Crowdsourcing Basics
What is crowdsourcing, and will it work for your collection?

Julia Colby
Vertebrate & Invertebrate Zoology Collections Manager

Alyssa Caywood
Collection Digitization Manager
Julia Colby
Vertebrate & Invertebrate Zoology Collections Manager

Alyssa Caywood
Collection Digitization Manager
Howe (2005):
Simply defined, crowdsourcing represents the act of a company or institution taking a function once performed by employees and outsourcing it to an undefined (and generally large) network of people in the form of an open call. This can take the form of peer-production (when the job is performed collaboratively), but is also often undertaken by sole individuals. The crucial prerequisite is the use of the open call format and the large network of potential laborers.
Citizen science

The collection and analysis of data relating to the natural world by members of the general public, typically as part of a collaborative project with professional scientists.
But for collections...

Crowdsourced transcription projects put the call out to a (hopefully) large network of people interested in science, to help digitize the dark data in our collections.

- Label transcription
- Field notes
- Diaries
- Georeferencing
- Camera trap data
- etc
Smithsonian Transcription Center
https://transcription.si.edu
Notes from Nature
https://www.notesfromnature.org

Zooniverse
https://www.zooniverse.org
Decipher our collections, discover hidden archives and contribute to knowledge
Join our community of 1,000+ volunteers

Get Involved ▼ Learn more

JOIN A VIRTUAL EXPEDITION OF...

Collection labels
Capture data from specimen and object labels to make it accessible for scientific and cultural research.
See all labels

Historical documents
Transcribe text and capture data from historical documents to make them digitally accessible.
See all field journals

Images
Identify and tag images of animals and collection objects to support information discovery and research.
See all camera traps

DigiVol
https://volunteer.ala.org.au/
Is this right for my collection?
How soon do you need the data?
What do you need?

Some collections are more appropriate than others.

You should consider:

- What information do you need to capture?
- How complex are your labels?
- How Darwin Core compliant is your data management system?
- How much time do you have to build and maintain expeditions:
  - Create image derivatives
  - Create expedition templates
  - Upload data and images
  - Create tutorials
  - Answer forum questions
- What is your plan for clean up/data prep for repatriation to your database?
Potential drawbacks

- Timeline will vary depending on time of year
- Volunteers may not read your tutorials/incorrectly format data (requiring more clean up)
- Building new expedition templates & writing tutorials can be time consuming
- You will need to produce image derivatives for the expedition
- You will need to clean the data (even with good instructions and conscientious volunteers)
Potential benefits

- Exposure--your collections will be seen by people who may never visit your museum
- Your labels will get transcribed
- Participation in events like WeDigBio provide opportunities for educational outreach
- Provide opportunities for collaborations with other collections/institutions
- It’s a chance to let your collections shine
Evaluate your labels

Simple labels are better for everyone--your tutorials will be easier to write and transcribers will be more accurate.
Map your data

Planning in advance what information you are getting from labels and other sources (catalog cards, ledgers, etc) will help you determine what information is appropriate for crowdsourcing.
Compare transcription platforms

We evaluated our goals, and our collections, and then assessed platforms.

Some projects only need a few pieces of information from transcribers and have a game-like interface; other projects require volunteers to parse complex information.
Have a team

Be prepared to divide and conquer:

● Who will build an expedition?
● Who will create/prep the images?
● Who will prepare the data/image upload?
● Who will write the tutorials?
● Who will interact with online volunteers and answer questions?
● Who will validate transcriptions?
● Who will maintain expeditions, including activation/deactivation?
Data Flow

You've built and activated an expedition--what happens? For DigiVol:

Part 1: Transcription

Online volunteer enters label information in the form.

Part 2: Validation

Validation volunteer proofs the data entered in the form.

Part 3: Clean-up

You download a csv of your validated data. Museum volunteers and staff work on final clean-up, standardization, and organization of data for import into the museum database.
In the expedition, click “View Tutorial” to learn how to transcribe labels.
Plan your clean up

Repatriation of the label information into your institution's or department's database is a vital part of the process.
- 6 Expeditions
- 2432 images transcribed
- 75 participants

Parts 1 & 2: ~5 months
Part 3: In progress

---

- 6 Expeditions
- 3044 images transcribed
- 28 participants

Parts 1 & 2: ~3.5 months (111 days)
Part 3: In progress
Intrigued? Try it!

