

# Social Media and Crowdsourced Transcription of Historical Materials at the Smithsonian Institution: Methods for Strengthening Community Engagement and Its Tie to Transcription Output

Lesley Parilla and Meghan Ferriter

## ABSTRACT

Crowdsourced transcription has grown in popularity as a tool for generating transcribed data and public engagement. This method of making digitized materials available on online platforms designed for volunteers to transcribe content works particularly well with science and historical materials. A well-designed site can offer volunteers a chance to interact with collections while providing the cultural institution with a new access point for researchers in the form of searchable text; a well-designed program of engagement can support sustained activity and unexpected positive outcomes. Many questions remain about how best to engage the public and the quality of resulting transcription. Many institutions design their sites to provide carefully structured experiences for volunteers. These projects organize materials around a research goal or subject and often provide detailed templates for the transcription. By fashioning a highly structured experience, are we fully engaging volunteers to interact with the materials? What happens if an institution creates an online environment that allows volunteers more choices and control? Would this affect the online community and transcription output? And what would be the impact of a structured engagement?

© Lesley Parilla and Meghan Ferriter. 

## KEY WORDS

Crowdsourcing, Transcription, Volunteer, Online engagement, Citizen science, Social media, Life-long learning, Outreach, Museums

The Smithsonian Institution is a diverse complex of nineteen museums and nine research centers located across the Americas, including Panama. With more than 138 million objects and 153,000 cubic feet of archival materials, the Smithsonian has many stories tucked in its collections. Digitization is creating increased access to these materials. In support of wider digitization efforts at these units and to make an online environment that can support contributions from any of them, Smithsonian's Office of the Chief Information Officer (OCIO) created the Transcription Center in June 2013. Designed to be a flexible, dispersed environment to manage transcription activity, the Transcription Center is a Web application and website that can support a wide variety of digitized materials and make them available online for volunteers to transcribe.<sup>1</sup> These materials include natural sciences specimens, photographs, correspondence, and diaries, just to name a few. All the contributed projects are available and easily browsed from within the same interface on the same site.

The Field Book Project was one of the first contributors to the Transcription Center. The project is a joint effort of Smithsonian Institution Archives, Smithsonian Libraries, and the National Museum of Natural History. It set out to identify, describe, and digitize natural sciences field documentation. Field books are primary source documents describing the events leading up to and including the collection of specimens. They provide rich data for researchers to understand how biodiversity has changed over time and space. The data can be difficult for researchers to discover and utilize due to their varying levels of description and diverse formats.<sup>2</sup> As collection items have been digitized, more questions about discovery and research "within" these items have emerged. One solution to increase access and improve the data associated with digitized collections is to crowdsource transcription.

## Meeting Motivations and Project Goals

How might we achieve the goal of transcribing materials to improve access to and engagement with collections? We suggest engaging the help and contributions of the public. And if we invite the public, why might they want to volunteer their time to transcribe? To begin to answer these questions in the context of Smithsonian Institution collections, we need to know more about participation and motivation in crowdsourcing cultural heritage and citizen science.<sup>3</sup>

Once we understand the models and motivations, we will consider what can be learned from the ways archives, libraries, and museums engage their audiences online. From there, we will consider the ways these institutions use social media to support this online engagement. Connecting these spheres and best practices will also ground the analysis of the ways the Transcription Center embraces contemporary challenges through collaborative transcription.

Any participatory project sponsored by archives, museums, and other cultural heritage organizations must balance its goals with the appeals to the public who will contribute to achieving those goals. Understanding the motivations that compel participation offers useful starting points from which to design a project and to maintain an engaged audience. Online and crowdsourcing projects focusing on tangible outcomes have explored participation and motivation.<sup>4</sup> Several studies have also emphasized the importance of project design and the ways it must coordinate with the goals of the project.<sup>5</sup> We still need to understand how to leverage project design to create tangible outcomes, as well as the intangible scaffolding that sustains the connection between participants, goals, and the organization.

Many of the existing studies of volunteer motivations explore the facets of the four categories of motivation identified by C. Daniel Batson, Nadia Ahmed, and Jo-Ann Tsang.<sup>6</sup> These motivations are egoism, altruism, collectivism, and principlism. For example, Jason Reed et al. and Dana Rotman et al. explored these motivations in the context of citizen science.<sup>7</sup> Nicolas Kaufman et. al and Vickie Curtis expanded on these models to account for participation based on being part of a community and the promise of future returns.<sup>8</sup>

Civic projects, volunteering, and learning or educational projects supported by the activities of volunteers feature motivations with intrinsic and extrinsic qualities.<sup>9</sup> Intrinsic motivations are connected to the inherent satisfactions of “doing hobbies, enjoyment in learning, mastering new skills and practicing existing skills, recognition, community, and passion for the subject.”<sup>10</sup> As Richard Ryan and Edward Deci described, intrinsic motivation is pervasive and important to human activity.<sup>11</sup> Intrinsic motivation figures prominently in discussion of and appeals for participation in citizen science and crowdsourcing activities. These kinds of appeals may invite participants to “learn more,” “have fun,” or indulge an interest in the subject matter.

By contrast, extrinsic motivations relate to instrumental value of activity.<sup>12</sup> Extrinsic motivation may also relate to social incentives, such as parties outside the crowdsourcing project to recognize one’s crowdsourcing activity as valuable. Furthermore, several crowdsourcing approaches incorporate explicit competition to meet extrinsic motivations; Mary Flanagan et al. detailed incorporating overlapping motivations in the design of Metadata Games.<sup>13</sup>

C. Daniel Batson et al. also proposed altruism and principlism as volunteer motivators.<sup>14</sup> Altruistic motivations are based on a desire to improve the welfare of others. Principlism includes the motivation of upholding one or more principles held dear.<sup>15</sup> The Transcription Center brings together cultural heritage and citizen science in the same space, and motivations to participate in these types of projects have some overlap. Motivations for participating in citizen science projects relate to the desire to contribute to the specific research goals

of the project, background interest in the subject, or existing experience with the subject.<sup>16</sup> Motivations for participating in crowdsourcing cultural heritage projects often begin as intrinsic and altruistic, remaining so throughout the contribution. These motivations may include an interest in the subject matter, a desire to learn, and a desire to help the institution with wider goals.<sup>17</sup> Indeed, interest in or passion for the subject matter and the opportunity to contribute to a project's goals function as leading motivations in cultural heritage and citizen science alike.<sup>18</sup>

After reviewing the motivations of the potential audience, the next step is to connect these motivations to the project goals with a specific set of engagement principles. No single roadmap exists—indeed, for Tim Causer and Valerie Wallace, the mark of a successful crowdsourced project is the ability to accommodate and address multiple motivations at the same time.<sup>19</sup> Open communication with participants about the development process is critical to maintaining trust in these cases and also to empowering participants to take ownership in guiding the growth of the project.<sup>20</sup>

Furthermore, understanding the motivations of participants is particularly important as many citizen science and crowdsourcing projects have a low retention rate.<sup>21</sup> Crafting specific engagement experiences to increase retention or, alternatively, to generate excitement and higher participation within the volunteer life cycle, is crucial.

