The Power of Engaging Citizen Scientists for Scientific Progress

Jeanne Garbarino¹ and Christopher E. Mason²,³,⁴*

¹Science Outreach Program, The Rockefeller University, New York, NY 10065,
²Department of Physiology and Biophysics, Weill Cornell Medical College, New York, NY 10021,
³The HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine,
⁴Weill Cornell Medical College, New York, NY 10021,

Citizen science has become a powerful force for scientific inquiry, providing researchers with access to a vast array of data points while connecting nonscientists to the authentic process of science. This citizen-researcher relationship creates an incredible synergy, allowing for the creation, execution, and analysis of research projects that would otherwise prove impossible in traditional research settings, namely due to the scope of needed human or financial resources (or both). However, citizen-science projects are not without their challenges. For instance, as projects are scaled up, there is concern regarding the rigor and usability of data collected by citizens who are not formally trained in research science. While these concerns are legitimate, we have seen examples of highly successful citizen-science projects from multiple scientific disciplines that have enhanced our collective understanding of science, such as how RNA molecules fold or determining the microbial metagenomic snapshot of an entire public transportation system. These and other emerging citizen-science projects show how improved protocols for reliable, large-scale science can realize both an improvement of scientific understanding for the general public and novel views of the world around us.

INTRODUCTION

The practice and scope of science, and of the scientist, are constantly evolving. With the emergence of new tools and techniques that can rapidly connect thousands, or tens of thousands, of citizen scientists comes an unprecedented ability to answer large-scale scientific questions. However, these opportunities challenge the traditional scientific research model, whereby clearly defined research questions and hypotheses are implemented by a small, highly trained team of researchers at an academic, private, or government institution. Indeed, a vast phalanx of citizen scientists can readily surpass several previous limitations of science, including the amount information that can be collected and the speed at which it can be processed. Particularly in scientific fields that require large datasets for scientific progress, some researchers are looking beyond the traditional scientific team into the realm of citizen science.

The concept of citizen science, or public participation in scientific research, is not new. For example, the National Weather Service launched the still-ongoing “Cooperative Observer Program” in 1890 to help understand national meteorological changes in real time, and to aid long-term climate studies (24). Similarly, the National Audubon Society began their “Christmas Bird Count” in 1900 to help raise awareness about the growing need for ecological conservation (13). Now in its 116th year, the Christmas Bird Count has helped scientists to understand a variety of aspects of bird ecology and natural history (22, 23, 27). However, these early projects revealed some of the challenges of large-scale citizen science, the importance of consistent training, data uniformity, and data reproducibility for efficacious use of the collected data. For example, in the context of the bird count, data had been collected on the number of deceased birds found along roadsides without providing a comparison with the total number of birds estimated to be in that geographical region, making it difficult to accurately calculate the health or death rates of specific bird populations. These issues have since been addressed, and new models have been implemented to provide more accurate measurements of bird ecology and population health (27).

Regardless of their limitations, citizen-science projects have transformed the practice of science by democratizing access to scientific methods, encouraging scientific literacy, and engaging the next generation of scientists early in their
life. For over a century, researchers have been finding new and more efficient ways to solve relevant scientific problems by tapping into the vast pool of science enthusiasts from the general public. Historically, these citizen-science projects have been primarily associated with making skilled observations, particularly in archeology, ecology, and astronomy. This is most likely due to the accessibility of these scientific disciplines. For instance, citizen-science projects in astronomy might simply involve going out at night to observe stars—an activity that does not require sophisticated equipment or a specialized environment (such as a laboratory). However, since the early 2000s, the citizen-science movement has been greatly enhanced by “the Internet of Things,” wherein internet-connected devices and mobile software platforms have become ubiquitous, especially for smartphone applications. Project information dissemination, detailed protocols for participation, and data collection methods are therefore available now to the majority of the world’s population, giving rise to more sophisticated and democratized roles for citizen scientists, and a more effective strategy for completing large research projects (28). This has opened up avenues for citizen-science projects in other disciplines, such as biology, bioinformatics, and microbiology, which have been historically inaccessible to the average citizen due to resource limitations.

