

A Sensor Network Approach to Managing Data Quality in Citizen Science

Andrea Wiggins

University of Maryland
4121G Hornbake, South Wing
College Park, MD 20742 USA
wiggins@umd.edu

Carl Lagoze

University of Michigan
4444 North Quad
Ann Arbor, MI 48109 USA
clagoze@umich.edu

Weng-Keen Wong

Oregon State University
1148 Kelley Engineering Ctr
Corvallis, OR 97331 USA
wong@eecs.oregonstate.edu

Steve Kelling

Cornell Lab of Ornithology
158 Sapsucker Woods Rd.
Ithaca, NY 14850 USA
stk2@cornell.edu

Citizen science is a type of human computation in which volunteers contribute to scientific research. Most citizen science volunteers take one of two roles. The first is *sensor* for which the primary contribution is recording observations of (natural) phenomena. The second is *processor*, which entails manipulation, classification, or information extraction from an artifact; e.g., an image. The nature of the role influences several aspects of project design, including data verification processes. In this paper we discuss our approach to addressing the data quality challenge of establishing the accuracy of human sensors at the tasks of detection and identification.

In processor tasks, participants classify or make decisions about an artifact, such as an image. These projects can take advantage of task repeatability to establish consensus as a primary approach to quality control, with multiple participants evaluating the same artifact. There are then multiple methods for obtaining reliable consensus such as voting, weighted voting, and dynamic assignment of artifacts to individuals. Other types of processor tasks include problem solving and puzzles that often rely on specific skills, such as spatial pattern matching. Many of these projects engage a large group of people in cognitive piecework; solving complicated tasks that are broken down into manageable segments that are easily accomplished and often game-like (Schmidt 2013). Problem-solving tasks may also have established criteria or lab-based methods for evaluating solutions for the given problem space. This is potentially advantageous because it can enable an algorithmic evaluation of contributions.

For many sensor-type citizen-science projects like eBird, a large citizen-science project that engages volunteers to submit observations of birds, there is no ground truth. Volunteers collect data in contexts (e.g., location, time, weather, etc.) that are not repeatable and do not necessarily gen-

erate artifacts like photos to document an observation. Bird identification is a challenging knowledge-based task engaging volunteers in what we call cognitive craftwork. It requires first detecting a bird, then seeing or hearing the bird clearly enough to identify the bird—either through memory or with identification aids—to the species level. eBird takes advantage of human cognitive abilities to identify birds, a task that automated sensors and computers cannot accomplish. With several straightforward data collection protocols, eBird is generating one of the largest and fastest growing biodiversity databases in existence.

Because there is no objective ground truth, however, eBird has a major data quality challenge—how do we know whether eBird observers made correct identifications of species? A novel approach to addressing variability in observers' data quality in citizen-science is to consider how data quality issues are handled for autonomous sensor networks (Rachlin, Negi, and Khosla 2011). There are multiple parallels between citizen science and sensor networks. Both consist of numerous individual actors that perform a limited task: humans following well-defined protocols for citizen science projects and transducers taking simple measurements for sensor networks. Both can be deployed at a relatively low cost and record the occurrence of phenomena across broad spatial extents and periods of time. Most importantly, both can collect large quantities of inherently noisy data.

Part of the data quality challenge for eBird (Royle 2004) and autonomous sensor networks (Ediriweera and Marshall 2010) is how to handle outliers or anomalies reported by individuals so as to distinguish between erroneous data and accurate, scientifically-relevant observations. Standard practice in sensor network design is to define a “typical” data output profile to represent the expected behavior of a sensor in the network (Rachlin, Negi, and Khosla 2011).

This profile describes the usual pattern of data output from nearby sensors where false responses are likely to be unrelated while correct responses are likely to be correlated (Zhang, Meratnia, and Havinga 2010).

In eBird, we take a sensor network approach to identify the expected data output profile of participants in a specific region. First, based on historic data and expert input we develop a specific list of species that are likely to occur for a particular region and date (Kelling et al. 2011). Currently, more than 2000 regional and temporal bird lists exist in eBird. Based on these lists, a contributor to eBird is presented with a spatiotemporally appropriate checklist with which to record observations. If a contributor wants to submit a species that is not on the checklist for a specific region or date, they must take an additional step to report an unexpected species, and the record is then flagged for review. Members of eBird's network of more than 800 regional experts contact those individuals who submitted flagged records to obtain additional information, such as field notes or photographs, in order to confirm unusual records. In 2013 more than 44 million records were submitted to eBird, almost 2.4 million (5%) were flagged for review, with just under half of the flagged records marked as invalid following review. All records, their flags, and their review history are retained in the eBird database.

In eBird and other citizen science projects in which volunteers act as intelligent sensors, data quality cannot be determined through comparison to an objective ground truth nor through consensus. Instead, eBird takes an approach used in autonomous sensor networks to establish the reliability of volunteer-generated observation data by evaluating the believability of the data based on prior evidence (such as data output profiles) and leveraging expert judgment to evaluate exceptions.

Acknowledgements

This work was funded by The Leon Levy Foundation, The Wolf Creek Foundation, The Seaver Institute, and the following U.S. National Science Foundation Awards: ABI Sustaining: DBI-1356308 and SOCS: IIS-120589.

References

- Ediriweera, D. and Marshall, I. 2010. Understanding and managing large sensor networks. *Drinking Water Engineering and Science Discussions*, 3:149-175.
- Kelling, S.; Yu, J.; Gerbracht, J.; and Wong, W. K. 2011. Emergent Filters: Automated Data Verification in a Large-scale Citizen Science Project. In Proceedings of the IEEE eScience 2011 Computing for Citizen Science Workshop.
- Rachlin, Y.; Negi, R.; and Khosla, P. K. 2011. The Sensing Capacity of Sensor Networks. *Information Theory, IEEE Transactions*, 57:1675-1691.

Royle, J. A. 2004. Modeling Abundance Index Data from Anuran Calling Surveys. *Conservation Biology*, 18:1378-1385.

Schmidt, F. A. (2013). The Good, The Bad and the Ugly: Why Crowdsourcing Needs Ethics. In Proceedings of the Third International Conference on Cloud and Green Computing (CGC), 531-535.

Zhang, Y.; Meratnia, N.; and Havinga, P. 2010. Outlier detection techniques for wireless sensor networks: A survey. *Communications Surveys & Tutorials*, 12:159-170.