



TOWARDS
A NATIONAL
COLLECTION



Arts and
Humanities
Research Council

FOUNDATION PROJECTS

SUPPORTING DOCUMENTATION

People and machines: co-creating with heritage collections

Engaging Crowds: citizen research and
cultural heritage data at scale

Mark Bell, Ashleigh Hawkins, Louise Seaward, Pip Willcox
The National Archives

Contents

Introduction	1
Scope	2
Summary of key recommendations	3
Workshop participants	4
Discussions summary	4
Tasks	4
Motivation	6
Data	7
Final provocations	9
Key recommendations	10
Conclusion	11
Acknowledgements	11
References	12

Introduction

Machine-learning technologies such as Handwritten Text Recognition (HTR) are beginning to be used to open and transform cultural heritage material, enabling search, discovery and the creation of linked data. Whereas in 2014, Ridge described cultural heritage crowdsourcing projects as asking the public to undertake tasks that cannot be done automatically,¹ today, some projects are asking volunteers to undertake tasks which could be completed by a machine with a high level of speed and accuracy. A deeper exploration of the opportunities and challenges presented by increased incorporation of machine-learning technologies into crowdsourcing projects has yet to be undertaken. This workshop was designed to start these discussions.

The 'People and Machines' interdisciplinary workshop explored the best routes to fuse digital innovations with the dedication and enthusiasm of volunteers through discussion of what makes an effective crowdsourcing task, how to maintain volunteer motivation, and methods of supporting volunteers to produce useful data, both as part of traditional crowdsourcing, and as part of workflows incorporating machine learning.

This report uses the term 'machine learning' to refer to a range of technical approaches which use statistical models and algorithms to analyse and draw inferences from patterns in data. Machine learning models and algorithms 'learn' from the data they are given, and autonomously adapt and improve their accuracy in response. Machine learning, which includes supervised learning, unsupervised learning, and deep learning, is a subfield of Artificial Intelligence.

¹ Mia Ridge, 'Crowdsourcing Our Cultural Heritage: Introduction' in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Abingdon: Ashgate Publishing, 2014), pp. 1-13 (p. 2).

Scope

This workshop was part of a series of workshops within the project Engaging Crowds: citizen research and heritage data at scale: a collaboration between The National Archives, the Royal Botanic Garden Edinburgh, Royal Museums Greenwich and Zooniverse at the University of Oxford. Engaging Crowds is a Foundation Project within the Arts and Humanities Research Council-funded Towards a National Collection programme.

The workshop was co-organised by The National Archives and Aarhus City Archives. It took place in person at The National Archives in December 2019.

Participants at this workshop considered a set of research questions relating to the efficiency, enjoyment, and ethics of combining machine learning and crowdsourcing in cultural heritage collections.

In order to examine these questions, expressions of interest were welcomed from cultural heritage practitioners and researchers from a range of disciplines including historians, web designers, Human-Computer Interaction researchers, social scientists, data scientists, computer scientists and more, who worked within the academic, industry or heritage sectors. There were a total of 42 participants, including the organisers.

The workshop was structured in two halves, both of which followed the same format of a brief introductory presentation followed by discussion in small groups of an assigned topic. In the morning session, participants considered which tasks are most effective for engaging volunteers, how to keep volunteers motivated, and how to help the crowd produce useful data. In the afternoon session, participants focused on the impact of integrating machine learning into heritage crowdsourcing on supporting tasks, motivating volunteers, and producing useful data.

This report contains a summary of discussions at the event and a set of recommendations for the incorporation of machine learning in crowdsourcing projects.

Summary of key recommendations

Participants at the workshop emphasised the many outstanding questions regarding the ethics and practicalities of incorporating machine learning into crowdsourcing workflows. The scenarios in which machine learning could be beneficial to a project need to be established, and user-friendly tools need to be developed. Volunteer's appetite for co-creating with an algorithm is yet to be determined. At the same time, workshop participants questioned at what point it becomes exploitative to rely on volunteers to undertake repetitive tasks which could be undertaken by a machine. From the discussions, three recommendations emerged towards bringing people and machines closer together as part of cultural heritage crowdsourcing:

Technological Transparency: project owners need to be transparent around the development, training, and use of a machine-learning tool, why it is being used in a specific project, and how the role of the volunteer interacts with that of the machine.

