Conference Abstract

Setting Collections Data Free with the Power of the Crowd: challenges, opportunities and a vision for the future

Margaret J Gold ‡

‡ Natural History Museum, London, United Kingdom

Corresponding author: Margaret J Gold (m.gold@nhm.ac.uk)

Abstract

The Natural History Museum in London has embarked on an epic journey to digitise 80 million specimens from one of the world's most important natural history collections. Publishing this data to our open Data Portal will give the global scientific community access to unrivalled historical, geographic and taxonomic specimen data gathered over the last 250 years. We have been involving the public in this large task by asking digital volunteers to help transcribe information written on specimen labels, via an online crowdsourcing interface, such that anyone in the world can participate. In this talk we will share our experience with these projects over the past year.

Within the Digital Collections Programme at the Natural History Museum London we have so far digitally imaged approximately 200,000 of our total holdings of around two million microscope slides, and from these have selected two discrete collections for crowdsourcing the transcription effort. Miniature Lives Magnified (now complete) contained 6,285 digitised Chalcidoidea slides from our Hymenoptera collection, and Miniature Fossils Magnified (not yet complete at time of writing) contains 2,000 Foraminifera slides from our Micropalaeontology collection. Both projects are hosted on the Notes from Nature platform, which has been built on the Zooniverse Panoptes open-project platform.

As the lead partner of the Crowdsourcing task within SYNTHESYS3, an EC-funded project creating an integrated European infrastructure for natural history collections, we have partnered with our fellow consortium members to help them design and launch crowdsourcing projects of their own. These have included an Amaranthaceae collection with 444 specimens and a Primulaceae collection with 3,093 specimens hosted on Notes from Nature, and a Brachiopod collection with 1,810 specimens built directly on Panoptes.

In this talk we will share the key insights gained through practical experience with this wide range of specimens, specimen data, and label-styles. In particular, we have gained insights into the design of the workflow and interface, such as the considerable reduction in human error when drop-down menu options are introduced where possible, rather than free data entry fields.

These five projects provided us with a unique opportunity to compare the dedicated Notes from Nature platform, which has significant advantages due to the size and engagement-level of the existing community, to the open project-building Panoptes platform, which has storytelling advantages in terms of the capacity to provide more information about a specific collection, its subject, and its underlying scientific importance.

A crucial element of running successful crowdsourcing projects is building an engaged community of digital volunteers. We compared the use of social media channels with more traditional Museum communication channels (such as e-newsletter and website), and found that the latter had the most reach in terms of raising awareness of the projects, but that the former enabled more frequent and varied engagement with a potential volunteer audience. However, when examining which metrics are the most important to track in assessing the success of various initiatives, we found that the highest impact on the ultimate volume of transcription were in-house volunteering days run in person, rather than online.

In reaching out and engaging with a diverse range of volunteer audiences, we found evidence of the major sources of motivation that are described in the existing citizen science literature, but also more nuanced insight into behaviours such as pursuing independent learning, the desire to enter all of the information even when not requested, and preferring tasks that can be performed by rote. Our efforts to support and nurture the existing Notes from Nature community confirmed the importance of the principle of 'giving-back', and gave us insight into how to do this when research results emerge over a longer timeline than is typical of field-based citizen science projects.

And finally, we will share our experience with the behind-the-scenes elements of crowdsourcing - the parts the 'crowd' doesn't see - such as data quality assessment, data ingestion, data publication, and the flow of data between internal systems.

In conclusion we will propose some visions of the future, such as moving towards a global platform for specimen label transcription with a shared underlying database infrastructure, how to deepen the engagement of digital volunteers from transcription tasks to scientific

observations, and ways to bring online crowdsourcing and field-based citizen science together in a more streamlined way.

Keywords

Citizen Science, Crowdsourcing, Public Engagement, Community Building, Website Interface Design, UI/UX, Data Quality, Data Management, Collections Management

Presenting author

Margaret Gold, Natural History Museum London

Hosting institution

Natural History Museum London

Author contributions

Margaret Gold, Natural History Museum London