CITIZEN SCIENCE DATA FACTORY

A Distributed Data Collection Platform for Citizen Science

Part 1: Data Collection Platform Evaluation

Prepared by

The Citizen Science Data Factory report includes four parts:

- Part 1: Data Collection Platform Evaluation
- Part 2: Technology Evaluation
- Appendix A: Wireframe Designs
- Appendix B: Cloud Computing Performance Testing

It is available for download and distribution from:

- http://www.azavea.com/research/company-research/citizen-science/
- http://www.scistarter.com/research/

Funded by:

The report was funded in part by a grant from the Alfred P. Sloan Foundation. The opinions expressed herein are those of the authors and do not represent the opinion of the Foundation.

This report was researched and developed by Azavea and SciStarter, including contributions from:

- Darlene Cavalier, SciStarter
- Robert Cheetham, Azavea
- Rob Emanuele, Azavea
- Jeff Frankl, Azavea
- Mary Johnson, Azavea
- Tamara Manik-Perlman, Azavea
- Josh Marcus, Azavea
- Michael Tedeschi, Azavea

This report also benefited from input provided by both the project representatives interviewed and input from the Citizen Science Community Forum.

Cover Photography Attribution:

- Gypful by Pierre Dalous
- Giant Group by Diliff

Azavea is a B Corporation that creates civic geospatial software and data analytics. Our mission is to apply geospatial technology for civic and social impact and to advance the state-of-the-art through research.

SciStarter brings together citizen scientists; thousands of potential projects offered by researchers and organizations; and the tools, resources, and services that enable people to find, pursue and enjoy these activities.

Published by:

azavea

scistarter.com

Azavea
340 N 12th St, Suite 402
Philadelphia, PA 19107
Telephone: +1 215 925 2600
Web: www.azavea.com
Email: info@azavea.com

SciStarter
306 Delancey St
Philadelphia, PA 19106
Telephone: +1 267 253 1310
Web: www.scistarter.org
Email: darlene@scistarter.com

Table of Contents

Executive Summary

With the rapid proliferation of online citizen science websites, Azavea, a geospatial software engineering firm, and SciStarter, an online community for citizen science, have collaborated to evaluate a representative set of online citizen science software tools. Supported by the Alfred P. Sloan Foundation, this evaluation effort has analyzed existing online citizen science websites for their technology, extensibility, visualization, authentication, gamification, and other features in order to better understand their ability to support a diverse and growing catalog of citizen science projects. The approach to this evaluation was threefold:

1) Explore development of a unified methodology for motivation, outreach, data collection and storage that transcends individual STEM domains and projects.
2) Evaluate feasibility of a flexible toolkit for building customizable data collection research projects within a larger cloud-based framework.
3) Examine a user interface design that will enable citizen scientists to engage in research across multiple topics of interest without having to search through and manage multiple websites.

While existing citizen science platforms provide important benefits for specific needs and audiences, none yet address the much broader need for a generalized data collection, sensor aggregation, and visualization platform that would enable researchers to design their own projects across a range of domains and make them available to a large and pre-solicited pool of volunteers. In order to advance the development of such a system, we must first look at what is working and what is not working in citizen science. To that end, we have put together a pair of white papers that examine the current state-of-the-art from two important and inter-related perspectives:

1) **Existing Platforms:** This, the first white paper, provides a review of existing citizen science platforms and makes recommendations for the features and functionality necessary for a generalized data collection, sensor aggregation, and visualization system upon which cross-disciplinary research can be efficiently solicited, organized, performed and disseminated. Specifically, it includes an in-depth overview of six data collection projects, as well as a summary review of four additional projects that are well-regarded in the genre.
2) **Technology Evaluation:** The second white paper provides a similar review of citizen science technology frameworks, cloud computing platforms, and sensor aggregation and visualization tools. Specifically, it evaluates factors ranging from open source software components and performance metrics to implementation and carrying costs in an effort to determine the type of generalized technology platform that is best suited to geographically-dispersed data collection projects.

Together, these papers define the technology blueprint for a multi-disciplinary citizen science platform focused on data collection and sensor input aggregation and visualization by way of the following specific features:

1) A flexible and scalable cloud-based computing infrastructure that can support both web-based and mobile data collection projects.
2) A means for recruiting participants from a broad pool of individuals who have already expressed interest in a particular type of project, topic or activity.
3) An incentive framework that will enable students and other volunteers to accumulate reputation points and badges, or see their names on leaderboards; these incentives would recognize work on both individual projects and cumulative contributions to all projects on the site.
4) Tools to input and aggregate volunteer contributions of data through geographically distributed networks of sensors.
5) Training methodologies to support and document learning, including a mechanism to test requisite knowledge for participation in more advanced projects.
6) Integration with existing social media for secure user authentication and data validation, as well as dissemination and outreach activities.
7) The ability to emulate popular social networking features including "following" or "liking" other contributors or subsets of data.
8) Geographic search, visualization, and data processing tools that enable citizen scientists to explore individual or collective data contributions.
9) Methods for effectively tracking, motivating, and com-

municating with all volunteers on a project.

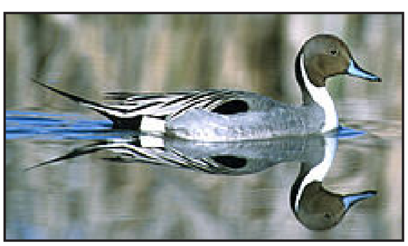
10) An identity management system to enable participants to track their own interests, contributions and accomplishments across projects on the site with a vision towards tracking across platforms.

The "Citizen Science Data Factory" thus envisioned is expected to increase public participation in citizen science research, broaden understanding of the natural world, and provide the foundation for meaningful contributions that advance scientific discovery, particularly for the growing number of research projects that require data input across multiple geographic and temporal extents. A set of user interface design documents has been included in this paper to graphically illustrate the user workflow.

In addition to informing citizen science, the mechanisms documented in these white papers will also be useful for rapidly mobilizing social networks for crisis response, public safety, and civic engagement activities that support local government, non-profit, and academic interests.

Introduction

Public participation in scientific research is not a new concept. To date, however, each citizen science project has generally started from scratch, resulting in a series of data silos and disparate collection efforts that can be frustrating to both project organizers and their potential volunteers.



One of the longest-running known citizen science projects is the Audubon Society's Christmas Bird Count, which was first held in 1900 and has been repeated annually ever since (Droege, 2007). The first Christmas Bird Count drew 27 volunteers, mostly from the northeastern United States. Subsequent counts have increased dramatically in both participation and geographic range; indeed, bird-watching projects in the United States now actively engage more than 15 million "citizen ornithologists" each year and reflect bird sightings from every state and territory (Bhattacharjee, 2005).

Citizen science projects continue to expand in scope and breadth, moving beyond ornithology to include everything from measuring snowfall to classifying galaxies (Cohn, 2008; Bonney et al, 2009). Public participation has proven particularly effective for projects involving ecosystem monitoring or biodiversity studies, which have traditionally required massive amounts of data at large geographic scales. One of the greatest facilitators of public participation in such a wide range of scientific domains has been the proliferation of the internet.

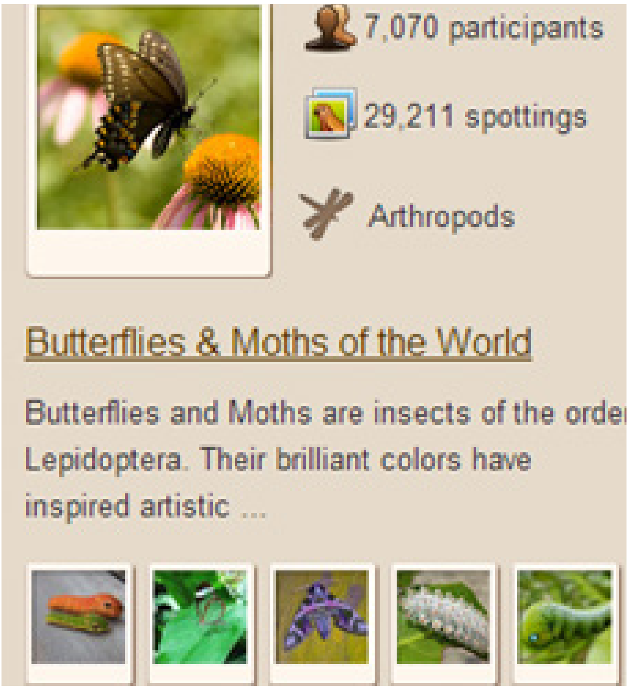
Internet access and the increasing availability of computers, tablets, and smart phones have enabled scientists to conduct far-reaching public participation projects more easily, and to visualize and share their results on a global basis. The visualization of results on the computer screen can be a powerful tool for both citizen science engagement and project success, particularly for research with a geospatial component.

In a report on public participation in scientific research that was funded by the National Science Foundation (Bonney et al, 2009), the Center for the Advancement of Informal Science Education (CAISE) identified three major categories of citizen science projects:

1) **Contributory projects**, which are generally designed by scientists solely for public data contribution.
2) **Collaborative projects**, which go beyond data collection by enabling the public to actively refine the project, analyze data, or disseminate results.
3) **Co-created projects**, which are designed by scientists and the public in tandem, and wherein the public is actively involved, to a greater or lesser extent, through all stages of the project.

For the purposes of this report, we are focusing largely on contributory data collection projects, as these are the most prevalent and the most likely to involve large and diverse participant communities. The report does not assume that scientists are the only authors of contributory data projects, and believe that strong projects may also be designed and promoted by members of the general public.

Regardless of category, most citizen science projects are focused on a single scientific question or environmental issue that can



Project Noah's Mission Area

only be addressed by analyzing data that were collected over multiple geographic and/or temporal extents (Bonney et al, 2009). In the case of the Christmas Bird Count, for example, it was an effort to address concerns about the declining bird population in the United States caused by over-hunting. Such projects require the collective input of many individuals in many different locations at once, and could not otherwise be completed by scientists without a public participation component.

By enabling the public to contribute to the ongoing generation of scientific knowledge, researchers are able to harness the "cognitive surplus" of a large community of interested participants (Shirky, 2010), gather useful data, and motivate public interest in Science, Technology, Engineering, and Mathematics (STEM) topics. The goal of this research, then, is to determine the features and functionality necessary for a single, destination resource for locating, building, managing, and participating in citizen science projects across multiple domains.

Methodology

A broad and multi-faceted review of existing projects and platforms was undertaken in an effort to learn from their successes and shortcomings. While the technology paper looks at citizen science from a software engineering and computing perspective, this paper takes a more general approach.

First, we selected a small subset of citizen science projects and platforms for in-depth or summary evaluation. These projects were selected because they are well documented in the literature and/or have particular utility or audience appeal. With the exception of Zooniverse and Crowdcrafting, two popular platforms for collaborative data classification projects, our in-depth evaluations were centered on contributory data collection projects.

To standardize our evaluations, we developed a series of metrics that could be applied to any citizen science project, regardless of size or domain, to help us determine its utility. In the course of our telephone interviews with the administrator(s) of each project, we asked a number of questions conforming to these metrics, as outlined in Table 1. We also visited each online project from a citizen scientist's or project coordinator's perspective to review its utility firsthand and take illustrative screenshots to inform our research.

Second, we compiled a list of features, functionality, and standard practices that comprise the current state-of-the-art. For the purposes of this paper, they have been broadly categorized as: (1) data contribution; (2) data use; (3) social, marketing and incentives; or (4) other considerations. The SciStarter website was an invaluable resource for this portion of the study, as it provides both summary data and easy access to more than 600 projects and platforms for review and comparison.

Third, we performed a review of the literature and found a number of articles and reports to both inform and broaden our efforts. In particular, a 2009 report on public participation in scientific research that was published by the Center for the Advancement of Informal Science Education (CAISE) and funded by the National Science Foundation, and a 2012 report on citizen science and environmental monitoring by the United Kingdom Environmental Observation Framework that encompassed 600 projects. Similar to our research, each report focuses largely on contributory data collection projects, since most citizen science projects fall under this model (Bonney 1996, Krasny and Bonney 2005, Bonney 2007).

Table 1: Evaluation Metrics and Interview Questions Used in the Study

Evaluation Metrics	Interview Questions
General	1. Considering the array of online citizen science data collection projects, what is distinctive about yours? 2. What, if any, are the innovative User Interface/User Experience features? 3. What are some of the shortcomings of the system? 4. If you could do it all over again, what would you do differently? 5. What new features are in the pipeline? 6. Is there a business model to support the sustainability of the project?
Flexibility	1. Is the platform focused on a few projects in a single domain, or can it accommodate data collection across multiple domains? 2. Can project leaders customize their data collection instruments? 3. If there is a standard format, how did you decide on the structure of the data? 4. Does it support multiple submission methods, including spreadsheet, web form, or mobile? 5. What proportion of submissions comes from each method? 6. If the project supports mobile submissions through a native app, how did you select which platform(s) to develop for (I.e. Android or iOS)? 7. Did you consider developing a responsive website (HTML5-based) instead? 8. Did you do the development in-house, or contract the work? 9. Do you consider mobile data submission to be central to your project's functionality? Do you anticipate this changing in the future?
Display, Visualization, and Publication	1. Is it possible for participants to see the results, or do they simply submit data? 2. What types of visualizations are available? Maps? Charts? 3. Are the visualizations interactive? 4. Can participants see just their own data or the data submitted by others as well? 5. Do you conduct any analysis based on aggregated data? 6. Can citizen scientists export their own data or other data from the site? 7. If so, is there a public API? 8. What other formats are used? Tables (CSV), spatial data (KML), images? 9. How are rights to data submission managed? Is there a single license agreement, or can users choose the license? 10. If the data isn't available to the public, who does have access? 11. Is the data being used by academics or other researchers? Are publications using the data shared in the site?

Technology Platform	1. Is it an open source platform that others can build on? If so, under what license was it released? 2. Where is the source code available? 3. Is there an open and documented API? If so, is there a community of developers or other organizations using the API to build new tools? 4. What was the composition of the team that built the platform? 5. How were the underlying technologies chosen? 6. Would you change the architecture if you could do it over? 7. What technology components were used to build the project? Operating systems, cloud platforms, languages, databases, software frameworks? 8. How long did development take? Were there distinct phases (planned or ad hoc)? Is there a roadmap for future development? 9. What are the technology gaps or other challenges you confronted? 10. To what extent is the platform scalable?
Social, Marketing, and Incentives	1. What is the process whereby data contributors are recruited? 2. How was the user community cultivated? 3. What social aspects do you support? Groups, following, messaging? 4. Do you integrate with existing social networks? 5. If so, do you use third party authentication? 6. Can people link to their Facebook account and post directly from your application? 7. Are pictures simply links to pictures hosted on Flickr/Picasa, or do they need to upload to your site? 8. How are data contributors incentivized? Leaderboards, reputation points, badges, ranks, money? 9. How do you support existing groups or communities? 10. Do you use a custom system or standard library or platform? Mozilla, Open Badges?
Data Quality	1. How do you ensure or quantify the data quality? 2. Are there automated checks, or is it all manual? 3. How transparent is the system and rating to users and researchers using the data?
Cost for a New Project	1. What is the required infrastructure? Web server, other? 2. What are the required human resources and/or specialized knowledge (e.g. database engineer, Drupal Developer)? 3. How many hours and/or dollars does it take to get a new project up-and-running? What skills are required?

Finally, we hosted an online discussion on Cornell University's Citizen Science Community Forum:

http://www.citizenscience.org/community/blog/2012/11/09/what-tools-and-technologies-are-powering-new-frontiers-for-your-citizen-science-projects/. We asked the community to tell us about the software tools they are using or planning to use to support their citizen science projects, including any custom tools they may be developing. We encouraged questions and comments that would help us better understand the useful features that different platforms have to offer, and specifically, their technology, extensibility, visualization, and engagement features. Comments from this forum were largely used to support our in-depth and summary evaluations, as well as our look at the current state-of-the-art.

Overview of Evaluation Results

Contemporary citizen science activities range from tracking the effect of climate change on plants and animals to monitoring earthquake intensity and measuring rainfall. With such a wide range of topics to consider, there have been relatively few attempts to develop generalized software platforms that can support multiple projects; those that do generally have a specific project focus. The Citizen Science Alliance, for example, which is funded in part by the Alfred P. Sloan Foundation, owns and operates Zooniverse, a popular collaborative platform for text, image and sound classification projects, including Galaxy Zoo. Indicia, a product of the National Biodiversity Network, is an open source toolkit for constructing species identification websites, and the Cornell Lab of Ornithology has been instrumental in the development of regional and national birding portals.

Ideally, a generalized data collection platform would be designed to help scientists and project organizers more effectively tap a pool of motivated students, teachers and members of the public; make it easier for these individuals to get involved in scientific research across multiple domains and activities; help organizers of citizen-science projects communicate outcomes and plan outreach; provide a means for recruiting participants; and ultimately increase public engagement in science.

While many projects or platforms provide one or more components of this vision, our evaluation has suggested that (at the time of the survey in December 2012) no single project or platform currently provides them all. That being said, however, we can extract a number of "lessons learned" from these efforts that can inform future development work. Most notably:

1) **Open source** – The use of open source components can reduce development and carrying costs for new projects as well as provide assurance of long-term availability of the source code.
2) **Cloud hosting** – In tandem with open source technology is the need for a cloud hosting platform that will mitigate internal hosting costs and provide scalability for processing large data sets and applying data science techniques.
3) **Data visualization** – Citizen science projects are more meaningful to participants if they can visualize the outcomes.
4) **Gamification –** The deliberate integration of game play into the design can have a significant impact on the ability to attract, engage and retain motivated participants.
5) **Multiple projects, single sign-on** – Most current projects require a separate login for each project. A unified authentication system that can be shared across many projects is far more convenient than multiple, separate logins for each project. Further, it creates an opportunity to accumulate aggregate data on engagement, make suggestionsto users for related projects, and enable contributors to track and build upon their own interests, contributions and accomplishments.
6) **Social networks** – Powerful tools to engage volunteers, validate data, encourage outreach, and cultivate new content.

The following tables summarize the current state-of-the-art and other pertinent information as reflected in each evaluated project.

Current State-of-the-Art

The goal of most citizen science projects is to obtain meaningful data from volunteers that will inform scientific research and address real-world issues. In its model for developing a citizen science project, the Cornell Lab of Ornithology outlined a nine-step

Table 2: In-Depth Evaluation Projects – Current State-of-the-Art

✔ *Currently Available* △ *Incomplete or “coming soon”* ⊖ *Not available*

Current State-of-the-Art	CitSci.org	eBird	Indicia	National Geographic FieldScope	National Phenology Network	Wild Knowledge
Focus	Invasive species	Birds	Species identification	Community geography	Species life cycle	Nature, other
Open Source						
Cost	Free if not cus-tomized	Annual fee	Up to £2,000	Not specified	Not specified	Varies – Up to £20,000
Data Contribution: Platforms						
Web	✔	✔	✔	✔	✔	△
Mobile	△	✔	✔	✔	✔	✔
Data Contribution: Formats						
Text	✔	✔	✔	✔	✔	✔
Coordinates	✔	✔	✔	✔	On set-up only	✔
Other Measurements	✔	✔	✔	Varies by project	✔	✔
Photos	△	✔	Varies by project	✔	Not from phone	✔
Audio/Video	△	✔	⊖	⊖	⊖	✔
Sensor	⊖	GPS, camera	⊖	⊖	⊖	GPS, camera
Data Use						
License Terms	Creative Commons	Not Specified	Non-Commercial	Not Specified	Creative Commons	Not Specified
Quality Assurance	✔	✔	✔	✔	✔	✔
Sharing	✔	✔	✔	✔	✔	Varies by project
Visualization	✔	✔	Varies by project	✔	✔	✔
Social, Marketing, and Incentives						
Social Network Integration	✔	✔	⊖	✔	✔	△
Social Mechanisms	✔	✔	✔	✔	✔	✔
Gamification	⊖	✔	⊖	△	✔	Varies by project
Financial Incentives	⊖	⊖	⊖	⊖	⊖	⊖
Other Considerations						
Authentication	Custom	Custom	Custom	Custom	Custom	Future Facebook
Recruiting	Custom	Custom	Custom	Custom	Custom	Custom
Training	✔	✔	✔	✔	✔	Varies by project

Table 3: In-Depth Evaluation Projects – Current State-of-the-Art

✓ *Currently Available* △ *Incomplete or "coming soon"* ⊖ *Not available*

Current State-of-the-Art	EpiCollect	Every Aware	iNaturalist	Project Noah	Zooniverse
Focus	*Epidemiology*	*Sensors*	*Wildlife*	*Wildlife*	*Image/sound classification*
Open Source	✓	⊖	✓	⊖	✓
Cost	*Varies by project*	*Varies by project*	*Varies by project*	*Varies by project*	*Varies by project*
Data Contribution: Platforms					
Web	△	△	△	✓	✓
Mobile	✓	✓	✓	✓	✓
Data Contribution: Formats					
Text	✓	✓	✓	✓	✓
Coordinates	✓	✓	✓	✓	*Not applicable*
Other Measurements	*Varies by project*	*Varies by project*	✓	✓	*Varies by project*
Photos	✓	⊖	✓	✓	*Classification only*
Audio/Video	⊖	⊖	⊖	✓	*Classification only*
Sensor	*GPS, camera*	*External networks*	*GPS, camera*	*GPS, camera*	⊖
Data Use					
License Terms	*Not Specified*	*Not Specified*	*Optional by user*	*Not Specified*	*Zooniverse owns data*
Quality Assurance	*Varies by project*	⊖	✓	✓	✓
Sharing	✓	*Varies by project*	*Varies by project*	*Varies by project*	✓
Visualization	✓	✓	*Varies by project*	✓	*Varies by project*
Social, Marketing, and Incentives					
Social Network Integration	✓	✓	✓	✓	⊖
Social Mechanisms	✓	✓	*Varies by project*	✓	✓
Gamification	⊖	*Varies by project*	*Varies by project*	✓	*Varies by project*
Financial Incentives	⊖	⊖	⊖	⊖	⊖
Other Considerations					
Authentication	*Custom*	*Custom*	*Facebook, other*	*Facebook, other*	*Custom*
Recruiting	*Custom*	*Custom*	*Custom*	*Custom*	*Custom*
Training	*Varies by project*	*Varies by project*	*Varies by project*	✓	✓

process that included recruiting participants, collecting and managing data, interpreting and visualizing data, and disseminating project results to the research community (Bonney et al, 2009). In order to evaluate the utility of representative citizen science data collection systems, we must first understand how existing and emerging technologies are being used address these processes and otherwise support a successful project. To that end, we focused our evaluation of state-of-the-art in four specific areas:

1) **Data Contribution**
    - Data Platforms
    - Media Formats
2) **Data Use**
    - License terms
    - Quality Assurance
    - Sharing
    - Visualization
3) **Social, Community Building, Marketing, and Incentives**
    - Integration with Other Social Networks
    - Social Mechanisms
    - Gamification
    - Financial Incentives
4) **Other Considerations**
    - Authentication
    - Contributor Ecosystem and Recruiting
    - Training

The strengths and weaknesses of each practice are assessed through hands-on evaluations, a review of the literature, and input from our citizen science forum. Data contribution and visualization mechanisms are also assessed from a technology perspective in our technical evaluation white paper.

Data Contribution

Scientific research provides many opportunities for public involvement, and data contribution is one of the most important. Indeed, most citizen science projects include some type of data collection component. Project Noah, for example, asks volunteers to collect data about local wildlife, while National Geographic FieldScope solicits data that is geographically driven. There are a number of mechanisms being used to acquire, organize, and maximize the value of contributed data. Data collection platforms and media formats are important considerations when planning a new project or establishing data collection policies.

Data Collection Platforms

The quality and rate of data collection in contributory citizen science projects can be greatly influenced by platform selection. While some local citizen science projects still rely solely on manually collected data through hardcopy notebooks or forms, our evaluation focuses instead on web, mobile, and sensor network technologies as potential platforms for data collection. Internet access in particular, and the increasing availability of computers, tablets, and smart phones, has transformed citizen science through their ability to support geographically distributed data collection efforts and to visually summarize the results. The use of wireless sensor networks for collecting data is also on the rise.

