

Creek Watch: Pairing Usefulness and Usability for Successful Citizen Science

Sunyoung Kim¹, Christine Robson^{2,3}, Thomas Zimmerman², Jeff Pierce², Eben Haber²

¹Human-Computer Interaction Institute
Carnegie Mellon University
5000 Forbes Avenue, Pittsburgh, PA
sunyoung.kim@cs.cmu.edu

²Mobile Computing Research Group
IBM Research - Almaden
650 Harry Road, San Jose CA 95120
{tzim, jspierce, ehaber}@us.ibm.com

³Department of Computer Science
University of California, Berkeley
2599 Hearst Ave., Berkeley, CA 94720
crobson@eecs.berkeley.edu

ABSTRACT

Citizen science projects can provide a wealth of data for environmental programs, but that data is only helpful if it is actually used. While previous citizen science research has focused on how to design effective capture interfaces and incentive mechanisms, in this paper we explore the application of HCI methods to ensure that the data itself is useful. To provide a focus for this exploration we designed and implemented Creek Watch, an iPhone application and website that allow volunteers to report information about waterways in order to aid water management programs. Working with state and local officials, as well as private groups involved in water monitoring, we conducted a series of contextual inquiries to uncover what data they wanted, what data they could immediately use, and how to most effectively deliver that data to them. We iteratively developed the Creek Watch application and website based on our findings and conducted an evaluation of it with users that are both contributors and consumers of water data: scientists at the city water resources department. Our study reveals that the data collected is indeed useful for their existing practices and is already in use in water and trash management programs. Our results suggest that the application of HCI methods to design the data for the end users is just as important as their use in designing the user interface.

Author Keywords

Citizen Science, Participatory Sensing, Mobile, Smart Phones, iPhone Application.

ACM Classification Keywords

H.5.2 User Interfaces, J.4 Computer Applications: Social and Behavioral Sciences.

INTRODUCTION

Citizen science is, at its core, a way of empowering people to help with large-scale scientific problems of interest to

them. Typically, citizen science projects engage volunteers in data-gathering activities around a project in their natural environment (e.g., ecology, biology, astronomy, etc). This approach fulfills two broad goals: engaging and educating the public about scientific issues, and collecting data that would otherwise be difficult or impossible to obtain.

Perhaps the most successful citizen science project to date is the Christmas Bird Count (CBC), a project started over a century ago by the Audubon Society, which asks volunteers to spend a few hours outside during the holidays counting the number of birds in each species they sight. The CBC is successful in two ways: it inspires the involvement of tens of thousands of volunteers, and the results, collected at thousands of sites each year, form one of the most critical data sources used by ornithologists to track bird populations and distributions across the globe. The volunteers involved in the CBC are mostly hobbyist birdwatchers and not full-time scientists; however, the results of their joint efforts have had significant scientific impact. [18]

While data collection for the CBC was traditionally done with paper forms, marked in pencil and sent by post to the Audubon Society, modern citizen science is no longer so limited. Web data submission is now common, and the proliferation of mobile phones in general (five billion in the world today [31]) and GPS and camera-equipped smart phones in particular (fifty million in the USA [9]) holds great promise for enhancing citizen science participation, providing a convenient means to collect and send time- and location-tagged data. Smart phones are a particularly powerful tool, combining an array of sensors that can be paired with rich user interfaces in a personal device that is continually carried by millions of people around the world.

Previous research in citizen science and participatory sensing has focused largely on the problem of engagement; recruiting volunteers and maintaining their interest in participation [5,12,24]. Incentive structures, competition, entertainment and educational approaches have all proven successful to some extent. Yet engagement is only half the measure of success. While engagement is essential to educate, motivate and retain citizen's participation, it has scientific impact only when the data is used. The disconnect between the ability to gather data and the ability to use it has resulted in several prominent citizen science projects that, while well-advertized and widely employed, have produced few if any useful scientific results [22,31].

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05...\$5.00.