Despite many valuable typological surveys and analyses of initial volunteer motivations, most studies “do not go further to offer a practical model for how such a project may be developed and implemented to ensure it is viable, successful, and sustainable.”<sup>22</sup> What remains underexplored are “factors affecting efficiency, accuracy, initial motivation, and long-term engagement” in transcription- and annotation-based projects.<sup>23</sup> Explorations of the ways motivations change, evolve, and relate through the life cycle of participation are also needed.<sup>24</sup> We hope our exploration will be added to the examples and ongoing analysis of crowdsourced transcription in cultural heritage and archival projects.<sup>25</sup>

As Rose Holley suggested, “If libraries want to stay relevant and valued, offer high quality data and continue to have a significant social impact they must develop active engagement strategies and harness crowdsourcing techniques and partnerships to enhance their services.”<sup>26</sup> Archives and museums face similar challenges with high stakes.<sup>27</sup> Opportunities are emerging to bring together engagement techniques, to extend social engagement through crowdsourcing, and to connect with larger goals of delivering service and increasing knowledge and access to it.

The Transcription Center was designed to accommodate many types of archival materials, museum collections, and more. Additionally, the Field Book Project materials are very diverse; they reflect the complexity of physical

arrangement and the diversity of Smithsonian Institution activities. With this in mind, it was necessary to create a “flexible” tool that facilitates a coherent experience with Smithsonian collections, accompanied by many concurrent, but specific, engagement approaches.

Any public-facing project may encounter hurdles that may be overcome with planning and iterative approaches. Some issues common to transcription and crowdsourcing projects include recruiting participants, assuring transcription quality, and balancing the volume of collections or projects and the quantity of resulting data. In this article, we will discuss one strategy for jumping the hurdle of sustained engagement: issuing “challenges”—time-bound, goal-based events—to volunteers.

The Transcription Center features a unique space where all projects can be discovered; its design allows volunteers to move flexibly between projects, using as few clicks as possible. The Transcription Center coordinator introduces structured experiences in the form of page review challenges and coordinated social media campaigns. We believe that this combination of approaches keeps the community growing and sustains the online community as a whole, as volunteers eventually cycle out of active participation. This article examines how the Transcription Center has worked with Smithsonian Institution content and Smithsonian’s units to create a flexible experience, combining structured experiences with a flexible site structure, which meets many motivations and sustains engagement. It then discusses lessons learned through the Transcription Center by investigating its methods and the relationship between outreach and output through materials contributed by the Smithsonian Field Book Project.

### **The Smithsonian Transcription Center: A Malleable Structure for Creating Data**

The Transcription Center is a website and program of engagement for the public. It also provides a service for museum units and researchers by inviting the public to transcribe digitized collections. It was created to be flexible and extensible, utilizing a Drupal module that can accommodate many material types based on its tight connection to and capacity to leverage existing Smithsonian Institution collection and information management systems. The Transcription Center, for example, hosts images organized in the digital asset management system (DAMS) and delivered by the image delivery service (IDS). The completed and reviewed transcribed text is then indexed, connecting the text and images via the item’s catalog or collection record in Smithsonian Institution’s central Collections Search Center. The result is text that can be indexed for searching, which makes the content of digitized collections more readily discoverable by the public and researchers alike. Thus, the Transcription

Center is one of a wider set of services and information technologies supporting collections management at the Smithsonian Institution, and it helps build access through existing infrastructure.

The Transcription Center can present numerous paths to volunteers when they work within the website interface. A volunteer can browse projects from several locations on the Web page—the homepage banner, “Latest Projects Updates,” featured projects, or a dropdown menu of represented archives and museums. Registered volunteers can return to projects to which they have contributed using their “My Work” page.



FIGURE 1. This is a screenshot of the Smithsonian Transcription Center website. The homepage banner offers several ways to navigate to projects.

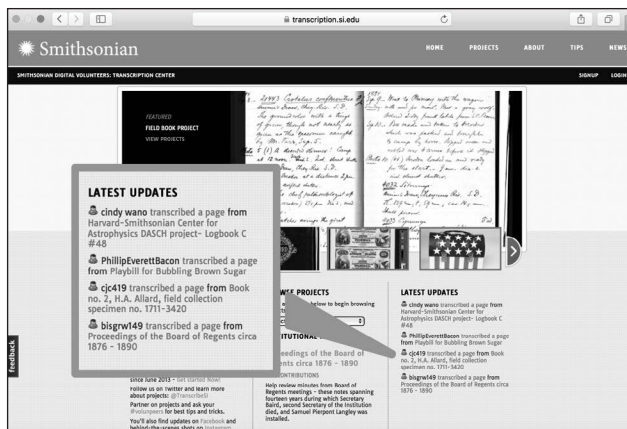
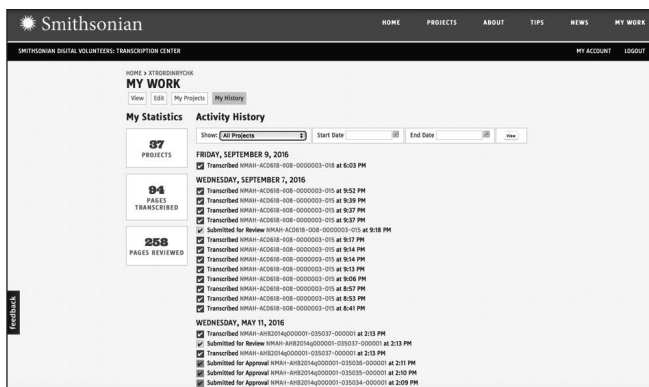


FIGURE 2. This is a close-up view of “latest project updates” as seen on the Smithsonian Transcription Center website.



**FIGURE 3.** This is a screenshot of a “my work” page on the Smithsonian Transcription Center website. This page enables volunteers to track their own work on projects.



**FIGURE 4.** This is a screenshot of a project page on the Smithsonian Transcription Center website. Volunteers can see work progress regarding transcription and review for each image as well as the project as a whole.

Typically, a project in the Transcription Center is a digital facsimile of a physical object or archival arrangement, for example, a folder of correspondence, a diary, a logbook, or a drawer of bumblebees. The project consists of images of each page or object in that arrangement (e.g., 132 pages of correspondence from a folder or 500 botanical specimen sheets). Each image is known as an *asset*. When a volunteer selects a project, the project page displays a project description and thumbnails of all assets. A progress bar indicates the status of the project as a whole and each asset; status points include whether assets need transcription, review, or are complete.