Internet/Web

The internet has been widely hailed for its ability to engage the public in scientific research. All of the projects evaluated for this paper have dedicated websites; while some projects use websites strictly as a means of promoting their endeavors, others facilitate data collection, all or in part, through a web-based interface accessible through their sites. Cornell University's eBird and Project Feeder Watch projects, for example, use the internet to coordinate the data collection efforts of thousands of citizen ornithologists across North America that are contributing to a growing database of bird sightings. For collaborative data classification projects such as Galaxy Zoo, wherein galaxy images must be painstakingly classified according to pre-designated criteria, the internet has also become a tool for harnessing distributed thinking power that can advance the pace of academic scholarship and democratize scientific discovery (Cook, 2011). At the same time, it is an ideal vehicle for public outreach, recruiting, and project dissemination activities, as shown in Figure 1.

In 2012, a joint study of 340 existing contributory, collaborative, and co-created citizen science projects registered on SciStarter was completed by Carnegie Mellon University and the University of California (Kim et al, 2013). Of the 340 projects studied, 53% used websites as their primary point of data submission. Accord-

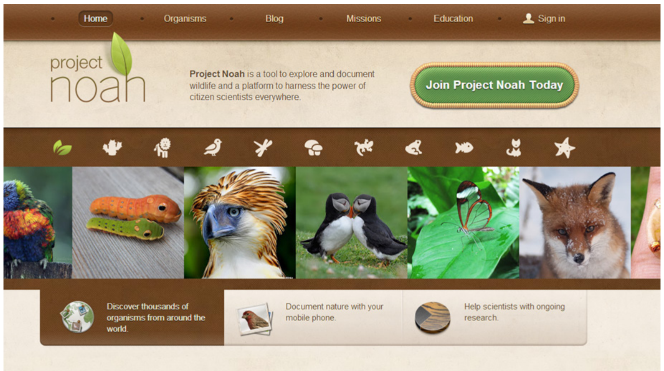


Figure 1: The Project Noah website's home page invites citizen scientists to learn more about documenting wildlife and offers easy access to registration and supporting materials.

ing to a semi-systematic review of environmental monitoring projects by the UK Environmental Observation Framework (2012), the numbers are even higher. The study determined that 80% of 211 active environmental monitoring projects were collecting data through a website. Custom online forms, or existing web services such as SurveyMonkey or Twitter, were used to facilitate data entry on these websites.

One of the major challenges to web-based citizen science projects is the cost of purchasing and maintaining database servers needed to support potentially large amounts of web traffic. Galaxy Zoo, for example, was receiving 70,000 classifications per hour within just twenty-four hours of being launched. In fact, it became so popular so quickly that the surge in traffic literally overwhelmed the server hosting the images for the site. The spike in traffic eventually melted a computer cable and shut the site down. The issue was ultimately resolved by putting all the servers in the cloud. The advantage of being "in the cloud" is that projects can scale to meet demand without being burdened by large amounts of hardware when the demand falls. The Galaxy Zoo site now has sufficient bandwidth to handle multiple data classification projects simultaneously as well as the increasing number of citizen scientists that want to participate.

Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction (Mell and Grance, 2011). The best known such infrastructure is Amazon Web Services, but other options include Rackspace Cloud and Google App Engine. These services provide an ability to add extra "machines" that are billed per minute, per gigabyte transferred/stored and provide a cost-effective way to enable citizen science projects to handle very large numbers of volunteers as the user community grows. One of the most important benefits of deploying citizen science projects in the cloud is that it provides access to the same systems that power Amazon, Google, Microsoft, or other parent organizations/cloud providers without requiring a large and ongoing investment in hardware infrastructure upgrades or internal IT management staff. Projects Like Galaxy Zoo are able to access additional processors, storage, and network bandwidth on an as-needed basis to support project growth and expansion over time, but will only pay for the resources that are actually used.

Mobile Devices

Mobile devices of all types are particularly well suited for data collection projects and allow more data to be shared through digital media. A typical mobile application is shown in Figure 2. The 2013 Pew Internet Project reported that 56% of adults and 78% of students in the United States now own a smartphone (Smith, 2013). Tablet devices have also risen in popularity: more than 31% of all internet users in the United States owned tablet devices by the end of 2012, up from only 12% the year before. In spite of this promise, however, only 11% of 340 projects registered on SciStarter offer mobile applications for field data collection (Kim et al, 2013).

There are a variety of issues that may be contributing to these lower numbers. First, the cost to develop a mobile application is relatively high, particularly for multiple operating systems such as iOS (iPhone) and Android. In addition, rapid advances in mobile technology require ongoing maintenance and support in order to keep pace with new operating system releases.

One way for citizen science project organizers to mitigate these concerns is by developing a mobile-optimized web application rather than native or cross-platform mobile applications. Mobile websites are optimized for smaller screens and touch interfaces. Most importantly, they are essentially universal to all contemporary smartphone browsers, which make them both flexible and cost-effective for many contributory data collection and collaborative data classification projects. In the past two years, new frameworks, such as Bootstrap and Foundation, have appeared that support "responsive" design through which the layout of a web application adapts to the size and orientation of the screen. This enables a project to avoid the need to create a web application specific to tablets and smartphones. Nonetheless, if a web application is not designed in this manner from the outset, it is rarely easy to retrofit.

From a positive perspective, mobile applications or mobile websites enable citizen scientists to enter data directly from the field as it is collected. They also facilitate "participatory sensing", an approach to citizen science and data collection in which participants use their smart phones and tablet devices to collect and share information about the world around them. The powerful sensors built in to most mobile and tablet devices can be used to collect sounds, images, and location-based data that can inform scientific research in ways simply not possible with a web-based application alone.

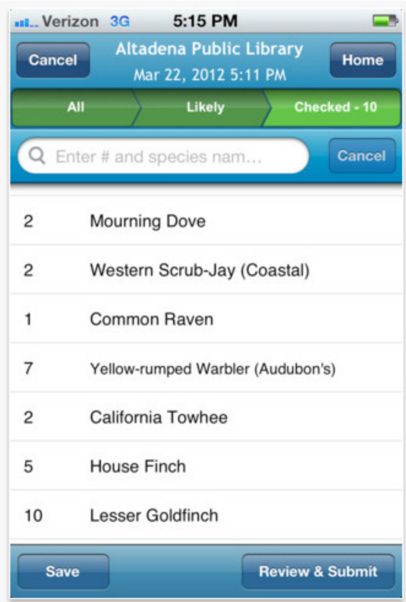


Figure 2: BirdLog works with the eBird website to enable field data entry of bird observations.

Sensor Networks

A sensor is a device that detects or measures some type of input from the physical environment. A sensor network is a series of sensor devices that detect or measure these inputs at diverse locations. Sensors can be used to enhance awareness of environmental quality and physiological changes on a real-time basis. In the realm of citizen science, sensors are being used to measure inputs ranging from air pollution to stream flow and feed them directly to computers. The Quake Catcher Network, for example, is a collaborative project sponsored by the National Science Foundation asking citizen scientists to install seismic sensors in locations that have a computer with an internet connection. These sensors will provide researchers with a greater understanding of seismic activity across the United States.

Perhaps the most accessible source of sensor data for citizen science projects is the smartphone. Contemporary smartphones are packed with built-in sensors that include an accelerometer, digital compass, gyroscope, GPS, microphone, and camera. They can also be fitted with plug-in external sensors capable of detecting inputs such as radiation levels or airborne chemicals. As smartphones have become more powerful, affordable, and accessible in countries around the world, they have introduced a collective research capacity that would not otherwise be possible without significant financial resources (Goldman et al, 2008). Participatory sensing is the concept of deploying sensor networks on a potentially global basis by harnessing the data-collecting enthusiasm of citizen scientists and their smartphones. In parallel with growing smart phone capability, hobbyists and inventors are extending them with new sensors. The Sensordrone project funded on Kickstarter is one example among several in the past two years.

One such project is WikiSensor, a software platform that enables citizen scientists to assess their local environmental quality using their iPhones and share the data with the broader user community. WikiSensor, shown in Figure 3, also offers crowdsourced maps of environmental quality. Street Bump is a mobile application specific to the Boston area that lets citizen scientists collect road condition data as they drive. Likely road problems are submitted to the City via its Open311 system. While the data provided by projects like these is definitely useful, the automated collection process may limit citizen engagement and education potential (Roy, et al, 2012). Another smartphone data collection project, Sensr, is discussed in our technical evaluation white paper.

Participatory sensing has raised privacy concerns, since it often requires participants to stream real-time personal data about their

Figure 3: The WikiSensor app.

locations, routines, or health behaviors (Anthony et al, 2007). The quality and extent of data collected in participatory sensing projects may also be impacted by the variations in sensor sensitivity across differing smartphone platforms, models, and years. At the same time, it offers tremendous potential for large-scale surveillance of critical environmental issues, including climate change (Roy et al, 2012).

Media Formats

Whether the data platform is a website, mobile application, or some combination of the two, data must be contributed to citizen science projects in one of several media formats. Participation often involves the completion of manual or online forms and surveys, the submission of geographic coordinates and digital photographs, or taking measurements.

Text attributes

Textual data entry was an important discussion point at the 2007 Citizen Science Toolkit Conference hosted by the Cornell Lab of Ornithology (McEver et al, 2007). The overwhelming consensus among its academic participants was that entering and managing textual data online provides a level of compliance, uniformity, standardization, and quality control that is unrivaled by manually completed paper forms. Online forms can be specifically designed with data entry protocols that prevent them from being submitted unless all fields have been correctly completed, as shown in Figure 4. This eliminates the need to check back with participants about potentially missing or non-compliant data (Prytherch et al, 2006). Cornell's eBird site, for example, uses customized data entry forms containing only those species typically found in a specific geographical area at a specific time of year in an effort to minimize erroneous entries. Further, online forms eliminate the time-consuming and potentially error-prone process of transferring manually collected textual data to the computer. The UK Environmental Observation Framework study (2012) also noted that online entry of textual data tended to reduce errors and improve data capture rates.

Coordinates

Since the first Christmas Bird Count in 1900, location data has been critical to many data collection projects (Aoki et al, 2009; Kim et al, 2013). Geographic coordinates can be used to describe a location even in remote areas where no physical street address is available, so they are particularly suited to environmental monitoring projects. The coordinates are generally expressed as latitude and longitude – in the case of 340 North 12th Street in Philadelphia, for example, as 39.958718, -75.158395, or N 39 ° 57' 31.4" latitude, W 75 ° 9' 30.2" longitude. Once these geo-

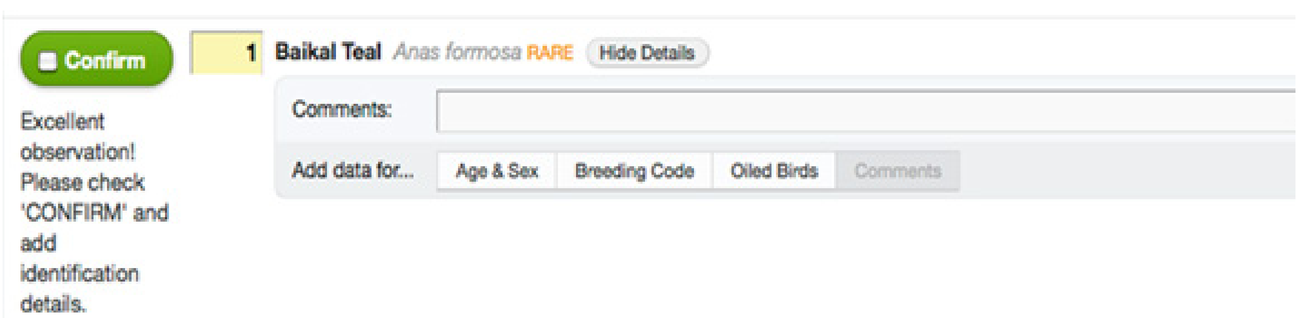


Figure 4: When eBird volunteers report a bird that is considered rare or unusual, they are automatically prompted by the system to provide additional information and/or confirm their observation.

graphic coordinates have been established, they can be stored in a database, displayed on a map, or transformed through data processing or spatial statistics routines. This process of turning descriptive locational data into an absolute geographic reference has become a critical component of the citizen science workflow. Global Positioning Systems (GPS) devices, particularly those in most smartphones, are a helpful means of gathering or "sensing" this data and attaching it to textual, photographic, or other observations collected in the field. Of the 211 environmental projects surveyed by the UK Environmental Observation Framework (2012), 92% required some type of location data.

Report Details

Location name:	Salton Sea
Observation date:	1/12/05
Duration:	3 hour(s) 0 minute(s)
# of people in birding party:	3
Are you reporting all the species you identified?	Yes
Total # of species:	71
Observation type:	Traveling Count
Start time:	7:00 AM
Distance covered:	77.6 mile(s)
Area covered:	N/A
Weather had a negative effect on my ability to collect birding information:	No

Figure 5: Temporal and other measurements in an eBird report.

Other Measurements

Temporal data adds important context to geographic coordinates. The time of day, week, or year when an observation was made, its duration, and the interval between multiple observations at one or more locations can have important scientific bearing. Projects like Journey North, for example, monitor the location and timing of monarch butterfly migrations each spring to help researchers better understand the impact of habitat destruction, climate change, pesticides, and other anthropogenic factors on the butterfly population (Donnelly, 1999). Project Budburst similarly monitors the temporal leafing, flowering, and fruiting of plants at different geographies as a measure of climate change. eBird accepts multiple types of temporal and other measurements, as shown in Figure 5.

The **size** of an object, which can be measured as its height, depth, length, breadth, or diameter, is another important observation that citizen scientists are asked to provide. The Snow Tweets Project, for example, asks citizen scientists to measure and share snowfall totals on the web. Rainlog similarly documents backyard rainfall measurements. The Tag A Tiny Tuna Fishing Program is somewhat more ambitious. It encourages anglers to record the length of juvenile Atlantic Bluefin tuna and tag them with ID tags before releasing them back into the sea.

A number of projects that explore climate change or weather ask citizen scientists to record and submit **air temperature** measurements at pre-determined intervals. Temperature Blast, for example, is a Baltimore-based project that asks citizen scientists to collect live and archived temperature data to compare readings around the city. The data is then used to test models of temperature patterns citywide to illustrate the Heat Island effect and aid in urban planning. Volunteer water monitoring projects require the measurement of **water temperature** in streams or rivers, and may also require the measurement of pH levels using test strips.

Perhaps the simplest and most basic measurement is **quantitative**. How many items or events can be observed in a particular location on a particular day or time? The Great World Wide Star Count is an example of an international citizen science project that asks participants to locate certain constellations overhead on designated evenings and count the number of stars they see. The results are being used to measure light pollution (Ohab, 2012). The eBird project asks participants to enter the total number of birds for each species they observe.

Photographs

Contributory citizen science projects are increasingly making use of photographs to advance research efforts in multiple domains. Species identification projects like Project Noah and iNaturalist, for example, ask citizen scientists to submit photographs of local wildlife or plants to help monitor how species distribution has changed over time. As shown in Figure 6, the Project Noah user profile page for each user enables them to display the photographs they have submitted and provides a running total of their contributions to the site. The Encyclopedia of Life, which is funded in part by the Alfred P. Sloan Foundation, is leveraging contributions from citizen scientists for a growing collection of images that aims to document all of Earth's 1.9 million known species. The goal of this project is to increase awareness and understanding of living nature by sharing the collective knowledge of its participants.

Photographs have become an important resource for many environmental monitoring projects. The UK Environmental Observation Framework study (2012) indicated that 32% of the 211 projects it surveyed required a photograph to be submitted as part of the data collection process. Smartphones have made it easy for citizen scientists to submit geotagged photographs directly from the field, thus making the data verification and validation process more efficient for project administrators.

In collaborative data classification projects, the photographs have already been collected by others and citizen scientists are asked to sort or catalogue them in some way. The projects on the Zooniverse platform, including Galaxy Zoo and Moon Zoo, are examples of data classification projects. Galaxy Zoo has proven particularly popular; within its first year, citizen scientists had classified more than 50 million galaxy images on the site. It is interesting to note that a study comparing the galaxy classifications made by citizen volunteers with those completed by experts showed a high level of similarity between the two (Lintott et al, 2008).

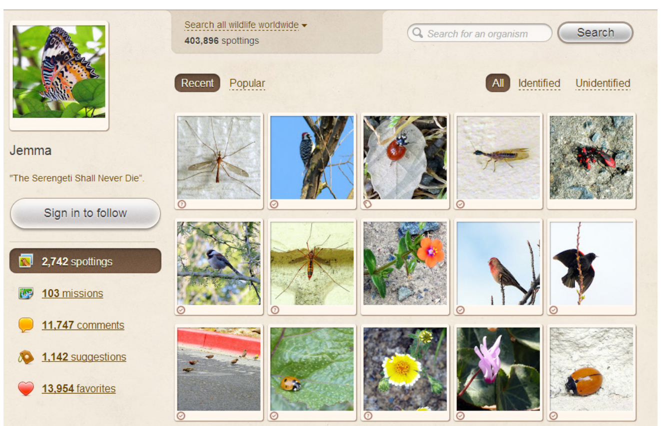


Figure 6: Project Noah harnesses the power of citizen scientists and their smartphone cameras to collect ecological data and preserve global diversity. Volunteer images are a critical component of the project.

While data classification projects have demonstrated the importance of human interpretive capabilities in citizen science, other projects are using images to extend these interpretive capabilities to computers. The Leafsnap project encourages citizen scientists to submit photographs of leaves using their smartphones. A pattern-matching algorithm is being developed and "trained" to automatically identify the tree species based on the photographs and return that information to the submitters. It works like an automated field guide. Leafsnap is currently limited to tree species of the northeastern United States, but is working with citizen scientists to extend its recognition capabilities to species across the country. It remains to be seen if connectivity issues or lengthy processing times will limit the appeal of image recognition analysis projects for the citizen science community (Roy et al, 2012).

Audio and Video

The public's fascination with user-created audio and video content is undeniable. In fact, more audio/video content is uploaded to YouTube in any sixty-day period than was produced by the three major television networks over the past sixty years. This fascination has yet to translate itself to citizen science, however. In its study of 211 environmental monitoring projects, the UK Environmental Observation Framework (2012) determined that only 5% required video or sound submissions as part of the data collection process, and only 9% used videos for tutorial or outreach purposes.

WildKnowledge, eBird, and Project Noah were the only projects in our evaluation that accepted audio or video contributions, often by linking to YouTube or Vimeo content, as shown in Figure 7. In our review of the literature, one project of note was Cornell University's Macaulay Library, which bills itself as "the world's largest and oldest scientific archive of biodiversity audio and video recordings." Thanks to contributors worldwide, site visitors can hear sounds made by three quarters of the Earth's birds, as well as a large number of insects, mammals, fish and amphibians. Citizen scientists are encouraged to add to the collection. Another citizen science project that is currently employing video is the Horowitz Dog Cognition Lab's Play with Your Dog. To participate, citizen scientists need to provide a 30-60 second video of themselves and their dogs playing together, upload their videos to the project website, and complete a short survey. The stated goal of the project is to learn more about human relationships with dogs.

There are also a number of data classification projects that present rather than collect audio and video content. Bat Detective is an online citizen science project on the Zooniverse platform that allows participants to take part in wildlife conservation by sorting recorded bat calls from all over the world. Digital Fishers asks

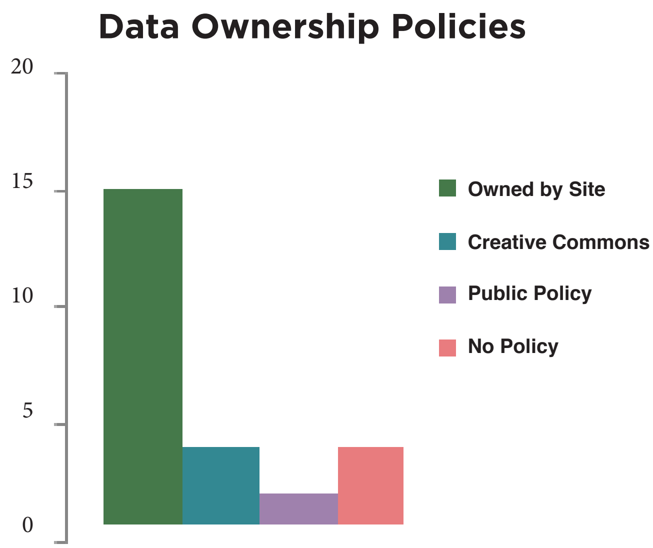


Chart 1: The results of a survey of 25 citizen science projects on the SciStarter website. Most projects surveyed claim ownership of all data submitted by participants.

participants to watch 15-second deep-sea videos on their home computers and click on some simple responses. The site also uses video for project demonstrations and tutorials.

During our citizen science forum, one participant extolled the need for videos as part of any future citizen science platform, though mostly as a means of showcasing a successful project. The CAISE study (2009) agreed that video can be an important addition to any citizen science project, particularly as a means of providing context or instruction for new participants.

Data Use

One of the discussion threads in our citizen science forum involved the desire for community administration of citizen science data. In reality, however, the terms of data ownership, administration, and use tend to vary widely from project to project.

License Terms

The legal and cultural frameworks associated with data sharing and "open" data are undergoing significant changes and remain in a state of flux. Many citizen science projects struggle to strike a balance between openly sharing their data and reserving specific rights. A Creative Commons license offers a variety of options for sharing creative media, including "Share-Alike" or "Non-Commercial", which clearly define the conditions under which the data will be made available. Creative Commons is a nonprofit organization dedicated to expanding the range of creative works available for others to legally build upon or share. However, Creative Commons was primarily designed for application to creative output, such as music, images, and text, and some groups now see it as inappropriate for structured data, the most common form of measurement data. To meet the needs of the OpenStreetMap (a wiki-style map of the planet), the Open Knowledge Foundation has developed the Open Data Commons and three flavors of Open Database licenses that more effectively cover database content. OpenStreetMap was the first major project to adopt the new license. While we do not know of any citizen science projects yet using this license, the concerns with Creative Commons license applied to data lead us to suggest that the Open Database License could be usefully applied to projects that seek to encourage open data sharing.

A survey of 77 small to medium-sized citizen science projects by Syracuse University (Wiggins, 2009) noted that 31% of these projects attributed data ownership to the public, potentially through some type of licensing mechanism. Another 20% of projects attributed data ownership to the researchers conducting the project, and 16% to project contributors. Another 14% indicated that they currently have no data ownership policies in place.

Of the projects we evaluated for this paper, two have a Creative Commons License, one specifies non-commercial use only, one claims ownership of all data, one offers users a Creative Commons option, and the rest do not specify a policy. Zooniverse, for example, uses a data license policy that is similar to many open source software projects. Each individual contributor has a copyright on the material they contribute, but the Citizen Science Alliance, which operates the site, has the right to reproduce, use, adapt and publish the data set as a whole. If participants do not agree to these terms, they are unable to interact with the site. The iNaturalist project outlines similar terms, and this is the most common arrangement for many of the most mature open source software projects. On the other hand, CitSci.org uses a Creative Commons license for its data, while the National Phenology Network provides open and universal access to its data as long as its data attribution policy is followed. Citizen scientists using Proj-

Figure 7: eBird users can embed audio recordings or videos in their submissions to help reviewers confirm their rare sightings.

ect Noah are granting the project a non-exclusive license to their content, where it will be made available to anyone using the site.

To supplement these results, we surveyed 25 projects on the SciStarter website, as shown in Chart 1. The majority of projects surveyed took ownership of all participant data, often under the provisions of copyright law. Other projects have multiple licenses in place. For example, while most of the data in the Encyclopedia of Life project is licensed under some version of the Creative Commons license, the license terms vary from page to page and image to image. The site's terms and conditions encourage participants to reuse its content, but at the same time warns them that each page on the website may consist of multiple data elements, and each element may only be reused under the specific license that applies to it. This is a confusing arrangement for consumers, but is not dissimilar from consumer-oriented web sites like Flickr and is a reflection of the complex intellectual property landscape that currently prevails.

Quality Assurance

Data quality is an important concern in citizen science projects because of the diversity of backgrounds and skillsets among its contributors. For the purposes of this paper, data quality is defined as the fitness of data for scientific research and concerns its completeness, validity, consistency, precision, and accuracy (Wiggins et al, 2011). All but one of the projects we evaluated for this paper described some type of quality assurance practices.

