Crowdsourcing and Citizen Science in Cultural Heritage

Mia Ridge

2025-05-21

A quick guide offering practical tips on how anyone can get started with crowdsourcing in their own library or cultural heritage institution.

Introduction

Crowdsourcing in cultural heritage is a method for enabling meaningful public participation in research or practical tasks based on cultural heritage collections or knowledge (adapted from the Collective Wisdom Handbook. In practical terms, a typical example involves asking people outside the institution to contribute effort via tasks such as transcribing, tagging, researching or sharing artefacts. Crowdsourcing projects tend to focus on creating enjoyable, inherently rewarding experiences that lead to high quality data with minimal risk of errors. This means that crowdsourcing platforms can also be a great way for staff to work with their own collections.

For example, you could ask locals to share their memories, stories or artefacts about your region, or you could work with the public to co-create an exhibition or collaborate with international experts on research tasks. Some volunteers want follow their interests in specific topics, while others enjoy serendipitious glimpses into varied collections.

So perhaps more importantly than the data collected or enhanced, crowdsourcing in cultural heritage is a way of inviting people to spend time with collections and institutions that they might never encounter otherwise. For example, the Living with Machines project engaged over 5000 online volunteers on the citizen science platform, many of whom had no previous experience with library or humanities research, and no previous relationship with the British Library.

While crowdsourcing in other fields might involve payment for tasks completed (for example, via Amazon's Mechanical Turk), the rewards for contributing to crowdsourcing in cultural heritage are usually intrinsic or altruistic. Thinking of it as a form of online volunteering can be helpful - participants can help enhance collections metadata, research objects or stories,

contribute their own knowledge and experience, etc, with opportunities to learn or socialise during the process. But unlike traditional volunteering, crowdsourcing isn't limited to a venue's location or hours of operation - online projects can be open to anyone, anywhere in the world, 24 hours a day.

Volunteers are often motivated by their interest in a topic or source type (e.g. beautiful maps or interesting photographs), the challenge of completing a task well (e.g. deciphering old handwriting), developing their skills and knowledge (e.g. becoming more accurate as they practise palaeography, or learning more about a collection or research question). They often stay motivated because they get feedback from project teams, and gain a sense of purpose or community.

Increasingly, libraries might combine crowdsourcing with machine learning and AI tools. For example, by building workflows that automate work such as pre-selecting relevant records, distributing tasks to volunteers with matching skills and interests, then supporting staff in quality checking and formatting the results. Automating tedious (and easily verifiable) tasks with machine learning can help create enjoyable volunteer or staff experiences.

It can also be used to 'scale up' the results of volunteer work. For example, the 'Ad or not' task we created for Living with Machines relies on people's ability to understand the purpose of short pieces of text - was it an advertisement or part of an article? The results of this task became the 'ground truth' dataset for training an experimental machine learning model that could assess whether other texts were ads or not.

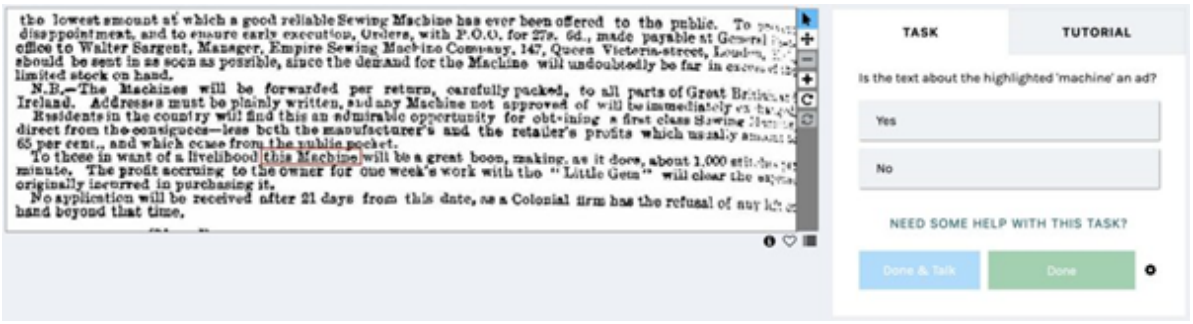

Figure 1: Screenshot of the Zooniverse task interface, with an historical newspaper image on the left, and options to label it 'yes' or 'no' in response to a question asking if it is an advertisement

Screenshot of an early version of the 'ad or not' task on Zooniverse. This simple 'yes or no' format also meant that the task was available in the Zooniverse app, further increasing its reach.

What's in a name?

'Crowdsourcing' is an awkward name, with implications of 'outsourcing' and anonymous crowds, but to date it is the most widely recognised term. Other terms used for similar work include digital public participation, community-generated digital content (CGDC), online volunteering, and variations such as 'niche-sourcing' for small or invitation-only projects.

'Citizen science' is another commonly used term with a lot of overlap with crowdsourcing. Citizen science projects might include natural history observations in the world (for example, recording wildlife, or monitoring water quality) or screen-based tasks such as counting penguins in photographs. With roots in a broader concept of 'science' (Wissenschaft) as knowledge or areas of study, 'citizen science' also includes citizen history, humanities (Geisteswissenschaften), social sciences and any other field that works with knowledge about the world. As inherently interdisciplinary spaces that welcome specialists and the public, libraries are excellent places to host crowdsourcing and citizen science projects.