In our research we address this disconnect by focusing on the usability and usefulness of collected *data*, rather than just on the *interface* presented to users. We wanted our design to be driven by the scientists who need the data as well as by the volunteer groups that support their investigations. By first understanding the needs and abilities of these groups, we are more likely to build a citizen science application that is useful in addition to being usable.

In order to find suitable groups of scientists and volunteers, we interacted with a variety of organizations. Meeting with the California Environmental Protection Agency helped us identify watershed health as a critical area that is also amenable to data gathering by average citizens. Interviews and brainstorming sessions with the California State Water Resources Control Board and the City of San Jose Environmental Services Water Resources helped us understand their priorities and need for data. We then met with local watershed monitoring organizations to understand the capabilities and limitations of their current citizen monitoring practices.

Having established the data needs of the stakeholders and accessed the abilities of volunteers we designed and developed Creek Watch. Creek Watch combines both an iPhone application and website. The iPhone application enables citizens to contribute water flow and trash data about creeks and rivers. The website, www.creekwatch.org, allows citizens to view contributed data, and provides scientists access to all of the collected data.

We then conducted a user study with 10 scientists from the City of San Jose Environmental Services Water Resources team that were not involved in our earlier study or design sessions. We tested the Creek Watch interface, assessed the usefulness of data it collects, and identified opportunities for improvements in the application and website. Overall participants were very pleased with Creek Watch, and they are already using the data it collects. The interviews and user studies enabled us to determine what data is useful to the scientists and conforms to the standards that they must comply with.

In the following sections we describe our series of contextual inquiries with state and local water control boards and with volunteer watershed monitoring groups and the importance and challenges in citizen science. Next, we explain our design of the Creek Watch iPhone application and website, how we took the needs of these different groups and converged upon a set of data and collection standards, and our user study with members of our local water control board. We discuss lessons for citizen science design and key implications of our work, and conclude by identifying future directions for research.

PRIOR WORK

The recruitment of citizens to participate in large scale monitoring has a long history, particularly in observing phenomena in the environment. Animal population research

has been aided by citizen science in a number of ways: in addition to the Christmas Bird Count [18], Cornell Lab's bird monitoring program [6] attracts over 200,000 volunteers each year. A similar project exists for counting fish [21]. Technology has enhanced the participation of citizens providing communication, documentation and measurement tools. The CONE project at UC Berkeley allows volunteers to contribute to bird sighting numbers using autonomous robotic cameras deployed in birding hotspots [24]. In the Lost Ladybug Project [15] citizens submit digital photos of Ladybugs along with location and weather information.

Citizen science provides a mechanism for participants to learn about and improve their community. Participants in the GarbageWatch project contribute pictures of garbage to help figure out where to place recycling bins [29]. Urban citizens equipped with sensors track neighborhood air [13,27,33] and noise [16] pollution. Citizens in Virginia have organized a volunteer group to monitor, consolidate and share data on local stream and rivers to maintain and improve the health of the Rivanna watershed [26]. Many cities use the non-emergency number 3-1-1 to implement a Citizen Service Center for residents to report needed roadwork, abandoned vehicles and other neighborhood code enforcement issues [4]. On-line versions of 3-1-1 developed by several cities provide citizens the ability to digitally interact with city services [19].

The value of mobile phones as tools for citizen science has been recognized, with examples such as air quality [20], and the Oil Reporter service which allows volunteers to report on areas affected by the Gulf Coast oil spill [11].

Citizens' role is not just limited to data collection. Technology makes it possible for citizens to volunteer their tools and brainpower. Idle time on volunteers' home computers is being used to search for extraterrestrial life [28] and study protein folding [25]. The Galaxy Zoo [23] leverages the human brain's exceptional pattern-recognition skills, recruiting citizens to classify galaxies from photos taken by a robotic telescope. This type of citizen science is a form of crowd sourcing, wherein people can perform distributed tasks towards a greater goal [10].