The United States Geological Survey partnered with several universities to perform a random survey of 128 citizen science projects involving invasive species. The survey, which focused on data management and data quality practices among these projects, determined that only 39% have some type of quality assurance methods in place (Crall et al, 2010). The UK Environmental Observation Framework study of 211 environmental monitoring projects (2012) was more positive, indicating that quality assurance practices seemed to be evident, to a greater or lesser extent, in all projects it surveyed. Similarly, the Syracuse University survey of 77 small to medium-sized citizen science projects (Wiggins, 2009) found that most respondents employed at least two methods of data validation in their projects. Participants in our citizen science forum singled out the iNaturalist project as doing a good job of quality assurance. One of their quality assurance methods is shown in Figure 8.

There are a number of studies indicating that citizen scientists can collect data comparable to that collected by professional researchers when specific quality assurance methods are put in place (Au et al., 2000; Canfield et al. 2002; Fore et al., 2001; Delaney et al., 2008). This is particularly true for projects that involve the collection of quantitative rather than qualitative data (Gommerman and Monroe, 2012). Among the projects surveyed, quality assurance efforts included such practices as volunteer training, expert validation, location validation, and deletion of suspect data and/or data that might introduce spatial or temporal bias. Depending on the project, these practices were employed before, during, and/or after the data was collected.

While some quality assurance methods require personal interven-

Figure 8: If iNaturalist can't match the species name provided by a volunteer, it will provide a warning like this.

tion by project organizers and professional researchers, others are implemented in a manner that enables either automated detection of problems or is a structured part of the user interaction – essentially, the crowd monitors the crowd. For example, the Open Air Laboratories (OPAL), which developed the Indicia project, surveys biodiversity data in the United Kingdom using a combination of methods for data quality assurance that include online tests to evaluate participant expertise, as well as pairing citizen scientists with skilled professionals as they undertake their first surveys. iSpot, a project of Open University in the UK, combines the accumulated reputation of each editor, their self-assessed confidence level, and the collective reputation of confirmations from other users to derive a composite species identification certainty score. Cornell University's ornithology projects provide participant support through web-based and print educational materials. These are supplemented by more personalized help from project staff regarding questions, species identification, and data entry protocols. Cornell also uses expert validation to document rare and unusual reports, including species seen out of their typical ranges. Further, most of their online data entry systems automatically trigger warning messages when unexpected data are entered.

Sharing

The data collected in citizen science projects is shared in a variety of ways. While some projects limit data access to members only, many others provide access to anyone with an interest in the subject matter. Indeed, one of the hallmarks of a successful citizen science project is the dissemination of contributed data to inform scientific research.

One of the most important methods of data sharing is through the publication of papers in peer-reviewed journals. The Galaxy Zoo project, for example, has generated some of the most highly cited peer-reviewed papers in the field of astronomy. The National Phenology Network project has also resulted in a number of peer-reviewed papers, which are listed on the project website.

There appears to be a direct correlation between public participation and data sharing in successful citizen science projects. All of the contributory data collection projects we evaluated for this paper offer one or more methods for sharing volunteer data, and our citizen science forum reiterated the importance of making project data available to the volunteer community. Further, the UK Environmental Observation Framework study of 211 environmental monitoring projects (Roy et al, 2012) found that 80% of these projects shared data with project members or the general public. One of the primary reasons given for data sharing was to encourage ongoing participation and underscore the value of volunteer contributions.

Visualization

The integration of data visualization tools with citizen science projects can provide a powerful means of community engagement and learning. Dynamic visualizations in particular provide a sense of instant gratification for citizen scientists, since they are able to see their contributions almost immediately in the context of a larger project.

Most citizen science data includes some type of geospatial component, which makes maps an effective means of data visualization for many projects. Each contribution can be displayed as a point on the map and some provide charts for aggregate data. Few projects move beyond this basic visualization capability, however, and we believe this is an important area for advancing the state of the art. Two exceptions, however, include the eBird project at the Cornell Lab of Ornithology and the National Phenology Networks. The former provides animated maps to display daily and seasonal patterns of relative abundance for individual bird species over designated multi-year timeframes. The latter provides a basic map of contribution locations.

While not as prevalent as maps, other types of data visualization include charts, graphs, and histograms. Projects such as CitSci.org and EpiCollect also use APIs to make volunteer data available for visualization outside the project in GIS and other software applications.

EBird is also innovative in the sense that it provides not only visualization of observations but also of user contributions. The relationship between data visualization and public participation was perhaps most graphically illustrated by changes to the eBird site in April 2006. The website was upgraded to enable participants to track their own observations through "life lists", graphs, maps, histograms, and other data visualizations, and to compare their observations with those of other participants. The number of individuals contributing data to the eBird site nearly tripled within just a few weeks of implementing these changes (Sullivan et al, 2009), and feedback continues to be very positive. One stunning example is provided in Figure 9.

Social, Community Building, Marketing, and Incentives

In order to maximize the benefits of citizen science for a diverse and disparate community of participants, it is important to understand what motivates them. A number of studies have been made on volunteer motivation (Nov, 2007; Nov et al, 2011; Bradford and Israel, 2004), and these can be summarized as follows:

- Altruism and the desire to assist a larger cause
- Advancing fields of research
- Learning opportunities on a topic of interest
- Community engagement and social networking
- A desire to publicly display one's knowledge
- Ideological beliefs, particular regarding the need for freely accessible data
- Guilt, political correctness, or other social concerns
- Preparing for a career change
- Personal enjoyment or friendly competition

Learning opportunities tend to be inherent in citizen science projects and vary only by topic or intensity, as further explored elsewhere in this paper. Altruism, egoism, ideological beliefs, social concerns, and preparing for career changes, however, are not factors that are easily controlled or integrated into citizen science projects to increase participation and engagement. Aspects of community engagement, social networking, and friendly competition, on the other hand, are more readily employable and show particular promise for contributory data collection projects.

Integration with Other Social Networks

More than 2 billion people around the globe are connected to the internet, and one in every nine individuals is on Facebook. More than 3,000 images are uploaded to Flickr each minute, and more than 5 billion images are currently hosted on the site. Twitter is adding 500,000 users each day, and Google+ reached a record 10

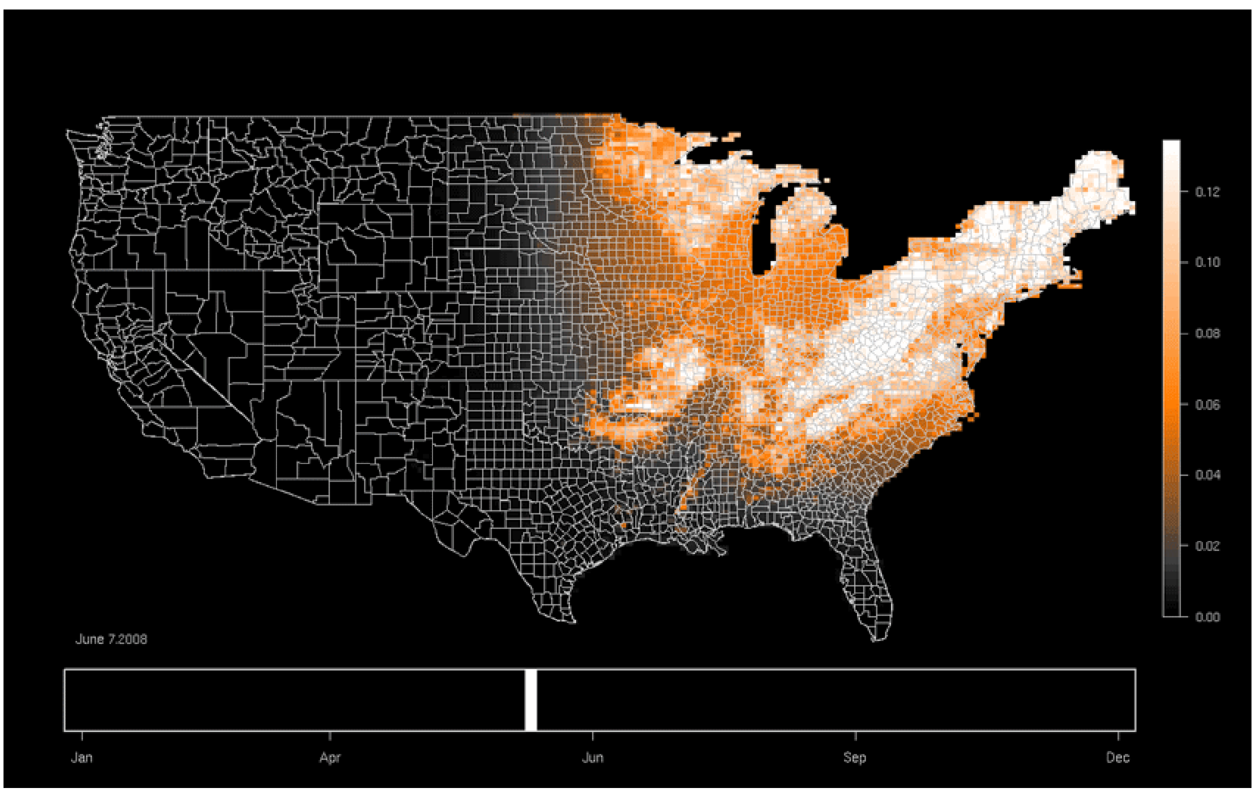


Figure 9: Animated occurrence maps are available on the eBird site for some species. This dramatic map shows the checklist sightings of scarlet tanagers by all contributors at a point in time between May and June of 2009.

million users in its first sixteen days of operation (Bullas, 2011). Many citizen science projects are leveraging the power and ubiquity of these social networks to engage volunteers, validate data, encourage outreach, and cultivate new content.

Social media can facilitate a greater sense of community among citizen scientists, even when they are geographically dispersed. The National Phenology Network and eBird are two examples of projects that host both Facebook and Twitter accounts for peer-to-peer communication and project outreach activities. Facebook, Twitter and Google+ offer several methods for integration with citizen science projects:

- **Authentication:** The Facebook, Twitter, or Google login can be used for secure user registration and authentication. Citizen scientists simply use their social media user name and password to log in to a citizen science project. The login is consistent across web and mobile devices. Project coordinators can also use the login to request permissions to access user profiles and "friend" lists to support outreach and recruiting efforts.
- **Social Plugins** can be used to add value and content. The familiar Facebook "Like" and Google "+1" buttons enable users to share their favorite citizen science web pages with friends; the Activities Feed enables project coordinators to display a stream of recent likes and comments from participants; the Recommendations Plugin enables participants to share personalized page recommendations with other members of the community.
- **Sharing** activity with Pinterest, Twitter, Facebook and Google+ is a common way to celebrate one's contributions as well as engage others with the activity. This can both build engagement and support outreach activities.
- **Analytics:** Both Facebook and Google offer sophisticated web site visitor analytics tools that can provide detailed analyses of user community demographics for project coordinators. These can be used to plan outreach activities and support recruiting efforts as well as make contributions to social science research.

Linking to external social media content can also be a simple way to increase available data for scientific research without the need to increase hosting capacity. For example, iNaturalist can link directly to a user's Flickr or Picasa account to add his/her photos to other observations on the iNaturalist site. Project Noah supports links to YouTube or Vimeo content.

The use of external authentication mechanisms as the only means of identifying users can also present problems, however. While Facebook is broadly used by many people, some people have a great deal of unease regarding the service's privacy-related behavior and similar concerns have been stronger in European countries and will continue to present challenges. We recommend that citizen science data collections provide the ability to leverage these social networks, but also provide a conventional registration and login for those who have opted out of the large social networks.

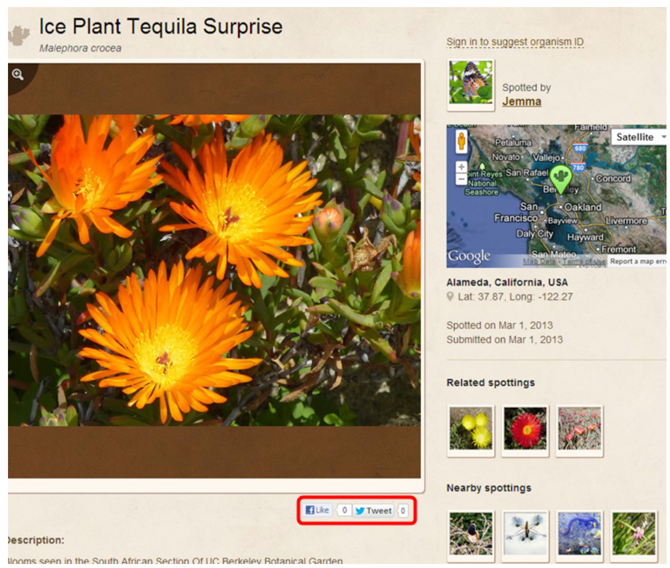


Figure 10: The detail page for the Ice Plant Tequila Surprise displays social mechanisms that include links to related and nearby spottings, and the opportunity to tweet or like the image.

Social Mechanisms

A recent study by Boston University (Nadkami and Hofmann, 2012) described two primary reasons why individuals use Facebook and other social media: (1) the need to belong to a larger group and (2) the need to present oneself to others in a positive way. Facebook satisfies these needs through its user profile pages, groups, and the ability to "follow" others. These social mechanisms can be applied to citizen science projects to increase community engagement and encourage ongoing participation.

Of all the projects evaluated for this paper, Project Noah is perhaps the most successful at integrating Facebook-like appeal into its citizen science work. Volunteers are able to create their own profile pages that will be populated with images of their most recent animal and plant observations. They can also add links to personal websites, blogs, or other related content. Volunteers can “follow” each other’s pages and click through the individual images there to a separate detail page. Once there, they can “like” or comment on the image, as shown in Figure 10. Volunteers are also able to view and follow lists of nearby or related observations by others that will help them learn more about the world around them. In addition to being part of the larger Project Noah community, volunteers can sign up for specific “missions” that interest them. Missions may be regional or species-specific in scope, and generally involve a dedicated subset of volunteers with a shared focus or interest.

It is important to note that these same social mechanisms are also being used to validate data on Project Noah. By clicking “Help me ID this species” when submitting a photo, volunteers are able to solicit help from the larger community and pool their collective knowledge to identify or verify their observations.

Gamification

Gamification is the concept of applying leaderboards, scoring, rewards, teaming, and other game-like features in a non-game environment to encourage competition, engagement, collaboration, interaction, and other gaming behaviors. The reasoning behind gamification is simple: if a non-game project is perceived as a game, it is more likely to attract and retain motivated participants just as an actual game might do.

Many projects institute some type of rewards system as part of a gamification and user engagement strategy. In a Syracuse University survey of 77 small to medium-sized citizen science projects (Wiggins, 2009), rewards ranged from role advancements to tee shirts and other promotional items, with public acknowledgment being by far the most prevalent. Nature’s Notebook, a National Phenology Network project, addresses the issue of public acknowledgment with a system of leaderboards: one leaderboard provides the first names and home states of the top 100 contributors since the inception of the program, while other leaderboards celebrate top contributors on a weekly, monthly, or annual basis.

There are potential concerns about some of the most common gamification features, however. Leaderboards, for example, have been found to motivate the people at the top of the board but demotivate the rest of the user community. Some research has suggested that “laddered” leaderboards that only show the few spots above and below the user mitigate some of this effect, but some gamification experts are suggesting that leaderboards have little net effect on user engagement.

For any citizen science project introducing a leaderboard or other scoring system, consideration must be given to the possibility of unfair “gaming” of the system through the use of “bots” – automated tools that could be used to carry out repetitive tasks, such as invalid data entry, to increase a participant’s ranking. A reputation-based system that enables users to earn points for dedicated entry while discouraging inappropriate, erroneous, or malicious entries is one potential development approach. It is also important to note that scoring may lead to disincentive in some cases, particularly if new volunteers feel that they are unable to compete with individuals that are more knowledgeable or have more time invested in a project. Resetting the leaderboards on a weekly or monthly basis can help alleviate this concern by giving new or infrequent participants repeated opportunities to score.

A digital badge or patch is another online acknowledgment of a volunteer’s skill or accomplishment that can increase gaming behavior. Project Noah currently issues citizen science merit patches, as shown in Figure 11, and the National Phenology Network is planning to do so in the future. The OpenTreeMap project also has a reputation point system that will be implementing a badging framework in early 2014. The Mozilla Open Badge Infrastructure is taking the concept of badges and recognition one step further by enabling learners to collect badges from multiple educational projects and share them with the world through social networking channels and other personal sites while also providing a secure verification framework.

Open Badges are currently being issued by schools and universities, community and nonprofit groups, government agencies, libraries and museums, and a range of other organizations. Since the program is not proprietary, any citizen science project could potentially create, issue, and verify digital badges, and any citizen scientist could then earn, manage, and display them to validate skills and competencies across multiple, diverse projects. SciStarter is currently evaluating the potential to combine Open Badges with an identity management system across projects and platforms.

The most successful efforts to integrate game-play into a citizen science context have relied on a combination of UI/UX design, careful attention to the learning goals, and a focus on motivational structures that enable users to gain new knowledge and new skills. We believe that scores, badging, quests, and "leveling up" will be an important and fruitful mechanism for improving user engagement and a unlocking a significant amount of "cognitive surplus" (Shirky, 2011), but we also recommend that the process be treated in a deliberate manner that integrates gamification design professionals, rather than slapping on a leaderboard and "badging up" the project and then crossing gamification off on the todo list.

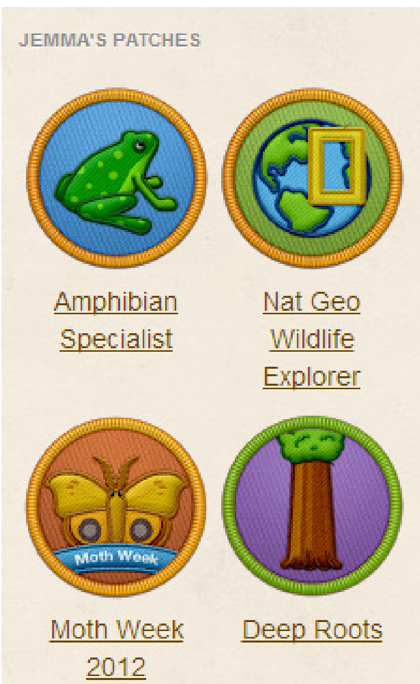


Figure 11: Examples of patches awarded to Project Noah participants.

Financial Incentives

One of the primary advantages of citizen science for project coordinators is the ability to tap an eager and dedicated group of individuals that will freely volunteer their time and effort to advance scientific research. Indeed, monetary incentives are contrary to many of the key motivations for volunteerism and citizen science, including altruism, guilt, ideological concerns, and simply having fun. The only exceptions to this seems to be the "challenge" projects, wherein a substantial monetary award is offered for solving a problem or submitting the most data within a pre-specified period. The 2009 DARPA Network Challenge, which explored the roles of the internet and social networking in collaborative problem-solving, is an example of this type of project. To mark the 40th anniversary of the internet, a $40,000 award was offered to the first team to locate ten randomly-placed weather balloons moored at fixed locations across the United States. A team of faculty, researchers, and students from MIT's renowned Media Lab ultimately identified the locations of the ten balloons in less than nine hours. The DARPA Network Challenge showed the world that with the proper incentives, whether they be financial compensation, intellectual and academic pursuit, or fostering a greater public or community good, today's "networked society" is indeed eager to come together virtually to solve a challenge.

There are a number of crowdsourcing efforts unrelated to citizen science that offer financial incentives to participants. Most notably is Amazon's Mechanical Turk, which bills itself as a marketplace for work that requires human intelligence, as shown in Figure 12. While not adapted to the needs of scientists – the service is primarily used by commercial firms that need to have data entry and visual image-tagging performed – it is an important analog as it incorporates several features that a science-focused system would require, including tools for designing tasks, a scalable infrastructure, and a mechanism for testing participants for skills and knowledge before they can participate. Tasks range from writing short articles to classifying Arabic text by dialect; "workers" complete the tasks for a small, pre-designated payment. Thousands of tasks are posted at any given time, and remuneration ranges from a few cents to about $10 per task. Projects like Field Agent and GigWalk have a similar business model, but are focused on tasks that can be completed with smartphones.

An important concern about including financial incentives as part of any citizen science business model is the potential for participants to “game” the system by completing as many paid tasks as possible without fully engaging in them (Downs et al, 2010). This has been the case with Amazon’s Mechanical Turk. Citizen science projects that offer financial or other incentives often need to add quality assurance measures to protect against bad data. Other strategies include pre-screening participants or using aggregated data from multiple sources to increase accuracy levels. Further, we see great value to using monetary prizes for projects that may require long periods of engagement in order to solve a problem. New services, such as Kaggle, InnoCentive, and ChallengePost, are being developed to enable sponsors of complex data science tasks to globally recruit teams of scientists, statisticians and software engineers to solve problems. To date, these efforts have largely supported improvements in proprietary commercial problem sets, but we believe there is significant opportunity to reward solutions to complex scientific, engineering and math problems by recruiting teams of people that work toward a prize. There is long history of this type of incentive structure and recent efforts, such as the Ansari X Prize series, are raising the profile of this type of approach, but we believe there is potential for smaller scale prizes to incentivize development of solutions to difficult science and civic problems.

Figure 12: From the Amazon Mechanical Turk Home page, where financial incentives are offered for completing small, repetitive tasks. .

None of the projects we evaluated offer financial incentives to engage their participants, but some examples of these smaller scale challenges include cash prizes offered by the recent MyHeartMap Challenge, a project developed by the University of Pennsylvania School of Medicine to develop a better map of Automatic External Defibrillators (AEDs) in the Philadelphia region.

Other Considerations

The lack of a unified authentication system that can be used across multiple, diverse projects, regardless of domain or originator, is a key consideration for citizen scientists as well as project coordinators. Other considerations include recruiting practices, IRBs, privacy polidies, and options for volunteer training.

Authentication

A single authentication system that can be shared across many projects is far more convenient for citizen scientists than having to maintain separate logins for every project. The Cornell Lab or Ornithology team is addressing this issue by enabling its participants to create a single user account that can access any of the Lab’s citizen science projects, including eBird. However, these user accounts are exclusive to Cornell projects and cannot be used, for example, to access a citizen scientist’s account on WildKnowledge or Project Noah. Similarly, Zooniverse, CitSci.org and iNaturalist also provide a single login for projects on their platforms but do not provide a mechanism for federating authentication across multiple platforms or for an ID on one platform to be used on another.

Beyond the obvious potential of social networking to build online presence and reach new audiences, Google, LinkedIn, Twitter, and Facebook make available authentication systems that would enable use of a Facebook or Google account to be used to create an account with citizen science projects. Broader use of these trusted systems could provide a cost-effective solution for project administrators that need to engage large user communities and validate contributed data. Further, it is important to note that there are existing open standards for authentication, with OAuth being the most common.

Contributor Ecosystem and Recruiting

How can researchers successfully recruit citizen scientists for their projects? A dissertation on citizen science from the University of California at Berkeley (Robson, 2012) found that recruiting volunteers through social networks, including Facebook and Twitter, was just as successful at increasing user registrations as recruiting through traditional media channels, such as press releases and news articles. By starting their recruiting campaign early, Berkeley had a ready contributor base well in advance of releasing its new CreekWatch water quality mobile app. SciStarter recruits participants, in part, by serving up projects aligned with popular activities such as hiking, fishing, or spending the day at the beach. These promotions take place via Twitter, Facebook, on SciStarter's website and via its distribution partners including Discover magazine and the National Science Teachers Association. The Cornell Ornithology Lab finds recruitment more efficient when solicitations are directed at an audience with a pre-specified interest in the study topic, which, in their case, is birding.

The ability to recruit participants from a broad pool of individuals who have already expressed interest in a particular type of project, activity, or subject matter has considerable appeal for both project coordinators and potential volunteers. For coordinators, it provides the ability to target limited recruiting resources to optimal effect. For volunteers, it offers advance notice of new projects in multiple interest areas from a single source without having to search multiple sites.

