full list of the featured projects can be found in the Appendix. In July 2021, members of the *Engaging Crowds* project team coordinated a workshop at the Discovering Collections, Discovering Communities (DCDC) conference to gather insights from practitioners and researchers with experience of enabling and supporting online volunteering. The discussions from this workshop are also summarised in this report.

A broad range of experience of citizen research in the UK is included in this report, from small-scale short-term projects, such as those set up to move pre-established volunteering projects online in response to the onset of the Covid-19 pandemic, to large-scale long-term citizen research projects. What is evident from all responses received to the call for information is that online citizen research projects bring many benefits to both volunteers and organisations, and also present a number of challenges. Successes relate to volunteer recruitment and engagement, as well as opportunities for increased data production and data quality. Data quality, however, was also reported by a number of respondents as an area of challenge, together with issues relating to chosen citizen research platforms and open access. In synthesising the experiences reported by responding projects and workshop participants, this report has identified a number of recommendations for future citizen research projects.

The findings of this report are limited to insights gained from the projects that responded to our open call and from workshop participants. For a more comprehensive overview of citizen research, we recommend *The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage*, an open access book co-authored by practitioners and offering an authoritative guide to crowdsourcing and digitally-enabled participation projects in the cultural heritage sector.[1] A member of the *Engaging Crowds* project team, Sam Blickhan at Zooniverse, was an investigator on the project behind this book (funded by the National Endowment for the Humanities in the US and the Arts and Humanities Research Council in the UK), and another team member, Pip Willcox at TNA, was a co-author of the book.[2]

---

[1] Mia Ridge, Samantha Blickhan, Meghan Ferriter and others, *The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage* (online, 2021) <https://doi.org/10.21428/a5d7554f.1b80974b>.

[2] Collective Wisdom, 'Collective Wisdom: The state of the art in cultural heritage crowdsourcing', *Collective Wisdom,* 2022 <https://collectivewisdomproject.org.uk/> [accessed 15 February 2022].

Thank you to all the people, projects, groups and organisations who responded to our call for information about digitally enabled citizen research projects in cultural heritage. We are indebted to your generosity and grateful for the ethos of sharing expertise that permeates the international crowdsourcing community.

Introduction

Across galleries, libraries, archives and museums, there is an established tradition of volunteering going back at least two centuries. From the 1990s the concept of 'citizen science' began to gain ground, with an emphasis on citizens producing or co-producing valid research, and creating new knowledge as partners in that pursuit. During the 2000s, the spread of digitisation and digital technologies, coupled with the growing availability of increasingly high-speed internet access, created the possibility for volunteering to shift online. The term 'crowdsourcing' was coined in 2006 to describe the practice of companies outsourcing work traditionally undertaken by employees to the public.[3] This term came to be applied to an expanding array of online volunteering projects, where members of the public were asked to perform small tasks in pursuit of a broader research objective. One area where this took off was cultural heritage, where it has become a vital route to engage a wide public in tasks (such as transcription, classification, and description of digitised collections), to increase access to and types of engagement with heritage collections, and to bring new perspectives to bear on our shared collections.

Heritage crowdsourcing (or citizen research: we use the terms interchangeably) projects are driven by a range of purposes: some investigate a specific research question, for example, while others aim to increase access to a particular collection.[4] Through such projects, additional possibilities for research are created, contributor networks are established, and both contributors and cultural heritage organisations benefit. We acknowledge at the same time that crowdsourcing projects are informed by the historical legacies of empire, slavery, prejudice and other inequalities that form part of the histories of cultural heritage organisations and the collections they hold,[5] making it necessary for crowdsourcing practitioners to develop inclusive projects and work actively to engage a diverse 'crowd' of participants. Indeed, Owens emphasises the need for the cultural heritage community to consider the impact on their values and ethics of crowdsourcing.[6]

Ridge et al argue that a number of factors make cultural heritage citizen research projects different from other fields, exploring in their publication areas of difference including the purposes behind project creation, the projects' content and data outputs, and motivations for participation.[7] There is extensive evidence that volunteers engaged in citizen research are motivated by a wide range of factors. Crowston and Fagnot found that, as participants become more involved, they find they agree with the project's ideology, and come to develop both a sense of social obligation and of community spirit.[8] Previous surveys of contributors on Zooniverse, the world's largest citizen

---

[3] Jeff Howe, 'Crowdsourcing: A Definition', (2006) <https://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html> [accessed 24 February 2022].

[4] Mia Ridge and others, 'What is crowdsourcing in cultural heritage?', in *The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage – community review version*, ed. by Mia Ridge and others (online, 2021), Chapter 2 <https://doi.org/10.21428/a5d7554f.1b80974b>.

[5] Ridge and others, 'What is crowdsourcing?'.

[6] Trevor Owens, 'Making Crowdsourcing Compatible with the Missions and Values of Cultural Heritage Organisations', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Abingdon: Ashgate Publishing, 2014), pp. 269-280 (p. 269).

[7] Ridge and others, 'What is crowdsourcing?'.

[8] Kevin Crowston and Isabelle Fagnot, 'Stages of motivation for contributing user-generated content: A theory and empirical test', International Journal of Human-Computer Studies, 109 (2018), 89-101 <https://doi.org/10.1016/j.ijhcs.2017.08.005>, (p. 93).

research platform and project partner, have found that, firstly, people engage in transcription projects because they enjoy contributing to knowledge and research. The second most frequent motivation is an interest in the subject matter. After this, and much lower in stated motivations, come 'entertainment' and 'distraction'.[9] These findings are borne out in our project's survey results (see Annex 7), with an additional category of supporting the particular organisation scoring similarly to entertainment as a motive for engaging in a project. Elsewhere it has been reported that, for volunteers, benefits are intrinsic (i.e. a result of the task itself), such as 'subject interest and curiosity, competence in the transcription task, and an enjoyment derived from taking part in the project'.[10] Participants' identification with the goal of the project is a strong motivation too. Extrinsic factors, such as 'interaction with other volunteers, e.g. status gained for expertise or high quality work', were found to be less motivating by comparison.[11] However, as Ridge et al emphasise, it is important to acknowledge that 'participants are not a monolith with the same needs and motivations.'[12]

[9] Zooniverse, 'Who are the Zooniverse Community? We asked them…', Zooniverse, 5 March 2015 <https://blog.zooniverse.org/tag/survey/> [accessed 16 February 2022].

