1-1-2015

Identifying Opportunities in Citizen Science for Academic Libraries

Cynthia Cohen
Claremont University Consortium

Liz Cheney
University of California - Los Angeles

Khue Duong
California State University - Long Beach

Ben Lea
University of Southern California

Zoe Pettway Unno
California State University - Fullerton

Recommended Citation

This Article is brought to you for free and open access by the Library Publications at Scholarship @ Claremont. It has been accepted for inclusion in Library Staff Publications and Research by an authorized administrator of Scholarship @ Claremont. For more information, please contact scholarship@cuc.claremont.edu.
Identifying Opportunities in Citizen Science for Academic Libraries

Cynthia M. Cohen
STEM Librarian; Reference and Research Services Coordinator
Claremont Colleges
Claremont, California
cynthia_cohen@cuc.claremont.edu

Liz Cheney
Science and Engineering Librarian
University of California, Los Angeles
Los Angeles, California
echeneyl@library.ucla.edu

Khue Duong
Science Librarian
California State University, Long Beach
Long Beach, California
khue.duong@csulb.edu

Ben Lea
Science and Engineering Librarian
University of Southern California
Los Angeles, California
blea@usc.edu

Zoe Pettway Unno
Science Librarian
California State University, Fullerton
Fullerton, California
zpettwayunno@fullerton.edu

Abstract

Citizen science projects continue to grow in popularity, providing
opportunities for non-expert volunteers to contribute to and become personally invested in rigorous scientific research. Academic libraries, aiming to promote and provide tools and resources to master scientific and information literacy, can support these efforts. While few examples currently exist of academic libraries involved in citizen science, this article identifies potential roles in community building; data curation; scholarly communication reform; and provision of space, technology, and resources.

**Introduction**

This article has its roots in two programs organized by the Science & Engineering Academic Librarians-South interest group of California Academic & Research Libraries (CARL/SEAL-S), with which the authors are all affiliated. Through these programs, we introduced attendees to citizen science—a popular topic in the sciences that many librarians are still unfamiliar with—and investigated how academic libraries are already involved, or could become involved, in citizen science. This article summarizes, provides context for, and expands upon these prior educational and exploratory efforts.

**Overview of Citizen Science**

To give a sense of the scope of citizen science, this section outlines different types of citizen science efforts and enumerates project examples in a variety of disciplines and formats.

**Definition and History**

Citizen science, also called crowd-sourced science, is the practice of non-experts contributing to or engaging in scientific research. The Cornell Lab of Ornithology (which spearheads many citizen science efforts and acts as a national leader in this field) defines citizen science as "projects in which volunteers partner with scientists to answer real-world questions" (Defining Citizen Science 2015). These projects might include collecting or analyzing data, observing and recording natural phenomena, or developing software or instrumentation. Citizen science projects promote public engagement with science and bolster researchers' own efforts by providing more volunteers and/or computing power.

Citizen science as a partnership between professional scientists and lay volunteers is a relatively recent phenomenon. Early scientific efforts required no specific training or qualifications and rarely took the form of paid positions; instead, investigations were often carried out by individuals in possession of other sources of income and sufficient leisure time (Haklay 2013). As scientists developed into a professional class beginning in the late nineteenth century, certain fields such as archaeology, ornithology, and astronomy made heavy use of non-professional
volunteers or enthusiasts (Haklay 2013; Silvertown 2009). More recently, people in industrialized countries have enjoyed shorter working hours, more leisure time, and an increase in basic scientific education; this gradual shift has contributed to the profusion of volunteers for citizen science (Haklay 2013).

**Scope of Citizen Science Activities**

Citizen science programs cover a broad spectrum of academic disciplines and vary widely in what is required of participants in terms of time commitment, travel, and use of technology. The citizen science community can therefore accommodate volunteers with different interests, skill sets, and amounts of free time. Consequently, one benefit of citizen science is its openness to people of all ages and walks of life. Certain projects involving simple counting or identification can make use of school-age (or even preschool-age) children as volunteers. Other projects rely on older children, teenagers, or adults but do not necessarily require subject expertise. Some programs provide orientations, training, or reference materials to ensure that participants are providing reliable data.