We see an enormous opportunity to ease recruitment challenges for smaller projects by providing a ready pool of motivated participants and making recommendations to potential contributors based on past activity. This type of proactive recommendation system, driven by user activity and Bayesian statistics has the potential to apply some of the latest practices in the private sector to make major advances in citizen science.

Institutional Review Boards (IRBs)

An institutional review board (IRB) is a committee formed to approve, monitor and review biomedical and behavioral research involving humans. Their primary purpose is to protect human research subjects from physical or mental harm. They provide oversight to ensure that research that works with human subjects is ethical and weighs risks of harm against potential scientific benefit. While IRBs were developed in order to avoid past mistakes and abusive treatment of human subjects in medical trials, they have also received a great deal of criticism for both being too lax, on the one hand, and for being too restrictive by requiring them for scenarios for which they were never intended. For example, some institutions have interpreted requirements for IRBs to be applied for opinion surveys, oral histories and unstructured interviews, well beyond the original intent aimed at biomedical and psychological treatment.

IRBs are relevant for the design and development of a citizen science data collection platform because some researchers may be required by their institutions to incorporate the data collection process in the IRB review and to treat contributors as human subjects. There are also several exemptions under which IRB rules do not apply. This is a complex topic, and this report will not address this question in greater detail, but we note it here as an important consideration for some citizen science projects, particularly if they involve collection of data about human health or mental state or are funded by government agencies that may require IRB review.

Privacy

Contributions to citizen science projects are an active engagement and production of new information. As people move from passive consumer to active producer, the way in which their location, date/time and other sensor data is used becomes more important. Critically, sharing that information with a researcher may be central to the science question. But there is a fundamental conflict sharing data and protecting privacy. These concerns will become increasingly important as more people engage with citizen science and the importance of online privacy grows.

Most citizen science projects address privacy at some level and many have an explicit privacy policy. This is usually adequate for many projects, but there are some important exceptions. As outlined above in relation to IRBs, data collected about human health may require anonymous contributions, encryption, and other measures or else be explicitly excluded from a generalized platform.

Contributions by children to citizen science data collection efforts are another important privacy concern. Children are often enthusiastic participants in citizen science projects, uploading data, taking measurements and interacting with others through online applications. In the United States online privacy for children is governed by the Children's Online Privacy Protection Act of 1998 (COPPA). The law was passed to protect the safety of children by limiting the data that can be collected from children. It applies to children under 13 and outlines privacy policy, parental consent, and other requirements. While it has not been consistently enforced, it applies to some citizen science projects, particularly if they have a commercial component (non-profit organizations are exempt) or operated by a commercial entity. COPPA specifically governs "personal information" that the child may enter, including name, home address, email, telephone or other unique identifier.

A 2013 update to the law also added geolocation and photos, video or audio that contain the image or voice of the child. Many citizen science projects collect email addresses, name and age of volunteers, as well as location, photos and other data. This may be acceptable if the data is not published, but many citizen science projects value openness and try to share their data broadly as well as recognizing their volunteers by name. Most citizen science web sites do not address this and welcome contributions from any age group. A small number, such as BudBurst, are more careful and specifically require that registered users certify that they are 13 or older. The BudBurst project takes additional steps as well. While public contributors need to be 13 years or older, it provides an ability to set up an "Educator" account that is able to add accounts for younger students as well as resources and curricular materials for educators working with different age groups.

The global nature of some projects also raises privacy concerns. The privacy requirements in the European Union, for example, are significantly different from the United States, where a self-regulatory approach has generally been preferred. In order to support collection of data within the EU, it may be necessary for a US-based project or platform to seek certification under the US-EU Safe Harbor program. More broadly, there are significant differences in the cultural norms regarding privacy in different countries, and citizen science data collection platforms may need to begin addressing this question in a more comprehensive manner in the future.

Finally, recent disclosures about government surveillance practices and active collaboration by commercial organizations to support mass surveillance are raising the level of both public anxiety and discourse about the appropriate limits of privacy. As is often the case, legislation significantly lags changes in technology and near term potential for even more invasive surveillance is frightening. Future online citizen science platforms will need to be both mindful of this important issue and take steps to articulate clear policies and enable researchers to protect the privacy of contributors when appropriate. There will inevitably be tradeoffs between other valid objectives. For example, the ability to cultivate a community, engage contributors in a discussion, and develop a social network may be in direct conflict with a desire to preserve the privacy of contributors. We believe this will be an important and evolving concern over the next decade that will require careful technology design as well as crafting of new policy as our society moves toward a new consensus on appropriate levels of public disclosure and privacy in these new contexts.

Training

Volunteer training can be an important means of ensuring data quality in citizen science projects. Indeed, an in-depth analysis of sampling protocols in data collection projects demonstrated that citizen scientists can be as accurate as professional scientists when collecting data, as long as they are given appropriate training to guide them (Nerbonne et al, 2008). Training can also instill confidence in new participants and help them increase their skills and knowledge.

All of the projects evaluated for this paper provide some type of training or supporting materials: methods ranged from workshops and online tutorials to field guides and train-the-trainer sessions. For example, CitSci.org offers a tiered train-the-trainer model for all new project administrators, while Indicia offers a forum and chat room. In its model for developing a citizen science project, the Cornell Lab of Ornithology ranked training participants as the fifth step in a nine-step process that also included visualizing data and disseminating project results (Bonney et al, 2009). Cornell projects, including eBird, offer extensive training materials that include recorded bird sounds, regional field guides, and in-depth project tutorials. Other examples include Zooniverse, where each project begins with a short training exercise.

In-Depth Evaluations

Six citizen science projects and platforms were selected for in-depth evaluation. These projects were selected because they are well documented in the literature and/or have particular utility, interest, or audience appeal that can inform future development, as noted below:

1) **CitSci.org** – Unlike many contributory data projects and platforms, which share all of the data collected in a single map, CitSci.org is focused on supporting individual projects rather than aggregating data. There are two primary implications of this model: first, there is much more customization of data collection forms; second, there is a greater focus on engagement and management of project participants though direct email and feedback forms. In addition, it offers some rudimentary ability to create custom charts or graphs of project data.
2) **eBird –** This is one of the most mature citizen science projects, and the longevity of the community might offer some unique insights. Unlike some of the mobile-focused applications, users can submit data in multiple formats including web forms and uploading spreadsheets. In addition, multiple third-party mobile applications have been built around the import and export of eBird data, which appears to be unique among citizen science platforms. Additionally, eBird offers some of the more sophisticated data visualizations, perhaps because of the limited scope of the data types it handles.
3) **Indicia** – An open source project dedicated to species identification. It requires more developer resources to implement, but offers significantly more customization. It also offers the iRecord tool, which allows projects to launch more quickly. It would be helpful to understand how they decided what features to offer in this version.
4) **National Geographic FieldScope** – FieldScope was initially developed as part of a general-purpose GIS designed for use in schools and universities. It has since evolved into a platform to support "community geography" citizen science projects. Members can upload georeferenced data and use FieldScope's web-based GIS tools for data visualization and analysis, so it may provide insight to inform data use.
5) **National Phenology Network** – This application is aimed at phenology data, and therefore offers considerably more attribute options given the variations in life cycle data by species. It may present a good use case for understanding how data can both be diverse and conform to standards. The visualization tools also offer some of the most sophisticated time-series visualizations.
6) **WildKnowledge** – WildKnowledge identified the engagement potential of mobile devices long before smartphones were truly popular. They have since created a suite of mobile toolkits that can be used to build customized data collection projects to record information in the field. They offer both free and subscription-based services.

Each project or platform was evaluated based on the following metrics, as further outlined in our Methodology section:

1) **General –** What is distinctive about this project? If it were completed all over again, what would be done differently?
2) **Flexibility** – Is the platform focused on a few projects in a single domain, or can it accommodate data from different types of data collection efforts across many domains?
3) **Display, visualization and publication** – Is it possible for participants to see the results, or are data submissions a one-way street?
4) **Technology platform** - Is it an open platform that others can build upon? Or is it a closed, proprietary system? Is the source code made available? Are there technology gaps or challenges faced by the project team? Is there an open and documented API?
5) **Social, marketing, and Incentives** – How are participants nudged toward participation? Are incentives offered?
6) **Data quality** – How is data quality ensured or quantified? Is it an automated or manual process?
7) **Cost for a new project** – How easy is it for a research project to get up and running on the platform? How many hours or dollars will it take?

Telephone interviews with the administrator(s) of each project were supplemented with a review of the literature and a hands-on evaluation of the project website from both a citizen scientist's and project administrator's perspective.

Data Collection Projects

Within the National Science Foundation typology, the six projects we selected for in-depth evaluation are "contributory data collection" projects. As defined in the 2009 study by the Center for the Advancement of Informal Science Education (CAISE):

"Contributory projects are researcher-driven data-collection projects. Scientists ask questions for which answers require the accumulation of large amounts of data collected over wide geographic areas and/or over long spans of time, and members of the public collect relevant data following protocols predetermined by the scientists."

Most projects currently labeled "citizen science" fall into this model (Bonney 1996; Krasny and Bonney 2005; Bonney 2007).

CitSci.org

CitSci.org is a website supporting citizens who monitor invasive species. CitSci.org provides a mapping program that allows citizens, school groups, and professionals to enter invasive species observations into a global database. The observations are then used for natural resource management, scientific studies, and environmental education. CitSci.org offers an opportunity for students and volunteers to perform field studies that contribute to its collective biological databases. The data is made available for analysis to answer local, regional, and/or global questions.

General

The project was funded by the National Science Foundation as part of a three-year study that is now in its third year. Its goal is to enable and empower citizen science projects and help organizers go through the scientific process. The administrators were determined to emphasize flexibility for researchers without becoming a "willy-nilly hodgepodge of observations". However, not all scientists use the same protocols; they have different methods, transects, and shapes. This can make it difficult to merge data from one effort with another.

CitSci.org is providing a "growing yet vetted set of measurements", such as the number of black spots on the right wing of a fly, or the percentage of woody versus herbaceous land cover. Administrators offer a tiered train-the-trainer model and will meet with project coordinators to help them organize their projects or point them to helpful resources on the site. For example, they have created a list of measurable attributes for both species and site characteristics and work with project coordinators to select those most suited to their projects. The goal is to provide detail while still making the attributes general enough that they can be used in multiple projects: using "sex" versus "sex of fox" or "sex of bear", for example. Trainers will also help project coordinators determine what attributes they want to measure, such as the length of a fox, or the temperature of a stream, in order to further their project goals.

Citizen scientists really want to know what is being learned by the data that they are contributing. To that end, CitSci felt that there were several improvements they could make to the project that would make it more appealing to participants:

- **Better analysis and visualization:** Currently, the site offers a simple pie chart to visualize data. A feature listed as "coming soon" will let citizen scientists pick a dependent and independent variable and visualize it in the chart type of their choice. For example, visualizing the percent of nitrogen through time, or the influence of water temperature versus extent of slope.
- **Day-to-day coordination of volunteers:** Another idea for the future is to offer a calendar of events, such as upcoming training sessions. Links to YouTube videos, PowerPoints, and other content are currently not available, but these features are all listed on the site as "coming soon".

CitSci.org felt that its initial design for the project limited organizers to a single survey model. If they could redesign the site, they would go beyond a simple date/time/coordinates data entry model to include options for defining a monitoring structure and applying it to multiple geographic locations. In that way, participants could easily enter data from multiple locations as part of a single project by just selecting them from a dropdown menu or picklist. Similarly, for citizen scientists who are consistently doing backyard monitoring work, it would only be necessary for them to enter their geographic coordinates once rather than having to re-enter them each time they submit an observation.

Administrators would also like to enable participants to do more things directly from their profiles. Currently, participants registered on the site can individually request permissions to join individual projects from their profile pages. They can also save data sets, downloads, jobs, and locations of interest there. However, they are unable to connect to Google Calendar to note dates and times for projects of interest, or fill in forms using Survey Monkey. Ideally, CitSci.org would like to build service-based widgets that participants could embed on their profile pages to provide customized access to features they regularly use. However, since the site was built using custom code, improvements of this magnitude would tend to be costly and not easily made.

CitSci employs grant funding to support the sustainability of the project. Its ecological subject matter has made it possible to obtain funding from multiple agencies, including the United States Department of Agriculture and the United States Geological Survey. Other partners include NASA. Grant funding has been used both to add new features and to underwrite simple deployments.

Flexibility

Many project coordinators want to add legacy data as well as newly-provided participant data to the database. CitSci.org enables batch uploads of data, but administrators note that it requires a great deal of patience. The system parses the file, but the submitter needs to identify the attribute type for each column in the file. In the future, CitSci.org would like to enable standard tab-delimited data.

The system converts picklists to checkboxes and left justifies forms when they are printed. The sPDF library is used to convert database-driven content to PDF. Approximately 60 to 80% of all projects on CitSci.org originate with static paper forms in the field. Participants then re-enter the data from their static forms into the online system and send their original forms to project coordinators, who will use them to verify the entries.

CitSci developed Palm and Windows PDA applications for participants, and the PDA application is still available through the site. However, they are abandoning these efforts in favor of contemporary mobile applications. A friend of the project is developing an iPhone application at no cost, and a student is using $5,000 of NREL funding to develop an Android application. There is no timeframe noted for when these applications will be available.

On the subject of flexibility, a comment from our citizen science forum indicated that CitSci.org did not provide the flexibility desired when formatting the data entry page.

Display, Visualization, and Publication

As shown in Figure 13, data visualizations are provided through canned pie charts generated by jQuery. These can be accessed through the "statistics" button on each project's profile page. A new customizable charting system is currently under development that will enable coordinators to design their own bar charts.

A custom-built desktop application provides maps using Google Maps or satellite imagery as a base map. The application is Java-based and leverages work by Oregon State. It displays United States projects only. Maps can be used to visualize point data for all projects collectively, or participants can select a specific project and view detailed data for that project only by clicking through to its profile page. Participants selecting the "add a point" option will

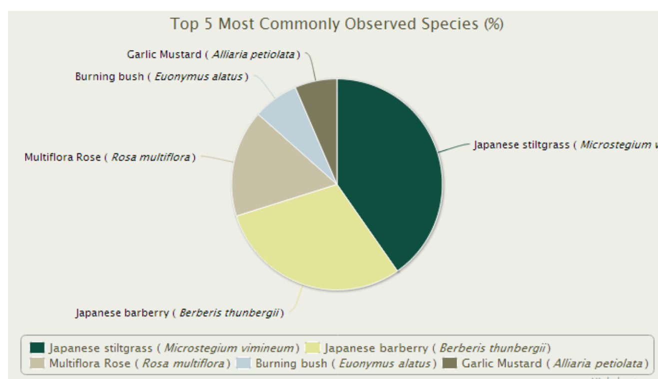


Figure 13: An example of the CitSci.org pie chart for an invasive species project.

access a survey form that enables them to add new data directly to the map.

As shown in Figure 14, the Advanced Map, which provides a global view, includes links to each project's profile page directly from the legend. Administrators indicated that map tiles are only re-rendered when new data is added in order to enhance performance; however, our hands-on evaluation noted that it took a considerable amount of time for the Advanced Map to load on each occasion. Several tools are provided for navigating the map.

Data is made available through a Creative Commons License in text, CSV, or shapefile format from each project profile page. Images are available in JPG, PNG or GIF format.

Technology Platform

The main development focus is currently on engaging and retaining volunteers. Project coordinators can send e-mails to their participants directly through the website, reminding them, for example, to bring their GPS to a particular project location. While coordinators can see their total list of members and pending members, they cannot view information on recent activities or number of submissions.

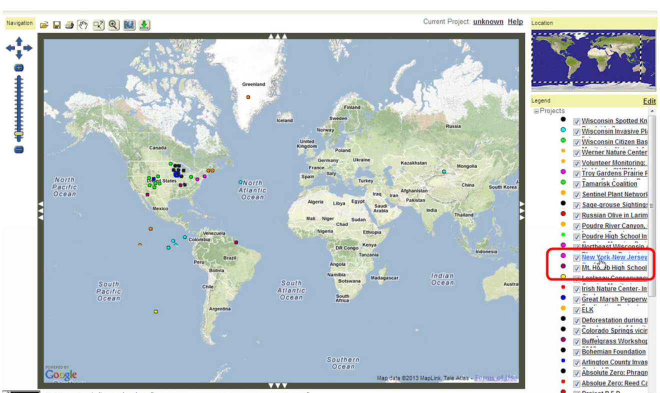


Figure 14: The Advanced Map interface shows the location of each project and provides links to project profile pages through the map's legend.

APIs are available for external developers that want to integrate data from the site into their own applications. Currently, these APIs are prototypes that support RESTful web series delivering JavaScript Object Notation (JSON) data. In the future, they hope to offer both XML and JSON Support.

The first iteration of the website was the National Institute of Invasive Species Science: www.niiss.org.

The project was originally built in 1991 using ASP, a predecessor of ASP.Net. A back-end interface with a GUI was developed for adding webpages or menu items. It didn't require programming experience for administrative users.

In 1995, the site was migrated and rewritten in PHP, following a "make versus buy" evaluation of ASP, JSP, and PHP that looked at overhead, speed of development, and speed of deployment. For the database, Jim Graham came from Hewlett Packard and did a database comparison: he dumped millions of records and compared Microsoft SQL Server, Oracle, MySQL and Postgres. They decided on Postgres and SQL Server because of their spatial data support and well-known text.

The site currently uses Microsoft SQL Server 2008, but there are plans to migrate to a newer version in the future. ODBC (Open Database Connectivity), a standard C programming language middleware API, is used to access the database management systems from PHP. Colorado State University, where the project originated, had a Microsoft license, which helped sway developers away from the LAMP model they had been previously considered (LAMP is the acronym for Linux operating system, Apache HTTP Server, MySQL database, and PHP, Perl, or Python components). All the code is open source. Jim Graham, now with Oregon State University, has integrated the code with the Postgres project.

Development was completed by a small team with just a few staff. The National Science Foundation funded graduate researchers to assist with the project, and development was completed using a cyclical, waterfall-style approach rather than Agile methodologies. The team developed for three weeks at a time, followed by a week for testing. Upon completion, a formal usability study was done in the usability lab at Colorado State University. Two graduate students and two full-time developers currently maintain the server and source code. They try to integrate user feedback on an ongoing basis, and they are moving toward a service-based architecture.

Social, Community Building, Marketing, and Incentives

While CitSci.org does not currently provide an incentive system, it does take advantage of social networking tools for project outreach and recruiting, as shown in Figure 15. Users are able to create their own profile pages complete with photos of themselves. They can join individual projects directly through these profile pages, and when they submit data, it is attributed to the user name from their profile.

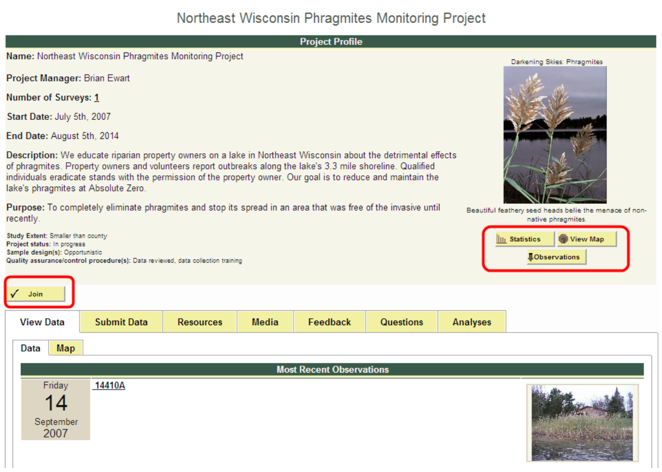


Figure 15: An example of a project profile page shows mechanisms for joining the site and accessing project data.

Each individual project also has a profile page where users can view or download data, or click to join. A CitSci.org Facebook page and Twitter feed disseminate pertinent information to current and potential users. CitSci.org also links to Flickr for access to image data. For recruiting purposes, the site offers project coordinators the ability to approve new members, view current members, and edit user roles, as well as send e-mails to their members. CitSci.org also enables project organizers to share project descriptions with the Project Finder on SciStarter.

Data Quality

Data quality is ensured through a variety of means. Authentication is required before a participant can submit data. Each participant logs into his or her account, selects a project to contribute to, receives authorization from the project coordinators, and can then submit data using the customized form created for the selected project. Very few participants contribute to multiple projects: perhaps 5% at most.

Project coordinators are responsible for performing their own data validation. This can be facilitated by setting the attribute fields on the data entry form to accept minimum/maximum entries, and by following guidelines on the website to determine whether stratified random sampling or opportunistic design is best suited to a particular project. Coordinators can also check a box to post an observation on the iNaturalist site and use that community to verify the species identification provided. The APIs can communicate with each other and send back verification once a designated number of participants have vetted it.

The Creative Commons License currently used in the site is relatively new. A custom set of data usage terms had been employed prior to that. There is no indication how the data is being used by others. Site administrators believe that they will have collected enough data within the next two years to support meaningful academic publications.

Cost for a New Project

Project coordinators need to contact site administrators at CitSci.org about creating a new project, and then work with them to determine measurable attributes and customize their data collection forms. Launching a project on CitSci.org is totally free, but if project coordinators want custom skins or new features, there might be discussion about securing grant funding to pay for these.

eBird

A real-time, online checklist program, eBird has revolutionized the way that the birding community reports and accesses information about birds. The focus of eBIrd is on bird distribution patterns across broad temporal and spatial extents. Launched in 2002 by the Cornell Lab of Ornithology, which also hosts a number of other ornithology projects, eBird provides rich data sources for basic information on bird abundance and distribution.

General

eBird began as a typical citizen science project, with researchers interested in getting birdwatchers to help them gather data to save birds. The initial project motto was, "Birding with a purpose." Rick Bonney, Director of Program Development and Evaluation at the Cornell Laboratory of Ornithology, has written several papers distinguishing eBird, a contributory data collection project that gets volunteers to actively go out into the environment to make observations, from projects like Zooniverse, which are collaborative data classification projects. Data collection was the major focus of the project between 2002 and 2003, using funding from the National Science Foundation. During that period, only a small volume of data was coming in. Cornell understood that if they were going to collect the large volumes of data needed for analytical research, they were going to have to take a different approach. They hired Chris Wood and Brian Sullivan, who were leaders in community, and worked with them to develop tools, features, and functionality that would be appealing to amateur ornithologists. Rather than just a place to collect data, eBird became a site where birdwatchers could create life lists of birds, compare observations with other birders, and see their names and observations on a map. eBird continues to update and improves its offerings to increase public participation.

eBird encourages year-round observations and repeated surveys from its participants. Related ornithology projects by Cornell include BirdSleuth, Celebrate Urban Birds, Great Backyard Bird Count, Project FeederWatch, NestWatch, and YardMap.

Flexibility

All tools and functionality available through the eBird project have been developed to appeal to beginning, intermediate, and advanced birders in multiple geographic locations. The eBird website is available in English, Spanish, Portuguese, and French. Participants can geographically submit data through the website by selecting from existing birding hotspots or adding new reporting locations to a global map.

Birders can also review or submit data through a pair of mobile applications by BirdsEye Birding. The BirdsEye mobile app, shown in Figure 16, leverages eBird data to help participants locate ideal birding sights near them, and even plan remote birding expeditions. The BirdLog app is essentially a mobile data entry form that enables participants to enter bird sightings directly from the field. BirdsEye is optimized for use with iPhone and iPod Touch Platforms, while BirdLog is available for Android, iPhone, iPad, and iPod Touch. Cornell University worked very closely with the developers of these mobile applications, since project administrators are very careful about how data should be entered into eBird.

BirdLog was released in spring 2012, and feedback has been very positive. Approximately 10% of all new data in eBird now comes from the app. Birders can plot the exact locations of their bird sightings using their smartphone's GPS, share checklists with friends via e-mail, and even create new checklists offline when an internet connection is not available.

