

Zooniverse Help

Getting Started

How to create a project with our Project Builder

Example Project

Glossary

Lab policies

Best Practices

Building, Launching, and Running Your Zooniverse Project:

Part I: Building a Great Project

Part II: The Launch Rush

Part III: In For the Long Haul

Appendix: Resources and Examples

Next Steps

[Use Caesar for more efficient data processing](#)

Data Analysis

Data Exports

Experimental Features

Hosting your own subject media

Markdown

Newsletters

[Subject Selection Process](#)

Transcription Project Guide

Getting Started

Project data

Workflows

Designing for volunteers

Beta testing

Project launch

Working with data

[Project maintenance and conclusion](#)

Acknowledgements and resources

Zooniverse Help

How to create a project with our Project Builder

A quick guide to building a project

A brief overview of all the steps you need to complete to set up your project.

1. Log in to your Zooniverse account.
2. Click on "Build a Project" and click "Create a New Project".
3. Enter your project details in the pop-up that appears. You'll then be taken to a "Project Details" page where you can add further basic information.
4. Add more information about your project via the blue tabs on the left-hand side of the project builder page. Guidance for each of these sections is provided on the page itself.
5. If you have data to upload, add your subjects via the "Subject Sets" tab of the Project Builder (detailed instructions are provided here),
6. Next, create your workflow. Click on the "Workflow" tab in the navigation bar on the left-hand side of the page. Use the "New workflow" button to create a workflow. Multiple workflows can be created.
7. Once you've created a workflow, the "Associated Subject Sets" section allows you to link your workflow to your subject sets. If you have no subjects, go to the "Subject Sets" tab and upload your data (see step 5 above).
8. Hit the "Test this workflow" button to see how your project looks.
9. Explore your project to figure out what works and what doesn't. Make changes, then refresh your project page to test these out.

10. Guidelines on how to design your project to maximize engagement and data quality are provided on the Policies page.
11. When you are happy with your project, set it to "Public" on the "Visibility" tab. Use the "Apply for review" button to submit it to the Zooniverse team for review.

Navigating the Project Builder

On the left-hand side of the project builder you will see a number of tabs which can be divided into three key sections; "Project", "Workflows" and "Subject Sets". These are terms you'll see a lot, and they have specific meanings in the Zooniverse.

- **Project** is pretty self-explanatory; Galaxy Zoo and Penguin Watch are examples of Zooniverse projects that could be built using the project builder.
- **Workflows** are sequences of tasks that volunteers are asked to do.
- **Subject sets** are collections of data (typically images) that volunteers are asked to perform tasks on.

For more Zooniverse definitions, check out the Glossary page.

Project

The tabs listed below are where you enter descriptive information for your project.

- **Project Details:** Here you can add information that generates a home page for your project. Start by naming and describing your project, add a logo and background image.
- **About:** Here you can add all sorts of additional pages, including *Research*, *Team*, *Results*, *Education*, and *FAQ*
- **Collaborators:** Add people to your team. You can specify their roles so that they have access to the tools they need (such as access to the project before it's public).

- **Field Guide:** A field guide is a place to store general project-specific information that volunteers will need to understand in order to complete classifications and talk about what they're seeing.
- **Tutorial:** This is where you create tutorials to show your users how to contribute to your project.
- **Media:** Add images you need for your project pages (not the images you want people to classify!)
- **Visibility:** Set your project's "state" - private or public, live or in development, and apply for review by the Zooniverse. You can also activate or deactivate specific workflows on this page.
- **Talk:** Create and manage discussion boards for your project.
- **Data Exports:** Access your raw and aggregated classification data, subject data, and comments from Talk.

Workflows

A workflow is the sequence of tasks volunteers are asked to perform. For example, you might want to ask volunteers to answer questions about your images, or to mark features in your data, or both. The workflow tab is where you define those tasks and set the order in which volunteers are asked to do them. Your project might have multiple workflows (if you want to set different tasks for different image sets). See the detailed Workflow section for more information.

Subjects

A subject is a unit of data to be analyzed. A single subject can include more than one image. A "subject set" consists of both the "manifest" (a list of the subjects and their properties), and the images themselves. Subjects can be grouped into different sets if useful for your research. See the Subject Details section for more on subjects.

Project building in detail

Detailed instructions on how to use the pages described above.

Project details

- **Name:** The project name is the first thing people will see and it will show up in the project URL. Try to keep it short and sweet.
- **Avatar:** Pick an avatar image for your project. This will represent your project on the Zooniverse home page. It can also be used as your project's brand. It's best if it's recognizable even as a small icon. To add an image, either drag and drop or click to open your file viewer. For best results, use a square image of not more than 50KB, but at minimum 100x100 pixels.
- **Background:** This image will be the background for all of your project pages, including your project's front page. It should be relatively high resolution and you should be able to read text written across it. To add an image, either drag and drop or click to open your file viewer. For best results, use images of at least 1 megapixel, no larger than 256 KB. Most people's screens are not much bigger than 1300 pixels across and 750 pixels high, so if your image is a lot bigger than this you may find it doesn't look the way you expect. Feel free to experiment with different sizes on a "typical" desktop, laptop or mobile screen.
- **Description:** This should be a one-line call to action for your project. This will display on your landing page and, if approved, on the Zooniverse home page. Some volunteers will decide whether to try your project based on reading this, so try to write short text that will make people actively want to join your project.
- **Introduction:** Add a brief introduction to get people interested in your project. This will display on your project's front page. Note this field renders markdown, so you can format the text. You can make this longer than the Description, but it's still probably best to save much longer text for areas like the About, Research Case or FAQ tabs.
- **Workflow Description:** A workflow is a set of tasks a volunteer completes to

create a classification. Your project might have multiple workflows (if you want to set different tasks for different subject image sets). Add text here when you have multiple workflows and want to help your volunteers decide which one they should do.

- **Checkbox: Volunteers choose workflow:** If you have multiple workflows, check this to let volunteers select which workflow they want to work on; otherwise, they'll be served randomly.
- **Researcher Quote:** This text will appear on the project landing page alongside an avatar of the selected researcher. It's a way of communicating information to your volunteers, highlighting specific team members, and getting volunteers enthusiastic about participating.
- **Announcement Banner:** This text will appear as a banner at the top of all your project's pages. Only use this when you've got a big important announcement to make! Many projects use this to signal the end of a beta review, or other major events in a project's life cycle.
- **Discipline Tag:** Enter or select one or more discipline tags to identify which field(s) of research your project belongs to. These tags will determine the categories your project will appear under on the main Zooniverse projects page, if your project becomes a full Zooniverse project.
- **Other Tags:** Enter a list of additional tags to describe your project separated by commas to help users find your project.
- **External and Social Links:** Adding an external link will add an entry in an External Project Links section that appears at the bottom of the project landing page. Adding a social link will add an entry in the same section along with an appropriate media icon in the link list. You can rearrange the displayed order among external links and social links by clicking and dragging on the left gray tab next to each link in the project builder interface.
- **Checkbox: Private project:** On "private" projects, only users with specified project roles can see or classify on the project. We strongly recommend you keep your project private while you're still editing it. Share it with your team to get

feedback by adding them in the Collaborators area (linked at the left-hand side of the Project Builder). Team members you add can see your project even if it's private. Once your project is public, anyone with the link can view and classify on it.

About

This section contains pages where you can enter further information for Research, Team, Results, Education and FAQ. All of these pages use Markdown to format text and display images.

- **Research:** Use this section to explain your research to your audience in as much detail as you'd like. Explaining your motivation to volunteers is critical for the success of your project – please fill in this page (it will display even if you don't)!
- **Team:** Introduce the members of your team, and the roles they play in your project.
- **Results:** Share results from your project with volunteers and the public here. This page will only display if you add content to it.
- **Education:** On this page, you can provide resources for educators and students to use alongside your project, such as course syllabi, pedagogical tools, further reading, and instructions on how the project might be used in an educational context. This page will only display if you add content to it.
- **FAQ:** Add details here about your research, how to classify, and what you plan to do with the classifications. This page can evolve as your project does so that your active community members have a resource to point new users to. This page will only display if you add content to it.

Collaborators

Here you can add people to your team. You can specify their roles so that they have access to the tools they need (such as access to the project before it's public).

- **Owner:** The owner is the original project creator. There can be only one.
- **Collaborator:** Collaborators have full access to edit workflows and project content, including deleting some or all of the project.
- **Expert:** Experts can enter “gold mode” to make authoritative gold standard classifications that will be used to validate data quality.
- **Researcher:** Members of the research team will be marked as researchers on “Talk”.
- **Moderator:** Moderators have extra privileges in the community discussion area to moderate discussions. They will also be marked as moderators on “Talk”.
- **Tester:** Testers can view and classify on your project to give feedback while it’s still private. They cannot access the project builder.
- **Translator:** Translators will have access to the project builder as well as the translation site, so they can translate all of your project text into a different language.

Field Guide

A field guide is a place to store general project-specific information that volunteers will need to understand in order to complete classifications and talk about what they're seeing. It's available anywhere in your project, accessible via a tab on the right-hand side of the screen.

Information can be grouped into different sections, and each section should have a title and an icon. Content for each section is rendered with Markdown, so you can include any media you've uploaded for your project there.

Tutorial

In this section, you can create a step-by-step tutorial to show your users how to use your project. You can upload images and enter text to create each step of the tutorial. You can add as many steps as you want, but keep your tutorial as short as

possible so volunteers can start classifying as soon as possible.

In some cases, you might have several different workflows, and will therefore need several different tutorials. In the Workflows tab, you can specify which tutorial shows for the workflow a volunteer is on.

Media

You can upload your own media to your project (such as example images for your Help pages or Tutorial) so you can link to it without an external host. To start uploading, drop an image into the grey box (or click "Select files" to bring up your file browser and select a file). Once the image has uploaded, it will appear above the "Add an image" box. You can then copy the Markdown text beneath the image into your project, or add another image.

Visibility

This page is where you decide whether your project is public and whether it's ready to go live.

- **Project State and Visibility:** Set your project to "Private" or "Public". Only the assigned collaborators can view a private project. Anyone with the URL can access a public project. Here, you can also choose whether your project is in "Development", or "Live". Note: in a live project, active workflows are locked and can no longer be edited.
- **Beta Status:** Here, you will find a checklist of tasks that must be complete for your project to undergo beta review. Projects must complete review in order to launch as full Zooniverse projects and be promoted as such. Once these tasks are complete, click "Apply for review".
- **Workflow Settings:** You will see a list of all workflows created for the project. You can set the workflows to "Active", choose what metric to measure for completeness statistics, and whether those statistics should be shown on your

project's Stats Page. For more information on the different project stages, see our Project Builder policies.

Talk

"Talk" is the name for the discussion boards attached to your project. On your Talk, volunteers will be able to discuss your project and subjects with each other, as well as with you and your project's researchers. **Maintaining a vibrant and active Talk is important for keeping your volunteers engaged with your project.**

Conversations on Talk also can lead to additional research discoveries.

You can use this page to set up the initial Talk boards for your project. We highly recommend first activating the default subject-discussion board, which hosts a single dedicated conversation for each subject. After that, you can add additional boards, where each board will host conversation about a general topic. Example boards might include: "Announcements," "Project Discussion," "Questions for the Research Team," or "Technical Support."

Data Exports

In this section you can request data exports for your Project Data (CSV format) and Talk Data (JSON format). Note that the Zooniverse will process at most 1 of each export within a 24-hour period and some exports may take a long time to process. We will email you when they are ready. For examples of how to work with the data exports see our Data Digging code repository.

Workflow Details

Note that a workflow with fewer tasks is easier for volunteers to complete. We know from surveys of our volunteers that many people classify in their limited spare time, and sometimes they only have a few minutes. Longer, more complex workflows mean each classification takes longer, so if your workflow is very long you may lose volunteers.

- **Workflow title:** Give your workflow a short, but descriptive name. If you have multiple workflows and give volunteers the option of choosing which they want to work on, this Workflow title will appear on a button instead of "Get started!"
- **Version:** Version indicates which version of the workflow you are on. Every time you save changes to a workflow, you create a new version. Big changes, like adding or deleting questions, will change the version by a whole number: 1.0 to 2.0, etc. Smaller changes, like modifying the help text, will change the version by a decimal, e.g. 2.0 to 2.1. The version is tracked with each classification in case you need it when analyzing your data.
- **Tasks:** There are two main types of tasks: questions and drawing. For question tasks, the volunteer chooses from a list of answers but does not mark or draw on the image. In drawing tasks, the volunteer marks or draws directly on the image using tools that you specify. They can also give sub-classifications for each mark. Note that you can set the first task from the drop-down menu.
- **Main Text:** Describe the task, or ask the question, in a way that is clear to a non-expert. The wording here is very important, because you will in general get what you ask for. Solicit opinions from team members and testers before you make the project public: it often takes a few tries to reach the combination of simplicity and clarity that will guide your volunteers to give you the inputs you need. You can use markdown in the main text.
- **Help Text:** Add text and images for a pop-up help window. This is shown next to the main text of the task in the main classification interface, when the volunteer clicks a button asking for help. You can use markdown in this text, and link to other images to help illustrate your description. The help text can be as long as you need, but you should try to keep it simple and avoid jargon. One thing that is useful in the help text is a concise description of why you are asking for this information.

Create Tasks

Create tasks with the "Add a task" button. Delete tasks with the "Delete this task" button under the "Choices" box.

Task Content

Tasks can be Questions, Drawings or Transcription. All types have "Main Text" boxes where you can ask your questions or tell users what to draw, as well as provide additional support for completing the task in the "Help Text" box.

Questions

Choices: This section contains all your answers. The key features of this section are:

- **Required:** if you select this, the user has to answer the question before moving on.
- **Multiple:** if you select this, the user can select more than one answer - use this for "select all that apply" type questions.
- **Next Task:** The "Next task" selection (which appears below the text box for each answer) describes what task you want the volunteer to perform next after they give a particular answer. You can choose from among the tasks you've already defined. If you want to link a task to another you haven't built yet, you can come back and do it later (don't forget to save your changes).

Drawing

This section contains all the different things people can mark. We call each separate option a "Tool" and you can specify a label, colour, and tool type for each option. Check out the Aggregation documents to understand how multiple volunteer answers are turned into final shapes for your data analysis. The tool types are:

- **bezier:** an arbitrary shape made of point-to-point curves. The midpoint of each segment drawn can be dragged to adjust the curvature.
- **circle:** a point and a radius.

- **column:** a box with full height but variable width; this tool *cannot* be rotated.
- **ellipse:** an oval of any size and axis ratio; this tool *can* be rotated.
- **line:** a straight line at any angle.
- **point:** X marks the spot.
- **polygon:** an arbitrary shape made of point-to-point lines.
- **rectangle:** a box of any size and length-width ratio; this tool *cannot* be rotated.
- **triangle:** an equilateral triangle of any size and vertex distance from the center; this tool *can* be rotated.
- **grid table:** cells which can be made into a table for consecutive annotations.

Transcription

This section deals with projects which require user-generated text. Tasks can range from adding keywords or extracting metadata to full text transcriptions.

- **Keyword tagging** is helpful when teams want to create a list of all of the things volunteers see in a given image, to make that object more discoverable in a database or online collection. In these cases diversity of opinion is helpful. Setting a retirement rate of 5 to 10 people will help capture diverse opinions.
- **Full text transcription** is more cumbersome and diversity of opinion is less helpful. Teams are usually trying to capture exactly what is on a page, so it will help to set a relatively low retirement rate for each image (i.e. 3 or 5) and be very clear in the tutorial how you would like volunteers to transcribe. Should they preserve spelling and punctuation or modernize it?
- Zooniverse does not currently offer aggregated classifications for text subjects. We can only report what each user transcribed for each subject. Before embarking on a transcription project be sure you have in-house expertise or access to expertise for combining multiple independent transcriptions into a single reading that you could use for research or to upload into a library or museum catalogue or content management system. For more information on Project Builder Data, please visit our Data Digging code repository as well as our Data

processing Talk board.

Linking the workflow together

Now that all the tasks have been created, we've got to string them together by specifying what happens *next*. Set your first task using the "First Task" drop-down menu below the "Add Task" button. Then, using the "Next task" drop-down under the "Choices" box, specify *what comes next*. In question tasks, you can specify different "Next Tasks" for each available answer (provided users can only select one answer).

Multi-Image options: If your tasks require users to see multiple subjects per task (like on Snapshot Serengeti), decide how users will see them. The Flipbook option means users have to press a button to switch between subjects, while separate frames mean that each subject will be visible for the duration of the classification task.

Subject retirement: Decide how many people you want to complete each task. You can change this number at any point (particularly after beta review). We suggest starting out high, between 10 and 20.

Subject Sets

On this page, you can add groups of data to be classified.

To do so, drag and drop items onto the drop zone in the browser and then upload. You can give each set a name so that you can easily distinguish between them.

Subject sets can be pretty powerful, and sometimes complex. You can have a single subject set that you add to over time, or have multiple subject sets, say, from different years or places. You can have different subject sets for different workflows, but you don't have to. You can even have multiple images in a given subject. For more details and advice on creating and structuring subject sets and associated manifests, check out <https://www.zooniverse.org/help/example> and scroll down to DETAILS - Subject sets and manifest details, a.k.a. "What is a manifest?"

Further Help

If you'd like some further information, check out the documentation behind building Kitteh Zoo, that talks you through building this project in the Project Builder.\n\nIf this doesn't help, get in contact with the Zooniverse team via the contact page.

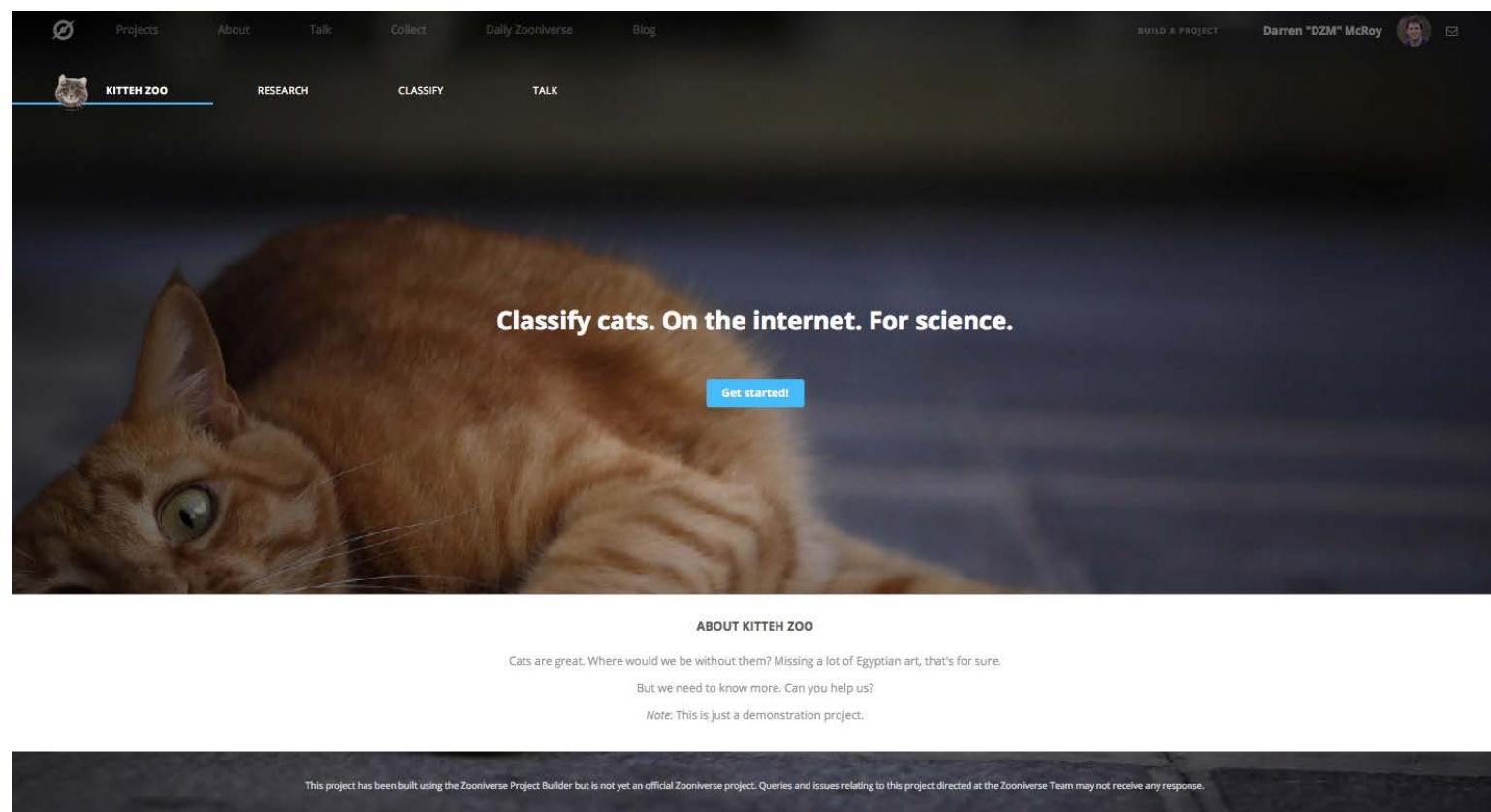
Made with Material for MkDocs

Zooniverse Help

Example Project

So you want to build a project using the Zooniverse Project Builder? This tutorial will help walk you through the process, using Kitteh Zoo as an example. You can explore the actual project.

If you'd like to read more on strategy on building and running a project, such as what to plan for when building your project and what to do after launch, check out the Best Practices pages.



Getting Started

To get started building, go to the Project Builder home page and log in to your Zooniverse account, then click the "Build a Project" button in the top right. Here you

can see all of the projects you own and collaborate on. Click on "Create a project" to start building.

The screenshot shows the Zooniverse Project Builder interface. At the top, there's a navigation bar with links for Projects, About, Talk, Collect, Daily Zooniverse, and Blog. To the right of the navigation is a "BUILD A PROJECT" button, a user profile for Darren "DZM" McRoy, and a mail icon. Below the navigation, there's a list of three projects: Project #2, Project #3, and Project #1, each with an "Edit" and a "View" button. Further down, there are three blue buttons: "Create a new project", "Project building policies", and "How to build a project". The footer contains the ZOONIVERSE logo, links to Projects, Collections, About Us, Education, Our Team, Publications, Contact Us, Talk, Zooniverse Talk, Daily Zooniverse, and Blog, along with social media icons for Facebook, Twitter, and Google+.

Start building: Now you're in the Project Builder itself. This is where the magic happens. On the left-hand side, you've got your main menus: Project, Workflow, and Subjects. These are terms you'll see a lot, and they have specific meanings in the Zooniverse. Project is pretty self-explanatory; Galaxy Zoo, Penguin Watch, and of course, Kitteh Zoo, are all examples of Zooniverse projects that you could build using the project builder. A workflow is the sequence of tasks that you ask volunteers to do, and subjects are the things (usually images) that volunteers do those tasks on.

PROJECT #587

Project details

- Research
- Results
- FAQ
- Education
- Collaborators
- Media
- Visibility

WORKFLOWS

New workflow

SUBJECT SETS

New subject set

Delete this project

Input the basic information about your project, and set up its home page.

Avatar

NAME

Kitteh Zoo

The project name is the first thing people will see about the project, and it will show up in the project URL. Try to keep it short and sweet. Your project's URL is [#/projects/DZM/dzm/kitteh-zoo](#)

DESCRIPTION

Classify cats. On the internet. For science.

This should be a one-line call to action for your project that displays on your landing page. Some volunteers will decide whether to try your project based on reading this, so try to write short text that will make people actively want to join your project.

INTRODUCTION

Cats are great. Where would we be without them? Missing a lot of Egyptian art, that's for sure.
But we need to know more. Can you help us?
***Note:** This is just a demonstration project.

Add a brief introduction to get people interested in your project. This will display on your landing page.

WORKFLOW DESCRIPTION

Add text here when you have multiple workflows and want to help your volunteers decide which one they should do.

TAGS

Tags:

Enter a list of tags separated by commas to help users find your project.

External links

Define your project.

The first thing you'll want to do is fill in some basic information about your project on the Project Details page. Just click and type in the relevant boxes. We've added a short description that will be formatted using the markdown language. The avatar and background image for Kitteh Zoo are in this folder. Download these images to your computer. Now add these images by dragging and dropping or by clicking on the relevant boxes (like in the image above). You can come back and add more details at any time while building your project.

Building a workflow

This is where you build the tasks that volunteers actually do. When you first get to this page, you'll see there is a sample task (specifically a question) already in place.

Projects About Talk Collect Daily Zooniverse Blog BUILD A PROJECT Darren "DZM" McRoy 

PROJECT #587

- Project details
- Research
- Results
- FAQ
- Education
- Collaborators
- Media
- Visibility

WORKFLOWS

Untitled workflow

New workflow

SUBJECT SETS

New subject set

Delete this project

A workflow is the sequence of tasks that you're asking volunteers to perform. For example, you might want to ask volunteers to answer questions about your images, or to mark features in your images, or both.

WORKFLOW TITLE

Untitled workflow

If you let your volunteers choose which workflow to attempt, this text will appear as an option on the project front page.

TASKS

Ask your first question here. (first)

Add task Question Drawing

First task Ask your first question here.

A task is a unit of work you are asking volunteers to do. You can ask them to answer a question or mark an image. Add a task by clicking the question or marking buttons below.

Version 1.1

Version indicates which version of the workflow you are on. Every time you save changes to a workflow, you create a new version. Big changes, like adding or deleting questions, will change the version by a whole number: 1.0 to 2.0, etc. Smaller changes, like modifying the help text, will change the version by a decimal, e.g. 2.0 to 2.1. The version is tracked with each classification in case you need it when analyzing your data.

ASSOCIATED SUBJECT SETS

Choose the set of subjects you want to use for this workflow.

This project has no subject sets. Add an example subject set.

Subject retirement Classification count: 15

How many people should classify each subject before it is "done"? Once a subject has reached the retirement limit it will no longer be shown to any volunteers.

View this workflow

Delete this workflow

MAIN TEXT

Ask your first question here.

Describe the task, or ask the question, in a way that is clear to a non-expert. You can use markdown to format this text.

HELP TEXT

Add text and Images for a window that pops up when volunteers click "Need some help?" You can use markdown to format this text and add images. The help text can be as long as you need, but you should try to keep it simple and avoid jargon.

CHOICES Allow multiple Required

Yes

Next task (End of classification)

The answers will be displayed next to each checkbox, so this text is as important as the main text and help text for guiding the volunteers. Keep your answers as minimal as possible – any more than 5 answers can discourage new users.

The "Next task" selection describes what task you want the volunteer to perform next after they give a particular answer. You can choose from among the tasks you've already defined. If you want to link a task to another you haven't built yet, you can come back and do it later (don't forget to save your changes).

Remove task

We want to create this:

Projects About Talk Collect Daily Zooniverse Blog BUILD A PROJECT Darren "DZM" McRoy 

KITTEH ZOO

RESEARCH CLASSIFY TALK



How many cats are there in this image?

None
 1
 2
 3+

Need some help?

Back Done

This project has been built using the Zooniverse Project Builder but is not yet an official Zooniverse project. Queries and issues relating to this project directed at the Zooniverse Team may not receive any response.

ZOONIVERSE

Projects Projects Collections

About About Us Education Our Team Publications Contact Us

Follow:   

We'll start by replacing the sample text with our question, which asks people how many cats are in the image. We add more answers using the "+" button under the

"Yes" answer. Use the screenshot below to fill in the workflow details (you may need to zoom in!)

The screenshot shows the Zooniverse project settings page for 'Cat Me!'. On the left, there's a sidebar with project details like 'Project #587', 'Research', 'Results', 'FAQ', 'Education', 'Collaborators', 'Media', and 'Visibility'. Under 'WORKFLOWS', 'Cat Me!' is selected, with options to 'New workflow' or 'Delete this project'. Below that is 'SUBJECT SETS' with 'New subject set' and 'Delete this subject set'. The main content area is titled 'A workflow is the sequence of tasks that you're asking volunteers to perform. For example, you might want to ask volunteers to answer questions about your images, or to mark features in your images, or both.' It shows a 'WORKFLOW TITLE' of 'Cat Me!', a note about letting volunteers choose which workflow to attempt, and a 'TASKS' section containing a single task: 'How many cats are there in this image? (first)'. Below this are tabs for 'Add task', 'Question', and 'Drawing'. A note says 'First task' is 'How many cats are there in this image?'. A 'Version 6.10' note explains versioning. An 'ASSOCIATED SUBJECT SETS' section shows 'This project has no subject sets' and a link to 'Add an example subject set'. A 'Subject retirement' section shows a count of 15. A note says 'How many people should classify each subject before it is "done"? Once a subject has reached the retirement limit it will no longer be shown to any volunteers.' At the bottom are 'View this workflow' and 'Delete this workflow' buttons.

MAIN TEXT

How many cats are there in this image?

Describe the task, or ask the question, in a way that is clear to a non-expert. You can use markdown to format this text.

HELP TEXT

Tell us **how many cats** you see. You should include:

- Actual cats (not drawn or simulated)
- Cats of any species

We will ask many people the same question about this image, so don't worry if you aren't absolutely sure. *Just give us your best guess.*

Add text and images for a window that pops up when volunteers click "Need some help?" You can use markdown to format this text and add images. The help text can be as long as you need, but you should try to keep it simple and avoid jargon.

