Crowding the Library: How and why Libraries are using Crowdsourcing to engage the Public

Sarah Severson
Digital Initiatives Projects Librarian
University of Alberta Libraries
sarah.severson@ualberta.ca

Dr. Jean-Sébastien Sauvé
École de bibliothéconomie et des sciences de l'information
Université de Montréal
jean-sebastien.sauve@umontreal.ca

Keywords

digital technology; crowdsourcing

Abstract

Over the past 10 years, there has been a noticeable increase of crowdsourcing projects in cultural heritage institutions, where digital technologies are being used to open up their collections and encourage the public to engage with them in a very direct way. Libraries, archives, and museums have long had a history and mandate of outreach and public engagement but crowdsourcing marks a move towards a more participatory and inclusive model of engagement. If a library wants to start a crowdsourcing project, what do they need to know?

This article is written from a Canadian library perspective with the goal to help the reader engage with the current crowdsourcing landscape. This article's contribution includes a literature review and a survey of popular projects and platforms, followed by a case study of a crowdsourcing pilot completed at the McGill Library. The article pulls these two threads of theory and practice together, with a discussion of some of the best practices learned through the literature and real-life experience, giving the reader

practical tools to help a library evaluate if crowdsourcing is right for them, and how to get a desired project off the ground.

Introduction

April 2018 saw the much anticipated launch of the Library and Archives Canada Co-Lab, a brand new crowdsourcing tool, which comes with the promise that it will "allow contributors to reveal our history, page by page" ([Library and Archives Canada](#), 2018). While still in beta, this tool lets anyone with an internet connection explore Library and Archives Canada's digitized collections and then tag, describe, translate, or transcribe them. This is a great example of how an institution can extend the reach of their digital collections by making a meaningful and proactive effort to gather input and encourage interaction from the public.

This announcement is just one in a long line of projects where cultural heritage institutions are using digital technologies to open up their collections and encourage the public to engage with them directly. Libraries, archives, and museums have long had a history and mandate of outreach and public engagement, but crowdsourcing marks a move towards a more participatory and inclusive model of engagement.

In this paper, we will explore what the current crowdsourcing landscape looks like for libraries, archives, and museums with a focus on Canadian case studies and examples in an attempt to help answer the following question: if and how Canadian libraries can apply crowdsourcing at their institutions. We will start with a definition and background on the topic before presenting a survey of the pertinent literature in both French and English and review a number of crowdsourcing tools. We will conclude with a report on a pilot done at McGill Library and some best practices that we have learned for doing digital crowdsourcing work. The aim of this paper is to look at crowdsourcing with an eye to the Canadian context which includes smaller institutions, with unique cross and inter-cultural collections and distributed communities.

Background

Coined in 2006, the term 'crowdsourcing' was first introduced in a *Wired* business magazine article where it was defined as "the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call" (Howe, 2006). As the practice of crowdsourcing has moved beyond its function and business origins and into the cultural heritage sector we have seen the understanding of the term evolve. As Estellés-Arolas and González-Ladrón-de-Guevara (2011) have pointed out, this emerging and cross-disciplinary practice is difficult to restrict to a single definition. In this article, we have adopted a broad definition of crowdsourcing as a practice increasingly focused on how to involve the expertise and motivation of the public to work on tasks that help better describe collections or make them more usable. We are not looking at the social internet applications used in libraries to encourage interaction or outreach like HistoryPin or Wikipedia, but rather focusing on crowdsourcing projects that are focused on task-based interactions that specifically enrich a collection.

Applications of crowdsourcing range from intellectually engaging work, such as transcribing handwriting, to more micro tasks, such as identifying if an animal is present in a picture. When looking at the variety of existing crowdsourcing projects there are any number of tasks, including tagging images, identifying people or places, proofreading texts, transcribing handwritten records, text encoding, translation, or even the creation or contribution of original content.

Libraries and crowdsourcing

Recent large-scale digitization efforts have made millions of pages and images available online, but making those same collections discoverable in a meaningful way is time consuming and often cannot keep pace with these digital transformations. Many projects are explicit in acknowledging that without the help of the public the sheer scope of work in transcription or description would be virtually impossible. Another added effect of crowdsourcing work is the transfer of labour from paid staff at the institution to the voluntary public. While this can sometimes be seen as a method of saving money for the institution, we will argue later in our case study and best practices that any labour cost that is saved is then spent on managing and engaging the project and community.

Crowdsourcing the transcription of handwritten text or the tagging of images can dramatically increase the findability of these often hidden digital collections by making them searchable. In 2016, the Royal BC Museum launched *Transcribe*, which focused on making letters, diaries, journals, and other written historical materials available for transcription. Launched with a transcription project focused on letters and diaries written by British Columbians during the First World War, the projects have now expanded to include a diverse set of records and letters relating to British Columbian history.

A major emerging advantage of transforming digitized collections into machine-readable data is that it opens up the door to new forms of scholarship that leverage digital technologies and computer networks. One such example is Data Rescue: Archival & Weather (DRAW), which is a new platform that crowdsources transcriptions of the historical weather logs captured at the McGill weather observatory since 1863. The project goal is to transcribe the weather observations into a database to help atmospheric scientists develop a better understanding of how climate has been changing over the last hundred and fifty years of record keeping.