There is some debate in the citizen science community about whether projects relying on simple categorization or data analysis fall within the realm of citizen science. In early 2013, Open University faculty member and researcher Adam Stevens argued that such projects are more akin to "data crunching" than the difficult work of interpretation that makes up true science (Stevens 2013). Others, such as Caren Cooper, a Research Associate at the Cornell Lab of Ornithology, assert that such a narrow definition of citizen science overlooks the inclusive, collaborative nature of this type of research (Cooper 2013). In this article, we take an expansive view of citizen science in order to give readers a feeling for the breadth and variety of citizen science projects and related fields.

Given this broad definition, we can classify projects into active and passive categories. Active projects include those in which participants gather information; for instance, they might make observations, gather data that are later analyzed by trained researchers, or help design software or instrumentation. For passive projects, volunteers contribute by offering physical space or computing power to an investigation. This might involve installing software or connecting a sensor to their personal computers, or placing traps, sensors, or other types of instrumentation in their yards.

Data for citizen science projects can be gathered indoors or outdoors, thanks to the great variety of citizen science research and the near-ubiquity of mobile technology. While many projects rely on volunteers gathering data out of doors or traveling to a specific location to take part in research, others ask participants to play games or complete tasks through a web interface or using their mobile phones. This creates a low barrier to entry (Newman et al. 2012) as well as increased visibility and outreach for researchers.
Web-based projects are sometimes called "citizen cyberscience," a term coined by François Grey in 2009. Citizen cyberscience projects can be further divided into three categories, as described by Muki Haklay (2013) in *Crowdsourcing Geographic Knowledge*: volunteered computing, volunteered thinking, and participatory sensing. Examples of these subcategories are given in the following section.

**Project Examples and Outcomes**

Numerous citizen science projects make use of volunteers nationwide to gather data. One of the best-known and longest-running projects is the Great Backyard Bird Count (GBBC), sponsored by the National Audubon Society and the Cornell Lab of Ornithology. The GBBC began in 1998 and has had over 100,000 volunteers over the years. During this annual event, participants record bird observations online and contribute to scientific knowledge about the number and distribution of birds. The 2014 GBBC highlighted the effects of the polar vortex on bird populations in the United States (Great Backyard Bird Count 2015).

Another example of a large, global citizen science effort is iNaturalist, which functions as a platform for a large number of citizen science projects and hosts a web presence and mobile app for both iPhone and Android devices. The platform allows users to record observations of plants and wildlife in order to monitor changes in biodiversity by using a common interface to participate in a variety of different projects, vastly increasing capacity to collect environmental science and ecology data in widely dispersed areas difficult for individual researchers to visit on their own. iNaturalist was founded by then-students at the UC Berkeley School of Information (iNaturalist.org 2014); over the years, it has partnered with a number of organizations, including the California Academy of Sciences and the Global Biodiversity Information Facility, in order to maximize its outreach. Projects that make use of iNaturalist's mobile interface are an example of participatory sensing activities, in which citizen science volunteers make use of sensors (including mobile phones) to gather information (Haklay 2013).

One international project in the physical sciences is SETI@Home. The first example of volunteered computing, in which participants are asked to download software or attach sensors to their computers (Haklay 2013), SETI@Home began in 1999 and is hosted by the Space Sciences Lab at UC Berkeley. Participants can download software that will monitor incoming radio telescope data from the Arecibo Observatory in Puerto Rico. The software is designed to look for narrow bandwidth signals which don't occur naturally. Today, the SETI@Home project has over 1.5 million individual users (BOINCstats 2015). In 2011, the data from SETI@Home enabled researchers to identify "125 million potential detections" of non-naturally occurring radio signals (Korpela et al. 2011).

In contrast to high-profile citizen science projects focusing on the physical and life sciences, Games With Words employs non-expert volunteers in a "web-based
research laboratory" focusing on language. Founded by a graduate student at Harvard University's Psychology Department, Games With Words is staffed by graduate students and researchers from across the United States (GamesWithWords.org n.d.). Projects include games such as "Ignore That" (which tests how well users can answer questions when potentially distracted by irrelevant information) and "VerbCorner" (which investigates the meaning people apply to various verbs in the context of a fanciful story). Games With Words illustrates the final category of citizen cyberscience, volunteered thinking, in which participants are asked to classify images or information or otherwise use their analytical skills to answer questions (Haklay 2013).

