A brief introduction to crowdsourcing on the Zooniverse platform

Dr. Samantha Blickhan
Zooniverse Co-Director & DH Lead
Adler Planetarium, Chicago

Zooniverse – a beginner's guide
Archives West Midlands
16 November 2022

@snblickhan
@the_zooniverse3
@sam@hcommons.social

https://www.zooniverse.org

ZOONIVERSE

What is crowdsourcing?

ZOONIVERSE

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. (2021) *The Collective Wisdom Handbook: Perspectives on Crowdsourcing in Cultural Heritage* (underlines mine)

Book: https://britishlibrary.pubpub.org
Project: https://collectivewisdomproject.org.uk/

Why crowdsource?

Discoverability	Engagement
Data processing/transformation	Opportunities for exploration
Increased access to collections	Encourage institutional transparency around the digitization process
Results in new, or expanded, reusable datasets	Work with communities toward a shared goal

ZOONIVERSE

What is Zooniverse?

THE ZOONIVERSE WORKS

708,761,484

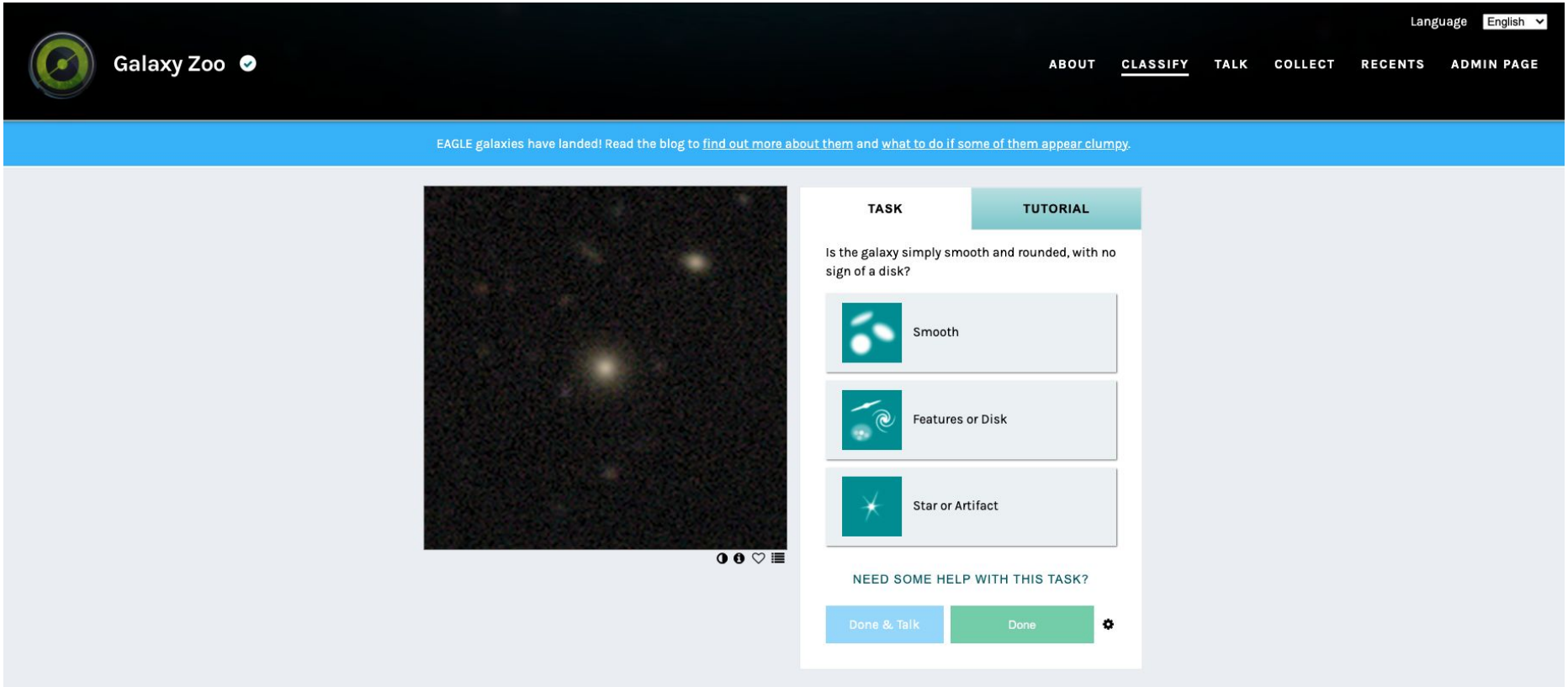
CLASSIFICATIONS SO FAR BY
2,540,706 REGISTERED VOLUNTEERS

A vibrant community. Zooniverse gives people of all ages and backgrounds the chance to participate in real research with over 50 active online citizen science projects. Work with 1.6 million registered users around the world to contribute to research projects led by hundreds of researchers.