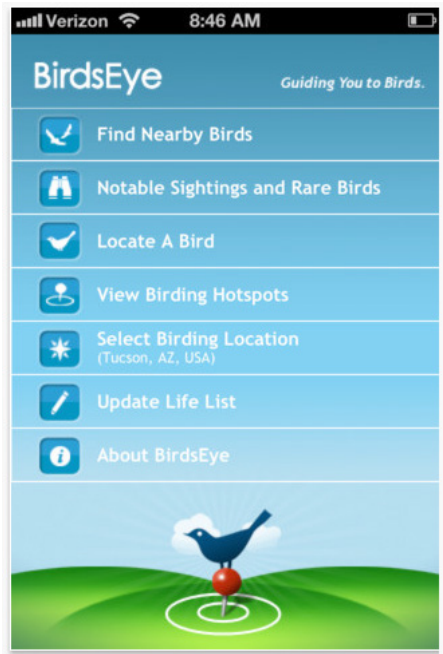


Figure 16: The BirdsEye app helps eBirders find birding hotpots.

Display, Visualization, and Publication

As illustrated in Figures 17 through 20, eBird offers several types of data visualizations, which are publicly available to anyone accessing the website. Birders can explore interactive range maps by species or sub-species and zoom in for greater detail. Bar charts of hot spots – important bird areas, or bird conservation regions – can be created to display species occurrences for a particular area of interest over the course of a year or range of years. Line graphs display bird observations by species over a designated data range, and include information on sighting frequencies, abundance, and

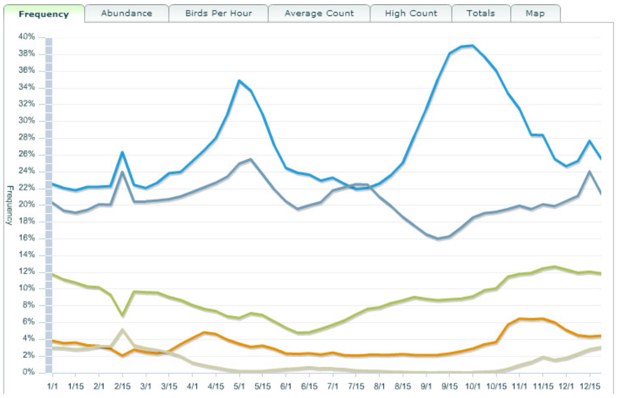


Figure 17: Line graphs display factors such as abundance and average count for up to five species.

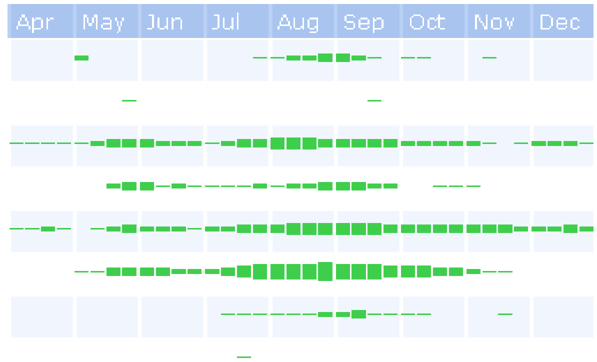


Figure 18: Bar charts display what birds to expect throughout the year in a region or location.

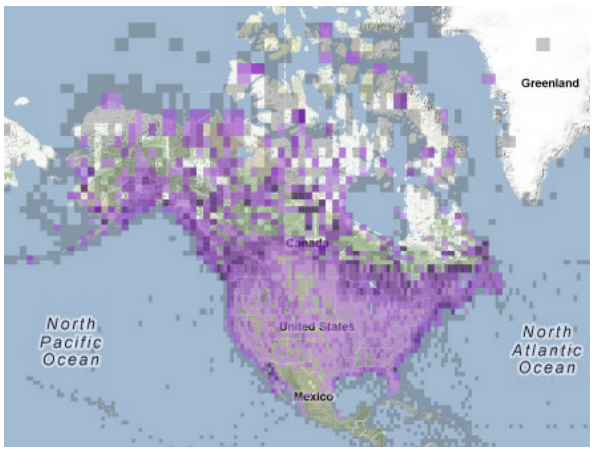


Figure 19: A range map displays the habitat ranges for the same five species featured in the line graph above (Figure 17).

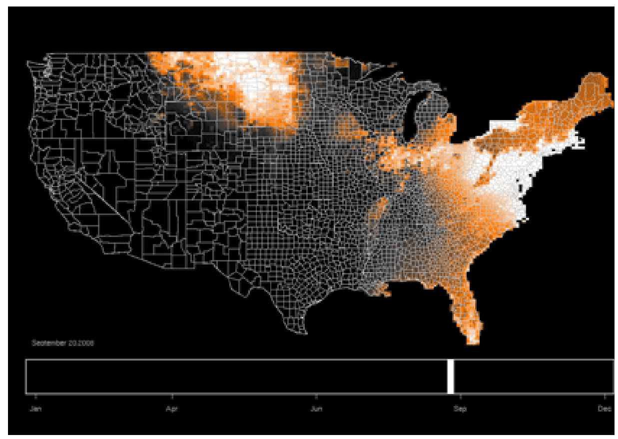


Figure 20: Animated occurrence maps are available for a variety of species and are based on aggregated checklist data.

average counts. A submission map displays real-time sighting data on a global basis. Animated occurrence or migration maps for the lower forty-eight states are also available for some species.

Each visualization results from a Java application using Oracle Spatial for pulling source data from particular geographic regions. More than one million observation locations are currently available, and that total is growing at a rate of about 30 to 40% annually.

The underlying premise of eBird is that all data for any species should be available for display. However, this is not always possible for private lands and may not be advisable for sensitive species. New infrastructure is being built into the site that will enable some data to be protected, particularly when it involves endangered or threatened species, such as spotted owls. This is not only important in the United States, but in other countries as well.

Approximately 60 to 80 publications have already resulted from eBird data, with subjects ranging from bird population to community development practices. All data and visualizations must be properly cited.

Raw data is available at no cost for non-commercial use, but potential users must submit an e-mail request with a description of their affiliations and proposed projects. They must also agree to eBIrd's terms of use. The project receives three to four requests each day for raw data. If the raw data is to be used for commercial purposes, a licensing agreement for revenue is required.

Technology Platform

An Oracle database has been used for the entire life of the eBird project. Because the project handles as many as 10,000 participant checklists per day, it requires an enterprise, business-level, high performance database to ensure that everything operates smoothly. One of the biggest challenges for the project is its content management system (CMS): more and more people are requesting their own regional portals, and the ability to remotely add information. The project is currently using the open source Plone CMS. Plone provides support for all major operating systems, which is important to eBird administrators. The project is actively looking at ways to switch to something that is more extensible and robust. Application development was completed with Java.

Another challenge involves data output. Hundreds of different customized views have been developed, and these are updated as frequently as every few seconds. eBird has approximately 100,000 total contributors and perhaps 15,000 "power users". Up to 250,000 IP addresses connect to eBird's data output each month. In past years, there have been as many as 6 million observations recorded annually on up to 2 million individual checklists.

The overall development team consists of database administrators, application developers, and web designers who divide their time between eBird and other projects. The mobile application development team was smaller, and consisted of one manager and one or two developers. Front-end testing for the mobile applications involved a group of "power users" who provided constant feedback on a beta version in order to inform its development.

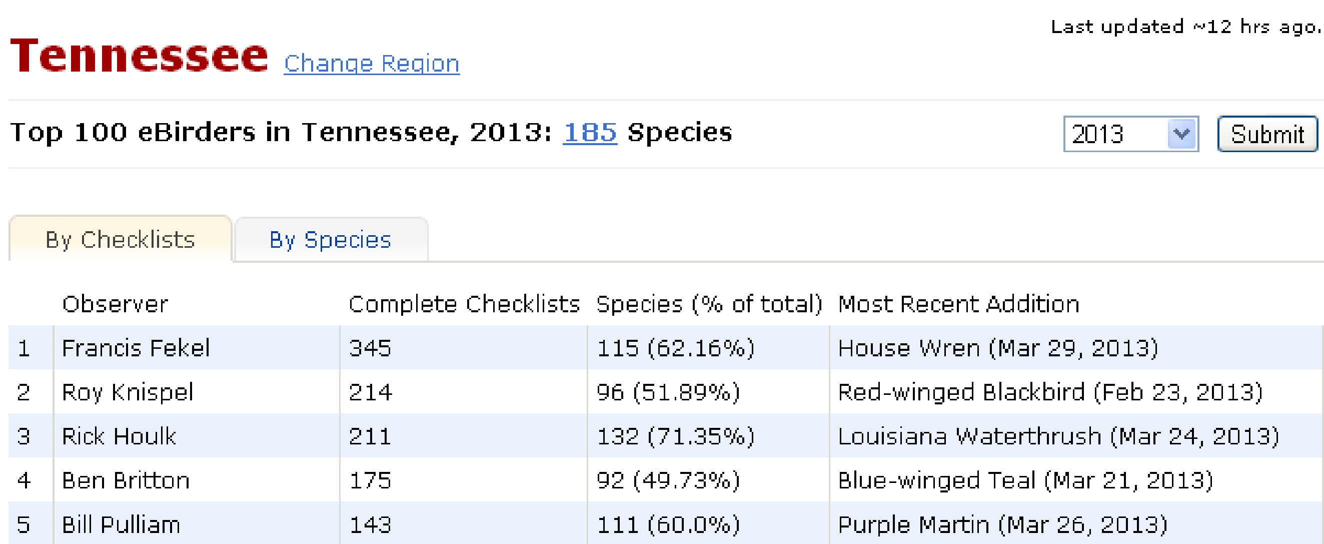
Social, Marketing, and Incentives

eBird provides a single account and authentication system that can be used for all Cornell Lab of Ornithology projects. A Top 100 list enables eBirders to compare themselves with the top eBirders in their geographic regions. The Top 100 is updated nightly and provides both species totals and checklist totals to encourage competition. eBirders can also keep track of the species seen in their favorite birding places, or sign up for e-mail alerts about rare species or species they haven't seen.

Birding tends to be highly competitive. Therefore, a strong incentive for eBird participants is the ability to see their names and birding observations on the website and to compare them with the observations of others. The eBird project does not currently have the time to implement additional incentives, such as badges, but organizers remain open to the idea and have discussed it in the past.

Data Quality

eBird uses a two-stage system for data quality assurance that includes both automated and manual methods. As birders begin the data entry process, a series of automated filters generates a checklist of species they are most likely to see based on their reporting date and location. Birders can simply enter counts for the bird species on their checklists, or add new species to their checklists based on unusual or unexpected observations. The


Last updated ~12 hrs ago.

Tennessee Change Region

Top 100 eBirders in Tennessee, 2013: 185 Species 2013 Submit

By Checklists By Species

	Observer	Complete Checklists	Species (% of total)	Most Recent Addition
1	Francis Fekel	345	115 (62.16%)	House Wren (Mar 29, 2013)
2	Roy Knispel	214	96 (51.89%)	Red-winged Blackbird (Feb 23, 2013)
3	Rick Houlk	211	132 (71.35%)	Louisiana Waterthrush (Mar 24, 2013)
4	Ben Britton	175	92 (49.73%)	Blue-winged Teal (Mar 21, 2013)
5	Bill Pulliam	143	111 (60.0%)	Purple Martin (Mar 26, 2013)

Figure 21: An excerpt from leaderboard of Top 100 birders in Tennessee. Top 100 lists can be generated by state or county, and viewed by checklist totals or species totals.

same automated filters that generated the checklists will then check the user-entered data against the database and flag potentially erroneous or suspect entries based on the following criteria:

1) Species rarity in a particular geographic region.
2) A species reported out of season for a particular region and date range.
3) High counts that exceed what a birder might see on a particular date in a particular region.

Flagged data are reviewed by a network of 500 regional experts, who will attempt to verify each observation. As part of that process, reviewers may request photographs or field notes from eBird participants to confirm a flagged observation. If there are no photographs or other supporting data available to confirm a sighting, the data is kept on file in the eBird database but is not made available for public display.

The provision of new tools, including the BirdLog mobile application for field data entry, has had a significant impact on both the quantity and quality of the data submitted to the site. In fact, eBird generated more new checklists in May 2012 alone than it did during its entire first three years of data collection. Because of its data quality practices, eBird data is now being used by the United States Department of the Interior, the Bureau of Land Management, the United States Forest Service, and other agencies.

The eBird project has received a new National Science Foundation grant that is funding a study of emerging techniques in data quality for data collection projects.

Cost for a New Project

In addition to the core eBird site, the eBird project hosts a number of regional portals that are specific to states or countries. These portals require an annual licensing fee. The BirdsEye and BirdLog apps require a fee from each user that is payable at download. Other than that, participation is free.

Indicia

Indicia is an open source toolkit that simplifies construction of new websites for recording wildlife data. Indicia was developed by the National Biodiversity Network and funded by Open Air Laboratories (OPAL) in the United Kingdom with grants from the Big Lottery Fund. It is part of an OPAL effort to get individuals to explore, study, enjoy, and protect the local environment. Indicia provides a set of services, tools, and examples that can be added to almost any website to provide rich online recording functionality.

General

Development has been primarily focused on the value of the data collected, and ways to mobilize the data more effectively for use by researchers. However, researchers were not drivers of the project originally. The Natural Environment Research Council has a Centre for Ecology and Hydrology, which, in turn, has a Biological Records Centre. The Biological Records Centre is committed to the Indicia project through at least 2017.

Indicia bills itself as a “kit car” as opposed to a manufactured car: all the pieces are provided to create a project website, but the project coordinators are responsible for assembling the parts. Individuals without coding skills can integrate it into an existing content management system (CMS) such as Drupal. Individuals with coding or website development experience can access provided programmer’s components to build their own data collection forms and otherwise customize their implementations.

Indicia websites consist of an online recording website and a data store known as Indicia Warehouse. Project coordinators can host their own Indicia Warehouse or use one provided by another organization, such as the Biological Records Centre. A new product, Instant Indicia, integrates with the Drupal CMS to enable coordinators to select from a list of features they want on their websites, such as input forms, reports, and maps, and quickly set up a website without having to perform any coding.

Future development efforts are centered on improving reporting and download capabilities, and making it easier and more cost effective for coordinators to set up forms so that they can draw summary charts or maps.

Flexibility

Indicia is currently web-based. There is no toolkit available for building mobile applications for any platform, and the web application is not mobile-optimized. Indicia administrators believe that a mobile application is essential for the ongoing success of the project, and they are actively seeking funding for its development.

A team at Bristol University built their own mobile application in 2012 that integrates with the iRecord implementation of Indicia. Their Plant Tracker application is available for iPhone and Android smartphones. It is used for tracking invasive plants. A new app for ladybird recording is in development by the same team and is expected to be released in Spring 2013.

Display, Visualization, and Publication

Because it is written in PHP, Indicia provides the ability to create point maps with symbolization as long as a query can be written. Indicia can also be integrated with a charting library to create line, bar, or pie charts of species data. The Indicia team did not try to engineer a complicated report-writing tool, since it would have been difficult to meet the needs of all users. The report grid control outputs tabular data that can include thumbnails of the photographs associated with each record. It also supports filtering and sorting. Report data can be output onto maps or charts as well. Data can be downloaded in CSV format, NBN Exchange format, or KML format for use in other applications.

There is no formalized license for data use. Current terms and conditions make the data available for non-commercial use only. The Indicia team would like to move to a Creative Commons License in the future.

Technology Platform

Indicia is an open source project. Its source code is available for use at no cost under the GNU General Public License. Indicia is written primarily in PHP with JavaScript for the front-end. PHP was selected because it was easy to work with and there were a number of PHP developers available at reasonable rates. In addition, data collection is coordinated by local record centers in each region, and many of these centers have staff with limited programming skills. Using PHP was a good way not to lose out on those individuals. Indicia was aware of many "elegant solutions" that simply didn't advance because the code was too hard to get into.

Indicia is designed for use with a Drupal CMS. Drupal has a higher learning curve than other CMS, but is a powerful solution and widely used in the biodiversity field. It also provides a library of free modules that that extend its core functionality for a variety of purposes.

The open source PostgreSQL with PostGIS was selected for the database, since it provides strong support for mapping data. The Indicia team had also considered MySQL, but felt its support for mapping data was too limited. The OpenLayers library is used to enable integration with a variety of base maps, including Google, Bing, and GeoServer.

The Indicia development team was composed of a key developer that was supported by two developers from the Biological Records Centre that worked full-time writing recording schemes, as well as three independent developers. They began work at the end of 2008. Two developers were used for six month to develop a proof of concept. Informal testing and feedback were performed regularly throughout development, and unit tests were developed using Selenium.

Approximately 75% of Indicia implementations have been completed by individuals or organizations that had previous experience with it. The Indicia team runs a three-hour training session to help new project coordinators get their sites up and running. There is also a considerable amount of developer documentation available. All implementations are locally hosted.

Social, Marketing, and Incentives

Through the OPAL site, Indicia provides links to Twitter, Facebook, and other social networking sites to encourage volunteers to share project information through their social media accounts. OPAL also offers a number of activities by geographic region, including educational conferences where volunteers can meet like-minded people and participate in enjoyable projects. Indicia is designed to engage participants from all age levels, including students as young as elementary school. To that end, the OPAL Kids Zone provides some online videogames, including Earthworm Frenzy, shown in Figure 22, that are specifically aimed at engaging school-aged citizen scientists.

While Indicia itself does not offer an incentive program, registered OPAL participants that submit results for three or more Indicia surveys can claim a free "OPAL Explorers Pack" consisting of a notepad, pencil, and magnifier to help them record local wildlife, as well as a certificate and stickers for each survey completed. Classrooms and school groups that complete three or more surveys as a class or group are eligible for a larger class pack as a reward.

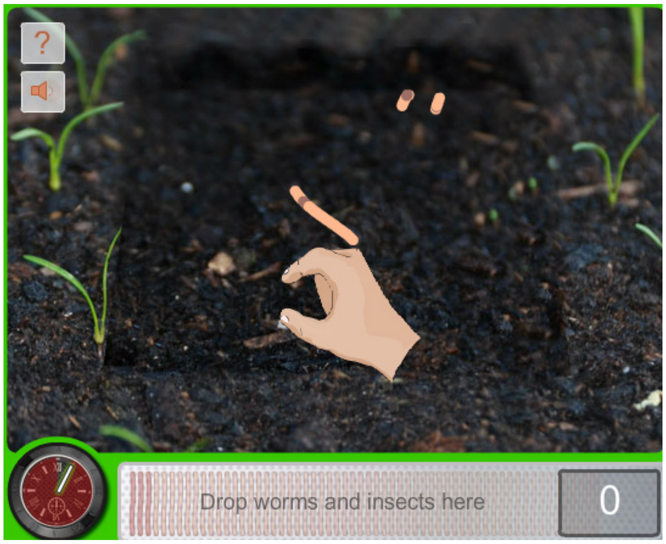


Figure 22: The Earthworm Frenzy video game engages young citizen scientists in the natural world by asking them to catch earthworms before they disappear back into the soil. A Top 10 leaderboard is displayed at the end of the game. The game is an introduction to an outdoor survey that children can perform.

Data Quality

Indicia supports expert verification of incoming records, including bulk verifications of data by species or all data submitted by a designated participant. An optional Data Cleaner module, which is Indicia's version of the National Biodiversity Network's Record Cleaner tool, supports the verification process. It works with the Indicia Warehouse to verify species identifications against the database and applies descriptive comments to any suspicious or potentially erroneous records. The comments are then flagged for manual verification by experts.

Cost for a New Project

There are three options for indicia implementations: building a website using Indicia, hosting survey forms on the iRecord website, or using iRecord to provide forms that can be inserted into an existing website using an iFrame. Building an Indicia website from scratch will require as long as two months and will likely cost between £1,000 and £2,000 if a contractor is employed to perform the work. Building a site provides the greatest potential for branding, customization, and user management, but will require a web service provider or server. Hosting a survey on iRecord can take as little as two weeks and will cost £100. This option does not require a web service provider or server, so it may be a good option for organizations facing resource constraints. Using iRecord via iFrame on an existing website also takes as little as two weeks and will cost £150.

To install Indicia for use with the Indicia Warehouse, PHP version 5.2 or 5.3 is recommended, and the cUrl PHP extension needs to be enabled. PostgreSQL 8.4 or higher is required for the database, and the PostGIS extension must be installed. Installing GeoServer to support spatial data requires a Java SDK.

A forum and chat room are available to provide guidance, and there is extensive documentation available.

National Geographic FieldScope

National Geographic FieldScope is a web-based mapping, analysis, and collaboration tool designed to support geographic investigations and engage citizen scientists in investigations of real-world issues—both in the classroom and in outdoor education settings. National Geographic FieldScope enhances scientific investigations by providing rich geographic context—through maps, mapping activities, and a rich community where fieldwork and data are integrated with that of peers and professionals, adding analysis opportunities and meaning to student investigations.

General

National Geographic is using GIS technology to facilitate "community geography." Community geography connects people to the places they care about by encouraging them to conduct fieldwork and share observations that can both generate and answer scientific questions. The FieldScope project supports this goal by enabling citizen scientists to upload their own georeferenced field data and media objects to a centralized website where they can use GIS tools to perform online data visualization and analysis. Current projects include Chesapeake Bay Water Quality, in which students on both sides of the bay collaboratively investigate water quality issues on local and regional scales and analyze the data to inform action. A more recent project, BioBlitz, is a partnership with the National Park Service. Citizen scientists and students collect data in a designated national park and use Fieldscope for visualization.

FieldScope works with non-standard data and helps citizen scientists do useful things with it. National Geographic considers FieldScope to be focused on visualization and analysis; they want the tool to be a platform for analysis.

FieldScope is distinguished from other citizen science projects by three factors:

1) It is designed to support learning and education.
2) It provides full visualization and analysis even for projects with limited funding.
3) It supports a broad range of disciplines.

FieldScope is currently expanding under a National Science Foundation grant to include new social networking tools and special "project builder" tools that will make it easier for non-programmers to create their own FieldScope projects. In the new interface, users will be able to pick the variables they want to look at or what layers they want to work with, as well as access pre-created maps. FieldScope is currently working with two citizen science programs—Project BudBurst (National Ecological Observatory Network) and FrogWatch, USA(Association of Zoos and Aquariums) to help build and test FieldScope's new capabilities, as shown in Figures 23 through 26.

Flexibility

FieldScope provides support for many domains: the common denominator is georeferenced data. The system accepts geographic location as well as date and time. There is no funding available for mobile applications at this time: National Geographic has applied for several grants but they have been unsuccessful. A project for Olympic National Park – Naturebridge – was able to obtain some Google volunteers who used 20% of their time to build a mobile app specific to that project.

Figure 23: The Project BudBurst site is using FieldScope to display volunteer observation data and provide several options for analysis.