By describing this small number of the citizen science projects in existence, we hope to have provided a sense of the diverse opportunities in citizen science for prospective volunteers or facilitators. Information on other citizen science projects can be found through local hubs (e.g., http://www.nhm.org/nature/citizen-science for Southern California or http://www.calacademy.org/science/citizen_science/ for Northern California), as well as broader resources such as SciStarter.

**Ensuring the Mathematical Accuracy of Large-Scale Citizen Science Projects**

While citizen science projects take many forms, one underlying concern is ensuring that lay volunteers gather high-quality data. Scientists who have never employed a citizen science research methodology may perceive citizen science projects to not be "scientific" enough due to the level of training the citizen data collectors receive. Since faulty data collection methods can lead to faulty data, there may be an assumption that the results generated by these programs are unreliable. However, the Law of Large Numbers suggests that those concerns are mitigated. In any scientific measurement or observation, unlikely outcomes are called "outliers." The Law of Large Numbers states that if many measurements are taken, one outlier will not significantly influence the overall outcome of the investigation. An outlier is also more likely to be balanced by another outlier at the opposite extreme. (For a more rigorous definition and proof, see *A Primer in Probability* (2nd ed.) by Kathleen Subrahmaniam (1990, p. 232). Many other statistics texts offer derivations.)

Until recently, collecting thousands or millions of results was of limited value because the computing power to process that much data was not readily available on a scientist's desktop. As an alternative to waiting until supercomputer time was available, researchers saw statistical sampling as an efficient way to work around that shortcoming. While one takes for granted that a random sample should have the same general properties as the whole of the data does (assuming the sample is truly representative), this very assumption is the main source of error when sampling: The 2% of the data analyzed in the study may not be accurately
representative of the 98% of the data that weren't included. The 1948 U.S. presidential election provides a well-known example: When many pollsters assumed that the 15% of voters still undecided would either not vote or break down into the same percentages as the 85% who did express an opinion, newspapers published the inaccurate headline "Dewey Defeats Truman" (Ross 1968, p. 251).

While analyzing all the collected data eliminates that source of error entirely, invalid data collection methods become more of a concern. To allow for this, many projects (including some featured at the CARL/SEAL-S Fall program, detailed below) have ways of filtering out the "bad" data (i.e., ways to build quality control into project design), resulting in numerous scholarly publications based on data collected by citizens (typically in collaboration with professional scientists) that have passed the scrutiny of the peer-review process (Bonney et al. 2014). Even in situations where data perceived to be flawed are not filtered out of the resultant dataset, the Law of Large Numbers shows that the data can still be usable.

Consider the following thought experiment: A coin is flipped a billion times to determine whether it is evenly weighted. Using a statistical sampling approach, one might observe 20 randomly selected flips and accurately record those results; as luck has it, 15 of those 20 flips landed tails. This particular outcome should occur randomly about 1.5% of the time. This statistical analysis concludes that the coin must be unevenly weighted. Using a more "big data" approach, one can observe and record every flip, but assume one misreads one's handwriting 5% of the time. Even with those data points that might be wrong due to error, one might still observe that the coin landed heads about 50% of the time. This statistical analysis concludes that the coin probably is evenly weighted.

Intuitively, we know that the latter analysis is the correct one--which is why large-scale citizen science projects can generate meaningful results even with non-experts gathering the information. The Law of Large Numbers allows the number of data points collected to outweigh the possibility that some of those data are invalid.

**Academic Libraries and Citizen Science Projects**

Citizen science projects hone science literacy skills in observing, inferring, graphing, predicting, and making sense of the collected data. These outcomes align with the mission of libraries in promoting and providing tools and resources to master scientific and information literacy. By connecting researchers with eager citizen scientists, the library becomes a conduit for citizen science projects as well as an intellectual hub: a place to consume scientific information as well as to create and engage in scientific endeavors. Promoting citizen science highlights
the idea that students, armed with basic science concepts, can engage in research and extend their intellectual curiosity outside of the classroom. This section presents two examples of existing connections between citizen science and academic librarianship.

