



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

Crowdsourcing


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...

Crowdsource 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



JOIN US!

LEARN HOW TO TRANSCRIBE

Become a Smithsonian Digital Volunteer and help us make historical documents and biodiversity data more accessible.

Join 6,920 volunteers to add more to the total 202,362 pages of field notes, diaries, ledgers, logbooks, currency proof sheets, photo albums, manuscripts, biodiversity specimens labels that have been collaboratively transcribed and reviewed since June 2013 - **Get started Now!**

Follow us on Twitter and learn more about projects: @TranscribeSI

Partner on projects and ask your

BROWSE PROJECTS

Select a category below to begin browsing projects.

Select a Category

FIELD BOOK PROJECT X ARCHIVES MONTH

H.G. Deignan - Field Catalog, Ceylon, 1944

88 CONTRIBUTIONS

Get insight into the birds, mammals, and other wildlife of Sri Lanka with ornithologist Herbert G. Deignan's 1944 field catalog. Deignan would go on to

LATEST UPDATES

@SiobhanLeachman transcribed a page from STRI Pollen Cards (Set 02)

Jessica Powers marked for review a page from IZ - Division of Marine Invertebrates Cards (Set 01)

@SiobhanLeachman marked for review a page from STRI Pollen Cards (Set 02)

@SiobhanLeachman transcribed a page from STRI Pollen Cards (Set 02)

@SiobhanLeachman marked for review a page from STRI Pollen Cards (Set 02)

Jessica Powers marked for review a


feedback

Smithsonian Transcription Center

https://transcription.si.edu

Notes from Nature

<https://www.notesfromnature.org>








Get Started About Talk Blog

NOTES FROM NATURE

Notes from Nature

TRANSCRIBE MUSEUM RECORDS

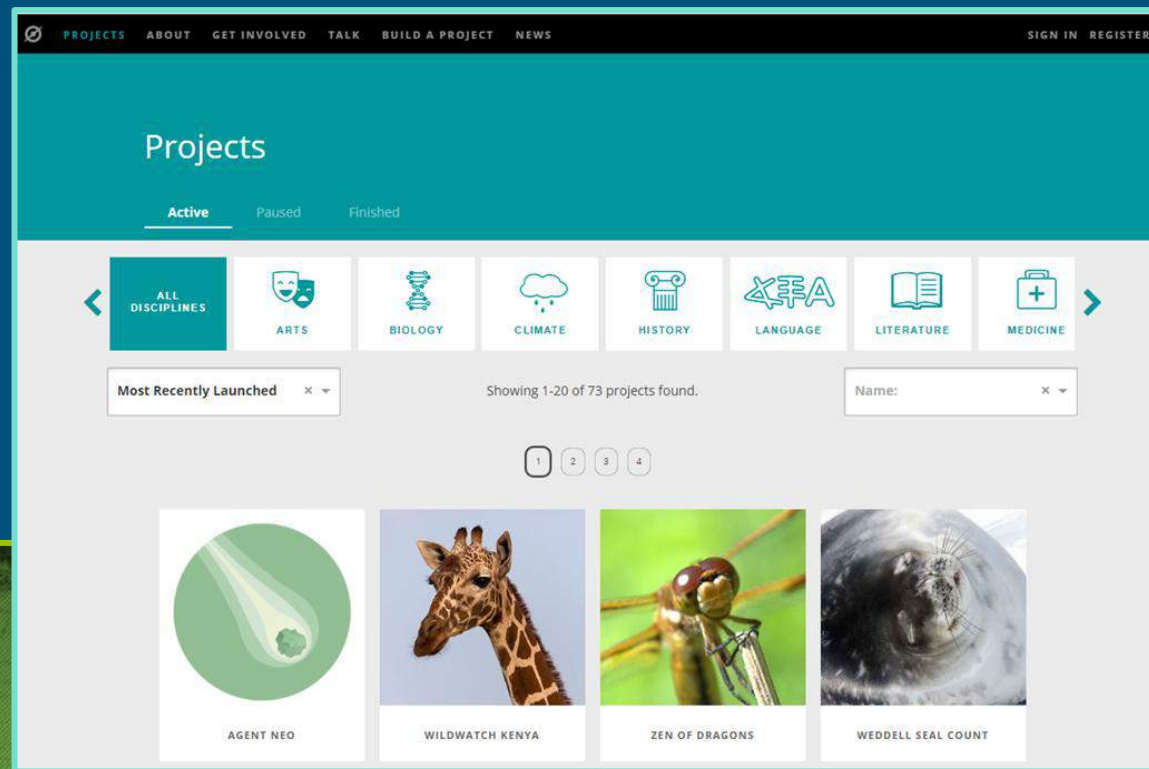
Choose a Group and Start transcribing!