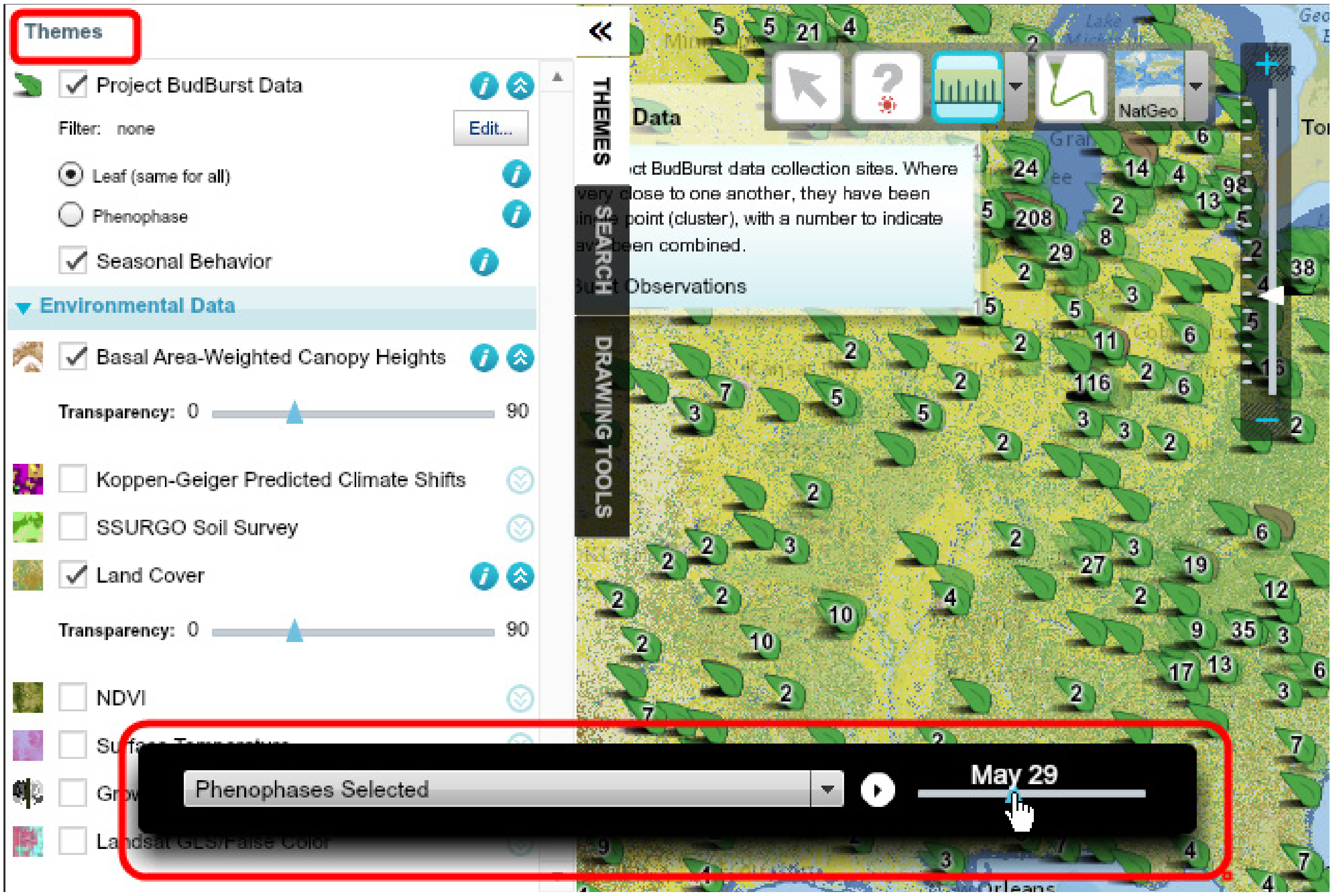


Figure 24: By accessing the Themes panel, Project BudBurst users can add and filter data for analysis at various geographic levels. They can also use a slider bar to visualize phenophases by date.

Display, Visualization and Publication

The new version of FieldScope will offer several types of data visualization, as demonstrated by Figures 23 through 26 from the Project BudBurst site. Participants can view their project data using interactive maps, tables, and graphs, as well as scatterplots and time series animations. Depending on the project, data can be downloaded in spreadsheet, CSV, or text format, or as a summary report. An API is also available. Uploaded images go to the FieldScope Flickr account directly from the application. There is currently no license referenced on the National Geographic FieldScope site for use of the data; however, the Project BudBurst site restricts data usage to non-commercial and research purposes.

Technology Platform

The initial project team consisted of only one developer/analyst between 2008 and 2011. Later, two additional developers came on that had experience with Django and Flex. There are currently 2.5 full-time equivalents working on the project. Bing was used at one point, and Google Maps for rudimentary graphs. The Esri Flex API was used for the original Chesapeake Bay project, but project developers have since rewritten the user interface using open source PureMVC framework. While the original Chesapeake Bay project was not released under an open source license, far more of the recent work is now being developed and released under open source licenses and can be found on GitHub at https://github.com/NGFieldScope/.

Social, Marketing, and Incentives

Once the new interface is complete, which will be sometime in late 2013, National Geographic will be examining incentives such as leaderboards to engage the user community. FieldScope is using grant funding to add social networking tools over the next few years to encourage collaboration and discussion among users.

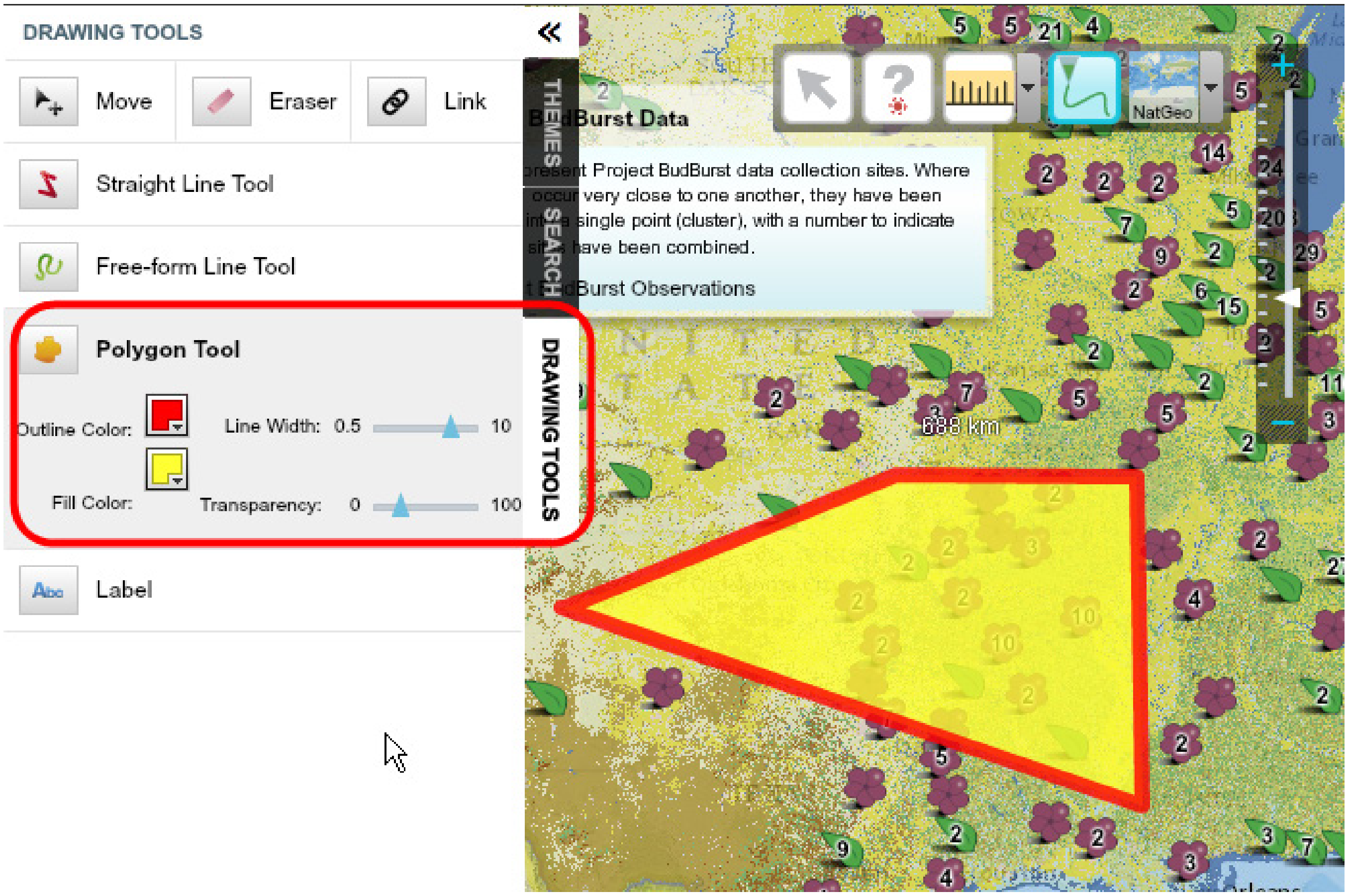


Figure 25: The Project BudBurst Drawing Tools panel provides line and polygon tools to select a customized area of study. Users can save their map views for future reference.

Data Quality

There are no automated processes in place for data quality assurance. For any given project, all data is either trusted or untrusted by default. Trusted data appears on the map, but it can be flagged for review by other users if there is a problem. It is then sent to the Project Manager for review.

Cost for a New Project

If a project does not require any new software features, and the project coordinators are able to provide GIS data in the correct format, it can be set up in as little as a few hours. A GIS Analyst is recommended to clip, reproject, and add symbology to the data. An important consideration is determining who will hold the canonical copy of the data. Data for the FrogWatch project, for example, is held by National Geographic, while data for BudBurst is held by the project.

While grant funding has supported the project in the past, going forward, large organizations will be subsidizing smaller projects, potentially through an annual fee.

National Phenology Network

The National Phenology Network developed Nature's Notebook, a project focused on collecting standardized ground observations of phenology by researchers, students and volunteers. They also foster phenology communities of practice, and the development of tools and techniques to support a wide range of decisions made routinely by citizens, managers, scientists, and others, including decisions related to allergies, wildfires, water, and conservation.

General

The National Phenology Network promotes status monitoring, which is the process of visiting a site regularly to take observa-

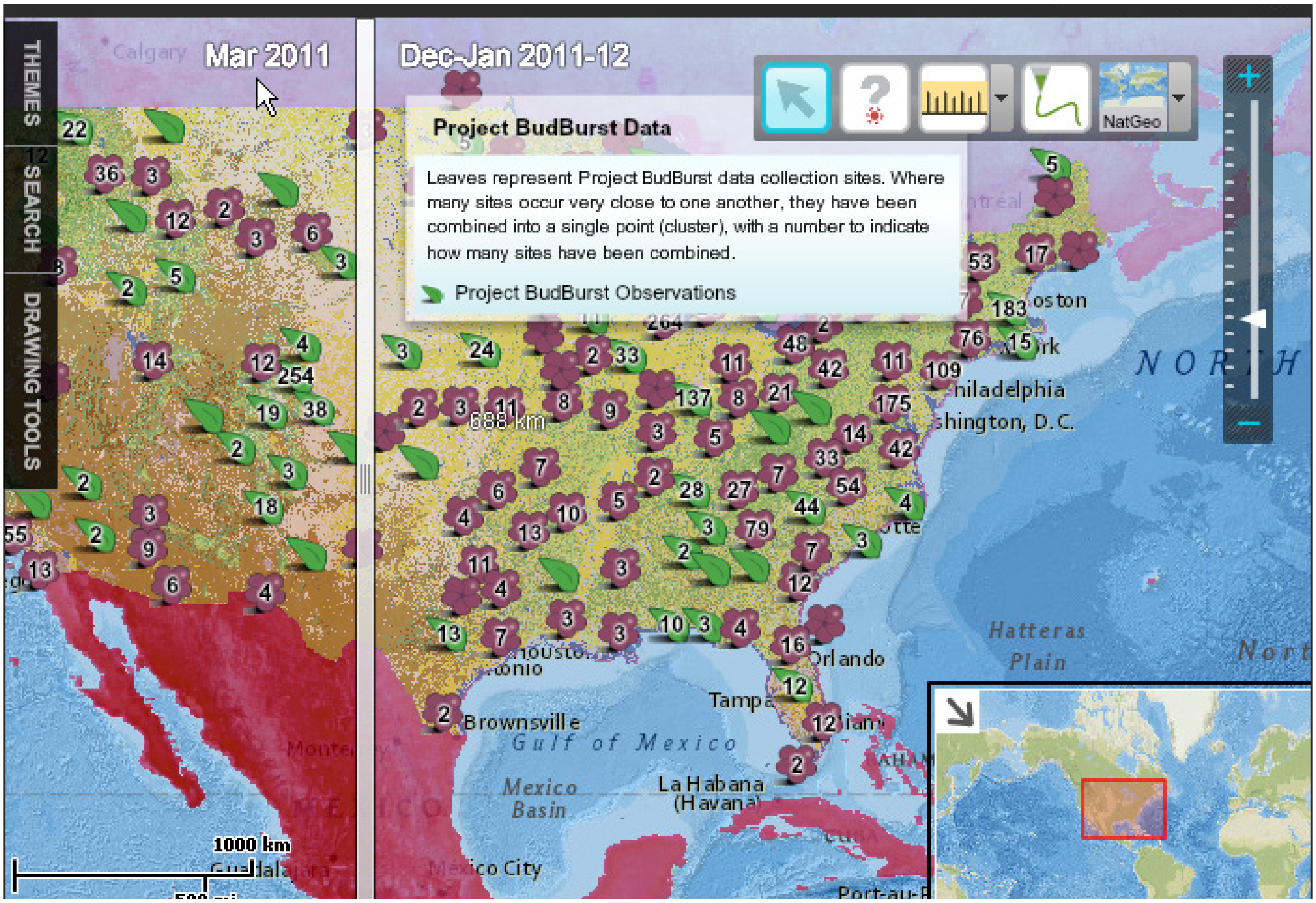


Figure 26: The Project BudBurst implementation of FieldScope enables users to compare surface temperature ranges using a "swipe" tool to move between map layers.

tions. The project puts science first and education second. It is primarily built for researchers and backyard naturalists that know a few basic species. It is considered by its project team to be more difficult to use than other projects, which may be the reason participation rates are lower. Funding was provided by the United States Geological Survey. The project team is interested in a diversified funding model that would include 10% private/foundation funding and 20% federal/non-USGS funding.

The project currently offers a visualization tool that enables users to select and map a species and add climate data to perform comparison studies, but the project team would like to take this to the next level by building a green-out wave visualization feature. They would also like to make it possible for users to download more data directly from the website, particularly climate data. Another important focus is on historic data integration: they have observational data going back to 1955 and would like to make more of it available.

Flexibility

The National Phenology Network's Nature's Notebook project offers smartphone apps for iPhone and Android users, as shown in Figure 27. The apps are recommended for field data entry, since they eliminate the need to transcribe data between manual forms and the project's web-based form. Currently, the apps include textual information and file images only and do not facilitate user-submitted photos.

Protocols for 35 plant and animal functional groups found in the United States have been developed for independent use by any researcher or organization interested in phenology. In addition, the Nature's Notebook program offers status monitoring, online

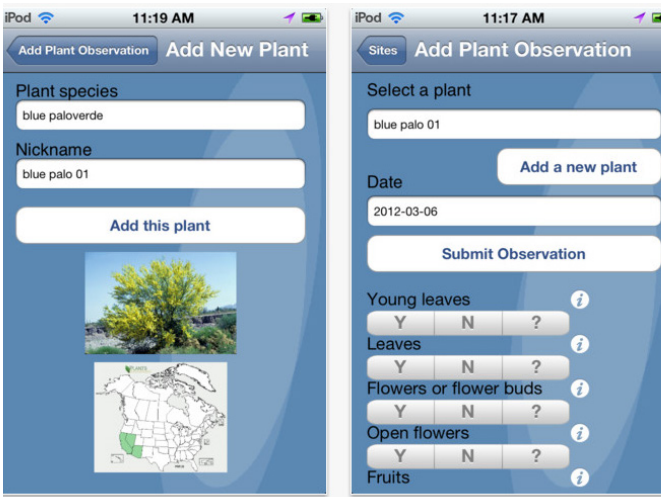


Figure 27: The Nature's Notebook iPhone app enables field data entry with a simple user interface and yes/no/unknown response protocols.

training, and species information about 900 plants and animal species. Data can be entered and accessed through this project.

Display, Visualization, and Publication

The Nature's Notebook project offers a mapping interface with several compelling data analysis and animation features, as seen in Figure 28 and 29. Data can be downloaded directly from the website by completing an online form. Data is available in Excel 2007 format, as a CSV file, or as a ZIP file. A public REST API enables users to add and access available data. A bulk upload tool is expected to be released in early 2013, mostly for federal partners and cases where citizen scientists are using the mobile app to collect data, but have no signal. Data is also being integrated with the YourGardenShow.com website.

Many peer-reviewed papers have been published as a result of the National Phenology Network, and these are listed on the website.

Technology Platform

The web application was developed between 2006 and 2007. The Oak Ridge National Laboratory and graduate students at the University of Tennessee built the website. It was built using MySQL and Drupal. It went live in Summer 2008. If the project team had it to do over, they would still choose Drupal but would build a custom module instead. Hosting is at the University of Arizona, with weekly backups to Oak Ridge National Laboratory servers at the Distributed Active Archive Center (DAAC). Rainlog developers thought the Java code they had written for that project could be adapted for use with the National Phenology Network site if paired with a session cookie to Drupal; this might be a future consideration, but the project team feels it would be hard to go back at this point.

The current staff consists of one full-time programmer, one director who also provides IT support, a part-time server administrator, part-time students, four to five external advisors, and an outreach coordinator. They have been using the same servers since 2007, but would like to upgrade in the future.

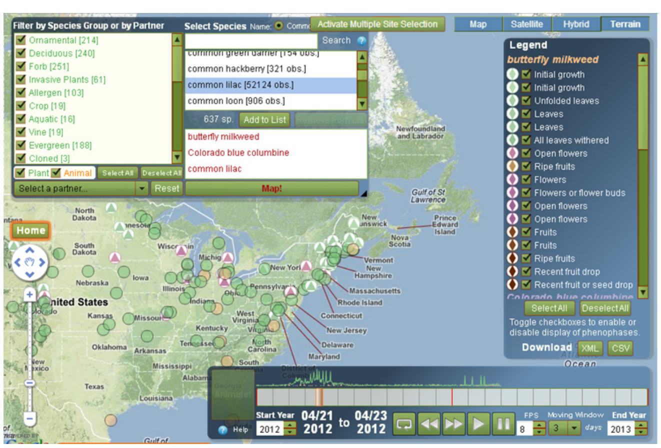


Figure 28: The Nature's Notebook site offers new data visualization tools that include time lapse phenology animation for user designated timeframes. Up to three species can be animated at once.

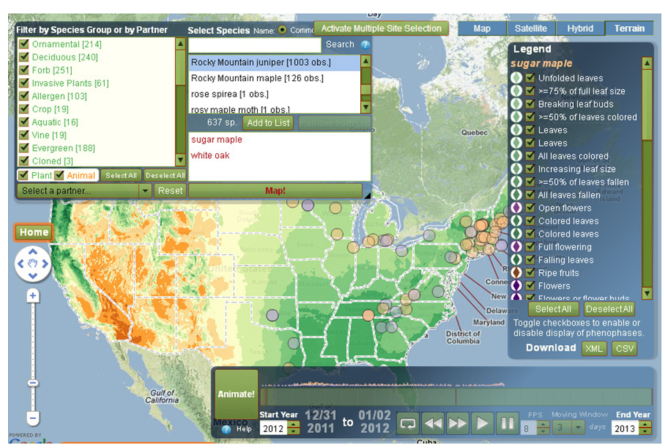


Figure 29: Climate data can be added to the map to show minimum/maximum temperature or precipitation rates over a thirty-year timeframe. Climate data can be animated alongside the phenology data.

Social, Marketing, and Incentives

The Nature's Notebook project hosts a Facebook page that provides information about the project; they also have a Twitter feed. They do not offer authentication through these sites.

A leaderboard on the Nature's Notebook site provides the first names and home states of the top 100 observers since the inception of the program, based on total number of observations. Other leaderboards celebrate top observers on a weekly, monthly, or annual level, as well as top animal and plant species observed. Future plans for incentives include adding badges for activities such as recruiting new observers or cloning a plant. Engagement tends to be better through group projects.

A number of outreach and training resources are available through the website. These are mostly aimed at partner organizations. Curricular resources are available for K-12 as well as higher education. Quarterly newsletters are available to partners, observers, and educators.

The project team currently has about 200 core sites, but they would like to have 1,000 or more. They would also like to have five to ten years of consistent longitudinal data for Autumn and Spring. In order to create a history, they encourage volunteer observations from research stations, national parks, and other places where land tenure is secure.

Data Quality

The National Phenology Network has a series of quality assurance and quality control measures in place. Quality assurance measures are targeted at improving data quality before it even enters the database and include detailed monitoring instructions as well as species profile pages that provide a photo, range map, and written description of the species to help reduce identification errors. Quality control measures include crowdsourced review of photos submitted with observational data.

Transcription errors between manual forms and the online interface are reduced through automated validation, or de-duplication of selected fields; for example, entering a date in the future is not allowed. Species names, abundance measures, and other pertinent data fields are presented as picklists on the data entry form to ensure consistency and conformity. In-person and online workshops offer training opportunities, and a variety of printable handbooks, brochures, and PowerPoint presentations are also available as reference.

Cost for a New Project

Collaborating projects are typically short-term, have specific goals, and are grant-funded. Each project owns its own data, but the National Phenology Network is granted all rights to share it publicly. Costs vary by project.

WildKnowledge

WildKnowledge is a spinout company from Oxford Brookes University, whose research had identified the "engagement" potential of mobile devices before the dawn of the iPhone. Individuals or organizations can subscribe to the portal to make their own content or commission WildKnowledge to create it for them. Existing clients include Wildlife Trusts, National Geographic, Archeological Trusts, and the British Museum.

General

WildKnowledge is passionate about mobile learning. It began as a research project in 2004 that explored the potential of devices fitted with GPS and interactive keys for recording wildlife in a more engaging way. Early projects used PDAs, but iOS was later adopted because it is popular in the school market. All WildKnowledge projects endeavor to use the full functionality of contemporary mobile devices to capture new data in the field and share knowledge with users.

WildKnowledge offers a suite of toolkits that can be used to receive or create data:

- **WildKey** facilitates decisions based on prompts and images, as shown in Figure 30.
- **WildForm** is a customizable recording form.
- **WildMap** facilitates multimedia experiences ranging from field trips to treasure hunts.
- **WildImage** enables the creation of multimedia, interactive diagrams.

Registered users can create their own projects using these toolkits and make them available for public or private use. Content is limited without a paid subscription; for example, three fields per form or three points of interest on a map or image. Project coordinators also have the option of branding their projects as their own for an additional fee.

Flexibility

WildKnowledge supports data collection on virtually any topic. The "wild" in WildKnowledge means "unrestricted". However, most projects are related to the environment or involve healthcare assessment and feedback. The toolkits offer a variety of options for data collection customization. For example, there are 20 different options for creating forms, including drop down lists, note taking, linking to mobile device cameras or GPS, and even quiz questions. WildKey, WildForm, WildMap, and WildImage are currently available for download in the Android Market; similar iPhone apps are coming soon. The general protocol for each project is that users download the forms and maps and store them locally, store data on their devices, and upload data when there is a connection.

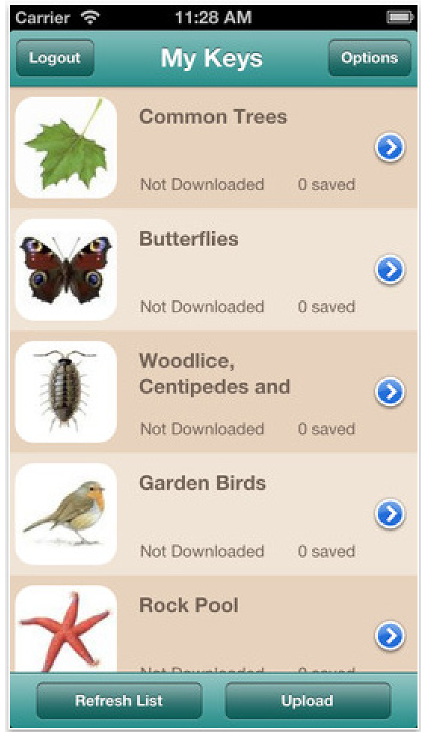


Figure 30: An example of the WildKey interface, which can be customized to meet the needs of each project.

Display, Visualization and Publication

The WildKnowledge website can be used to upload, store, and visualize data as maps, charts, and reports. These can be kept private or shared with the entire community as open projects, as shown in Figure 31. Many of the multimedia apps created by the WildKnowledge team for client projects can be downloaded from the Android Market or iTunes for use by anyone. Data collection projects listed on the WildKnowledge site are sharing data as final reports or through dedicated websites. Since each project is created individually, either by the WildKnowledge team or by user organizations, there is no standard for downloadable data formats, public access, or rights to submitted data.

Technology Platform

Development of the WildKnowledge website was outsourced, and the mobile toolkits were developed by the same external team. Django was used for the web application and data management, JavaScript for the web forms, and MySQL for the database. Two web developers currently maintain the website and mobile toolkits. The toolkits are available through the Android Market, with iPhone iTunes release expected in the near future. OpenStreetMap is the recommended source for mapping content.

Media content size is currently not restricted, but the team is aims to add some type of setting in the future that will control camera resolutions within the mobile applications in order to reduce file sizes.

Social, Marketing, and Incentives

The WildKnowledge team is currently exploring the possibility of a unified login system that would enable users to link to their Facebook accounts and post content directly from WildKnowledge mobile applications. Recruitment and community cultivation efforts seem to be largely completed through individual project websites. There is no standardized system of incentives for WildKnowlege projects.