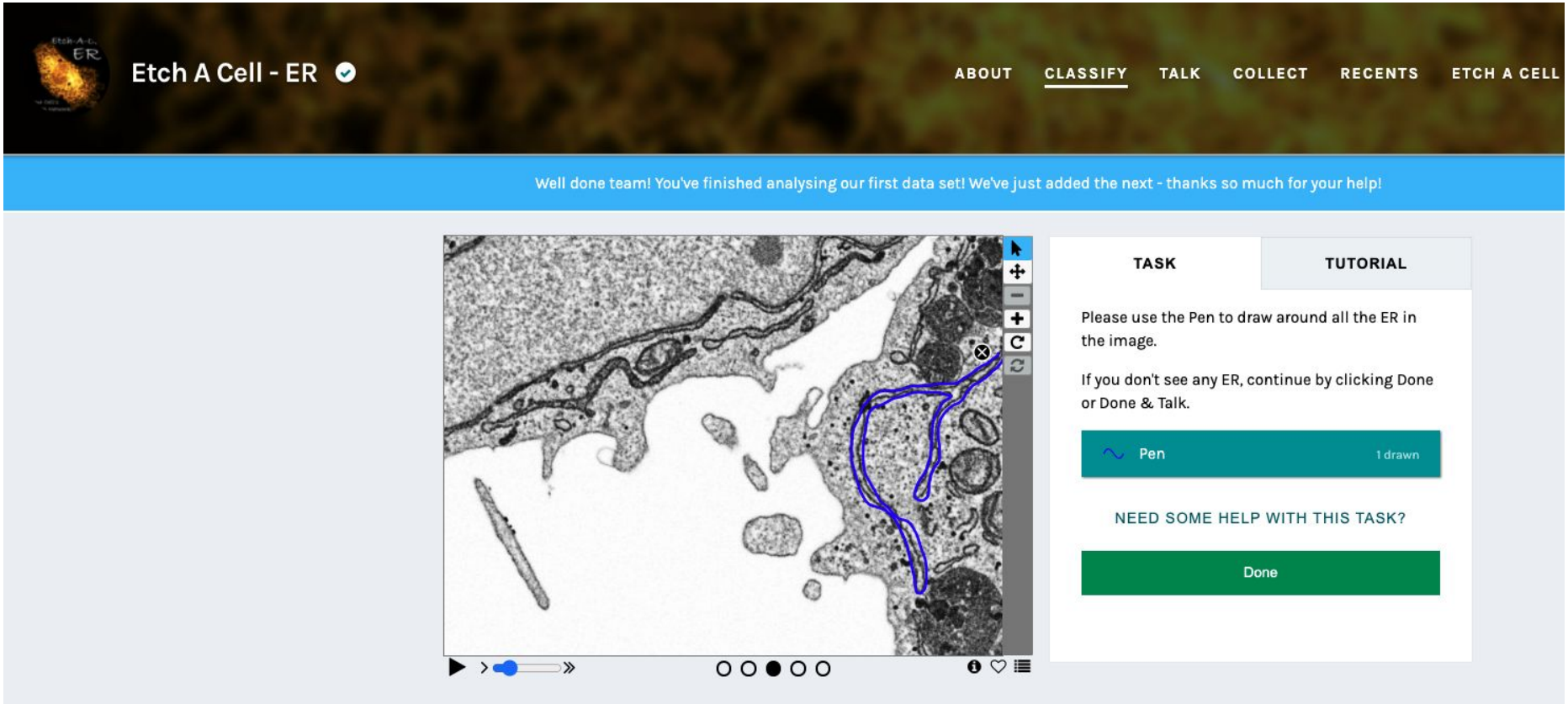
SIGN IN OR REGISTER TO GET STARTED

Sign in | Register

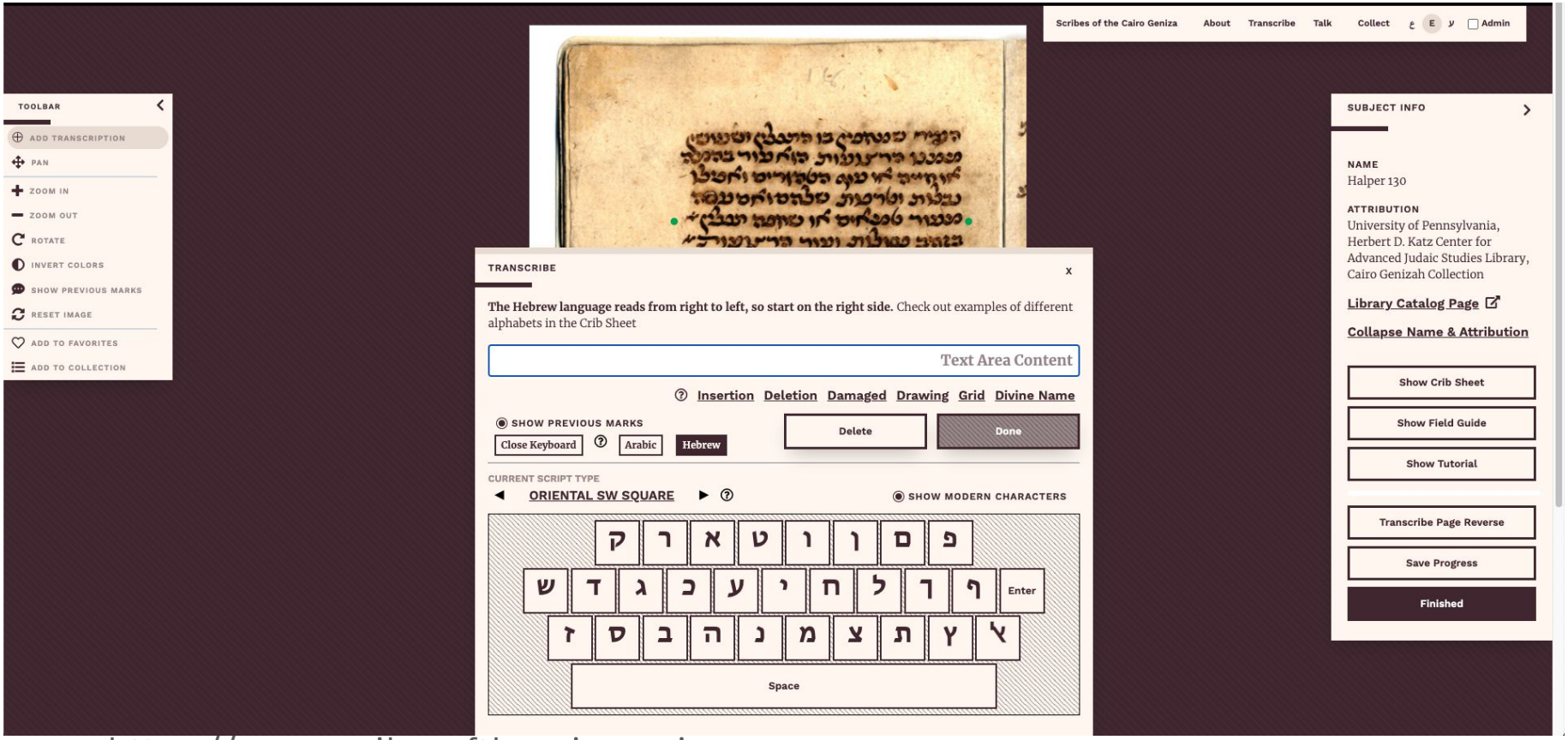
https://www.zooniverse.org



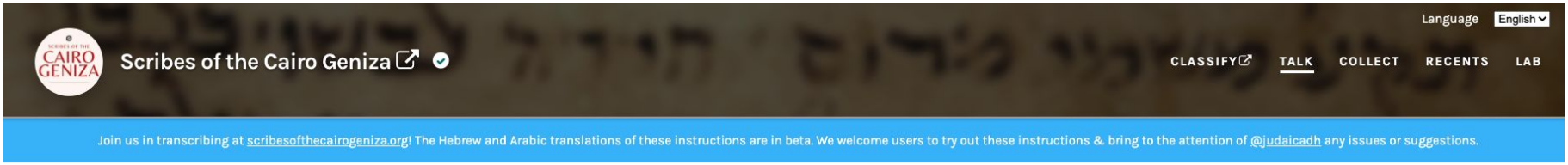
https://www.galaxyzoo.org



https://www.zooniverse.org/projects/h-spiers/etch-a-cell-er



https://www.scribesofthecairogeniza.org



Scribes of the Cairo Geniza Talk

Search or enter a #tag

Moderator Controls

Notes/Subject Discussions
General comment threads about individual subjects
citsci-rancho Subject 37169022 *2 days ago*
597 Participants
16934 Discussions
28538 Comments