[10] Alexandra Eveleigh and others, 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science', in CHI 2014 One of a CHInd: Conference Proceedings, Toronto, Canada | April 26 - May 1, 2014: The 32nd Annual ACM Conference on Human Factors in Computing Systems, Book 3, ed. by Matt Jones and others (New York: Association for Computing Machinery: 2014), pp. 2985-2994 <https://doi.org/10.1145/2556288.2557262> (p. 2986).

[11] Eveleigh and others, 'Designing for Dabblers', p. 2987.

[12] Mia Ridge and others, 'Understanding and connecting to participant motivations', in The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage – community review version, ed. by Mia Ridge and others (online, 2021), Chapter 6 <https://doi.org/10.21428/a5d7554f.1b80974b>; Owens, 'Making Crowdsourcing Compatible', pp. 269-270.

Types of project

Citizen research in cultural heritage is broad and diverse. The key unifying thread between the different projects reported to us was that volunteers are asked to work with cultural heritage materials selected and provided by organisations.

Our report includes an array of projects ranging from small – with few volunteers – to large scale. Small-scale projects tend to be locally focused and research or information driven – the data sought is qualitative and/or based on information gathering around a single event or place. Volunteers on this type of project are often those with a local connection. Large-scale projects are more likely to be quantitative, with the tasks tending towards annotation or transcription rather than more detailed descriptions. In other words, for larger projects cultural heritage institutions usually already hold the data and seek volunteers to transcribe or annotate it, rather than generate completely new information. Both general and specialist volunteers participate in projects such as these, for example Art UK's *Art Detective* attracts volunteers with either expertise or interest in resolving questions about artwork held in public collections in the UK, and *GB1900*, a joint project coordinated by the University of Portsmouth, National Library of Scotland, the Royal Commission on the Ancient and Historical Monuments of Wales, and The National Library of Wales, appealed to those with more locally-focused interests.

The tasks undertaken as part of online citizen research projects included in this report fall into four general categories: transcription, information gathering, annotation (tagging images or text), and undertaking historical or collections-based research. Projects involving transcription were the most numerous, followed by tagging. Examples of each project type include:

Transcription: *Digital Unlocking Nature's Archive* (Natural History Museum Library and Archives), *Science in the Making* (The Royal Society)

Information gathering: *Art Detective* (Art UK), *Enrich the List* (Historic England)

Annotation: *Britain from Above* (Historic England, Historic Environment Scotland and Royal Commission on the Ancient and Historical Monuments of Wales), *Cynefin* (The National Library of Wales), *GB1900* (University of Portsmouth, National Library of Scotland, Royal Commission on the Ancient and Historical Monuments of Wales, and The National Library of Wales), *Georeferencer* (The British Library)

Research: *Liberating the Collections* (UCL), *Oaks Disaster 1866-2016* (Dearne Valley Landscape Partnership), *Shout Out Loud* (English Heritage)

Volunteer motivations

The responses we received from cultural heritage institutions reflected the findings of published literature, as illustrated by some of the examples below.

In Art UK's *Sculpture* project, one volunteer stated:

'It's an excuse to be out and about and explore different areas – and to get a better understanding of various types of art sculptures, features, history, materials used and the location chosen. It's very interesting to know the detail of each individual piece of art. I feel a satisfying sense of achievement when seeing the images after I've transferred them onto my computer's hard drive.'[13]

Volunteers in this project have also spoken about being excited to use their existing photography skills and develop new ones in the process of their work.[14] In the *Liberating the Collections* projects, volunteers noted that they appreciated being able to contribute from home in a convenient way.[15]

Competition motivates some volunteers but seems unimportant to others. In the *GB1900* project, some volunteers stated that they enjoyed the sense of competing against themselves or other contributors. However, it must be noted that gamification can also have the opposite effect, with some volunteers reporting feeling overwhelmed by the huge amount of material transcribed by other volunteers.[16] Other *GB1900* volunteers expressed collaboration and the ability to contribute to something useful as key motivators.[17]

---

[13] Akhtar Khan, 'Volunteers' Week: Akhtar Khan', Art UK, 7 June 2019
<https://artuk.org/discover/stories/volunteers-week-akhtar-khan> [accessed 15 February 2022].
[14] Dewi Owens, 'Volunteers Week: Dewi Owens', Art UK, 1 June 2021
<https://artuk.org/discover/stories/volunteers-week-dewi-owens> [accessed 15 February 2022].
[15] 'Liberating the Collections'.
[16] Aucott and others, 'Citizen science', pp. 12-13.
[17] Aucott and others, 'Citizen science', p. 8.

Successes – what went well for organisations?

Organisations reported several signs of success for their projects, relating both to their organisational aims and the impact on volunteers. In particular, projects focused on volunteer recruitment, retention and engagement levels as markers for success, as well as the completion of tasks and accuracy of the data produced. Many projects also emphasised the importance of recognising the contributions of volunteers for ensuring a project's success.