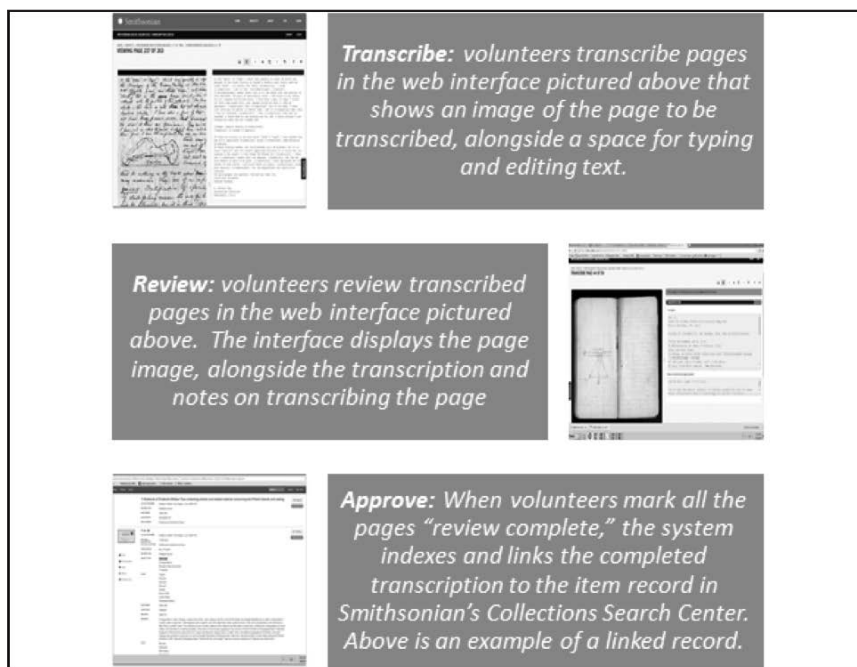
When transcribing in the Web interface, a volunteer sees the asset image on the left and a transcription window or fields on the right for entering the transcribed data. The project-specific instructions remain accessible via a button above the transcription window. The asset image can be rotated, reset, and zoomed in for closer reading.



The Transcription Center uses a three-step, peer-review process for transcription workflow: transcribe-review-approve. Anyone, anonymously or with an account, may transcribe; registered volunteers with accounts may transcribe and review. When performing these tasks, volunteers will see different options to save, edit, or complete pages; they also receive messages of gratitude for their contributions and, so their time is not wasted, a caution when another volunteer is already transcribing a page. Smithsonian staff approve pages that have been fully transcribed and reviewed. Once volunteers completely review all pages of an item, it is made available as a PDF from the Transcription Center website and made searchable on Smithsonian Collections Search Center (CSC).<sup>28</sup> CSC is the online search platform that makes available records from the Smithsonian's units. Currently, over 9.3 million records populate the search platform for museum objects, archives, and library materials, including more than 1.4 million online media files.

### The Transcription Center Three-Step, Peer-Review Process: Transcribe-Review-Approve

Transcriptions completely reviewed on the Transcription Center are made text-searchable for the first time, improving the collections information. Volunteers who share their time in what they consider a flexible, collaborative



**FIGURE 5.** This image highlights the three-step process used at the Smithsonian Transcription Center.



experience make this level of accessibility possible. Since the time of the beta launch of the Transcription Center in June 2013, 6,288 volunteers have completely transcribed and reviewed 165,520 pages of field notes, diaries, logbooks, biodiversity specimen labels, ledgers, photo albums, manuscripts, and more.

We describe the Transcription Center's structure as flexible because it accommodates these many collections in the same interface and allows volunteers to move between pages and projects. Compare, for example, other transcription platforms like *Old Weather*,<sup>29</sup> which presents a clearly delineated topic theme with a highly structured interface; this design can compel continued activity at the cost of movement through adjacent projects. Volunteers state that they like the open experience of the Transcription Center, which does not restrict them to one type of material; the flexible nature of the website encourages "wandering" and may contribute to the long visits. Volunteers compare the online experience to a "choose your own adventure."

Communication with volunteers is critical to the success of this model—to gather, to inform, to encourage, and to support participants. The Transcription Center project coordinator reaches out to volunteers through registration emails, responding directly to feedback contact form entries, via social media, and by campaign email. The Transcription Center is represented on social networking platforms that include Twitter, Facebook, Tumblr, Instagram, and YouTube. Some social media messages are "blasts" of information regarding projects; other means of communicating are tailored, sharing specific messages with groups of volunteers who have participated in select projects or activities on the Transcription Center. Within this framework of communication, the types of outreach detailed below can boost activity through focused activity at specific times. Furthermore, ongoing promotion of transcription opportunities by the archives and museum groups who steward the collections being transcribed augment this communication. This combined approach contrasts with traditional recruitment strategies in that it is not a single call to action, but rather an open, ongoing invitation to join in—moving toward dialogue with participants.

### The Field Book Project

The Smithsonian's Field Book Project was an early proponent of the Transcription Center. Smithsonian staff endeavor to reconnect field book content with the specimens it documents as well as with research and any resulting publications. Reestablishing this connection can be achieved through digital approaches; the impetus for inclusion in the Transcription Center is to make the field books text-searchable to assist researchers and the public with exploring historical science.

The Transcription Center has become an important part of making field book content accessible to researchers, as these materials are often difficult to read because of the handwriting. Diaries, maps, and photograph captions are often handwritten by authors who never expected someone else to read them. When the Field Book Project started, it focused on identifying and cataloging the materials dispersed geographically throughout the institution. Once identified and cataloged, the project developed the goal of creating a prototype workflow for scanning the materials. Once materials were scanned, the project looked for collaborations within and outside the Smithsonian Institution to make the digitized materials available as widely as possible. Collaborations are essential because the project is small, but also more plausible because of the relationships the Smithsonian Institution already has with cultural institutions around the world. The project makes materials available through online platforms like the Internet Archive and the Biodiversity Heritage Library, and it connects with audiences through social media platforms like Twitter, Typepad, and Flickr.

Collaboration with the Transcription Center offers unique benefits for the project. While it makes the field books available to new audiences, it also increases the ways they can be searched. Before the Transcription Center, researchers had two choices. A researcher could search a field book catalog record to determine if content was pertinent to his or her research. These catalog records are much like library catalog records, comprised of title, summary, and controlled subject headings. Researchers could also read the digitized field notes, though handwriting might be difficult to decipher. Now, volunteers create text that is indexed, allowing a researcher to search the text itself. From November 2013 to January 2016, volunteers transcribed 189 field books, more than 25 percent of those digitized. Additionally, completed Transcription Center projects like field notes—those that have been completely transcribed and reviewed—may be downloaded as PDFs by the public for free.

Collaboration with the Transcription Center also enables the Field Book Project to leverage existing social media for greater engagement with the field notes. One of the most powerful ways to explain the breadth and depth of field book content has been by highlighting stories found in them. Volunteers often respond to the same type of content that staff do: stories of exploration, challenges, and discovery mixed with personal anecdotes and delights. In some cases, there are preexisting blogs or Flickr images. Volunteers spend significantly more time with individual items than staff do and are able to investigate and expound on the information they find and share through social media. The stories thus gain a renewed immediacy.