News
Updates on the Project
ajohnst2 Research Assistance Request- Mental Health Benefits of Zooniverse *a month ago*
16 Participants
31 Discussions
175 Comments

Introductions
Volunteers are Encouraged to Say "Hi" and Introduce Themselves
judaicadh RESEARCHER MODERATOR TEAM Introductions *2 months ago*
71 Participants
2 Discussions
109 Comments

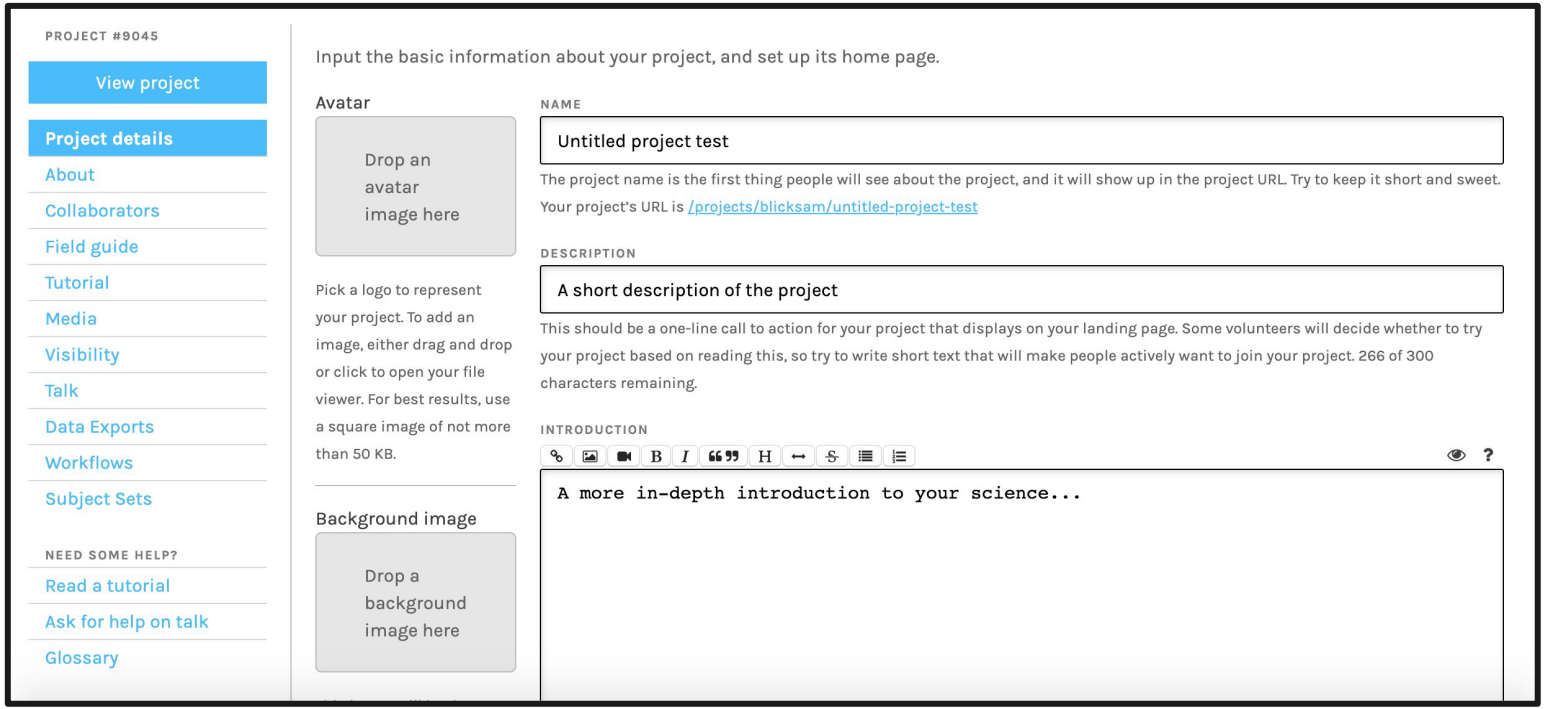
Questions for Research Team
Ask the Experts
BlueEyes074 ligature *2 months ago*
65 Participants
96 Discussions
406 Comments