 Plants	 Bugs	 Magnified	 Fossils	 NFN Labs
15 Expeditions	3,829 Volunteers	308,635 Classifications	116,000 Subjects	94,991 Completed

Why should you get involved?

To improve our world
Museum records contain historical biodiversity data. Scientists and researchers can use the data to conduct new research and make better conservation decisions.

Transcribe with pride



PROJECTS ABOUT GET INVOLVED TALK BUILD A PROJECT NEWS SIGN IN REGISTER





Projects

Active Paused Finished

ALL DISCIPLINES ARTS BIOLOGY CLIMATE HISTORY LANGUAGE LITERATURE MEDICINE

Most Recently Launched x Showing 1-20 of 73 projects found. Name: x

1 2 3 4

 AGENT NEO	 WILDWATCH KENYA	 ZEN OF DRAGONS	 WEDDELL SEAL COUNT
---	--	---	---

Zooniverse

<https://www.zooniverse.org>

Notes from the Researchers

"The digitized data you are creating will help advance research related to species extinction, ecosystem changes, environmental health and even human health."



Decipher our collections, discover hidden archives and contribute to knowledge

Join our community of 1,000+ volunteers

Get involved ↓

Learn more

Image by John Tann from Flickr

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

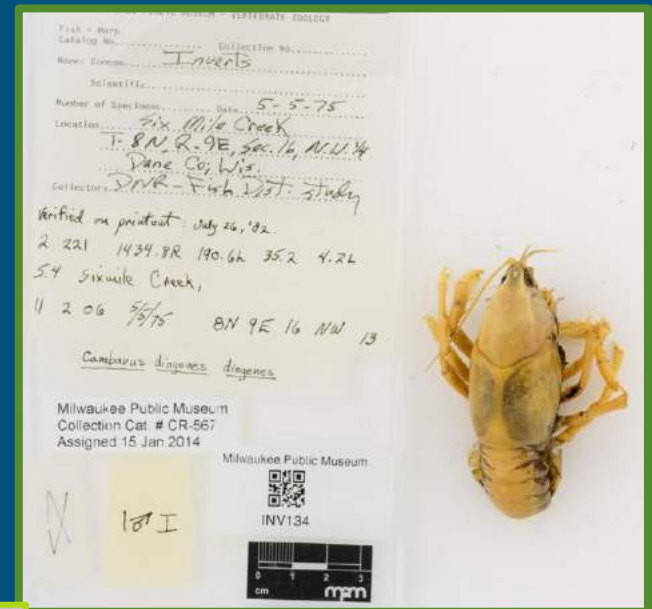
- Timeline will vary depending on time of year
- Volunteers may not read your tutorials/incorrectly format data (requiring more clean up)
- Building new expedition templates & writing tutorials can be time consuming
- You will need to produce image derivatives for the expedition
- You *will* need to 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.

Bee Data Entry Map



Catalogue Module:

ZOO SPEC

Collection Details
Department: 1 Invertebrate Zoology Specimen Count: 1 No. of Pieces: 1
Catalogue: 1 Insecta Original Count: 1

Identification
Taxon: 1 Rubus - Rosaceae Type Status:

Collection Event
Collection Event: North America, United States, Wisconsin, Milwaukee, Metzsch Park

Specimen Details

No. of Spec...	Spec. Type	Life Stage	Sex	Remarks
1	Whole Body	Adult	Unknown	

Incubation Stage: Plummage:

Set Mark: Parasites: Indicators: On Loan: Tissue: Voucher:
On Exhibit: MM:

Dates/Acq. Affiliations Measurements Description 1 Description 2 Materials Stratigraphy Provenience Zoo. Specimen

Display | Object 1 of 1 caywood Botany Level 4 emuteat

BeeMu Project
Default/ Record
Template (~40
fields)

MPM Staff
Parsing/Open
Refine

Online Volunteer
Transcription (~20
fields)

Online Volunteer
Transcription/
Parsed by MPM
staff (~4 fields)

ZOO SHARED

Preparation Details

Prep. type	Count	Medium	Prepared By	Date	Location	Prep. remarks
1 Pinned	1					

Related Objects

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.