CHOICES Allow multiple Required

Nose	Next task (End of classification)
1	Next task (End of classification)
2	Next task (End of classification)
3	Next task (End of classification)

+ Add choice

We added both text and images into the *Help Text* box using the markdown language (learn more about markdown).

Here is the markdown for the help text:

Tell us **how many cats** you see. You should include:

- Actual cats (not drawn or simulated)
- Cats of any species

We will ask many people the same question about this image, so don't worry if you aren't absolutely sure. *Just give us your best guess.*

Here are some examples of cats:

```
![Sink Cats Relax](http://zooniverse-resources.s3.amazonaws.com/bigblogfiles/cat_demo/cat_bloonet.jpg)
![Only 1 of these cats counts.](http://zooniverse-resources.s3.amazonaws.com/bigblogfiles/cat_demo/cat_valentina_a.jpg)
![Wet kitteh iz not amused](http://zooniverse-resources.s3.amazonaws.com/bigblogfiles/cat_demo/cat_joeltelling.jpg)
```

And here are some examples of not-cats (you can ignore these):

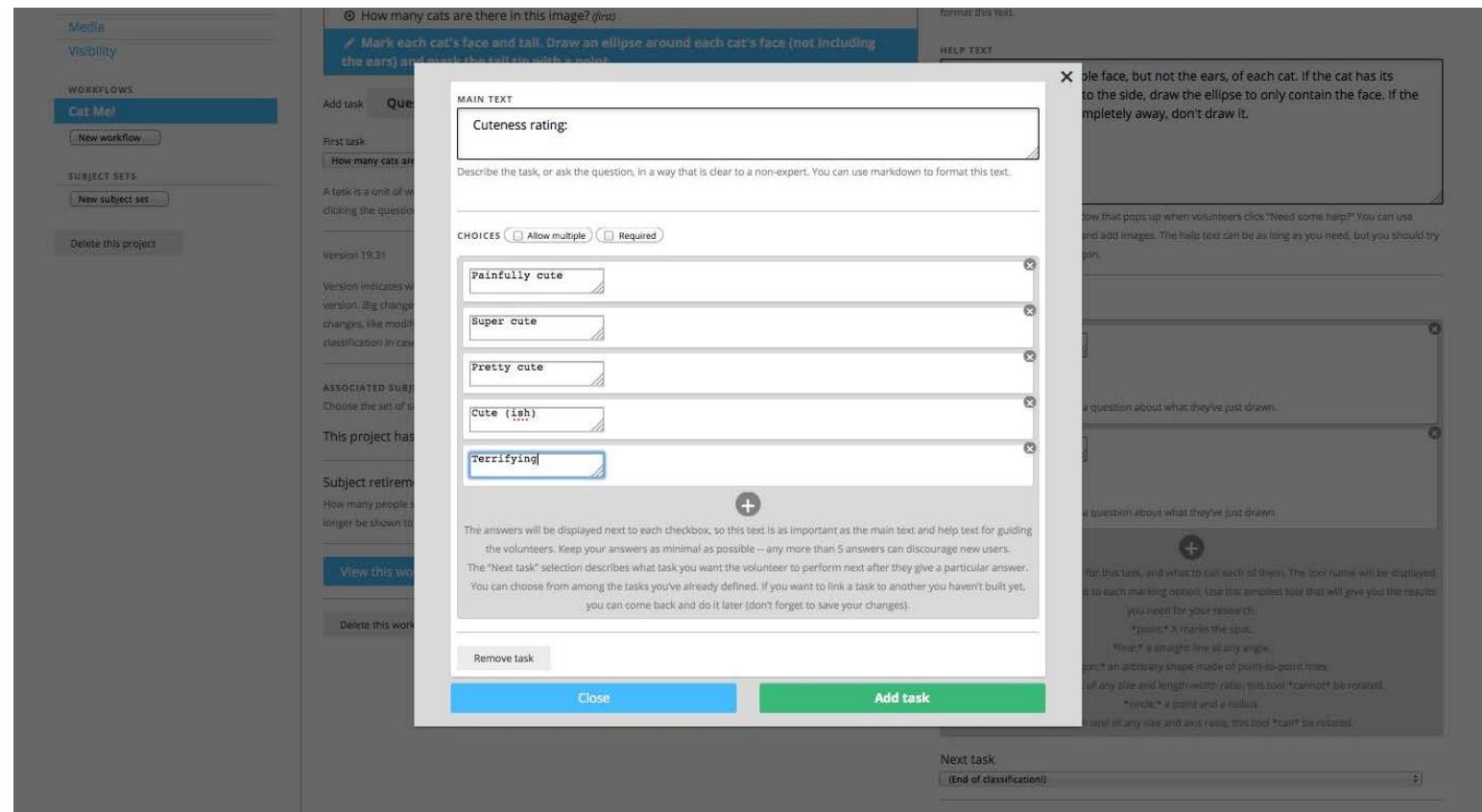
! [A sample of things that are not cats.] (http://zooniverse-resources.s3.amazonaws.com/bigblogfiles/cat_demo/notcats.png)

We can set subsequent tasks to depend on the answer to this question. Right now we haven't made any other tasks, so the only option is "End of Classification." Once we create more tasks, we'll go back through and link them. Note that this question is **required** (people can't move on until they've answered it) and only one answer is allowed.

Now we want to draw circles around the cat's faces and mark a point on their tails. Why? Because we can. (For your own project you'd obviously want to think carefully about the reasons for adding tasks to a workflow, and what you want to get from the answers/marks.)

So under the **Task** list, we'll click on **drawing**. We're asking folks to draw (with ellipses) around the cats' faces, as well as mark their tail tips with a point. We've changed the color on the Cattail points so they stand out more too. As usual, the main text gives people basic instructions on what we want, and the help text provides some more explanation on how to do the task.

In addition to marking all the cat faces, we want to know just how cute they are. So every time someone marks a cat-face, we've added a pop-up question to ask just that. Add this question by clicking on the *sub-tasks* button below the *Type* and *Color* task specifications.



When building your own project, you can combine any number of tasks in any order. You can start with a drawing task instead of a question. You can add sub-tasks for any drawing tool you make.

In general, keep in mind that people are more likely to complete more classifications if the workflow is short and simple. Try to keep the workflow as simple as possible to achieve your research goals, and definitely try to only request tasks that cannot be accurately accomplished by automated methods.

Linking the workflow together

Now that all the tasks have been created, we've got to string them together by specifying what happens *next*. Right now, this means you kind of need to work backwards. The drawing task is the last task in this workflow, so we'll leave the "Next

"Task" button as the default "end of classification." But we'll have to go back to our first question.

The first question, "How many cats are in this image?" only allows one answer, so you can specify the next task depending on the answer. If folks say "None" for the number of cats, the classification ends. But if they say there's at least one cat, then they go on to the next question.

Upload subjects

To really get started building a project, you need images to work with. Normally you would add your own images by clicking on the "New Subject Set" button on the left hand side of the screen. This is one of the trickier steps in project creation -- for the purposes of this tutorial you can simply copy the Kitteh Zoo subject set, but check out the next section "Uploading subjects -- the nitty gritty" if you want to practice the full approach. To do this go to the workflow you created and under the *associated subject set* section click on *add an example subject set*. You should now see the 'kittehs' subject set selected.

The screenshot shows the Zooniverse Project Builder interface. On the left, a sidebar lists project details, research results, FAQ, education, collaborators, media, and visibility. A 'WORKFLOWS' section is expanded, showing a 'Cat Me!' workflow with a 'New workflow' button. Below it are 'SUBJECT SETS' and a 'Delete this project' button. The main workspace is titled 'PROJECT #587' and contains a 'WIDGETS' section with a 'Workflow Title' set to 'Cat Me!', a 'Tasks' section with a note about letting volunteers choose which workflow to attempt, and a 'Help Text' section with a note about asking many people the same question. A 'MAIN TEXT' box contains the question 'How many cats are there in this image?'. A 'CHOICES' section includes options for 'Allow multiple' and 'Required'. A 'OPTIONS' section lists 'None', 'Next task (End of classification!)', and '1', with a note about marking each cat's face and tail.

CONGRATULATIONS!

You should have successfully created Kitteh Zoo! To view it, go back to the *Build a Project* page (by clicking the button in the top right of the page) and then click on the view button next to the new project you have just made.

Project #2	Edit	View
Kitteh Zoo	Edit	View
Project #3	Edit	View
Project #1	Edit	View

[Create a new project](#)
[Project building policies](#)
[How to build a project](#)

Uploading subjects - the Nitty Gritty

When you actually build your own project, there won't be an example set of images already loaded for you. Get started uploading a set of subjects for your project by clicking on the "New Subject Set" button on the left hand side of the screen. That will

bring you to the *Subject Uploader*.

If every subject you want classified is a single image, and if the image file name contains enough information for you to be able to precisely match the subject with the rest of your data later, then this is pretty easy: you can just drag and drop the images into the Subject Uploader (or click within the "Upload Subjects" box and choose the files in the pop-up menu) and it will automatically detect that it needs to make subjects out of each file.

Much of the time, though, project builders need to keep some data associated with their images in the classification interface. For example, in Kitteh Zoo all the images are Creative Commons licensed and need to have attribution attached when they are displayed. With a simple file called a *manifest*, we can associate the needed information with each image. Below you can find more details about the manifest in the section *DETAILS - Subject sets and manifest details, a.k.a. "What is a manifest?"* For now we will assume the manifest is created and the files are ready. You can see an example manifest file in the "Kitteh" zip file.

People often find it easiest to have all of the subject images in a single folder along with a manifest file; you will upload both at the same time. However you organize your files, pay attention to the details and keep records, as it may be important for your data analysis later (e.g. when you download your classifications of subjects and need to know how to correctly link those back to all the rest of your metadata).

Click on the "Upload Subjects" box and navigate to the "Kitteh" folder you downloaded. The easiest thing to do is simply hit *cmd + a* (on Mac, or *ctrl + a* on Windows) to select everything in that folder. Note that the subject uploader ignores the excel file and the other folders. Click "open" to select those images and the manifest file for upload. The project uploader now indicates that the "Cat project manifest.csv" has 30 subjects for upload. (If we had more subjects to upload, we'd do them in batches of no more than 1,000 at a time.)

Click "Upload" to start the process. When everything is uploaded, you'll see a list of all of the subjects. The numbers to the left are unique identifiers for each subject, and

the icons to the right let you preview or delete each subject.

Subject sets can be pretty powerful, and sometimes complex. You can have a single subject set that you add to over time, or have multiple subject sets, say, from different years or places. You can have different subject sets for different workflows, but you don't have to. You can even have multiple images in a given subject. For more details and advice on creating and structuring subject sets and associated manifests, check out the details section below.

PROJECT BUILDER'S MANUAL – THE DETAILS

- **Project:** This holds all your project level details. The project name, the people involved, and all the extra content (e.g. text and pictures) you want to share, are all here.
- **Project Details:** This is your project's "behind the scenes" home page. Start off by naming and describing your project, add a logo and background image.
- **Research Case, FAQ, Results, and Education:** You can add a lot of information in these pages to help volunteers better understand the motivation for your project, the best approaches for classifying, and the outcomes of your project.
- **Collaborators:** Add people to your team and specify what their roles are so that they have the right access to the tools they need (including access to the project before it's public).
- **Workflows:** A workflow is the sequence of tasks that you're asking volunteers to perform. For example, you might want to ask volunteers to answer questions about your images, or to mark features in your data, or both. The workflow is where you define those tasks and set out the order in which the volunteers will do them. Your project might have multiple workflows (if you want to set different tasks for different image sets).
- **Subjects:** A subject is a unit of data to be analyzed. A subject can include one or more images that will be analyzed at the same time by volunteers. A subject set

consists of a list of subjects (the "manifest") defining their properties, and the images themselves. Feel free to group subjects into sets in the way that is most useful for your research. Many projects will find it's best to just have all their subjects in 1 set, but not all: larger projects often find it essential to group subjects into multiple sets.

DETAILS - Workflows

Note that a workflow with fewer tasks will be easier for volunteers to complete. We know from surveys of our volunteers that many people classify in their limited spare time, and sometimes they only have a few minutes. Longer, more complex workflows mean each classification takes longer, so if your workflow is very long you may lose volunteers.

Workflow Name: Give your workflow a short, but descriptive name. If you have multiple workflows and give volunteers the option of choosing which they want to work on, this name will appear on a button instead of "Get started!"

Version: Version indicates which version of the workflow you are on. Every time you save changes to a workflow, you create a new version. Big changes, like adding or deleting questions, will change the version by the number to the left of the ".": 1.0 to 2.0, etc. Smaller changes, like modifying the help text, will change the integer to the right of the ".", e.g. 2.0 to 2.1. The version is tracked with each classification in case you need it when analyzing your data. *Note:* the version looks like a decimal number, but *it is not*: it is a string with two numbers separated by a "." character. So, for example, version 2.1 is not the same as version 2.10.

Tasks: There are two main types of tasks: questions and drawing. For question tasks, the volunteer chooses from a list of answers but does not mark or draw on the image. In drawing tasks, the volunteer marks or draws directly on the image using tools that you specify. They can also give sub-classifications for each mark. Note that you can set the first task from the drop-down menu.

Main Text: Describe the task, or ask the question, in a way that is clear to a non-

expert.

The wording here is very important, because you will in general get what you ask for. Solicit opinions from team members and testers before you make the project public: it often takes a few tries to reach the combination of simplicity and clarity that will guide your volunteers to give you the inputs you need.

You can use markdown in the main text.

Help Text: Add text and images for a pop-up help window. This is shown next to the main text of the task in the main classification interface, when the volunteer clicks a button asking for help. You can use markdown in this text, and link to other images to help illustrate your description. The help text can be as long as you need, but you should try to keep it simple and avoid jargon. One thing that is useful in the help text is a concise description of why you are asking for this particular information. This help information often overlaps with information in the Field Guide, but it is an opportunity to provide extra detail about each specific task.

DETAILS - Project Details:

Name: The project name is the first thing people will see and it will show up in the project URL. Try to keep it short and sweet.

Avatar: Pick an avatar image for your project. This will represent your project on the Zooniverse home page. It can also be used as your project's brand. It's best if it's recognizable even as a small icon. To add an image, either drag and drop or click to open your file viewer. For best results, use a square image of not more than 50KB, but at minimum 100x100 pixels.

Background: This image will be the background for all of your project pages, including your project's front page, which newcomers will see first. It should be relatively high resolution and you should be able to read text written across it. To add an image, either drag and drop or click to open your file viewer. For best results, use images of at least 1 megapixel, no larger than 256 KB. Most people's screens are not

much bigger than 1300 pixels across and 750 pixels high, so if your image is a lot bigger than this you may find it doesn't look the way you expect. Feel free to experiment with different sizes on a "typical" desktop, laptop or mobile screen.

Description: This should be a one-line call to action for your project. This will display on your landing page and, if approved, on the Zooniverse home page. Some volunteers will decide whether to try your project based on reading this, so try to write short text that will make people actively want to join your project.

Introduction: Add a brief introduction to get people interested in your project. This will display on your project's front page. Note this field (renders markdown) [<http://markdownlivepreview.com/>], so you can format the text. You can make this longer than the Description, but it's still probably best to save much longer text for areas like the Research Case or FAQ tabs.

Checkbox: Volunteers choose workflow: A workflow is a set of tasks a volunteer completes to create a classification. Your project might have multiple workflows (if you want to set different tasks for different image sets). Check this to let volunteers select which workflow they want to work on; otherwise, they'll be served workflow-subject pairs randomly.

Checkbox: Private project: On "private" projects, only users with specified project roles can see or classify on the project. We strongly recommend you keep your project private while you're still working out its details. Share it with your team to get feedback by adding them in the Collaborators area (linked at the left). Team members you add can see your project even if it's private. Once your project is public, anyone with the link can view and classify on it.

DETAILS - Additional Content

Research, FAQ, Results, and Education: These pages are where you really get to share all the cool things about your project. All of these pages use Markdown (see link above) to format text and display images.

This page is for you to describe your research motivations and goals to the volunteers. Feel free to add detail, but try to avoid jargon. This page renders markdown, so you can format it and add images via the Media Library and links. The site will show your team members with their profile pictures and roles to the side of the text.

RESEARCH

This is a demonstration project using the project building tools of the Zooniverse.
It is also about cats.

Research: Explain your research case to your audience here in as much detail as you'd like. This page displays no matter what, since explaining your motivation to volunteers is critical for the success of your project!

Results: Once your project has hit its stride, share the results of your project with your volunteers here. Volunteers really value feedback about how their inputs were used to help the research. This page will only display if you add content to it.

FAQ: Add details here about your research, how to classify, and what you plan to do with the classifications. This page can evolve as your project does so that your active community members have a resource to point new users to. This page will only display if you add content to it.

Education: If you are a researcher open to collaborating with educators you can state that here, include educational content, and describe how you'd like to help educators use your project. Also, if your project is primarily for educational purposes you can describe that here. This page will only display if you add content to it.

DETAILS - Media

You can upload your own media to your project (such as example images for your

help pages) so you can link to it without an external host. To start uploading, drop an image into the box (or click it to bring up your file browser and select a file).

Once the image has uploaded, it will appear above the "Add an image" box. You can then copy the markdown text beneath the image into your project, or add another image.

DETAILS - Visibility

This page is where you decide whether your project is public and whether it's ready to go live. For more information on the different project stages, see our project builder policies.

DETAILS - Collaborators

Add people to your team and specify what their roles are so that they have the right access to the tools they need (including access to the project before it's public).

The screenshot shows the 'Collaborators' section of the Zooniverse Project Builder. On the left, there's a sidebar with links like 'Project details', 'Research', 'Results', 'FAQ', 'Education', 'Collaborators' (which is highlighted in blue), 'Media', 'Visibility', 'Workflows', 'Cat Me!', 'New workflow', 'Subject sets', 'Kittehs', 'New subject set', and 'Delete this project'. The main content area has a 'PROJECT OWNER' section stating 'You are the project owner.' Below that is a 'COLLABORATORS' section listing three users: 'brooke' (Collaborator, Expert, Researcher, Moderator, Tester), 'aliburchard' (Collaborator, Expert, Researcher, Moderator, Tester), and 'zookeeper' (Collaborator, Expert, Researcher, Moderator, Tester). There's also an 'ADD ANOTHER' section with a 'Username:' input field and a list of role checkboxes: 'Collaborator' (selected), 'Expert', 'Researcher', 'Moderator', and 'Tester'. A note below explains each role: Collaborators have full access, Experts can enter gold mode, Researchers are marked as scientists, Moderators have extra privileges, and Testers can view and classify. At the bottom is a green 'Add user role' button.

Owner: The owner is the original project creator. There can be only one.

Collaborator: Collaborators have full access to edit workflows and project content,

including deleting some or all of the project.

Expert: Experts can enter “gold mode” to make authoritative gold standard classifications that will be used to validate data quality.

Researcher: Members of the research team will be marked as researchers on “Talk”

Moderator: Moderators have extra privileges in the community discussion area to moderate discussions. They will also be marked as moderators on “Talk”.

Tester: Testers can view and classify on your project to give feedback while it’s still private. They cannot access the project builder.

Translator: Translators will have access to the project builder as well as the translation site, so they can translate all of your project text into a different language.

DETAILS - Subject sets and manifest details, a.k.a. "What is a manifest?"

The condensed answer:

A manifest is a file that tells our software how to combine the images you have into units of data (subjects) to be classified. The manifest also allows you to link your classifications back to the rest of your data. A manifest is formatted as a CSV file with 1 line per subject, with the names of images to be associated with a subject on each row (with additional information often included in other fields as well). There is an example in the "Kitteh" zip file.

The full answer:

What we call a "manifest" is really just a plain text file with a specific format to each line.

To understand the format, let's start with the first few lines from the Kitteh Zoo manifest:

```
image,!origin,link,attribution,license,#secret_description
6672150457_420d61007d_b.jpg,Flickr,https://www.flickr.com/photos/aigle_dore/6672150457,Mo
Brenn,Creative Commons - share adapt attribute,sleepy striped kitteh is unsuspecting of
paparazzi
8300920648_d4a21bba59_z.jpg,Flickr,https://www.flickr.com/photos/aigle_dore/8300920648,Mo
Brenn,Creative Commons - share adapt attribute,grandfather kitteh has ear hair. a lot
of it
6713782851_82fc8c73e5_z.jpg,Flickr,https://www.flickr.com/photos/hellie55/6713782851,heha
Commons - share adapt attribute,juvenile kittehs practice break-in at the catnip
factory
```

The first line of the file is a header line that specifies the name of each of the manifest fields. In this case, our manifest has 6 fields (or columns), called "image", "!origin", "link", "attribution", "license" and "#secret_description". They are separated by commas: this is what's known as a "comma separated values" file, or CSV file.

After the first line, each row of the file contains information about 1 subject. The first field, which aligns with the "image" header, contains the name of the image that's associated with that subject. This field is critically important as it tells the project builder which images to make into subjects.

All the other fields are optional, but in general having more information in the manifest is better. Most projects include additional information in the manifest that helps them match the classifications and subjects to the other data they need for their research. For example, many projects include a separate ID that they have generated and which can be used later as a unique identifier to match classification data to their own metadata (for other projects, the unique identifier is the filename). The additional information in the manifest can also be made available to volunteers as they classify or in the Talk discussion tool (some very keen volunteers find this extremely useful). Any fields with names that begin with "#" or "//" will never be shown to volunteers, such as the "#secret_description" field in Kitteh Zoo. These hidden fields will still be returned to you in the classification file, so you can use these to include information helpful to your research without worrying about whether it might affect the classifications themselves. There might be information that could be helpful for the volunteers in discussions or further exploration in the Talk discussion tool, but could bias the classifications if visible to volunteers while classifying. Any fields with names that begin with "!" will not be accessible to volunteers in the

classification interface, but *will* be available on Talk after classification (such as the "origin" field above). Information in fields that *don't* begin with a "#", "//", or '!' will always be accessible to volunteers in both the classification interface and the Talk discussion tool.

For now, let's assume you're just including minimal information, like:

```
my_own_id,the_image
1,kitteh_in_box.jpg
2,kitteh_stalking.jpg
3,kitteh_losing_balance.jpg
```

Note the field names have changed from the previous example. That's because, aside from marking whether a field is hidden or not, it doesn't actually matter to the Zooniverse what the fields are called (or what order they're in), so you can name and order them according to whatever works best for your project.

Using a manifest CSV file also makes it very easy to create subjects with multiple images:

```
my_own_id,image1,image2
1,kitteh_in_box.jpg,kitteh_eating_box.jpg
2,kitteh_stalking.jpg,kitteh_pounced.jpg
3,kitteh_losing_balance.jpg,kitteh_falling_off_sofa.jpg
```

If you upload this manifest plus the 6 images named in it, the Zooniverse software will create 3 subjects with 2 images each. When these subjects come up in the classification interface, volunteers will be able to flash between the images or switch between them manually. Later on, when these subjects are classified, the subject_id assigned by the project builder can be used with a subject export to match each classification back to this information.

You can create a manifest file in a simple text editor (such asTextEdit or Notepad), although this method is prone to errors like missed or extra commas. People often find it easiest to create manifest files using spreadsheet software such as Google Sheets, iWork Numbers or Microsoft Excel. Creating and maintaining a manifest using a spreadsheet makes the manifest easy to read, and you can export it to CSV format

when you're ready to upload your subjects. You can also open existing CSV files in spreadsheet software.

Note: if you have a large subject set it may be cumbersome to manually create a manifest. We suggest using a command-line or other tool to copy-paste a directory list of files into a spreadsheet to help you get started.

Made with Material for MkDocs

Zooniverse Help

Glossary

A collection of definitions for terms that are used across the Zooniverse. The terms are split into three different categories:

- General Terms
- People
- Project-Specific Terms

If you'd like to see a definition that currently isn't on this page, get in touch and we'll look into adding it.

General Terms

- **Classification** - A classification is all the data associated with a volunteer's response to an item of data (or subject) they're presented with whilst going through a project. In essence, a classification is the core unit of human effort produced by the Zooniverse community.
- **Collection** - A collection is similar to favoriting an image, but allows volunteers to additionally link together groups of subjects. These collections can be set to either public or private, and you can see all of the public collections on the Collections page. Volunteers can only add to their own collections, and not those of others.
- **Favorites** - If a volunteer finds an image on a project that they like and wants to be able to see again in the future, they can mark it as a 'Favorite' by clicking the heart icon underneath it. These images then show on the volunteer's Favorites page.

- **Newsletter** - a brief email message about your project sent to your project's registered volunteers or to a subset of registered Zooniverse volunteers. Newsletters are a great way to update volunteers on the your project's progress and also help give the daily classification rate a boost. If you'd like a newsletter to be sent out for your project, get in touch with Grant Miller, the Zooniverse Communications Lead.
- **Project** - A project is a way for our volunteer community to engage with a specific research goal or question, using data provided by the researchers. This gives the researcher data to work with and helps progress science. It takes shape in the form of a website, which includes the main classification interface and the Talk discussion tool.
- **Project Builder** - This is the web tool that researchers use to create Zooniverse projects. There is documentation to help with this process on the Project Builder page.
- **Zooniverse** - the Zooniverse encompasses the team, website, projects and our codebase. The Zooniverse is an open web-based platform for large scale citizen science research projects, capable of supporting tens of thousands of simultaneous users. The Zooniverse hosts the largest collection of online citizen science projects in the world, supporting over 1.5 million registered users and containing projects in astronomy, ecology, humanities, physics, and beyond.

People

- **Collaborator** - Collaborators are people with permission on a project with full access to edit workflows and project content, including deleting some or all of the project. You can add a collaborator to a project you own through the Collaborators section of the Project Builder.
- **Display name** - This is the name that shows up in the Talk boards. You can edit this at any time by going to your Settings page and entering into the appropriate box what you'd like to change it to.
- **Expert** - Experts can enter "gold mode" to make authoritative gold standard data

that will be used to validate data quality. You can add an expert to a project you own through the Collaborators section of the Project Builder.

- **Moderator** - this is a role given to allocated by the project owner, and typically given to a member of the project's volunteer community. The moderator is given extra privileges in the project's Talk discussion tool (such as the ability to create new discussion boards), and they oversee and help moderate the Talk discussions. You can add a moderator to a project you own through the Collaborators section of the Project Builder.
- **Project owner** - This is the person who has built the project using the Project Builder. They are typically the researcher working with the data.
- **Real name** - this is the name that will be published alongside any research which involves Zooniverse data from projects you've been involved with.
- **Researcher** - Members of a project's research team will be marked as researchers on Talk. You can add a researcher to a project you own through the Collaborators section of the Project Builder.
- **Testers** - Testers are people who can view and classify on your project to give feedback while it's still private. They cannot access the project builder. You can add testers to a project you own through the Collaborators section of the Project Builder.
- **User name** - This is the name that you log in with. This is currently permanently associated with your Zooniverse account and your classifications.
- **Volunteer** - the Zooniverse's preferred term for a member of the public who is participating in and contributing to a Zooniverse project.

Project-Specific Terms

- **Aggregation** - Each subject in a Zooniverse project is independently classified by multiple people. Aggregation is the process of combining these multiple assessments together.

Annotation - Annotations are markings, drawings, answers or data about a subject provided by volunteers as part of the classification process.

- **Classification interface** - This is the web interface where volunteers are review subjects (subjects being data that volunteers are presented to in projects) and perform the desired project assessments and tasks. Researchers can add in various things to this interface such as tutorials on how to use it, and also mini-courses that give you information about the science behind the project. Each workflow has its own separate classification interface. You can access the classification interface through the main landing page of the project.
- **Export** - An export is how to get the volunteer classifications and other relevant information about your project that is stored in the Zooniverse databases. It is essentially a data dump from the Zooniverse database. Project owners and collaborators can request data exports from their projects through the Project Builder. Project data is supplied in CSV/JSON format and Talk data is supplied in JSON format, and the two most commonly used exports are the subject data export, providing all the information stored about the subjects you've uploaded for your project, and the classification export, which has information stored for each project classification. You can request a data export by going to the Project Builder, going to the Data Exports tab and then selecting from where whichever data export you'd like, provided you have the right permissions.
- **Field Guide** - A field guide is a place to store general project-specific information that volunteers will need to understand in order to complete classifications and talk about what they're seeing. It's available anywhere in your project. It's different to the tutorial in that the information is generally more about the science behind it, and is a way of sharing knowledge with your volunteers. Field guides are optional and generally contain more information than tutorials.
- **Gold standard data** - This is data from classifications made by 'Experts'. 'Expert' is a role assigned by the project owner, and their data can be used as a standard to compare the rest of the data against. There is currently an experimental feature which allows you to create a training set and provide in-classification feedback using gold standard data.