THE POWER AND REACH OF CITIZEN SCIENCE

Several striking examples have recently been published that show the power of utilizing modern-day citizen science to provide meaningful contributions to scientific progress. One of the earliest and largest success stories from such work by the general public was FoldIt—the protein folding prediction game from the University of Washington. In 2010, an analysis of FoldIt showed that the game’s 57,000 players outperformed advanced computer algorithms for predicting protein structure (7) and, in the following year, contributed to solving the crystal structure of the Mason-Pfizer monkey virus (M-PMV) retroviral protease (15). Similarly, over 100,000 “citizen neuroscientists” from over 130 countries are playing the citizen science game EyeWire, significantly accelerating the work of traditional scientists in their quest to map the mouse retinal connectome (16).

Recording observations and playing games over the Internet are seemingly natural fits for engaging large populations of nonscientists for the purposes of scientific progress. However, as researchers realize the potential inherent in the citizen-science movement, this approach is being combined with emerging technologies and collection devices to create new opportunities. Specifically, there has been a large increase in the number of citizens engaged in studies regarding the microbiome (the ecosystem of microbes) or metagenome (the collection of genomes from a sample). Driven in part by the steady drop in next-generation sequencing costs (21), along with the increasing public interest in the “quantified self” and personalized medicine movements (19), which aim to provide individuals with a deeper understanding of their own biology, the general public has shown a great curiosity to learn about the microorganisms living in, on, and around them—to the delight of many scientists conducting research in this field. For most studies related to the microbiome, we are just starting to catalog the molecular basis of our environments (11). Widespread participation enables us to more quickly and efficiently survey various landscapes and systems in terms of microbial organisms, which is essential if we are to truly understand the implications of specific microbial profiles on or in a given sample.

The desire to learn more about the elusive human microbiome has spurred the creation of many services that provide both individualized microbiome reports and population- and city-scale microbiome measurements. For instance, American Gut, which is part of the Human Food Project, is a service that primarily focuses on providing participants with information on their intestinal microflora. Similarly, uBiome is a personalized microbiome service that can assess the gut, nose, mouth, skin, and genital microbiome. In addition to providing individualized reports to users, who typically send their samples to these services through the mail (as per user kit instructions), both American Gut and uBiome are using the collective datasets to more fully understand the impact of the human microbiome on health. Academic groups have also used citizen scientists to create reports on the microbial communities pertaining to humans, such as the microbiome of the belly button (14), and a survey of flu infections and outbreaks (12) for people when they begin to feel sick. Other projects include a focus on the mouth or intestinal microbiome, such as the Human Oral Microbiome Database (HOMD, 6) and MetaHIT (Metagenomics of the Human Intestinal Tract, 18), whereas still others focus on areas of the land, sea, and air as part of the Earth Microbiome Project (EMP, 11), a collection of microbes from extreme environments called the eXtreme Microbiome Project (XMP, 9), and the Metagenomics and Metadesign of Subways and Urban Biomes (MetaSUB, 9). Also, the work of the global Ocean Sampling Day (OSD, 17) enables a global effort to catalog the diversity of life across the world’s oceans. These and a wide variety of citizen-science projects can be found on the SciStarter website (http://scistarter.com/). Table 1 provides a summary of microbial citizen-science projects.

INCREASED SCOPE AND EXCHANGE OF SCIENTIFIC RESEARCH

Often, there are research projects or strategies that could provide a clear picture for the understanding of a specific scientific question, but scientists may not have the required manpower or funding (or both) to make it happen. However, by utilizing citizen science, scientists are gaining myriad research team members requiring little or no funding, who can collectively survey large geographical areas, analyze large datasets, or provide large sample numbers (8). As alluded to above, having widespread participation
in microbiome-themed citizen-science projects helps to greatly enhance the statistical power of data analysis and introduces nonscientists to more detailed information pertaining to microbiology.

For instance, with regard to the microbiome, there is a widely held assumption that all bacteria are bad, or there are misunderstandings related to results from microbiome experiments. An example of the disconnect between research results and the media interpretation of said results was revealed through the New York City (NYC) Pathomap project, which surveyed the microbiome of all 468 stations across NYC’s subway system (1). Even though the vast majority of bacteria found were harmless, the general media (newspapers, TV, online) focused primarily on sequences