Crowdsourcing Transcription Tools Comparison							
		Smithsonian Bumblebee Project	Zooniverse/ Notes From Nature-- CalBug Expeditions	Zooniverse/Notes From Nature-- NY Botanical Garden Mycological Collection	DigiVol-- U of Hawai'i Insect Museum	Digivol--New Zealand Arthropod Collection Project: Plant pests of the Australasian-Pacific region	MPM WI Crayfish
2							
34	Host				x	v	
35	Locality	v	v	v (includes habitat info)	D/V ("Other Locality Info")	v	v
36	Country	D	D	D	x	x	
37	2 letter region code					x	
38	State/Province	x	D	D (if US; otherwise free text)	x		
39	District/County	x	D	D (if US; otherwise free text)	x		x
40	Municipality/City/Town				x		x
41	Township						x
42	Range						x
43	Section						x
44	Quarter						x
		V (info that goes beyond country/state/district/I at-long)					
45	Precise Locality						
	Location Information						
46	Nearest Named Place						
47	Island				x		
48	Waterbody						
49	Elevation (from/to)	x					
50	Elevation (single field)		x		v	v	
51	Elevation Unit (feet or meters)	D					
53	Coordinate Uncertainty in Meters				D		
	Numbers						
54	Barcode/ USNMMENT number	x					
55	Catalog Number				x	x	
56	Other Number (s)				x		
	Notes						
57	Collection Notes						For verbatim collection method details.
		"Label Note" --for additional info; if information is on separate labels, you can create additional notes fields	For additional information on labels only.				Other Label Notes; Nearest named place island name waterbody, water temp, weight, depth, elevation, lat/long, etc
58	Notes					For other text and numbers.	
59	Notes on Transcribing this Page (optional)	x			x	x	
60	Total Number of fields (Required and optional).	15	9	5	25	19	18

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.

Back

A DIGIVOL Expedition

My Profile ▾



Milwaukee Public Museum Crayfish Expedition 3

The Milwaukee Public Museum holds the largest collection of crayfish from Wisconsin. These specimens formed the foundation of *The Crayfish and Shrimp of Wisconsin* by Joan Jass and H. H. Hobbs III. As aquatic ecosystems are impacted by humans, crayfish could potentially become invasive species as they populate new habitats, or native populations may decline, risking extinction. This expedition will give scientists the tools they need to answer these questions.

These labels can be challenging to interpret--please see our updated tutorials. Click on "View tutorial" to download the PDFs. Thank you for helping us transcribe our labels!



Crayfish Expedition 3

Admin

Settings

Validate tasks

Get Started →

View tutorial ←

Visit Project Forum »



4
Volunteers

497
Tasks

In the expedition, click "View Tutorial" to learn how to transcribe labels.

DIGIVOLMilwaukee Public Museum- Bees Expedition 1
MPM_ENT_333_1_cropped.jpg

Skip

Create Forum Topic

View Tutorial

Catalog Number ENT333



Institution: Milwaukee Public Museum Project: Milwaukee Public Museum- Bees Expedition 1 Catalog Number: ENT333 Taxa: Ceratina dupla

Institution: Milwaukee Public Museum Project: Milwaukee Public Museum- Bees Expedition 1 Catalog Number: ENT333 Taxa: Ceratina dupla

1. Collection Notes



Copy values from a previous task

2. Collection Location

Collection Date (from) Year MM DD ?
(to) Year MM DD ?

Country United States ?

State Province WI ?

County Milwaukee Co. ?

Municipality

Precise Location

Verbatim Latitude

Verbatim Longitude

3. Miscellaneous

Catalog Number ENT333

Other Number ?

Identified By ?

Date Identified (from) Year MM DD ?
(to) Year MM DD ?

Collector(s) ?

4. Notes Record any comments here that may assist in validating this task

Shrink

Your Notes

Submit for validation

Save unfinished record

Skip

Create Forum Topic

Plan your clean up

Repatriation of the label information into your institution's or department's database is a vital part of the process.

The screenshot shows a Google Sheet titled "Bee Expeditions Data" with a task list. The sheet is organized into columns for Task, Assigned, Status, and Date Completed. The tasks are grouped into four main categories: Spreadsheet maintenance, Data Cleaning, Import Prep, and Import. The tasks are color-coded: green for Spreadsheet maintenance, orange for Data Cleaning, blue for Import Prep, and grey for Import.

Task	Assigned	Status	Date Completed
Data download	Alyssa	Done	5/30/2017
Data compiled in original spreadsheet	Alyssa	Done	5/30/2017
Creation of working Master spreadsheet/upload to Google workspace	Alyssa	Done	5/30/2017
Preliminary changes (connection of formatting errors from transfer, ex; male/female symbols, carriage returns)	Alyssa	In Progress	
Addition of in-house transcription data	Alyssa	Done	5/30/2017
Continuity/record quality control	Alyssa	In Progress	
IRNs (if needed)	Alyssa		
People/Parties--Collectors	Anna	Julia	
Create new parties upload/add new parties to EMu	Julia/Debra if needed	In Progress	
Back Match Parties w/IRNs	Chris		
Dates (Collected and Identified)	John	DONE	6/30/2017
People/Parties--Determiners	John	Julia Needs to check people	
Parsing numbers	Alyssa, John		
Lat/Long	John	DONE	6/30/2017
Mapping	John	DONE	6/30/2017
Geopolitical Location	Anna	DONE	6/30/2017
Precise Location	Julia	In Progress	
Location Notes/Habitat	Julia	In Progress	
Collection Notes	Julia/Alyssa/John		
Follow up on Transcriber/Validator notes.	Julia		
Update Column headers	Alyssa		
CSV download	Alyssa		
Final Cleaning	Alyssa/Chris		
Test Import/Troubleshooting	Debra/Alyssa/Julia/Chris		
EMu Import	Debra		