- **Landing page** - front page of your project's website. This is where people are directly when they go to your project's url. The landing page is always accessible from your project's website by clicking on the project's avatar.
- **Marks** - for drawing tasks, volunteers are asked to highlight content on an image by drawing circles, boxes, etc. around it. These drawings are referred to as marks.
- **Mini-course** - this is an educational course that is embedded into the project. It is designed to help teach users more about the science behind projects that they are interacting with. Gravity Spy has one such mini-course.
- **Project Avatar** - the project's logo. This image shows in the top left corner of for the project website. If the project becomes Zooniverse approved, the project avatar will also be listed on the Zooniverse projects page
- **Project Tags** - These help define which field of research your project belongs to, and determines which category your project will sit under on the Projects page in the categories section (it will still appear on the main Project page regardless). Users can also search by tag to find projects.
- **Subject** - The chunk of data/thing a volunteer on a Zooniverse project is being presented with and asked to review and analyze. It typically is an image, graph, photo, audio recording, video, or a collection of these different things.
- **Subject Set** - This is a group of subjects (subjects being data that volunteers are presented to in projects). Subjects are uploaded into subject sets through the Project Builder, and it is subject sets that can be linked to workflows in order to get the desired subjects showing on your project's webpage. You can group subjects into subject sets however you wish. You might want to group Subjects together, for example to represent a season's worth of images in Snapshot Serengeti or a particular cell dye staining as in Cell Slider.
- **Talk** - is the object-orientated discussion tool associated with your project. Talk enables volunteers to comment on the subjects they've reviewed and promotes discussion amongst the volunteer community. Talk is also a place where the research team and project volunteers can interact. Talk has a series of message

boards for longer discussions. Additionally, each subject has a dedicated page on its project Talk where a registered volunteer can write a comment, add searchable Twitter-like hashtags, or link multiple subjects together into groups called collections

- **Talk Tags** - Tags are used to help note something as relating to a particular topic. For example, one way of using it would be the following: you've found an image that you think would be good for the Daily Zoo, and so you post it in Talk and then include the tag #dailyzoo in the post. This makes it easier for it to be found as a suggestion, because you can do a search for a particular tag using the search bar. A list of popular tags is displayed on the right-hand side of the Talk page and clicking one will take you to instances of that tag in the Talk boards. So for example if you clicked on #dailyzoo, you'd then be shown all the different suggestions that people have made for it that they've tagged.
- **Tasks** - A task could be listing how many of a particular thing a volunteer sees in an image and then drawing circles around them, identifying the various animals they can see in an image or identifying where abouts in an image something is. There are a wide variety of tools to help create a wide variety of different tasks in the Project Builder tool. One or more tasks make up a workflow.
- **Drawing Task** - A task where the volunteers are asked to directly highlight or mark something on an image (e.g. drawing a circle around a penguin if visible in the image presented)
- **Question Task** - A task where the volunteers are asked to assess the image and respond to a multiple choice question. In the Project Builder, a question task can allow the volunteer to choose a single response or select multiple answers to the question posed.
- **Sub-task** - Sometimes when you are asked to do a task, such as drag a circle around an element in a picture, you are then asked a further task and what is contained in the circle. This is a subtask. For example, you may be asked to circle penguins in an image, a sub-task would be identifying whether the penguin circled is an adult or chick.

Survey task - A survey task is a task where you identify something by selection from many options and then are asked a variety of questions about what you've just identified, like behaviour, number or color. For example, you could be asked to identify an animal in an image and then answer questions on how many legs you can see, which way it is facing and whether it is an adult or baby. An example of such a project is Camera CATalog.

- **Tools** - tools enable volunteers draw or highlight a particular area of a subject image presented on the classification interface (e.g. draw a circle, draw a line, place a pointer). You can design your project such that a single task in the Project Builder can have one or more tools available for volunteers to mark or identify different features found in your subject images.
- **Tutorial** - is a very brief walk-through explaining the main goals and aims of your project. It quickly introduces and explains to the volunteer how to do the requested tasks. This is created in the project builder and is presented to first-time volunteers of your project. Project tutorials are optional.
- **Workflow** - This is a series of tasks and assessments that a volunteer is asked to do when presented with data in a Zooniverse's project classification interface. This can be either one task or multiple tasks, depending on the project.
- **Active workflow** - these are workflows that are currently being presented to volunteers. They can't be edited whilst they are active. You can make a workflow active and inactive by ticking the relevant box on the Visibility tab on the Project Builder page.
- **Default workflow** - The default workflow is the workflow that appears if the volunteers doesn't choose one from the offered list on the front page of the project but just clicks on the Classify tab. If a project has more than one active workflow, the project owner and project collaborators can choose which one will be the default workflow.
- **Inactive workflow** - Inactive workflows are ones that aren't currently being shown to volunteers, but can be edited. You can make a workflow active and inactive by ticking the relevant box on the Visibility tab on the Project Builder

page.

- **Zooniverse approved project** - this is a project that has undergone beta testing and been approved by the Zooniverse team. Details of this process can be found on the Policy Page. Zooniverse approved projects that are currently live can be seen on the Projects page.

Made with Material for MkDocs

Zooniverse Help

Lab policies

How to Launch your Project and Zooniverse Policies

This page is for people who have used our Project Builder, and would now like to share their project with the Zooniverse community or their own crowd of collaborators or volunteers.

How to Launch your Project

A live project can exist in three different states, it may be a: 1. Public Project 2. Review Project 3. Zooniverse Project

Each of these project states is discussed below.

1. Public Project Do you have your own crowd of volunteers, or only want members of your collaboration to classify your data? You can make your project public simply by sharing your project's URL with those you want to take part.

If you're planning a big public launch, or you think your project might attract significant attention, you should let us know so we can support you properly. Please note, there is a limit of 10,000 subjects per user - contact us if you'd like to host more.

2. Review Project If you would like the Zooniverse volunteer community to contribute to your project, the first step is to apply for review. You can do this in the project builder by clicking the "Apply for review" button within the "Visibility" section. At this point, your project is considered a "Review Project".

Your project will then be subject to two stages of review – firstly, the Zooniverse

team will check your project complies with Zooniverse rules (see below). Typically, we will respond to you within two weeks. If successful, your project will then be shared with a group of volunteer project testers who will provide feedback. We will contact you when a slot becomes available for volunteer review (this currently takes one to four weeks). Once your project has been shared with our volunteer project testers, the majority of project feedback is typically received within a week. This will be made immediately accessible to you.

During this process you should address the formal feedback provided by the volunteer project testers, as well as any comments on your project's Talk discussion pages. Any technical issues can be reported to us via the GitHub repository, but please consider carefully any comments on workflow design.

It is important you review the results produced at this stage to ensure they are of sufficient quality for your research. If they aren't, please consider re-approaching your project design, make any changes, and submit for review again.

3. Zooniverse Project Once your project has been successfully reviewed and you have a final version of your project, you can apply for your project to be launched as an official Zooniverse project. To do this, notify the team by email at contact@zooniverse.org.

Your project will then undergo a final review process. This may include requesting information from you about your plan for using the results, and examining the results from the review phase. We may also carry out a peer review of your project. This review process can take a variable amount of time, from weeks to months.

Once the project has passed review, it will appear in the Zooniverse project list. We will consider including your project in our Zooniverse Newsletter and other promotional material. If you want your project to be promoted, note that we favour projects whose researchers are proactive in engaging with their community, on the Talk discussion forums and elsewhere.

If you'd like your project to launch on a particular day, perhaps to coincide with a press release, please get in touch.

Zooniverse Policies

At present, we provide the Zooniverse Project Builder software and hosting for free.

We reserve the right to remove content for any reason whatsoever. We will remove content and projects where:

- The content is not legal.
- The content is likely to cause offense, or is suitable only for an adult audience.
- The copyright on material uploaded to the site is not clear; please only use content you have the right to use.

We reserve the right to decide which projects appear on the main project page and which are promoted to the Zooniverse community. In particular, if you have a project that is very close to an existing Zooniverse project, please contact us to discuss.

Projects promoted to the Zooniverse community **must**:

- Have the goal of producing useful research; your study needs to be well designed, and you must intend to analyze and write up your results as a formal publication.
- Make their classification data open after a proprietary period, normally lasting two years from project launch.
- Communicate research findings to their communities, via open access publication, a blog or elsewhere.
- Acknowledge Zooniverse in any publications. Please use the following text: "*This publication uses data generated via the Zooniverse.org platform, development of which is funded by generous support, including from the National Science Foundation, NASA, the Institute of Museum and Library Services, UKRI, a Global Impact Award from Google, and the Alfred P. Sloan Foundation.*"
- Report publications using Zooniverse-produced data to us via this form.

The Zooniverse is not typically a platform for surveys which ask for audience opinion, or about them; our projects normally invite volunteers to actively assist in research. Therefore, we typically do not support survey-style projects. If you aren't sure

whether your project goals fit this model, please contact us to discuss.

Any changes made to the standard Zooniverse policies for a particular project **must be** (1) approved by the Zooniverse team in writing, in advance of launch, and (2) communicated to volunteers via the Announcement Banner, which must be displayed for the duration of the project.

In addition, participants reasonably expect to have access to the subjects (e.g. images) after the project has finished, in order to provide classification data and discussions with context. Exceptions to this requirement need to (1) be approved by the Zooniverse team in writing, prior to launch, and (2) must be communicated to volunteers throughout the project via the Announcement Banner, which must be displayed for the duration of the project.

If the project dataset is unable to be shared publicly and the data needs to be removed from the Zooniverse platform to comply with institutional practice, the requirements are as follows:

1. The project data will not be removed until the entire project is complete.
2. Upon the project's completion, the project page will be replaced by an Archive page (see: <https://anno.tate.org.uk/#/>), which must include the following information:
 - If available, a link to the project data (e.g. images hosted behind a paywall)
 - A link to the classification data when available (i.e. the project results), nb: this can be delayed for up to 2 years to allow teams to publish results, etc., in line with project policies detailed above
 - Links to any publications about the project or using project data

If you have any questions relating to the content of this page, please get in touch.

Made with Material for MkDocs

Zooniverse Help

Building, Launching, and Running Your Zooniverse Project:

Best Practices for Engagement & Success

Introduction: "Best Practices," What and Why?

Welcome! If you are reading this, you are probably interested in starting a Zooniverse project using the Project Builder. Perhaps your images/subjects are prepared, you have a compelling research case, and you know what information you are looking for... all that you need now is to process those images and turn them into data, so that you can make the breakthrough of the century in your field. We say, great—you've come to the right place!

First, however, take a moment to consider the Zooniverse slogan: **People-Powered Research.**

Why is that so important? Well, because it's true. Simply put, **volunteers are the lifeblood of a Zooniverse project**, and only with their efforts can your project accomplish its research goals. Your project's amateur researchers are not a data-processing machine: they are people with many options for how to spend their free time, who (may) freely choose to help you. Thus, it is critical to the success of your project that you **engage** your participants as much as possible.

Volunteer engagement takes many forms, and may differ from project to project. To get you started, this document contains a set of basic "best practices" for engagement that new research teams should consider as they create and manage a

project. These best practices were identified by experienced volunteers, researchers, and Zooniverse staff at a Sept. 2015 workshop, funded by the Alfred P. Sloan Foundation, at Chicago's Adler Planetarium. They are organized into three categories:

Building your project to be appealing and volunteer-friendly. Launching your project and recruiting volunteers. Managing your project from post-launch to completion.

Although these best practices are for the most part **non-mandatory suggestions**, they are drawn from significant experience, and **we strongly recommend reading through this entire article before you start building your project!** It may be useful to refer back to it later on as well.

Please read on for how to Build a Great Project, and best of luck with your project!

— *The Zooniverse Team*

Made with Material for MkDocs

Zooniverse Help

Part I: Building a Great Project

Know that you are making a commitment! Building, launching, and (especially) maintaining an engaging project takes time and sustained effort. Before you even start, your project team should be prepared to remain personally involved with your project for its entire lifespan. Reduce your individual workloads by delegating responsibilities to each member of your team.

Understand your goals and expectations from the start. The end goal of a Zooniverse project generally should be to produce new research from volunteers' efforts. Be sure that your project design and messaging reflect the tangible goal that you are working towards.

Consider reaching out to Zooniverse volunteers early on. Some of our most successful projects in terms of engagement have been those built in cooperation with volunteers. You might find volunteers interested in helping with your build on Zooniverse Talk's Project Building board.

Keep everything short and simple. Your volunteers are smart, but most will not be experts. Avoid jargon and complex questions. Make the tasks fairly basic, and consolidate them when possible. Ideally, a 10- or 12-year-old child should be able to understand and do your project.

Help text should help volunteers answer the question. When volunteers click "Need Some Help?" they are seeking help in answering that question. There are other areas in the project to describe your research; the help text is for visual examples and question-specific guidance. You can (and should) discuss your research goals in depth on your project's Research page.

Use images as examples wherever possible. Visual examples of what you are

asking volunteers to look for are extremely important to have in addition to text descriptions. Consider offering a variety of examples of "correctly" classified data to educate volunteers.

Create informative content and a solid name and tagline. The title of your project should be short and punchy; your tagline should try to "hook" volunteers. Use your static pages (Research, FAQs, etc.) to explain your project clearly and resolve the most common questions about it.

Be prepared to test and make changes. However good your project's "first draft" is, volunteer testing (including the formal Zooniverse "project review" stage) will reveal things to alter. You should first test your project with contacts, friends, or family who are trying it for the first time—and watch them as they do it! Study your feedback and raw data, and make adjustments.

Make your project visually clean and interesting. Choose imagery that fits your project's aesthetic and that you have permission to use. Make sure that text is readable over your background. Also be sure to do a review to make sure that all of your text is displaying properly.

Made with Material for MkDocs

Zooniverse Help

Part II: The Launch Rush

Understand the importance of the launch period! The first few days following the launch of your project are critical. Thousands of volunteers will be trying the project and deciding if they want to become regular classifiers. Your classifications per day will likely spike, then fall and level out; your level of long-term engagement depends largely on what you do during launch.

Write an amazing newsletter. For new Zooniverse projects, a newsletter will announce the project to over a million volunteers; it can bring huge traffic. Your newsletter should be short and personal, clearly describing the project and its goals. Include links to the project and associated sites (blog, social media, etc.). Don't use styling or images—plain-text newsletters do best.

Have a promotion plan in place. The most successful projects will recruit new volunteers from outside the Zooniverse volunteer pool. Coordinate with your organization's communications department. Try to promote to any existing networks or known interest groups; seek out press attention. Also see if you can leverage connections with related organizations for promotion.

"Talk" to your volunteers! One appealing element of the Zooniverse is that volunteers can engage directly with professional researchers on the Talk discussion boards. Your presence in these first few days is especially important. By answering questions, you're helping volunteers feel more comfortable and laying the foundation for those volunteers to teach future newcomers.

Prepare your blog, social media, and Talk, and be active. These are critical for developing deeper engagement, especially during your launch. Make sure that several boards are set up on Talk and that research team members are prepared to quickly

respond to volunteers. During launch, consider daily blog posts and up to five posts per day on social-media accounts.

Have a plan to identify and appoint moderators. Moderators have additional authority on Talk and are a resource for new volunteers. They are appointed by you—typically after asking them personally. Choose mods wisely. You might start with a volunteer who is a mod on another project, then add first-timers later. Be sure that your moderators understand your expectations.

Expect things to go wrong, and be ready to fix them. In the attention that your project receives shortly after launch, you may discover additional issues that you will need to address immediately. Broken projects, or ones with messaging issues, do not engage and retain volunteers well. Update your FAQ with questions that you find yourself frequently responding to.

Plan your transition from launch to operating your project. After about a week, the launch frenzy will subside and less of your time will be required. However, you will still need to continue to engage with your volunteers and oversee your project. This is discussed further in Part III.

Made with Material for MkDocs

Zooniverse Help

Part III: In For the Long Haul

Even after launch, running your project demands a time investment. You should dedicate time each week to volunteer engagement. Set up a plan in advance for who on your team will do what, and how often. Commit to your plan, and hold each other to it!

Send occasional reminder emails to your volunteers. Sometimes volunteers do forget about projects. The Zooniverse Communications team can help you send newsletters to your volunteers. Newsletters are proven to bring people back to projects and boost classifications.

Give people a reason to join your community. Volunteers who discuss your project on Talk, social media, etc. often remain more devoted. You may be able to entice more people to do this if, in addition to your posting regularly, you give them a specific purpose. For instance, volunteers could collect examples of rare objects, or find the same animal in multiple images.

Make it fun to be a part of your project. Your research may be serious, but that doesn't mean taking part can't be fun! Instead of competitive gamification (a turnoff for some), try 'gamizing' instead, with wry commentary, puns, captions, or other whimsical non-competitive amusement.

Don't be afraid of volunteers! Just as your volunteers are better at classifying than they think, you are better at communicating to them than you think. As a professional researcher, you will command an automatic degree of respect to volunteers; use it to educate and engage.

Report back to volunteers on what they've accomplished. In a survey, over 90% of Zoo volunteers said they classify to help advance research. Use blogs and

social media posts to illustrate how their work is having a tangible effect and helping make discoveries. Encouraging new volunteers is especially important. Updates don't have to be detailed; brief reports are fine!

Pay close attention to your project metrics. Your metrics are a valuable tool in evaluating the success of your engagement initiatives. Use them to inform decisions. Currently, a classification counter is available, and you can obtain additional information by querying your database.

Consider giving volunteers more ways to contribute. Some Zooniverse volunteers have conducted in-depth data analyses, written sorting algorithms, and become co-authors with researchers on peer-reviewed papers. Volunteers can also help with blogs and social media.

End your project and do right by your volunteers. When your project has been completed, thank your volunteers and explain what happens next. We expect researchers to use the results of the project in peer-reviewed research, and to share results with the community. Classification data should also be made open after a proprietary period. More information can be found [here](#).

Made with Material for MkDocs

Zooniverse Help

Appendix: Resources and Examples

Useful Links, Blog Posts, and Documents:

- Project Builder Policies — Official Zooniverse policy about user-built projects.
- Project Review: Best Practices & Flowchart — Overview of project review process and timeline, with helpful diagram and list of suggestions for projects entering review process.
- Examples of Strong Newsletters — Samples of newsletters we have used.
- Overview of Talk — Best practices for Talk board usage and descriptions of associated team roles.
- What Are Moderators For? — Learn what moderators are and why you want them.
- Measuring Success in Citizen Science Projects: Part 1 and Part 2 — An analysis of past (and current) Zooniverse projects, demonstrating the importance of engagement.
- Who Are The Zooniverse Community? We Asked Them... — A summary of results from a survey of Zooniverse volunteers conducted as part of a master's thesis in 2014-2015.
- Zooniverse GitHub — Zooniverse's code base. You can submit issues and pull requests.
- Zooniverse Researchers & Moderators — Private Facebook group that you may join.
- Zooniverse Backchannel Blog — Private blog you may request to join. Posts are relatively rare and often concerned with technical details of handling Zooniverse

data.

- 17 Tools for Social Media — Social media is important, but can also be time-consuming. These tools can help you pre-schedule, analyze, and get the most out of your posts.
- The Power of a Challenge — A challenge that got 650,000 classifications in a weekend.
- Zooniverse "Meta" Publications — for those wishing to know a lot more about research into the practice of citizen science. (Click the "Meta" link on the left side of the page.)
- Why We Don't Have an 'I Don't Know' Button — Volunteers often ask for some way to express uncertainty. Using Snapshot Serengeti, this explains why we want a best guess.
- Federal Crowdsourcing and Citizen Science Toolkit — Running a citizen science project from within the federal government comes with added constraints. This kit can help.

Examples of Project Builder Projects:

- Fossil Finder — Good example of content: name, tagline, help text/images, info pages.
- Whales as Individuals — Good example of tasks/questions with multiple workflows.
- Planet Four: Terrains — Has a strong focus on researcher/volunteer interaction on Talk.
- "Kitteh Zoo" — Lighthearted sample project illustrating many of the Builder's capabilities.

Examples of Gamized Behavior:

- Chimp & See Talk — This popular Talk is also used for the "Chimp ID" process: a

good example of non-competitive gamized behavior that also serves a research purpose.

- My Galaxies — Write text using galaxies as letters; built with volunteer help.
- Cookie recipes for Seafloor Explorer, Worm Watch Lab and Notes From Nature.
- Snapshot Serengeti Meme Generator — allowed users to caption photos (now defunct).
- Zooniverse project cocktails for Advent, part one and part two.
- Just for Fun — Various "fun" posts on the Zooniverse blog, many about gamizing.

Made with Material for MkDocs

Zooniverse Help

Use Caesar for more efficient data processing

<https://caesar.zooniverse.org/workflows/>

The Caesar system receives a copy of each classification submitted by users and can act on this data as it flows through the system.

Project owners can configure the caesar system to:

- Extract parts of information from the classification
- Reduce (Aggregate) extracted data into answers
- Add Rules to activate based on Reduced data
- Add Actions to attach to Rules to:
- Retire a Subject
- Add a Subject to another subject set (which can be used to advance subjects a next workflow which asks more detailed questions).

If a project has Extracts and Reducers setup in Caesar, this can be used to request exportable flat CSV files of both the extracted classification data and the Reduced(Aggregated) answers.

One other feature of the Caesar system is it allows for external third parties to hook into the Extract and Reduction phases. Data can flow out of the Caesar system to the another web service that can provide custom Extract and Reduction functionality. External systems can also be called as an Action when the reduced data matches rules.

The caesar system is documented at <https://zooniverse.github.io/caesar> and <https://github.com/zooniverse/caesar#readme>

For inbuilt Extractors see

<https://github.com/zooniverse/caesar/tree/master/app/models extractors> For inbuilt Reducers see <https://github.com/zooniverse/caesar/tree/master/app/models reducers>

Additionally, all of the aggregation code in

<https://github.com/zooniverse/aggregation-for-caesar> is available to be set up as (Zooniverse-hosted) extractors and reducers. The documentation on how to do that can be found at <https://github.com/zooniverse/aggregation-for-caesar#installing-for-online-use>

Made with Material for MkDocs

Zooniverse Help

Data Analysis

People have written code that can help you.

Because Zooniverse projects hosted on the Panoptes platform share a codebase, their outputs should be similar in format. That means that code written by other researchers to aggregate contributions from other users, and perhaps also to weight different users by their performance, may work for your projects too. There is a collection of such code available here: https://github.com/vrooje/panoptes_analysis and here <https://github.com/zooniverse/Data-digging> in repositories curated by Brooke Simmons (Lancaster University/Galaxy Zoo).

For details of more complex analysis, we suggest looking at papers by teams behind Space Warps (<http://mnras.oxfordjournals.org/content/455/2/1171.full.pdf>) or Snapshot Serengeti (<http://onlinelibrary.wiley.com/doi/10.1111/cobi.12695/abstract>).

Made with Material for MkDocs

Zooniverse Help

Data Exports

Zooniverse projects provide a large amount of data to research teams. These data can be exported from the `Data Export` tab on a project's `Lab` page.

Classification export

This `.csv` file has one row for every classification submitted for a project. This file has the following columns:

- `classification_id` : A unique ID number assigned to each classification
- `user_name` : The name of the user that submitted the classification. Non logged-in users are assigned a unique name based on (a hashed version of) their IP address.
- `user_id` : User ID number is provided for logged-in users
- `user_ip` : A hashed version of the user's IP address (original IP addresses are not provided for privacy reasons)
- `workflow_id` : The ID number for the workflow the classification was made on
- `workflow_name` : The name of the workflow
- `workflow_version` : The major and minor workflow version for the classification
- `created_at` : The `UTC` timestamp for the classification
- `gold_standard` : Identifies if the classification was made on a gold standard subject
- `expert` : Identifies if the classification was made in "expert" mode
- `metadata` : A `JSON` blob containing additional metadata about the classification (e.g. browser size, browser user agent, classification duration, etc...)

- `annotations` : A `JSON` blob with the annotations made for each task in the workflow. The exact shape of this blob is dependent on the shape of the workflow.
- `subject_data` : A `JSON` blob with the metadata associated with the subject that was classified. The exact shape of this blob is dependent on the metadata uploaded to each subject
- `subject_ids` : The ID number for the subject classified

Subject export

This `csv` file has one row for every subject uploaded to a project. This file has the following columns:

- `subject_id` : A unique ID number assigned to each subject as they are uploaded
- `project_id` : The ID number for the project
- `workflow_id` : The workflow ID the subject is associated with
- `subject_set_id` : The ID of the subject set the subject is connected to
- `metadata` : A `JSON` blob with the subject's metadata
- `locations` : A `JSON` blob with the URL to each `frame` of the subject
- `classifications_count` : How many users have classified the subject
- `retired_at` : If the subject is retired this is the `UTC` timestamp for when it was retired
- `retirement_reason` : The reason why it was retired
- `created_at` : The `UTC` timestamp for the creation of the subject
- `updated_at` : The `UTC` timestamp for the latest update to a subject

Workflows export

This `csv` file has the information for every major version of a workflow. This file has the following columns:

- `workflow_id`: The ID number for the workflow
- `display_name`: The display name for the workflow
- `version`: The major version number
- `active`: `true` if the workflow is active
- `classifications_count`: How many classifications have been made on the workflow
- `pairwise`: `true` if selection behavior is set to compare subjects against each other (not typically used)
- `grouped`: `true` if selection behavior set to select subjects by set (not typically used)
- `prioritized`: `true` if selection behavior shows subjects in a given order (not typically used)
- `primary_language`: The language code for the workflow
- `first_task`: The task key for the first task
- `tutorial_subject_id`: A default subject linked to the tutorial (not typically used)
- `retired_set_member_subjects_count`: The number of retired subjects from the workflow
- `tasks`: A `JSON` blob showing the full workflow structure
- `retirement`: The retirement rules for the workflow
- `aggregation`: Information passed to downstream aggregation services (deprecated)
- `strings`: A `JSON` blob containing all the text associated with the workflow
- `minor_version`: The minor version number

Zooniverse Help

Experimental Features

If you see a feature on another Zooniverse project but you don't see it as an option in the Project Builder, it may be an experimental feature we haven't enabled for everyone yet. Most new Project Builder features are tested out as experimental features before we make them available as standard. If you see a feature that you think would be useful for your project, email us at contact@zooniverse.org and ask for it to be enabled.

Made with Material for MkDocs

Zooniverse Help

Hosting your own subject media

We'll happily host your media (images, videos, etc.) for you, but if you don't want to upload them all to us you can host them on your own server. E.g. for copyright reasons, or to save time (if you already have them online and don't want to spend the time uploading them).

All you need to do is include the URLs in the manifest.csv file and then upload the manifest using the Panoptes command-line client using the panoptes subject-set upload-subjects command. You'll need to specify a couple of extra options to the upload-subjects command:

1. `-r column_number` – this option tells the client which column(s) in your manifest contains the media URL(s). You can specify this option more than once to include more than one piece of media. Replace `column_number` with the number.
2. `-m file_type` – this option tells the client what the file type of your media is. Unlike when you upload files to us, the client can't automatically detect the file type. If you don't specify this option, it will default to `image/png`. It's important to make sure the type you specify matches the media files you're using, as otherwise they may not be displayed correctly when people visit your project. The type should be specified as a standard MIME type string.

Important: Make sure your media URLs start with `https` rather than `http`. Zooniverse.org is served over a secure connection, so in order for your media to be loaded correctly it will have to be served securely too.

So, for example, if your manifest.csv looks like this:

URL	Catalogue	Location
-----	-----------	----------

number

https://example.com/images/image1.jpg	img-003-01	Oxford, UK
---	------------	------------

https://example.com/images/image2.jpg	img-003-02	Chicago, USA
---	------------	-----------------

You would use this command to upload the manifest:

```
panoptes subject-set upload-subjects -r 1 -m image/jpeg manifest.csv
```

Made with Material for MkDocs

Zooniverse Help

Markdown

Here's a few tips that work everywhere we support the Markdown syntax, e.g. pages, tutorials, the project description...

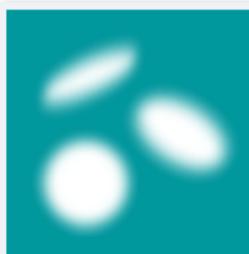
Displaying and Resizing Embedded Images

You can upload images to Panoptes and display them anywhere the Project Builder allows markdown. Tutorials, help, the field guide, and even as icons. Here's Galaxy Zoo adding icons to each answer using transparent .png images:

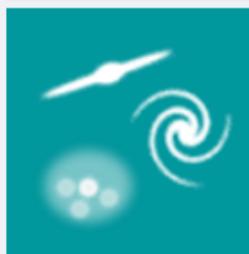
TASK

TUTORIAL

Is the galaxy simply smooth and rounded, with no sign of a disk?



Smooth



Features or Disk



Star or Artifact

But what if your image is stupidly large to use as an icon, or much too small for the field guide?

Panoptes has a special version of markdown that lets you specify the image dimensions to display. Copy the markdown string from the Media tab, then add ={pixel_width}x{pixel_height}, or blank for square just before the last bracket. For example:

```
![image.png](https://panoptes-uploads...748eb.png =60x)
```

will display a square image resized to 60 by 60 pixels.

This works for an image you host yourself (see [Adam's section]) - just update the URL.

Made with Material for MkDocs

Zooniverse Help

Newsletters

Did you know you can send a newsletter to your volunteers once your project is launched as an official Zooniverse project? Just send plain text copy to contact@zooniverse.org and we will edit and send to your registered volunteers.

Newsletters are one of the key ways of engaging with your community. Updates let volunteers know that their hard work is appreciated and being used by your research team. Newsletter will bring volunteers back to your project more than any other means of communication.

Made with Material for MkDocs

Zooniverse Help

Subject Selection Process

When you first set up a project, the process for selecting which subjects get shown to volunteers is very simple: it randomly selects an (unretired, unseen) subject from the linked subject sets for that workflow.

