Dataset Review—*Zooniverse, the online repository for Citizen Science*

Author/Compiler: Chris Lintott, PI (Oxford University), Laura Trouille, co-I (Adler Planetarium)

Repository: http://zooniverse.org - Operated by Citizen Science Alliance

Zooniverse advertises itself as "the world's largest and most popular platform for people-powered research." Its origins are with Galaxy Zoo, a project that drew on public participation ("Citizen Science") to identify patterns in images of galaxies. The initial project was led by Chris Lintott and Kevin Schawinski of Oxford University. That program, launched in 2007, met with immediate success and soon led to a generalized platform, "Zooniverse." Lintott has maintained a leadership role throughout the dramatic expansion of Zooniverse, which is now anchored at the Adler Planetarium in Chicago as well as at Oxford University. As of this writing, Zooniverse has 78 separate projects of data collection and analysis, displayed in 9 disciplinary categories. Individual projects are listed in more than one discipline, but the rough totals are: 12 projects in Physics and Space, 41 projects in Nature, Biology, and Climate, 3 projects in Medicine, 7 projects in Social Science and History; and 6 projects in Arts, Language and Literature. (There is a further overlap of the projects in social sciences and humanities.)

The opening page of the site leads viewers readily into a list of the 9 disciplines and then to the projects within each discipline. Thus, one finds that the "Snapshot Grumeti" project, classifying camera-trap images from the Singita Grumeti Game Reserve in Tanzania, has 547 volunteers who have completed 28,444 classifications of 23,816 subjects, resulting in 200

completed subjects. Viewers may join, get video introduction to each project, and then get started on work, aided by tutorials.

This review focuses on Zooniverse projects in History, Social Sciences, and Humanities. They are relatively small in number, but they benefit from Zooniverse staff support and certain of them have become very successful. Those to be discussed briefly in this review include "League of Nations in the digital age," "Fossil Finder," "Old Weather," "Scribes of the Cairo Geniza," "Plastic Tide," and "Shakespeare's World." These projects address widely varying topics within history, social science, and humanities, with time frames ranging from the immediate present to early human evolution.

"The League of Nations in the Digital Age" is a project of archival scanning and documentation. Its purpose is to create a full digital version of the League of Nations archives (1919 – 1946). Overseeing this project is the Geneva-Tsinghua Initiative and the Citizen Cyberlab of CERN, the University of Geneva, and the United Nations Institute for Training and Research (UNITAR), in association with the UN Sustainable Development Goals (SDG). So far, only a small portion of League of Nations records have been digitized; of those 750,000 digitized pages, roughly one third are now online. The project relies on Optical Character Recognition for digitization but uses crowdsourcing and artificial intelligence to index the documents. That is, the work of Citizen Scholars is necessary to make the observations that will enable automatic indexing of the scanned pages. The project benefits from a second global institutional nexus: not only are the League of Nations archives held in Geneva, in close association with the United Nations Library and the University of Geneva, but the CERN project, the world's leading subatomic particle accelerator, is also in Geneva and is sharing its high-level computing capacity with the project.

"Fossil Finder" is a project in archaeology and paleontology. Based at the University of Bradford, it conducts aerial scans of shores of Lake Turkana in northern Kenya. The Lake Turkana region is rich in deposits of human and animal fossils, tracing back to 6 million years ago. It is in this region that human tools have been dated back to 3.3 million years ago. Project researchers have created numerous aerial images of the land, using drones, and Citizen Scientists are to explore the images in search of sites that are likely to contain additional human fossils. The project is funded by the Arts and Humanities Research Council of the United Kingdom.

The "Old Weather" project is to document weather observations from the Arctic, mostly from the nineteenth century. It asks Citizen Scientists to transcribe and document ships' logs from voyages to the Arctic, collecting information on weather patterns. A parallel project within Old Weather focuses on ships' logs from whaling voyages, seeking especially to document descriptions of sea ice.

The "Plastic Tide" is to document environmental degradation brought by plastic refuse along ocean shores. It traces the location of plastic refuse on the beaches of Great Britain. The project is based on the concern that, within the past ten years, the volume of plastics on British beaches has increased by some 250%; further, it is feared that the plastic located so far is perhaps 1% of the total that has reached the sea. Project researchers send drones to photograph beaches and ask Citizen Scientists to identify patterns in the distribution of plastic, with the objective of developing an algorithm for automatic analysis of the images. The UK project is designed as a trial analysis for a larger-scale project. In addition, the Plastic Tide group sponsors "beach-cleans," in which volunteers collect plastic waste from selected beaches.

"Scribes of the Cairo Geniza" is a project in social history from the tenth to thirteenth centuries. It is to classify manuscript texts in Hebrew and Arabic, held in a Cairo Geniza or storeroom where literate Jews deposited all texts that made reference to God. These records, which are being used increasingly as a source for social history of Cairo and all the regions with which Cairo was in contact, are being classified and digitized, and may thus be rendered even more valuable as a result of this Citizen Science work.

Finally, "Shakespeare's World," a project in transcription of English-language manuscripts of the sixteenth and seventeenth centuries, has achieved striking success. The project, based at the Folger Library in Washington, DC, grew out of a project on Early Modern Manuscripts Online (EMMO), which conducted "Transcribe-a-thon" workshops beginning in 2014. At the end of 2015, the group launched "Shakespeare's World" on Zooniverse, and soon found that users in multiple countries were creating finding aids and editing metadata as well as transcribing texts. In one discovery of the project, the term "white lie" was observed in a 1567 letter, at a time two hundred years earlier than the date reported for the term by the Oxford English Dictionary.

The range of activities and strategies pursued in Zooniverse is so wide—even within the history/social science/humanities disciplines—that it is difficult to pick out the most important new directions. But an interesting research tactic has been to rely on photographic information from drones, having Citizen Scientists identify patterns on the images, and using their observations to develop machine-learning algorithms so that the process can be automated. This technique was pioneered in Chris Lintott's "Sky at Night" project, which works with satellite images of the earth, and is used in the "Fossil Finder" and "Plastic Tide" sprojects just described.

Zooniverse's dual institutional base, at Oxford University and the Adler Planetarium, is backed by a robust organizational structure. It is owned by the Citizen Science Alliance, which in turn is governed by a board of 7 institutional representatives. These are the Adler Planetarium, the Johns Hopkins University, and the University of Minnesota in the U.S.; the National Maritime Museum, the University of Nottingham and Oxford University in the United Kingdom; and

Vizzuality. The latter is a development firm, based in Madrid and Cambridge, that has operated the Zooniverse portal and describes itself as a science and technology company focusing on sustainable development.

Zooniverse is already quite large as a research enterprise, though it is clearly just a tiny portion of what it could become. For readers of JWSR, it will be worthwhile to explore Zooniverse and its projects, to see how this resource could meet the needs of researchers in world systems, world history, and historical sociology. It is a platform that provides structure and guidance yet allows for great variety in the topics and methodologies with which it is associated. Zooniverse is not, however, a one-stop-shopping service for online research. It is clear that, of the Zooniverse projects that have experienced great success, each online project is closely tied to a network of researchers with common interests, in which substantial investment and innovative project design have strengthened those networks.

Patrick Manning
Professor of History, Emeritus
University of Pittsburgh
pmanning@pitt.edu
http://www.manning.pitt.edu