Crayfish

(9 species)

- 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

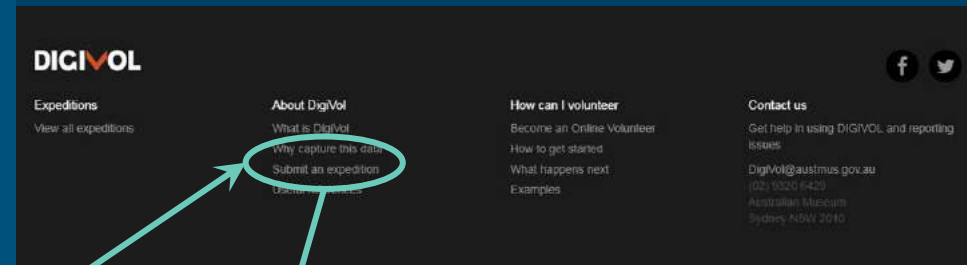
Bees
(110+ species)

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.



SUBMIT AN EXPEDITION

DigiVol is open to any institution or individual who has a project that would be well suited to DigiVol volunteers.

Any proposed expedition will need to conform to an existing DigiVol template (fields may be added or removed) or have sufficient funds to enable the development of a new task template.

If you are interested in posting an expedition on DigiVol and you have material that would be suitable please contact us (DigiVol@austmus.gov.au)



VISIT NOTES FROM NATURE

CLASSIFY

TALK

COLLECT

BLOG

FIELD BOOK

Research

The Team

FAQ

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.

ZOOIVERSE

Projects
Collections
Build a Project
How to Build
Project Policies
FAQ

About Us
Education
Our Team
Publications
Acknowledgements
Contact Us

Zooniverse Talk
Daily Zooniverse
Blog





Connect your projects!

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....

The screenshot shows the WeDigBio website interface. At the top, there is a navigation bar with the WeDigBio logo and menu items: Home, Get Involved, Resources (highlighted with an orange arrow), About, and Team Members. Below the navigation bar is a green header with the text "Worldwide Engagement for Digitizing Biocollections" and "Admin Login". The main content area has a breadcrumb trail "Resources / Event Planning" and a large heading "Event Planning". Below the heading, there is a paragraph of text: "These documents are intended to provide guidance for leaders of WeDigBio on-site transcription blitz events. Feel free to edit or modify to suit the needs of your institution. Also see Austin Mast's plenary talk on WeDigBio at the 2015 Society for the Preservation of Natural History Collections Conference for additional ideas and context." Below this text are two links: "Event Planning" (highlighted with a green arrow) and "Tasklist and Timeline" (highlighted with a green arrow). At the bottom of the page, there is a blue footer with a grid of links: Home, Get Involved (Event Signup, Calendar, Find a Project, Event List, 2016 Events, 2015 Events), Resources (Educational, Exercises, Event Planning, Games, Press Kit), About (FAQ, Contact us), and Team Members (WeDigBio Team, Members). To the right of the footer links are social media icons for Facebook, Twitter, YouTube, and RSS. Below the footer links are logos for iDigBio, Australian Museum, Smithsonian Institution, UF, and the University of Queensland. At the very bottom, there is a small text block: "WeDigBio is funded by a grant from the National Science Foundation's Advancing Digitization of Biodiversity Collections Program (Cooperative Agreement EF-1115210). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

We went to 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





MPM DataBlitz 2016



DataBlitz Totals

	Participants	Tasks Transcribed
St. Joan's Students	18	51
Public	15	181
MPM Employee	3	10
MPM Volunteer/ Intern	2	25
	38	267

MPM DataBlitz 2016

MPM Milwaukee Public Museum

PLAN A VISIT RESEARCH & COLLECTIONS PROGRAMS JOIN & SUPPORT

What are you looking for?

DataBlitz

Get FREE Museum Admission and Help Us Transcribe Scientific Data from Our Collections at MPM's DataBlitz!

Be a part of Citizen Science and Help MPM Digitize our ...

SCIENCE NEEDS YOU!

MPM is digitizing our natural history collections as part of an international endeavor to broaden education and research. We want to see how many natural history specimen records can be transcribed in a single day and we are inviting you to get in on the action!

When: Saturday, October 22, 2016 from 11:00 a.m.-3:00 p.m.

Registration is required. To learn more or sign-up click here.

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



Worldwide Engagement for Digitizing Biocollections



@WeDigBio

Questions?



Julia Colby
colby@mpm.edu



Alyssa Caywood
caywood@mpm.edu

Resources

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