Per workflow weighting

However, we also have support to tweak this process. It's possible to assign weights to subject sets, so subjects in set A have a proportionally higher chance of getting shown compared to set B. Because subjects retire and subject sets have different amounts of subjects in them, it is possible to set weighting up in two different styles:

1. A fixed chance per subject set. For instance, subjects in set A will always have a 60% chance of being shown to a volunteer, and subjects in set B will always have a 40% chance of being shown, regardless of how many subjects actually are in set A. A practical use of this would be to show gold-standard training images to volunteers for 30% of the shown images, even though there might only be 20 training images in the project. Once those 20 training images have all been seen by a user, they would of course not be shown again, providing a natural automatic stop to training.
2. A fixed chance for each individual subject in a set. This is still configured at the per subject set level, but the weight is now applied at a per-subject basis, i.e. relative to all of the other available subjects in the workflow. With this you can make it twice as likely for a subject to be shown. A practical use for this is when you have new data available, but the previous dataset isn't fully complete yet. By increasing the relative weight of the old subjects, you can make sure that the

subjects in the old subject sets have a high chance of still getting their last few classifications needed for them to proceed to retirement, while still making the new subjects available for classifying as well.

Per user weighting

It's also possible to set these same parameters up on a per-user basis. This overrides the configuration at the workflow level. A practical use for this is to use heuristics to detect users with a high accuracy in certain types of subjects, and giving these users a higher chance of seeing certain subjects. This is an advanced use case which combines various features.

How to set this up

Unfortunately, we haven't yet had time to build any user interface for project owners to configure this behaviour themselves. If you'd like to make use of this on one of your projects, please let us know at contact@zooniverse.org and we'll be happy to set it up for you.

Made with Material for MkDocs

Zooniverse Help

Getting Started

1.1 Who is this resource for?

This guide is aimed at people who want to transcribe a large amount of handwritten text using volunteer crowdsourcing. You will either have an existing collection of digital images, or have identified manuscript or archival sources that you plan to make digital images of, which will be transcribed through the effort of volunteers. This guide is a comprehensive introduction to the lifecycle of a crowdsourced transcription project using the Zooniverse platform, covering technological and social aspects of such projects.

You may be interested in crowdsourcing because the volume of material you have is larger than you or your research team could transcribe alone. If you plan to ask volunteers to participate, you need to make the transcription process meaningful and rewarding. Thus we carry through this guide an assumption that the material you are interested in transcribing will be of interest to others beyond your immediate use case.

Crowdsourcing text transcription is of interest to a wide range of people. We have tried to write this guide with an awareness of the variety of interests and materials that may have brought you here. Our advice is not tailored to assumptions about your disciplinary background, institutional affiliation, or educational status, though those may well affect how you go about implementing the advice we give. Handwritten pages are important sources for meteorologists and medievalists, physicists and poets, economists and ecologists. Text transcription projects often originate from teams in museums, libraries, universities, government agencies, archives, genealogical organizations, and scholarly societies, but can be hosted by individuals or other organizations, too. A project to crowdsource the transcription of text may take a

year or more to complete, and so we assume that you—individually or collectively—can sustain your project over a long period of time, and be comfortable with uncertainty about how long it takes.

This guide was developed as part of an NEH-funded Institute for Advanced Topics in Digital Humanities, Building Capable Communities for Crowdsourced Transcription, that ran from 2021 to 2023. The authors of the guide worked with 15 project teams in seven online meetings and one in-person meeting to build crowdsourced transcription projects on the Zooniverse platform. This guide was informed by the questions and experiences of those projects, as well as our experience designing and teaching the institute.

This guide was written collectively by the Co-Directors of the institute, and its two research assistants.

We have written this guide in the first person, addressing our advice to “you” as a member of a team building a crowdsourced transcription project. When we refer to technical terms related to the Zooniverse we have capitalized the words.

Co-Directors

Samantha Blickhan, Zooniverse Co-Director & Humanities Lead, Adler Planetarium

Evan Roberts, Assistant Professor, History of Medicine and Population Studies, University of Minnesota

Benjamin Wiggins, Senior Lecturer of History and Library & Archive Studies, University of Manchester

Research Assistants

Treasure Tinsley, PhD Candidate, History, University of Minnesota

Trevor Winger, PhD Candidate, Computer Science, University of Minnesota

1.2 How to use this resource

We have written this guide to take you through the entire lifecycle of a crowdsourcing project. We encourage you to read the entire guide as you consider whether crowdsourcing is the right approach for your project and sources, and whether Zooniverse is the appropriate platform for your needs. As you develop, launch, sustain, and wind down your project in the months and years to come, we hope you will come back to relevant sections to consult in more detail.

We encourage you to begin by using this guide to assist you in creating a strategic plan for your project. While there are important common elements to all crowdsourced transcription projects, every project is unique in some way. You must make decisions about what will work best for your source material, goals, and team of people. This guide aims to help you narrow down those choices and understand the tradeoffs.

As you read this guide you will notice that we offer advice about different choices you can make in various phases of your project. Take the time to write down your thoughts about what you will do, and review these before you start the serious work of collecting and organizing images and developing Tasks and Workflows. You should think about when you need the data and how likely delays will affect your other work. Your volunteer community is not working on your timetable. Revisit your strategic plan as you make and implement decisions.

1.3 What is Zooniverse?

Zooniverse is an online platform for people-powered research. Though the platform's origins are based in astronomy research, the founders of Zooniverse recognized that many research challenges in the modern social sciences and humanities are similar: there is abundant digital media containing information that people want to transform into a format that allows easier manipulation. There is a fundamental similarity between an astronomer asking, "What shape is this galaxy?" and a medieval historian asking, "What words are written on this page?"

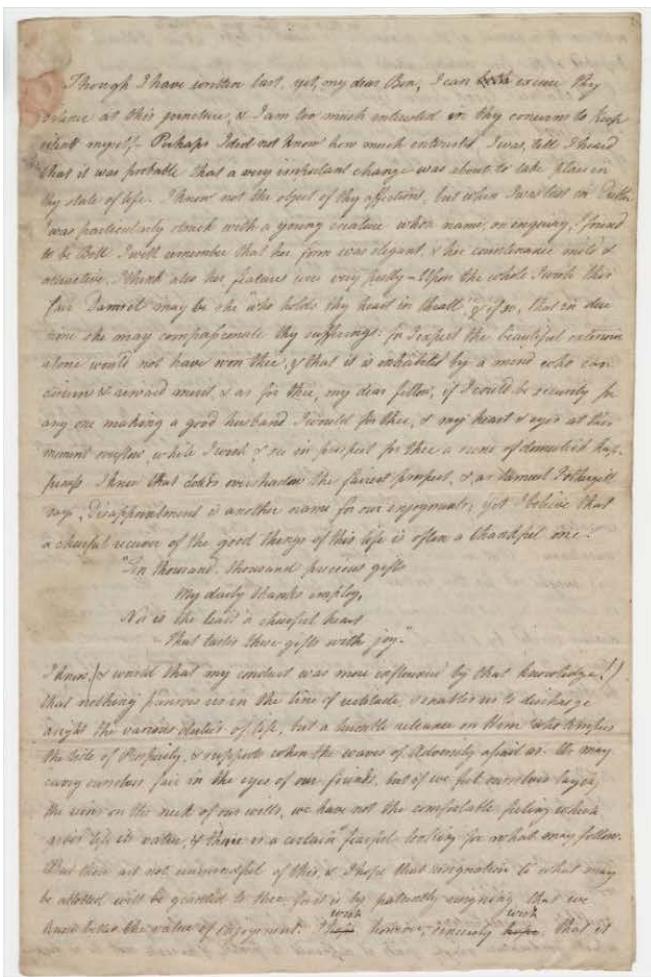
Because digital images can be shared easily on the internet, it is possible to ask people all over the world to help us answer questions about those images. The Zooniverse Project Builder is a browser-based tool that allows anyone to build what are essentially sequences of questions posed to people about the image at which they are looking. We will describe the technical aspects of using the Project Builder and how they apply to transcription in this Guide. You should also consult the general Zooniverse guide to building projects at help.zooniverse.org, which will contain the most up-to-date technical information about the Project Builder.

Projects on Zooniverse ask volunteers for help in systematically examining images (or more generally, media) and recording answers about them because the questions being asked about the image are not yet possible for a computer to answer with acceptable accuracy. The boundaries between “what can a computer do quickly and accurately?” and “what can people do more accurately?” are shifting rapidly. Before asking volunteers for help with transcribing, consider whether some of your material could be transcribed by a computer. Computers are becoming better at accurately recognizing handwritten text—some handwritten documents can already be read accurately enough by computers—but for other documents the human eye still rules. Data from Zooniverse and other transcription projects is already being used to improve how computers read handwriting. As you develop your project, we encourage you to be aware of these possibilities and collaborate where you can with people trying to improve these methods. For more information, see “Preface” (Wanamaker, Frydman, and Dahl), which discusses and links to recent work in computer recognition of handwriting.

Zooniverse is also a community of project creators and volunteers who connect online about the projects on which they are collaborating. Developing a community of volunteers who will spend time on the project that you create is a distinct challenge from designing and building the project itself. Finding and encouraging people to spend their time on your project, and to keep them engaged and motivated, is a social challenge. We encourage you to think about who might be interested in the material you want to transcribe. When a project launches publicly, Zooniverse will

publicize it to an email subscription list of several million people. Every project launch is a chance to engage existing Zooniverse volunteers as well as new volunteers who join the platform because they are interested in that new project.

Transcription projects often draw on audiences interested in history, genealogy, literature, the culture of the place the writing comes from, and the history of particular institutions, organizations and topics.



Though I have written last, yet, my dear Ben, I can [deletion] both[/deletion] excuse thy silence at this juncture, & I am too much interested in thy concerns to keep silent myself - Perhaps I did not know how much interested I was, till I heard that it was probable that a very important change was about to take place in thy state of life. I know not the object of thy affections, but when I was last in Dublin I was particularly struck with a young creature whose name, on enquiry, I found to be Bell. I well remember that her form was elegant, & her countenance mild & attractive. I think also her features were very pretty - Upon the whole I wish this fair Damsel may be she "who holds thy heart in thrall," & if so, that in due time she may compassionate thy sufferings: for I expect the beautiful exterior alone would not have won thee, & that it is inhabited by a mind who can discern & reward merit, & as for thee, my dear fellow, if I could be security for any one making a good husband I would for thee, & my heart & eyes at this moment overflow, while I wish, & see in prospect for thee a scene of domestick happiness. I know that clouds overshadow the fairest prospect, &, as Samuel Fothergill says, Disappointment is another name for our enjoyments; yet I believe that a cheerful receiver of the good things of this life is often a thankful one.

"Ten thousand, thousand precious gifts

My daily thanks employ,

Nor is the least a cheerful heart

That tastes these gifts with joy."

I know, (& would that my conduct was more influenced by that knowledge!) that nothing preserves us in the line of rectitude, & enables us to discharge aright the various duties of life, but a humble reliance on Him who tempers the tide of Prosperity, & supports when the waves of Adversity assail us. We may carry ourselves fair in the eyes of our friends, but if we feel ourselves laying the reins on the neck of our wills, we have not the comfortable feeling which gives life its value, & there is a certain fearful looking for what may follow. But thou art not unmindful of this, & I hope that resignation to what may be allotted will be granted to thee, for it is by patiently resigning, that we know better the value of enjoyment. I [deletion] hope[/deletion] [insertion] wish[/insertion], however, sincerely [deletion] hope[/deletion] [insertion] wish[/insertion], that it

Above: Mary Leadbeater to Benjamin Haughton, July 16, 1804. University of California Santa Barbara MSS4, Box 6, folder 8, 001. From *Corresponding with Quakers*, Principal Investigator and Project Director Rachael Scarborough King.

1.3.1 Project Builder broad overview

The elements of a Zooniverse project are simple. You have a collection of images which contain information that is best extracted by asking people to look at the image and do something. The Project Builder allows you to upload the images, ask people a

set of questions about the image, and export the answers that people provide. Zooniverse projects feature multiple, (typically) independent classification, meaning more than one person will respond to the questions being asked about each image. By comparing the answers that different people give to the same question about the same image, you can be more confident you are getting an accurate answer. In general, crowdsourcing has been proven to be a reliable method for generating high-quality data, even at expert levels, and a number of Zooniverse projects have resulted in publications demonstrating the reliability and usefulness of crowd-generated data.

The Zooniverse platform stores the answers that different people give, and makes it available to you for download as a CSV file. You will use your knowledge of the material to decide how to combine the answers that different people give into a single answer that best represents the material you are transcribing: this is called Aggregation.

Because text transcription is a popular and complex task, Zooniverse has developed additional tools for aggregating text data and reviewing transcription project results. For more detail on data aggregation, see Section 7.

1.3.2 Outlining the Zooniverse project process

This graphic represents seven high-level steps required to build and launch a public Zooniverse project (i.e. one that is listed on our Projects page).

BUILD A ZOONIVERSE PROJECT



1.3.3 Project input: overview

To build a project in the Project Builder you will need a collection of digital images of the material you want transcribed. You do not need all of your images right away. You should know the provenance and usage rights for each image, and have a plan in place for organizing information about the data you are using, such as a spreadsheet. If you are using images of material in a library or archive, the file reference of the

images is information you should plan to save along with the image names.

1.3.3.1 What is a Subject Set?

Transcription projects often involve collections of related pages that you likely want transcribed together. For example, it is not very helpful to have one random page of a set of multi-page letters transcribed. You probably want related pages transcribed together. Subject Sets allow you to group related images to be transcribed at the same time, so you can examine the output (text) from those images together. See Section 2 for more detail.

1.3.4 Project output: overview

Zooniverse projects allow you to collect multiple transcriptions of the same text. We find that 2-5 people transcribing an image is sufficient to produce a generally accurate transcription without wasted effort. Not every volunteer will write the same thing, so you will need to aggregate the data into a harmonized or consensus version that best represents the original material. Zooniverse provides a tool for aggregating lines of text transcription which can be used for certain transcription project types, and also provides Python scripts to assist with aggregation of data produced by other tools available in the Project Builder. As a project creator your knowledge of the material and plans for using the output should shape how you collect and aggregate data. For example, if you are producing an authoritative edition of a manuscript you will be less tolerant of spelling mistakes than if you are coding or classifying text into numeric categories for quantitative analysis.

The interests of project creators will shape the way that data is initially collected and aggregated. Zooniverse requires that you make your data publicly available within two years of your project's completion. Therefore, you should ideally collect and aggregate data in a way that will be useful to other people interested in the material you are transcribing. Try to anticipate how others might want to use your data when putting together your strategic plan.

The aggregated consensus output (i.e. the 'approved' transcription data) is likely to

be most useful to other people interested in your material. However, when you conclude your project you should also make available the larger, unedited dataset of transcriptions and the code used to transform these into aggregate output.

1.4 What is crowdsourced text transcription?

This section will provide context around methods and history of crowdsourcing, with a focus on transcribing text. It will dive into ethical approaches, motivations for transcribing, and best practices for developing goals and objectives for your project.

1.4.1 What is crowdsourcing?

Historically, crowdsourcing was conceived of as a way to outsource labor to the crowd via an open call (Howe). This model of crowdsourcing persists today as microwork (see for example Amazon's Mechanical Turk, Appen or Clickworker, all of which pay gig workers on a per-task basis to produce training data for machine learning models). Since the term was first used in 2006, research crowdsourcing has largely deviated from the original definition in the sense that it is not paid work, but a way to involve the public in research in meaningful ways.

For a text transcription specific definition of crowdsourcing, we look to The Collective Wisdom Handbook: perspectives on crowdsourcing in cultural heritage: "At its best, [crowdsourcing] is a form of digitally-enabled participation that creates meaningful opportunities for the public to experience collections while undertaking a range of tasks that make those collections more easily discoverable and usable by others." (Ridge, Blickhan, Ferriter, et al.)

The most common humanities crowdsourcing task is text transcription, in which volunteers help to create digital, searchable versions of text found in digital images of archival records.

1.4.1.1 The ethics of crowdsourcing

Because there is no financial reward for taking part in crowdsourcing, it is up to practitioners and project creators to produce “human-centered projects that deliver benefits to all involved” (Ridge, Blickhan, Ferriter et al.). The best way to ensure you are creating an ethical project is by establishing values and letting those values drive your project design and testing processes as well as the ways you involve and engage with volunteers.

Values are broad, and can include everything from committing to only publish project research in open-access journals, to including warnings on projects dealing with potentially traumatizing content, to clearly articulating how you plan to acknowledge volunteer effort in publications. Values can and should evolve over the course of your project. A good project includes places where you and your team are able to take a moment to check in on your values and either update them based on your experiences, or actively choose to keep them. Your project cannot be values-agnostic; choosing not to articulate your values is de facto support of existing power structures (Ridge, Blickhan, Ferriter et al.).

Human-centered projects mean keeping people in mind at every stage. When designing your crowdsourcing project, consider this: how do volunteers benefit from helping transcribe this text? What can you do to ensure a positive user experience, and make sure that volunteers’ labor is being used efficiently? If you are working with students, how are you accounting for the additional power dynamics in place within the student/teacher relationship? Zooniverse requires all project creators to make their data publicly available within two years of the project’s completion—how does your data management plan account for this requirement? These framing questions can help you get started thinking through the ethical considerations of your project; further ethical issues are considered in Section 6: Acknowledging Volunteers.

1.4.2 Why transcribe?

Before getting into the specifics about why you want to transcribe, this table will help you think more broadly about why to crowdsource more broadly.

WHY CROWDSOURCE?

Scale	Discoverability	Engagement
Too much to do on your own	Data processing & transformation	Opportunities for exploration
Cannot be automated	Increased access to collections	Encourage institutional transparency
	Create or expand reusable datasets	Work with communities
Broaden, deepen, and diversify corpora		



1.4.2.1 Identifying your needs

If your motivations are only coming from one of the above three categories (Scale, Discoverability, or Engagement), there's a good chance you're going to run into problems. Either you are too focused on data and not focused enough on people, or you've put too much of a focus on the public project itself and haven't given enough thought to your data management practices or whether you even need to ask the public for their help.

1.4.2.2 Is Zooniverse the appropriate platform for your project?

Zooniverse projects benefit from 1) the ease of using the Project Builder; and 2) the size of the existing volunteer community. With that said, designing, building and running a crowdsourcing project requires a lot of effort and resources, especially in the form of staff time. Planning out your project in advance will help you confirm that you have the resources and bandwidth to take on this responsibility. For more information on selecting platforms, see The Collective Wisdom Handbook, Ch. 7, Aligning tasks, platforms, and goals.

Case Study: New tools for transcription at scale

The People's Contest Digital Archive is a project led by the Richards Center at the Pennsylvania State University to transcribe Civil War era documents from Pennsylvania. Transcribing through the Zooniverse platform was an opportunity to broaden access to documents in collections at Penn State and elsewhere in Pennsylvania.

Kevin Clair, Project Coordinator:

"As a librarian tasked with digital collection stewardship, my team and I are always interested in new ways to provide access to our collections for users and to engage with our community in new and interesting ways. For us, publishing our collections as data opens them up to new methods of analysis and can allow connections to be made between materials in the collections that can't be made through resource description or close reading alone. Working with Zooniverse, particularly with a project that already had a degree of existing transcription, gave us the tools and the impetus to find a methodology for publishing transcriptions at scale that we could refine for future projects like this."

1.4.3 Develop research goals and objectives

Goals are high-level ideas, and objectives are the actions you take in service of those goals. You should identify your project goals and objectives early on in the project planning process.

1.4.3.1 What are your broad research goals?

Your goal here is likely to create digital, searchable text through crowdsourced transcription. Do you have additional research goals that are related to the content of the transcribed text? It helps to map out the full extent of your project goals at the start, as this should drive your decision-making process.

1.4.3.2 Breaking your project into smaller pieces (objectives)

For the goal of creating digital, searchable text through crowdsourced transcription, objectives might include identifying team members, creating a data management

plan, identifying and securing funding, designing Workflows, and getting buy-in from institutional administrators.

Kevin Clair, Project Coordinator, *The People's Contest Digital Archive*:

"Methods for measuring engagement are also something that digital collections programs have found elusive. Having engagement visible as directly as it is with a community transcription project – through the Zooniverse Talk boards, or even through the steady stream of subjects awaiting final approval – is something we can take directly to supervisors and administrators to document the impact that our digitization program is having on visibility of digital content from the library."

1.4.3.3 What should be done through Zooniverse?

As the list above shows, not all of your goals and objectives will take place on Zooniverse, though there will be many points where the success of your 'offline' objectives directly impacts your Zooniverse project. When mapping out your goals and objectives, take care to consider what tasks truly need to be crowdsourced, and what can be done internally—often, the labor required to work with crowdsourced data can outweigh what would be required for institutional staff to do the work themselves. You should also explore what parts of your project could be performed by a computer—what is technologically feasible is changing rapidly and will vary across different sources. Thinking through the full scope of your project can help you to identify what scale of project you are able to take on.

Case Study: Developing a strategic plan

Deciphering Secrets: Unlocking the Manuscripts of Medieval Spain brings medieval Spanish manuscripts online, and provides volunteers with the opportunity to learn paleography before participating in transcription.

Roger Martinez, Project Director:

"I think the challenge here is that, as humanists start to work more closely with

technology, we weren't trained to use these types of materials and then we don't know the critical kind of steps you have to run through. It's not like in history, where we know that you're going to do something like a historiographical session, we're going to present an argument... you don't have the same kind of guideposts.

You don't know what you don't know. So I think the steps have to be related to, starting with the highest level, what are the goals and objectives of the project? And from there, it's really, how are you going to measure and assess your progress towards those goals? Then also, for the next step, really thinking through, 'Okay, so what are some of the specific strategies I might implement that will move in this direction?' And then, what do the deliverables look like at the end. **You have to think about this as a strategic planning process.**

This is part of the conversation that's happening out there right now is there's this expectation that by using this crowdsourcing technology or processes that it will make it easier for you to do things and I think what we realize is though, it doesn't actually make things easier, it just facilitates a type of greater engagement of many people working on a common project. So collective action."

1.5 Early considerations

This section will signpost some things to think about early on as you begin to conceptualize and design your project.

1.5.1 What are your short- and long-term goals?

A clear understanding of your broader project goals is vital as you develop your plan for crowdsourcing the transcription of handwritten text. We noted earlier that we make few assumptions about your disciplines or background. But whatever discipline you are working in, Zooniverse requires that projects have the goal of producing useful research. Across the diversity of your research fields it is likely that your goals are to:

1. Produce data (text) that you share so that other people can study the text, and present and publish results.
2. Produce data that you will initially analyze, present, and publish. You will then make the data public in accordance with Zooniverse policies.

Crowdsourcing takes time. Sometimes it proceeds more quickly than expected, and other times more slowly. You should identify short-term analysis, presentation, and publication objectives that can be achieved with only a small share of the text transcribed and aggregated. What is a particularly interesting subset of your data that you can write about? In transcription projects these subsets of the data are often identified by the project leads. Some science projects might be able to use a random sample of images, but this probably doesn't work for letters, diaries, and drafts of literary works. Identify your highest priority material, and prioritize those images for transcription. Using early results to demonstrate that volunteer work is contributing to discoveries and being shared is tremendously encouraging for your volunteers, and can help with volunteer retention, especially in long-running projects.

Showing initial results and the viability of your work is also important in seeking more funding for your project. Grant funds can help provide your team with the additional skills and resources it requires (see Section 8).

Working with volunteers can take your work in new and unexpected directions, as you collaboratively discover unexpected things via the initial transcriptions. Only you can decide whether to pursue these new directions by evaluating them in relation to your long-term goals.

Articulating long-term goals is important for all transcription projects. For projects led by one or two people, clearly stating long-term goals will help you prioritize your work as your project progresses. Crowdsourcing projects led by a small team can falter when there are other demands on people's time. Articulating long-term goals at the beginning of a project can help larger collectives and institutions with structured discussion of where a project is going, and help with aligning staff work with project needs.

We encourage you to build towards your long-term goals with a series of short-term achievable objectives that can be communicated to volunteers and other collaborators.

1.5.2 Where will your project and its data live?

Zooniverse provides all projects with an online transcription interface (where the work of volunteers is done) and a dedicated message board, called Talk, where you can interact with your project community (see Section 4.3). You should also consider how to develop a presence for your project in other ways, both online and offline (see Section 6 on working with volunteers).

Many projects have established separate websites that give further background on the broader goals of the project. These are most often hosted by affiliate institutions. Project-specific social media accounts can also be useful in calling attention to your work, recruiting volunteers, and publicizing data and publications. Because the social media landscape is rapidly changing and project needs vary, we do not give advice on which platforms to use. We do recommend that you establish an identity for your project that is separate from the personal or professional accounts of project leaders (though they can and should amplify project messages).

When you have data to share, a project website can be an additional venue for publicizing the availability of data by linking to a trusted data repository. It is vitally important that you have robust systems for storing your images, programs and data. Because the work of crowdsourcing projects is collaborative, you should develop a system that allows data and programs to be shared and accessed by all members of your team, and a workable system for communicating about your project.

When you have developed final aggregated versions of your transcription data, you must share this publicly (see Section 8).

1.5.3 What types of support do you have? What types of support should you seek?

This chart lists the seven main steps of building a Zooniverse project (the same steps shown in the chart in section 1.3.2). Under each is a bulleted list of skills that are helpful at that step.

PROJECT TEAM SKILLS



Crowdsourcing text transcription requires the application of a range of different skills, potentially including digital image creation (photography and digital preservation), organizing and manipulating large volumes of digital images (file management),

designing an appropriate Workflow for transcription (user experience design), recruiting and working with volunteers (social movement leadership), data cleaning and analysis (file management, scripting and programming), and presenting and publishing analyses.

This is a wide range of different skills, and it is a rare person who will be skilled in all of the above listed areas. We encourage you to build collaborations with people who have complementary skills, particularly if you are reading this as a lone researcher, or a smaller group of two to three people. If your team does not have the skills required, do you have the time and resources to allocate for a team member to learn them, or can you form a collaboration to bring these skills to your project?

Many crowdsourcing project creators (and potential project creators) are at universities or based in libraries, archives, and museums. Resources for learning these skills are likely available to you via your institution. You are also likely to have colleagues with complementary skills. Collaboration with students or interns who have skills relevant for your project can be mutually beneficial when approached in the right way. If you work with students or interns, ensure they receive appropriate credit and reward for their work. Recruiting collaborators is valuable practice for recruiting volunteers, as you must articulate why your project is interesting, and why other people should become involved.

Identifying the skills and resources you have, what you need, and the gap between them is an important part of your strategic plan for your project. As you read this guide, try to identify areas where there is a gap between your current team's skills and resources and what you need to construct useful and usable text data. How will you acquire those skills and resources by the time you need them?

Case Study: The benefits of collaboration

Letters to and from authors provide scholars and readers with new insights into what authors were thinking about that didn't make it into print. In the *Maria Edgeworth Letters* project, transcribing letters held in one place makes the material more widely

accessible and provides volunteers with exciting opportunities to learn more about the thoughts behind the printed word.

Hilary Havens, Co-Principal Investigator:

"I'm very fortunate to be part of a project that has a lot of collaborators. We have four librarians who frequently give us advice. And having them help produce things like the personography, or check our metadata, check our TEI template, make sure we're linking to name authorities . . . having this knowledge that would be very hard for an English professor to acquire is great. And having someone who knows so much about metadata and name authorities takes a lot of pressure off other people doing other things."

Made with Material for MkDocs

Zooniverse Help

Project data

2.1 What is a Subject? What is a Subject Set?





Above: A subject from *The Gravestone Project*. Photo taken by Rebecca Schneider on December 10, 2022, in Watrous Cemetery near Watrous, NM.

In Zooniverse parlance, a Subject is a unit of data to be analyzed. For transcription projects, Subjects are images of the text you want your volunteers to transcribe. A single Subject can include more than one image. For example, multi-image Subjects can group pages from the same letter or group pairs of images that are known to be the front and back of a single sheet of paper.

A Subject Set consists of a group of Subjects (the images) as well as a Manifest (a list of the Subjects and their metadata). You can put all the Subjects in your project into a single Subject Set, or you may choose to break up your data into multiple Subject Sets. You can ask volunteers to perform the same actions on all Subject Sets, or you can assign different actions for each Subject Set, making it an effective way to produce different types of data.

Creating multiple Subject Sets allows you to group your Subjects appropriately for any particular nuance of partitioning that may be useful in your Subject analysis or to create a better volunteer experience. You may choose, for example, to create Subject Sets based on the date of creation, the structure of the data output, or the data source. In one instance, you may choose to group a set of related images—such as all the images from one archival folder or manuscript volume (dozens to hundreds of pages)—together in a Subject Set.

Transcription projects often have natural divisions that make for convenient Subject Sets based on the organization of the material in archives or libraries. However, you can choose not to subdivide into Subject Sets based on those divisions if it makes more sense for your project. If a Subject Set contains images from multiple archival folders or manuscript volumes, your image naming conventions can still indicate these distinctions, or you can include this information in the Subject metadata, to enable you to maintain the connection back to the archival organization.

2.1.1 What makes a good dataset for crowdsourced transcription?

2.1.1.1 Content

Content should be the driving force in your choice of material to transcribe. Here are a few things to think about when choosing datasets:

- How do the Subjects relate to one another?
- What is the unit of Subject: a single image? Multiple images (e.g. a multi-page letter)?
- Can longer Subjects be separated in a way that will ease volunteer work without removing crucial context?
- Can Subjects be divided meaningfully to alleviate unnecessary steps in analysis later on?