Collaboration:

With volunteers: involve volunteers in technology development, to train machine learning models and validate their outputs.

Between heritage organisations: organisations need to work together to develop and share adaptable and shareable tools.

Between people and machines: project owners will need to ensure that machine learning improves the volunteer experience.

Communication: communication is essential to help motivate volunteers, to develop a sense of community and of being valued, to explain the use of machine learning, to develop clear instructions and to provide regular feedback.

Workshop participants

Organisers: Mark Bell (The National Archives), Soren Poder (Aarhus City Archives), Louise Seaward (The National Archives), Pip Willcox (The National Archives).

Presenters: Samantha Blickhan (Zooniverse, University of Oxford/Adler Planetarium), Grant Miller (Zooniverse, University of Oxford), Mia Ridge (British Library).

Discussions summary

The following questions were discussed in groups of seven, designed to include representation across the sectoral, professional and disciplinary range of workshop participants in each group, and made up of participants, organisers, and presenters. Each question was discussed by two groups: the summaries presented here synthesise their discussions.

Tasks

Which kind of tasks are most effective for engaging volunteers with heritage crowdsourcing?

Participants discussed the differences between micro-tasks (e.g. correction) and macro-tasks (e.g. transcription of a chunk of text) noting that micro-tasks bring higher levels of accuracy, with macro-tasks bringing more errors and, consequently, increasing the need for moderation and correction. Micro-tasks are good for engaging time-poor volunteers but need to be made meaningful and, if possible, varied. Project owners need to appreciate that different volunteers will have different motivations, and frame projects so that they target a particular audience. It is necessary to match tasks to the motivations of the crowd, for example by providing different tasks to target different audiences. Enabling volunteers to change and develop their roles as they gain experience is also important, as longer term volunteers are often more enthusiastic, and produce better data. There is a clear link between engagement in the project and empowerment, with suggestions from participants including: giving volunteers the ability to give and withdraw consent easily, to have an element of control over their data, and the right to anonymity.

In order to guide volunteers to become better transcribers or taggers it is best to start with easy tasks, before offering more complex tasks based on volunteers' experience. Some level of feedback needs to be incorporated into the workflow, for example, automatically via accuracy scores for each user, or by providing feedback targeted at either the community as a whole, or tailored to volunteers individually.

Communication is central to ethically encouraging volunteers to engage with projects and return to them repeatedly. It is important to explain a project's motivations for seeking user contributions and to make explicit what a project owner wants from the crowd, to make research impacts clear, to be transparent about how data and other outputs will be used, and to promote both outputs and volunteer contributions, for example via social media. It is the organisation's responsibility to give back to

volunteers, for example, by showing gratitude, ensuring the project includes an element of investing in the training and upskilling of volunteers, or providing volunteers with a platform to share their expertise.

It is also beneficial to ensure that a project develops a network, rather than solely making use of free labour. Including community building as part of a crowdsourcing project enables volunteers to see (and learn from) what others are doing, and provides opportunities for them to chat and share their thoughts. This is useful for the organisation too, who can have a better understanding of volunteers, gain immediate and regular feedback from contributors, and keep project owners involved in actively engaging volunteers by responding to comments. For projects which engage larger, more diverse, and new crowds, the benefits to volunteers need to be kept in mind throughout the project to ensure that it isn't just providing a platform for expert super-contributors to share their knowledge. For this, open channels of communication are vital.

How can we use machine learning to support heritage crowdsourcing tasks?