### TABLE 1.
Summary of selected microbial-based citizen-science projects.

| Project | Description | Role of Citizen Scientist |
|---------|-------------|--------------------------|
| American Gut | Open-source, open-access research project within the Human Food Project initiative that aims to characterize human gut microflora through wide-scale participation. | Participants purchase a kit and send personal samples through the mail. A personalized report is generated, comparing an individual’s microflora with that of a larger population. |
| uBiome | Personal microbiome service provider that characterizes microbial ecosystems on mouth, ears, nose, gut, and genitals, with an aim to characterize microbiomes at the population level. | Participants purchase a kit, fill out a health survey, and send samples through the mail. Again, a personalized report is generated, and users are free to compare their data with population data through a web interface. |
| Belly Button Biodiversity | First citizen-science project to study a human microbiome, aimed at characterizing the specific microbial ecosystem present in our navels. | Samples from participants were collected at in-person events in the Raleigh-Durham region (NC), or through the mail, and were cultured on plates, as well as sequenced. Participants could see collective culture data, and browse through a database of microbial species. |
| GoViral | New York University-based project to study how community viral symptom information can be used to understand an individual’s risk of disease. | Participants can enroll in this study free of charge, and upon feeling sick, they will contribute a nasal, saliva, or fecal sample for analysis. In addition to receiving a personalized report, participants can access population data through the Open Humans ecosystem, www.openhumans.org. |
| Human Oral Microbiome Database (HOMD) | The goal of this project is to characterize the microbial populations found in the human oral cavity. | HOMD is an open access global database with up-to-date user statistics and microbial classifications. Sourcing oral microbial sequences from the Human Microbiome Project, citizen scientists can perform search queries and data analysis using tools inherent to the website. |
| Earth Microbiome Project (EMP) | The EMP aims to systematically characterize microbial biodiversity from around the world. | The open access format of the EMP allows for a variety of entry points for citizen science activities, including sample contribution and data analysis. |
| Metagenomics and Metadesign of Subways and Urban Biomes (MetaSUB) | MetaSUB is a global collaborative project aimed at characterizing the molecular portrait of public transportation systems. | Citizen scientists are provided with a detailed protocol and a kit to sample surfaces in a subway or equivalent. These samples are mailed to the appropriate laboratory, where they are sequenced. Participants can view collective data via a web interface. |
that mapped to potential pathogens. However, this was also a learning opportunity for the Pathomap team, highlighting the importance of transparent and thorough explanations for citizen-science projects. While Pathomap participants were provided with necessary information about the precise experimental protocol and instructions for the real-time mobile app that recorded each part of the process of data and sample collection, it was important to also communicate the difference between identifying a match between a DNA fragment in a sample and a known sequence versus finding actual living microbial organisms, including the publication of an update to the original paper (2). By having an open channel for public engagement via citizen science, including Twitter, blogs, and online forums (25), researchers can improve communication on the current understanding of complicated project data and help dispel fear related to misinterpretation of data and/or the misunderstanding of the research process. (20, 30).

**ADDING VALUE TO THE PUBLIC UNDERSTANDING OF SCIENCE**

In a 2015 study conducted by the Pew Research Center, it was reported that the public’s limited knowledge in science, technology, engineering, and math (STEM) is a problem for scientific progress. Furthermore, it was revealed that scientists and the general public see scientific issues through different lenses, creating a gap—sometimes a chasm—between the scientific consensus and public belief (10). While the reasons for these results are many and complex, we suggest that citizen science is a valid mechanism to help rectify shortcomings in the public’s understanding of science.

For example, participation in citizen-science projects can introduce the framework for authentic science research practices. As there seems to be a disconnect between the actual pace and the public’s perceived pace of scientific research, a well-designed citizen-science project can provide clarity to nonscientists when it comes to the general nature of science research, including concepts that contribute to a sound experimental design (such as the inclusion of control groups, sample size and representation, and appropriate statistical analysis), and the nature of science as a process of sifting through types of evidence. This is particularly beneficial in providing the tools to discern between “good” and “bad” science, and possibly contribute to enhanced critical thinking, which is favorable regardless of the subject matter (29).

Evaluations on a variety of citizen-science projects have suggested a positive impact on participants’ awareness of specific scientific issues and their content-knowledge gains, as well as improved skills related to scientific inquiry and critical thinking (3). Additionally, citizen-science projects have a large potential for making positive impacts when it comes to attitudes toward science. For instance, citizen-science opportunities can provide a mechanism for participants to become invested in science research in a way that is familiar to them (either through providing personal samples or by sampling areas in their neighborhood), which can be empowering. Furthermore, by encouraging inclusivity and openness, citizen science can break down the fear about or perceived distance from science, making science more accessible. This can be particularly important when it comes to efforts to improve diversity in science, since the message inherent to citizen science is that science is for everyone, regardless of personal, geographical, or socioeconomic background.