Whatever it's called, recognising the work that volunteers put into projects is important. Some projects do this by hosting events for volunteers, others ensure that they are named and credited in publications and datasets.

Is crowdsourcing right for you?

Crowdsourcing isn't for everyone nor the answer for every need. For example, running a successful project draws on a range of skills, and may require collaboration across many departments in an institution. The development of machine learning / AI methods and increasingly sophisticated crowdsourcing platforms can reduce the amount of work required to gather source material, review and quality control contributions and process the resulting data, but you will still need the resources and inclination for reviewing and sharing progress reports, social interactions with volunteers, and responding to questions. It may take a few iterations to create tasks that produce useful data via tasks that will attract volunteers. If you don't enjoy talking to volunteers, negotiating with colleagues and wrangling resources (or don't have any resources to spare), it might not be right for you right now.

Ideally, you would also be able to ensure that the source collections, desired data results and types of tasks available on your platform of choice are a good match by prototyping or piloting workflows before committing to a full project. That said, you don't have to limit your project to the types of tasks you've seen before - you can invent new tasks and workflows, and work with new technologies to meet your needs. For example, the Living with Machines project invented a 'close reading' task that asked volunteers to discern the sense in which specific words were used, supported by computational linguistic analysis.

Relevance to the Library Sector (Case Studies/Use Cases)

Cultural heritage institutions can support citizen science projects that ask participants to make observations about the natural world. For example, community science projects at the UK's Natural History Museum ask people to help investigate noise pollution and local pondlife.

Libraries, archives and museums can ask volunteers to help create or enhance collection data by transcribing handwritten text, entering data from catalogue or specimen cards into databases, or adding tags or labels to describe images. For example, Smithsonian Digital Volunteers: Transcription Center.

They might also ask volunteers to research objects or record information from their personal knowledge. For example, photos on Flickr Commons have been tagged with locations, personal names and histories, and specialist object labels identified by people with local or historical knowledge about photographs.

Library and other GLAM staff can initiate projects, help manage and run them, and check, process and ingest data from crowdsourcing projects. For example, the Library of Congress reviewed comments left on their Flickr Commons images, and updated some of their collections records with information provided by the public.

Libraries can support projects run on national portals, such as Latvia's iesaisties.lv. They can point language learners or people with local knowledge to projects in a range of languages and other national portals.

There's a strong relationship between IIIF and crowdsourcing. In 2017 the British Library used IIIF images for the 'In the Spotlight' project, in part to save hosting costs, and in part to explore the potential for IIIF annotations in crowdsourcing. This later inspired a collaboration with Zooniverse to support importing media into Zooniverse via IIIF manifests. Crowdsourcing platforms that support IIIF include the Digirati / National Library of Wales/Madoc platform and From the Page.

Hands-on activity and other self-guided tutorial(s)

The best way to learn more about crowdsourcing is to try a range of different projects. This will help you understand participant motivations, get a sense of the importance of great titles and instrutions in getting you started, and think about how data might move between GLAM systems and crowdsourcing platforms.

You can find projects to try at:

- https://www.zooniverse.org/projects
- https://fromthepage.com/findaproject
- https://scistarter.org/
- https://transcription.si.edu/

- http://lesherbonautes.mnhn.fr/

If your organisation has records in Europeana, you might be able to devise crowdsourcing tasks for them on the CrowdHeritage site.

Recommended Reading/Viewing

If you are interested in learning more about participating in, creating or using data from crowdsourcing projects, a recent comprehensive publication is the open access publication, The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage.

The Collective Wisdom Handbook is designed to answer the most common questions that people have as they think about starting, maintaining and using data from a crowdsourcing or citizen science project. Topics covered include:

- What is crowdsourcing in cultural heritage?
- Why work with crowdsourcing in cultural heritage?
- Identifying, aligning, and enacting values in your project
- Designing cultural heritage crowdsourcing projects
- Understanding and connecting to participant motivations
- Aligning tasks, platforms, and goals
- Choosing tasks and workflows
- Supporting participants
- Working with crowdsourced data
- Managing cultural heritage crowdsourcing projects
- Connecting with communities
- Planning crowdsourcing events
- Evaluating your crowdsourcing project

Finding Communities of Practice

When you're ready to think about creating crowdsourcing projects or working with crowdsourcing data it's useful to reach out to existing communities of practice for support and feedback.

The LIBER Citizen Science Working Group is a community of practice open to all LIBER members. They are currently producing a guide for Citizen science in Research Libraries. Topics published to date include:

- Skills: Citizen Science skills development for staff, researchers, and the public
- Infrastructures: As being active in the development of infrastructure for researchers to carry out Citizen Science

The low-traffic JISCMail discussion list on crowdsourcing has many interested members and is a great way to connect with others who have an interest and experience in crowdsourcing in cultural heritage.