

Making Collections More Visible with Zooniverse



Kelly Hovinga
Program Assistant
University of Michigan
Shapiro Design Lab

Learning Objectives

1. Explain the connection between promoting library resources/expertise and the involvement of library users in the creation of knowledge.
2. Be able to identify and assess collections or resources at your institution that could benefit from becoming Zooniverse projects.
3. Understand what goes into making a Zooniverse project

Background Information



University of Michigan Archive: researchers documenting birdsong

Citizen Science

Citizen Science = the crowdsourcing of scientific research to the general public; thus allowing greater scientific discovery and a closer working relationship between the non professional citizen scientists and professional scientists.

What is Zooniverse?

Zooniverse is an online platform for the creation and hosting of Citizen Science projects.

“The Zooniverse is the world’s largest and most popular platform for people-powered research. This research is made possible by volunteers — hundreds of thousands of people around the world who come together to assist professional researchers. Our goal is to enable research that would not be possible, or practical, otherwise.”

Quote from www.zooniverse.org/about



Image courtesy of Wikipedia

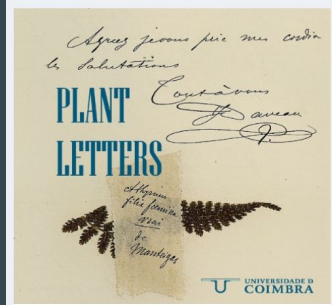
Projects related to library and archive collections



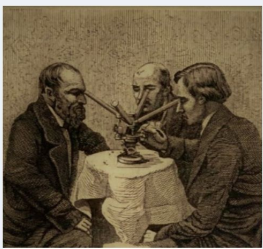
UNEARTHING MICHIGAN
ECOLOGICAL DATA



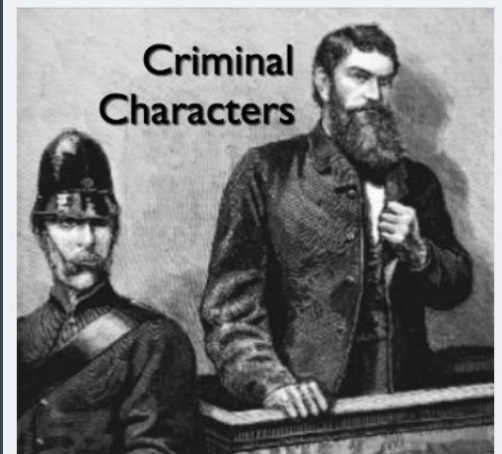
CASTAWAY



PLANT LETTERS



WORLDS OF WONDER



CRIMINAL CHARACTERS

www.zooniverse.org

What this means for your collection

1. Creation of metadata
 - a. Making your collections more accessible
2. Engages an audience in deep usership of the collection
3. Makes the collection more visible through other forms of media
4. Creates a community stake in the collection through collaborative ownership

Citizen Science, Education, and Engagement

Criminal Characters Zooniverse project, from the about page.



The Citizen, A Weekly Budget of Passing Events and Police News (Melbourne), 2 June 1877. Courtesy of State Library of Victoria.

What do I mean by engagement?

I am ascribing to the Appleton/Christenson theory that engaged learning is based on three metrics:

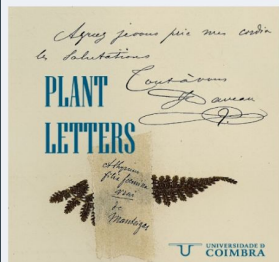
1. Social relationships
2. Emotional relationships
3. Academic relationships

Citizen Science and Engagement

Phillips, Ballard, etc.

- Citizen Science has all the potential for all the components of engaged learning, and so does the Zooniverse platform.
 - Social engagement
 - Talk section and newsletters
 - Emotional engagement
 - Fast classification methods and the larger impact of the research.
 - Academic engagement
 - Subject matter and the ability to interact with the material on both a surface and deep level.

The Project



PLANT LETTERS

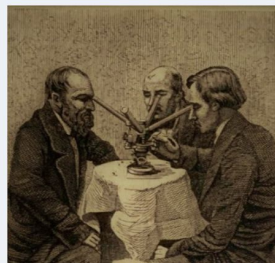


SNAPSHOT MADIKE



SNAPSHOT DEBSHAN

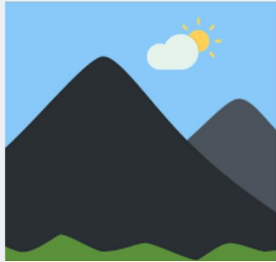
Help scientists understand the diversity and dynamics of wildlife populations in Central Zimbabwe!