UNDERSTANDING THE PROBLEM

We began the Creek Watch project with a discussion with the California Environmental Protection Agency of areas in which volunteers equipped with mobile devices could be empowered to help their environment. We conducted a series of interviews and brainstorming sessions with ecologists in the state agency to understand current practices and problems in environmental management. Our discussions identified the monitoring of watershed health as an ideal application of citizen science: waterways are numerous and geographically distributed, local water agencies lack the manpower to monitor all of them, and citizens are aware of the importance of water. While we choose to focus on waterways, the same basic issues of

large geographic scope and limited organizational manpower occur in many environmental domains.

Manpower and Access

Watersheds are geographically large, and monitoring all of the waterways within them would require a significant investment of time and effort that is beyond most water organizations. For example, the city of San Jose, California's third largest city, has over 700 miles of creeks – far too much ground for a local agency to cover, especially when creeks run through private land that officials cannot access without permission. In practice, this means many areas never get visited by officials and receive no monitoring.

However, many of these creeks run through parks and recreation areas visited by citizens every day; these citizens could perform some monitoring if the process were sufficiently convenient. Creeks that run through private land could be monitored by the land owners. As a result, watershed monitoring is well positioned to take advantage of the key potential of citizen science: there are many more people who care about the environment than just those who are professionals in the field.

Interacting Stakeholders

In addition to supplying manpower any solution must address data standardization and sharing among a variety of interacting stakeholder organizations. Efforts in this area must consider the needs of the different types of groups involved with environmental management:

1. Government organizations: these groups are charged with managing environmental resources at the state, regional, county and city levels. In the context of watershed health, these organizations are responsible for enforcement of regulations pertaining to water use, wastewater discharge, and pollution.
2. Private groups: these are typically consulting companies, which work closely with these government organizations. They are typically funded in part by grants for environmental programs. Private parties also often hire these groups, for example to conduct evaluations for environmental permits.
3. Volunteer groups: these are typically organized on a city or county level. These groups have varied levels of organization and coordination, from twice-a-year outings to pick up trash, to monthly monitoring programs with professional water testing equipment. Volunteers who act as citizen scientists are members of one of these groups.

Data sharing, re-use, and standards

Beyond collecting data, one of the biggest challenges faced by organizations is sharing data - if two groups do not gather data in precisely the same fashion, it may not be comparable. The way in which a test is performed is as important a part of the data as the test results. Quality

assurance (QA) techniques often differ between organizations. Standardization is critical, but with many different organizations collecting many different measures for different reasons, it is often lacking in practice.

Even for programs where QA has been standardized, the problem of data sharing typically remains unsolved. Historically each government organization has maintained its own databases with poor interoperability. Currently, data is most frequently exchanged as completed reports or as spreadsheets. Recent legislation in California has mandated a state-wide system to share this data, but the system is still under development.

Getting information from private and volunteer groups can be even more difficult. Watershed volunteer groups typically aggregate data they have collected into a yearly report which is made available to government organizations as a summary document. More sophisticated groups may include a spreadsheet of the original data. Because these groups typically have a specific driving mission (e.g., searching for mercury or other toxins local to their area, tracking water flow rates which affect fish spawning, etc.), the data they collect is focused on a specific problem and may not include information that other organizations need to make further use of the data.

Our discussions with the California EPA revealed that there are over 250 organized and recognized groups in California working on watershed health [8], yet their idiosyncratic data collection, QA, and sharing practices make it likely that only a few will have their data used by others. We conclude that for a citizen science project to succeed in having its data widely used by potential consumers, the data must be easily shared, easily understood, and have some quality assurance mechanisms. In our work we apply human-centered design methods to the design of the data collection and format to ensure usefulness by the scientists who are the primary end users of the data.