Recruiting volunteers

Being able to recruit and retain volunteers was a key marker of success for these citizen research projects. Many of the projects reported recruiting many more volunteers than they had aimed for or been able to attract for any previous in-person volunteering opportunities. Volunteer recruitment and retention are especially important as most citizen research projects are frequently reliant on a core group of especially active volunteers.[18] *GB1900* was very successful in recruiting volunteers and the project team attributes this success in part to having an easy-to-use platform that was welcoming to beginners.[19] *Operation War Diary* was similarly successful in attracting volunteers, in part because it had purposefully optimised volunteer tasks so that the barrier to entry was minimised, because it was a partnership with the world's largest citizen research platform with reach and familiarity (Zooniverse), and because it formed part of international centenary commemorations of the First World War.

For the *Science in the Making* project at The Royal Society, continued promotion was vital to ensure a sufficient pool of volunteers to keep the project going. The National Library of Wales found that having a running set of projects helped them to sustain a community of volunteers that could move from one project to another. The *Transcribing Scotland's Places* (Historic Environment Scotland), *Britain from Above*, and *Loughborough Road History* (LEAF Tenants and Residents Association) projects all stressed the importance of social media to reach potential volunteers, particularly those from a younger demographic. *Britain from Above*, English Heritage's *Belsay Hall* project, and Historic Environment Scotland's *Scottish Development Department Photographs* crowdsourcing campaigns engaged volunteers through the range of activities on offer and the multiple channels of possible contribution (email, social media and project website). The jointly coordinated Royal College of Surgeons (RCS), London School of Economics (LSE) and London Metropolitan Archives (LMA) *Unlocking the London Lock Asylum* project involved three public workshops, hosted online by LMA, in which archivists and historians from RCS, LSE and LMA helped members of the public to undertake research using the LMA dataset on the Ancestry platform. The online sessions quickly reached capacity, and around 90 members of the public attended in total.

---

[18] Lesandro Ponciano and Francisco Brasileiro, 'Finding Volunteers' Engagement Profiles in Human Computation for Citizen Science Projects', *Human Computation*, 1.2 (2014) <https://doi.org/10.15346/hc.v1i2.12>; Andrew Mao, Ece Kamar, and Eric Horvitz, 'Why Stop Now? Predicting Worker Engagement in Online Crowdsourcing', in *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, 2013, pp. 103–11 <https://econcs.seas.harvard.edu/files/econcs/files/mao_hcomp13b.pdf>.

[19] Humphrey Southall and others, 'GB1900: Engaging the Public in Very Large Scale Gazetteer Construction from the Ordnance Survey 'County Series' 1:10,560 Mapping of Great Britain', *Journal of Map & Geography Libraries,* 13.1 (2017), 7-28 <https://doi.org/10.1080/15420353.2017.1307305> (p. 20).

Data production

Success in citizen research projects is often measured quantitatively – projects, including *Art Detective* and *Science in the Making* referenced the completion of tasks in a timely fashion and the amount of data produced. The British Library's *Georeferencer* project acknowledged that the data produced as part of the project had previously been impossible to collect due to the associated expense.[20] The *Cynefin* project claimed success when volunteers georeferenced the allotted 1,100 tithe maps, transcribing 27,000 places in the process.[21] In the Aberdeen City Archives Revealed and Code the City collaborative project, *Aberdeen Harbour Board Arrivals Transcription*, volunteers have transcribed all their First World War arrival logs, totalling over 30,000 transcriptions. In The National Library of Wales' citizen research projects, volunteers completed the tasks set out in a shorter time than expected. Historic Environment Scotland undertook a communications campaign, starting in May 2020, asking users to identify images in their digital archive. As a result, over 1,200 Scottish Development Department records were revised. As of January 2021, the campaign has successfully identified over 1,700 previously unidentified images in the digital archive.

Data quality

Data quality is just as important as data quantity. Projects stressed that the data produced by volunteers must be usable and emphasised the need to test the extraction of the resulting data in the project development phase in order to ensure that data produced by volunteers will be usable after the project is finished — in effect that volunteers' work was not wasted. Early citizen research projects at The National Library of Wales reported working out what to do with the data as one of the biggest challenges, and ensured that subsequent projects had a plan and funding in place to make data publicly available, ensuring its widest possible use and reuse.

Organisations point to the relatively low need to moderate data, or low levels of data uncertainty, as a measure of success. Projects have set up different systems to validate results. Historic Environment Scotland established a network of experts and a review process to ensure an accurate digital record.[22] *Transcribing Scotland's Places* divided their volunteers into groups with assigned moderators who could check transcriptions and give tailored guidance. The *Liberating the Collections* project generated new data which highlighted underrepresented voices in the collections and this data will be used to enhance University College London's (UCL) Special Collections' catalogue data relating to female authors, and to develop new research guides.[23] *Science in the Making* encourages volunteers to check each other's work and this approach results in data of sufficient quality for the

---

[20] Kimberly C. Kowal and Petr Pridal, 'Online Georeferencing for Libraries: The British Library Implementation of Georeferencer for Spatial Metadata Enhancement and Public Engagement', *Journal of Map & Geography Libraries*, 8.3 (2012), 276-289 <https://doi.org/10.1080/15420353.2012.700914> (p. 287).
[21] Adrian Murphy, 'Cynefin project – National Library of Wales – repairing and digitising the country's tithe maps', *Museums+Heritage Advisor,* 3 March 2016 <https://advisor.museumsandheritage.com/features/cynefin-project-national-library-wales-repairing-digitising-countrys-tithe-maps/> [accessed 24 August 2021].
[22] Fredrick Alexander, 'Scottish Development Department citizen research campaign write-up' (Unpublished, Historic Environment Scotland, 2021).
[23] 'Liberating the Collections'.

