Laura Montini March 23, 2016

**Abstract**

Title: The Design Decisions that Shaped a Citizen Science Project

Author: Laura Montini

Contact: lmmontini@berkeley.edu

Academic Affiliation: University of California, Berkeley School of Information

Summary: The nature of citizen science projects, which can last indefinitely and involve an unlimited number of participants, means that they can potentially happen on a massive scale. This characteristic makes them well suited for tackling large projects to which organizations might not otherwise be able to dedicate resources. However, this same characteristic also means that citizen science projects can be especially difficult to form and manage. Both large numbers of human and information resources must be organized very intentionally. This case study details the process that The California Academy of Sciences took when organizing a mostly volunteer and amateur-staffed initiative to document all plant life in an 18,000 acre area. This account works through the design questions the Academy needed to answer in order to establish an effective citizen science project.

Acknowledgments: Thanks to The California Academy of Sciences, and especially to Alison Young and Ken-ichi Ueda for their time and guidance.

Keywords: citizen science, information organization, design decisions

**Overview**

In 2011, the Marin Municipal Water District (MMWD) asked the California Academy of Sciences for help documenting every plant species in the Mount Tamalpais watershed. It was immediately clear to the Academy that the project, which involved finding more than 900 plant species over 18,000 acres, was too large in scope and scale for its small team of professional botanists. Consequently, the Academy designed a citizen science project to complete the undertaking.

**What is being organized?**

The plant-focused biodiversity survey relies on volunteers’ participation. Alison Young, the Academy’s citizen science engagement coordinator, began building a team of citizen scientists by contacting people who were familiar with the area already -- for example, those who had previously volunteered at the watershed. From there, the initial volunteer list expanded by word of mouth and, in 2015, it included more than 200 people.

Before the outings, which are ongoing as of March 2016 and usually occur once per month, Young notifies volunteers via email. Each event receives a showing of about 30 to 50 volunteers. The surveyors are divided into groups of four, which include: a data recorder, a photographer, a plant specimen collector and a professional or amateur botanist who is able to identify plant species.

In addition to people, the project involves organizing data about individual plants. The recorded data about a plant observation produce a digital resource description. Ideally each resource description reaches research grade status. In order to qualify for this classification, a resource description must: be associated with a person (anonymous uploads are not accepted), include the date and location of the observation and include evidence of the organism -- in this case a picture. The document officially becomes research grade once a consensus about the identification of the plant it describes is reached.

**Why is it being organized?**

Once it is complete, the database of plant life on Mount Tamalpais will represent a closed resource collection (as opposed to an open resource collection in which items are regularly added or deleted). The survey is designed to depict a snapshot in time, and this is because the MMWD wants to use it as a baseline with which to track and compare future change. The database also provides current information about threatened, endangered and non-native plant species; the data have already informed plans for projects that remove invasive species.

**How much is it being organized?**

The project makes use of a mobile application and website called iNaturalist ([www.inaturalist.org](http://www.inaturalist.org)) that enables observation recording. The tool helps volunteers convert observations made in the field into a digital resource in a database. iNaturalist intentionally does not require granular data entry. This is largely because the app was designed with amateur users in mind, and its creators wanted to avoid as many obstacles to creating an entry as possible.

As a result, iNaturalist isn’t picky about the data it accepts. For example, users can classify a plant at any level, from kingdom all the way down to species. If they choose to enter only “plant,” that is sufficient. The app even allows this field to be left blank. In contrast, Calflora, an online database of plants in California, requires its users to enter the scientific name for every plant they observe. During its first year in 2012, the Mount Tamalpais watershed project used Calflora to log plant sightings, but it ultimately abandoned the tool in favor of iNaturalist. Young decided that the app was a better fit for the project’s mostly non-expert participants.

**When is it being organized?**

Though iNaturalist is the primary way in which volunteers ultimately enter resource descriptions into the database, they rarely use it to capture all of their data. This is because Mount Tamalpais lacks Wi-Fi access not to mention reliable cell phone reception. For this reason, volunteers go out into the watershed with paper datasheets that have fields that correspond to those on iNaturalist. At the end of the day, group members use the app to create resource descriptions about their observations, referring to their hand-written notes and pictures taken in the field.

Information is generally organized on the way in, when it is added to the collection (as opposed to on the way out, when someone or something later retrieves the resource for a particular purpose). However, the resource descriptions, particularly organisms’ identifications, are dynamic and can be updated later, hopefully to improve accuracy. These corrections and clarifications are usually the result of a communal effort and are enabled by the crowdsourcing aspect of iNaturalist’s platform.

**How or by whom?**

While users are permitted to enter broad classifications in iNaturalist’s identification field, the platform’s community of online users, including those beyond the Mount Tamalpais watershed project, typically help to refine those classifications. Considering the information provided, users can indicate what they believe the identity of an organism is. They can agree with the name that is already there, they can disagree, or they can suggest a more granular classification. When a consensus about the organism’s identification is reached, it officially changes in iNaturalist’s database. The platform doesn’t attach status to any user, which means that though someone like Young, a marine biologist, can provide her opinion on an undecided classification, her expert input carries no more weight than that of an amateur, or two.

**Where?**

 The observations are initially housed on iNaturalist, where they are created, and this is where members of the Mount Tamalpais watershed project interact with the data. From here, anyone can download the database. Young, for example, keeps an updated spreadsheet on her computer.

 Because research-grade data must meet standards that are recognized by several organizations, once the document reaches this classification, it can also live in other places. Outside of the project, iNaturalist shares research-grade observations of California plants with Calflora. Similarly, research-grade resource descriptions are shared with the Global Biodiversity Information Facility (GBIF), a network of biodiversity databases used mostly by scientists. Once the citizen science-created data is in the GBIF network, scientists all over the world can access it.

**Other considerations**

The Mount Tamalpais watershed project aims to find as many plants on a list of 900 known species as it can. This number comes from an herbarium collection amassed from more than 100 years of botanizing in the area. The project has documented about 750 of these species with photos. It has also added 50 new plants that weren’t on the original list.

One of the main questions for the future of the project is when it should end. The effort, originally designed to take one year, has now gone on for four years. Young acknowledges that the Academy and MMWD will never find all of the species on the original list; some are too rare, while others might not exist in the area anymore. Exactly when the volunteers can say they have completed their work in the watershed is something Young along with the Academy and MMWD still need to determine.

Young’s other unanswered question is how to ensure that an equally comprehensive survey happens again -- especially decades after the original volunteers have moved on. “What we're doing now, we would have loved if someone did this 100 years ago so we could really compare,” Young says. “It would be great if in 50 years or 100 years someone would go back.”