CONTEXTUAL INQUIRIES

Following on our brainstorming sessions with the California Environmental Protection Agency, we conducted a series of contextual inquiries [1] into creating a citizen science mobile application for watershed health monitoring. We conducted in-depth interviews, on the phone and in the field when possible, with eleven scientists, environmental workers, and water monitoring volunteers from across seven organizations. From the government side, we interviewed four ecologists in the California State Water Resources Control Board and the City of San Jose Environmental Services Water Resources. From the private sector, we interviewed three ecologists from San Francisco Estuaries Institute (SFEI) and the Southern California Coastal Water Research Project (SCCWRP). In the volunteer sector, we interviewed three volunteers from two local groups, the Stevens & Permanente Creeks Watershed Council (SPCWC) and the Alameda Creek Alliance. We also interviewed the volunteer outreach organizers of the

Water Environment Federation, an international federation of water quality experts with over 80'000 members, including most of the environmental scientists we interviewed.

The primary goal of these interviews was to identify (1) the data most needed by environmental organizations that could be gathered by citizen scientists, (2) how it would be used, and (3) how to best ensure that the data would actually be useful, including understanding the protocol and format required to make the data acceptable to the scientists, The challenge of getting useful data from citizen scientists and ensuring that it gets shared with existing programs that can leverage it was emphasized by our first interview subject:

“In the early days it was an activity to engage understanding in water quality. It didn’t matter what they did as long as they did something. Then we realized no one uses the data and they aren’t monitoring for things we care about.”
Interview #1, State Water Control Board

Critical Data: Water Flow and Trash

Problems in watershed management are numerous and varied, including the improvement of habitats for aquatic life, water collection for drinking and irrigation, flood and erosion control, and the identification and control of pollution. The data that could help these areas are equally numerous (e.g., water flow, turbidity, Ph levels, dissolved oxygen, measurements of contaminant levels, aquatic life surveys). While much of this data requires specialized equipment to measure, some very helpful data requires only simple observation. We identified water flow rate as a candidate datum from our interviews:

“Wherever people go, if there’s a creek, it would be great to get info on flow. Qualitative and quantitative: take a picture and send to a database, GPS tagged. This would be tremendous – we need data on flow.” Interview #1, State Water Control Board

Water flow data is needed by several of the stakeholders we interviewed. Water flow rate in creeks is critical to understanding the health of a watershed: these small waterways, which in aggregate hold most of the water in California, are habitat for aquatic life, lead to drinking water reservoirs and farm fields, and are unfortunately also a potential pathway for pollution. Without an accurate picture of when and where water is flowing in a watershed, it is very difficult to understand the health and capacity of waterways, making water management more difficult. Capturing water flow data using the format and protocol required by the scientists assures that this data could be used by the stakeholders. Unfortunately, water flow data available to environmentalists is spotty at best.

“We’re not getting a lot of [flow] data. Most is generated by discharge permits: when someone has a permit to discharge [waste water] into a stream, they are required to monitor flow, so we get that

information. But...there are many more places we can look then we have people or resources for. ... Flow is kind of hidden, but it’s really important. What happens is when we have to get it its too late to go back in time.” Interview #4, State Water Control Board

Fortunately, flow data is easy for citizen scientists to collect. No equipment is needed to determine if water is flowing or not, and if creeks are full or low. Furthermore, people pass by these waterways all the time in parks, on walks, and around their neighborhoods.

Each of the groups we interviewed had a different reason for wanting flow data. At the state-level, flow data is desired for water management and planning. At the city level, flow data is desired for tracking pollution. One of the local volunteer groups (SPCWC) wanted flow data for a study of fish spawning they were conducting. One of the private consulting groups (SCCWRP) wanted flow data for environmental assessments they conduct as a service for construction permits. The other (SFEI) wanted flow data for a longitudinal study of human effects on bay area water supply. These varied needs for the same flow data suggests that the data should conform to standards to ensure maximum re-usability.

We also identified data concerning trash as important to our stakeholders and amenable to citizen reporting.