Data Quality

Data quality can be manually verified by an administrator. Options such as dropdown lists and linking to device functions such as GPS and camera can also be integrated into each project as an additional means of quality assurance. The WildKnowledge team will also be working with the Biological Records Centre in the UK, which is preparing national distribution datasets of terrestrial and freshwater species based on volunteer observations.

Cost for a New Project

If an organization wants to brand a project as its own, it requires a subscription as well as up-front costs. Costs range from £8,000 for a mobile application to £20,000 plus an 18% annual maintenance and hosting fee for a branded mobile application that pushes data to a branded version of the WIldKnowledge web portal. Registered users may use limited versions of WildKnowledge applications, including forms, maps, and at no cost. However, a paid subscription is necessary to use the full version of each application. A discount is provided to educational users. Subscription costs range from £1.99 per Wildknowledge toolkit for an individual to £3,000 for all four toolkits and up to 100 users for commercial organizations.

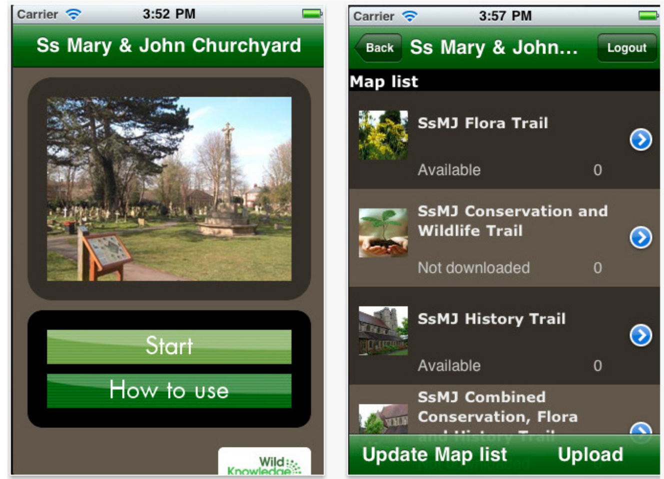


Figure 31: The Saints Mary and John Churchyard project used WildKey and WildMaps to create a downloadable tour application for visitors that integrates multimedia and map elements.

Summary Evaluations

To supplement our in-depth project evaluations, an additional four citizen science projects and platforms were selected for summary evaluation. These projects explore further aspects of the genre, including sensor networks and third-party authentication methods that were not previously touched upon:

1) **EpiCollect** – EpiCollect is aimed at form generation for data collection. It is one of the few citizen science platforms that offer significant flexibility in terms of the types of attributes that project creators can specify. This is aimed specifically at collection through mobile devices.
2) **EveryAware** – EveryAware is a sensor-based platform for mobile data collection. It may offer some insight into the challenges of building projects around sensor data, which typically requires hardware and calibration.
3) **iNaturalist** – The iNaturalist project is open source with a mobile component. It has a very explicit quality assurance process that was well-ranked in our citizen science forum. It also integrates multiple third-party authentication systems, including Facebook.
4) **Project Noah** – Project Noah is primarily of interest because of its strong social marketing and incentive mechanisms, including a system of digital patches and the ability to “follow” other participants on the site.

We also completed a summary evaluation of Zooniverse, a very popular collaborative data classification project. In addition to the familiar Galaxy Zoo, there are a number of interesting data classification projects on Zooniverse, including CycloneCenter, WhaleFM, Seafloor Explorer, Bat Detective, and Cell Slider.

1) **Zooniverse** – This site is distinct from the projects where users contribute data, in that the data—often including sound as well as image media—have already been collected and pre-processed by scientists. Citizen scientists provide classification and transcription services instead.

While we did not conduct telephone interviews as part of our summary evaluation process, a review of the literature was performed along with a hands-on study of the project websites.

Data Collection Projects

The four contributory data collection projects are evaluated first.

EpiCollect

Another Open Source project, EpiCollect is aimed at form generation for data collection. It is one of the few citizen science platforms that offer significant flexibility in terms of the types of attributes that project creators can specify. Similar to WildKnowledge, this project is based in the UK and aimed specifically at data collection through mobile devices. EpiCollect was originally designed for the collection of epidemiology data, but can also be used for other types of field survey work, including ecology and community data collection.

Geotagged data collection forms enable users to collect data in the field using their smartphones and submit it to a central project website. A drag-and-drop form builder enables several types of form customizations:

- A text input field for short entries, such as the answer to simple questions
- A Long Text element for responses that require more text
- A Select One element to limit the field of possible answers and require the selection of one
- A Select Multiple element with checkboxes for selecting one or more entries from a list
- GPS location and photo

All data specific to a project can be viewed on Google Maps or Google Earth using a web-based mapping interface. Project owners can customize the look of their project's homepage; public users can view and filter data only. Google Charts is used to display graphic summaries of the data. Clicking any point on the map will display the data entered for that record, as shown in Figure 32. Data can be filtered by pre-designated criteria, and a time slider enables users to display data for a specified time period for temporal as well as spatial exploration. Data can be downloaded in CSV or XML format, or viewed and filtered on a user's smartphone.

EpiCollect is hosted in the cloud using Google App Engine. The project uses web APIs to access and display Google Maps, Google Charts, Google Talk (for instant messaging), and the KML specification. It was built using the JavaScript libraries JQuery, script.aculo,us, ExtJS, and Mapstraction. Despite its reliance on Google technology, EpiCollect is available for both Android and iPhone devices. The source code for the mobile application is available under an open source Apache License.

EveryAware

EveryAware is funded under the European Union Seventh Framework Programme. It is somewhat unique among citizen science projects in that it collects environmental data from a network of sensors rather than through conventional data entry protocols. EveryAware has been designed to "facilitate the combination of sensing technologies, networking applications, and data-processing tools that will enable citizens to collect and visualize environmental information", as shown in Figure 33. Envisioning itself as a platform, EveryAware is comprised of the following main elements:

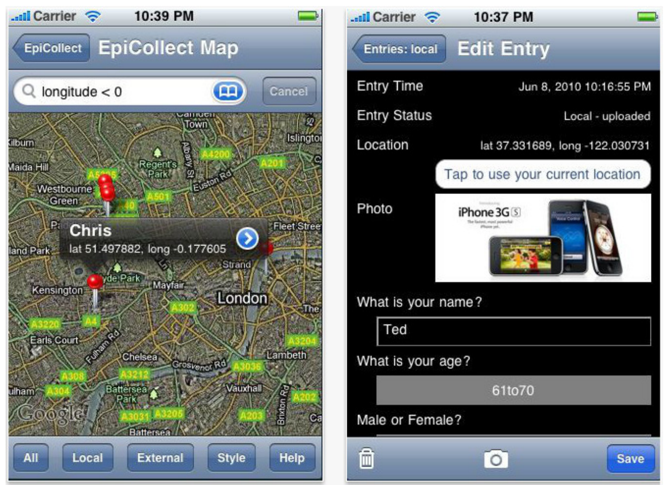


Figure 32: The EpiCollect mobile app can be customized for a variety of data collection projects, including epidemiology data. It integrates coordinate, photographic, temporal, and other data formats.

- A sensor box that includes sensors, energy source, and the ability to communicate recorded data from the sensors to smartphones on a local basis. The sensors offer minimal user interaction capabilities.
- Smartphones with data acquisition software installed to collect and transmit sensor data and provide a user interface.
- Backend technology that collects, processes, and publishes the collated sensor data.

EveryAware's focus on user-centered design is particularly noteworthy. User-centered design puts the user at the center of the development process and requires designers and software developers to clearly understand each feature of the application on the user's terms before any software code is written. It also takes into consideration the ways the application will fit in with other workflow or daily activities of each user. In addition, because EveryAware relies on smartphones, mobility was an important factor in the design process.

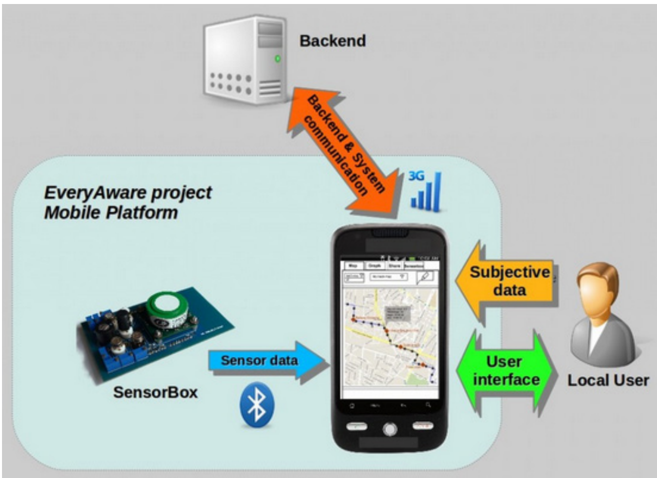


Figure 33: The Every Aware Air Quality project integrates external sensor boxes and smartphones to gather data. The workflow is diagrammed on the EveryAware website.

EveryAware appears to be a fairly immature project, as there have been only two pilot studies so far: a noise monitoring study and an air quality sensing study. The project team has published a number of white papers and reports that offer some insight in to the challenges of building projects around sensor data, which typically requires hardware and calibration. These are available on the project website.

iNaturalist

iNaturalist began as a Master's degree project at the University of California Berkeley in 2008. Its focus is on bringing together naturalists to record their observations, meet others who share their interests, and learn more about the natural world. It is an explicitly open source project that relies on open source software including Ruby on Rails, jQuery, MySQL, and Google Maps, as well as open source projects and public APIs including Catalogue of Life and uBio. The source code is available on GitHub.

iNaturalist was referenced in our citizen science forum as a preferred platform for gathering and submitting data that can be verified. The site's commitment to data quality assessment was considered a great value. All entries are considered "casual" grade unless the iNaturalist community agrees with the observer's species identification, and the observation has a date, is georeferenced, and includes a photo, as shown in Figure 34. Entries can also be flagged if considered out of range or if the species name entered does not match the information in the project's database.

It was also noted in our forum that while iNaturalist offers iPhone and Android apps to its users, many volunteers still preferred using paper data sheets, GPS-enabled cameras, and entering data after the fact rather than while in the field, which may indicate that the mobile apps are not as user-friendly as they could be. When compared to Project Noah, iNaturalist was considered better at identifying species that volunteers were not familiar with, while Project Noah was better for mapping species volunteers could readily identify.

As shown in Figure 35, the ability to use multiple authentication systems, including Facebook and Google, was a noteworthy feature. In addition, iNaturalist can link directly to a user's Flickr account to add his/her Flickr photos to other observations on the site. Users can decide if they want their uploaded images to be shared using a Creative Commons license.

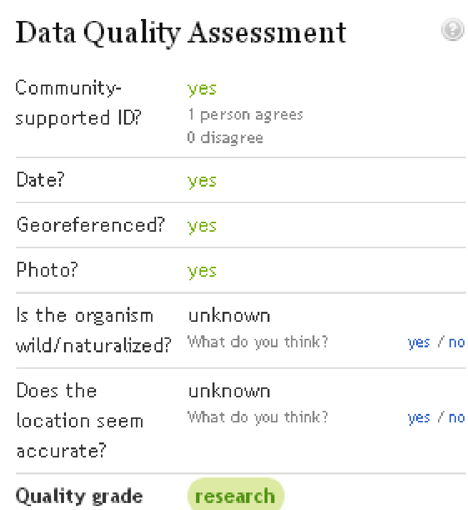


Figure 34: Quality assurance is detailed for each submission and subject to comment from the larger user community.

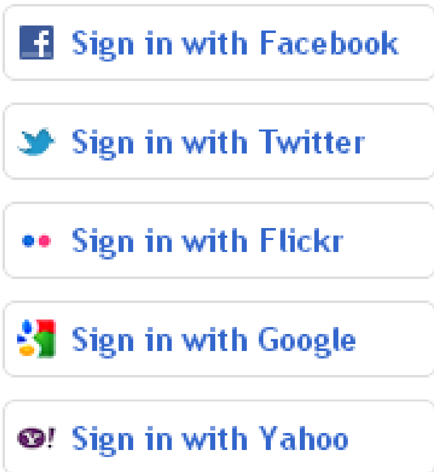


Figure 35: iNaturalist offers multiple third-party authentication options.

Project Noah

Similar to iNaturalist, Project Noah is a tool to explore and document wildlife and a platform to harness the power of citizen scientists everywhere. Project Noah was launched from New York University's Interactive Telecommunications Program and backed by National Geographic to leverage the growing popularity of smartphones to collect ecological data.

Like iNaturalist, Project Noah uses multiple authentication systems such as Google, Facebook, Twitter, Yahoo!, Windows Live, and AOL. In fact, participants must use one of these pre-existing authentication systems to access and contribute to the site. Similar to Facebook, users are asked to create a profile page that can be publicly accessed by the user community, and they can "like" or comment on observations made by others. Project Noah supports photos as well as links to YouTube or Vimeo video content.

Project Noah was referenced in our citizen science forum as more of a tool to get people outside to explore the world around them rather than actively contribute to scientific research. It was considered particularly useful for mapping species that are already familiar to the user.

Perhaps most notable about Project Noah is its polished social marketing mechanisms, and particularly its use of patches to recognize and encourage its members. The system of patches is based on the merit badges issued by the Boy Scouts and Girl Scouts. Spottings patches are earned by overall contributions; Missions patches are earned by working on specific missions or projects; Special Achievements are earned for unique accomplishments, such as providing observations from multiple countries; and Specialist patches can only be earned by participants who submit many observations to a specific wildlife category. Examples of specialist patches are shown in Figure 36.

Data Classification Projects

Projects such as Zooniverse, which ask citizen scientists to categorize or transcribe images or sound clips according to specified protocols, are part of the collaborative data classification model. The 2009 study by the Center for the Advancement of Informal Science Education (CAISE) states that:

> "In Collaborative projects, as in Contributory projects, scientists ask research questions, and members of the

Figure 36: Examples of the specialist patches available to Project Noah participant. Each participant's patches are displayed on their profile pages and can be linked to from the comments they provide on topical submissions.

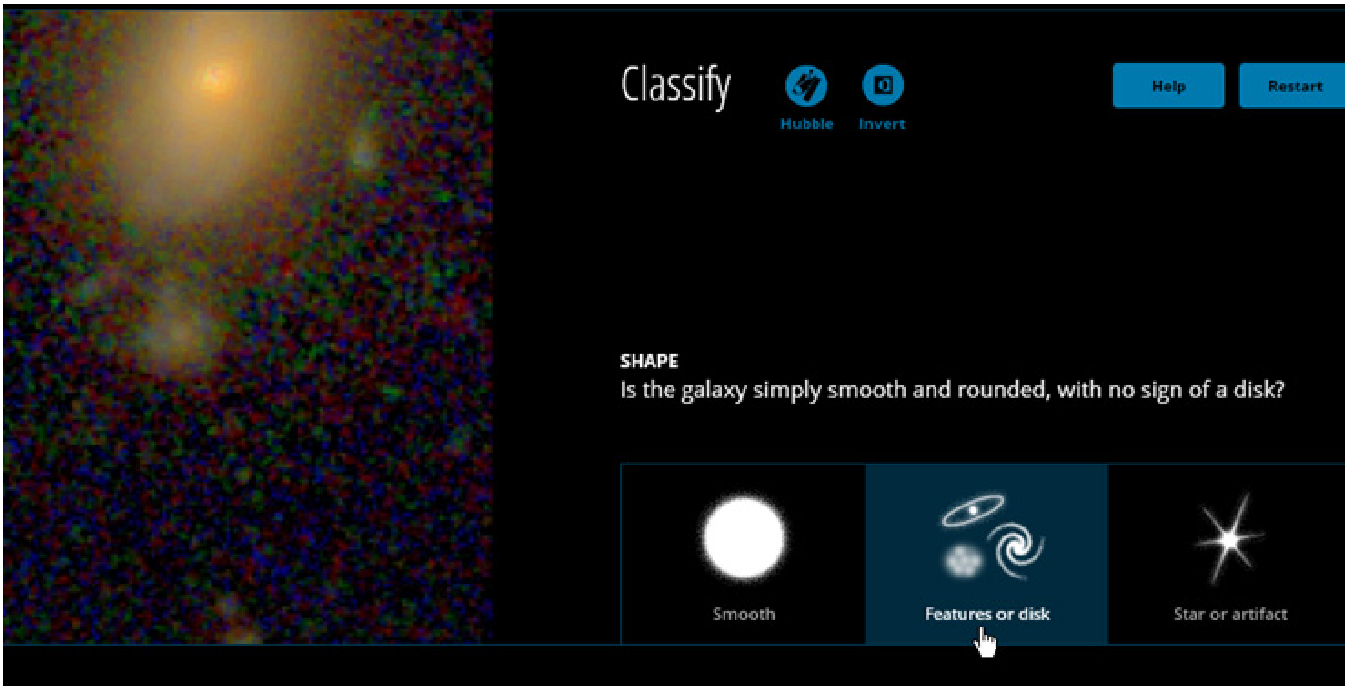


Figure 37: The Galaxy Zoo project integrates a tutorial with its classification interface to enable even beginning users to contribute to the project almost immediately.

public collect data to help answer those questions. However, project participants are actively involved in multiple research activities, including analyzing samples, interpreting data and drawing conclusions, and presenting results to other members of the public and/or to scientists and community decision-makers."

Zooniverse

The Zooniverse project and its open source Scribe application are covered more extensively from a technology perspective in our white paper on citizen science technology platforms. Because it is a data classification rather than data collection project, we did not perform in-depth evaluation of the project or interview its development team. However, its high profile, popularity, and success with citizen scientists warranted a summary evaluation from a user's perspective.

Zooniverse began in 2007 as a single project, Galaxy Zoo, which was an effort to recruit citizen scientists to classify images of approximately one million galaxies from the Sloan Digital Sky Survey. A web interface provided volunteers with a simple tutorial to help them classify galaxies according to their shapes or colors. Each image is classified by multiple volunteers, and each classification is weighted against established performance metrics for quality assurance purposes. More than 300,000 volunteers have contributed to the project so far, and Galaxy Zoo is now in its fourth iteration. More than 30 peer-reviewed papers have resulted from the project, and several new discoveries have been brought to light by participants. Most notably were Hanny's Voorwerp, an ionized gas cloud discovered by a Dutch teacher, and "Pea Galaxies," named for their small size and green coloration. Pea Galaxies appear to have very high rates of star formation.

Zooniverse now hosts more than a dozen citizen science projects in domains ranging from astronomy to climate change. Bat Detective, for example, is helping scientists characterize bat calls, and Cell Slider is helping scientists analyze cancer data. Each project has a dedicated team of professional researchers behind it that are responsible for using the results provided by the site. Projects can be developed by the researchers, by the Zooniverse team, or some combination of the two, which has results in an unfortunate backlog of projects. However, in December 2013, the project announced that it had received $ 1.8 million Google Global Impact Award that will enable Zooniverse to re-build its platform to support a more flexible system and the creation of projects by researchers.

Zooniverse is owned and operated by the Citizen Science Alliance, an international collaboration of scientists, software developers, and educators. The international collaboration enables the organization to solicit grants from public agencies in multiple countires. Citizen scientists retain ownership of their own contributions made to Zooniverse projects, and may use or re-distribute them. However, the Citizen Science Alliance retains ownership of the overall dataset for each project. This is a similar arrangement to the one used by many contemporary open source software projects.

In its study of citizen science projects, the UK Environmental Observation Framework (Roy et al, 2012) noted some important lessons learned as a result of the Galaxy Zoo project that can be applied to other citizen science projects as well. Motivation has been increased by integrating tutorials with the classification interface itself, thereby enabling immediate engagement in the classification process by answering a few simple questions about the image. In addition, the use of HTML5 and JavaScript has alleviated early issues with scalability and performance, particularly when web traffic is high. Cloud infrastructure has also assisted with both spikes in traffic and the continuing growth of the user community.

Recommendations

The review of the citizen science platforms is aimed at identifying gaps and making recommendations in terms of a future platform that can both learn from the best aspects of existing projects and advance the state-of-the-art. The following recommendations are broadly organized into social attributes and technical attributes. For some of these recommendations, we have also developed a set of wireframe designs that are included in an Appendix to the white paper.

Social Attributes

Release software under open source licenses

The rapid advancement of science over the past three centuries has been premised on transparent publication of results. Results are openly shared, enabling both replication and the ability to build on top of new knowledge and understanding. Open source software development processes embody the same underlying principles as the scientific method. When we share our work, we develop better software, and work that is shared enables others to scrutinize it and improve it. Many funded science projects build software that is not made available under these conditions and the community is poorer for it. For example, Leafsnap, an iOS application for identifying tree species, was developed using grant funding, but the source code was not released, and the result is a less sustainable project that has not continued to evolve. Releasing software under an open source license does not guarantee that it will build a community around it or receive contributions from the larger community, but it improves the odds, and since most software projects do not survive, it at least ensures that the resulting software could be used by others later and it will not be lost to the community. Releasing the source code does not preclude a project from generating revenue, any more than publishing a scientific paper prevents a scientist from using her results to build a business.

Share your data and do so under an open database licenses

There appears to be a direct correlation between public participation and data sharing in successful citizen science projects. All of the contributory data collection projects we evaluated for this paper offer one or more methods for sharing volunteer data, and our citizen science forum reiterated the importance of making project data available to the volunteer community. One of the primary reasons given for data sharing was to encourage ongoing participation and underscore the value of volunteer contributions, but it is also an important mechanism for engaging people that may want to contribute by working with the data but are not necessarily interested in the collection process.

Many of the projects we reviewed do not make clear statements concerning who owns the data that is collected. Others rely on Creative Commons licenses. We recommend that projects should both make clear statements about data ownership (even if they are claiming ownership) and how it can be used as well as adopt an open data license of some kind. The Open Knowledge Foundation has developed three such licenses that can be applied to different scenarios.

Make clear privacy statements

Whether citizen science projects are aimed at children, adults, or both, they should make clear statements about what data they collect about users and how they to use it. There are many legitimate and productive uses for user activity data that can help to both improve the functioning of the application and provide better information and recommendations to the end user. But project operators owe it to their contributors to be transparent about how the data is used. When children are the target audience, this is particularly important as many governments now have regulatory frameworks governing how data collected from children is used. Organizations like DataONE are making important contributions by developing policy guidelines.

Crowd validation

Most of the projects we reviewed have some mechanism for review and/or validation of data collected from contributors. In many cases, the project won't be able to scale with manual review of all results. The most scalable systems we reviewed had some component of "crowd validation". These include several tactics including:

- Zooniverse or Mechanical Turk projects which may provide the categorization task to multiple people and then use the consensus result, rather than a single response.
- iSpot, which provides a composite validity score for each species identification based on the reputation of the original contributor, their self-assessed level of confidence, and confirmations from others based on the photograph of the species.

There are lessons to be drawn from outside the citizen science realm as well. The Humanitarian OpenStreetMap Team (H.O.T.) has developed the OSM Tasking Manager for organizing crowd-sourced mapping projects on the tight timelines imposed by humanitarian crisis response. The Qatar Foundation's Computing Research Institute (QCRI) has recently released MicroMappers, a software platform for crowd-sourcing tools for sifting through a lot of data and separate the wheat from the chaffe.

Design incentive structures

The deliberate integration of game play into the design can have a significant impact on to attract, engagement and retention of motivated participants. However, our observation is that the gamification fad has encouraged the ad hoc addition of leaderboards, badges and scores on many projects without consideration for the learning goals that are being incentivized. When poorly implemented, gamification features can have a negative, demotivating effect or increase confusion. We recommend that gamification and design of incentive structures be given the same level of design expertise as the user interface and software architecture in citizen science projects.