## Weaving Together Data and Outreach with the Transcription Center

As evidenced by the Field Book Project's experience with the Transcription Center, crowdsourced transcription is an important tool for transcription output and public engagement in science. Transcription makes archival content text-searchable, which is imperative if contemporary researchers are to discover the information it contains. Yet, there are frequent questions about quality of output when engaging the public in crowdsourcing.

Indeed, quality is an important consideration. Many institutions may opt to pursue optical character recognition (OCR) as an alternative to taking on crowdsourced transcription. Currently, this is only a proven option for typed documentation. OCR software has an accuracy rate as low as 80 percent. Additionally, paid transcription services may not be an option because of budgetary constraints.

The Transcription Center has found that the quality of the data output correlates with the quality of engagement. Through the kinds of campaigns discussed below, volunteers recognize the importance of the work they are doing. Individual communications with volunteers support these findings (e.g., why they take part in the Transcription Center). These motivations result in a high level of transcription accuracy. A survey of 80 pages selected at random from 8 different transcribed documents demonstrated that only 8 pages out of 80 had errors, or a total of 8 errors for an estimated 5,803 words. Even including the number of guesses at words (22), that still presents an accuracy rate of more than 99%, which is comparable to paid transcription services.

Transcription is a powerful way to encourage the public to interact with museum collections. The number of visits to the Transcription Center site and the average length of these visits suggests vivid engagement with collections. Volunteers spend more time with and may therefore gain a deeper understanding of museum resources via their activity in the Transcription Center than they might through traditional online offerings like exhibits and social media outlets such as blogs or Flickr. When comparing Transcription Center visit statistics with more established methods of social engagement like blogs produced by the National Museum of Natural History (NMNH), the difference is significant. We compared two times (January 11–25, 2015, and March 8–22, 2015) when the Transcription Center maintained a typical level of social engagement to NMNH's main blog and its other blogs combined. The rates of visitation are comparable, but the lengths of visits vary greatly. Average time per visit on the Transcription Center was 17 minutes. Time spent on the NMNH blog was less than a minute (see Tables 1 and 2).

These data present a picture of engagement with archival and museum collections. They also suggest that asking volunteers to participate in citizen

science encourages them to stay longer and to develop a connection with the materials. Volunteers stay with the projects because of this connection, the information they discover, and the feelings of success they experience when completing pages and projects. Transcription becomes a powerful way to engage the public by helping to make collections more searchable for researchers, by offering them a chance to work directly with museum resources, and by asking them to invest time and effort.

**Table 1. Comparative Statistics for the Transcription Center, January 11–25, 2015<sup>30</sup>**

| Sites for January 11–25, 2015     | Transcription Center | NMNH Blog | All Other NMNH Blogs |
|-----------------------------------|----------------------|-----------|----------------------|
| Number of visits                  | 3,700                | 2,604     | 10,812               |
| Average length of visit (minutes) | 17:27                | 00:26     | 00:43                |

**Table 2. Comparative Statistics for the Transcription Center, March 8–22, 2015**

| Sites for March 8–22, 2015        | Transcription Center | NMNH Blog | All Other NMNH Blogs |
|-----------------------------------|----------------------|-----------|----------------------|
| Number of visits                  | 2,632                | 1,555     | 5,832                |
| Average length of visit (minutes) | 17:04                | 00:34     | 00:55                |

Based on what we can see about the length of visits to the Transcription Center and what can be inferred about the depth of engagement, what can be done to use social media to increase interaction—and therefore ensure high-quality transcription?

### **Beyond “Build It and They Will Come”: Challenges and Campaigns Are the Guided Tours that Encourage Volunteers to Stay**

The Transcription Center began blending structured experiences with the flexible nature of the website shortly after launch. Volunteers were transcribing quickly, but the transcription coordinator saw a lag in the number of pages reviewed by volunteers. The coordinator began crafting structured experiences in the form of week-long review challenges to encourage an increase in volunteer reviews of transcribed pages. She based these on what was learned by observing activity during regular volunteer engagement.

The first review challenges were issued to volunteers via emails, posts on Tumblr and Facebook, and interactions on Twitter. The success of the review challenges inspired the coordinated social media campaigns. For the coordinated social media campaigns, the coordinator increased and targeted communication with volunteers by making blog posts across the Smithsonian departments that corresponded to the campaign theme (the Transcription Center does not

have a blog); sending emails and tweets, making Tumblr posts, and offering rewards in the form of virtual contact with staff members who revealed “behind the scenes” details about the materials in the campaign.

The Transcription Center continues to use both types of structured experiences (review challenges and social media campaigns) to catalyze volunteers’ action on particular tasks. Why use challenges at all? While volunteers can always choose whether they would like to transcribe, review, or do both, challenges promote collaborative activity around the same focus. Additionally, volunteers report feelings of cohesion from these brief “collaborative competitions.”

To understand volunteer activity during structured engagement experiences, we analyzed statistics during two-week periods. A two-week period accounts for other activity on the website during this time and provides insight into the tail effect of a campaign. Larger Transcription Center site statistics, which are pulled on a weekly basis, also support the selected periods.

### Structured Engagement: Challenges

Challenges are specific, time-bound tasks that focus activity on a direct request. Seven-Day Review Challenges encourage volunteers to work together to review and, therefore, complete projects. Completely transcribing and reviewing projects is important because the text can only be indexed once the entire project (all of its pages) has gone through the peer review process. There are now two types of challenges: review challenges and “Contribute and Connect.”

Two challenges described below, the April 2015 Seven-Day Review Challenge and WeDigBio in October 2015, asked volunteers to transcribe or review “as many pages as possible.” For example, in a Seven-Day Review Challenge, the total pages completed each day are tallied as a cumulative benchmark and shared with participants. This type of targeted event is meant to encourage focused drives around activities with the goal of finishing projects.

In a Contribute and Connect challenge, volunteers are asked to complete a specific set of projects for a behind-the-scenes reward. The collaborative campaigns discussed in the next sections evolved from the Contribute and Connect approach; they featured a smaller, specific set of materials and a designated time frame, and were rewarded with behind-the-scenes discussions with Fossil Marine Mammals curator Dr. Nicholas Pyenson.

These challenges result in a significant increase in visits to the site, in pages transcribed, in active users, and in replies to the Transcription Center Twitter account from the public (see Table 3). Replies are significant because they represent responses from volunteers to the Transcription Center project coordinator and are thus a strong indication of volunteer engagement. In the April 2015 challenge, there was a 28% increase in visits to the site (see Table

3). During the October 2015 WeDigBio challenge, we saw a similar increase of 32% (see Table 4). Both campaigns seemed to offer residual benefit, as numbers remained higher than before the challenges. Unexpectedly, the length of visit fell in both cases.