WORLDS OF WONDER



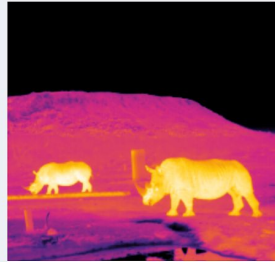
MUON HUNTERS 2.0



WEATHER RESCUE



UNEARTHING MICHIGAN
ECOLOGICAL DATA



ASTRO-ECOLOGY

Screenshot of the
Zooniverse project page.

What makes a good collection for a project?

1. Already digital
2. Should at least be 50,000 images for classification
3. Marketable
4. Data classification can be broken into small tasks

Take a moment to think about a collection you have that might qualify as a good data set for a zooniverse project.


What are the components of a good Zooniverse project?

- About page with clearly expressed goals and impact statement
- Tutorial
- Active, monitored chat section
- Simple workflow tasks
- Field guide
- FAQ page
- Legal agreements
- Data curation plan

About Page

About Criminal Characters

Crime has been central in shaping the history and society of Australia. This project will make a significant contribution to family, local, social and criminal justice history by revealing untold stories about the lives of people who committed crimes in Australia across time. It aims to discover new perspectives on the types of factors that led to individuals ending up in the prison system. In particular, it is hoped that the research will challenge existing ideas about what the label of 'criminal' has historically meant by revealing the diverse nature of the people who spent time in prison.

Notes from Nature - WeDigFLPlants 

ABOUT CLASSIFY TALK COLLECT RECENTS NOTES FROM

Research

The Team

The plant specimen data created by WeDigFLPlants enable conservation biologists, natural resource managers, ecologists, and others to understand changes to the distribution and abundance of the flora of Florida, especially the components that are rare/endangered or introduced/invasive or of high economic value. The data are also critical to curators of herbaria who manage the specimens for use in perpetuity.



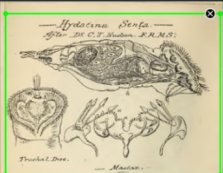
Melbourne Gaol, Russell Street, circa 1937. Courtesy of State Library of Victoria.

Transcription of these prison records offers unprecedented opportunities to discover how criminal offending fitted into the wider lives of offenders, as well as illuminating the involvement of individuals in a range of criminal activity from the end of the convict period through to the start of the Second World War. The creation of

About page tips

- Don't use jargon
- Use pictures
- Explain what is going to happen to the volunteers' work (why it matters)
- Clearly expressed goals

Tutorial



Hydraspis, Sericea
From DE C.T. Hudson, F.R.S.E.
Doubtful, Dec. 1890. ...

Hydraspis sericea, one of the beetles or weevils, anatomically is generally found in company with Byrrhus Verrilli. The Byrrhus are eaten at a rate of several individuals, and I have not yet seen how many I have seen consumed, for fear of being charged with exaggeration, but let us be content with formal a weevil, by a large beetle anatomically. Hydraspis enters the rougher, at least also enters there into his wide open mouth, much as a toadfish would swallow a spade, and instantly crushed them before your eye in the beetle's mouth, which whole whole, at the bottom of his throat.

TASK TUTORIAL

Choose the type of illustration, then draw rectangles around each illustration of that type.

- ☒ Drawing of microscopic specimen 1 drawn
- ☐ Photograph of microscopic specimen 0 drawn
- ☐ Drawing of instrument/tool 0 drawn
- ☐ Photograph of instrument/tool 0 drawn
- ☐ Technical drawing/construction manual 0 drawn
- ☐ Other 0 drawn

Mark any illustrations by choosing the type of illustration and drawing a rectangle around each of them. If you see a plate with multiple illustrations of the same type, you can draw a rectangle around all of them.

Continue

◀ ○ ● ○ ○ ▶

Cedar Waxwing, Bombycillus cedroris
Baltimore Oriole, Icterus galbula
Scarlet Tanager, Piranga olivacea
Black and White Warbler, Mniotilta varia
Chestnut-Sided Warbler, Dendroica pensylvanica
Blackburnian Warbler, D. fusca
Redstart, Setophaga ruticilla


What to do

Each subject is a photo of a single page from the Biological Station student papers. As you scan or read each page, look for the following: maps, graphs, tables of numbers, species listed or named, diagrams, or photographs. Pages may include multiple types of data, so click on all options that apply. If you don't think there is any data on the page, click the "there is no data on this page" option.