needs of the project. Even though the British Library's *Georeferencer* project acknowledged that the results obtained from citizen researchers were less precise and detailed than if the work had been undertaken by professionals, they were impressed by the low error rate with 97% passing their benchmark of acceptability for the project.[24]

New discoveries

Volunteers are uncovering a wealth of new information. For example, in the *Oaks Disaster 1866-2016* project to produce a complete list of fatalities of the 1866 Oaks Colliery Disaster, volunteers were able to identify 22 previously unknown victims.[25] Coordinators of Historic Environment Scotland's *Scottish Development Department Photographs* photograph identification campaign were impressed with volunteers' creativity and ingenuity, with contributors cross-referencing maps to provide further information from Google Maps or Street View, which contributed towards the successful identification of over 1,700 previously unidentified images in the digital archive.[26] *Art Detective* participants have made many discoveries about the UK's national art collection including identifying a lost Van Dyck masterpiece,[27] and correcting over 700 artwork records.[28]

Public engagement

Engaging new publics with heritage collections was often a central aim for the citizen research projects considered in this report, with the *Britain from Above* project regarding the level of public engagement as its key success. Similarly, Historic Environment Scotland, as part of its *Scottish Development Department Photographs* campaign, measured success by being able to increase the number of users and amount of time spent on their project.[29] In the first two months of the campaign there was a 17% increase in new users and session duration, as well as a 14% increase in the number of pages-per-session.

Engaging the public brings mutual benefit: the *Belsay Hall* project reported significant engagement with people in the local community who had knowledge of the historical site. These volunteers produced useful material based on their expertise that is being fed into new interpretations being developed for the site.

As part of their *Georeferencer* project, The British Library invited their top contributors on a behind-the-scenes tour of the Maps Collection.[30] In addition, participants of the project were invited to join the Georeferencer's user group in order to connect with other participants. Responding to the shift

---

[24] Kowal and Pridal, 'Online Georeferencing', p. 286.

[25] Dearne Valley Landscape Partnership, 'New research reveals higher death toll for 1866 Oaks Colliery Disaster' Dearne Valley Landscape Partnership, 2016 <https://discoverdearne.org.uk/research-reveals-higher-death-toll-1866-oaks-colliery-disaster> [accessed 22 July 2021].

[26] Alexander, 'Scottish Development Department'

[27] ART UK, 'Achievements', ART UK, [n.d.] <https://www.artuk.org/artdetective/about/achievements> [accessed 15 February 2022].

[28] Jade King, 'Art Detective: the story so far…', ART UK, 15 March 2016 <https://artuk.org/discover/stories/art-detective-the-story-so-far> [accessed 15 February 2022].

[29] Alexander, 'Scottish Development Department'.

[30] Kowal and Pridal, 'Online Georeferencing', p. 285.

to remote working with the onset of the pandemic, two English Heritage project teams at Wrest Park and Audley End devised online talks programmes that have been well supported by staff, other volunteers and local people. This is a new innovation which the organiser suspects will continue into the future as it has proved very successful. *Britain from Above* indicated that a mixture of online and in-person activity encourages participation and could lead to a more diverse set of volunteers.

Impact on volunteers

Related to institutional public engagement aims is the impact that these projects have on volunteers themselves. For volunteers, being part of a project gives a sense of community, the feeling that they have contributed to shared knowledge, and the opportunity to use existing and new skills to do something worthwhile. For others, the opportunity to develop new skills is key. For example, as part of the *Aberdeen Harbour Board Arrivals Transcription* project, Aberdeen City Archives collaborated with Code the City, an initiative which aims to develop the public's open data, digital, and coding skills. One participant of the *Unlocking the London Lock Asylum* project reported:

"being able to work with primary sources that have never been studied so early on in my university career really helped me to understand what the process of historical research entails."

Britain from Above found that a key benefit of the project was in up-skilling the public, for example, by encouraging older people to develop digital skills and confidence.

Early feedback for the *Shout Out Loud* project suggests the young people have found the project interesting, particularly the insights it has given them into heritage research, and the sensitivities around work on contested heritage. In the *Belsay Hall* project, the longer-term aim is that the volunteers will form a self-sustaining volunteer research team that can support onsite curators, provide material for guided tours and serve as a point of contact for members of the public wanting to share memories of the site. During the *Liberating the Collections* project, three volunteers presented talks as part of UCL's Rare Book Club Series, and one submitted a poster to an academic conference.[31] The British Library team at LibCrowds designed the *In the Spotlight* project to balance productivity with enjoyment and opportunities for engagement, whether discussing interesting playbills on a public forum or social media, or investigating aspects of theatre history.

Recognising volunteers

The organisations featured in this report were mindful of the fact that dedicated volunteers deserve to be supported and recognised. Although studies have suggested that not all volunteers are motivated by competition,[32] some projects favoured an online leader board for this purpose. The *Transcribing Scotland's Places* and the *Oaks Disaster 1866-2016* projects were able to showcase and acknowledge volunteer work in exhibitions. Projects also underlined the importance of regular communication in helping volunteers to feel included and appreciated.[33] In addition to project

---

[31] 'Liberating the Collections'.

[32] Alexandra Eveleigh, Charlene Jennett, Stuart Lynn, and others, '"I Want to Be a Captain! I Want to Be a Captain!": Gamification in the Old Weather Citizen Science Project', in *Proceedings of the First International Conference on Gameful Design, Research, and Applications* (presented at the Gamification '13, New York: ACM, 2013), pp. 79–82.

[33] Aucott and others, 'Citizen science', p. 152.

updates, the British Library also found that volunteers value feedback on the quality and accuracy of their work.[34] In the *Georeferencer* project, contributors can see the progress of the project and their own participation on social media. The *Britain from Above* project stressed the importance of volunteers knowing what they are contributing to, so that they recognise the value of sharing their knowledge.[35] Some organisations acknowledged that volunteers might need their recognition to be more formal, for example, Art UK is able to provide a record of volunteering service if required.