2.1.1.2 Data management planning

A data management plan is a document that outlines how you intend to organize and maintain all data related to your research. You can create a data management plan using a template tool like DMPTool. Keep your data management plan in mind when identifying potential transcription project datasets. How does the organizational structure of your data help or hinder your plans for data analysis and quality control?

2.1.1.3 Image quality

Zooniverse sets a strict per-image upper limit of 1MB. In practice, it is best to start

out with the best image quality for your Subjects and then manipulate images to meet requirements through standard image processing techniques (ImageMagick, Adobe, also a free-to-use processing notebook utilizing ImageMagick and Google Drive). Subjects must be in the appropriate file format (see the upload box in the Subject Sets tab of the Project Builder for a list of acceptable file types). Each user, by default, can upload 10,000 Subjects to the Zooniverse platform; this limit can be increased upon request by following the instructions (here)[<https://www.zooniverse.org/about/contact>].

If you are creating your own images you can find further information about digital image creation in Appendix B of “Breathing new life into death certificates” (Bailey, Leonard, Price et al.).

2.2 Where will your data come from?

This section will provide information on identifying appropriate datasets for use in crowdsourced text transcription projects.

2.2.1 How to source project datasets

Transitioning from a specific research interest to a dataset and project environment that is conducive to crowdsourcing can seem like a daunting task. As a project creator, you may source your datasets from one or more archives, libraries, museums, and other institutions, as well as materials you have digitized yourself.

The digital images created by or at cultural institutions are generally suitable for use in Zooniverse projects, provided you have the right to use them (see 2.2.2 Copyright concerns). You may need to do some preprocessing to make your images the right size for the Zooniverse requirements (see 2.2.1.3 Image quality).

If there are no digital images of the handwriting you are interested in, you will need to create your own images. There are many factors to consider in undertaking this work. See the supplemental appendix of this article on creating images for

handwriting recognition, for which similar considerations apply.

Case Study: Sourcing subjects for *The Gravestone Project*

The Gravestone Project (TGP) is a digital humanities collective that brings people together to think about the various ways that people memorialize the dead. As part of the BCCCT cohort, The Gravestone Project developed a Zooniverse project to record and transcribe gravestones. However, sourcing these images presented both unique challenges and opportunities.

Rebecca Schneider, Associate Director:

"There are several institutions, like Find-a-Grave, where we could have sourced subjects from. The crowdsourced transcription, that's really thrilling. I'm almost more thrilled by the crowd. That's a huge labor question solved, but I'm personally more excited about the submission of subjects. Because like you said, that means someone's standing in a graveyard. To me, that's the benefit or the contribution of physically being there on the ground, reading a gravestone and taking a photo."

Amy Giroux, Digital Historian & Technical Lead:

"The memorialization of the people is important. You know, holding that space. The other question that always pops up is we need to have them... say that they're not relinquishing their copyrights to that image, but we're allowed to use it. That was another ethical piece—we can't just pilfer a picture. We have to make that effort to go out and obtain permission to use it. And like you said, at the institutional level, it's a little different thing, you've got to have a repository or whatever, it's more of a legal thing that you have to deal with. But in our case, it's morally important."

Rebecca Schneider:

"For me it's also that feeling of being in community with those people. It's sharing this interest, but also this commitment to ... commemoration and memorialization. This is truly a collective effort. And we might be putting the most hours into

conceiving of the structure and the database and all that, but it wouldn't be what it is without the subjects and if it was just [our team], like even that many people can't achieve coverage."

2.2.2 Copyright concerns

Copyright issues can be a significant source of delays in bringing text transcription projects to the public. The original handwritten materials you seek to have transcribed may be the property of an archive, library, or museum who received it from another party. The images themselves may have been created by yet another entity. In short, copyright issues with handwritten material can be complicated. Some common copyright situations transcription projects encounter include:

- Material held by a government agency, and the images were created by the government. Often publicly available upon request.
- Material held by a government agency, but the images were created by private entities (e.g. Ancestry). May not be available for transcription, even if viewable on government websites or freely viewable on the private entities website.
- Material created by private individuals, and held by a university, library, private archive, museum or scholarly society have highly variable usage rights. Policies may differ at the collection level within an institution, if the content creator or donor has placed usage restrictions on the material.

You are responsible for determining whether you have the right to use images in a Zooniverse project. Directly asking the institution(s) who hold the material or images whether you can place the material on Zooniverse is the most straightforward way to proceed. You may need to enter into an agreement with the institution to place them on Zooniverse. Zooniverse policy is that you should be sure of your rights to use the images before you make a project public. It is a Zooniverse requirement that you make the data created in your project public. The transcriptions created by your project may be useful data for the institution holding the original images or manuscript material, and offering to share back the resultant transcriptions with the

holding institution can sometimes make it easier to get permission to use images.

If you do not have permission to share images publicly, you will need to consider whether Zooniverse is the appropriate platform for your project.

Roger Panetta, Project Director, *Shadows on Stone*:

"In those incidents where you are working with archival material or collection, be certain about the copyright clearance and the mechanics of the transfer to the Zooniverse site. The presence of private profit-making intermediaries can complicate this process. Do not despair! You can find workarounds and multiple tasks when such impasses occur."

2.3 How do I get my data into Zooniverse?

Before you can start building Workflows and iterating your project design, you first need to upload some data to the project. We strongly recommend that you begin by uploading a small subset of your data for use in workflow building and testing. This helps reduce the risk of needing to re-upload large amounts of data.

This section will provide you with information on uploading data to the Zooniverse, adding metadata to the Subjects in your project, creating data Manifests, and the circumstances which require specific information to be included in your Manifest to support specific project features.

2.3.1 How to upload

This section will cover simple and more complex methods of uploading data to your Zooniverse project.

2.3.1.1 Simple upload

This section describes how to use the simple method to upload data with or without metadata.

2.3.1.1.1 SIMPLE UPLOAD: WITHOUT METADATA

If every Subject you want transcribed is a single image, and if the image file name contains enough information for you to be able to precisely match the Subject to your data export later, then this is pretty easy: you can just drag and drop the images into the Subject Uploader (or click on the 'Upload Subjects' box and select images via your file browser). Please note that, when using the simple uploader, we recommend uploading your data in batches of no more than 1,000 at a time.

2.3.1.1.2 SIMPLE UPLOAD: WITH METADATA

Simple subject upload: with metadata By using a Manifest, you can associate related metadata with each image you plan to upload to your project. Above, you can find more information about creating a Manifest, but for now we will assume the Manifest is created and your files are ready.

Project creators often find it easiest to have all of their Subject images in a single folder along with their Manifest CSV file; you must upload both at the same time. However you organize your files, pay attention to the details and keep careful records, as it may be important for your data analysis later (e.g. when you download your data exports and need to link those back to the rest of your metadata).

Click on the 'Upload Subjects' box and navigate to the folder where you are keeping your images and Manifest CSV file. The easiest thing to do is to select everything in that folder (hit cmd + a on Mac, or ctrl + a on Windows) and click "open" to upload those images and the Manifest file. The project uploader gives you a preview, to indicate the number of Subjects that are ready for upload. If there is a problem with any of your files, or with your Manifest, the preview will indicate the number of files that, for whatever reason, have failed to be created. If this is the case, you will need to troubleshoot before continuing with the upload process.

Click 'Upload' to start the process. When everything is uploaded, you will see a list of all of the Subjects you have created. The numbers to the left are unique identifiers for each Subject (also called Zooniverse IDs), and the icons to the right let you preview or delete each Subject. When previewing a subject, you can also view its

metadata by clicking on the icon below the image shaped like the letter 'i'. Please note that, when using the simple uploader, we recommend uploading your data in batches of no more than 1,000 at a time.

2.3.1.2 Command-line interface + Python client

For teams with more than 1,000 Subjects to upload, the simple (web client) uploading method is not ideal. Zooniverse provides support for batch uploading through a command-line interface (CLI) and a Python library. Both methods offer the same functionality and are officially supported by Zooniverse. Command-line-interface documentation can be found [here](#) and Python documentation can be found [here](#).

2.3.2 Manifests and metadata

This section provides detailed information on how to include metadata in your Subject upload.

2.3.2.1 Why should I include metadata with my Subjects?

There are no requirements for uploading metadata to Zooniverse. You can include as much or as little metadata as you like. Much of the time, though, project creators want to keep some data associated with their images in the Classification interface. For example, if the images you are using need to be displayed with a license, that information can be included with the data you upload to the Project Builder, and viewed by volunteers by clicking on the Subject Info button (the small 'i') in the project Classification interface.

Additional information that can be helpful for volunteers while Classifying (if available) might include:

- * Attribution/authorship
- * Date created
- * Location or other geographic information
- * URL to view the collections record, and/or high-resolution image
- * Archival reference of the material

Some of this information may not be available for your dataset—generating it may

even be a goal of your project. However, if you do have metadata for your images, it can be useful for volunteers for several reasons. First, it can sometimes help with transcription. For example, if your project works with letters and the holding institution has already cataloged the date and information about the writer and addressee, that can be a helpful resource for volunteers trying to decipher possibly difficult signatures. Including the URL to a full catalog record in your Subject metadata can support volunteers who may be interested in learning more about a particular document or its content.

2.3.2.2 What is a Manifest

A Manifest is a file that allows the Zooniverse platform to associate metadata with the images you upload to the Project Builder. Your Manifest must be formatted as a CSV ("comma separated values") file in order to work properly. **Each Subject Set you upload must have its own Manifest.**

Each row of the CSV file should contain information about a single Subject that you're uploading to your project. The first row should include your headers (variable or column names). So, if you're uploading a set of 100 Subjects and want to include their metadata, your CSV should contain 101 rows.

The first column in your table should contain the image filename. This is how the Manifest finds the relevant images in your upload in order to assign the metadata appropriately. If there are filenames in your Manifest that do not correspond to an image file upon upload, it will cause an error. Similarly, if there is a file in your dataset that is not included in your Manifest, it will also cause an error. To avoid errors, double check to be sure the list of image filenames in your Manifest CSV exactly matches the filenames of the images you want to upload. We recommend that you avoid using special characters unless they have a specific use case, e.g. using # in front of a column heading to hide that information on the project front end.

2.3.2.3 Special features

Certain features of the Zooniverse platform require you to include specific information

in your Manifest. The examples below will walk you through the information you need to include in your Manifest to support these features, including links to example Manifests.

2.3.2.3.1 ALICE

ALICE stands for Aggregate Line Inspector and Collaborative Editor. It acts as a means for reviewing aggregations of your volunteers' transcriptions at a line-by-line level through a graphical user interface (see figure; for more information see, section 7). Please read through the ALICE About page for more information on setting up your project for use with ALICE.

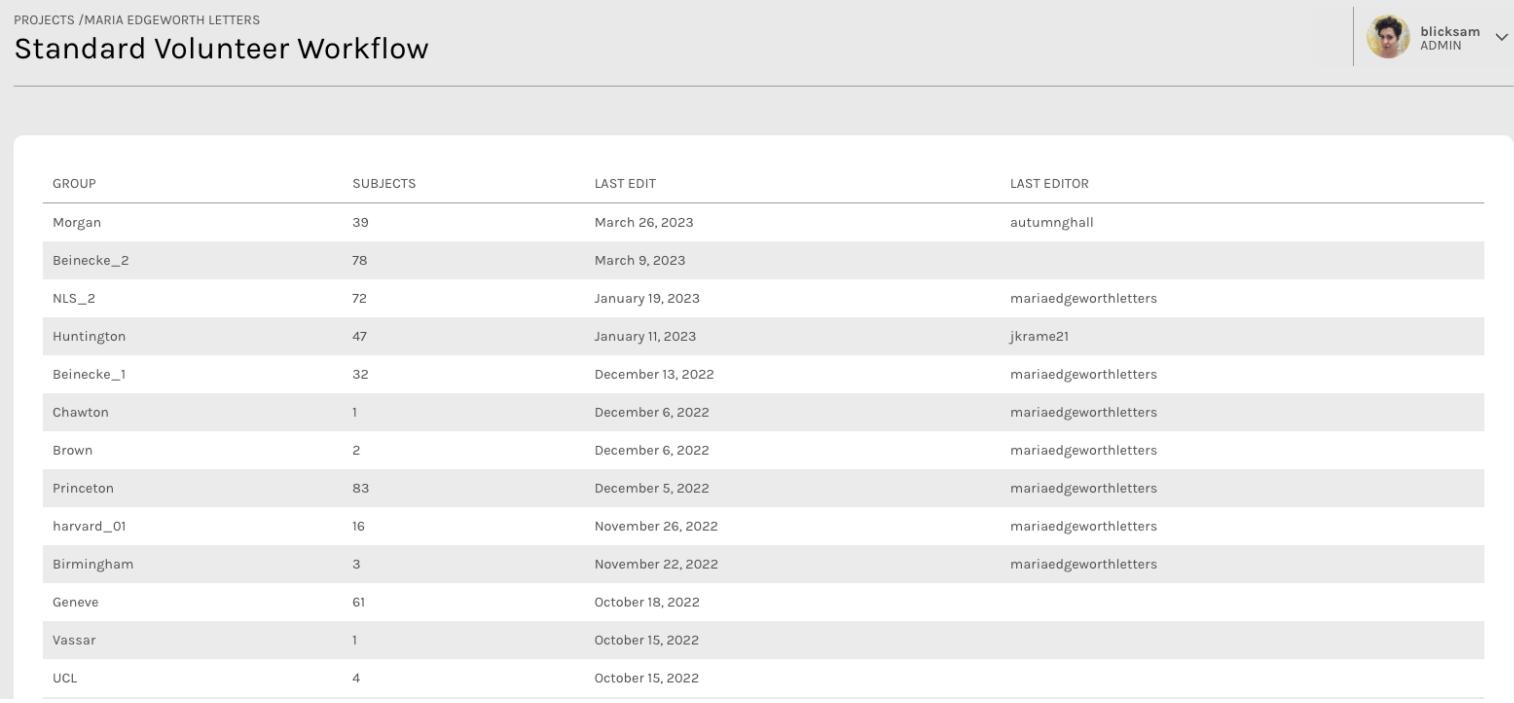
If you are planning to use ALICE, you must include an `internal_id` field in your Manifest. This ensures you'll be able to view and search for an image/item in ALICE by your institutional reference instead of the auto-generated Zooniverse ID. Avoid special characters and spaces.

If you are planning to use ALICE, you must also include a `group_id` field in your Manifest. This ensures you'll be able to view and export smaller groups of images instead of sending all your transcription data to the same ALICE folder. We recommend that you use the same groupings as you do for your Subject Sets in the Project Builder. Avoid special characters and spaces. Note: the `group_id` field will have no bearing on how data is presented to volunteers in your Zooniverse project, it will only impact how you view transcription data in ALICE.

For teams using ALICE, **we recommended that you limit your group sizes (and thereby also Subject Set sizes) to under 100 images**. This reduces the risk of timeout issues during the data export process.

PROJECTS	Maria Edgeworth Letters	blicksam	ADMIN		
WORKFLOW NAME	WORKFLOW ID	GROUPS	TOTAL SUBJECTS	LAST EDIT	LAST EDITOR
Volunteer Workflow - Primarily French Documents	22205	1	28		
Standard Volunteer Workflow	18542	23	517		

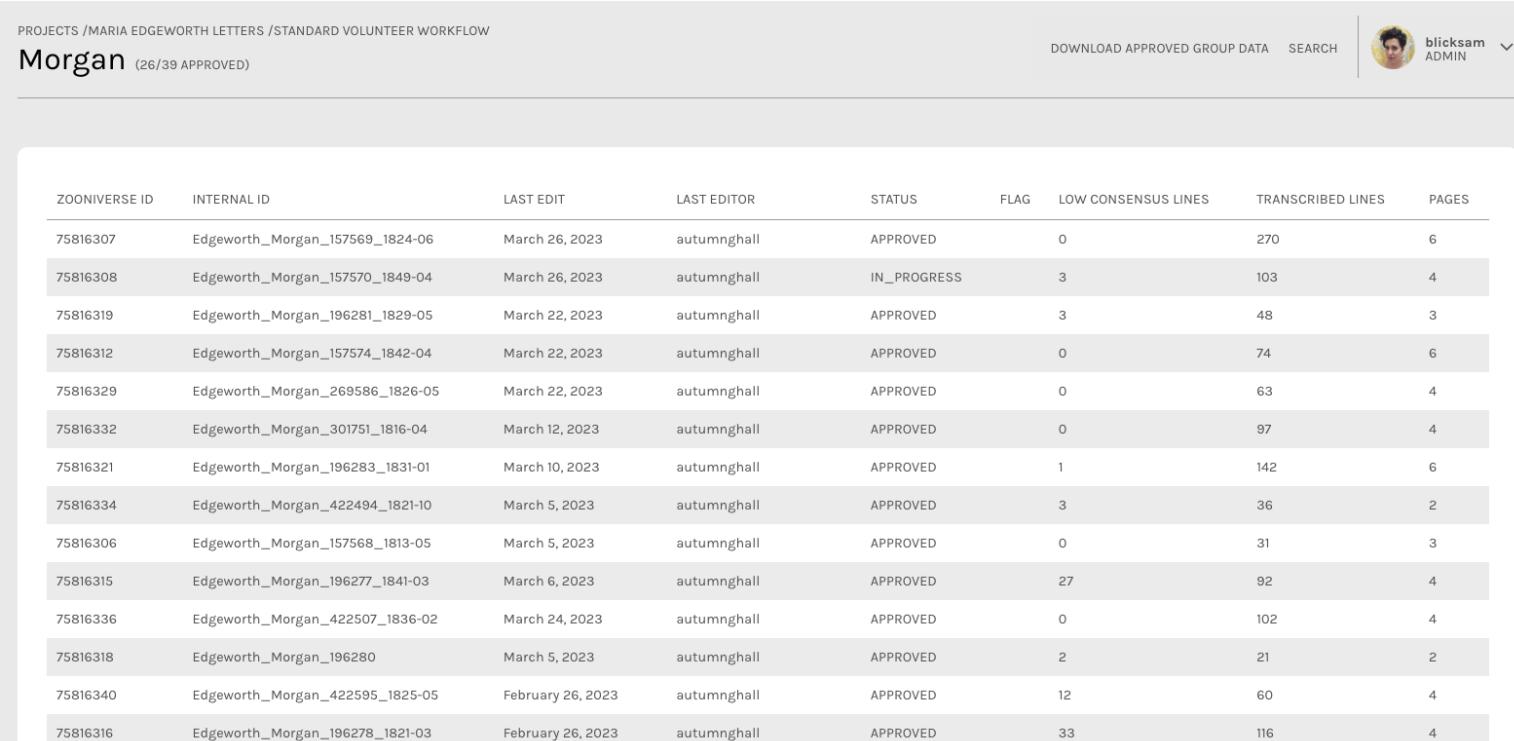
Above: A screenshot of the Workflow View for Maria Edgeworth Letters, Co-PI Hilary Havens.



This screenshot shows the 'Standard Volunteer Workflow' section of the project's data page. It displays a table with columns for GROUP, SUBJECTS, LAST EDIT, and LAST EDITOR. The groups listed include Morgan, Beinecke_2, NLS_2, Huntington, Beinecke_1, Chawton, Brown, Princeton, harvard_01, Birmingham, Geneve, Vassar, and UCL. Each group entry shows the number of subjects it contains and the date of the last edit made by the specified editor.

GROUP	SUBJECTS	LAST EDIT	LAST EDITOR
Morgan	39	March 26, 2023	autumngall
Beinecke_2	78	March 9, 2023	
NLS_2	72	January 19, 2023	mariaedgeworthletters
Huntington	47	January 11, 2023	jkrame21
Beinecke_1	32	December 13, 2022	mariaedgeworthletters
Chawton	1	December 6, 2022	mariaedgeworthletters
Brown	2	December 6, 2022	mariaedgeworthletters
Princeton	83	December 5, 2022	mariaedgeworthletters
harvard_01	16	November 26, 2022	mariaedgeworthletters
Birmingham	3	November 22, 2022	mariaedgeworthletters
Geneve	61	October 18, 2022	
Vassar	1	October 15, 2022	
UCL	4	October 15, 2022	

Above: A screenshot of the Group View for Maria Edgeworth Letters, Co-PI Hilary Havens.



This screenshot shows the 'Morgan' group view page. It includes a header with the group name and a download link for approved data. Below is a table of transcription records for the 'Morgan' group, showing details like Zooniverse ID, Internal ID, Last Edit, Last Editor, Status, Flag, Low Consensus Lines, Transcribed Lines, and Pages.

ZOONIVERSE ID	INTERNAL ID	LAST EDIT	LAST EDITOR	STATUS	FLAG	LOW CONSENSUS LINES	TRANSCRIBED LINES	PAGES
75816307	Edgeworth_Morgan_157569_1824-06	March 26, 2023	autumngall	APPROVED	0	270	6	
75816308	Edgeworth_Morgan_157570_1849-04	March 26, 2023	autumngall	IN_PROGRESS	3	103	4	
75816319	Edgeworth_Morgan_196281_1829-05	March 22, 2023	autumngall	APPROVED	3	48	3	
75816312	Edgeworth_Morgan_157574_1842-04	March 22, 2023	autumngall	APPROVED	0	74	6	
75816329	Edgeworth_Morgan_269586_1826-05	March 22, 2023	autumngall	APPROVED	0	63	4	
75816332	Edgeworth_Morgan_301751_1816-04	March 12, 2023	autumngall	APPROVED	0	97	4	
75816321	Edgeworth_Morgan_196283_1831-01	March 10, 2023	autumngall	APPROVED	1	142	6	
75816334	Edgeworth_Morgan_422494_1821-10	March 5, 2023	autumngall	APPROVED	3	36	2	
75816306	Edgeworth_Morgan_157568_1813-05	March 5, 2023	autumngall	APPROVED	0	31	3	
75816315	Edgeworth_Morgan_196277_1841-03	March 6, 2023	autumngall	APPROVED	27	92	4	
75816336	Edgeworth_Morgan_422507_1836-02	March 24, 2023	autumngall	APPROVED	0	102	4	
75816318	Edgeworth_Morgan_196280	March 5, 2023	autumngall	APPROVED	2	21	2	
75816340	Edgeworth_Morgan_422595_1825-05	February 26, 2023	autumngall	APPROVED	12	60	4	
75816316	Edgeworth_Morgan_196278_1821-03	February 26, 2023	autumngall	APPROVED	33	116	4	

Above: A screenshot of the Subject View for Maria Edgeworth Letters, Co-PI Hilary Havens.

2.3.2.3.2 SEQUENTIAL CLASSIFICATION

This feature will present the Subjects in a given set to your volunteers in the order you establish within your Manifest. You must add a `priority` field to your Manifest if you want volunteers to be able to classify Subjects in sequence. The contents of this field must be integers.

To complete setup, once you've uploaded your data with the required field, **you must send an email to contact@zooniverse.org requesting that Sequential Classification be turned on for all applicable workflows**. Please include your project ID number as well as the ID number of all the Workflow(s) you want set up for Sequential Classification.

2.3.2.3.3 MULTI-IMAGE SUBJECTS

Multi-image Subjects offer an alternative to Sequential Classification by presenting a group of related images as a single Subject. This option is particularly helpful for teams who are working with data that has existing sub-relationships within a given Subject Set, such as letters.

The number of image fields you need to include in your Manifest depends on the maximum number of images a Subject in a given set may have. If other Subjects in the set do not have the same number of images, leave those fields blank (e.g. if one Subject only consists of a single image but others in the Subject Set consist of five images, include the image name in the first field in that row, and leave the other four blank).

2.3.2.3.4 INDEXED SUBJECTS

Indexed Subjects allow volunteers to choose an individual Subject from a list and work through a Subject Set in sequential order (nb: Sequential Classification is required for teams who want Indexed Subjects). This option is helpful for teams who have useful metadata already available for their Subjects, and who think their volunteer communities will want to choose what they want to transcribe. Examples might include volunteers who want to find a particular author's work within a collection, or documents from a particular year.

To indicate which metadata fields should appear in the index, simply put a % in front of the heading of the column you want to include. To learn more about Indexed Subjects, read this blog post.

2.3.2.4 Resources for creating Subject Sets and Manifests

Scripts

Multi-Image Manifest generator (Python Script)

Manifest examples (Google Sheets - view only)

ALICE

ALICE + multi-image Subjects

ALICE + Sequential Classification

Indexing + ALICE + Sequential Classification

Indexing + Sequential Classification

Sequential Classification

2.3.3 Breaking down your data into different Subject Sets

This section provides guidance on how to organize your dataset into smaller Subject Sets before uploading these sets into your project.

2.3.3.1 Relating data to Tasks

Most often, your data will dictate the kind of Task or Workflow structure that should be created for your project. When designing Workflows, keep in mind:

1. Your research interest
2. Usability (from a volunteer perspective)

For transcription projects, a useful default for organizing Subject Sets is to mimic or slightly adapt the original archival or manuscript structure of the handwritten material. This could mean grouping by year, author, topic, etc. For example, your image collection may include 1,000 manuscript pages from an archival collection, originally held in two boxes with 10 folders each. In this example, you could break this dataset down into two Subject Sets (i.e. one for each box) or 20 (i.e. one for each folder).

If all the pages have a similar structure and layout you can send them all through the same transcription Workflow. If the pages vary widely, you may need to create different Workflows to facilitate capturing the page layout and content (e.g. registers will require a different transcription approach than diaries).

Depending on the types of information you want to extract from an image you may have multiple Workflows that connect to the same set of images. For example, an initial Workflow asking volunteers to classify an image by document type can be helpful for breaking down a large dataset into smaller Subject Sets, each with its own specific Workflow. Because document type recognition can rely on relatively coarse distinctions between images, this may be a task you can experiment with using computer recognition for.

Workflows can be connected with as many Subject Sets as you like, depending on the volume of data you want to have transcribed. For example, you may choose to launch the project with your entire dataset linked to a single transcription Workflow, or start with a subset of your total data, and 'release' more Subject Sets as others are completed.

Because transcription work can proceed quite slowly, transcription projects benefit from having many smaller Subject Sets, rather than fewer large Subject Sets. By organizing your material in this way, you are more likely to obtain complete transcriptions of related material sooner. Returning to a previous example, if you broke up the set of 1,000 images into two groups of 500 Subjects, it would take a long time for each to be completed. If you broke up the set of 1,000 images into 20 folders with 50 Subjects each, the small groups would be completed more quickly,

especially if you released the data in stages.

This type of Subject Set creation also ensures that volunteers will also encounter related material as they transcribe, which may increase engagement and retention.

2.3.3.2 Subject Set selection

If you want your Subject Sets to remain separate (i.e. you don't want to group them all together within the same Workflow), you may want to consider using Subject Set Selection. This is a platform feature that allows you to let volunteers choose what Subject Set to work on within a single Workflow. If you have a single transcription Workflow that has multiple Subject Sets associated with it, simply send an email to contact@zooniverse.org with your Project ID and the relevant Workflow ID, and we can turn this on for you.

Case Study: The ins and outs of creating Subject Sets

Poets & Lovers is a project to transcribe the diary of 'Michael Field,' created by two women novelists Katharine Harris Bradley and Edith Emma Cooper.

Peter Logan, Principal Investigator:

"Identifying and preparing subject sets is probably the most time-consuming task in building your project, and you want to do it right the first time, if possible, so go slowly. Below are some tips for thinking through the process.

1. *Type your subjects.* For example, pages with tabular data might be one type, prose pages another, audio-visual data a third.
2. *Look for subtypes within each group.* One of the most useful lessons I learned was to separate multi-page subjects, like correspondence, from subjects that could be treated as individual pages. In my case, those were diary pages. This led to the creation of two workflows, one that was tailored to multipage subjects, where keeping the images in order mattered (who was writing and to whom?), and a second designed for single diary pages (where we know that context).

- a. Note: Even the latter can be served up more or less consecutively, using a priority column in your manifest, but once subject #1 is completed and removed, volunteers will not be able to access it before transcribing subject #2.
3. *Establish the ideal quantity for subject sets.* With your basic types established, think about how many images to include in each subject set. I made my sets too large initially and learned that volunteers like to work toward a goal, such as contributing to completing a set, so you want a quantity that scales to that imagination, and not one so large that each day's work is an unrewarding drop in an oversized bucket.
4. *Remove unneeded images.* We had images of blank pages in our comprehensive collection, for example. It sounds obvious, but images that do not need transcription should be removed.
5. *Optimize your images.* It pays to have someone on your team who knows how to modify images, since volunteers cannot transcribe what they cannot read.
 - a. Which image dimensions best filled an average screen in the Zooniverse format? 900 x 400 worked well.
 - b. Which level of compression reduced file size the most while retaining high resolution when enlarged? For JPEG files, 65% showed no deterioration when enlarged on a big monitor.
6. *Create your manifest.* You can build in whatever program you want, but it has to be saved as CSV. (We learned the hard way that Zooniverse will not accept a manifest in Excel's spreadsheet format, .XSLX.) This was an easy process, but consider what data you want to share with volunteers. Extra columns in your manifest are visible to users, unless you hide them with a hashtag. We wanted to include a link to the original images in an external archive, for example. It now appears in the information window for the image, along with other relevant data, like the date and volume it came from.

