TAG YOURSELF

SARA STOUDT

As environmental research budgets get slashed, can amateurs fill in the gaps?

Note: This article was originally publised in Logic Magazine's Nature issue.

Since 1992, volunteers have tagged more than 1.5 million monarch butterflies. Tagging a butterfly involves capturing it in a net, attaching a label to its wing, and releasing it back into its habitat. The identifying information on the tag goes into a database that tracks the monarchs famous migrations to and from California and central Mexico. Chip Taylor, founder and director of Monarch Watch, the nonprofit that organizes this vast volunteer effort, says the process is easier than it sounds and that the monarch butterfly is hardier than it appears. Even students can participate, and they often do so as part of their science classes, guided by teachers. Taylor noted that 2018 held the record for the largest number of butterfly tags distributed, with over 320,000 mailed to interested volunteers across North America.

Monarch butterflies are unique. Not only are they resilient to being captured and tagged by volunteers, they are also widely loved. In addition to Monarch Watch, there are numerous Facebook groups dedicated to the orange and black beauties, including Monarch Butterfly Garden with over 50,000 followers, and The Beautiful Monarch with over 25,000 members. Volunteers are eager to study and protect them.

What about species that arent as charming? Who will account for the creepy crawlies and the drab species? These are some of the questions underlying the scientific communitys increasing reliance on crowdsourced environmental data. Monarch Watch helped pioneer the crowdsourcing model, but its analog, low-tech approach has since been overtaken by a wave of apps and platforms that have made the process of collecting data much more accessible. The proliferation of smartphones and the rise of the mobile web has enabled more people to contribute more observations, on more species. And while this phenomenon has been growing for years, it has acquired a new urgency since the election of Donald Trump. Under Trump's leadership, scientists have seen a decrease in funding for environmental research and an official denial of climate change despite mounting evidence. This means that research scientists need crowdsourced data more than ever before, incomplete as it may be.

"Community scientists" can help. In addition to not relying on government budgets, these nature-loving, albeit untrained and unpaid, members of the public have another advantage: they can use apps to collect data about more species, over a larger physical area, than the comparatively small number of professional environmental scientists can. But if community scientists completely drive scientific research agendas, society risks losing valuable information about critically important species. Community science efforts can only augment scientific research, not replace it.

Shrooms at Scale

The aughts saw an explosion of community science apps and websites that let users collect photos, dates, times, and geo-coordinates for different plant and animal sightings. Love birds? Join the hundreds of thousands of users on eBird, a database started in 2002 that now has hundreds of millions of bird observations. More of a mushroom person? Theres Mushroom Observer, which has spawned hundreds of thousands of observations from thousands of participants since it started in 2006.

There are online spaces for generalists as well. One example is iNaturalist, an app, website, and online community that launched in 2008. A user on a hike can upload a photo of a plant or animal and have immediate access to the platforms community of naturalists to help identify it. iNaturalist has amassed 25 million observations, over 10 million of which are research-grade. The platform is also a popular public engagement tool for museums and other institutions that use it as part of their programming.

The scale of data collection thats possible with these platforms surpasses anything a team of scientists could ever hope to match over the course of a career, even with ample funding. Users of iNaturalist and eBird have collectively recorded observations of over 200,000 species.

This data often makes its way into scientific research. Despite Mushroom Observer being dominated by amateurs, one of the sites developers, Joe Cohen, says that trained researchers actively participate in forum discussions, track particular species, and share specimens with other users. iNaturalist, Mushroom Observer, eBird, and Monarch Watch all submit user-collected data to the Global Biodiversity Information Facility (GBIF), a central repository that brings together data on species occurrences from tens of thousands of different data sources. Scientists can easily access the data filtered by their species of interest. Data collected by app users and accessed via GBIF has been cited hundreds of times in scientific research.

Burnt Out on Butterflies

But the apps and websites that make this large-scale data collection possible are not designed for conducting scientific research. iNaturalist, for example, makes

clear that its first priority is "to connect people to nature"–the breadth and volume of the environmental data collected is a fortuitous side effect of community-building. And since producing data specifically for scientific research is not what these platforms are for, sampling problems abound.

Data collection sites that are near users' homes or are easy to get to become hotspots for observations, regardless of their value for scientific research. While scientists often travel to field sites to collect data, Mushroom Observers, for example, typically collect data near where they live.

When community scientists do travel, they may be more interested in going to places where they can expect to find a particular species, either because the species is more prevalent or because the place supports data collection in some way. Chip Taylor of Monarch Watch remarked that monarchs are better represented in Iowa because its county conservation boards facilitate tagging efforts. Volunteers also prefer to report species they find interesting. Rare species may be overreported because people are excited to see them and may even travel specifically to see thema data collection pattern encouraged by some platforms design features, like eBirds rare bird alerts. In 2018, the monarch butterfly was the most observed species in iNaturalists research-grade observations in ten of the forty-eight states in the continental US, though its unlikely that monarch butterflies are the most prevalent species in any of those states. Meanwhile, relatively little data was collected by users on the less charismatic Bridgeoporus nobilissimus mushroom.

Misidentifications can also be a problem, even though companion apps for community scientists generally require multiple identifications before an observation is confirmed, and some apps like iNaturalist use computer vision to suggest identifications. In an effort to ensure the quality of their dataset, scientists using crowdsourced observations for research may treat the number of contributions a user has made as a proxy for data quality, and filter out users with a weak contribution history.

Since the charisma, visibility, rareness, and location of a given species can all affect data collection in ways that don't necessarily reflect the species' actual distribution, it can be difficult to determine whether an absence of observations corresponds to a real decline of the species or something else. Some platforms try to account for this in different ways, but others, wary of user attrition, are hesitant to add barriers to submitting observations. After all, what matters most to the platforms is attracting and retaining users. Helping out scientists is a secondary concern.

Bridging the Gap

Researchers do their best to account for the limitations of crowdsourced data. They add instructions to particular data collection efforts or work in tandem with volunteer data collectors on training initiatives about the need for high-quality data. "Data fusion" methods and integrated population models have also become popular tools to bring data together from different sources, weighing the strengths

and weaknesses of each dataset. Gaps in community science data can inform scientists future data collection, providing opportunities to improve sampling design and data collection efficiency.

Despite their limitations, then, platforms like iNaturalist, Mushroom Observer, and eBird are still valuable for scientists. The scale of biodiversity is such that scientists alone cannot record everything, particularly in an era of slashed research budgets and anti-science public policy.

Still, defunding science has serious consequences and we cant afford to narrow our focus at such a critical moment of ecological change. There are inevitable, irreparable gaps in data collected by community members. Going back in time to collect better data on a particular species or in a specific region is impossible. When scientists are more reliant on data collected by volunteers, the fluctuating interests of the public can destabilize research efforts. We still need data about boring species, and from faraway places. These gaps take infrastructure and resources to fill, and we ignore them at our peril.