By "unveiling important parts of our history," as Library and Archives Canada says in their press release (2018), for input and engagement, we also have a unique opportunity to engage with the contributions and input of a more diverse community. Bringing in more perspectives and knowledge when describing or tagging collections, a process that was previously typically controlled by experts, can bring in a wider diversity of languages and allow us a deeper understanding of our descriptive practices. One early example of this is Library and Archives Canada's *Project Naming,* which launched its site in 2004 as a part of a larger project. The website enabled Indigenous peoples to identify photos of their ancestors, relatives, and themselves online from the collection

and at the same time share their knowledge with each host institution to improve descriptions and understanding of their holdings (Smith, 2008).

The practicality of online distributive work is relatively new and linked to the rise of fast computer networks and cheaper computing terminals, which created the opportunity to break down complete tasks and send them out to a distributed workforce for processing (Ridge, 2014). While fast networks enable this work, it is also the emerging nature of a more social internet which brings together diverse populations, making it easier to turn to the public to solicit knowledge and contribution on any number of projects.

Literature review

Crowdsourcing has received considerable attention from professionals and scholars since Howe's original article in 2006. Crowdsourcing was first primarily related to commercial practices, which Howe expanded on in his book (2009), where he primarily focused on the power of the crowd to solve complex problems that the individual could not.

Early on, professionals working in the cultural heritage sector showed an interest in incorporating crowdsourcing into their activities to open up their collections. As a result, a body of literature on the subject has developed over the past 10 years that is specific to the cultural heritage context. Four major themes emerge from the literature surveyed: general crowdsourcing theory and definition, the use of crowdsourcing in the library context, specific aspects of crowdsourcing, such as gamification, volunteers' participation, and user experience, and detailed case studies of crowdsourcing projects.

Since the idea of crowdsourcing is emerging and cross-disciplinary, it is not surprising that many authors want to define and understand crowdsourcing within our institutional context. Through the surveying of 32 articles, Estellés-Arolas and Gonzáles-Ladrón-de-Guevara (2012) proposed a broad definition, which is built around eight general characteristics that address what is being done, who is doing it, and for what gain. This definition is widely accepted by the literature. In addition to definitions, scholars have also developed scientific typology, a categorization of crowdsourcing project types (Geiger & Schader, 2014; Oomen & Aroyo, 2001), and conceptual models (Pedersen et al., 2013; Dunn & Hedges, 2013; Hedges & Dunn, 2012; 2017) in order to better understand and conceptualize crowdsourcing.

According to Hedges and Dunn (2012) there is an observable pattern in crowdsourcing of four elements: assets (type of content, primary material like geospatial data, text, image, sound, video, ephemera or intangible heritage, numerical or statistical data), tasks (activity done by the volunteers to the asset), processes (combination of tasks including transcribing, correcting, tagging, categorizing, cataloguing, linking, contextualizing, recording/creating content, commenting/critical responses, mapping, georeferencing, translating), and outputs (what is produced in the end). They argue that understanding the interaction of these four elements is critical to a successful project.

General considerations about crowdsourcing have also been covered specifically for cultural heritage and LAMs. McKinley's (2012) seminal report, offers a general overview and identifies crowdsourcing as part of a shift towards participatory culture in museums, archives, and libraries. Similarly, the potential of crowdsourcing in libraries has been covered extensively in the literature. This ranges from Trevor Owens' (2012) series of thoughtful blog posts to a number of edited volumes that bring together various projects and voices. Mia Ridge, digital curator at the British Library since 2015, edited the volume *Crowdsourcing Our Cultural Heritage* (2014), which serves as one of the most complete overviews on crowdsourcing and its effectiveness within heritage collections and institutions. Her introductory text is followed by case studies of projects relating to cultural heritage and essays on some of the challenges and opportunities of cultural heritage crowdsourcing. Dunn and Hedges, who had a chapter featured in Ridges book, came out with their own edited volume in 2018 titled *Academic Crowdsourcing in the Humanities: Crowds, Communities and Co-production* and focuses on a theoretical framework to understand crowdsourcing. Internationally, Pauline Moirez's (2017) recent essay responded to her 2013 report for the Bibliothèque nationale de France in which the author utilized a similar theoretical framework.

A considerable amount of scholarship also focuses on specific aspects of crowdsourcing, such as the motivations of users and gamification. Volunteers' motivation is probably the most popular issue covered by scholars since their participation is at the core of a project's successful completion (Alan & Campbell, 2017; Hossain, 2012; Oomen & Aroyo, 2011; Reed et al., 2013; Ridge, 2013). According to Puah, Bakar, and Ching (2011), volunteers may take part in crowdsourcing projects solely because of their willingness to contribute to human knowledge. The authors bring elements from user experience and gamification into the discussion since they underline the importance of the platform capitalizing on the volunteers' motivation. Gamification appears especially as a recurrent subject in the literature. After Greenhill et al. (2015), Mohnheuser et al. (2017) completed an extensive literature review showing that gamification makes projects more successful. Throughout their analysis the authors were able to identify specific types of gamification according to different crowdsourcing project categories.