**Table 3. Seven-Day Review Challenge, April 2015<sup>31</sup>**

| Two-Week Periods                  | Before (March 9–22, 2015) | Seven-Day Review Challenge (April 6–19, 2015) | After (May 4–17, 2015) |
|-----------------------------------|---------------------------|---|------------------------|
| Visits to site                    | 2,443                     | 3,391   | 2,793                  |
| Average length of visit (minutes) | 17:26                     | 16:37   | 14:11                  |
| Pages transcribed                 | 3,805                     | 6,213   | 3,943                  |
| Active users                      | 214                       | 331   | 195                    |
| Replies                           | 73                        | 151   | 164                    |

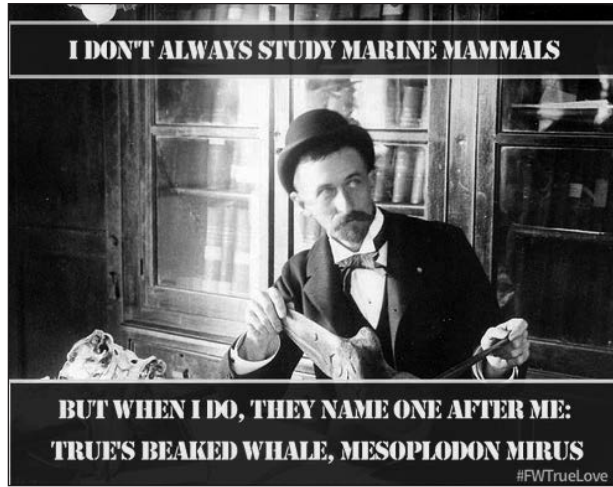
**Table 4. WeDigBio Challenge, October 2015**

| Two-Week Periods                  | Before (September 21–October 14, 2015) | WeDigBio (October 19–November 11, 2015) | After (November 16–29, 2015) |
|-----------------------------------|--|---|------------------------------|
| Visits to site                    | 2,526                                  | 3,673                                   | 3,994                        |
| Average length of visit (minutes) | 15:28                                  | 17:28                                   | 13:44                        |
| Pages transcribed                 | 5,065                                  | 8,186                                   | 4,856                        |
| Active users                      | 171                                    | 297                                     | 256                          |
| Replies                           | 26                                     | 160                                     | 77                           |

## Structured Engagement: Coordinated Social Media Campaigns

Coordinated social media campaigns developed as unit staff became more actively engaged with Transcription Center volunteers. These campaigns include making blog posts across the Smithsonian departments that correspond to the campaign theme; increasing and targeting communications from the Transcription Center coordinator in the form of emails, tweets, and Tumblr posts; and rewarding participants with online contact with a staff member to provide “behind-the-scenes” details about the materials in the campaign.

The #FWTrueLove campaign developed around the field notes of Frederick William True, who worked at the Smithsonian from 1881 until 1914, in the field of paleontology. Staff from the Smithsonian Libraries, the Department of Paleobiology, and the Smithsonian Institution Archives worked with the Transcription Center coordinator during the campaign.



**FIGURE 6.** This image was used for the February 2015 #FWTrueLove Campaign.

Comparing three two-week periods, we continued to see increases in engagement elements similar to the changes seen during the campaign. Pages transcribed, active volunteers, and communication on other social media platforms increased. Though the length of an average visit dropped, the number of active users rose by 14%, and the number of pages transcribed increased by more than 10%. Though the length of visits fell, the total amount of time increased from 1,067 hours to 1,159 hours, an increase of 9% (see Table 5).

**Table 5. #FWTrueLove Campaign, February 2015**

| Two-Week Periods                  | Before (January 5–18, 2015) | #FWTrueLove (February 9–22, 2015) | After (March 9–22, 2015) |
|-----------------------------------|-----------------------------|-----------------------------------|--------------------------|
| Visits to site                    | 3,580                       | 4,907                             | 2,443                    |
| Average length of visit (minutes) | 16:31                       | 13:32                             | 17:26                    |
| Pages transcribed                 | 1,467                       | 1,910                             | 1,903                    |
| Active volunteers                 | 129                         | 172                               | 107                      |
| Replies                           | 117                         | 136                               | 73                       |

Interaction increased most dramatically. “Replies” indicates responses from volunteers to the Transcription Center coordinator. These communications increased 47%. Other forms of social media engagement like retweets, favorites, and comments are difficult to compare, as the campaign’s designated hashtags are not used during noncampaign times. However, increased rates of retweets and favorites corresponded with social media, such as blog posts, project launches, and virtual tours.



FossilFossick was a collaborative campaign kick off to Fossil Week, built around the field notes of three paleontologists and paleobotanists, leading into the national campaign for #FossilStories. This campaign built on the earlier success of #FWTrueLove in February 2015 and was developed in coordination with the staff at the Biodiversity Heritage Library and the Smithsonian Department of Paleobiology, the Smithsonian Libraries, and the Institution Archives. During this campaign, we found similar behavior to #FWTruelove: the amount of time on-site dropped as pages transcribed increased. The number of active users also increased (see Table 6).

Unlike the previous campaign, #FossilFossick was immediately followed by another campaign #WeDigBio. A globally coordinated transcription event, WeDigBio drew attention to liberating data from digitized biodiversity specimens. The WeDigBio focus for the Transcription Center and campaign with the National Museum of Natural History focused on the remaining 12,000 bumblebees (of an original 44,414 total) with the goal of transcribing the specimen labels and improving the utility of the collection. Pages transcribed continued to rise as replies and number of active users rose slightly.

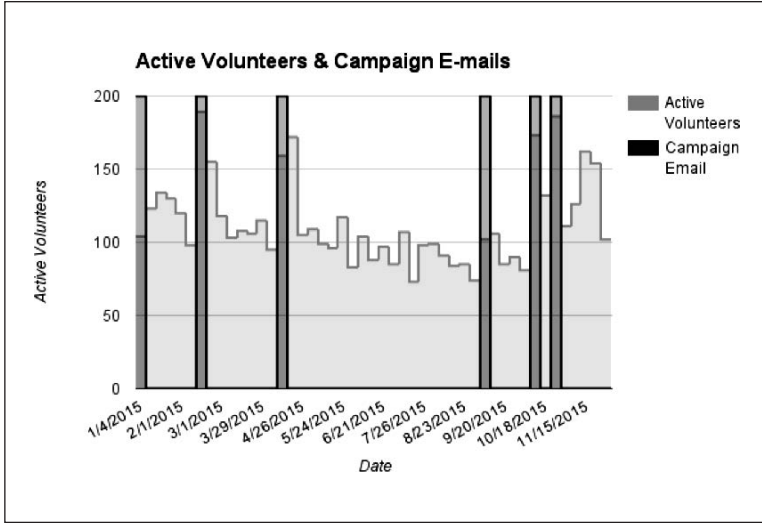
**Table 6. #FossilFossick Campaign, October 2015**

| Two-week period                   | Before (September 7–20, 2015) | #FossilFossick (October 5–18, 2015) | After (November 9–22, 2015) |
|-----------------------------------|-------------------------------|-------------------------------------|-----------------------------|
| Visits for site                   | 2,557                         | 3,655                               | 5,032                       |
| Average length of visit (minutes) | 18:31                         | 16:12                               | 13:14                       |
| Pages transcribed                 | 2,071                         | 2,261                               | 2,891                       |
| Active users                      | 96                            | 153                                 | 158                         |
| Replies                           | 63                            | 57                                  | 58                          |

### Effects on Long-Term Online Activity

Each time the Transcription Center conducted a structured experience, numbers increased and returned to a consistent level during “normal” engagement. The Transcription Center has been able to maintain consistent numbers of visitors and output due to this influx of new users and excitement, as seen in Figure 7, which illustrates online activity from January to November 2015. Campaign- and challenge-related communications create a consistent increase in activity. When communications dropped during May to August, activity slowly began to decline.