---

[34] Kowal and Pridal, 'Online Georeferencing', p. 284.
[35] Morris, Hargreave and McIntyre, 'A bird's eye view: Evaluation of Britain from Above' (Unpublished, 2014), p. 14.

Challenges – what was difficult for organisations?

The institutions we heard from appreciated that citizen research projects are not simply an easy way to create data and engage new publics. Running a citizen research project requires a significant amount of resources and needs buy-in from a range of internal stakeholders. Some of the organisations dealt with uncertainty about the impact of their citizen research projects. In English Heritage's *Britain from Above* project, they faced questions internally about the impact of the project and decided that a clearer end goal would have made it easier to measure success.[36] In the *Belsay Hall* project, they reported that the work took up a significant amount of staff time but the outputs produced by volunteers were relatively small. Projects reported navigating some of the following challenges.

Covid-19 pandemic

Although the onset of the pandemic created some opportunities for new online connections, and led to the formulation of new projects, such as the *Aberdeen Harbour Board Arrivals Transcription* project, it also created challenging conditions for team working. English Heritage found some issues with paywalls and volunteers not being able to access the correct software while working remotely. The *Shout Out Loud* project team noted that they sometimes found it difficult to keep volunteers motivated to work remotely, with attendance at project meetings sometimes being patchy. *Belsay Hall* found it quite hard to develop a cohesive volunteer team remotely, and expressed concerns about whether the group would be sustainable once Covid restrictions were lifted. In *Liberate the Collections*, they found that volunteers wanted more opportunities to connect remotely during the project via discussions and show-and-tell events.[37] Some projects, such as *Art UK Sculpture* which requires volunteers to go out to take photographs of statues *in situ*, were greatly impacted by legal restrictions on movement and concern for volunteer physical well-being.

Platform

The technical infrastructure required to start and sustain a citizen research project is considerable and the choice of platform is central. In the *Cynefin* project, the team encountered issues with the security of their website and had to tackle spam contributions to their online community functions. The *Operation War Diary* and *GB1900* projects pointed out that sufficient follow-up resources (time and funding) are required for any project to keep servers operating and software updated.[38] *GB1900* also found that the slowness and inflexibility of their web system was off-putting for volunteers.[39] To ensure project sustainability, The Royal Society's *Science in the Making* project incorporated crowdsourcing opportunities into their search and discovery platform.

---

[36] Morris, Hargreave and McIntyre, 'A bird's eye view', p. 120.
[37] 'Liberating the Collections'
[38] Southall and others, 'GB1900', p. 17.
[39] Aucott and others, 'Citizen science', p. 12.

Project iteration

Several projects explained the necessity of having space to iterate the project after launch. The *GB1900* project tried to keep their parameters for transcription flexible so that ambiguous data could be transcribed by volunteers. They also noted that their original software needed to be modified after launch to encourage more volunteers to validate data, rather than just transcribe from scratch.[40] In the *Operation War Diary* project, the team learnt the importance of continued user testing and shifting processes if they were not generating the required data.

Data quality

Some projects had an issue with getting the data in the format they needed. Organisations are keen to produce data that is the least uncertain (and therefore requiring the least moderation) and best suited to address the project aims. There is a sense of challenge in framing data so that it adapts to the needs of three audiences: volunteers, organisation, researchers – in other words making it specific enough to be interesting to volunteers but broad enough for it to be reusable in other contexts. Our *Engaging Crowds* project identified a fourth audience: data for computational use including training data for machine learning.

When it came to the *Enrich the List* project, the text produced by their volunteers was not authenticated by Historic England or the Department for Digital, Culture, Media and Sport, and therefore cannot be part of the formal description in Historic England's catalogue.[41] The Royal Society, in their *Science in the Making* project, stressed that public engagement is not enough and projects need to produce data that will improve collections. In *Operation War Diary*, The National Archives found that it was difficult to reconcile volunteers' transcriptions of uncertain or unreadable words. It is important to standardise how users should indicate uncertainty so that the returned data is consistent. In the *GB1900* project, there was a disconnection between the citizen research project and the process of aggregating the data back into internal systems.

Sharing data

While it is generally considered incumbent on project organisers to make the results of citizen research available, it can be a complex undertaking. In *Shout Out Loud*, project organisers recognised that it was important for their young volunteers to have tangible outputs they could point towards, as some of their work was positioned in sensitive reports. One project was criticised at its launch for promising data would be available under a Creative Commons (CC)[42] licence but having no data actually available. In 2017, a year after the project's launch, a data dump was made available under

---

40 Southall and others, 'GB1900', p. 25.
41 Matthew Saunders, 'Towards a Strategy for the National Heritage list for England – A View from the Amenity Sector' (Online: Historic England, 2021) <https://historicengland.org.uk/research/results/reports/27-2021/TowardsaStrategyfortheNationalHeritageListforEngland_AViewfromtheAmenitySector> (pp. 47-48).
42 https://creativecommons.org/

a CC0 licence[43] although it was incomplete.[44] The National Library of Wales aims to apply CC0 licences where it can, although it points out that some documents are subject to agreements with third parties.[45] In the *Cynefin* project, further funding facilitated the post-project reuse of data, and it is now made available online.[46] For the early unfunded projects at The National Library of Wales, like the *Aberystwyth Shipping Records* project, the transcription was done using Excel spreadsheets which, although it allowed the volunteers to quickly transcribe the records, caused problems when making the data available: the data was made available for research use, but without a dedicated website the volunteers were left disappointed they couldn't see the results of their efforts.[47] This led to the development of Madoc, a IIIF-powered[48] collection showcase and crowdsourcing tool, part funded by the Welsh Government, to build a crowdsourcing pipeline of projects which incorporated an easy route for making the content available.