Tip: Species names are often underlined or italicized in paragraphs.

Continue

◀ ○ ● ○ ○ ○ ○ ▶



Wildebeest

Long-legged, with dark grey-brown fur with faint stripes of darker fur on neck and flank. Large head with a rounded black nose and white 'beard'. Stands higher at the shoulder than the rump. Both sexes have horns that extend out sideways then curve sharply back towards head.

Often confused with [Buffalo](#)

Looks good!

By comparing the animal to the photos here and reading the description below, we can confirm whether we have the right animal. You can flip through the photos to see examples of the animal from multiple angles using the bullets

Continue

◀ ○ ○ ○ ○ ● ○ ○ ○ ○ ▶

Tutorial tips

- Keep it short (4-6)
- Use pictures and notate them
- Don't write too much
- Use simple, explicit language
- Greet and thank them

Talk section

Galaxy Zoo Talk

Search or enter a #tag



Notes

General comment threads about individual subjects



[miai](#) Subject 32499347 [14 minutes ago](#)

1827 Participants

25696 Discussions

46401 Comments

Help

A place to get general help for using the site.



[Ine Theunissen](#) Possible transient in IC 2392 and NGC 606 ? [5 hours ago](#)

170 Participants

311 Discussions

1453 Comments

Science

A place to discuss the science behind the site, and inspired by the results



[CeciliaB](#) Voorwerpje Hunt 2019 [a day ago](#)

84 Participants

260 Discussions

1771 Comments

Chat

A place for general chat



[JohnF.](#) "Neutron star and black hole binary mergers: the first results of the LIGO-Virgo era" - Live-Streaming RAS conference (Friday 9th May 2019) [3 days ago](#)

106 Participants

170 Discussions

740 Comments

Recent Comments

Popular Tags:

[starforming](#)

[spiral](#)

[agn](#)

[starburst](#)

[disturbed](#)

[ring](#)

[merger](#)

[edge-on](#)

[bar](#)

[overlap](#)

[irregular](#)

[barred-spiral](#)

[star-forming](#)

[dustlane](#)

[hot](#)

[elliptical](#)

[decals](#)



[asteroid](#)

Talk tips

- “There isn’t a button for that”.
- Unexpected discoveries.
- Feedback and forums - establish expectations from the onset, draft a set of rules and stick to them, confrontations take place outside the forum, ignore conspiracy theorists, encourage side projects.
- You have to talk to them, and so do the experts.

Workflow


Definition: Set of tasks that citizen scientists complete to create data





TASK

TUTORIAL

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



 Smooth

 Features or Disk

 Star or Artifact

NEED SOME HELP WITH THIS TASK?


Next →




TASK

TUTORIAL

Could this be a disk viewed edge-on?

 Yes - Edge On Disk

 No - Something Else

NEED SOME HELP WITH THIS TASK?

Back

Next →

This project is 4 -7 tasks long.

Workflow tips

- Keep the number of tasks under 5 (7 max if your subject is a big crowd pleaser)
- Make the tasks short/simple (but not too simple)
- Use truncation
- Focus on only one or two types of data you want to collect
- Use drop downs or multiple choice whenever you can

Field Guide

FIELD GUIDE

What is the prisoner's weight? If missing, write "Missing".

10f

-

BCH or BR

CE or C of E

NEED SOME HELP WITH THIS TASK?

CUM or CUM L

No

D&D or DR & DIS

NEED SOME HELP WITH THIS TASK?

DO or DITTO

Next →

False Repres

F'ROY

FIELD GUIDE

Diagram/Illustration

TORIAL

Map

Graph

Table of numbers

Species listed or named

Photograph

Data that doesn't fit

Illegible and blank pages

Field Guide tips

- Have one
- Use pictures that exemplify variations that are seen in the classifications
- Bold the important information
- Don't be afraid to provide deeper information

FAQ page

[Research](#)[The Team](#)[FAQ](#)

What is the purpose of this project?

The purpose of this project is to reuse environmental and ecological data going back a century. With this data, researchers will be able to better identify locations, species, and topics for future studies and place their work within a long term narrative of ecological changes. Additionally, the reused data can inform policy decisions for environmental management.

What if I get it wrong? Will it mess up the data?

No it will not! We show the image to lots of people, and combine all of those answers. So even if one person isn't quite sure, there is a bunch of people to check the work. We think that overall, people will be pretty good at identifying data types. For images that result in a lot of disagreement, they are sent to and reviewed/identified by an expert.

Isn't everything data to someone? So shouldn't I be saying "other" for all the information in the paragraphs?