 Another major theme in the literature about crowdsourcing is the prevalence of detailed case studies on specific projects, typically written by people working at the respective institutions. Notably, there were no detailed case studied on Canadian projects. These include Transcribe Bentham (Causer et al., 2012; Moyle et al., 2011), Operation War Diary (Grayson, 2016), Decoding the Civil War (Durkin, 2016), or platforms like Zooniverse (Banks, 2013; Cox et al., 2015; Feder, 2016; Simpson et al., 2014), and Digitalkoot (Chrons & Sundell, 2011). These articles usually trace the different steps the professional faced while undertaking crowdsourcing projects within their own institutions. They present a selection of results regarding the volunteers' participation or the workflows themselves, along with tips or best practices. These articles are helpful to gain a detailed understanding about a specific project, but they often focus on large-scale projects that are difficult to translate into more general contexts.

Review of crowdsourcing projects and platforms

Many of the early and successful crowdsourcing projects were multi-year custom built systems but over the past 10 years there has been a movement towards standardizing the application of these tools through open source software and ready to use platforms that you can leverage to build a project with little to no programming support.

Projects

One of the first library projects to show the success of crowdsourcing was the National Library of Australia's Trove project. Launched in 2009 as a delivery and discovery platform for National Library of Australia digitized newspapers project, one of the most notable features was the ability to correct the text generated by Optical Character Recognition software. By 2013 users had corrected 100 million lines of text, which they estimate is the equivalent of 270 standard work years of effort (Ayers, 2013). The National Library of Finland approached a similar project with a large newspaper collection by creating Digitalkoot, which gamified the act of correcting the newspapers' text (Chrons & Sundell, 2011).

The New York Public Library (NYPL) has also built a number of custom projects to work on specific areas of their collection to great success. In Emigrant City, you could work to uncover what kinds of history are found in the bond and mortgage records from The Emigrant Savings Bank during the years 1841–1933. Volunteers were asked to identify types of information, transcribe words and numbers, then act as adjudicator between different transcriptions to help normalize the data. The data is then periodically merged and uploaded to the website for anyone to use.

Another popular example is NYPL's What's on the Menu? project, which asked users to transcribe historical restaurant menus. Like Emigrant City, the final product from the project is a set of downloadable data files that describe food dishes, prices, and restaurant locations that anyone can use for research (Lascarides & Vershbow, 2014). A good example of how this data can be leveraged in pedagogy can be found in the work of Trevor Muñoz, Assistant Dean for Digital Humanities Research at the University of Maryland Libraries, who has used data compiled from the What's on the Menu? project to teach graduate students about digital humanities data curation, and written about his process extensively on his blog (Muñoz 2013a; 2013b; 2014).

Other libraries, archives, and museums are building crowdsourcing into their entire collections. These projects typically have a wider diversity of materials and tasks, which gives them greater breadth in data for both researchers and the community. This approach must be balanced by the potential for low participation if the goals are too vague or the volume of work is daunting. Depending on the structure of the project this might be an acceptable trade off. Library and Archives Canada has tried to strike this balance by having their entire collection available to crowds, but they promote smaller groups of thematically related materials as challenges. For example, Rosemary Gilliat (Eaton)'s Arctic diary and photographs were featured in order to help connect with potential groups of volunteers and help ensure that some smaller collections can be

found and worked on. Alexandra Haggart, head of the initiative at Library and Archives Canada, also recognizes that transcription and translation work is already being done by researchers for their own use, and by providing a platform that encompasses their entire site, they are not limiting what people will work on (Johnson, 2018). The University of Iowa DIY History site has seen almost 87,000 pages transcribed at the time of writing and is part of a larger long-term commitment to new forms of public engagement and scholarly publishing. Project Transcribe Bentham (Causer et al, 2012; Causer & Terras, 2014; Moyle et al, 2011) is another example of a successful awarding winning site that was launched in 2010 and built with open source code.

Platforms

One of the outcomes of these early crowdsourcing projects was the acknowledged need to develop software platforms that could be easily used to build a project, thus opening up the option of crowdsourcing. Most platforms allow a project to be built quickly with little to no programming required and can support a number of different task types.

One such early example includes the 2008 Metadata Games, an open source platform that focused on the gamification of descriptive tagging for cultural institutions. Currently, Metadata Games is being used at 11 institutions including the British Library, Boston Public Library, Digital Public Library of America, and the American Antiquarian Society. Another open source platform that supports the transcription of tasks is Wikisource, created by Wikimedia and aimed at crowdsourcing a digital library of transcribed texts. The Bibliothèque et Archives nationales du Québec has been actively working with Wikimédia Canada to contribute digitized copies of public domain books from their collections and support their transcription with monthly workshops at the main library, La Grande Bibliothèque.

FromThePage is an open source software platform where you can upload scanned documents for transcription, correction, and translation for a monthly fee. Easy to use and fast to set up, FromThePage is suitable for a smaller library or research team focused on the process and task rather than public engagement. In the latest version, they added support for project management, TEI-XML export, and Omeka or Internet Archive integration.

The Zooniverse platform is one of the most popular crowdsourcing platforms with over 83 active projects at the time of writing in 2018. It was originally built for a single project in 2007 to support an astrology project that worked to classify millions of astronomical images. In 2009 Oxford University took what they learned from this project and turned it into a platform where anyone could build their own crowdsourcing project This mature platform now supports a variety of project types and different disciplines ranging from the humanities to the sciences. Zooniverse has an active and diverse community of developers and volunteers that have helped develop the user friendly interface.