2.3.4 When should I upload my Subject Sets into Zooniverse?

Start by uploading a small dataset (<50 Subjects), to use for building and testing Workflows. You do not need to have all your Subjects before you get started. Starting with a fraction of your total dataset allows you to experiment with different aspects of your project—remember, your Workflow design should influence how you break your full dataset into Subject Sets. In the Subject Set you use for building and testing your Workflows, you should include examples of the different formats and layouts volunteers can encounter when working through your full dataset. A full data upload should be the final thing you do before your project launches.

2.3.5 Beta testing Subject Sets

Beta testing requires a small Subject Set. For most projects, this is usually between 150 and 300 Subjects. For projects using the Transcription Task, it is usually about 25 Subjects. Beta tests last for one week, so the aim is to include enough data to last more than a few days, but not so much that nothing is retired by the end of the week-long test. The goal of beta testing is to validate the results from your Workflow(s), while also verifying that volunteer expectations are clear throughout the project and the user experience is positive.

Made with Material for MkDocs

Zooniverse Help

Workflows

3.1 What is a Workflow?

Volunteers on Zooniverse projects contribute to data creation by examining a piece of media and performing activities that systematically identify features of the media. For transcription projects, the media is an image of text, typically handwritten, which cannot be read by a computer.

You will ask volunteers examining your images to do one or several small things when examining each image. As well as transcribing the text, these might include answering questions about content on the page. Each of the actions volunteers perform when examining an image (or more generally, any other type of media) are known as a Task. Tasks are then connected in a logical sequence to create a Workflow.

Definitions: **Task:** Any action a volunteer performs when examining a Subject.

Workflow: This is a series of Tasks that a volunteer is asked to complete when presented with data in a Zooniverse project. This can be one Task or multiple Tasks, depending on the project.

See glossary for further elaboration on definitions

Effective Zooniverse projects consider the following principles in designing Tasks and Workflows:

- * Tasks are simple and can be understood with minimal guidance.
- * The successful completion of a single Task by a volunteer contributes useful data even if the volunteer submits their Classification without completing the entire Workflow.
- * Tasks are connected in such a way that nearly all relevant information a volunteer might encounter can be Classified for the research team.
- * The collection of complex or extensive data from a single Subject can be addressed by developing multiple

Workflows.

3.2 Designing Workflows for the data output you want

This section contains information on the relationship between your desired project output and the process of designing your project Workflow(s).

3.2.1 Creating a plan for use

The Zooniverse Project Builder provides creators with significant flexibility in how they ask volunteers to identify information (or data) on an image. There are often several ways to organize the collection of the information you want. You should familiarize yourself with the different Tools in the Project Builder by reviewing existing projects and experiencing them as a volunteer. Experiment with these Tools in the Project Builder to see how different Tools work with your images.

As a project creator you should be familiar with the text you are asking volunteers to transcribe. Content drives Workflow design, so even if discoverability is a goal of your project, you should put in the effort beforehand to look through the materials to develop a general understanding of what your dataset contains. For example, if you are transcribing letters you may want to have volunteers separately identify text that is dates, names, or addresses, so that this information can be easily identified.

We strongly recommend that before beginning to create Tasks and Workflows in the Project Builder you:

1. Identify the data that you want to produce at the end of the project.
2. Identify Tasks in the Project Builder that volunteers can use to provide this information. At this stage, it is extremely helpful to identify existing projects that have similar Tasks and Workflows to the ones you want to use.
3. Outline the Tasks and Workflow on paper or in software that allows you to sketch the sequential flow between Tasks. Consider whether Tasks might need to be

repeated to collect multiple examples of the same type of information.

4. Practice the Workflow with your data and consider whether it actually produces the data you identified as your ideal output. Request a Data Export from your test Workflows to confirm you are able to get what you need.
5. Consider what information can be identified by programmatic post-processing of the data. For example, instead of asking volunteers to identify dates, names and addresses in letters you might be able to programmatically identify these through their content. Dates are generally recognizable as a string of numbers, or contain numbers and words that do not appear in other contexts. Dates, and names and addresses of senders and recipients may also appear in distinct locations on a letter that can be identified via the coordinate locations of the text.

3.2.2 Balancing the volunteer experience with prioritization of data quality

When designing Tasks and Workflows:

- Try to see the material from the perspective of someone inexperienced but eager to learn. As an expert you may have identified explicit or implicit structure in the text that you wish to have represented in your transcriptions. However, volunteers will not necessarily know the structures you wish to extract. The Tasks and Workflows you design should guide volunteers through the transcription process.
- Consider how often volunteers will encounter a particular type of text. A situation that volunteers may encounter in 1 of every 50 images is relatively rare, but worthy of consideration in your design. Tasks, Task options, and Workflows to address situations that volunteers encounter once in 1,000 images may compromise the volunteer experience by asking volunteers to perform a lot of additional work for little apparent reward.

Well-designed Tasks and Workflows can help volunteers navigate a large amount of complex text and transcribe it into a format that supports your research needs.

However, even with well-designed Workflows you should remember that volunteers have limited time. Long Workflows are more likely to be left incomplete by volunteers. If your material can be divided into several shorter, distinct Workflows you may be able to reduce the amount of incomplete transcriptions submitted.

Workflow design is also an ethical consideration (for more on this, see section 4.1.1). Putting in additional effort during the design stage can result in a better experience for volunteers once the project has launched. When you design Workflows with humans in mind, you communicate that you are respectful of the effort volunteers are donating to your project.

3.3 How to design a Workflow

3.3.1 Working from your content

The Zooniverse Project Builder includes several Task types that can be presented to volunteers to capture textual information. The Tasks that are frequently used in crowdsourced transcription projects are:

- Question Tasks
- Dropdown Tasks
- Transcription Tasks

3.3.1.1 Question Tasks

The screenshot shows a Zooniverse project page for 'Corresponding with Quakers'. The main content is a handwritten letter from 1841, transcribed with blue circular annotations. The sidebar includes navigation links like 'Home', 'Help', and 'CORRESPONDING WITH QUAKERS', and project-specific links like 'ABOUT', 'CLASSIFY', 'TALK', 'COLLECT', and 'RECENTS'. A 'FIELD GUIDE' section lists categories such as 'Quakerism/religion' (which is highlighted in teal), 'business/finance/household accounts', 'travel', 'education', 'slavery/abolition/colonialism', 'gender/women/female preaching', 'American Revolution/French Revolution', 'politics', 'books/reading/authorship', 'letter writing/the past', 'family and friends' (also highlighted in teal), and 'other'. At the bottom right, there are buttons for 'Done & Talk' and 'Done'.

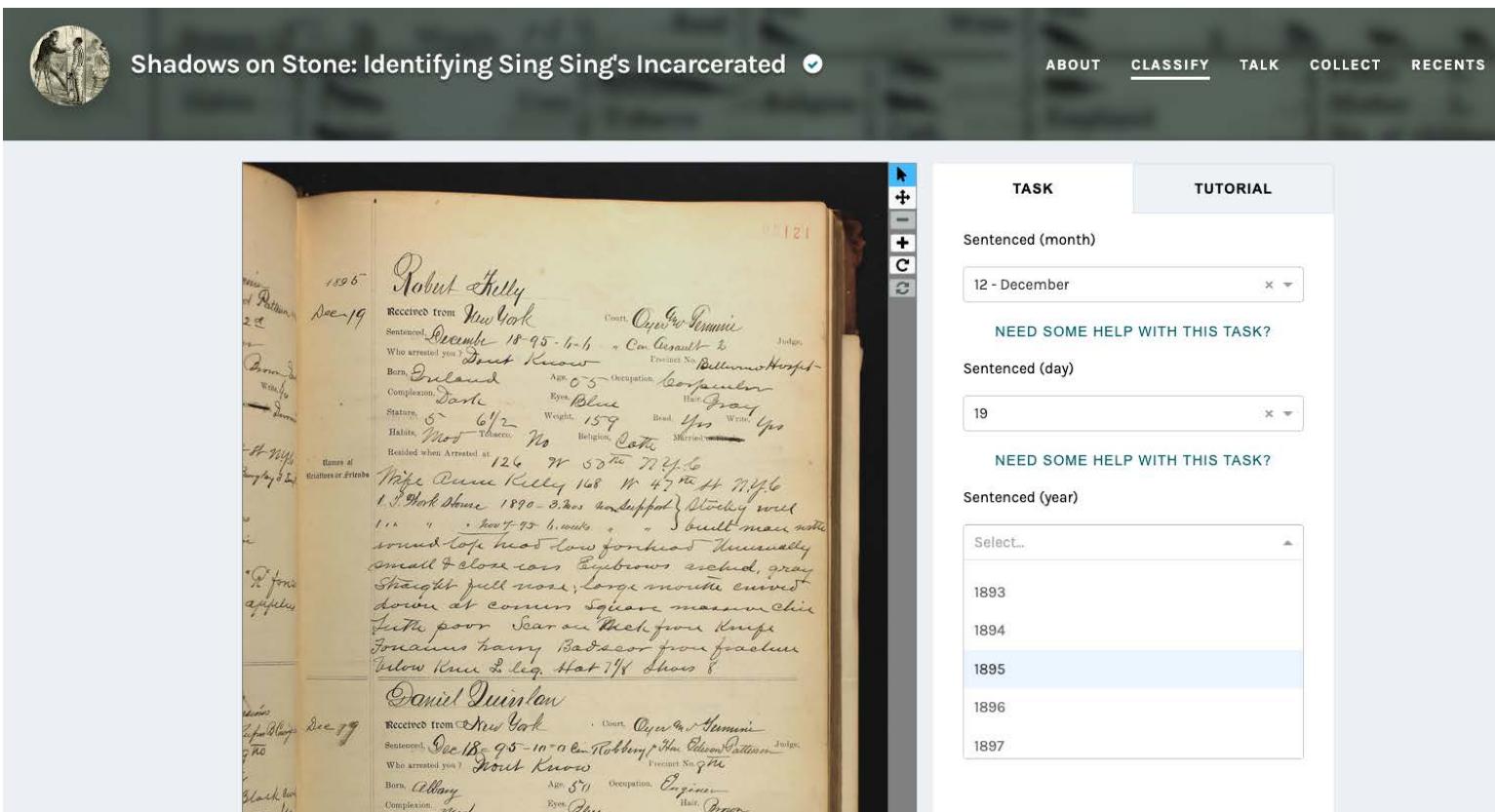
Above: A screenshot of a question task on a Zooniverse project page. From *Corresponding with Quakers*, Principal Investigator and Project Director Rachael Scarborough King.

Question Tasks present users with a question and a set of predefined answers. The minimal number of answers to present is two—if there was only one option, you would know the answer already. We recommend that you try to avoid giving users too many options (less than 6 is ideal), though there are situations where this may be required, especially for ‘select all that apply’ questions (the screenshot above is a good example). The default behavior is to only let volunteers choose a single response—if you want volunteers to select all that apply, be sure to choose ‘allow multiple.’

You can use Question Tasks to create different paths through a Workflow depending upon a volunteer’s response. For example, if you are transcribing manuscript material which occasionally has pictures, you may ask volunteers a yes/no question about whether a picture appears on the page. If their response is ‘Yes,’ you may ask a follow-up question about the content of the picture. This branching structure hides

irrelevant Tasks from volunteers. Read more about how to link Tasks in a Workflow via the main Help page.

3.3.1.2 Dropdown Tasks



Above: A screenshot of a dropdown task on a Zooniverse project page. From *Shadows on Stone: Identifying Sing Sing's Incarcerated*, Project Director Roger Panetta.

Dropdown Tasks allow volunteers to efficiently enter standardized text entries. For example, you may have material with a limited list of place names or personal names that occur within the material. From your knowledge of the material you can define a list of pre-specified options that volunteers can select from. Dropdown menus are an efficient way of controlling the entry of elements such as dates for which you want to limit the values that may be selected. For example, if you are transcribing letters from a restricted time period the range of years for which you will accept a letter being written may be limited. Dropdown Tasks constrain users to these values, and potentially lead to less error.

3.3.1.3 Transcription Tasks

The screenshot shows a transcription task on the People's Contest Digital Archive. On the left, a scanned page from a journal is displayed. The page contains handwritten text with several underlined sections. On the right, a digital interface for transcribing the text is shown. At the top right are links for 'ABOUT', 'CLASSIFY', 'TALK', 'COLLECT', and 'RECENTS'. Below that is a 'TASK' tab and a 'TUTORIAL' tab. A pop-up window titled 'Transcribe the line of text that you've marked.' contains the text 'Jan. for prayer besides the row of chairs arranged in the front of the room. We had our warmest seasons'. Below this, there are 'Text modifiers' buttons for 'UNCLEAR', 'DELETION', and 'INSERTION'. A text input field contains 'Jan. for prayer besides the row of'. At the bottom of the pop-up is a question: 'Have all the volunteer-made underline marks turned grey?' with 'Yes' and 'No' buttons. The overall interface is designed for collaborative transcription.

Above: A screenshot of a transcription task on a Zooniverse project page. From *People's Contest Digital Archive*, Project Coordinator Kevin Clair.

The Transcription Task is used to allow volunteers to collaboratively mark (underline) and transcribe lines of text. The results of this Task type include positional data as well as the transcription itself. This allows the research team to not only process the transcribed text, but also its location relative to other lines of text. For example, volunteers presented with a page from a journal can transcribe each line of that journal and the research team will be able to see that they are sequentially located on the page.

Once a line has been transcribed, it will persist on the image, meaning that other volunteers will see it when they encounter that Subject. The transcription can then be reviewed by other volunteers, who will have the option to either select it from a dropdown menu—indicating that they agree with the initial transcription—or input a new transcription either by selecting and editing the previous transcription or typing an entirely new one. Once the text has been agreed upon by the set number of volunteers (see Retirement Limits in section 6) the drawn line will turn gray, indicating it has reached consensus and no longer needs additional review. No additional

transcriptions may be submitted for a line once it has turned gray.

This is the most common Task type used in transcription projects. Projects featuring this Task will also be able to use the Aggregate Line Inspector and Collaborative Editor (ALICE) to inspect and edit crowdsourced transcriptions after they have reached their retirement limit.

For detailed instructions on how to set up your project using the Transcription Task, see this Google Doc.

Made with Material for MkDocs

Zooniverse Help

Designing for volunteers

4.1 Who are the volunteers?

Volunteers are the most important part of any crowdsourcing project. This section will provide guidance around user-centered project design.

4.1.1 Building an ethical project

Research crowdsourcing differs from traditional micro-work platforms (e.g., Amazon's Mechanical Turk) in that volunteers are not paid for their labor. Because of this, it is your job as a project creator to consider how your volunteers benefit from taking part in your project. Below are some important ethical considerations to bear in mind while crafting your project.

Rebecca Schneider, Associate Director, *The Gravestone Project*:

"Backward, or User-centered design principles helped us to conceptualize who else might use the data we hoped to generate and how that data could be used in other disciplines, as well as by non-academic stakeholders. It was thrilling to consider design choices with these other users in mind. It also struck me as an essentially ethical concern to plan for both anticipated and unanticipated consumers of the data (to the extent that's even possible)."

4.1.1.1 Know your values

According to The Collective Wisdom Handbook (Ridge, Blickhan, Ferriter et al.), reflection on the power dynamics present within crowdsourcing projects can help to empower crowdsourcing project creators to "create equitable and transformative

spaces, communication, and activities." We strongly encourage teams to establish their values at the start of their project planning, and to revisit these values throughout the project lifecycle. As the Handbook also states, "Failing to articulate what is important, and how you will demonstrate it, will lead to you missing a vital component of how your project is organized."

4.1.1.2 People-forward thinking

Keep volunteers in mind at all times. Part of creating a good crowdsourcing project is considering the perspective of the people who will be taking part in your project. How can the design choices you make create a positive user experience? A great way to ensure you're engaging in people-forward thinking is to try out other projects – creators who take part in other crowdsourcing projects as volunteers tend to have an easier time thinking about the volunteer experience.

4.1.1.3 Don't waste volunteers' time

You must have a plan for working with your data before you launch your project.

The most common problem for transcription projects is teams not being prepared to use their resulting data. Before you involve the public, you must have a plan for how you will aggregate, review, clean, and store the transcriptions produced through your crowdsourcing project. If you do not know how you will do one or more of these things, you risk asking the public to donate their time, only for the data never to be used. Wasting volunteers' time is the cardinal sin of crowdsourcing, and something to avoid at all costs.

4.1.1.4 Always experiment and iterate (even before Beta testing)

Putting in the additional effort early and often in the process will require more labor for you and your team, but it will almost always result in a better experience for your volunteers—and higher-quality results. You should walk through your Workflows and Tasks at every step of your process (even before building your project in Zooniverse). It is almost always helpful to also have people less familiar with your data go through

the Workflow to uncover assumptions you may make based on your own expertise.

Case Study: Working with difficult materials

The Crocker Land Expedition focuses on a 1913 exploration of the existence of a huge island called Crocker Land that famous explorer Robert Peary claimed to have seen. Although we now know the land does not exist, the expedition to the area produced important scientific research. As part of the BCCCT cohort Rebecca Morgan explored how the Zooniverse could be used to create a platform for transcribing the diverse material gathered by the expedition, much of which is held at the American Museum of Natural History.

Rebecca Morgan, Project Director:

"I wonder about building a community of people that are working on documents that might potentially over you know, days or an hour or whatever, affect them. The [subject sets] may seem innocuous, innocuous on the outside, but after immersing yourself in however many profiles of Sing Sing prisoners or a whole tome on interactions with Indigenous folks and animals, they can take a toll. Especially if you're not ready for it, it can take its toll. And I'm not saying necessarily that it's a bad toll and seeing how important the close reading, skill development... These materials demand some perspective, some thoughtfulness, but I think if someone's just jumping into this to pass the time, or for fun, I don't know if we need trigger warnings on stuff, but we need to figure out a way to work with it and what we owe our community as far as being explicit about that."

4.1.2 Finding your community

One benefit of using Zooniverse for your crowdsourcing project is the built-in community of millions of potential volunteers. However, not all registered Zooniverse volunteers will take part in your project—you should expect to do some recruiting of your own. Ideally, you want a mix of people familiar with the Zooniverse platform as well as those who bring existing expertise or interest in your subject matter.

Publicizing your project on subject-specific listservs, community message boards and via social media accounts can help create interest from others with an existing interest in your project topic. Finding a blend of these groups will help to create a well-balanced community who will not only help your project, but can assist one another with technical questions as well as tricky transcriptions. Read more about building communities in Section 6.

4.2 Home page

A major part of designing a successful project is writing effective copy. On the Zooniverse platform alone, there can be dozens of other projects to choose from, let alone the rest of the internet—what makes yours special? Why should people donate their time to helping you?

Take time choosing a project name and writing the project description and Introduction. These are the first things people will see when they visit your project, so make sure they are accessible and interesting. Avoid long exposition and project history, as this information can go in the About page. Use action words, avoid jargon, and make sure to mention your main Task type. For example, in this case, “Help us transcribe XYZ...” is more effective than “We are digitizing XYZ and want to create digital searchable text via crowdsourced transcription.” For more on writing an effective call to action, see The Collective Wisdom Handbook chapter 9, “Supporting participants.”

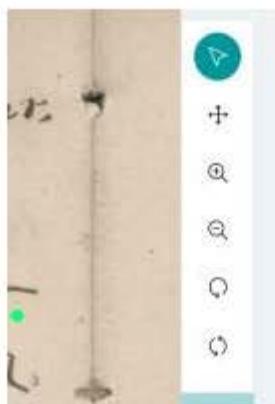
4.2.1 Making use of images

The images you include in your project (illustrative images that aren’t Subjects to be Classified) should help volunteers better understand what you are asking them to do (e.g. screenshots of the project’s user interface), or illustrate concepts you are trying to explain (e.g. page layout, script types, etc.). Upload images via the Media tab of the Project Builder. You can resize images before uploading, or after uploading using Markdown. We recommend keeping your Media images under 500KB, to avoid long

load times. Modifying screenshots of your user interface can quickly help volunteers understand what you are asking them to do.

TASK

TUTORIAL

**Image manipulation tools**

The buttons at the top right allow you to move (pan), zoom, and rotate the image. Hovering your mouse over each tool will tell you what it does.



Above: A Tutorial screenshot from *People's Contest Digital Archive*, Project Coordinator Kevin Clair.

4.2.1.1 Media

The Media tab of the Project Builder allows you to upload media to your project (such as example images for your Help pages or Tutorial) so you can link to it without an external host. To start uploading, drop an image into the gray box (or click 'Select files' to bring up your file browser and select a file). Once the image has been uploaded, it will appear above the 'Add an image' box. You can then copy the Markdown text beneath the image into your project, or add another image.

4.2.2 Markdown

The Project Builder supports Markdown, a markup language that allows you to create formatted text via a plain-text editor. The Project Builder features a short Markdown guide, which you can reach by clicking the '?' icon in the upper right corner of any text entry field. A more detailed Markdown guide is available at <https://help.zooniverse.org/next-steps/markdown/>

4.3 Talk

'Talk' is the name given to the message board system that comes with every Zooniverse project. This section will explain what Talk is, and how to use it to the advantage of your project as well as your community.

4.3.1 What are Talk boards for?

'Talk' is the name for the discussion boards attached to your project. On Talk, volunteers can discuss your project and data with each other, as well as with you and your team. **Maintaining a vibrant and active Talk is essential for keeping your volunteers engaged with the project.** Conversations on Talk can also lead

to additional research discoveries, as volunteers bring their own interests and expertise to the project.

4.3.2 How to build Talk boards

In the Talk tab of the Project Builder, first activate the default Subject-discussion board, which hosts a single dedicated conversation for each Subject you upload to the project. Then add additional boards for conversations about a general topic. Examples might include: 'Announcements,' 'Project Discussion,' 'Questions for the Research Team,' or 'Technical Support.'

To get email notifications when volunteers post on Talk, make sure you 1) have your Notification preferences turned on, and 2) have checked your Talk email preferences. Both can be adjusted at zooniverse.org/settings (this page is also accessible by clicking on your username in the top right corner of the screen).

4.3.3. Talk examples

PLACEHOLDER

Case Study: Keeping up with Talk

The People's Contest Digital Archive is a project led by the Richards Center at the Pennsylvania State University to transcribe Civil War era documents from Pennsylvania. Transcribing through the Zooniverse platform was an opportunity to broaden access to documents in collections at Penn State and elsewhere in Pennsylvania.

Kevin Clair, Project Coordinator:

"It is important to keep up with the Talk boards in your Zooniverse project every day, especially at the outset (after your launch). There will be a great deal of activity, and most of it will be questions about Zooniverse itself, or about the materials you have

made available for your transcription project. The Zooniverse team are very quick with responses to technical questions if you forward them along, but only you will be able to answer questions about your materials (and you will get them!).

You will notice, after your project has been around for a while, that your Talk boards will take on a life of their own and you will not need to monitor them as closely. Most of the responses we have received in the People's Contest Digital Archive have been to point out blank pages for others, or to note pages that have been damaged or are otherwise unreadable. Also, some volunteers will also be contributing to other digital humanities projects on Zooniverse, and will be able to provide answers to people's questions or help to fix problems they are having with the platform—in many cases even before you or the Zooniverse team know that they're a problem...

Our Talk boards have also led to improvements to the project itself. Once enough people have asked the same questions about how to transcribe something, it's time to add it to our Tutorial or Field Guide so we have a resource to which we can point people."

4.4 About pages

The About pages are a space for you to share details about your project, its history and outcomes. This section will provide details on what type of information to include, and how About pages can be a source of support for project volunteers.

4.4.1 Research

This section should include the project background (including any funding sources), brief historical context (what will this work contribute to the field?), project goals (including any related work happening outside the Zooniverse project), and planned research outputs (in particular, where volunteers can plan to access project results). Essentially, this is where you have space to justify to volunteers: 1) why you need their help; and 2) how their efforts will affect the research.

4.4.2 Team

Who is working on this project? This section should list all project team members, their institutional affiliation, research specialization, and specific role within the project. Photos are encouraged, if team members are comfortable sharing. Additionally, if there is a specific email address volunteers should use to contact the team, this is a great place to put that information (if you haven't yet, you may want to consider establishing an email address for your project). The Zooniverse username of all Collaborators will be listed here as well.

4.4.3 Results

This section should include links to results when available, and will ideally also include a short description of project results. Before results are available (e.g. at the project launch), we encourage you to signpost where you are planning to host the project results. This communicates to potential volunteers that you understand the platform requirements for making data publicly available and are committed to doing so.

4.4.4 Education

This section is where you can share resources for educators, students, or anyone interested in learning more about your research topic. It is a great place to include additional reading or considerations. You can also post resources that may be used in conjunction with your project, such as syllabi and pedagogical tools.

4.4.5 FAQ

This section should list frequently asked questions about your project. You can start by including anticipated questions, but the expectation is that you will add to this list after your initial beta test, as well as after the project's public launch.

- Keep FAQ items concise and easy to read.

- Bulleted lists are preferable to paragraphs of text.

4.5 Tutorial, Field Guide, and Help Text

You can support volunteers with additional information in three places. There may be some overlap in the content of these resources, and volunteers will encounter them at different points in their work.

Content for each of these resources is rendered with Markdown, so you can use any media you've uploaded for your project in this space. Upload media for your project (illustrative images that aren't Subjects to be classified) in the Media tab of the Project Builder. Visuals help tremendously in making your support resources easier to understand.

4.5.1 What is a Tutorial?

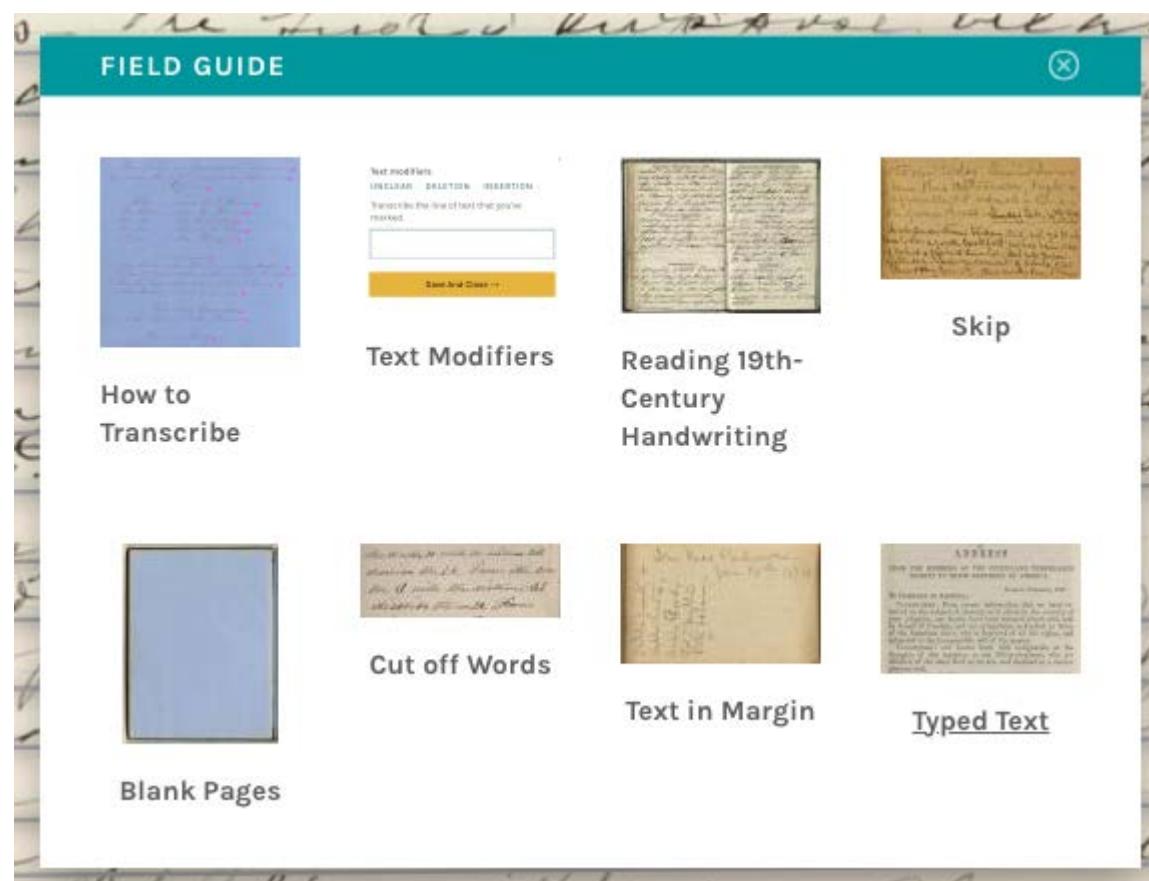
Tutorials appear the first time a volunteer opens a Workflow on your project. You can use the Tutorial tab to build a step-by-step guide to participating in your project. Tutorials are Workflow-specific; if your project has multiple Workflows, you should design a Tutorial for each one. Volunteers should come away with an understanding of your basic project goals (why are you doing this?) and what they will be asked to do if they choose to participate. You can upload images and enter text for each step. There is no limit to how many steps you can add, but we encourage you to aim for 6 steps or fewer. Remember: the longer your Tutorial, the less likely it is that volunteers will read the whole thing.