Participants emphasised that machines cannot replace the human ability to think laterally, answer a range of questions, or find different interests in a single object. At the same time, they questioned whether it is ethical to submit volunteers to a repetitive task that a machine can do, and at what point it becomes exploitative not to invest in machine learning.

It is necessary to determine in what scenarios machine learning will be beneficial to a project, for example the project must have enough of the right kind of data to feed the machine learning model, and be able to ensure that its use will result in data with a high enough level of accuracy for its intended purpose. We note that the range and extent of suggested tasks will alter over time as technologies are further developed and their reliability and trustworthiness improved. Suggested tasks in which machine learning could be incorporated to improve heritage crowdsourcing projects include:

Allocating tasks based on volunteers' interests and strengths.

Inferring or validating answers from the crowd.

Predicting whether a contributor will provide useful data.

Pre-processing (e.g. using Optical Character Recognition (OCR) or HTR) of data to create micro-tasks, or to redact sensitive documents before users can see them.

Classifying images.

Identifying connections and similarities across datasets.

Volunteers could be involved in training machine learning algorithms and validating their outputs, but at the same time it is vital to ensure that inexperienced volunteers are not feeding inaccurate data into the machine learning algorithm. Machine learning could help optimise the value of expert volunteers by presenting tasks to users according to their interests and expertise, giving people more challenging tasks the more they contribute.

To facilitate the reuse of machine learning models for heritage crowdsourcing, the model's development and training must be documented in detail and collaboration between organisations and projects is

necessary so that models, data, and opportunities for volunteers can be shared more easily. However, there are a number of issues making the reuse of machine learning models developed for crowdsourcing particularly challenging, including their specificity, the use of many data standards to represent the same data across different contexts within heritage, and issues relating to intellectual property rights and other access restrictions.

Motivation

How can we keep volunteers motivated?

In order to keep volunteers motivated throughout the duration of a project it is necessary first to discover the motivations of the crowd. User surveys and a discussion forum or chat function were suggested as useful ways in which project owners can get the feedback they need to identify the range of motivations of their volunteers.

Each volunteer will have their own motivations and it is necessary to design a project so that it can meet the needs of individuals in the crowd. Participants noted their experience that the majority of people will want lighter interactions, for which it is necessary to contribute a small amount of time, with a minority of volunteers wanting deeper interactions, for which a greater time commitment is required. Projects should be designed so that volunteers can participate as much as they want, whenever they want. Similarly, volunteers will respond best to different methods of incentivisation. Providing volunteers with the opportunity to opt in to the incentive structure they prefer can have a positive impact on their level of contribution: for example, some will be motivated by the introduction of a competitive element, whereas others respond negatively to gamification, either by it making the stakes seem lower, or making the tasks seem more intimidating.

Suggestions for motivating volunteers included:

Providing training so that volunteers can progress.

Making clear what volunteers are contributing to and the impact of their contribution.

Using time-bound challenges with a new focus each month.

Offering a range of tasks, including tasks with low completion barriers; providing 'super-users' the opportunity to take on larger roles.

Building a community with an appointed community manager, and using regular communication to foster a sense of community.

Giving volunteers some level of ownership over what they have done, for example by being able to curate a subset of the data they have contributed.

Again, communication is key to motivating volunteers. A clear communication strategy throughout the duration of the project (e.g. via a blog or forum) can provide regular positive feedback to volunteers and raise awareness of their contributions more widely. It is vital that volunteers feel valued and that credit is given to them for their contribution in order to create a sense of usefulness and meaningfulness.

Organisations need to be open about their decision making, and, where possible, involve user communities in technology design.

How can we use machine learning to motivate volunteers?

Participants suggested that machine learning could be incorporated into crowdsourcing workflows in order to motivate volunteers by, for example, delivering a more interesting and diverse set of tasks to volunteers, or providing suggestions for words to assist during transcription tasks. Use of machine learning tools could be seen as a means of liberating volunteers from tedious and repetitive parts of a project so that their time can be spent on more complex and interesting tasks.