With each challenge and campaign, we saw consistent levels of interaction and output. Numbers increased during the challenge or campaign and returned to previous levels afterward. This has consistently occurred during the two years of the website’s life. We postulate that these campaigns are necessary to assure a new influx of volunteers as others leave.



**FIGURE 7.** This chart illustrates volunteer activity on-site and campaign activity by week, January to November 2015. The value assigned to campaign email is arbitrary (set at 200) to highlight the weeks when campaign emails were distributed.

**Table 7. Volunteer Transcription Activity Levels**

| Total Transcription Activities | Volunteers with This Level of Activity |
|--------------------------------|--|
| 500+                           | 3.07%                                  |
| 100-499                        | 7.76%                                  |
| 51-99                          | 7.31%                                  |
| 21-50                          | 14.31%                                 |
| 11-20                          | 15.17%                                 |
| 5-10                           | 19.40%                                 |
| 1-4                            | 32.98%                                 |

Extending the transcription life cycle of volunteers is an important task for sustaining activity and the robustness of the Transcription Center. As shown in Table 7, nearly 68% of volunteers have performed 20 or fewer transcription activities (transcribing, completing a page, or reviewing a page). Thirty percent of volunteers have participated in 4 or fewer.

### Lessons Learned

These structured experiences might appear to work in direct opposition to the flexible, customizable nature of the Transcription Center. When one looks at the life cycle of volunteer activity and the Transcription Center’s consistent online community growth over the last two years, these structured experiences

appear to sustain the strength of the community as a whole, bringing in new volunteers as others eventually cycle out of active participation. Challenges and coordinated social media campaigns increase volunteer interaction, transcription output, and the number of active users on the site. Average length of visit dips demonstrably during the campaign. We hypothesize this may be because new active users are specifically coming to spend time on the projects highlighted in the challenge or campaign. This varies from other periods of time on the site, when volunteers come to work on one project but move onto unrelated projects that catch their attention because of the site's open structure.

The flexible design of the website also offers important benefits for institutions and researchers. Highly structured transcription websites make it easy to promote a type of transcription project; however, a narrow theme developed around a format or subject would not work well with the Field Book Project. The Field Book Project focuses on exposing field documentation across disciplines and material formats. The Transcription Center structure assures that any of the field book materials can be contributed for transcription. This means that an entire collection can be transcribed, even if it includes many material formats.

The process of developing these campaigns highlights that the Transcription Center is a partnership with the volunteers, with both sides benefiting from invested time and effort. It can take a great deal of time and preparation to assure that content and staff communications foster the online community. Staff must also approach interactions with spontaneity and intensity. However, the intermittent nature of coordinated social media campaigns offers both staff and volunteers breaks from the intensity of the campaigns. Staff get a break, and volunteers have time to form their own ways of connecting with materials.

The campaigns may serve the vital purpose of introducing new volunteers to the online community and offering a sort of tour that demonstrates the types of experiences a volunteer could have in the Transcription Center. A campaign shows a specific method to answer multiple and shifting motivations by orienting volunteers to the benefits of taking part, providing opportunities to learn and work together with others, and highlighting what to expect in the standard, noncampaign experience.

Interacting with volunteers also means being prepared to capture enthusiasm. Transcription Center staff have learned that once a campaign is complete, volunteers may look for the next opportunity to build on what they have completed. This may be because they have developed a new interest in a scientist whose name came up in papers that were transcribed; they may have a desire to connect data with other knowledge repositories (e.g. Wikipedia, JSTOR Global Plants, or eBird).<sup>32</sup> Volunteers may also become interested a theme that transcends the campaign, such as Women in Science. Staff can offer volunteers the chance to continue the experience of transcribing, by preparing related

materials beforehand and releasing them as they fit volunteers' expressed interest. Staff may also offer interaction with subject matter experts as an evolution of campaign experience—whether virtually or in-person.

One of the most important lessons learned regards the benefits of asking volunteers to do something, but being careful not to ask them to do everything. We have learned to build on their interest. Volunteers are often willing to explore materials in ways that staff may not expect. In other cases, departments want transcription or other tasks completed that do not inspire volunteer interest. From the institutional perspective, all goals may not be met. However, the goals of the partnership are met—volunteers and the institution both benefit from the experience of sharing time and resources.

## Conclusion

At the beginning of this discussion, we posited several questions regarding the effectiveness of a highly structured experience and the effect on volunteer engagement. The Transcription Center's work over the last three years offers answers to some of these questions and glimpses of answers to others. Does a highly structured experience enable an institution to fully engage its volunteers to interact with the materials? The Transcription Center's experience over the last three years seems to indicate it does. The Smithsonian Institution can expect and actively encourage volunteers to engage with the materials by providing a loosely structured site that periodically offers structured experiences to guide volunteers, especially new ones, to discover content of interest. This flexible structure offers increased opportunities for Smithsonian units to contribute and enables volunteers to browse the breadth of content. They find their own connections and define their own journeys. Volunteer activity increases with more choices and control.

When structured experiences are periodically used in this environment, they can be powerful tools, especially for building an online community and increasing transcription output. The structured experiences can be used to increase communication with volunteers and thus provide a way for a site to allow them more of a voice. The Transcription Center's structured experiences increased volunteer communication with staff and enabled the Transcription Center to better identify and target volunteers' numerous motivations. This also reinvigorates the online community by bringing in new volunteers as others cycle out. A structured experience can act as an informal introduction to the Web interface, materials, and online community, demonstrating the kinds of experiences volunteers can create for themselves. By limiting the time and extent of the structured experience, the Transcription Center can tailor future campaigns and challenges to fit identified and evolving volunteer motivations.

The time limit also enables staff to set aside the time needed for increased communication with volunteers during a campaign or challenge.

The structured experiences intensify interactions and output. Volunteers feel a connection to the materials that translates to learning and a high degree of transcription accuracy. This could be seen in the increase in the number of replies from volunteers and pages transcribed during each structured experience discussed above.