Remember you are a part of a community. You won’t have to reinvent the wheel—you can modify existing templates for your collection projects. If you like a particular project, ask for advice from the project contact.
Can my institution join?

Yes! Scroll to the bottom of the DigiVol main screen, click “Submit an expedition.” Halfway down the “About DigiVol” page, you will see the contact address--send an email inquiry to get started.
Do you have a biodiversity transcription project? How can you make your own Notes from Nature project?

Notes from Nature welcomes anyone who wants to build their own project. Before you contact us for more details, first ask whether Notes from Nature is the right platform for you. Do you have imaged biodiversity specimens with label text (or just labels) that you need transcribed? If so, onwards.

For now, the first step is to contact the Notes from Nature project manager, Michael Denslow (notesfromnature.pm@gmail.com), who can provide you with the full details needed to get a project together. In a nutshell, Notes from Nature is built on top of the Zooniverse’s Project Builder, which greatly simplifies creation of projects. With our new system, you can spin up a project in minutes to hours. All you need are images you want transcribed and a clear idea of what fields you want to capture.
Once you’ve picked a platform, built your expeditions, and launched them, consider other ways to connect your digitized collections to volunteers:

- **Transcription events**
  - WeDigBio
- **Social media**
  - Facebook
  - Twitter
  - Blogs
- **Member events**
Signing up was easy....

We went to [www.wedigbio.org](http://www.wedigbio.org) and used the Resource menu to begin planning.
MPM DataBlitz 2016

Small pilot event October 21-22, 2016

- Invited two classes from a local all-girls high school for a closed event on Friday
- Opened to the public on Saturday (coincided with a large annual MPM archaeology event)
- Promoted on MPM website, in member newsletters, social media, and on local television
Promotion helps!

Most participants who signed up on our website heard about the event from a 3 minute break in we did for the local morning news.
Get feedback!

Although our event was small, we know what participants liked...

- “That I was helping the scientific community”
- “The importance of our tasks was well explained and we were given the complete picture of the project”
- “I can do it from home now”
- “Contributing to the improvement of the museum's data”
- “Learning about the whole process of transcription and why it's important for science and the world.”
- “Getting to contribute to MPM and developing skills”
- “Finding out about crowdsourcing transcription opportunities”
- “Knowing that we helped with this large undertaking”
- “Knowing that I contributed to the advancement of science made doing the activity satisfying”
- “I enjoyed learning new things and realizing that everything in the ecosystem is valuable”
...And what we should consider (or keep doing) for our next event.

- “MPM could offer annual transcription sessions for high school students who want to participate”
- “I think the tours telling us why all of the collecting mattered was important”
- “Going to see behind the scenes collection was very important”
- “Have some snacks so people can stay refreshed”
- “Have one at the library”
- “Getting to meet the staff was important”
- “More advertisement through social media or posters”
Would we do it again?

YES.

Next time:

- More social media/promotion
- Different space (and music!)
- Keep the tours!
- Connect to other institutions in real time
- Plan for a more robust survey with follow up to evaluate impacts/ retention of participants as online volunteers
- Incentives (swag)
- More staff to help answer transcription questions
- Encourage participants to register for DigiVol ahead of time
October 19-22, 2017
To host an event or sign up for the mailing list, visit www.wedigbio.org
Questions?

Resources

- iDig Bio: [www.idigbio.org/tags/crowdsourcing](http://www.idigbio.org/tags/crowdsourcing)
- WeDigBio: [www.wedigbio.org](http://www.wedigbio.org)
- DigiVol: [http://volunteer.ala.org.au](http://volunteer.ala.org.au)
- Notes from Nature: [www.notesfromnature.org](http://www.notesfromnature.org)
- Smithsonian Transcription Center: [https://transcription.si.edu/](https://transcription.si.edu/)
- Zooniverse: [www.zooniverse.org/](http://www.zooniverse.org/)

Questions?

Julia Colby
colby@mpm.edu

Alyssa Caywood
caywood@mpm.edu