In order to ensure that the incorporation of machine learning does not discourage users, it is necessary to be transparent about its use and how the role of the volunteers interacts with it. To overcome volunteer disengagement, projects will need to identify and communicate the motivation for using machine learning so they can separate out what it is that machines can and cannot do, and to get people to focus on what they are experts at. As the process may be even less co-creational than traditional crowdsourcing, it may be necessary for the rewards of participation to be more explicit, and for project owners to place more emphasis on articulating the value of volunteer contributions. Community managers are more important in projects incorporating machine learning, in order to help users understand that what they are doing is worthwhile.

Explaining machine learning to volunteers in an understandable way will also be key. One approach would be to portray machine learning as a tool which will save time in some areas but will also create new areas of work on which more people's time will need to be spent. With the help of machines, better services can be built; nevertheless every effort should be made to overcome the 'black box' connotations of machine learning by being as transparent as possible about the tools, how they work, and how they were developed.

Data

How can we help the crowd to produce useful data?

Participants defined useful data as 'data that is accurate and can be shared'. To ensure that data created through crowdsourcing is shareable, it must include provenance metadata, indicate that it was produced by the crowd, and how it was created. Such data supports trust in the crowdsourced data, credits how the work was done, and makes the volunteers' contribution and the infrastructure transparent. Collaboration between organisations, and the creation of tools that are adaptable and shareable, will help to improve both the accuracy of data and the volunteer experience.

A number of suggestions were made to encourage volunteers away from mistakes, including:

Foregrounding the goals of the project.

Providing clear objectives, training, online tutorials, guidance, support etc.

Enabling experienced volunteers to become moderators.

Maintaining communication with volunteers so they can provide, and be provided with, feedback.

Providing real-time responses to volunteer questions.

Appointing a community manager.

Incorporating verification into the workflow.

Using relevant data standards where possible, and providing information about them to volunteers.

Breaking the project down into small tasks with varying levels of difficulty.

To create accurate data whilst also valuing the contribution of volunteers, it is important that communication with volunteers is maintained throughout a project. Volunteers need to be given the opportunity to communicate with, and learn from, each other. Project owners need to recognise the costs and trade-offs of using volunteer labour while aiming to retain editorial and institutional authority to ensure the resulting data is useful and usable. It should be remembered that members of the public can be more expert than the project owners, whose role it is to harness and value this expertise.

How can we use machine learning to produce useful data from heritage crowdsourcing?

Participants noted that the definition of what is useful changes when machine learning is incorporated into heritage crowdsourcing. Different things are possible when you work with people or with machine learning. Machine learning can be used to improve the accuracy of volunteer-generated data by integrating tools and moderation into workflows. Although machine learning could enable checking data at a much greater scale it is likely to have a lower accuracy than could be achieved by volunteers. Users may tend towards exaggerated trust in content generated by machine learning. It is necessary for project owners to identify gaps in the training data used: for example, by comparing with other models on commonly used test data or using federated models or models produced by trusted consortia, in order to ensure the data used has a provenance; and identifying biases within the data that the model is trained on so that training datasets can be augmented as necessary. Project teams and volunteers need to be trained in identifying gaps and made aware of the importance of both their role, and that of the data, in the machine-learning pipeline.

Organisations working with others and being open about a machine learning model's workings and development is seen as vital. It was suggested that institutions could share models and data with each other through a shared platform, with federated machine learning enabling institutions to train machine learning models collaboratively while keeping their data safe. To embed explanation of the processes used to produce the data in its metadata it is necessary to cite the model, weights, and which training data was used. Project owners could use containerisation technology, which would enable them to distribute a full machine learning pipeline run in a virtualised environment, bundled with all necessary software and data for maximum reproducibility, and data provenance. Participants suggested that a 'crowdsourcing with machine learning network' could productively communicate with intellectual property rights experts, and that lessons could be learnt from commercial companies in this space.

Final provocations

At the close of the workshop participants were asked a final set of provocative questions. A selection of their responses is included below:

Is crowdsourcing exploitative?