It remains to be seen whether the Smithsonian's Transcription Center structure and approach can inform the activity of other institutions. The Smithsonian Transcription Center models useful tools for offering volunteers choices and for utilizing structured experiences to inspire dialogue and excitement around a shared goal. The website structure is designed to meet the needs of Smithsonian units and thus emulates the loose structure of the Smithsonian Institution. The system is also designed to support and to integrate with Smithsonian Institution information management technologies as part of a suite of tools and approaches for managing and creating access to diverse collections. It may be difficult to compare the Transcription Center's experience to other crowdsourcing projects in terms of format or scope, but it is important to add the lessons learned here to those of other crowdsourcing projects. It would be interesting to compare the Transcription Center's experiences to crowdsourced transcription projects that are more narrowly focused due to the collection holdings. We share these approaches with the hope that others can put them to work within their own organizational strategies.

## ABOUT THE AUTHORS

---



**Meghan Ferriter** is an interdisciplinary researcher who facilitates engagement and advises on workflow as Project Coordinator for the Smithsonian Transcription Center. She works closely with multiple museums, archives, volunteers, and partners to balance objectives of digital historical and citizen science projects. Her recent research explores information sharing and collaboration in digital platforms including Tumblr and Twitter, as well as crowdsourcing and citizen science best practices. Ferriter holds an MA in history from Old Dominion University and a PhD in sociology from the University of Glasgow.



**Lesley Parilla** is the cataloging coordinator for the Smithsonian Field Book Project, a joint initiative at the Smithsonian Institution Archives, Smithsonian Libraries, and the National Museum of Natural History to increase accessibility to field book content that documents natural history. She is the principal cataloger, and manages metadata standards, record contribution, and outreach through social media. As coordinator, she works closely with archives, library, and museum staff throughout the workflows for collection selection, processing, conservation, and digitization. She earned an MLIS from the University of Hawaii at Manoa, with a certificate in special collections and archives.

## NOTES

- <sup>1</sup> The Smithsonian Transcription Center can be accessed at <https://transcription.si.edu>
- <sup>2</sup> Sonoe Nakasone and Carolyn Sheffield, "Descriptive Metadata for Field Books: Methods and Practices of the Field Book Project," *D-Lib Magazine* 19, nos. 11–12 (2013), doi:10.1045/november2013-nakasone.
- <sup>3</sup> The Citizen Science Association defines citizen science as "the involvement of the public in scientific research – whether community-driven research or global investigations." Citizen Science Association. "The Power of Citizen Science". (2016). <http://citizenscience.org/>
- <sup>4</sup> Daren C. Brabham, "Moving the Crowd at iStockphoto: The Composition of the Crowd and Motivations for Participation in a Crowdsourcing Application," *First Monday* 13, no. 6 (2008); Rick Bonney, Caren B. Cooper, Janis Dickinson, Steve Kelling, Tina Phillips, Kenneth V. Rosenberg, and Jennifer Shirk, "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy," *BioScience* 59, no. 11 (2009): 977–84; Johan Oomen and Lora Aroyo, "Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges," in *Proceedings of the 5th International Conference on Communities and Technologies* (Brisbane, Australia: Association for Computing Machinery, 2011), 138–49; Andrea Wiggins and Kevin Crowston, "From Conservation to Crowdsourcing: A Typology of Citizen Science," in *Proceedings of the 2011 44th Hawaii International Conference on System Sciences (HICSS)* (IEEE, 2011), 1–10, <http://www.computer.org/csdl/proceedings/hicss/2011/4282/00/07-04-02-abs.html>.
- <sup>5</sup> Donelle McKinley, "Design Principles for Crowdsourcing Cultural Heritage," Non-Profit Crowd (2015), <http://nonprofitcrowd.org/crowdsourcing-heuristics>; Jason Reed, M. Jordan Raddick, Andrea Lardner, and Karen Carney, "An Exploratory Factor Analysis of Motivations for Participating in Zooniverse, a Collection of Virtual Citizen Science Projects," in *Proceedings of the 2013 46th Hawaii International Conference on System Sciences (HICSS)*, 610–19, <https://www.computer.org/csdl/proceedings/hicss/2013/4892/00/4892a610-abs.html>.
- <sup>6</sup> C. Daniel Batson, Nadia Ahmad, and Jo-Ann Tsang, "Four Motives for Community Involvement," *Journal of Social Issues* 58, no. 3 (2002): 429–45.
- <sup>7</sup> Reed et al., "An Exploratory Factor Analysis"; Dana Rotman, Jen Hammock, Jenny Preece, Derek Hansen, Carol Boston, Anne Bowser, and Yurong He, "Motivations Affecting Initial and Long-Term Participation in Citizen Science Projects in Three Countries," *iConference 2014 Proceedings* (Urbana-Champaign, Illinois: iSchools, 2014).
- <sup>8</sup> Nicolas Kaufmann, Thimo Schulze, and Daniel Veit, "More than Fun and Money: Worker Motivation in Crowdsourcing—A Study on Mechanical Turk," in *Americas' Conference on Information Systems* 11 (2011): 1–11; Vickie Curtis, "Motivation to Participate in an Online Citizen Science Game: A Study of Foldit," *Science Communication* 37, no. 6 (2015): 723–46.
- <sup>9</sup> Richard M. Ryan and Edward L. Deci, "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," *Contemporary Educational Psychology* 25, no. 1 (2000): 54–67.
- <sup>10</sup> Mia M. Ridge, "From Tagging to Theorizing: Deepening Engagement with Cultural Heritage through Crowdsourcing," *Curator: The Museum Journal*, 56, no. 4 (2013): 435–50.
- <sup>11</sup> Ryan and Deci, "Intrinsic and Extrinsic Motivations."
- <sup>12</sup> Ryan and Deci, "Intrinsic and Extrinsic Motivations."
- <sup>13</sup> Mary Flanagan, Sukdith Punjasthitkul, Max Seidman, Geoff Kaufman, and Peter Carini, "Citizen Archivists at Play: Game Design for Gathering Metadata for Cultural Heritage Institutions," in *Proceedings of DiGRA 2013: DeFragging Game Studies* (Atlanta, GA: Digital Games Research Association, 2013).
- <sup>14</sup> Batson et al., "Four Motives for Community Involvement."
- <sup>15</sup> Sultana Lubna Alam and John Campbell, "Crowdsourcing Motivations in a Not-for-profit GLAM Context: The Australian Newspapers Digitisation Program," in *ACIS 2012: Location, Location, Location: Proceedings of the 23rd Australasian Conference on Information Systems 2012* (Melbourne, Australia: Deakin University, 2012).
- <sup>16</sup> M. Jordan Raddick, Georgia Brace, Pamela L. Gay, Chris J. Lintott, Carie Cardamone, Phil Murray, Kevin Schawinski, Alexander S. Szalay, and Jan Vandenberg, "Galaxy Zoo: Motivations of Citizen Scientists," *Astronomy Education Review* 12(1) (2013), accessed <http://portico.org/>