Assign Tutorials to your Workflows via the Workflows tab of the Project Builder. More info on writing Tutorials is available on the Zooniverse blog.

IMAGE PLACEHOLDER

4.5.2 What is a Field Guide?

The Field Guide is a space for you to provide information on the project content that helps volunteers become more proficient at interpreting your source material. Examples might include frequently-used abbreviations, lists of common places and names, or explanations of period-specific writing conventions (e.g. the long s, cross-writing, etc.) Information can be grouped into different sections; each section should have a title and an icon. The Field Guide can also host a general background to your material and research Subject.



Above: A screenshot

of a Zooniverse Field Guide topic list from *People's Contest Digital Archive*, Project Coordinator Kevin Clair.

4.5.3 What is Help Text?

You need to provide Help text for all Tasks in your project Workflow(s). This information should help volunteers complete the Task they are currently being asked to complete. It is a space for specific instructions on how to use the tools provided for that particular step of the Workflow, not necessarily how to interpret the content (that's what the Field Guide is for).

Case Study: Talk support

Letters to and from authors provide scholars and readers with new insights into what authors were thinking about that didn't make it into print. In the *Maria Edgeworth Letters* project, transcribing letters held in one place makes the material more widely accessible and provides volunteers with exciting opportunities to learn more about the thoughts behind the printed word.

Hilary Havens, Co-Principal Investigator:

"One of the most challenging and time-consuming, but also one of the most rewarding parts of the project has been interacting with volunteers on the Talk Boards. This becomes relevant as soon as your project enters beta testing, but you should work on the Talk Boards even before then. My collaborators and I thought we did a good job with our Tutorial, Help Text, and Field Guide, but once we started seeing comments from the volunteers, we realized how wrong we were! Many of the volunteers – especially the beta testers - were much more experienced than us in dealing with Zooniverse projects and knew exactly what we were missing. Our Help Text for the transcription tool needed to be more specific and have more pictures. We needed to cover many more edge cases in the Field Guide (such as tables, damaged pages, and the British pound sign). The Talk Boards helped us improve our project immeasurably. They also take a tremendous amount of time – especially in the beta period and early weeks of the projects – as the volunteers help the project team explain and improve things.

The Talk Boards also helped us provide guidance to the volunteers beyond the Field Guide, Tutorial, and Help Text. Maria Edgeworth and her circle have difficult handwriting, and we devoted an entire section of the boards to a handwriting guide that an undergraduate research assistant undertook as a project. The guide was very popular with volunteers and editors alike adding numerous entries linked to weird writerly quirks, like Edgeworth's k's that look like they have an "e" at the end.

Overall, the Talk Boards have been the place to actually engage with the public regarding the project and share our interest in the material. We have had marvelous

discussions with our contributors on topics like Edgeworth's interest in lace, her views of other writers, her opinions on deafness and disability and how those would be viewed today. Our contributors have found and continue to find allusions to people, places, works, and other things that will be incredibly helpful as we take the project to the next stage: correct the transcriptions, encode the letters, and add notes. We have a dedicated research assistant to help extract information from their posts and find a consistent way to credit contributions from the Talk boards, which is the next step in our plans."

Made with Material for MkDocs

Zooniverse Help

Beta testing

5.1 What is beta testing for?

Testing your project and making iterative changes based on feedback from others is vital to a successful crowdsourced transcription project. By soliciting and reviewing feedback on your project's Workflow, interface, and Subject Sets, you will ensure that volunteers from diverse backgrounds and levels of proficiency are able to effectively contribute to your project.

Beta testing is a required phase of all publicly launched and listed Zooniverse projects, in which a set of volunteers are asked to test your project Workflow(s) and provide feedback on their experience. This is a formal process facilitated by the Zooniverse team.

Prior to formal beta testing, you should also test your project with friends or family who are trying it for the first time—and watch them as they do it! See where they struggle, ask them for feedback, and review the data they submit.

As you receive feedback and implement changes, you will iterate through different versions of your project. This will help you produce a Workflow that balances both an intuitive and enjoyable volunteer experience and the best quality data.

5.1.1 The project review process

In order to launch your project publicly and be listed on the Zooniverse project page, you will need to go through project review, which includes an internal review by the Zooniverse team as well as beta review by Zooniverse volunteers.

Before you apply for review, you should create a beta testing Subject Set and upload it specifically for this purpose. For Transcription Task projects, we recommend creating a beta testing set with around 25 Subjects, depending on how much text each image contains. The aim is to include enough data to last for five to seven days of beta testing, but not so much data that Subjects don't get retired. That way, your beta test will produce enough completed classifications to be able to test your aggregation and data analysis methods. Your beta test set will need to be linked to an active Workflow.

In addition, you will need to ensure that your project page, including the About, Team and FAQ sections, contain enough information for your reviewers to gain a sense of your project. Additionally, your project must have a Tutorial, Field Guide and Help text, and your Talk boards must be set up. Apply for review through the 'Visibility' tab of the Project Builder. The most up-to-date requirements will be listed there.

Before we send your project to our volunteers for beta testing, your project will first go through the internal review process, in which a member of the Zooniverse team will review your project based on specific criteria. Your internal review feedback will be delivered via a private Talk board, which your internal reviewer will set up. Make sure you are subscribed to Talk email notifications, so you don't miss this notification and delay your review process.

The project will then be sent to Zooniverse volunteers who have signed up to be beta testers. They will submit beta classifications to your project as well as provide feedback through a standard Google form which is set up for you by the Zooniverse team and linked on your project via the Announcement Banner.

Please note that this process may take some time to complete. Based on past experience, the entire process can vary from eight-to-sixteen weeks from applying to beta to full launch.

5.1.2 When to apply for review

It is important to go through the review process early enough to have enough time to implement any needed changes to both your Subject Sets and your Workflows. You should spend time reviewing and editing your project with contacts, colleagues, or friends so that you can go into the review process with a fairly robust project.

Depending on the feedback you receive from your internal review and your beta testing, you may elect to go through another round of testing with beta testers after implementing changes, especially if those changes result in a significantly different Workflow requiring different data analysis or quality control methods. You may also decide to test out new Workflows after your project has already launched, though additional beta tests are not required to add new Workflows to a project after it has launched.

5.1.3 What feedback is good feedback?

Unfortunately, you will never be able to please everyone, and beta testing is no exception. Much of the feedback you receive will be invaluable, but some of it will be about things you cannot change.

For example, if you have asked volunteers to transcribe an 18th-century text, helpful feedback may be to include examples of obsolete abbreviations in the Field Guide. Feedback that is less helpful may be that the handwriting is difficult to read. While this is good to know, in that it will help you better prepare volunteers for engaging with difficult scripts, it is also not something that can be resolved through editing your project Workflow.

Be careful not to overfit your post-beta edits to a small number of people; make sure that you're considering the full scope of the feedback you receive, including those that are in contrast with one another. If one person says they aren't interested in your project, but 10 people say they can't wait to participate, don't waste time trying to appeal to one person.

5.2 Beta: step by step

If you want your project to launch publicly, you must go through the review process. A public launch means that 1) we will send out a launch newsletter to our full volunteer community; and 2) your project will be marked as an 'official' Zooniverse project and listed at zooniverse.org/projects. If you do not want to launch publicly, you do not have to apply for beta review. Just build a project, set it to public, and send out your URL to your audiences. Please note that choosing to run a private project (and therefore skipping beta review) may limit our ability to assist you in later project stages – beta testing is a great way to catch issues with your project early on, while they are still easy to change.

Read our overview of the project review process for additional details.

5.2.1 Beta checklist

When you are ready for beta review, go through the checklist found in the Visibility tab of the Project Builder. At this stage, your project should be ready for volunteers to use – don't submit something for beta testing that is unfinished, or that you wouldn't send to the public.

5.2.2 Internal review

When you initially apply for review, your project will first be reviewed by a member of the Zooniverse team. The team member will deliver their feedback using our standard internal review form. Once they have finished their review, the team member will write to you via a Talk post on a private (Team + Zooniverse) board they set up as part of the internal review process. **Make sure you are signed up to receive Talk emails for your project, or you may miss this step.** Once you have read the feedback, you can respond to the reviewer via the review form. Note that there is also a section for you to fill out, too.

5.2.3 Beta review

After you have passed the internal review, it is time to schedule the beta review. Every Tuesday we send projects to the Zooniverse beta participant community for review. The max we send out on any given Tuesday is 3 projects. Your internal reviewer will tell you the next available slot and give you the opportunity to delay if needed. Beta lasts for about 1 week, and you will receive a link to your feedback via your private Talk board. Once the beta test is over, it is your responsibility to review your feedback and make any necessary changes.

As you prepare for beta, be sure to look at an example of the Feedback Form we will send to our beta testers.

Made with Material for MkDocs

Zooniverse Help

Project launch

6.1 Prepare for launch

When you have finished implementing the changes from your beta test, you are ready to launch your project. As you head toward launch, you want to ensure that your project is prepped, you have an engagement strategy, and that you and your team are prepared for the human effort required. The first few days after launch are critical, as you will typically experience your highest number of volunteers in that time. However, as the initial rush levels out, you will also need a long-term engagement plan (see section 8) to build a lasting community of volunteers around your project.

To make sure that your project will run as smoothly as possible after launch, make sure that you have reviewed all of the content on your project and that you are happy with it. You should still expect issues to arise after launch, but careful pre-launch review will help ensure that you are prepared to fix anything that comes up later.

Before you apply for launch, make sure you've done the following:

- Carry out a final 'check' of your project, including:
 - Reviewing all volunteer resources (Tutorial, Field Guide, etc.)
 - Testing your workflow from start to finish
 - Clicking on any links you have included in your project text, to make sure they work
- Upload your full dataset
- Plan your own PR efforts via social media, listservs, etc.

- Make a launch schedule: how often will you check Talk in the first few weeks? In the first month?

6.1.1 Apply for launch

To apply for launch, you will post on the private Talk board that the Zooniverse team set up during project review. At this point, the Zooniverse team will also carry out a final quick check of your project to ensure it is launch ready.

6.1.2 Prepare your data

After the final review, you will upload the remaining Subject Set(s) you have for launch. Be sure to attach the appropriate Subject Set(s) to your Workflow(s). You can always upload additional Subject Set(s) later on, but you want to be sure you have plenty of data available at your initial launch so the early wave of volunteers does not run out of items to transcribe.

6.1.2.1 Visibility of metadata

If you want to include metadata with your Subject upload, you can do so via a Manifest. Instructions for how to upload data (with or without metadata) is available in section 2.

6.1.2.2 Retirement limits

When you first launch your project, keep an eye on your early Data Exports (or initial data in ALICE) to make sure that the quality of the output is meeting the expectations set via your beta testing results. If it isn't, you may need to make additional adjustments to your project Tutorial, Help text, or even your Workflow.

6.1.3 New project announcements

New projects are announced to the Zooniverse community via an email newsletter,

typically sent out on Tuesdays, which reaches hundreds of thousands of people. This provides a large surge of volunteers at the beginning of a project, known as the Launch Rush. Managing this initial surge may require a significant amount of time in the first days or weeks after launch. Not only will you need to be prepared to answer questions from volunteers, but you should expect some issues to surface and be prepared to fix them quickly.

When you apply for launch, you will submit copy about your project for the launch newsletter. You will submit this via the private Talk board the Zooniverse team set up during beta. You can also read our guide to writing an amazing newsletter for additional suggestions. In short, be short. Also, be personal, clearly describe the project and its goals, and include links to associated sites (social media, other web homes of your project, etc.). Example Newsletters

You may also choose to do a 'soft launch,' in which your project is set to 'Approved' and listed at zooniverse.org/projects without sending out a newsletter. This can give you time to work out any kinks in your project management infrastructure, while still leaving a public announcement as an open possibility for the future. If you are interested in a soft launch, be sure to mention this to the Zooniverse team member carrying out your internal project review.

6.1.4 Prepare your team

Before you launch, it is important to think and discuss how people's time and energies will be divided up both before and during the initial Launch Rush as well as in the long term.

If you are working on a team, take some time to discuss the division of duties: Who will be responsible for checking Talk boards? When will they be monitoring and responding to questions and comments? Who is in charge of keeping an eye on your stats and early results to spot any potential problems? After the Launch Rush, how will these roles change?

If you are working alone, try to plan out the time you will spend on monitoring your

project during and after launch. How will you balance this work with your other commitments? Is this amount of effort realistic and sustainable? If not, you may want to consider delaying your launch until you can find some additional team members to work on your project. For both large and small teams you may find it useful to schedule regular time during the week to check on Talk, engage with your volunteers, and follow up on any technical issues that have arisen.

6.2 Build a community

Building a healthy community of volunteers is vital to both an ethical and successful crowdsourced project. You will need a community of people interested in volunteering their time and energy to your project and you also need to maintain that community, interest, and trust over time. There are many different ways to build community around your project. By building a project on the Zooniverse platform, you are tapping into the millions of people who regularly use Zooniverse or receive the newsletter, which offers a significant amplification of your project in its first days. However, to build long-term engagement, you will want to craft a sustainable plan to reach new audiences and inspire returning volunteers (see section 9).

6.2.1 Develop a strategic engagement plan

Your first step will be to determine who may be interested in the content or methodology of your project. For example, does your project relate to specific interest groups, such as academic, political, or cultural organizations, local communities, or organizers? How might your project be a useful teaching tool that could appeal to teachers or students? Does your project appeal to a particular academic discipline or profession that may be tapped? Identify as many different potential audiences as you can think of.

Next, determine your team's capabilities for outreach to these groups, and create your outreach plan. This includes both technical capacity and labor bandwidth. Possible outreach strategies could include traditional media, social media, listservs

and newsletters, targeted outreach to interest groups, and live events. If you want teachers to incorporate your project into the classroom, you can design sample lesson plans. If you are located at an organization with a marketing or communications team, coordinate your plans with them. Many Zooniverse projects have received TV and newspaper coverage in local markets.

Case Study: Engaging the public

Deciphering Secrets: Unlocking the Manuscripts of Medieval Spain brings medieval Spanish manuscripts online, and provides volunteers with the opportunity to learn paleography before participating in transcription.

Roger Martinez, Project Director:

"Engaging the public in citizen science initiatives can be enhanced by identifying the natural affinity groups that may have personal, cultural, regional, thematic, and professional interests in your materials. Often we think about our materials in one dimensional terms where we are focused on a distinct perspective. For example, the Deciphering Secrets: Unlocking the Manuscripts of Medieval Spain project is specifically focused on uncovering cathedral and municipal manuscripts that pertain to Jewish, Christian, and Muslims relations. This niche theme can be marketed to a broader range of individuals and organizations.

First, U.S.-based genealogical societies that focus on LatinX populations are a prime candidate as the crowdsourced manuscripts offer the opportunity for individuals to find new clues about their surnames and places of origin. Similarly, cultural institutions that are dedicated to distinct Spanish religious populations, like the American Sephardi Federation, provide an opportunity for Sephardic Jews and their supporters to recover lost elements of their pre-1492 expulsion history. Regional and local historical groups, especially those intent on promoting cultural heritage knowledge, similarly offer the avenue of building cadres of crowd-sourcers tying their local American heritage to their ancestral origins.

Thematically, institutions and their members who are focused on issues like

interreligious and interfaith initiatives, international collaboration, and comparative studies, are also ideal to reach out to with invitations to participate.

Lastly, there are professional organizations that might have peculiar interests in your distinct materials, such as graphic artists and typographers who are seeking unique and intriguing ways to inject historical features into their works. Think broadly about your materials to maximize opportunities to find new types of collaborators."

6.3 Sustain your community

Finding a community of volunteers is one thing; sustaining a community is very different. This section will provide guidance on how to support your project community in the long term.

6.3.1 Talk

Your primary point of contact with volunteers will be your project's Talk boards. Talk is Zooniverse's message-board space. Each project has its own Talk boards that act as a forum for volunteers to interact with one another as well as the research team. Talk is often one of the first places volunteers will go if they have questions about either your project or its content.

It is important to dedicate time to regularly review questions and comments on Talk. During your initial launch rush, this may mean daily or even more frequent review. As volunteer participation levels out after its initial spike, you will get a sense for the rhythm of Talk and your volunteers' posting frequency. It is often best to designate specific team members and/or specific times for engagement with volunteers on Talk boards. Ideally, you should aim to respond within a few days, if not sooner. If you plan to take breaks for holidays, etc., use the Announcement Banner to let your volunteers know it may take longer than usual for you to respond.

6.3.1.1 Technical questions

Volunteers will often pose questions about Workflows, Tasks, and Subject Sets on Talk. Be prepared to answer these questions succinctly and in jargon-free language as much as possible. You should do your best to respond to technical questions before reaching out to the Zooniverse team for help (we're a small team, so it will take us longer to respond). If a lot of volunteers are struggling with technical issues, you might consider whether your Tutorial or Help text needs to be edited for clarity. Some issues may be solved by a volunteer updating their web browser, clearing their internet cache or cookies, or even just refreshing the page.

If you are unable to resolve the issue, please send an email to contact@zooniverse.org with the following information:

- The project ID number
- A brief explanation of the problem
- Links to any relevant Talk threads

6.3.1.2 Discoveries

Talk is often a place where volunteers will post about things they find interesting in the Subject Sets they are working with. These are important ways to build interest and long-term engagement with your project and can even inspire new directions for your research. Sharing discoveries from Talk on social media or project blogs (with permission of the original post author) is a simple way to highlight volunteer effort as well as share interesting project content with the broader public.

6.3.2 Project updates

Part of ethically engaging with your volunteers is sharing regular updates on the project progress, both in terms of how the transcription project is progressing and how the larger project is using the data produced by those same volunteers. However, you should think carefully about when and how you share project updates – the last thing you want is for volunteers to get discouraged by what seems like a lack of progress.

Project updates can be shared on Zooniverse via Talk and through project-specific newsletters (which are distinct from the Zooniverse-wide newsletter that announces new projects). These are emails that you can send to your registered volunteers after your project has launched. Newsletters should share your project progress and will often bring people back to your project. To send out a newsletter, email plain-text copy to contact@zooniverse.org and we will send it out for you.

You can also share your project updates outside of the Zooniverse platform, through social media, a project website, listserv, etc.

6.4 Acknowledging volunteers

Many projects choose to acknowledge volunteer work in their publications and other research outputs. Volunteers generally appreciate and have increasingly come to expect formal acknowledgement of their effort in crowdsourced science. There is no one way to effectively acknowledge volunteers and their contributions, but the best place to start figuring out an appropriate acknowledgement of their efforts is to ask them. You can begin this conversation by posting a Talk thread which seeks advice about how volunteers would like to be acknowledged. Since some volunteers may not have considered this question before, you might start the conversation by providing examples of how other researchers have recognized volunteer labor.

For instance, the team of researchers behind the *Bash the Bug* project acknowledged its volunteer community generally in the acknowledgements section of its most recent publication while the *Exoplanet Explorers* project listed every volunteer contributor by either their username or real name (if the user opted in to share it) under the Team section of their Zooniverse project. There are even projects like *Radio Galaxy Zoo* in which volunteers have made such significant contributions they were invited to become co-authors on peer-reviewed publications in scientific journals.

As the European Commission's Parthenos Project points out, such formal acknowledgements in scientific publications are important, but may not be the most important form of acknowledgement to the volunteers themselves. Creative ways of

recognizing volunteer participation through such means as in-person or virtual events and portable badges might mean more to a given volunteer community. Your volunteer community and individual volunteers will always be the place resource to guide you on how to acknowledge their contributions in a way that's meaningful to them.

Made with Material for MkDocs

Zooniverse Help

Working with data

7.1 Designing for data output

The primary objective of your Zooniverse project involves transforming an existing dataset (images of handwriting) into another form of dataset (lines and perhaps columns of text) that can be analyzed more efficiently to achieve the broader goals of your project. This transcription data may take on various forms depending on your plans for use. In this section, we'll consider how your data can be processed, stored, analyzed, and shared during the course of your Zooniverse project and after it is completed.

Before creating your project, you should know what kind of data you want to collect. Your final data set will comprise a consensus for each Task in your Workflow for a given Subject. Keeping this in mind allows you to think of an ideal architecture for building your project and the provided tools to onboard your historical information to a usable digital dataset.

7.2 Data exports

7.2.1 How to export data from the Project Builder

You can request a data export via the Data Exports tab of the Project Builder. In this space, there are several types of data you can request.

Export Type	Contents	Format
Classification	all task data from all workflows for your classification	JSON in

export	entire project	CSV
Workflow classification export	all task data from a specific workflow	JSON in CSV
Workflow export	structural information about how your workflow looks, with version info	JSON in CSV
Subject export	a list of all subjects you have uploaded to the project, with any metadata included	JSON in CSV
Talk comment export	all comment data from Talk	JSON
Talk hashtag export	all hashtag data from Talk	JSON

7.2.2 When to export data from the Project Builder

You can request a data export at any point, even if you have only submitted test transcriptions to your project. If you have not submitted any test transcriptions on your workflow(s), your classification data export will be empty. The Data Exports tab will also show when a Data Export was last requested for each export type.

7.2.3 How to export data from ALICE

For information on exporting data from ALICE, please refer to the ALICE documentation. Note that only 'approved' transcriptions can be exported from ALICE.

7.2.4 Data export examples

classification_id	user_name	user_id	user_ip	workflow_id	workflow_name	workflow_version	created_at	gold_standard	expert	metadata	annotations	subject_data	subject_ids
487254189	bwiggins	1598199	68c814866599a09448e8	23893	Workflow 3	95.110	2023-05-02 21:29:16 UTC			{"source": "api", "session": "0a5420483e8ed0", [{"task": "T50", "value": "", "task_label": "Date"}, {"task": "T50", "value": "1897", "task_label": "Year"}], {"id": 114694939, "filename": "79958562", "retired": null}}	{"source": "api", "session": "79958562", "id": 114694939, "filename": "79958562", "retired": null}	79958562	
488696460	evanrobertsz	321294	13a458e2ef8787268521	23893	Workflow 3	95.110	2023-05-09 18:08:16 UTC			{"source": "api", "session": "dff491f8df60a0ad6", [{"task": "T50", "value": "April 14 1897", "task_label": "Date"}, {"task": "T50", "value": "1897", "task_label": "Year"}], {"id": 114694831, "filename": "79958562", "retired": null}}	{"source": "api", "session": "79958562", "id": 114694831, "filename": "79958562", "retired": null}	79958562	
488701729	evanrobertsz	321294	13a458e2ef8787268521	23893	Workflow 3	95.110	2023-05-09 18:22:41 UTC			{"source": "api", "session": "dff491f8df60a0ad6", [{"task": "T50", "value": "December 28 1898", "task_label": "Date"}, {"task": "T50", "value": "1898", "task_label": "Year"}], {"id": 114694939, "filename": "79959029", "retired": null}}	{"source": "api", "session": "79959029", "id": 114694939, "filename": "79959029", "retired": null}	79959029	

Above: the header structure of a Project Builder workflow classification export. All examples below demonstrate the task-specific structure of data found in the annotations column, unless otherwise noted.

This section will cover task types that are commonly used in Text Transcription projects. For information on other task types, please see the main 'How to create a project' documentation.

7.2.4.1 Question task (Project Builder export) - single response

This is the data export format for a question task where only a single response has been submitted, from the project *Corresponding with Quakers*:

```
{"task": "T5", "task_label": "What is the letter about? Choose a few keywords.", "value": ["health/illness/death"]}
```

In the above case, a volunteer has selected the option "health/illness/death" for the question What is the document about? Choose a few keywords.

TASK	TUTORIAL
What is the document about? Choose a few keywords.	
<input type="checkbox"/> health/illness/death	
<input type="checkbox"/> childbirth/pregnancy/breastfeeding	
<input type="checkbox"/> Quakerism/religion	
<input type="checkbox"/> business/finance/household accounts	
<input type="checkbox"/> travel	
<input type="checkbox"/> education	
<input type="checkbox"/> slavery/abolition/colonialism	
<input type="checkbox"/> gender/women/female preaching	
<input type="checkbox"/> American Revolution/French Revolution	
<input type="checkbox"/> politics	
<input type="checkbox"/> books/reading/authorship	
<input type="checkbox"/> letter writing/the post	
<input type="checkbox"/> family and friends	
<input type="checkbox"/> other	

NEED SOME HELP WITH THIS TASK?

7.2.4.2 Question task (Project Builder export) - multiple responses

This is the data export format for a question task where multiple choices are allowed, from the project *Corresponding with Quakers*:

```
{"task": "T5", "task_label": "What is the letter about? Choose a few keywords.", "value": ["letter writing/the post", "family and friends"]}
```

In the above case, a volunteer has selected the options "letter writing/the post" and "family and friends" for the question What is the document about? Choose a few keywords.

This is what the task input area looks like for this specific example (i.e. what the volunteer sees when entering data):

TASK	TUTORIAL
What is the document about? Choose a few keywords.	
<input type="checkbox"/> health/illness/death	
<input type="checkbox"/> childbirth/pregnancy/breastfeeding	
<input type="checkbox"/> Quakerism/religion	
<input type="checkbox"/> business/finance/household accounts	
<input type="checkbox"/> travel	
<input type="checkbox"/> education	
<input type="checkbox"/> slavery/abolition/colonialism	
<input type="checkbox"/> gender/women/female preaching	
<input type="checkbox"/> American Revolution/French Revolution	
<input type="checkbox"/> politics	
<input type="checkbox"/> books/reading/authorship	
<input type="checkbox"/> letter writing/the post	
<input type="checkbox"/> family and friends	
<input type="checkbox"/> other	

NEED SOME HELP WITH THIS TASK?

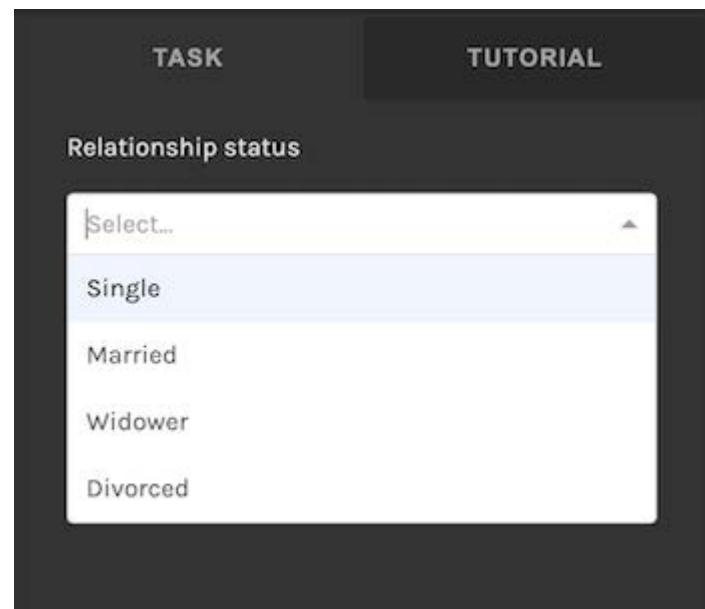
7.2.4.3 Dropdown menu - single (Project Builder export)

This is the data export format for a single dropdown task, from the project *Shadows on Stone: Identifying Sing Sing's Incarcerated*:

```
{"task": "T30", "value": [ {"select_label": "Relationship status", "option": true, "value": "6e68e5fc7071a", "label": "Married"} ]}
```

In the above case, a volunteer has entered "Married" into the "Relationship status" field via selecting an entry from the options provided in the dropdown list.

This is what the task input area looks like for this specific example (i.e. what the volunteer sees when entering data):



7.2.4.4 Dropdown menu - multiple (Project Builder export)

This is the data export format for three dropdown tasks, from the project *Shadows on Stone: Identifying Sing Sing's Incarcerated*, presented together in the same workflow step:

```
{"task": "T52", "task_label": null, "value": [ {"task": "T15", "value": [ {"select_label": "Sentenced (month)", "option": true, "value": 3, "label": "3 - March"} ]}, {"task": "T16", "value": [ {"select_label": "Sentenced (day)", "option": true, "value": 7, "label": "7"} ]}, {"task": "T17", "value": [ {"select_label": "Sentenced (year)", "option": true, "value": 1898, "label": "1898"} ]} ]}
```

This is what this data looks like when parsed:

```
[{"task": "T52", "task_label": null, "value": [{"task": "T15", "value": [{"select_label": "Sentenced (month)", "option": true, "value": 3, "label": "3 - March"}]}, {"task": "T16", "value": [{"select_label": "Sentenced (day)", "option": true, "value": 7, "label": "7"}]}, {"task": "T17", "value": [{"select_label": "Sentenced (year)", "option": true, "value": 1898, "label": "1898"}]}]}
```

In the above case, a volunteer has entered "March" for Sentenced (month), "7" for Sentenced (day), and "1898" for Sentenced (year), meaning that, according to this record, this person was sentenced on March 7, 1898.

This is what the task input area looks like for this specific example (i.e. what the volunteer sees when entering data):

The screenshot shows a dark-themed Zooniverse task interface. At the top, there are two tabs: 'TASK' (selected) and 'TUTORIAL'. Below the tabs, there are three dropdown menus labeled 'Sentenced (month)', 'Sentenced (day)', and 'Sentenced (year)', each with a 'Select...' placeholder. Between the first two dropdowns is a link 'NEED SOME HELP WITH THIS TASK?'. Between the second and third dropdowns is another link 'NEED SOME HELP WITH THIS TASK?'. At the bottom left is a 'Back' button, at the bottom center is a 'Next →' button (which is highlighted with a yellow border), and at the bottom right is a gear icon.