**CHALLENGES ASSOCIATED WITH CROWDSOURCING FOR SCIENCE RESEARCH**

While citizen-science projects have great potential for making positive impacts on both the science research enterprise and the general public, there are challenges related to crowdsourced research. The most obvious of these challenges is the varied skill level of citizen-science participants, which could lead to inconsistencies among data points. Therefore, it is essential that citizen-science project leaders provide clear directions and expectations to participants and provide some type of feedback mechanism, akin to “customer service,” should there be questions or the need to troubleshoot. This is often easy to do on the websites of crowdsourced projects and usually includes FAQs (frequently asked questions), video-based guides with step-by-step training, and contact personnel for logistical or technical questions. As projects increase in their size and scope, these steps of coordination and uniformity become more essential for the data to be useful long-term.

Another challenge related to citizen science is the effective communication of “big picture” ideas, and having citizen-science participants fully understand the scope and limitations of the research project. As previously discussed, this type of challenge was evident with the NYC Pathomap project (1), particularly when it came to the limitations of the data analysis—even when clearly delineated. Depending on the nature of the research project area, addressing this can require a lot of work, and even still, efforts may not fully yield positive impacts among all participants. Moreover, it is equally important that scientists communicate the ethical issues related to crowdsourcing research, especially when it pertains to human research (4, 5). The founders of uBiome reported examples of these issues, including the challenges associated with obtaining Institutional Review Board (IRB) approval, which helps to protect citizen-science participants when providing personal samples (26).

Lastly, one can argue that the greatest strength of citizen science can also be its greatest weakness. Because the science research is not entirely conducted by “experts,” some may discount the validity of the data or overall value of the project in promoting scientific progress. These issues can be addressed through proper training of citizen scientists; incorporating things like geo-tagging, time-stamping, photo documentation, and sample tracking into the data collection mechanism can also help normalize the process. For example, sample double-checking was performed as part of
the first project of MetaSUB (the NYC Pathomap), wherein all data were manually validated by geo-coordinate, photo, and time-stamp as having been accurately collected as reported. Since mobile phones today are almost all smartphones enabled with cameras, the Pathomap project was the first to be able to use photo-based sample tagged data redundancy to confirm sample provenance and accuracy (1), but this will likely be a continuing trend.

CONCLUSION

As citizen science becomes better integrated into the science research enterprise, it is important that we continue to determine and refine best practices. Furthermore, by engaging citizen scientists, researchers are helping to make real connections to both scientists and authentic scientific practice. Too often, we see a stereotypical portrayal of scientists painting a picture that scientists are unapproachable or unrepresentative of the greater population. These stereotypes can also hinder the public’s trust in the scientific process, subject matter, or results, and cloud judgment related to public health and policy, since the results are deemed too abstract or convoluted. This could, in turn, affect funding or other factors related to the progress of scientific research. Thus, the incorporation of citizen science into the research model can benefit scientists in many ways. In addition to the direct contributions pertaining to a specific research project, the inclusion of citizen science can help scientists dispel misinformation associated with particular scientific topics and provide a channel for more effective public engagement, thus improving the science as well as the utilization of the scientific results to address medical and social problems.

ACKNOWLEDGMENTS

For CEM we would like to thank the Sloan Foundation (2015-13964), the Bert L. and N. Kuggie Valle Foundation, the Irma T. Hirschl and Monique Weill-Caulier Charitable Trusts, the WorldQuant Foundation, the STARR Consortium (17-A765, 19-A9-071), support from the National Institutes of Health (R25EB020393, R01NS076465), and the Pershing Square Sohn Prize. The authors declare that there are no conflicts of interest.