Case study: Piloting Zooniverse at McGill Library

At the McGill Library, crowdsourcing has come up a number of times while brainstorming about new digital projects or outreach efforts with our digitized collections. Crowdsourcing was usually brought up as a good way to get people involved in our collections and to potentially transform large amounts of data, like the transcription of handwritten letters.

After these conversations, we realized we did not have enough information on how a crowdsourcing project would best fit with our collections or how resource intensive it would be for the library. In order to answer these questions, we turned to the literature and reviewed the Zooniverse website but found the information either too specific to an individual case or too broad to help understand the full impact of a crowdsourcing project.

While we knew we could have done more research and created reliable enough resource estimates and a project plan, as an organization the McGill Library, and the Digital Initiatives department particularly, learns best by doing. In the winter of 2018 we had the opportunity to work with a master's student from Université de Montréal École de bibliothéconomie et des sciences de l'information on a project for 3 months. We decided to take this opportunity to do a small internal crowdsourcing project as a pilot with one of our digitized collections. The crowdsourcing platform Zooniverse was selected early as an obvious choice for our pilot since we wanted to focus on something out of the box that did not need any custom development. The goal of the pilot was to gain a better understanding of how the Zooniverse platform worked, what kinds of data we could get from it, and how it could be used for future projects.

The first stage of the pilot was to establish the scope and the materials for the project. We approached the Rare Books and Special Collections department and asked them to identify a few different collections they would be interested in using for crowdsourcing. The first two collections were composed of William Osler's handwritten memo books and a set of manuscript cookbooks. These were considered because we were most interested in learning about transcription as we have extensive manuscript collections and an active research community on campus doing text analysis. These two collections were not selected in the end because the handwriting was considered difficult to read, and there was a concern that it would be difficult to get people engaged with the collections. The third collection that was considered was a set of ledgers from the Montreal Stock Exchange fonds in order to explore how Zooniverse handled numerical information. Because of the 3-month timeline, we only considered collections that had already been digitized.

Figure 1. Photograph of a silver cup from the Traquair collection management system.

The fourth collection that was considered and then chosen was a digital collection of just over four hundred black and white photographs from professor and architect Ramsay Traquair of Québec of silver artefacts that he used in his published volume *The Old Silver of Quebec* in 1940. The photographs were mounted in albums and had captions written by Traquair on both the back and front, which included drawings of the different silver marks. We chose this collection because the captions were small and with the handwriting and the drawings, we would be able to test two different tools in the Zooniverse interface. Finally, this collection had contextual information in the form of essays on the digital collection website, which could be easily repurposed for the texts in the crowdsourcing project and was already earmarked for migration into our new archival

Building the project and designing the workflows

With the materials chosen, the next step was to start building the Zooniverse project and design the workflow, or the series of tasks, we were going to ask our volunteers to do. Based on the information on each photograph we determined two sets of workflows were needed. One workflow asked the volunteer to identify where the caption text was on the photograph by underlining it and then transcribe the handwriting. The workflow was to be completed on both the front and back of the photographs.
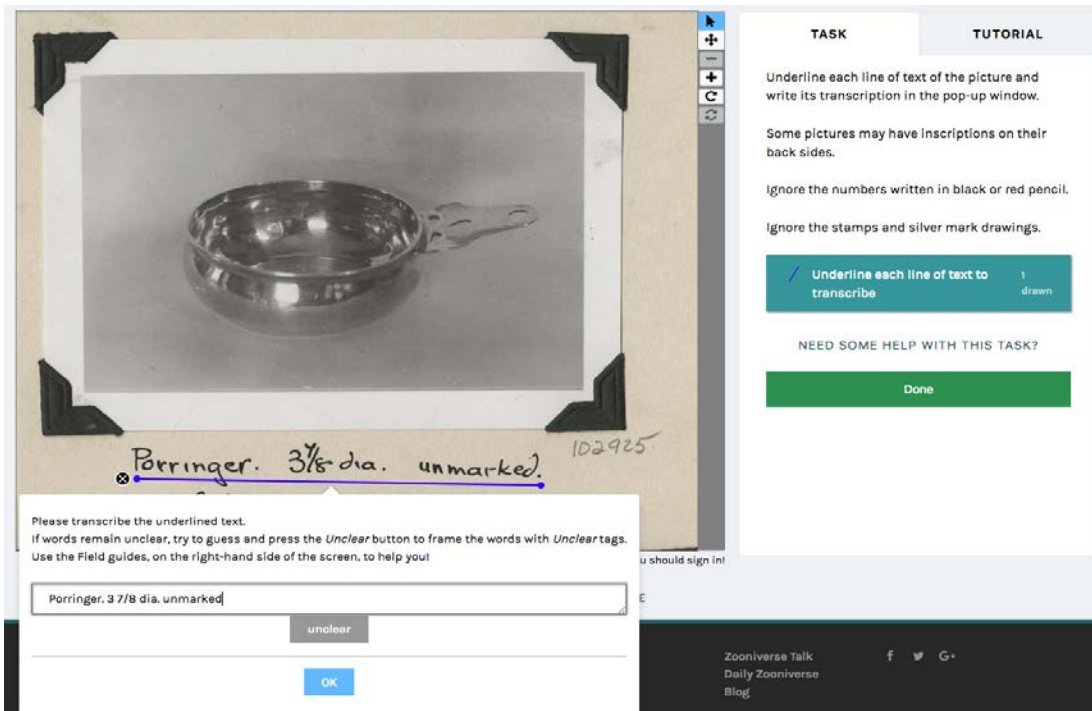