- stable?au=pgg3ztfcv7h ; Dana Rotman, "Collaborative Science across the Globe: The Influence of Motivation and Culture on Volunteers in the United States, India, and Costa Rica" (Ph.D. diss., University of Maryland, College Park, 2013); Rotman et al., "Motivations Affecting Initial and Long-Term Participation in Citizen Science Projects in Three Countries"; Curtis, "Motivation to Participate in an Online Citizen Science Game."
- <sup>17</sup> Ridge, "From Tagging to Theorizing"; Alam and Campbell, "Crowdsourcing Motivations in a Not-for-profit GLAM Context."
- <sup>18</sup> Stuart Dunn and Mark Hedges, "Crowd-Sourcing Scoping Study: Engaging the Crowd with Humanities Research," Arts and Humanities Research Council (2012), <http://www.humanitiescrowds.org>; Alexandra Eveleigh, Charlene Jennett, Ann Blandford, Philip Brohan, and Anna L. Cox, "Designing for Dabblers and Deterring Drop-outs in Citizen Science," in *Proceedings of the 32nd Annual Association of Computing Machinery Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada: Association of Computing Machinery, 2014), 2985–94; Curtis, "Motivation to Participate in an Online Citizen Science Game"; Raddick et al., "Galaxy Zoo"; Rotman et al., "Motivations Affecting Initial and Long-Term Participation in Citizen Science Projects in Three Countries."
- <sup>19</sup> Tim Causer and Valerie Wallace, "Building a Volunteer Community: Results and Findings from Transcribe Bentham," *Digital Humanities Quarterly* 6 (2012), <http://www.digitalhumanities.org/dhq/vol/6/2/000125/000125.html>.
- <sup>20</sup> Chern Li Liew, "Collaborative Construction of Digital Cultural Heritage: A Synthesis of Research on Online Sociability Determinants," *D-Lib Magazine* 21, no. 11 (2015): 5. On evolving a project's approach to include motivations, see also Nicole Saylor and Jen Wolfe, "Experimenting with Strategies for Crowdsourcing Manuscript Transcription," *Research Library Issues* (December 2011): 9–14; Jen Wolfe, (2015). "Diary of a Transcription Crowdsourcing Project: Or, How to Transform Your Institution, One Page at a Time," *Archival Outlook* (January/February 2015): 8–9, 30.
- <sup>21</sup> Oded Nov, Ofer Arazy, and David Anderson, "Technology-Mediated Citizen Science Participation: A Motivational Model," in *Proceedings of International AAAI Conference on Web and Social Media* (2011), <http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2802>.
- <sup>22</sup> Julia Noordegraaf, Angela Bartholomew, and Alexandra Eveleigh, "Modeling Crowdsourcing for Cultural Heritage," in *Proceedings of MW2014: Museums and the Web 2014* (February 17, 2014), <http://mw2014.museumsandtheweb.com/paper/modeling-crowdsourcing-for-cultural-heritage>. A notable exception is Rose Holley, "Many Hands Make Light Work: Public Collaborative OCR Text Correction in Australian Historic Newspapers" (Canberra, Australia: National Library of Australia, 2009).
- <sup>23</sup> Elizabeth R. Ellwood, Betty A. Dunckel, Paul Flemons, Robert Guralnick, Gil Nelson, Greg Newman, Sarah Newman et al., "Accelerating the Digitization of Biodiversity Research Specimens through Online Public Participation," *BioScience* 65:4 (2005): 383–396.
- <sup>24</sup> Raddick et al., "Galaxy Zoo"; Dana Rotman, Kezia Procita, Derek Hansen, Cynthia Sims Parr, and Jennifer Preece. "Supporting Content Curation Communities: The Case of the Encyclopedia of Life." *Journal of the American Society for Information Science and Technology*, 63:6 (2012): 1092–1107.
- <sup>25</sup> Alam and Campbell, "Crowdsourcing Motivations in a Not-for-profit GLAM Context," 1–11; Laura Carletti, Gabriella Giannachi, Dominic Price, Derek McAuley, and S. Benford, (2013). "Digital Humanities and Crowdsourcing: An Exploration," *Museums and the Web 2013 Conference*, April 17–20, 2013, Portland, OR (2013); Tim Causer, Justin Tonra, and Valerie Wallace, "Transcription Maximized; Expense Minimized? Crowdsourcing and Editing the Collected Works of Jeremy Bentham," *Literary and Linguistic Computing* 27, no. 2 (2012): 119–37; Rose Holley, "Crowdsourcing: How and Why Should Libraries Do It?," *D-Lib Magazine* 16, nos. 3–4 (2010); Nicole Kearney and Elycia Wallis, "Transcribing between the Lines: Crowd-Sourcing Historic Data Collection" (presented at Museums and the Web Asia 2015 Conference, October 5–8, 2015, Melbourne, Australia); Mia M. Ridge, ed., *Crowdsourcing Our Cultural Heritage* (Farham, Surrey, UK: Ashgate Publishing, Ltd., 2014); Causer and Wallace, "Building a Volunteer Community."
- <sup>26</sup> Rose Holley, "Crowdsourcing and Social Engagement: Potential, Power and Freedom for Libraries and Users," in *Proceedings of Pacific Rim Digital Library Alliance (PRDLA) Annual Meeting and Conference:*



- Libraries at the End of the World: Digital Content and Knowledge Creation* (Place of pub: publisher, 2009),1-28.
- <sup>27</sup> Alexandra Eveleigh, (2014). "Crowding Out the Archivist? Locating Crowdsourcing within the Broader Landscape of Participatory Archives," in *Crowdsourcing Our Cultural Heritage*, 211–29; Mary Flanagan and Peter Carini, "How Games Can Help Us Access and Understand Archival Images," *The American Archivist* 75, no. 2 (2014): 514–37.
- <sup>28</sup> Once fully transcribed and reviewed, the text created by volunteers is stored in the MySQL database and indexed in the Smithsonian's Enterprise Digital Asset Network (EDAN). This indexing allows searches to be made against the data and for individual pages to be discovered in the Smithsonian's Collections Search Center. The Collections Search Center can be accessed at <http://collections.si.edu/search/>
- <sup>29</sup> oldWeather can be accessed at <https://www.oldweather.org/> as part of the Zooniverse.
- <sup>30</sup> "Number of visits" relates to the number of individual visits to the website, but does not specify unique visitorship. This is important partly because the Transcription Center aims to have repeat visitors. "Average length of visit" specifies average length of time an individual user spends on the Transcription Center site.
- <sup>31</sup> "Number of visits" indicates the number of individual visits to the website, but does not identify unique visitorship. This is important partly because the Transcription Center aims to have repeat visitors. "Average length of visit" specifies average length of time an individual user spends on the Transcription Center site. "Active Users" specifies number of unique users who worked on the projects. "Replies" identifies the number of comments and questions sent by volunteers to the Transcription Center coordinator.
- <sup>32</sup> JSTOR Global Plants is the world's largest database of digitized plant specimens and supports international scientific research and collaboration; it is accessible at <http://plants.jstor.org/>. eBird is a National Science Foundation-funded citizen science resource established by Cornell Lab of Ornithology and National Audubon Society for birders featuring tracking, evaluation, and exploration of data. It is accessible at <http://ebird.org/content/ebird/>.