Integrate with social networks, but handle with care

More than 2 billion people around the globe are connected to the internet, and one in every nine individuals is on Facebook. More than 3,000 images are uploaded to Flickr each minute, and more than 5 billion images are currently hosted on the site. Twitter is adding 500,000 users each day, and Google+ reached a record 10 million users in its first sixteen days of operation (Bullas, 2011) and is now growing at a faster rate than Facebook. Contemporary social networks are powerful tools for engaging volunteers, validating data, encouraging outreach, and cultivating new content. They can also help solve pragmatic, technical challenges such as user account authentication. They can support citizen science projects in several ways:

Social networks can also suggest useful ideas that can be adopted to support citizen science projects, even without using the network in question. We highlighted Project Noah as an particularly successful of this. Volunteers are able to create their own profile pages that will be populated with images of their most recent animal and plant observations. They can also add links to personal websites, blogs, or other related content. Contributors can "follow" each other's pages and click through the individual images there to a separate detail page. Once there, they can "like" or comment on the image. Combining both gamification and a social component, volunteers are able to view and follow lists of nearby or related observations by others that will help them learn more about the world around them. In addition to being part of the larger Project Noah community, volunteers can sign up for specific "missions" that interest them. Missions may be regional or species-specific in scope, and generally involve a dedicated subset of volunteers with a shared focus or interest.

We recommend leveraging social networks where it makes sense – authentication, social endorsements, sharing activity, and analytics - but both provide alternative mechanisms and be cognizant of the privacy implications for project contributors.

Technical Attributes

Integrate data visualization

Citizen science projects are more meaningful to participants if they can visualize the outcomes, but visualization can also be an important source of discovery and new science. Some of the projects we reviewed are integrating maps and charts to help participants visualize the results, but we believe that contemporary software and cloud processing infrastructure enable this to be advanced in a significant way. The eBird project is the best example of what is possible. Visualization tools on eBird enable participants to see the impact of their work in aggregate but also discover how climate change is impacting species, attract new cadres of observers and make new ornithological discoveries. We also believe there is a significant opportunity to provide online data processing and data science tools that provide an ability to search for correlations and visualize change. In summary, we are advocating for more than putting dots on a map and to aim, instead, for data processing that helps contributors achieve new learning and understanding.

Multiple projects, single sign-on

There are several impediments for contributors to participate in multiple projects, but two significant ones are related to discovering the projects and the need to maintain several user accounts. SciStarter has created a catalog of citizen science projects, media partnerships to distribute and promote projects, and a diverse community of citizen scientists. We believe this is a good start, but we also believe there is a significant opportunity to do more. Such improvements would provide a single infrastructure that can provide an affordable platform for many smaller projects while also enabling users to discover related projects that share a common workflow, user interface standards and account management. A common platform such as this will not be a solution for every project, but we believe it could expand the number of projects that can engage the public in the collection of science data in an affordable manner.

Use cloud infrastructure

Technology infrastructure is an important challenge for most projects. Sever hardware is not very expensive to purchase for a limited project, but it is expensive to maintain and support over time. Further, it requires a great deal of technical expertise to implement and manage. The efficiencies gained by working with a shared infrastructure platform are substantial. Further, in recent years, the economies of scale are beginning to make managing your own infrastructure less attractive, even if an organization has the technology staff to maintain their own infrastructure. Finally, many of the cloud technology providers, such as Rackspace, Microsoft Azure, Google Compute Engine, Amazon Web Services, and Linode, provide metered billing for increments of computing, storage and network capacity. This enables projects to be designed that can start up several dozen servers for an hour in order to crunch through a lot of data, and then shut down the machines when the task is complete. This is an extremely compelling capability and maintaining dozens of servers for these types of situations will be unaffordable for most projects.

In addition to cost and scalability, there are additional advantages to using cloud infrastructure. Many of the large providers provide both secure facilities, lower energy costs, and far better security practices than even the most well-funded universities and corporations. A second white paper will examine different cloud infrastructures in more detail.

Plan for internationalization

Many citizen science projects could potentially attract a global audience, but are not designed to accommodate non-English speakers. The content and data management tools being used by many of the projects we reviewed support internationalization and multiple languages, but few of the projects are designed to accommodate this need. Once again, eBird is an exception, but it is the proverbial exception that proves the rule. The OpenTreeMap project is an example of a citizen science platform that was originally developed using a platform, Python and Django, that supported internationalization and localization but was not designed that way, and an entirely new version had to be written to integrate this type of capability. This is a feature that we recommend teams design into their projects from the outset. The payoff will be the ability to use the tools in a truly global way.

Implement Open Web APIs

Application Programming Interfaces (APIs) may seem like a topic that is more appropriate for technical specifications, but we believe they deserve some attention. Apart from eBird and a handful of other projects (OpenTreeMap, CUAHSI Hydrological Information System and others), none of the citizen science projects we reviewed are providing open APIs. An API is a bit like a tap for a water infrastructure or an electrical outlet for power – they are a standardized way for software components to interact, including accessing and updating data in a database. Why is this important? We believe there are several reasons including: flexibility, engagement and collaboration.

First, whether implementing a web or mobile application, the end user's experience is generally limited to the workflow defined by the developer. An API provides a mechanism whereby applications can be created that support alternative application workflows. For example, the OpenTreeMap project provides a web application as well as an API. The API has enabled the project team to implement both Android and iOS applications without re-building the database access and data management tools – the user experience is a flexible façade created on top of the core functions. And if someone wishes to build a new mobile application that uses and updates the tree data but is optimized for utility vegetation management, rather than street tree data collection, only the façade needs to be created, rather than an entirely new application. This results in faster development and more flexibility.

Second, we believe that the potential population of data collection volunteers is far larger than the current one. But we also believe that there are many people that wish to contribute to projects but are not interested in collecting data per se. For example, a growing population of software developers, designers and data analysts refer to themselves as "civic hackers". They get together in evenings and on weekends to create software together than improves their community in some way. In most cities, there are now several "hackathons" and "data-palooza" events held each year that gather people together for a weekend to solve problems and create new software and visualizations. New organizations, such as Code for America, are organizing volunteer "brigades" in communities around the United States to work on these types of projects. In spring 2013, the second International Space Apps Challenge was organized by NASA. The two-day event was the largest hackathon in history, and gathered 10,000 people around the world to build new applications using open data APIs and data sets from NASA. Open, machine readable APIs are one of the key ingredients that are making this type of mass collaboration possible. We believe there is significant potential for engaging this "civic hacking" community to make the data more available and in more contexts.

Third, APIs can mitigate the silo-ing of projects. Most citizen science projects are focused on a single, specific domain or task. However, when a research question crosses domains, it becomes difficult to bring together data from multiple projects. Export capabilities enable this sort of collaboration, but APIs enable the ability to create interactive tools that would draw on the data from multiple citizen science projects.

On the Horizon

The world of crowd-sourcing and citizen science has evolved significantly even as we wrote the report. There are new platforms and announcements of new platforms on an ongoing basis. A few of the most recent developments were made prior to publishing the report but were not discussed at length.

CrowdCurio

The National Science Foundation has funded a new project, CrowdCurio (http://crowdcurio.com), to create a crowdsourcing platform that "connects interested citizens with researchers to help answer important questions in the science and humanities." Curio will be designed to enable researchers to create their own crowdsourcing projects. The project is led by Edith Law, who has a background in machine learning and computer-human interaction, two domains that could have a significant positive impact on citizen science efforts.

PyBossa and Crowdcrafting.org

PyBossa is an open source platform that allows image transcription. PyBossa, developed in conjunction with the Citizen

Cyberscience Centre, is designed as a framework for developing data collection applications, not as an application or hosted service itself.

Crowdcrafting.org is an application that is powered by PyBossa with many excellent features for creating a citizen science community. It provides a common place where anyone can create a new project, import tasks for human categorization or transcription, and a place for interested volunteers to discover projects to engage with. There are some limitations and we outline them in the companion technical paper, but the new platform is a strong showing.

Zooniverse as a Platform

In December 2013, the Zooniverse project announced that it had received $ 1.8 million Google Global Impact Award that will enable Zooniverse to re-build its platform to support a more flexible system and the creation of projects by researchers. Zooniverse has a great deal of experience creating visually attractive, engaging, and successful classification and transcription projects, and we look forward to seeing the new platform as it develops.

Data Sharing, Best Practices and Standards

Standardization has been an important foundation for growth and innovation in the technology ecosystem. Web standards, such as HTML, CSS and Javascript, underpin much of the development of the internet. Data-related standard like FGDC (metadata), GTFS (transit), JSON (web data), OGC (geospatial web services), and OAuth (authentication) have made significant contributions to interoperability. There are several efforts underway aimed at standardization, integration and data sharing. Of the several reports and guidelines for citizen science, several resources being developed and released by the DateONE organization in the past year deserve special mention including:

- *Data Policies for Public Participation in Scientific Research: A Primer*, DataONE Public Participation in Scientific Research Working Group, August 2013, http://www.dataone.org/sites/all/documents/DataPolicyGuide.pdf
- *Data Management Guide for Public Participation in Scientific Research*, DataONE Public Participation in Scientific Research Working Group, February 2013, http://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf
- *Primer on Data Management: What you always wanted to know but were afraid to ask*, Carly Strasser, Robert Cook, William Michener, Amber Budden, http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Conclusions

The number of citizen science programs has grown enormously over the past several years. Public participation has become a vital component of many academic and government research projects, driving our understanding of subjects ranging from astronomy to climate change to species migration. With the growth of web and mobile computing, citizen science projects and the dissemination of their results have become far more common and available to the public. While the first Christmas Bird Count engaged most of its 27 volunteers from a single geographic region, the Cornell Lab of Ornithology now engages millions of volunteers each year from a global base of ornithology buffs.

However, despite this growth, much of the potential of citizen science remains unrealized. More projects are available than ever before, but there is no single repository where potential volunteers can efficiently search for multiple projects across a diversity of academic disciplines, sign up for one or more projects using a unified authentication and ID management system, visualize their data contributions, communicate with others, and earn collective acknowledgements or incentives that transcend individual projects and organizers. At the same time, project administrators are struggling to make the best use of emerging technologies with potentially limited budgets and to understand the process required to recruit, engage, and recognize the citizen science community. Finally, social scientists and practitioners struggle to research the collective field of projects and participants. At the intersection of these issues is the opportunity to transform the current state-of-the-art in terms of scientific productivity and contributor satisfaction. The resulting effort will provide a broad foundation upon which new levels of public participation can be built, shared, and enjoyed.

Our evaluation of existing projects and platforms has also demonstrated the need for a flexible, open source citizen science toolkit upon which project administrators can build customizable data collection projects within a larger cloud-based framework. While a number of citizen science toolkits currently exist, they tend to be limited in scope and are generally focused on specific types of projects or topics. The use of open source components to build a more generalized toolkit can reduce development and carrying costs for new projects; it also complements the public accessibility and transparency standards that are the hallmarks of many citizen science applications. Most importantly, open source technology provides the opportunity to create a vibrant community of software developers and other users that will not only download and use the toolkit to implement projects for themselves and others, but can also be relied upon as a collaborative resource for managing additions, revisions, and bug fixes that will keep the toolkit robust and current over time. Using a cloud-based platform to jointly host these projects provides access to a shared pool of configurable computing resources, such as networks, servers, storage, applications, and services, that can be rapidly provisioned over the internet with minimal management effort or service provider interaction. Properly implemented, cloud services are scalable to support both large and small projects, and they potentially provide a cost-effective "pay-as-you-go" billing system that can be aligned with limited resources. Cloud hosting also reduces the cost of acquiring the maintaining computer hardware. There are some significant efforts underway to create this type of platform, including Crowdcrafting, CrowdCurio and the new Zooniverse effort.

Recruiting and retaining participants is a constant source of concern for both existing and new projects. While some of most popular projects, such as eBird and Zooniverse, have established communities to tap, most new projects have to start from scratch to develop new communities of participants. Social networks offer powerful tools for recruiting volunteers, validating data, encouraging outreach, and cultivating new content. The ability of these tools to bring people together for the common good was dramatically demonstrated during the recent political uprisings in the Middle East (Callari, 2011), as well as during the aftermath of natural disasters such as the 2010 earthquake in Haiti (Biewald, 2011) and the 2011 earthquakes, tsunami and nuclear crisis in northern Japan (BBC News, 2011). Hundreds of thousands of people collaboratively shared their experiences in real-time to increase society's understanding of these transformative and evolving issues.

The most prevalent social networks, including Facebook, Twitter and Google+, provide options for a unified authentication system that could potentially be shared across many citizen science projects, regardless of originator or subject matter. An alternative to this approach would be to develop a unified authentication system specific to the citizen science. Having a single login and password that can be used to access and contribute to any project of interest is far more convenient for citizen scientists than managing multiple, separate logins. For project administrators, it helps validate the source of data submissions and adds an extra level of security to both web- and mobile-based projects.

The introduction of game-like features that incite friendly competition seems to have a significant positive impact on the level of volunteer engagement, particularly by younger people. Scoring, leaderboards, integration with social media to share challenges and progress with others, public recognition for prolific or expert contributors, and the ability to define groups that can band together and aggregate their scores are examples of features and practices that support gamification. For greatest impact, methods should be put in place to enable Incentives and other forms of recognition to be accumulated and displayed across multiple, diverse projects.

The increase in eBird participants as a result of new data visualization tools underscores the importance of a visualization element in any citizen science project. Indeed, all evaluated projects offered some type of visualization for participants. Maps are particularly engaging, as they provide a geospatial connection to contributed data. Other visualizations include charts, graphs, and histograms. The open source Google Charts, which is used in the EpiCollect project, is a possible solution for displaying data graphically in charts.

Our papers and design documents provide a blueprint for an open source citizen science platform that reflects a multi-faceted evaluation of the current state-of-the-art.

Acknowledgements

This is Part 1 of a four part report. The full report includes the following:

- Part 1: Data Collection Platform Evaluation
- Part 2: Technology Evaluation
- Appendix A: Wireframe Designs
- Appendix B: Cloud Computing Performance Testing

This report was funded in part by a grant from the Alfred P. Sloan Foundation.

We respectfully acknowledge the individuals and projects that contributed to its development of this study; we could not have completed this document without their invaluable expertise and input. We are particularly grateful for the valuable insights provided by both the project representatives we interviewed and the input from the Citizen Science Community Forum.

These recommendations are meant to both celebrate their accomplishments and acknowledge the need for continued growth and improvement that will benefit everyone. To that end, we welcome additional comments from the citizen science community such that we may further refine this vision and move it forward toward reality.

Bibliography

Anthony, D., Kotz, D., & Henderson, T. (2007). Privacy in location-aware computing environments. *Pervasive Computing, 6*(4), 64–72.

Aoki, P. M., Honicky, R. J., Mainwaring, A., Myers, C., Paulos, E., Subramanian, S., Woodruff, A. (2009). A vehicle for research: using street sweepers to explore the landscape of environmental community action. In Proc. of CHI.

Au, J., P. Bagchi, B. Chen, R. Martinez, S. A. Dudley, and G. J. Sorger. (2000). Methodology for public monitoring of total coliforms, Escherichia coli and toxicity in waterways by Canadian high school students. Journal of Environmental Management 58 (3):213–230.

Bhattacharjee, Y. (2005). Ornithology: a half-century of interaction with bird watchers has evolved into a robust and growing collaboration between volunteers and a leading ornithology lab. Science, 308, 1402–1403.

Bonney, R. (1996). Citizen science: A Lab tradition. *Living Bird*, 15(4), 7–15.

Bonney, R. (2007). Citizen science at the Cornell Lab of Ornithology. In Yager, Robert E. and Falk, John H. (Eds.), *Exemplary Science in Informal Education Settings: Standards-Based Success Stories* (pp. 213–229). Washington, D.C.: NSTA Press.

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., and Wilderman, C. C. 2009. Public Participation in Scientific Research: Defining the Field and Assessing Its Potential for Informal Science Education. A CAISE Inquiry Group Report. Washington, D.C.: Center for Advancement of Informal Science Education (CAISE).

Bonney, R.; Cooper, C.B.; Dickinson, J.; Kelling, S.; Phillips, T.; Rosenberg, K.V.; Shirk, J. (2009). Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy. BioScience 59: 977–984, American Institute of Biological Sciences.

Bradford, B. M., and G. D. Israel. 2004. *Evaluating volunteer motivation for sea turtle conservation in Florida.* University of Florida: Agriculture Education and Communication Department, Institute of Agriculture and Food Sciences, AEC 372.

Bui, L. Keeping the Kids Involved: A Look at COPPA and Citizen Science. http://scistarter.com/blog/2014/01/keeping-kids-involved-look-coppa-citizen-science/ Accessed 1/7/2014.

Bullas, J. (2011). 20 Stunning Social Media Statistics Plus Infographic. http://www.jeffbullas.com/2011/09/02/20-stunning-social-media-statistics/

Canfield, D. E., C. D. Brown, R. W. Bachmann, and M. V. Hoyer. 2002. Volunteer lake monitoring: Testing the reliability of data collected by the Florida LAKEWATCH program. Lake and Reservoir Management, 18(1): 1–9.

Cook, G. 2011. How Crowdsourcing is Changing Science. *The Boston Globe*.

Crall, A.; Newman, G.; Jarnevich, C.; Stohlgren, T.; Waller, D.; Graham, J. (2010). Improving and integrating data on invasive species collected by citizen scientists. *Biological Invasions.*

DataONE Public Participation in Scientific Research Working Group. August 2013. *Data Policies for Public Participation in Scientific Research: A Primer*. http://www.dataone.org/sites/all/documents/DataPolicyGuide.pdf

DataONE Public Participation in Scientific Research Working Group. February 2013. *Data Management Guide for Public Participation in Scientific Research*. http://www.dataone.org/sites/all/documents/DataONE-PPSR-DataManagementGuide.pdf

Delaney, D. G., C. D. Sperling, C. S. Adams, and B. Leung. 2008. Marine invasive species: validation of citizen science and implications for national monitoring networks. Biological Invasions 10:117–128.

Donnelly, E. 1999. Journey North: tracking the migration over the Internet. In J. Hoth, L. Merino, K. Oberhauser, I. Pisantry, S. Price, and T. Wilkinson, eds., The 1997 North American Conference on the Monarch Butterfly, pp.347–48. Montreal: Commission for Environmental Cooperation.

Downs, J.S., Holbrook, M.B., Sheng, S., and Cranor, L.F.. Are your participants gaming the system?: screening mechanical turk workers. In Proceedings of the 28th international conference on Human factors in computing systems, CHI '10,pages 2399{2402, New York, NY, USA, 2010. ACM.

Fore, L. S., K. Paulsen, and K. O'Laughlin. 2001. Assessing the performance of volunteers in monitoring streams. Freshwater Biology 46:109–123.

Goldman, J.; Shilton, K.; Burke, J.; Estrin, D.; Hansen, M.; Ramanathan, N.; Reddy, S.; Samanta, V.; Srivastiva, M. (2008). Participatory Sensing: A citizen-powered approach to illuminating the patterns that shape our world. Center for Embedded Networked Sensing, UCLA.

Gommerman, L.; Monroe, M. Lessons Learned from Evaluations of Citizen Science

Programs. University of Florida, FOR291, one of a series of fact sheets from the School of Forest Resources and Conservation, Florida Cooperative Extension Service, Institute of Food and Agricultural Sciences, University of Florida. Original publication date: May 2012.

Hand, Eric. (2010). Citizen science: People power. Published online 4 August 2010 | *Nature* **466**, 685-687 (2010) | doi:10.1038/466685a.

Kim, S.; Mankoff, J.; Paulos, E. (2013). Sensr: Evaluating A Flexible Framework for Authoring

Mobile Data-Collection Tools for Citizen Science. Association for Computing Machinery (ACM).

Krasny, M. E. and Bonney, R. (2005). Environmental education through citizen science and participatory action research. In E. Johnson and M. Mappin (Eds.), *Environmental education or advocacy: Perspectives of ecology & education in environmental education* (pp. 292–319). Cambridge, UK: Cambridge University Press.

Lintott C.J.; Schawinski, K.; Slosar, A.; Land, K.; Bamford, S.; Thomas, D.; Raddick, M.J.; Nichol, R.C.; Szalay, A.; Andreescu, D.; Murray, P.; van den Berg, J. 2008. "Galaxy Zoo: Morphologies derived from visual inspection of galaxies from the Sloan Digital Sky Survey." *Monthly Notices of the Royal Astronomical Society* 389(3), 1179.

McEver, C., Bonney, R., Dickinson, J., Kelling, S.,Rosenberg, K., and Shirk, J. (Editors). 2007. Proceedings

of the Citizen Science Toolkit Conference. Cornell Laboratory of Ornithology, Ithaca, NY, June 20-23, 2007.

McGonigal, Jane. *Reality Is Broken: Why Games Make Us Better and How They Can Change the World.* Penguin Books. 2011.

Mell, P.; Grance, T. 2011. The NIST Definition of Cloud Computing. National Institute of Standards and Technology, U.S. Department of Commerce. Special Publication 800-145.

Nadkamia, A.; Hofmann, S.G. Why do people use Facebook? *Personality and Individual Differences*, Volume 52, Issue 3, February 2012, Pages 243–249.

Nerbonne, J. F., Ward, B., Ollila, A., Williams, M., & Vondracek, B. (2008). Effect of sampling protocol and volunteer bias when sampling for macroinvertebrates. Journal Of The North American Benthological Society, 27(3), 640-646. doi:10.1899/07-101.1

Nov, O. (2007). What motivates Wikipedians. *Communications of the ACM*, 50 (11) 60-64.

Nov, O.; Arazy, O.; Anderson, D. (2011). Dusting for Science: Motivation and Participation of Digital Citizen Science Volunteers. *ACM iConference 2011*.

Ohab, J. (2012). The Top Citizen Science Projects of 2011. Shareable: Science and Tech. http://www.shareable.net/blog/the-top-11-citizen-science-projects-of-2011

Prytherch DR, Smith GB, Schmidt P, Featherstone PI, Stewart K, Knight D and Higgins B (2006) Calculating early warning scores - A classroom comparison of pen and paper and hand-held computer methods. Resuscitation 70: 173-178.

Robson, C. (2012). Using Mobile Technology and Social Networking to Crowdsource Citizen Science. Technical Report No. UCB/EECS-2012-195: http://www.eecs.berkeley.edu/Pubs/TechRpts/2012/EECS-2012-195.html

Roy, H.E., Pocock, M.J.O., Preston, C.D., Roy, D.B., Savage, J., Tweddle, J.C. & Robinson, L.D. (2012)

Understanding Citizen Science & Environmental Monitoring. Final Report on behalf of UK-EOF. NERC

Centre for Ecology & Hydrology and Natural History Museum.

Shirky, Clay. (2010). Cognitive Surplus: Creativity and Generosity in a Connected Age. Penguin Press, ISBN-10: 1594202532.

Smith, A. (2013). Smartphone Ownership, June, 2013. http://pewinternet.org/Reports/2013/Smartphone-Ownership-2013.aspx

Strasser, C., Cook, R., Michener, W., Budden, A. *Primer on Data Management: What you always wanted to know but were afraid to ask*. http://www.dataone.org/sites/all/documents/DataONE_BP_Primer_020212.pdf

Sullivan B.L.; Wood, C.L;;Iliff, M.I.; Bonney, R.E.; Fink, D.; Kelling, S. (2009). eBird: A citizen-based bird observation network in the biological sciences. BiologicalConservation 142: 2282–2292.

Werbach, Kevin. *For the Win: How Game Thinking Can Revolutionize Your Business.* Wharton Digital Press. 2012.

Wiggins, A. Citizen Science Project Survey Results. (2009). Published by Syracuse University, with funding from the National Science Foundation.

Wiggins, A.; Newman, G.; Stevenson, R., Crowston, K. (2011). Mechanisms for Data Quality and Validation in Citizen Science. Proceedings of "Computing for Citizen Science" workshop at the IEEE eScience Conference.

Evaluated Projects

Links to the websites of each project evaluated in this paper are provided below.

In-Depth Evaluations

CitSci.org
http://www.citsci.org/

eBIrd
http://ebird.org/content/ebird/

Indicia
http://www.indicia.org.uk/

National Geographic FieldScope
http://education.nationalgeographic.com/education/program/field-scope/?ar_a=1

National Phenology Network
https://www.usanpn.org/

WildKnowledge
http://www.wildknowledge.co.uk/

Summary Evaluations

EpiCollect
http://www.epicollect.net/

EveryAware
http://www.everyaware.eu/

iNaturalist
http://www.inaturalist.org/

Project Noah
http://www.projectnoah.org/

Zooniverse
https://www.zooniverse.org/