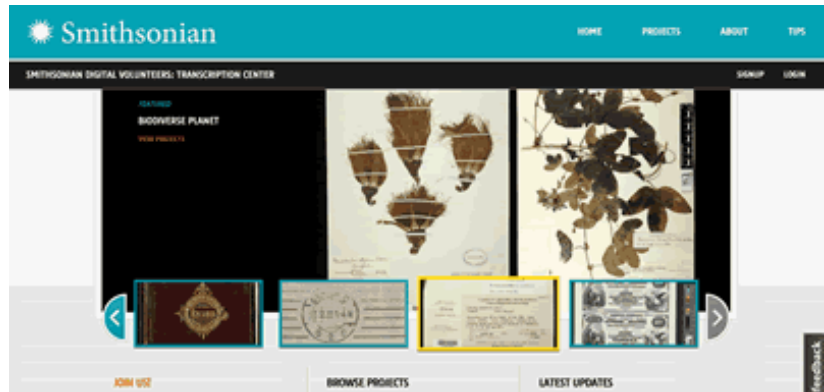# The Smithsonian Transcription Center: Crowdsourcing Document Transcription
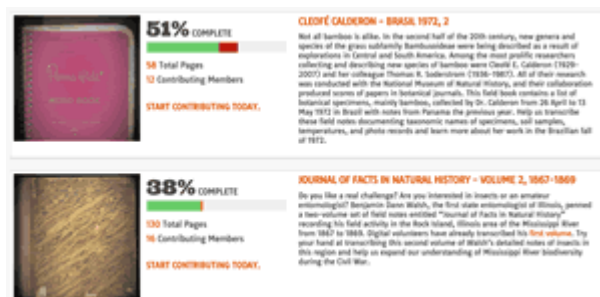
Case Study Overview

The Smithsonian Institution has 138 million objects and specimens as well as 2 million library volumes. Altogether, the Smithsonian has over 157,000 cubic feet of archival material in its various collections. Less than 1 percent are on display in its 19 museums, libraries, galleries, archives and research centers. What might we learn if we could search the thoughts of artists, scientists, inventors, explorers and revolutionaries as expressed in the documents and specimens of the Smithsonian's collections? Now, through digitization and transcription, ordinary citizens can help bring their insights to light!

In June 2013, partners in the Smithsonian Institution opened the Smithsonian Transcription Center. By visiting the website, volunteers are helping the center review and transcribe diaries, field notes, specimen labels, logbooks and more. The Smithsonian Transcription Center gets volunteers from across the United States and around the world.

Project Description

Participants in the project volunteer to type what they see in logbooks and on pages and specimen labels using instructions tailored to each document type. Volunteers can contribute anonymously or create an account to track their work. Once volunteers have an account, they may review the work of others and make edits where necessary.

Each week, new projects are added, with new discoveries waiting in the pages! The Smithsonian Transcription Center uses social media, email and blog posts to connect volunteers to each other and provide details about collections featured in the center.

When volunteers are transcribing a page, they may download it for free as a PDF. Once a project is complete, the entire project can be downloaded as a PDF — again for free! The text that volunteers create goes into the Smithsonian's database, searchable in the Collections Search Center. As volunteers transcribe specimen labels, they create data to make new

collection records. As they work, volunteers are creating greater access to and more useful information for Smithsonian Institution collections — which is seriously amazing!
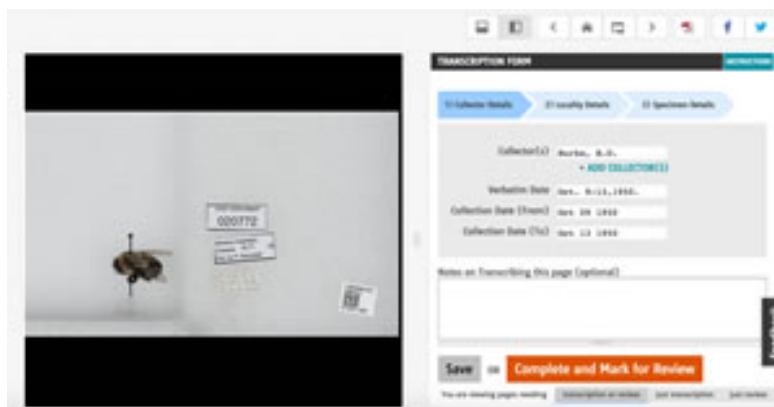
Challenges