---

[43] https://creativecommons.org/publicdomain/zero/1.0/
[44] Aucott and others, 'Citizen science', p. 5.
[45] Murphy, 'Cynefin project'.
[46] https://places.library.wales/
[47] https://github.com/LlGC-NLW/shippingrecords
[48] IIIF: International Image Interoperability Framework, https://iiif.io/

Feedback from DCDC *Engaging Crowds* workshop

In July 2021 the *Engaging Crowds* project team ran a virtual workshop at the DCDC Conference to gather insights from practitioners and researchers with experience of managing and working with online citizen research projects. The 90-minute workshop was attended by 55 people, including facilitators and partners from various *Towards a National Collection* projects. Participants were asked to consider different elements of volunteer engagement, including motivations, activity, pastoral care and diversity, and discussed how to maximise the usefulness of volunteer efforts, including with regards to cleaning, shaping, presenting, sharing and reusing cultural heritage data.

After a brief introduction, participants were polled about their experience of online citizen research. Participants identified the benefits of working with citizen research as breaking down barriers and attracting greater diversity of volunteers, bringing in expertise from the public, building closer links with the community, providing flexible volunteer opportunities, and enriching information about collections. Challenges experienced by participants included finding volunteers with the necessary skills, engaging diverse volunteers, ensuring quality control, maintaining volunteer motivation throughout the duration of a project, finding time to manage volunteers, managing expectations, providing social interaction, and measuring volunteer engagement beyond time spent and amount of data created.

Participants were then randomly assigned to breakout rooms in small groups where they discussed five questions:

1.   How do we build and sustain a community of citizen researchers?

Participants discussed the importance of determining in advance the project's focus, the collections to be used, and/or the desired outcomes, in order to steer the community. It was suggested that using existing links to communities and individuals was easier than building a community through open promotion, but that there is a tension around using pre-established volunteer communities and engaging new volunteers and increasing volunteer diversity.

Participants emphasised that in order to sustain a community of citizen researchers it is necessary to offer a variety of different tasks a) to keep volunteers engaged; and b) to enable volunteers to work to their own strengths and interests. It was seen as vital to keep open channels of communication with volunteers throughout a project, updating them on progress and discoveries. It can be a challenge after a project has ended and organisers move onto other work to maintain regular contact with community members: if possible it should be encouraged as a demonstration of the ongoing value of their work, and to enable re-engagement for future projects. Providing rewards and organising other activities (e.g. trips, events) can help sustain the community.

2.   What are the best ways to guide and support citizen researchers to complete tasks?

A number of tips were suggested, including: identifying and recruiting those who are going to be motivated by the project's/organisation's mission; providing communities with some level of co-creation of the project and ownership over the work they do; and ensuring that all staff involved in the project appreciate that managing volunteers is, to an extent, everyone's responsibility.

Participants discussed how internet access and access to devices can be a barrier to online citizen research. If possible, access to, or resources for the procurement of, devices should be provided, and online digital know-how training can be incorporated into the project to overcome some of the

associated barriers and ensure that volunteers are gaining skills and experience as part of their participation.

3.   What challenges and opportunities arise with data produced by citizen researchers?

Participants suggested that challenges with data produced by citizen researchers centre around quality control. It takes time to manage the data produced and a system is required to check the data, whether that be automated or manual, such as through the use of volunteer moderators. A further challenge was identified around data formats and ensuring that the tool used to support the project exports data in a way that is compatible with other systems used by the organisation.

Participants reported that the greatest benefit of data produced by citizen researchers is that it can incorporate information and expertise from people outside of the organisation who are able to provide additional expert, local, personal and/or contextual information which enhances metadata and collections.

4.   What would you do differently next time if you were starting a citizen research project from scratch?

The majority of suggestions for what workshop participants would do differently relate to the pre-launch planning phase of a citizen research project. Participants suggested doing more detailed testing of tools and workflows, agreeing terminology, and identifying pitfalls so that they can be overcome before the project begins. Participants would also consider in more detail the scale of volunteer input, the data quality required, and how the data is to be managed before the project starts, and undertake a risk assessment taking into consideration the workload of those managing the project.

5.   What should happen when a citizen research project ends?

Participants identified three activities which should happen once a citizen research project ends. First, volunteer involvement and the impact on volunteers needs to be assessed. Well-being and ethics of care needs to be factored in so that organisations are able to evaluate volunteers' feelings about the project, their levels of engagement, and their awareness of how they have contributed to the project. Second, the project should be evaluated to assess whether it has met predetermined formal goals, to measure how successful it has been, and establish what has been gained as a result of the project. Third, feedback should be provided to volunteers to outline their contributions (as a community) and the overall outcomes of the project. Project outcomes should also be shared with the wider community and sector for accountability, and to celebrate citizen contributions.

Recommendations

Having explored the successes and challenges encountered by the organisations referenced in this report and the participants at the DCDC workshop, a number of recommendations can be made for future citizen research projects:

Don't underestimate the resources required to set up and run a citizen research project.

Establish required outcomes or Key Performance Indicators before the project begins and identify how these will be measured throughout. These can be used to report progress to internal stakeholders, and shared with project volunteers and the wider community to celebrate citizen contributions and new knowledge created.

Invest time in project design and set up. Thoroughly test tools, workflows, and data outputs as part of project development to ensure that they meet the needs of the organisation and make them as easy to use as possible for volunteers.

Be prepared to make changes to the workflow and continue testing throughout the project. Design the project to be iterated, updated, and extended on an ongoing basis.