At the 2013 Special Libraries Association All-Sciences Poster Session (Special Libraries Association Biomedical and Life Sciences Division 2013), Allison Scripa and Purdom Lindblad presented a model of library involvement with citizen science projects in a learning-community residence hall at Virginia Tech. As part of the Citizen Science Challenge, residents were divided into houses competing to earn the Citizen Science Cup (highest-scoring team) and tours of special labs and facilities on campus (all teams). Throughout the academic year, students engaged in meteorology, ornithology, and botany citizen science projects. With SatCam, students captured observations of sky and ground conditions as an earth-observation satellite passed overhead and submitted data to help improve satellite modeling. For the GBBC, mentioned above, students provided real-time observations of birds within the local community and contributed data (along with other volunteers across the continent) to help create a snapshot of birds in migration. The last project in the sequence, Project Budburst, used students' data to track individual plant species adaptation to climate change (Lindblad & Scripa 2012). The participation fostered a positive attitude towards scientific research and showed that students with minimal scientific background can engage in research projects in the early stage of their undergraduate studies.

Another example comes from the oldAstronomy project, which is using a citizen science platform (Zooniverse) to turn a well-established literature database (Astrophysics Data System (ADS)) into a data source, by connecting the ongoing Astrophysics Data System All-Sky Survey (ADSASS) with citizen science (Goodman 2013). Project output includes an all-sky "heatmap" of the density of published articles on the sky and an image layer where astronomical images previously available only as figures within ADS-held articles are now viewable, in context, on the sky (Goodman 2013). Managed by the Harvard-Smithsonian Center for Astrophysics, ADS contains over 11.2 million online records for publications in astronomy and astrophysics, physics, and the arXiv e-prints (Smithsonian Astrophysical Observatory/National Aeronautics and Space Administration n.d.). Using a technique like georeferencing (which assigns information to a specific time and place on earth), citizen scientists add value to ADSASS by "astroreferencing" images in the ADS full-text corpus, enabling "astrometric alignment of an image of the sky according to given coordinates, orientation, and pixel scale" (Pepe et al. 2011). Since automated computer astroreferencing successfully captures only 1% of the tested images in the ADS system, the ADSASS project team has determined that human judgment and annotation provided by citizen scientists could do a much better job at classifying images of the sky and determining presence of sufficient metadata. Presented with images extracted from ADS articles, citizen scientists can identify "if an
image contains a sky field, if it has coordinates, and if its caption has references to coordinates, pixels scale, orientation, etc." (Goodman 2013). Beyond demonstrating the value crowdsourcing can add, this project further illustrates the increasing potential for academic libraries in stewardship of research data. Libraries can support the community-driven, human curation of crowd-sourced data that occurs through citizen science, by acting as potential "data quality hubs" (Giarlo 2013).

**Librarians as Citizen Scientists: Educating Ourselves and Identifying Opportunities**

In 2013-14, CARL/SEAL-S programming centered around citizen science, with the goals of educating academic librarians about this popular topic and further investigating existing and potential academic library involvement. Our Fall 2013 professional development program, jointly planned with and held at the Natural History Museum (NHM) of Los Angeles County, featured a panel discussion followed by group participation in an outdoor citizen science activity. We revisited the subject at the California Academic & Research Libraries (CARL) 2014 biennial conference in San Jose, California, reporting conclusions from our Fall program and providing opportunities to participate in several indoor, hands-on citizen science activities.

Representing a diversity of organizations, project goals, and STEM subject areas, our Fall-program panelists were Lila Higgins (NHM), Kimball Garrett (NHM), Robert de Groot (USC and Southern California Earthquake Center), Aydogan Ozcan (UCLA Henry Samueli School of Engineering and Applied Science and the California Nanosystems Institute), and Katherine Pease (Heal the Bay). Two dominant themes emerged from a discussion of how to design rigorous research projects for non-experts. First, panelists emphasized the importance of thinking from the non-expert's perspective: Successful projects often employ common knowledge that participants already have and are location-based, in places where people already are. Rewards and incentives are also important motivators for non-expert participation. Second, panelists described several ways to build quality control into project design: For example, computer games can be designed to become more accurate as more people play. Alternatively, projects may rely on scientists to train citizens and/or check citizen work.