For crowdsourcing projects like the Smithsonian Transcription Center, upholding the quality and validity of transcribed data is important. Volunteers work with Smithsonian staff in a three-step peer review process: Anyone can transcribe; registered volunteers review; and then staff validate the submissions. Specialized instructions for each type of material help keep volunteers on track, and they can submit questions directly or through social media.

Sustaining volunteer engagement is another challenge; the Smithsonian Transcription Center shares the collective success of volunteers as frequently as possible. Volunteers also have special behind-the-scenes access to curators and collections managers, and they get first peeks at newly digitized collections. They also have the opportunity to share what they are learning as they transcribe.

Benefits and Outcomes

The Smithsonian Transcription Center continues to meet its goals and to grow in the size and scope of its featured collections. The center's interface, workflows and communications are also improving through helpful feedback from volunteers.

Volunteers can expect to discover hidden histories, learn about scientific collecting, and understand the variety of Smithsonian collections. They also have the opportunity to join quarterly behind-the-scenes talks, ask Smithsonian staff questions, and make requests for new material — all part of the Smithsonian Transcription Center's commitment to making the serious fun of transcription even better. Through their social media chat and shared discoveries, volunteers have revealed hidden histories of women in science; contributed to Wikipedia, eBird and other citizen science projects; and helped Smithsonian staff identify the collections to be shared next.

The *Smithsonian Transcription Center's* 5,250 digital volunteers have completely transcribed and reviewed over 113,016 pages — a total that includes 859 projects shared by 13 Smithsonian archives, museums and libraries. The "pages" include biodiversity specimens, from which data have been transcribed and used to create 27,004 new collection records for bumblebees and 23,488 new records for U.S. National Herbarium sheets. Data transcribed from logbooks for the Digital Access to a Sky Century@Harvard program have been used by astronomers to correlate glass plates and the passage of light over time through the galaxy, thereby identifying cosmic events and locating black holes. The *Smithsonian Transcription Center* has indexed its

transcriptions and made the discoveries in its transcribed pages useful, resulting in over 3,300 downloads of project PDFs.

Tips

The *Smithsonian Transcription Center* case study illustrates the following steps in the Federal Citizen Science and Crowdsourcing Toolkit:

- **Scope Out Your Problem — Engage Your Stakeholders and Participants**
  The *Smithsonian Transcription Center* was conceived and guided by a committee of representatives from its eight original participating museums and archives; the committee has since grown to include representatives from the Smithsonian Institution's 13 archives, museums and libraries. Each month, the project coordinator sends the group an email with status updates. The email outlines ongoing successes, volunteer feedback and upcoming events. An additional administrative email informs the group about progress with site development and new system features.

- **Design a Project — Build Your Model**
  After giving careful consideration to the needs and challenges of similar projects, the *Smithsonian Transcription Center* decided to use peer review rather than algorithmic matching and cumulative community progress rather than leaderboards. The project also uses easy-to-follow instructions for volunteers to help them overcome a long learning curve.

- **Sustain and Improve — Communicate Effectively**
  The *Smithsonian Transcription Center* constantly checks with participants through emails and multiple social networks and social media tools. The Center asks for volunteer and staff input and carefully analyzes project performance to detect the needs of participants and ensure that the project meets its goals. The center wants to keep volunteering fun, purposeful and rewarding.

Learn More

- Website: *Smithsonian Transcription Center*

Contact Information

Name: Meghan Ferriter
Email: ferriterm@si.edu

Name: Ching-hsien Wang
Email: wangch@si.edu