Help and Technical Issues
Ask for Tech Help, Report Bugs, Etc.
54 Participants
53 Discussions

Recent Comments

Popular Tags:

nikkud
name_of_god
faded
torn
colons
hebrew
hebrewscript
cantillation
microfragment
marginalia
largeletters
printed
damaged
arabicscript
parchment
ink_corrosion
asktheexperts
binding



https://www.zooniverse.org/lab
https://www.zooniverse.org/publications

Data export types (see Data Exports tab of PB)

- Classification export: all task data from all workflows for your entire project
  - JSON in .csv
- Workflow classification export: all task data from a specific workflow
  - JSON in .csv
- Workflow export: structural information about how your workflow looks (w/ version info)
  - JSON in .csv
- Subject export: list of all subjects you have uploaded w/ any metadata included
  - JSON in .csv
- Talk export: all comment and hashtag data from Talk
  - JSON
- Details at: https://help.zooniverse.org/next-steps/data-exports/

What is ‘raw’ vs ‘aggregate’ data?

- Zooniverse classification data exports are displayed in a spreadsheet where 1 row = 1 classification. This is your ‘raw’ data — thousands of rows of classifications. These can be exported for the entire project (all workflows), or on a workflow-by-workflow basis.
- In order to turn those rows of data into useful project results, you can use Python scripts to group the data together based on the workflow and task, and then combine the grouped classifications to produce aggregate results based on your needs.
    - ‘Results’ here may vary depending on your goals — they could be a tally of all total responses (e.g. tagging), or the majority response for a given task (e.g. a yes/no question)
- Once your data is aggregated, you can then join those results back to the original file uploaded to the project.

Using data exports to influence project design

- Reviewing your data early on in the project building process can help ensure you're getting usable data
- This is a good reason to generate data exports early and often
- Testing your data aggregation process using data from your beta test is a great way to make sure you are fully prepared for a full project launch
    - If you aren't able to work with the data from your beta test, this is a good sign you need to change your workflow and/or recruit additional team members to help with this part of the process. If you can't do this step, you really shouldn't feel comfortable asking volunteers to donate their labor to your project.

What if my data is ‘bad’?

- There are a few types of ‘bad’ data:
  - Data that is inaccurate or low-quality
  - Data that is unusable for your project goals
- If your early data exports include either of these ‘bad’ data types, it’s a good sign you need to revise something in your workflow.
  - This could be your design, your instructions (Tutorial, Help, etc.), or your data aggregation process.

ZOONIVERSE

How to start thinking about a crowdsourcing project

Adapted from *The Collective Wisdom Handbook*, ch. 5, ‘Designing cultural heritage crowdsourcing projects’

Getting started

"The earliest steps of a crowdsourcing project can be the most challenging." (*TCWH* ch.5)

1. Articulate motivations
2. Describe project data
3. Consider your goals *beyond* data
4. SWOT analysis/SMART objectives
5. Outcomes
6. Values
7. Identifying resources & skills

What are your motivations?

- What is the problem you are trying to solve?
- What does this work/project enable?
- What does the output data look like?
- What are example/possible research questions that could be answered using what is created through your project?
- Who is missing or excluded if you undertake this project in the way you imagine it now?
- Why is crowdsourcing the best approach? What does it bring that you would not get otherwise?

Describe your data

- What is your project data?
- What are you going to do to/with this data?
- What form is your data in (currently)? What form will it be in at the end of the project (ideally)?
- Where will you store the data produced through this project?
- Are there potential sub-groupings of your dataset that will impact your project?

What are your goals *beyond* data?

Consider the additional opportunities crowdsourcing can provide.

- Who are your audiences and how will you engage them beyond the main project task?
- How can you provide an educational opportunity for participants?
- How can you engage historically underrepresented communities?
- How can you offer mentoring or insight on career paths?
- How can you build support for your institution or partner institutions you may be working with?