At NHM and CARL 2014, we explored different ways to be a citizen scientist. Outdoors, in our Fall citizen science activity, participants used iNaturalist to contribute to a species inventory for the NHM Nature Gardens. Because iNaturalist provides a common interface for a variety of different projects, this activity afforded participants a simple skill set that applies to other NHM projects, including the L.A. Nature Map, to which they were able to continue contributing information long after the day of our program. This extended connection--and the subsequent online collaboration to identify the species we'd
observed—affirmed that citizen science is about more than data collection: Participation can also build community, and change the way we look at the world around us.

At CARL 2014, participants tried three indoor citizen science projects: Galaxy Zoo, Happy Moths (via Citizen Sort), and Ignore That! (via Games With Words). When several participants reported feeling uncertain (or even worried) about their ability to contribute sufficiently rigorous data to the hands-on activity, it was suggested that this might be analogous to how students feel when doing literature searches using library resources. It was also noted that the inclusion of a social-science project (Ignore That!) inspired ideas for citizen science applications to library and information science research.

**Future Directions: Potential Roles for Academic Libraries**

Through discussion with panelists and participants at our 2013-14 programs (Cheney et al. 2014; Higgins et al. 2013), we have identified several potential roles academic libraries could play in citizen science. Academic librarians could facilitate collaboration by making use of established relationships with faculty, staff, students, and others in the community to connect citizen scientists and researchers. Additionally, they could provide opportunities for students to refine their career objectives via hands-on research in their potential fields of study. Academic libraries could provide data curation services, internet access, technology training, and space for makerspaces and other citizen science work (Cheney et al., 2014; Higgins et al., 2013). The importance of outreach and communication—through formal publications as well as informal storytelling—creates further opportunities in data visualization, e-book production, and the capture and preservation of oral histories (Menninger 2014).

Because many citizen scientists lack the access to research materials provided by academic institutions, citizen science has the potential to impact scholarly communications and encourage open access publishing. During the research process (to the extent licensing agreements would allow), it may be possible for academic libraries to provide broader access to the published scientific literature. Enabling participants to contribute to that body of literature, through citizen science projects, enhances their learning experience and engenders a sense of pride (Dickinson & Bonney 2012, p. 22). Because citizen science projects rely heavily on the general public, researchers often feel obliged to make their publications available to readers without institutional privilege (Menninger 2014), resulting in a frequent preference for open access journals. Current efforts to create a new journal focusing on the study of citizen science as a practice (rather than on the results of citizen science projects) have established full open access as a priority (Citizen Science Association 2013).
Conclusion

By providing a way for non-experts to become personally invested in scientific research and contribute to the creation of new scientific knowledge, citizen science allows participants to feel they can make a difference in big issues like biodiversity, habitat preservation, and global health--crucial, overwhelming issues that often leave untrained individuals feeling powerless. By empowering people (including previously unrepresented constituencies) to help solve such problems using evidence-based solutions, and by building a much broader, personally invested community around such efforts, citizen science can generate social benefits in addition to providing researchers a means to collect or process large amounts of data. Increased access to, engagement with, and participation in scientific research aligns naturally with work already being done in academic libraries, as many services academic libraries currently provide could have specific applications to citizen science projects.

Acknowledgements

The authors would like to acknowledge our colleagues at the Natural History Museum (NHM) of Los Angeles County: Richard Smart (Citizen Science Program Coordinator), Lila Higgins (Citizen Science Manager), and Richard Hulser (Chief Librarian), who were instrumental in our event planning and have greatly enriched our understanding of citizen science.

References


Cooper, C. 2013. Pearls across the Zooniverse: When crowdsourcing becomes citizen science. Guest Blog on Scientific American. [Internet]. Available from:


Identifying Opportunities in Citizen Science for Academic Libraries


Menninger, H. 2014. Climate at hyper-local scales: Engaging the public to study the ecology of our homes and bodies. Presented at: Science Boot Camp for Librarians Southeast; Raleigh, NC.


Stevens, A. 2013. Citizen science. Echus overlook [Internet]. Available from: http://echusoverlook.net/citizen-science/
