



Zooniverse for Heritage Projects: Zooming through Zooniverse Data

Emily Long
ConCode
April 15th, 2021

Code and tutorials at:

<https://github.com/emilyrlong/HeritageCombs>

<https://rpubs.com/emilylong/zooniverse>

-
1. Introduction to the Heritage Combs Project
 2. Zooniverse - How to build a workflow
 3. How to clean data in R
 4. A highlight of some of the findings from the combs project (if we have time)



Heritage Combs:

A group project developed in the **Crowd-Sourced and Citizen Data for Cultural Heritage** module in the UCL Data Science for Cultural Heritage MSc programme.

Thank you to everyone who participated!



Group Members:

- Tianqi Yang
- Yijun Chen
- Irene Iriarte Carretero
- Wangxue Zhao
- Hongrong Wang
- Emily Long

Module Leads:

- Dr Josep Grau-Bové (project supervisor)
- Dr Pakhee Kumar
- Rosie Brigham



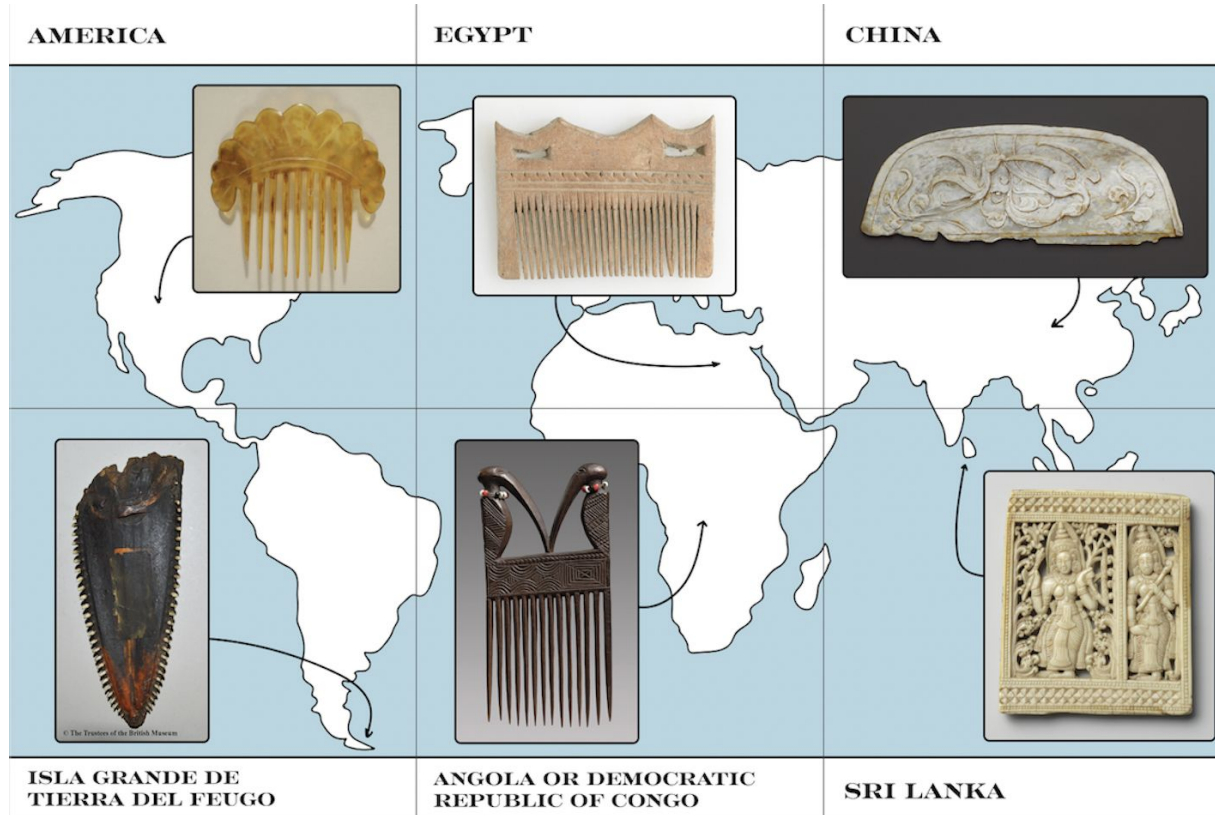
What is crowdsourcing?

[1] Estellés-Arolas, Enrique & González-Ladrón-de-Guevara, Fernando, 2012. Towards an integrated crowdsourcing definition. *Journal of information science*, 38(2), pp.189–200.

Definition [1]:

- A participative online activity
- The volunteers are a group with varying knowledge, heterogeneity, and number
- A flexible, open call
- Mutually beneficial:
 - the volunteer should gain something (economic benefit, skills, self-esteem, etc.)

How do age, material, decoration and degradation affect the perceived value of combs and their suitability for display?



Comb images courtesy of the Rijksmuseum, the Cleveland Museum of Art, the Trustees of the British Museum, the Metropolitan Museum of Art.

Degradation

None



Image courtesy of the Art Institute of Chicago.

A little



Image courtesy of the Rijksmuseum.

Some



Image courtesy of the Rijksmuseum.

Severe



Image courtesy of the Rijksmuseum.