Put citizen researchers at the centre of project design to ensure a positive impact on volunteers, for example by providing knowledge exchange opportunities and enabling them to develop new skills.

Use multiple avenues/platforms for recruitment to attract a diverse body of citizen researchers.

Offer citizen researchers variety: of tasks offered; of ways to engage with the project; of thanks and acknowledgment given; and avenues for feedback.

Communicate regularly with volunteers about project progression and achievements and acknowledge the contribution they make.

Decide on quality standards and establish a system to ensure that data is validated.

Make data generated by volunteers widely available accompanied by clear and easy to locate licence information.

Conclusions

Crowdsourced data makes an invaluable contribution to our collective understanding of our shared culture and history. Despite the known issues around, and barriers to, digital access, citizen research contributes to opening up cultural heritage collections, both in terms of relationships between the public and organisations through the project activity itself, and in terms of the results, for example through producing open access datasets or digitised versions of items that were previously unavailable.

Here, a democratic and transparent approach to promotion and process is a key lesson learnt: providing the public with multiple channels of entry (social media, hackathons, websites etc.) will ensure a more diverse crowd with which to engage. Being as generous as is feasible with context is also essential – the more informed the crowd, the better the data will be. Also, a key feature is that workflows should include a component that allows creativity, or for volunteers to contribute extra information, or get involved in ways not necessarily previously conceived of by the project organisers.

The project planning phase has been found to be key to a project's eventual success. Rigorous testing is required to ensure that the platform will be simple for volunteers to use, produces data that meets the needs of the organisation, and that can be incorporated into its other systems. It is vital that the data is of high enough quality to meet the needs of the organisation and its users and data moderation needs to be built into the project from the start with a plan in place for making the data available to the public.

Communication is vital both during and after the project. Volunteers, motivated as they are by a desire to contribute and who then develop active interest in the project, benefit most from regular feedback. Initial instructions, therefore, should be simple, but, at the same time, further enquiries should be responded to personally and with increasing detail. Organisers, in short, need to build responsiveness into their workflows and maintain open channels of communication throughout the duration of the project. At the end of the project, it is vital that volunteers are recognised and kept up-to-date with the outcome of their work, and how the data is, will, and can be used.

References

Alexander, Fredrick, 'Scottish Development Department citizen research campaign write-up' (Unpublished, Historic Environment Scotland, 2021).

ART UK, 'Achievements', *ART UK*, [n.d.] <https://www.artuk.org/artdetective/about/achievements> [accessed 15 February 2022].

Aucott, Paula, Humphrey Southall and Carol Ekinsmyth, 'Citizen science through old maps: Volunteer motivations in the GB1900 gazetteer-building project', *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 52.3 (2019), 150-163 <https://doi.org/10.1080/01615440.2018.1559779>.

Collective Wisdom, 'Collective Wisdom: The state of the art in cultural heritage crowdsourcing', *Collective Wisdom,* 2022 <https://collectivewisdomproject.org.uk/> [accessed 15 February 2022].

Creative Commons, 'CC0', *Creative Commons*, [n.d.] <https://creativecommons.org/share-your-work/public-domain/cc0/> [accessed 15 February 2022].

Crowston, Kevin and Isabelle Fagnot, 'Stages of motivation for contributing user-generated content: A theory and empirical test', *International Journal of Human-Computer Studies*, 109 (2018), 89-101 <https://doi.org/10.1016/j.ijhcs.2017.08.005>.

Dearne Valley Landscape Partnership, 'New research reveals higher death toll for 1866 Oaks Colliery Disaster' Dearne Valley Landscape Partnership, 2016 <https://discoverdearne.org.uk/research-reveals-higher-death-toll-1866-oaks-colliery-disaster> [accessed 22 July 2021].

Eveleigh, Alexandra and others, 'Designing for Dabblers and Deterring Drop-Outs in Citizen Science', in CHI 2014 One of a CHInd: Conference Proceedings, Toronto, Canada | April 26 - May 1, 2014: The 32nd Annual ACM Conference on Human Factors in Computing Systems, Book 3, ed. by Matt Jones and others (New York: Association for Computing Machinery: 2014), pp. 2985-2994 <https://doi.org/10.1145/2556288.2557262>.

Howe, Jeff, 'Crowdsourcing: A Definition', (2006) <https://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html> [accessed 24 February 2022].

Khan, Akhtar, 'Volunteers' Week: Akhtar Khan', Art UK, 7 June 2019 <https://artuk.org/discover/stories/volunteers-week-akhtar-khan> [accessed 15 February 2022].

King, Jade, 'Art Detective: the story so far…', *ART UK*, 15 March 2016 <https://artuk.org/discover/stories/art-detective-the-story-so-far> [accessed 15 February 2022].

Kowal, Kimberly C. and Petr Pridal, 'Online Georeferencing for Libraries: The British Library Implementation of Georeferencer for Spatial Metadata Enhancement and Public Engagement', *Journal of Map & Geography Libraries*, 8.3 (2012), 276-289 <https://doi.org/10.1080/15420353.2012.700914>.

'Liberating the Collections: A UCL Special Collections remote volunteering project - Evaluation report' (Unpublished, UCL, 2021).

Morris, Hargreave and McIntyre, 'A bird's eye view: Evaluation of Britain from Above' (Unpublished, 2014).

Murphy, Adrian, 'Cynefin project – National Library of Wales – repairing and digitising the country's tithe maps', Museums+Heritage Advisor, 3 March 2016 <https://advisor.museumsandheritage.com/features/cynefin-project-national-library-wales-repairing-digitising-countrys-tithe-maps/> [accessed 24 August 2021].

Owens, Dewi, 'Volunteers Week: Dewi Owens', *Art UK*, 1 June 2021 <https://artuk.org/discover/stories/volunteers-week-dewi-owens> [accessed 15 February 2022].