Figure 2. Screenshot of the transcription workflow

The second workflow asked them to indicate if there were any silver mark drawings on the photograph by drawing a box around it, then navigate to the field guide and match it to the silversmith's name and active dates. The field guide information came directly from a table in Professor Traquair's book *The Old Silver of Quebec* (1940).
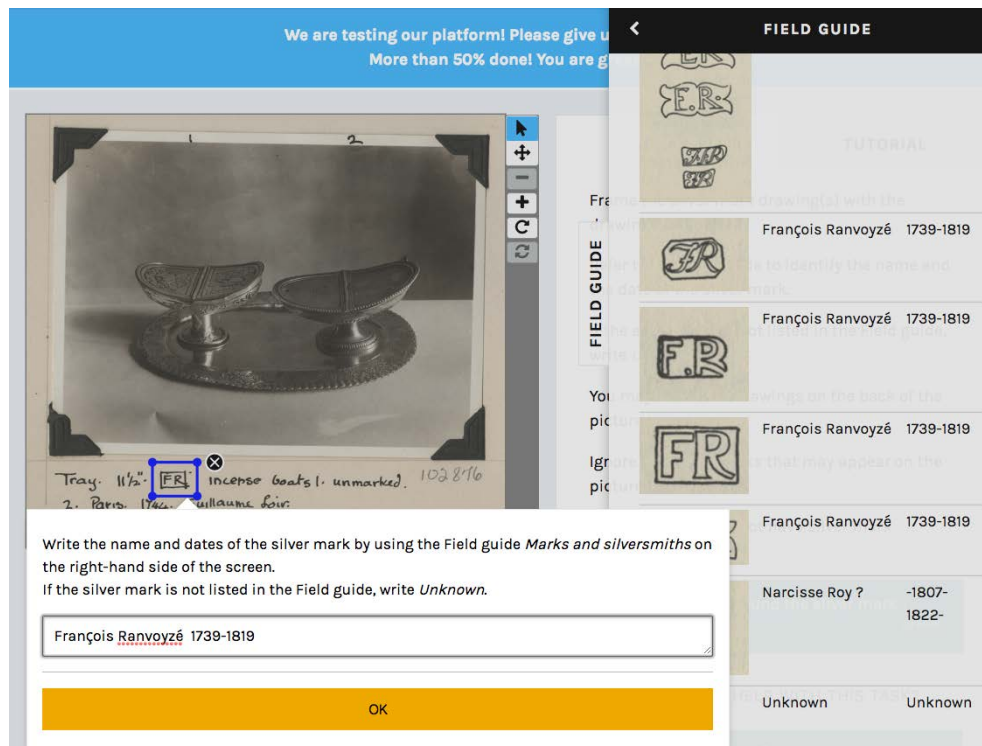

Figure 3. Screenshot of the silver mark identification workflow

Proposing two workflows to the volunteers was intentional as part of the pilot because it made it possible to compare different kinds of tools and questions in order to see which would be most easily completed by the volunteers. Also, the workflows would generate different kinds of data that would need to be handled differently.

Completion

In crowdsourcing, the term 'completion' is used to indicate that you have collected enough data. According to Zooniverse most of the tasks are considered complete after five to ten evaluations, depending on the difficulty of the workflow, and this number can be set for each type of task. For the purposes of the pilot we chose to set the required number of completed workflows at three since the project was small and the data was going to be double checked by a trained staff member when it was put into the archival catalogue. The completion number was important because even if we needed only three people to complete each workflow on four hundred images, that meant we needed over 1,200 workflows for the project to be considered done. This was a daunting number; when we looked at the literature on past crowdsourcing volunteer behaviour and motivation we saw the average person would only do a few tasks before moving on.

Understanding what the total number of tasks or completed actions were needed for the project to be completed turned out to be an important conversation to have early because it brought to light the numbers of participants we would need in order to complete even a small project like ours.

Using the research and texts from the original digital collection we were able to create the first iteration of the project using the Zooniverse project builder in two days. The first phase of testing was done locally by sending the project to our internal library staff, which numbered at just over 130 people. About six staff members completed some workflows and sent in some extensive feedback on the design and wording of the tutorials. This led to a fairly extensive rewrite, using simpler language and the elimination of some of the steps in the workflow, taking us from six clicks to just three.

The second phase of testing involved sending out the test project to a larger group for more input. We identified a number of small groups outside the library that might have interest in either crowdsourcing in libraries, such as The McGill University School of Information Studies digital curation classes and specialized art history groups and classes, that would be more interested in the collection. In total we estimate that emails were sent to about two hundred people, inviting them to participate in the project and send feedback. Considering no follow up was done beyond the one email, it was not that surprising we did not receive any feedback from this group.