That is a great point! Yes, all information can be used as data; but for this project, we want you to focus on the forms of data we have defined. Each of the images you classify is part of a larger document, so the data on one page is tied to the data from all the other pages in that document. Even if you mark a page as "no data", we will still know the species, measurements, and diagrams in the rest of the document. Researchers and policy makers will then be able to read through the correct documents for narrative data (the information in paragraph format) if they need it.

Why can't a machine do this?

Machines and machine learning for image recognition has yet to advance to the point where it can reliably identify images. Additionally, since many of these papers were done before the widespread usage of computers, the text and data are not in standardized formats. That is why we need your help to unearth the data!

What do I do if there are multiple data types on a page?

FAQ tips

- Two questions that are always asked: “why can’t a machine do this” and “what if I am wrong?”
- Change up what is in the FAQ after the first round of beta testing
- Don’t be afraid to repeat information in any of the other pages
- Ask yourself “what else will they want to know?”

Legal agreements

Your University's copyright office and legal council office.

Scassa and Chung, *Managing Intellectual Property Rights in Citizen Science*:

https://www.wilsoncenter.org/sites/default/files/managing_intellectual_property_rights_citizen_science_scassa_chung.pdf

Zooniverse Privacy Policy: <https://www.zooniverse.org/privacy>

Cit Sci.org:

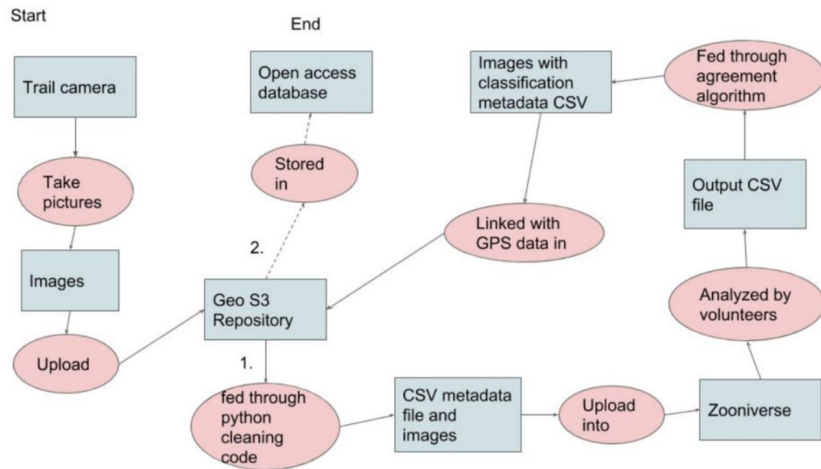
<https://www.citsci.org/CWIS438/Websites/CitSci/Data.php?WebSiteID=7>

Components of a data curation plan

- Place for the data to live in current format
- Place for the data to live in its raw format along with the programs and methods used to create the current format
- Plan and timetable for backing up the data
- Method for the public to access the data, but not the backup
- Staff to maintain and follow the curation plan
- Timetable for reassessment of data type and transfer to new formats

Map out the life of your data

For example, below is a simplified version of the data plan for **Michigan ZoomIN**:



Key:

————— Leads to
- - - - - Leads to,
second process

Object

Process

Links for technical backend

-Collection of code for Zooniverse projects:

<https://github.com/zooniverse>

-S3 Bucket:

<https://docs.aws.amazon.com/AmazonS3/latest/dev/UsingBucket.html>

-Zooniverse help section:

https://help.zooniverse.org/getting-started/?_ga=2.182895718.1320338486.1557771978-1144094771.1557771978

What this all actually looks like

<https://www.zooniverse.org/lab>

(This link is for me)

Advice from personal experience

- Start your discussion with a data curation specialist in the room as well as the collection specialist.
- Talk to your researchers and figure out what types of information they would use.
- If you plan on going the social media route, make sure you have the staff and stories to keep it active.
- Zooniverse has its own newsletter to advertise your project, but be prepared for outreach events when your volunteers start to slow down.

Conclusion

1. When people have a stake in the creation of information, they care more about it.
2. Engagement is community and activity based and so is Zooniverse
3. A good Zooniverse project is composed of numerous parts and requires maintenance
4. What you do with the data once it is created is just as important as the creation process
5. Zooniverse projects are still a great way of making a collection more visible and creating engagement

Questions?

Bibliography

Open Link:

<https://docs.google.com/document/d/1HCiHyjdC85Dve5Gw7-ae0qc7EBvS-NXZVFetil3-Tek/edit?usp=sharing>