Owens, Trevor, 'Making Crowdsourcing Compatible with the Missions and Values of Cultural Heritage Organisations', in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Abingdon: Ashgate Publishing, 2014), pp. 269-280.

Ridge, Mia, 'From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing', Curator: The Museum Journal, 56.4 (2013), 435-450 <https://doi.org/10.1111/cura.12046>.

Ridge, Mia, Samantha Blickhan and Meghan Ferriter and others, *The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage* (online, 2021) <https://doi.org/10.21428/a5d7554f.1b80974b>.

Ridge, Mia and others, 'What is crowdsourcing in cultural heritage?', in *The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage*, ed. by Mia Ridge, Samantha Blickhan and Meghan Ferriter and others (online, 2021), Chapter 2 <https://doi.org/10.21428/a5d7554f.1b80974b>.

Ridge, Mia and others, 'Understanding and connecting to participant motivations', in *The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage*, ed. by Mia Ridge, Samantha Blickhan and Meghan Ferriter and others (online, 2021), Chapter 6 <https://doi.org/10.21428/a5d7554f.1b80974b>.

Saunders, Matthew 'Towards a Strategy for the National Heritage list for England – A View from the Amenity Sector' (Online: Historic England, 2021) <https://historicengland.org.uk/research/results/reports/27-2021/TowardsaStrategyfortheNationalHeritageListforEngland_AViewfromtheAmenitySector>.

Southall, Humphrey and others 'GB1900: Engaging the Public in Very Large Scale Gazetteer Construction from the Ordnance Survey 'County Series' 1:10,560 Mapping of Great Britain', Journal of Map & Geography Libraries, 13.1 (2017), 7-28 <https://doi.org/10.1080/15420353.2017.1307305>.

Zooniverse, 'Who are the Zooniverse Community? We asked them…', *Zooniverse*, 5 March 2015 <https://blog.zooniverse.org/tag/survey/> [accessed 16 February 2022]

Appendix – list of projects

Aberdeen Harbour Board Arrivals Transcription, Aberdeen City Archives Revealed Project and Code the City https://codethecity.org/2020/05/02/aberdeen-harbour-board-arrivals-transcription-project/

Aberystwyth Shipping Records, National Library of Wales. https://blog.library.wales/aberystwyth-shipping-records-2/

Art Detective, Art UK.
https://artuk.org/about/about

Aviva Archive.

Belsay Hall Project, English Heritage. https://www.english-heritage.org.uk/visit/places/belsay-hall-castle-and-gardens/belsay-awakes/

Book of Remembrance Project, National Library of Wales.
https://www.library.wales/discover/digital-gallery/manuscripts/modern-period/the-welsh-national-book-of-remembrance#?c=&m=&s=&cv=&xywh=-580%2C0%2C5991%2C7178

Britain from Above, English Heritage and Historic Environment Scotland.
https://www.historicenvironment.scot/archives-and-research/archives-and-collections/britain-from-above/

Cynefin, National Library of Wales. https://www.library.wales/discover/projects/end-of-project/cynefin

Digitally Unlocking Nature's Archive, Natural History Museum Library and Archives.
https://www.nhm.ac.uk/take-part/citizen-science/digitally-unlocking-natures-archive.html

Durham History Centre.

Enrich the List, Historic England. https://historicengland.org.uk/listing/enrich-the-list/

From the Page.
https://fromthepage.com/

GB1900, University of Portsmouth, National Library of Scotland, Royal Commission on the Ancient and Historic Monuments of Wales and National Library of Wales.
https://geo.nls.uk/maps/gb1900/#zoom=6&lat=55.0000&lon=-2.5000&layer=0

Georeferencer, The British Library*.*
https://www.bl.uk/projects/georeferencer

In the Spotlight, The British Library.
https://www.libcrowds.com/collection/playbills/about

Liberating the Collections, University College London, Special Collections.
https://blogs.ucl.ac.uk/special-collections/tag/liberating-the-curriculum/

Loughborough Road History Project, LEAF Tenants and Residents Association.
https://loughboroughroadsw9histories.wordpress.com/

Hussein, Abira, and Mnemoscene, *Nomad.*
https://nomad-project.co.uk/#workshops

Madoc.
https://cultural-heritage.digirati.com/building-blocks/madoc/

Oaks Disaster 1866-2016, Dearne Valley Landscape Partnership. https://discoverdearne.org.uk/the-oaks-disaster/

Operation War Diary, The National Archives. https://blog.nationalarchives.gov.uk/tag/operation-war-diary/

Recogito.
https://recogito.pelagios.org/

Science in the Making, The Royal Society.
https://makingscience.royalsociety.org/s/rs/page/welcome

Scottish Development Department Photographs, Historic Environment Scotland.
https://www.historicenvironment.scot/about-us/news/members-of-the-public-are-asked-to-help-identify-scotland-s-archives/

Shout Out Loud, English Heritage. https://www.english-heritage.org.uk/learn/shout-out-loud/

The Zooniverse.
https://www.zooniverse.org

Transcribing Scotland's Places, Historic Environment Scotland.
https://www.historicenvironment.scot/archives-and-research/archives-and-collections/scotlandsplaces/

University of Michigan Library, *Behind the scenes: unearthing Michigan ecological data.*
https://daily.zooniverse.org/2020/02/20/behind-the-scenes-unearthing-michigan-ecological-data/

Unlocking the London Lock Asylum, Royal College of Surgeons, London School of Economics, London Metropolitan Archives. https://www.lse.ac.uk/News/Latest-news-from-LSE/2020/K-November/Unlocking-the-past

Wrest Park Project, English Heritage. https://www.english-heritage.org.uk/visit/places/wrest-park/