The third phase of testing was done with Zooniverse. We submitted the project to testing, and after their review and approval, it was sent out to their community of active volunteer testers that number over one thousand. Each project that is built on the Zooniverse platform is given access to this group of motivated testers who are primarily invested in the crowdsourcing platform. While it can be assumed this group had little to no familiarity with the subject matter of our project, they are an active and motivated group who were familiar with crowdsourcing. Along with the project URL they are sent a standard form with specific questions from Zooniverse, such as:

- How easy or difficult did you find the task?
- In your opinion, is this project suitable for the Zooniverse?
- If we decide to launch this project publicly, do you think you will take part?

The motivated group of testers and the form allowed us to gather significantly more feedback. In total, 30 people filled out the form, with 19 of them finding the workflow tasks to be either very or moderately easy and only two saying it was impossible or confusing. Overall, we left the project open to all of the testing groups for a month and we received over 50 comments and feedback through email, the form, and the forums.

We identified three major areas for consideration and feedback:

1. The Zooniverse platform (user interface, button placements, layout). Much of this could be addressed with custom programming but we considered most comments were minor.

2. The workflows (tasks, tools, tutorials, field guides). We were able to address with changes to the choice of photo examples and consistency of language.
3. The objects (context and content).

Feedback regarding the Zooniverse platform indicated that many users expected a more sophisticated user interface. For example, many noted they did not know how to choose between the two workflows. The link "classify" in the general Zooniverse menu leads to one of the two tasks automatically, and the buttons leading to the workflows are available only on the project main page. Also, some volunteers argued that the field guide tab (on the right-hand side of the screen) was hard to find. Other volunteers pointed out the absence of an "Undo" action on the platform. Similarly, another volunteer highlighted that there was no possibility to click on a button to skip an image that was less interesting and instead was expected to refresh the page. Unfortunately, none of these elements can be changed by the project owner without customized programming.

In offering feedback about our project's workflows, some users mentioned that they were not always certain what the objectives of tasks were: "It seems unclear if you want the dates written on the notes or the dates written in the field guide." Once users had decided what we were asking them to do, many encountered problems using the tools available to help them carry out their tasks. Many identified problems relating to the use of the drawing tool to underline the captions to be transcribed such as it being difficult to place and "the drawing tool was unpredictable." One user indicated that the platform's drawing tools did not work properly on their tablet: "[…] very fiddly on a tablet."

Finally, the specific collection of this project and its context posed particular challenges as well. The readability of the architect's handwriting was frustrating to many volunteers: "Getting used to Traquair's handwriting is the most difficult part; samples of his writing to read would be nice to understand how he formed different letters." The French text also caused concern for volunteers. Many pointed out that the English-speaking volunteers would not necessarily pay attention to the French language special characters: "This needs more information to non-French speakers." Furthermore, church-related objects appeared unpopular to some testers: "I am not enamoured of church objects or impressed by silver or valuables." All these comments were carefully considered, and any changes that occurred because of feedback were communicated to the volunteers.

Pilot outcomes

The pilot was designed as a learning exercise for us as an organization, so in the end we opted not to submit the project for inclusion in the public Zooniverse platform. Through targeted testing just over 70% of the photograph captions were transcribed and 30% of the silver marks identified. In the end, the silver mark identification was thrown out because too few were completed and the accuracy was so low. The captions were found to be about 50% correct but almost all needed some manual intervention before they were used to enhance the item level records for these objects in our archival management system.

Overall our experience with the Zooniverse platform was positive, and we learned a lot about the platform and how we might use it in the future. While it cannot do everything, what it did do, it accomplished well, and the ease of use and low barrier of entry were important. A small archives or library could easily take a collection and create a project in a short amount of time. The most important thing we learned was that it was critical to spend time designing the workflow tasks and thinking about the user experience in order to make sure it matched with the platform. It is also important to gain a better understanding of how the task choice can affect the resulting data. For example, in our case study we found that using the underline tool meant that the final transcribed data ended up in a number of different columns and was harder to normalize. Normalizing the data happens outside of the platform and, depending on the project, can be difficult to do. The Zooniverse community shares various programming scripts to help with clean up, but this was not accessible to people without programming knowledge.

Best practices

Based on our literature review, the project summary, and our own pilot experience, we have distilled four major areas of best practices.

Collection

Picking the right material for a crowdsourcing project is the first step towards success, and for each organization, it will be different. What we learned is that you need to select materials that are interesting to your community. That could be a highly specialized collection that appeals to a specific research group or an undergraduate class, or it might be an entire collection that has a broad appeal, as with Co-Lab.

The second major consideration when you are selecting material is to consider what kinds of questions you want answered which can be supported by the types of tasks done with crowdsourcing. For example, we found the identification of the hand drawn silver marks in our pilot project was both too specialized and too difficult of a task to make crowdsourcing effective. Another factor to consider is if you can get useful data out of the material. Ideally, the tasks can be performed on all presented objects. If volunteers are asked to transcribe textual information when the object does not have any, they will think that they misunderstood the task they are to fulfill.