Crowdsourcing should not be exploitative, but it can be if the project's true purpose is hidden. Crowdsourcing projects should be transparent about their aims. Volunteers need to have agency and be rewarded for their contributions. The expectations of participants and project owners need to be managed. The outputs of the project need to be of benefit to the public.

Does crowdsourcing produce useless data?

It is an ethical imperative on the people setting up the project to ensure that data is not useless. There are two scenarios that risk the creation of useless data: where platforms enable volunteers to 'make up' their own rules about the data they collect; and where the project designers have not considered how they (and others) will use the data collected.

With the amount of data growing exponentially, will we run out of potential volunteers?

Yes but ... this is why it is important to use both people and machines to carry out crowdsourcing projects.

No but ... we will run out of people with particular skills e.g. reading old handwriting.

Maybe but ... we imagine machine learning will become more efficient enabling machines to undertake a larger portion of the workflow.

Key recommendations

Having explored the various discussions of participants at the workshop, a number of recommendations can be made for the incorporation of machine learning in crowdsourcing projects:

Technological Transparency:

Project owners need to be sure, and able to articulate, that incorporation of machine learning into their crowdsourcing workflow will be beneficial to both the project and volunteers, and will result in the production of useful and usable outputs.

Project owners need to identify biases in the machine learning model so they can be mitigated, and augment any gaps in the training data.

Transparency around decision-making and the development/selection and training of machine learning models and tools is necessary in order to enable reuse of the tools, models and data, and to ensure transparency and openness with volunteers.

Datasets created as a result of crowdsourcing projects using machine learning will require additional metadata, for example, to document the model and training data used.

Collaboration:

Volunteers should be involved in developing technology, training machine learning models, and validating their outputs.

Heritage organisations need to collaborate to develop machine learning models and tools which are adaptable, can be shared, and meet the specific needs of crowdsourcing in a heritage context.

Machine learning should be incorporated in ways which improve the volunteer experience, for example by undertaking repetitive and tedious tasks in order to provide volunteers with more varied tasks, to support volunteers in tasks such as transcribing, or to automate the moderation process and improve the accuracy of data.

Communication:

Communication is central to the success of a crowdsourcing project incorporating machine learning: to explain the use of machine learning clearly, to help motivate volunteers, to encourage long-term engagement, to provide clear instructions, guidance and regular feedback, to emphasise the value of volunteer contributions, and to facilitate community building.

Project owners need to identify and communicate their motivations for using machine learning in order to explain how and why machine learning is being used, and to assign the tasks best suited to volunteers.

Community managers are important. There needs to be increased emphasis on articulating the value of volunteer contributions. Rewards need to be more explicit.

Conclusion

Participants at the workshop were positive about the possibilities of incorporating machine learning into crowdsourcing projects in ways which would provide useful outputs and meaningful interactions for volunteers. It is evident that increased collaboration (both with volunteers and cross-organisationally), communication, and technological transparency will be vital for the successful incorporation of machine learning into crowdsourcing projects.

There will be a number of additional challenges that those wishing to develop crowdsourcing projects incorporating machine learning will need to address. From concerns over volunteer exploitation, to considerations of the shareability of machine learning models, both ethical and practical questions have emerged. Moreover, for many institutions, technological barriers, such as infrastructure, data, and technical skills, mean that the incorporation of machine learning into crowdsourcing will seem far out of reach.

The workshop provided a useful starting point for these discussions, and more detailed research into the challenges and opportunities machine learning presents to crowdsourcing, as well the impact on volunteer interaction and engagement, will need to be explored further.

Acknowledgements

Thank you to all the participants and presenters who took part in this workshop.

References

Ridge, Mia, 'Crowdsourcing Our Cultural Heritage: Introduction' in *Crowdsourcing Our Cultural Heritage*, ed. by Mia Ridge (Abingdon: Ashgate Publishing, 2014), pp. 1-13.

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>>.