REFERENCES

1. Afshinnekoo, E., et al. 2015. Geospatial resolution of human and bacterial diversity with city-scale metagenomics. Cell Systems 1(1):72–87.
2. Afshinnekoo, E., et al. 2015. Modern methods for delineating metagenomic complexity. Cell Systems 1(1):6–7.
3. Bonney, R., et al. 2009. Public participation in scientific research: defining the field and assessing its potential for informal science education. A CAISE inquiry group report. Center for Advancement of Informal Science Education (CAISE), Washington, DC.
4. Callaghan, C. W. 2015. Crowdsourced ‘R&D’ and medical research. Br. Med. Bull. 115(1):67–76.
5. Cameron, P., D. W. Corne, C. E. Mason, and J. Rosenfeld. 2013. Crowdfunding genomics and bioinformatics. Genome Biol. 14(9):134.
6. Chen, T., W. H. Yu, J. Izard, O. V. Baranova, A. Lakshmanan, and F. E. Dewhirst. 2010. The Human Oral Microbiome Database: a web accessible resource for investigating oral microbe taxonomic and genomic information. Database 2010:bau013.
7. Cooper, S., et al. 2010. Predicting protein structures with a multiplayer online game. Nature 466(7307):756–760.
8. Crain, R., C. Cooper, and J. L. Dickinson. 2014. Citizen science: a tool for integrating studies of human and natural systems. Annu. Rev. Environ. Res. 39(1):641–665.
9. Ehrenberg, R. 2015. Urban microbes come out of the shadows. Nature 522(7557):399–400.
10. Funk, C., and L. Rainie. 2015. Public and scientists’ views on science and society. Pew Research Centers Internet American Life Project.
11. Gilbert, J. A., J. K. Jansson, and R. Knight. 2014. The Earth microbiome project: successes and aspirations. BMC Biol. 12:69.
12. Go Viral. 2015. [Online.] www.goviralstudy.com/.
13. History of the Christmas Bird Count. (nd). [Online.] www.audubon.org/content/history-christmas-bird-count. National Audubon Society.
14. Hulcr, J., et al. 2012. A jungle in there: bacteria in belly buttons are highly diverse, but predictable. PLoS One 7(11):e47712.
15. Khatib, F., et al. 2011. Crystal structure of a monomeric retroviral protease solved by protein folding game players. Nat. Struct. Mol. Biol. 18(10):1175–1177.
16. Kim, J. S., et al. 2014. Space-time wiring specificity supports direction selectivity in the retina. Nature 509(7500):331–336.
17. Kopf, A., et al. 2015. The ocean sampling day consortium. Gigascience 4:27.
18. Li, J., et al., and MetaHIT Consortium. 2014. An integrated catalog of reference genes in the human gut microbiome. Nat. Biotechnol. 32(8):834–841.
19. Mason, C. E., S. Porter, and T. Smith. 2014. Characterizing multi-omic data in systems biology. Adv. Exp. Med. Biol. 799:15–38.
20. Mason, C. 2015. The long road from data to wisdom and from DNA to pathogen. 2015. [Online.] http://microbe.net/2015/02/17/the-long-road-from-data-to-wisdom-and-from-dna-to-pathogen/. MicrobeNet.
21. Metzker, M. L. 2010. Sequencing technologies – the next generation. Nat. Rev. Genet. 11(1):31–46.
22. Niven, D. K., J. R. Sauer, G. S. Butcher, and W. A. Link. 2004. Christmas Bird Count provides insights into population change in land birds that breed in the Boreal forest. Amer. Birds 58:10–20.
23. North American Bird Conservation Initiative, US Committee. 2009. The State of the Birds, United States of America. Washington, DC.
24. NWS Cooperative Observer Program. 2014. What is the coop program? [Online.] http://www.nws.noaa.gov/om/coop/what-is-coop.html. NOAA's National Weather Service.

25. Petit, R. A., M. Ezewudo, S. J. Joseph, and T. D. Read. 2015. Searching for anthrax in the New York City subway metagenome. Zenodo.

26. Richman, J., and Z. Apte. 2014. Crowdfunding and IRBs: the case of uBiome. Guest Blog – Scientific American Blog Network.

27. Sauer, J. R., W. A. Link, W. L. Kendall, J. R. Kelley, and D. K. Niven. 2008. A hierarchical model for estimating change in American woodcock populations. J. Wildlife Manage. 72(1):204–214.

28. Silvertown, J. 2009. A new dawn for citizen science. Trends Ecol. Evol. 24(9):467–471.

29. Stein, B., A. Haynes, M. Redding, T. Ennis, and M. Cecil. 2007. Assessing critical thinking in STEM and beyond, p 79–82 In Eskaner, M. (ed.) Innovations in e-learning instruction technology, assessment and engineering education. Springer, The Netherlands.

30. Yong, E. 2 October 2015. There’s no plague on the NYC subway. No platypuses either. [Online.] http://phenomena.nationalgeographic.com/2015/02/10/theres-no-plague-on-the-nyc-subway-no-platypuses-either/. National Geographic.