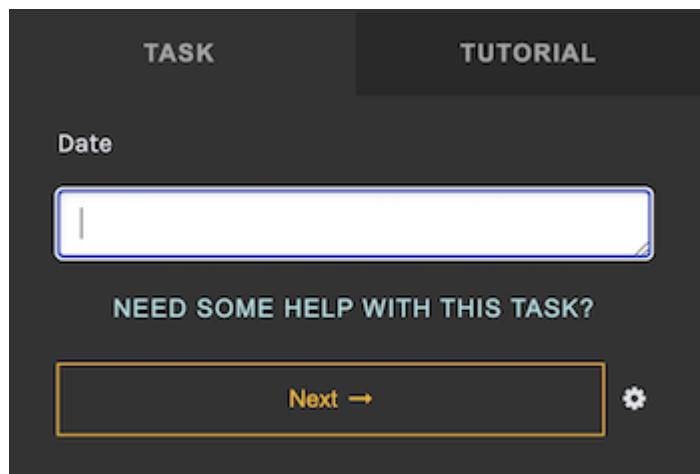
7.2.4.5 Free-text entry (Project Builder export)

This is the data export format for a free-text entry task, from the project *Shadows on Stone: Identifying Sing Sing's Incarcerated*:

```
[ { "task": "T50", "value": "March 24, 1899", "task_label": "Date" } ]
```

In the above case, a volunteer has entered "March 24, 1899" into the `Date` field provided.

This is what the task input area looks like for this specific example (i.e. what the volunteer sees when entering data):



7.2.4.6 Transcription Task (Project Builder export)

This is the Project Builder data export format for the Transcription Task, from the project *Corresponding with Quakers*:

```
[{"task": "T0", "value": [{"x1": 1460.4500246994899, "x2": 805.3800428124487, "y1": 96.11065371315667, "y2": 104.96295076568421, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1450.3331137823154, "x2": 1350.4286184752184, "y1": 166.92903013337732, "y2": 154.2828914869094, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 799.0569734892148, "x2": 135.13469454964604, "y1": 170.7228717273177, "y2": 174.5167133212581, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1474.3607772106047, "x2": 615.6879631154292, "y1": 282.00889181623586, "y2": 285.8027334101763, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1441.4808167297879, "x2": 79.49168450518695, "y1": 374.3257039354521, "y2": 386.9718425819201, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1475.6253910752514, "x2": 69.37477358801246, "y1": 461.58406059608114, "y2": 466.64251605466836, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1427.5700642186732, "x2": 63.05170426477869, "y1": 553.9008727152974, "y2": 566.5470113617654, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1384.573192820682, "x2": 71.90400131730621, "y1": 646.2176848345135, "y2": 655.0699818870411, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1485.7423019924258, "x2": 61.78709040013183, "y1": 756.2390910587849, "y2": 73.8267742466657, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}, {"x1": 1441.4808167297879, "x2": 57.99324880619139, "y1": 870.0543388769968, "y2": 877.6420220648777, "frame": 0, "details": [{"task": "T0.0.0"}], "toolType": "transcriptionLine", "toolIndex": 0}]
```

ndex":0}, {"x1":1473.0961633459578,"x2":171.80849662440326,"y1":957.3126955376256,"y2":969.9588341840936,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1460.4500246994899,"x2":168.01465503046276,"y1":1068.598715626544,"y2":1073.657171085131,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1417.4531533014988,"x2":56.72863494154467,"y1":1154.5924584225259,"y2":1178.6201218508152,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1487.0069158570725,"x2":61.7870904001318,"y1":1253.2323398649762,"y2":1282.3184587518526,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1476.8900049398983,"x2":83.28552609912745,"y1":1358.1952906306606,"y2":1383.4875679235963,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1490.800757451013,"x2":54.19940721225106,"y1":1459.3643998024043,"y2":1490.9797464185742,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1437.6869751358474,"x2":65.58093199407224,"y1":1541.564301004446,"y2":1578.238103079203,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1476.8900049398983,"x2":49.14095175366387,"y1":1626.2934299357817,"y2":1666.7610736044794,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1417.4531533014988,"x2":45.347110159723485,"y1":1702.1702618145891,"y2":1759.077885723695,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}, {"x1":1452.862341511609,"x2":56.72863494154453,"y1":1794.4870739338055,"y2":1838.7485591964435,"frame":0,"details":[{"task":"T0.0.0"}],"toolType":"transcriptionLine","toolIndex":0}], "taskType": "transcription"}, {"task": "T0.0.0", "value": "Mountmelick 4th of 3rd mo", "taskType": "text", "markIndex": 0}, {"task": "T0.0.0", "value": "1811", "taskType": "text", "markIndex": 0}, {"task": "T0.0.0", "value": "My dear Aunt Sally,", "taskType": "text", "markIndex": 1}, {"task": "T0.0.0", "value": "Thou sent to know if I would", "taskType": "text", "markIndex": 2}, {"task": "T0.0.0", "value": "chuse thee to write to me? certainly, \u00a0\u00a0\u00a0 for two", "taskType": "text", "markIndex": 4}, {"task": "T0.0.0", "value": "reasons thy letters would be most welcome, the", "taskType": "text", "markIndex": 5}, {"task": "T0.0.0", "value": "one, because they come from thee, \u00a0\u00a0\u00a0 the other", "taskType": "text", "markIndex": 6}, {"task": "T0.0.0", "value": "from Ballitore. I hear thou art with", "taskType": "text", "markIndex": 7}, {"task": "T0.0.0", "value": "Mary Bostello, it is a bad time for me to

write," , "taskType": "text", "markIndex": 8}, {"task": "T0.0.0", "value": "for my letter cannot recompense for taking", "taskType": "text", "markIndex": 9}, {"task": "T0.0.0", "value": "from such elegant \u0026 agreeable company", "taskType": "text", "markIndex": 10}, {"task": "T0.0.0", "value": "but thou has always very good company", "taskType": "text", "markIndex": 11}, {"task": "T0.0.0", "value": "but thou has not always Mary Costello, I", "taskType": "text", "markIndex": 12}, {"task": "T0.0.0", "value": "would be obliged to thee to remember me to her.", "taskType": "text", "markIndex": 13}, {"task": "T0.0.0", "value": "I suppose thou often sees my dear Lydia, alas", "taskType": "text", "markIndex": 14}, {"task": "T0.0.0", "value": "Beaty's foreboding fears were but too well founded", "taskType": "text", "markIndex": 15}, {"task": "T0.0.0", "value": "how are thy eyes \u0026 head? I hope thy are, at", "taskType": "text", "markIndex": 16}, {"task": "T0.0.0", "value": "least, no worse. Governess had not time to look", "taskType": "text", "markIndex": 17}, {"task": "T0.0.0", "value": "for the verse thou sent for, but I intend to", "taskType": "text", "markIndex": 18}, {"task": "T0.0.0", "value": "send it the next time. I am sorry I have", "taskType": "text", "markIndex": 19}, {"task": "T1", "task_label": "Have all the volunteer-made underline marks on all pages turned grey?", "value": "No"}

This is an example of what the first two lines of positional data look like when parsed:

```
[{"task": "T0", "value": [{"x1": 1460.4500246994899, "x2": 805.3800428124487, "y1": 96.11065371315667, "y2": 104.96295076568421, "frame": 0, "details": [{"task": "T0.0.0"}]}, {"x1": 1450.3331137823154, "x2": 1350.4286184752184, "y1": 166.92903013337732, "y2": 154.2828914869094, "frame": 0, "details": [{"task": "T0.0.0"}]}]}
```

This is an example of what the first four lines of text data look like when parsed:

```
[{"task": "T0.0.0", "value": "Mountmelick 4th of 3rd mo", "taskType": "text", "markIndex": 0}, {"task": "T0.0.0", "value": "1811", "taskType": "text", "markIndex": 1}, {"task": "T0.0.0", "value": "My dear Aunt Sally,", "taskType": "text", "markIndex": 2}, {"task": "T0.0.0", "value": "Thou sent to know if I would", "taskType": "text", "markIndex": 3}]
```

7.2.4.7 Transcription Task (ALICE Export)

This is the most basic of the ALICE data exports format (plain .txt) for the Transcription Task, from the project *Corresponding with Quakers* (first four lines of text only):

Mountmelick 4th of 3rd mo

1811

My dear Aunt Sally,

Thou sent to know if I would

This export version is only available after a subject is marked as Approved in ALICE.

For more information on how to use ALICE, read the documentation.

7.3 Data aggregation

Because Zooniverse projects collect multiple transcriptions per Subject, you will need to have a plan for how to identify a single transcription that best represents the material you are transcribing. This could be through manual review, a 'best of' determination, scripted data aggregation methods, or other.

In this section we describe existing methods for aggregating data from text transcription projects.

7.3.1 Text aggregation with ALICE

The following images demonstrate text transcription data processing using ALICE. The first shows the aggregate line view for a given page, while the second shows the view for reviewing and editing an individual line of text.

The screenshot shows the ALICE transcription interface. At the top, it displays 'FOLDER_40' and '79949303'. On the right, there are buttons for 'ALL CHANGES SAVED', 'MARK AS APPROVED', 'UNDO', 'LAYOUT', 'MORE', and user info ('tinsi037 Admin'). Below this is a 'Transcribed Text' table with columns for 'FLAG' and 'CONSENSUS SCORE'.

	FLAG	CONSENSUS SCORE
AGGREGATED TRANSCRIPTION		3/3
Philadelphia		3/3
My very Dear Dicky		45/5
After Acknowledging yr. many Favours and		3/3
Obligations I lie under too thee my Deserving		3/3
Friend. I acquaint thee of my safe arrival		3/3
after a passage of Ten Weeks, which was		3/3
Tedious and Disagreeable, narrowly Escaping		3/3
being Taken by a French Man of War		3/3
Who chased us for Several hours, & was		3/3
thought by some we had our wages for our		3/3
is Stalling but that I leave. Dear Friend		3/3
I know not what to write having but little		3/3
yet could not forbear letting thee know where		3/3
I am Gratitude obliging me therfore, and		3/3
may shortly let thee know we have very troublesome		3/3
times in this Land such as has not been known		3/3
here yet I do believe there is a Remnant yt. will		3/3
stand thier Ground and bear that Antient and		3/3
noble Testimony yt. our forefathers Did tho.		3/3
[unclear]/[unclear] 4. mo 1755 -		3/3

On the left, the 'Original Subject' is shown as a thumbnail of the handwritten letter. Below it are buttons for 'All Pages' and 'DELETE PAGES'.

Above: Letter from John Peters, Dec. 4, 1755. University of California Santa Barbara MSS 4, Box 12, folder 40, 001. A screenshot from the ALICE interface for

Corresponding with Quakers, Principal Investigator and Project Director Rachael Scarborough King.

The screenshot shows the ALICE transcription interface. On the left is a thumbnail of the original document page, which contains a handwritten letter from John Peters dated December 4, 1755. The letter is written in cursive and discusses various topics including his travels and experiences. To the right of the image is a table titled "Transcribed Text". The table has two columns: "AGGREGATED TRANSCRIPTION" and "SELECTED TRANSCRIPTION". The "AGGREGATED TRANSCRIPTION" column lists several lines of text with checkboxes next to them. The "SELECTED TRANSCRIPTION" column shows the line "After Acknowledging yr. many Favours and October 16, 2022 10:02 Vimy" selected, indicated by a red square checkbox. Other lines include "After Acknowledging yr. many Favours and October 24, 2022 03:28 Hippogonizer", "After Acknowledging yr. many Favours and November 7, 2022 04:01 JFK", and "After Acknowledging yr. many Favours and November 12, 2022 18:02 blackpapermon". The table also includes columns for "FLAG" and "CONSENSUS SCORE" (with a score of 4.5/5). At the bottom of the table, there are buttons for "ADD LINE BELOW", "DELETE LINE", "REPLACE WITH SELECTED", and "CLOSE".

Above: Letter from John Peters, Dec. 4, 1755. University of California Santa Barbara MSS 4, Box 12, folder 40, 001. A screenshot from the ALICE interface for Corresponding with Quakers, Principal Investigator and Project Director Rachael Scarborough King.

The Aggregated Line Inspector and Collaborative Editor (ALICE) is the main tool for researchers seeking to automate aggregation of multiple volunteer transcriptions into a single response. To use ALICE, you will need to be using the Transcription Task in your project to collect volunteer transcriptions. Please read through the comprehensive documentation about ALICE if you plan to use this tool, as there are restrictions on how to set up and build your project which require intervention from the Zooniverse team.

7.3.2 Other Task types

Zooniverse has a robust framework for aggregating data generated via other Task types (e.g. Question Tasks, Dropdown Tasks, etc.). The in-house code supports data

aggregation for nearly all the Project Builder Task types, by providing tools to process and aggregate classification data exported from Zooniverse projects.

A general overview of this process would be: configuring Panoptes to your exported data, extracting the pertinent data from a classification export, and then reducing data for a given Task based on the consensus from each Task's classification set.

For an example of Python data aggregation for a Question Task, see this notebook.

7.4 Data sharing requirement

Projects promoted to the Zooniverse community have the goal of producing useful research; your study needs to be well designed, and you must intend to analyze and write up your results as a formal publication. You must make volunteer classification data open within two years of a project's completion. You also commit to communicating research findings to your volunteer community. Finally, you must formally acknowledge Zooniverse in any publications.

Please report publications using Zooniverse-produced data to us via this form.

7.5 Unexpected data

One of the benefits of a crowdsourced transcription project is that you are engaging with volunteers who come from a wide variety of backgrounds and fields. While you have a specific end-goal of using your classifications, there are many other, often unexpected outputs of a crowdsourced transcription project.

Talk boards provide one site where volunteers can share data or findings with one another and the research team. For example, the letters of Poets and Lovers often referenced specific paintings or pieces of art. Volunteers on the project took it upon themselves to look up digitized versions of the artworks mentioned and posted them on Talk. This not only provided a fantastic resource for volunteers who wanted to learn more about the two women authors of the diaries, but also the research team

will try to incorporate these links into the published edition of the diaries. Talk boards can be exported from the 'Data Exports' tab of the Project Builder as a JSON for easier data analysis.

Case study: Poets and Lovers

Poets and Lovers is a project to transcribe the diary of Michael Field which was created by two women novelists Katharine Harris Bradley and Edith Emma Cooper.

Peter Logan, Principal Investigator:

The questions aren't hard and if I can't answer them, I tell one of my partners to get involved, and they'll know the answer. And some of the questions have been very interesting too. Some of them brought up things we didn't know about. So that was an unexpected outcome. We did actually learn things about what's in these diaries that we didn't know about. [The authors] talk a lot about paintings in the diaries, and I was going to write up [the volunteers] responses to the paintings. One of the volunteers found links to online images of the paintings, so we can incorporate them into the transcription. Which is exactly what we're going to do.

Project outputs are not limited to data. By virtue of the nature of crowdsourced transcription projects, research and data are opened up to more people. Many project teams have seen the engagement with the public as a vital output of their work, and have structured their research and project around that engagement. For example, the Mapping Prejudice team primarily worked with volunteers directly through meetings. They saw the work of educating people on the history of racial covenants through these in-person or small virtual events as an integral practice for their community work. Here, the *process* of crowdsourced transcription was as important of an output as the data.

7.6 Data curation

Depending on your affiliations, aspirations, and your project's goals for presenting

data, there may be several ways to preserve and share your project's results. The most straightforward and cost-effective way for you as a researcher would be to utilize an affiliate institution's data repository (ones with the CoreTrustSeal are best). These institutional repositories allow for consistent access and long-term presentation of your data while allowing you to organize your data in a clear and detailed manner. If you are not a part of an institution, good options for preserving and publishing your data are the Open Science Framework and Dryad. In the social sciences, a widely used data repository is Open ICSPR.

Examples of successful data publications from Zooniverse projects can be found here:
<https://www.zooniverse.org/about/publications#data>.

Made with Material for MkDocs

Zooniverse Help

Project maintenance and conclusion

8.1 Long-term projects

After your project is set up and has launched, you will need to determine how to best maintain your project and support your volunteers over the lifetime of the project, as well as best practices for sunsetting or 'ending' your project. By asking volunteers to donate their time and effort to your project, you are committing to the effort of project maintenance and data sharing.

Some projects have a finite set of data that can be transcribed using crowdsourcing over a relatively brief period of weeks or months. However, if your project perpetuates for months or years using new data or even new Tasks and Workflows, you'll need to consider how to sustain its community of volunteers and perhaps also financial support for administrative, technological, and intellectual labor.

Work with students (or other enthusiastic volunteers) when possible. I set up an internship for a group of students who helped me with the crowdsourcing project (and other documentary editing work). Their enthusiasm for the subject material was refreshing and enlivened my interest for what I already knew was a great project, but appreciated an injection of excitement into what I was doing. The students helped me select the best documents to work on for crowdsourcing. And while they were only around for a short time, I got a lot out of explaining crowdsourcing to them and thinking through the documents. - *Serenity Sutherland, BCCCT Cohort Member*

8.2 Maintaining volunteer communities

Communication (see Section 6) with current and potential volunteers is the

foundation of long-term project maintenance. Regularly updating your volunteer community about the progress and current needs of your project is critical for continued engagement. Routine newsletters, social media posts, and Talk updates about what's new and ongoing with the project helps volunteers to continue their transcription efforts. Moreover, press releases, media interviews, and other forms of broad outreach can help to engage new audiences particularly when it showcases progress made from initial volunteer contributions.

The usual output of Zooniverse projects is a peer-reviewed article, conference presentation or book. But there are many other possible research products and early results of your work can and should be communicated to your volunteers. Keep your volunteer community engaged by communicating what you, your team, or even volunteers have found that offers insight into your documents and the history they tell.

Project updates can be shared on Zooniverse via Talk and through project-specific newsletters (which are distinct from the Zooniverse-wide newsletter that announces new projects). These are emails that you can send to your registered volunteers after your project has launched. Newsletters should share your project's progress and will often bring people back to your project. To send out a newsletter, email a plain-text copy to contact@zooniverse.org and we will send it out for you.

8.3 Financial support for projects

Anyone is welcome to create and promote a project using the Zooniverse platform, but historically most of Zooniverse's project creators are professional researchers who come from a university, cultural institution, or governmental organization. For those in such roles hoping to maintain a project for months or years, it can be crucial to secure long-term support for a project from their organization. Such support is helpful before a project launches.

8.3.1 Institutional support

There is a great deal of labor involved in building a Zooniverse project; monetary or in-kind support from your employer from the beginning helps move projects from ideas to launch. Since the Project Builder allows teams and individuals to prototype projects with minimal technical assistance, it is also possible to get a project up and running in order to show your institution a functional proof of concept. Consider discussing your vision for the project with your institution's office of public engagement, marketing or communications department, or executive committee early on and keep them updated as your project progresses. The long-term funding needs of each project will vary significantly, but common contributions institutions make come in the form of assistance with external communications and design elements, administrative staff or research assistants, data processing and analysis staff and infrastructure, and dedicated workspace or even dedicated staff. Transcription projects tend to need more data processing and analysis support than other forms of crowdsourcing science and often seek help from information technology units in their institutions.

8.3.2 Grant and gift funding

While institutional buy-in always helps, it is rare for institutions to fund all the needs of long-term crowdsourcing transcription projects. Many researchers are expected to bring in external grants from governments or gifts from donors and foundations. Such external monies can accelerate the timeline of your project or sustain your project for a longer period than you or your organization could alone. The sources of grant funding vary by country, but an excellent primer for understanding the grant application process in fields where transcription projects tend to originate is Raphael Brewster Folsom's *How to Get Grant Money in the Humanities and Social Sciences*. Grant application cycles can take over a year, so begin thinking about your project's needs early and sketch out the aspects of the project that need the most labor or infrastructure in order for the project to succeed. Depending on your technical needs, you may want to consider reaching out to our team to collaborate on a grant application.

Obtaining gift funding is much more idiosyncratic than applying for grants, but the best place to begin is by speaking with someone in your organization who works in a 'development' or 'foundation' unit and guides relationships with existing and potential donors. You might be asked to draft a pitch of your needs and present that to them or even directly to donors. As with institutional support each project's needs differ, but to help your project thrive in the long term, begin generating your applications and pitches by outlining your overall objectives and working backward to assess the costs associated with activities you'll need to undertake to meet your goals for the project.

8.4 Sharing data and results

The transcriptions that your volunteers will produce are a significant output that must be made accessible to your volunteers and the public, generally. How teams share their transcription data varies, but you should strive to make it clear to anyone accessing your data what you did to generate it, how they can utilize it, and what you have done or plan to do to analyze it. There are many places available to post and preserve your data. If you are a member of a research institution, your library or archive likely hosts a data repository in which you can deposit your data and even your analysis of the transcriptions. There are alternative options that allow unaffiliated researchers to deposit their data, such as Dryad, and disciplinary-specific repositories such as OpenICPSR (social science) that enhance the likelihood that members of a specific field will engage with your data.

8.4.1 Recording your methods

Regardless of the repository you choose, you should work to provide strong contextual documentation about your project and the methods used to generate your transcriptions and find consensus among them. The best means for documenting the processes you utilized to create your project is to write a readme file. Readme files are text documents that accompany data and guide users through the who, what, when, where, why, and how the data was generated. Cornell University has put

together an indispensable reference for drafting quality readme files that you should consult when you're ready to write your own.

8.4.2 Packaging your data

In reductive terms, your project will generate some .json files. You could post those files in an open Google Drive and hope that interested parties now and well into the future know what they are and how to utilize and interpret them. But a much more effective means of data sharing is to package your data in a fashion that makes it findable, accessible, interoperable, and reusable (FAIR). This requires you to decide what to include in your data set (e.g., images, transcriptions, consensus transcriptions, etc.), document your dataset (in a readme, as in the step above), outline the reuse rights of your data (usually by selecting a Creative Commons license), and transform all files into non-proprietary formats (generic .txt files instead of Microsoft's .doc file).

The University of Reading has produced a detailed guide to preparing your data for a repository, and the guide produced by ICPSR is also useful. As you consider which repository to store your data, prioritize those known to have clear preservation plans that will keep files in working order and publicly accessible for decades or longer. Accrediting entities like the CoreTrustSeal can help you to evaluate repositories you're considering for your data set.

Additionally, you may want to consider publishing your dataset in a journal, such as the *Journal of Open Humanities Data*. The JOHD also provides a list of recommended repositories for their publication requirements.

8.4.3 Sharing your insights

Data repositories help to share and preserve your data for generations, but they act as places to preserve your evidence rather than your interpretations. The venues for making arguments from your transcriptions are countless and include books, articles, documentaries, exhibitions, courses, websites, apps, and many other media.

Moreover, you can also amplify your interpretive products through traditional and social media channels. Regardless of where you release your interpretations, it's critical to both acknowledge your volunteers' labor and to make your volunteers aware of how their labor is leading to your findings. Sharing your interpretive progress with your volunteers during and even after your project is an important part of giving back to your volunteer community and can be accomplished through updates to Talk, newsletters, and other project news sources like social media or a project website.

8.5 Wrapping up your Zooniverse project

Though Zooniverse projects can take years and your analysis of their results can take even longer, there often comes a point when your project will need to conclude. This section will give you insights into this process, so that you know what to expect.

8.5.1 Sunsetting your project site

Active Zooniverse projects are listed at <https://www.zooniverse.org/projects>. When your project runs out of data, it will be automatically moved to the "Paused" category. When you are sure your project is finished, you must reach out to the Zooniverse team and ask us to move it to the "Finished" category.

While projects can be deleted, it is our goal for Zooniverse-approved projects to keep an active, public site available as long-term reference to the past effort. The idea is to provide a record of the project, its Talk message boards, and a long-term link that can point to the project's outputs that are available and accessible to volunteer contributors. You may want to consider leaving one or more of your Workflows set to 'Active' to allow future visitors to experience the method(s) you used to collect your project data.

When you feel your project has been entirely completed, thank your volunteers via an email newsletter and explain what happens next. Update your Results page

accordingly, and don't forget to submit project publications to our Publications page. We expect researchers to use the results of the project in peer-reviewed research, and to share results with the community. As detailed in the step above, Classification data must be made accessible within 2 years of the project's completion, though sooner is preferable.

8.5.2 Preserving access to your activities

While your Zooniverse project site can stay 'live' indefinitely as a finished project, crowdsourcing projects and researchers' attendant analyses often produce other research outputs—from books to websites to apps to exhibitions and beyond—that should be thoughtfully organized and made accessible to volunteers and the public, generally. The Endings Project, a digital archiving effort from the University of Victoria suggests that researchers with projects that include significant digital elements should consider three questions as their project concludes:

1. How should the project conclude?
2. How should the project's dynamic features be preserved?
3. Where should the project be archived?

Each project must arrive at their own answers to those questions, but The Endings Project outlines a set of compliance principles to guide researchers through considerations that will help the digital elements of their work have a high likelihood of remaining accessible and even functional for decades. As you reflect on what you've accomplished through your Zooniverse project, imagine what you'd like someone to know about your process and results fifty years from now and work toward making those portions of your project sustainable to preserve and legible well into the future.

Made with Material for MkDocs

Zooniverse Help

Acknowledgements and resources

9.1 Additional project team

Professor Lucy Fortson, University of Minnesota

Dr. Cliff Johnson, Zooniverse Co-Director & Science Lead, Adler Planetarium

Dr. Laura Trouille, Vice President of Science Engagement, Adler Planetarium

Linda Greve, Program Director for Community Outreach and Grants Coordinator, University of Minnesota Libraries

Tawny Lane, Government Grant Manager, Adler Planetarium

Shannon Farrell, Data Curation Network

Wanda Marsolek, Data Curation Network

9.2 Cohort members and projects

Corresponding with Quakers

- Rachael Scarborough King (University of California Santa Barbara)

Deciphering Secrets

- Roger Martínez-Davila (University of Colorado Colorado Springs)

Maria Edgeworth Letters

- Heather Barnes (Wake Forest University)

- Susan Egenolf (Texas Agricultural & Mechanical University)
- Hilary Havens (University of Tennessee Knoxville)
- Carrie Johnston (Wake Forest University)
- Jessica Richard (Wake Forest University)
- Jessica Wilson-Saia (Wake Forest University)

People's Contest Digital Archive

- Kevin Clair (Pennsylvania State University)
- Haven Harrington (Pennsylvania State University)
- Yue Tang (Pennsylvania State University)

Poets & Lovers

- Carolyn Dever (Dartmouth College)
- Peter Logan (Temple University)

PRINT

- Amy Larner Giroux, University of Central Florida

Shadows on Stone

- Roger Panetta (Fordham University)

Additional Cohort Members

Valerie Achterhof (University of Illinois Chicago)

Amy Bailey (University of Illinois Chicago)

Kate Boyd (University of South Carolina)

Graham Duncan (University of South Carolina)

Heather Heckman (University of South Carolina)

Tom Hollenhorst (University of Minnesota Duluth)

Tyler Hoppenfeld (University of California Davis)

Suphan Kirmizialtin (New York University Abu Dhabi)

Rebecca Morgan (American Museum of Natural History)

Ashley Reichelmann (Virginia Polytechnic Institute and State University)

Rebecca Schneider (New Mexico Highlands University)

Emily Stanback (University of Southern Mississippi)

Serenity Sutherland (State University of New York Oswego)

Susannah Ural (University of Southern Mississippi)

David Joseph Wrisley (New York University)

9.3 Institute website

Visit the institute website.

9.4 Works cited

Authored publications

- Breathing new life into death certificate: Extracting handwritten cause of death in the LIFE-M project (Bailey, Leonard, Price, et al.)
- Engaging Crowds: New options for subject delivery & interaction (Blickhan)
- New developments in crowdsourced text transcription (Blickhan)

- The Zooniverse: A Quick Starter Guide for Research Teams (Blickhan)
- "Strangers in the Landscape": On Research Development and Making Things for Making (Blickhan, Granger, Noordin and Rother)
- How to Get Grant Money in the Humanities and Social Sciences (Folsom)
- Building Better Digital Humanities Tools: Toward broader audiences and user-centered designs (Gibbs and Owens)
- The Rise of Crowdsourcing (Howe)
- The Crowd and The Library (Owens)
- The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage (Ridge, Blickhan, Ferriter, et al.)
- Tutorial Designs and Task Types in Zooniverse (Rosser and Wiggins)
- Discovering features in gravitational-wave data through detector characterization, citizen science and machine learning (Soni, Berry, Coughlin, et al.)
- Top Ten Tips - Writing a Great Zooniverse Tutorial (Spiers)
- Preface to Explorations in Economic History: Methods (Wanamaker, Frydman, Dahl)

Web resources and tutorials

- Aggregation for Caesar
- ALICE: the Aggregate Line Inspector & Collaborative Editor
- Appen
- Clickworker
- Comma-separated values
- CoreTrustSeal
- Dryad
- The Endings Project

- Guide to Social Science Data Preparation and Archiving
- Guide to writing "readme" style metadata
- How to set up your project using the Transcription Task
- Mapping Prejudice
- Markdown
- Mechanical Turk
- Multi-Image Manifest generator (Python Script)
- Open ICSPR
- Open Science Framework
- Panoptes CLI
- Zooniverse Python client
- Zooniverse Question Task Aggregation

Made with Material for MkDocs