User experience and task design

Intentionally design the whole user experience from the beginning of the project, which includes the interactions of the organization, the material, and the volunteers. Best practices guides suggest keeping the tasks simple; for example Zooniverse recommends designing a task so that a ten year old could do them, and this was reinforced by the outcomes of our own pilot. We originally started out with the project that included 12 decision points. We quickly heard feedback that this was too many points, and people were not following the detailed instructions we had created. We quickly changed the workflow to have tasks that were much more simplified and this

helped not only engagement and completion but also had the benefit of giving us data that was simpler to manage in the end.

Community

Identifying and understanding your target community is important to the success of any project that seeks to engage with people. This is especially true of crowdsourcing projects where you are asking the public to engage in work for you. A smaller or more specialized community means you will be able to better target communication to volunteers and help keep them motivated. With a larger community base, your projects may be completed quicker, but it could be more challenging to connect with its members. We also found it was important to keep in mind to plan for the unexpected. The community that connects with your project might not be the one you expect, and that could be a big benefit in the engagement exercise and open up different exposure for your collections.

After you have identified and engaged with your community, you will need to continue to cultivate it well after the launch of the project. Plan to have regular communication with the public to encourage participation and be ready to engage with and work with the community. Depending on your project, the community can have valuable insight into the collection, especially if description or identification are the primary tasks. Cultivating community will not only help engage with the project, but also broadly impact your collections and organization. Volunteers that feel valued and useful are more likely to participate.

Resources

One of the common myths we encountered when first talking about crowdsourcing was that it was a great way to complete large amounts of repetitive work like transcription or description. While it is true that these projects directly utilize volunteers instead of paid staff, they require just as much institutional resources to ensure they are successful and managed well. When planning for a project, you should budget for dedicated hours to manage and respond to the community for the entire duration of the project. This includes directing the communication plan and the launch of the project, but also someone to continue to monitor and regularly engage with the community to maintain interest.

A second key resource to budget for is technical expertise. This could be in the form of custom programming to build your project to specification, but even if you plan to use one of the out of the box platforms, ensuring you have someone available to clean up and manipulate the data and import it into the destination system is important. Understanding what state you will receive the data in is important to project planning in order to ensure the data is useful.

Further reading

A number of resources are available online that address crowdsourcing best practices to ensure you have a successful project. McKinley (2016) created an accessible set of

21 principles to guide any successful crowdsourcing project. If you are more interested in learning about best practices from many different perspectives, the Crowd Consortium has published a complete report on their website, including videos and written notes of the conference *Engaging the Public: Best Practices for Crowdsourcing Across the Disciplines* held at the University of Maryland in 2015. Finally, Zooniverse they keeps a well-researched set of best practices up to date on their website. Even if you opt not to use their platform, these practices can be easily applied to other projects.

Conclusion

Crowdsourcing is an attractive and powerful means to engage with the public, and we can learn a lot about our collections from the crowd. Over the past decade, we have seen an increase in both number and types of crowdsourcing projects in the library, archives, and museum fields. We can learn a lot about best practices and practical applications, but the sheer volume of studies can be daunting. By moving forward with a small case study, the McGill Library was able to look at practical concerns and best practices. By combining a literature review, a crowdsourcing project, and platform survey with our case study we hope to create a more accessible entry point into how a library might work towards incorporating crowdsourcing at their institution. We still sit in on the occasional meeting and hear the phrase, "We'll just crowdsource it," but we now have more information from this research to give an informed answer on when it would make sense to realistically engage in a project and what kind of effort and resources are needed to make it successful.

References

Alam, S. L., & Campbell, J. (2017). Temporal motivations of volunteers to participate in cultural crowdsourcing work. *Information Systems Research*, *28*(4), 744–759.

Ayes, M. (2013). 'Singing for their supper': Trove, Australian newspapers, and the crowd. Paper presented at: IFLA WLIC 2013 - Singapore - Future Libraries: Infinite Possibilities in Session 153 - Newspapers.

Banks, M. (2013). Exploring the Zooniverse. *Physics World*, *26*(10), 65.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving. *Convergence: The International Journal of Research into New Media Technologies*, *14*(1), 75–90.

Causer, T. & Terras, M. (2014). 'Many hands make light work. Many hands together make merry work': Transcribe Bentham and crowdsourcing manuscript collections. In M. Ridge (Eds.), *Crowdsourcing Our Cultural Heritage* (57-88). Farnham: Ashgate.

Causer, T., Tonra, J., & Wallace, V. (2012). Transcription maximized; expense minimized? Crowdsourcing and editing *The Collected Works of Jeremy Bentham*. *Literary and Linguistic Computing*, *27*(2), 119–137.

Chrons O, Sundell S. (2011). Digitalkoot: Making old archives accessible using crowdsourcing. *AAAI Workshop Tech. Rep. AAAI Workshop - Technical Report*, *WS-11-11*, 20–25.

Cox, J., Oh, E. Y., Simmons, B., Lintott, C., Masters, K., Greenhill, A., … Holmes, K. (2015). Defining and measuring success in online citizen science: A case study of Zooniverse projects. *Computing in Science & Engineering*, *17*(4), 28–41.

Data Rescue: Archives and Weather. (n.d.).

Dunn, S., & Hedges, M. (2013). Crowd-sourcing as a component of humanities research infrastructures. *International Journal of Humanities and Arts Computing*, *7*(1–2), 147–169.