Decoration

Plain



Image courtesy of the Art Institute of Chicago

Handcraft



© The Trustees of the British Museum

Handcraft + Jewels (Both)

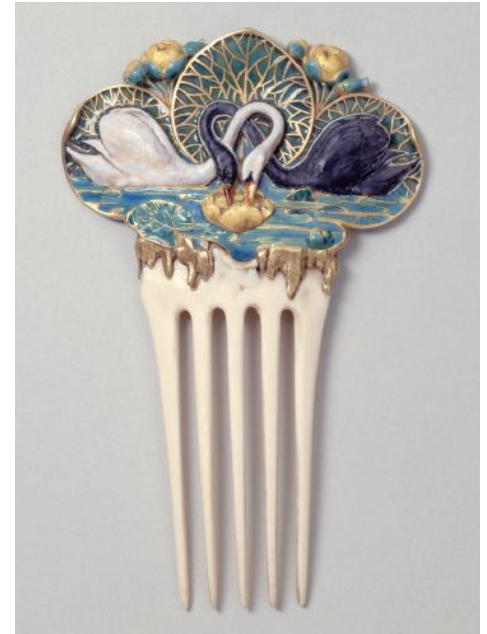


Image courtesy of Paris Musées

A/B Test for Labelling

A



Digital image courtesy of the Rijksmuseum.

B

c. 1800 - 1825, tortoise shell, Netherlands



Digital image courtesy of the Rijksmuseum.

Participant Engagement

- Data gathered from Feb. 23 - Mar. 6, 2021
- 149 participants from around the world
- 1,522 responses total
 - 798 for unlabelled images (A)
 - 724 for labelled images (B)
- Average of ~10 responses per participant
- 172 comments for optional question

c. 3900 - 3500 B.C.,
ivory, Egypt



Digital image courtesy of the
Metropolitan Museum of Art.

To the Zooniverse!

Heritage Combs: <https://www.zooniverse.org/projects/ucl-dot-dasch/heritage-combs>

Penguin Watch: <https://www.zooniverse.org/projects/penguintom79/penguin-watch>

Scribes of the Cairo Geniza Transcription: <https://www.scribeshofthecairogeniza.org>

Adler Planetarium Metadata Tagging: <https://www.zooniverse.org/projects/webster-institute/tag-along-with-adler>

Iberian Camera Trap Project: <https://www.zooniverse.org/projects/aicensusuhu/iberian-camera-trap-project>

On to the data!

Python Code from Zooniverse:

<https://github.com/zooniverse/Data-digging>

R Code from Emily:

<https://github.com/emilyrlong/HeritageCombs>

Problem: Non-Logged In User Data

	classification_id	user_name	user_id	created_at	session	session_id
row names	323860447	not-logged-in-36263ca2433b26bb9368	1	2021-04-13 11:00:49 UTC	dc54698ee38e710c3422a93e3744a5966a9c020666...	1
2	323860744	not-logged-in-fe37a93ebaa987c0e3a7	2	2021-04-13 11:02:59 UTC	9331b9bf152a5e37160ba4dcf146bac7e3051618636...	2
3	323860817	ucl.dasch	3	2021-04-13 11:03:33 UTC	b6ceb8505af8b3c45c4988f8861d1b77fe547de1cf02...	3
4	323860933	not-logged-in-09493619f2e13b0a3aa5	4	2021-04-13 11:04:20 UTC	dc54698ee38e710c3422a93e3744a5966a9c020666...	1
5	323860980	not-logged-in-09493619f2e13b0a3aa5	4	2021-04-13 11:04:44 UTC	dc54698ee38e710c3422a93e3744a5966a9c020666...	1
6	323861061	not-logged-in-59ee8feae6e9af10144d	5	2021-04-13 11:05:26 UTC	9331b9bf152a5e37160ba4dcf146bac7e3051618636...	2
7	323861122	not-logged-in-59ee8feae6e9af10144d	5	2021-04-13 11:05:58 UTC	9331b9bf152a5e37160ba4dcf146bac7e3051618636...	2
8	323861257	ucl.dasch	3	2021-04-13 11:06:50 UTC	b6ceb8505af8b3c45c4988f8861d1b77fe547de1cf02...	3
9	323861310	ucl.dasch	3	2021-04-13 11:07:12 UTC	b6ceb8505af8b3c45c4988f8861d1b77fe547de1cf02...	3
10	323861400	not-logged-in-59ee8feae6e9af10144d	5	2021-04-13 11:07:47 UTC	9331b9bf152a5e37160ba4dcf146bac7e3051618636...	2
11	323861549	not-logged-in-59ee8feae6e9af10144d	5	2021-04-13 11:08:46 UTC	9331b9bf152a5e37160ba4dcf146bac7e3051618636...	2
12	323861771	ucl.dasch	3	2021-04-13 11:10:04 UTC	b6ceb8505af8b3c45c4988f8861d1b77fe547de1cf02...	3
13	323861953	ucl.dasch	3	2021-04-13 11:11:08 UTC	b6ceb8505af8b3c45c4988f8861d1b77fe547de1cf02...	3

Key Takeaways:

- Encourage participants to log into Zooniverse!
- Increase the benefit for participants



Thank you!

If you have any questions, feel free to reach out via ConCode or my

Email: emily.long.17@ucl.ac.uk

Twitter: @EmilyAndTheData

LinkedIn: <https://www.linkedin.com/in/emily-long8297/>

Portfolio: <https://emilylongdata.myportfolio.com>