Goals vs. objectives

- Goals are the highest-level articulations of ideas, and are typically somewhat abstract.
- Objectives articulate concrete actions that advance a project toward goals. They are generally easier to translate into work plans and to evaluate progress.
- Goals are often what make it into communications — they are more of a rallying point. Focus on your goals initially.
- How do your project goals align with your institutional strategy? Alignment here can help motivate your institution to support your project.

SWOT analysis

Strengths | **W**eaknesses
Internal to your project or institution

- Subject matter expertise
- Curatorial/archival practice
- Existing audiences and/or volunteers
- Limited financial resources
- Limited tech/development support

Opportunities | **T**hreats
External to your project or institution

- External partnerships
- Can design new collaboration practices
- Resources
- Sustainability
- Management overhead

SMART objectives

SMART objectives

Specific

- **What** — what will this project create? (Data transformation, communities)
- **Why** — why does this project matter?
- **Who** — who will be invested in the project? (Goals, outcomes, communities)
- **Where** — where will this project's activities take place? (Original data included)
- **Which** — which needs will be vital to the project's success? (People, resources, funding, etc.)

Measurable

- Measurable goals help you track your progress and evaluate your final product
- These goals are based on your responses to the five Ws
- This can also help you make sure that you are tracking (and eventually acknowledging!) labor appropriately as you assign these tasks to your team

Achievable

- When faced with a list of actual tasks to complete, it's much easier to determine whether you have the necessary resources to complete the work
- Make sure you've considered external resources here, too. The primary example in this type of project is the size of your crowd — given your needs and goals, what will happen if your crowd is smaller than you anticipate? What if it's larger?
- Are any of your tasks or milestones tied to a specific funding schedule, e.g. research funding? These may come with hard deadlines.

Realistic

- Find the right balance for your team between ambitious and modest
- Make sure your goals reflect the needs of your data
  - e.g. you may find ‘extra’ data you want to process in a similar way, but if it doesn’t fit with the design choices you’ve made for your original dataset, you may wind up with unusable results
- Make sure your goals reflect the needs of your audience(s)
  - Similarly, if your community is interested in a particular subject, adding new data that is unrelated may cause some people to stop participating

Time-bound

- Your goals should have clear start and end dates
- These dates can shift, but the allocated effort should remain the same unless otherwise determined through experience

Outcomes

- What does success look like for your project?
- To determine success, you need to build in opportunities for evaluation *and* iteration
- What kind of results will demonstrate a successful project?
- If you have a specific research goal, how does this fit in with the broader success of the project? If you're *only* thinking about your personal research outcomes, you are likely missing at least one of the steps we've discussed today (most likely values!)

Values

- Once you have determined goals, motivations, and milestones, you should think about values
- Doing this early on in the process will empower you and your team to make decisions that support those values
- This will also empower all participants (team members and volunteers alike) to challenge actions that aren’t in line with those values

Clearly defined values allow you to:

- Prioritize activities, allocate resources, define success, enable participation
- Build trust among all project stakeholders through accountability
- Increase the likelihood of a successful project

Undefined values still exist

- Choosing not to articulate your values does not make you Switzerland
- If you are inheriting values from a parent institution (via a mission or vision statement) , you still need to make sure they are communicated to your external stakeholders (including volunteers)

Recognizing resources

Key project resources

- People
- Technology
- Finances
- Context/information/knowledge

People & skills inventory

- Identify stakeholders: who are the people involved?
- Of those people, what skill sets do they possess?
    - Content specialization, project or community management, communications, data management or analysis, design, software or web development, etc.

ZOONIVERSE

Resources

Exercise: the 5 Ws

- **What** — what will this project create? (Data transformation, communities)
- **Why** — why does this project matter?
- **Who** — who will be invested in the project? (Goals, outcomes, communities)
- **Where** — where will this project's activities take place? (Original data included)
- **Which** — which needs will be vital to the project's success? (People, resources, funding, etc.)

Next steps

- Try different projects! https://www.zooniverse.org/projects
- Read the docs https://help.zooniverse.org
- Familiarize yourself with the process: https://help.zooniverse.org/best-practices/4-resources/ (see especially “Project Review: Best Practices & Flowchart”
- Transcription tools overview: https://sites.google.com/zooniverse.org/new-developments-in-ctt/home
- Questions? Check the Project Building Talk Board: https://www.zooniverse.org/talk/18
- Still have questions? Email contact@zooniverse.org