Durkin, K. (2016, June 21). Decoding the Civil War.

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science, 38*(2), 189–200.

Falconer, E. (2007, August 23). Increase motivation.

Feder, T. (2016, January). Crowdsourcing platform gets results. *Physics Today*, *69*(1), 25–27.

FromThePage. (n.d.).

Geiger, D., & Schader, M. (2014). Personalized task recommendation in crowdsourcing information systems - Current state of the art. *Decision Support Systems*, *65*, 3–16.

Grayson, R. (2016). A life in the trenches? The use of operation war diary and crowdsourcing methods to provide an understanding of the British army's day-to-day life on the Western Front. *British Journal for Military History*, *2*(2).

Greenhill, A., Holmes, K., Woodcock, J., Lintott, C., Simmons, B. D., Graham, G., … Masters, K. (2016). Playing with science: Exploring how game activity motivates users participation on an online citizen science platform. *Aslib Journal of Information Management*, *68*(3), 306–325.

Hedges, M., & Dunn, S. (2012). Crowd-sourcing in the humanities: A scoping study. Arts and Humanities Research Council.

Hedges, M., & Dunn, S. (2017). *Academic crowdsourcing in the humanities: crowds, communities, and co-production.* Cambridge, MA: Elsevier Science, Chandos Publishing.

Holley, R. (2010). Crowdsourcing: How and why should libraries do it? *D-Lib Magazine, 16*(3/4).

Hossain, M. (2012). Users' motivation to participate in online crowdsourcing platforms. In *2012 International Conference on Innovation, Management and Technology Research* (pp. 310–315). Malacca, Malaysia.

Howe, J. (2006). The rise of crowdsourcing. *Wired Magazine, 6*(14), 1–5.

Howe, J. (2009). *Crowdsourcing: Why the power of the crowd is driving the future of business.* New York: Three Rivers Press.

Lascarides and Vershbow. (2014) What's on the menu? Crowdsourcing at the New York Public Library In M. Ridge (Eds.), *Crowdsourcing Our Cultural Heritage* (113-138). Farnham: Ashgate.

Newswire. (2018, April 18). Library and Archives Canada launches Co-Lab.

McKinley, D. (2012). *Practical management strategies for crowdsourcing in libraries, archives and museums.* Wellington: School of Information Management, Victoria University of Wellington.

McKinley, D. (2016, July). Design principles for crowdsourcing cultural heritage.

Metadata Games, (n.d.).

Moirez, P. (2017). Construire et enrichir ensemble les savoirs: crowdsourcing et patrimoine numérisé. In *Communs du savoir et bibliothèques* (pp. 129–139). Paris: Le Cercle de la Librairie.

Moirez, P., Moreux, J.-P. & Josse, I.(2013). *Etat de l'art en matière de crowdsourcing dans les bibliothèques numériques.* Paris: Bibliothèque nationale de France.

Morschheuser, B., Hamari, J., Koivisto, J., & Maedche, A. (2017). Gamified crowdsourcing: Conceptualization, literature review, and future agenda. *International Journal of Human-Computer Studies, 106*, 26–43.

Moyle, M., Tonra, J., & Wallace, V. (2011). Manuscript transcription by crowdsourcing: Transcribe Bentham. *Liber Quarterly, 20*(3/4), 347–356.

Muñoz, T. (2013a). What IS on the menu.

Muñoz, T. (2013b). Refining the problem.

Muñoz, T. (2014). Borrow a cup of sugar? Or your data analysis tools?

Oomen, J., & Aroyo, L. (2011). Crowdsourcing in the cultural heritage domain: Opportunities and challenges. In *C&T '11 Proceedings of the 5th International Conference on Communities and Technologies* (pp. 138–149). Brisbane, Australia.

Owens, T. (2012, March 10). Crowdsourcing cultural heritage: The objectives are upside down.

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., … de Vreede, G.-J. (2013). Conceptual foundations of crowdsourcing: A review of IS research. In *46th Hawaii International Conference on System Sciences* (pp. 579–588). Hawaii: IEEE Computer Society.

Reed, J., Raddick, M. J., Lardner, A., & Carney, K. (2013). An exploratory factor analysis of motivations for participating in Zooniverse, a collection of virtual citizen science projects. In *2013 46th Hawaii International Conference on System Sciences* (pp. 610–619).

Ridge, M. (2014). *Crowdsourcing our cultural heritage*. Farnham: Ashgate.

Simpson, R., Page, K. R., & De Roure, D. (2014). Zooniverse: Observing the world's largest citizen science platform. In *Proceedings of the 23rd International Conference on World Wide Web* (pp. 1049–1054). New York, NY, USA: ACM.

Smith, D. A. (2008). From Nunavut to Micronesia: Feedback and description, visual repatriation and online photographs of Indigenous peoples. *Partnership: The Canadian Journal of Library and Information Practice and Research*, *3*(1).

Transcribe Royal BC Museum. (n.d.).

Zooniverse. (2014, February 19). Snapshot Serengeti brings authentic research into undergraduate courses.

Zooniverse. (n.d.). How to create a project with our Project Builder.

Zooniverse. (n.d.). Building, launching, and running your Zooniverse project: Best practices for engagement & success.