“Many of the most serious water quality problems in California are associated with non-point source pollution. ... Trash is a severe non-point source problem. Trash clogs our waterways blocking fish migration paths, impairs aquatic life and poses a threat to many beneficial uses of our creeks and streams.” – Interview #1, State Water Control Board

While many groups regularly conduct trash surveys and cleanups, there is simply too much ground to cover. The California State Water Resources Control Board is required by the U.S. Environmental Protection Agency (EPA) to maintain a list of waters that are considered "impaired", i.e., they are too polluted to be used for swimming, fishing, drinking, or other beneficial uses [30]. In California, trash is the second most common impairment pollutant [7].

To identify impaired waters, the state water board must collect and assess water quality data and determine if it meets standards. The process starts with stakeholders who alert the water board of locations where water quality standards may not be met. Citizens can alert state boards of pollution but need to present evidence. Photographs, tagged by location and time, provide tangible evidence of trash in waterways. Photos can help identify the type of trash (e.g., bottles, paper, car parts, and medical waste) and potential sources (e.g., individual littering, illegal dumping of landscape and construction debris). Trash occurs most frequently in sites with high public access [17], which bodes well for citizen monitoring of trash levels. The culprits for watershed pollution are often unaware of the

extent to which their activities are damaging the environment.

“It used to be that industry was the problem. But the average resident (collectively) is now the problem. People’s pesticides, car washing, dog poop, garbage that we drop it’s a really big problem.” – Interview #7, City of San Jose Environmental Services

The problem of getting data on such simple measures as flow rate and trash is complicated by the fact that good news is often assumed not to be worth reporting.

“We have a citizen stream keeper program: people go out at least once a month to a creek to observe it ... but if they don’t see anything bad they don’t report, so we don’t know.” Interview #6, Local Water Monitoring Volunteer Coordinator

To be successful, we concluded that our citizen science application must equally support and encourage the reporting of both problems and situations where nothing seems amiss.

We concluded from these contextual inquiries that water quality data consumers would benefit from an application that enabled citizen scientists to report on water flow and trash, and that the data needed to be collected following a protocol and format acceptable to the end users (scientists). A smart phone is ideal for collecting this data, since it is always on hand, can collect pictures and accurate location data, and can simplify submitting data to a central organization.

More advanced needs

In addition to flow and trash data, we identified several other standard measures of water quality that would be useful, such as temperature, turbidity, dissolved oxygen, and Ph level. Unfortunately, for a citizen scientist to report on this data, they need to have water testing kits or more specialized equipment, complete with the training to collect data according to accepted protocol in order for this data to be acceptable for scientific studies. For example, the

Water Level	Flow Rate	Trash
Dry: No water Present	Still: Water is present but is not visibly flowing	None: no trash in the water and surrounding area
Some: Water fills less than 75% of the channel	Slow: Water is present but is barely moving	Some: A few items of trash such as cans, bottles
Full: Water reaches up almost to the top of the banks	Fast: Water is present and flow is easily detected	A lot: Ten or more items of trash

Table 1: Water observation definitions used in Creek Watch, from the California EPA’s Surface Water Ambient Monitoring Program

Stevens & Permanente Creeks Watershed Council volunteer group used tools ranging in price from several hundred to several thousand dollars, operated by a trained biologist, and even so would only take samples on a monthly basis due to limitations on lab time for sample analysis.

By limiting our application to data that can be collected by citizens without specialized training, and that complies with the format and protocol required by scientists, we design our program to provide useful and usable data. From our discussions with the scientists who vet citizen provided data, and from participating in data collection field trips with the volunteers, we learned that data needs to be sufficiently easy to collect in a reliable and predictable way, so scientists can trust and therefore use the data. We also observe that participation in a volunteer monitoring group does encourage learning. All of the volunteer groups we interviewed have a strong educational component, to raise awareness of watershed stewardship, but also to train the volunteers to make good quality measurements. It is our hope, as well as the hope of the scientists we interviewed, that this simple and easy data reporting will provide a bridge for participants to more advanced involvement in water quality measurement.

DESIGN & DEVELOPMENT

We designed and built Creek Watch to be a participatory creek monitoring system that allows citizen scientists to capture water flow and trash data and scientists to view and analyze the collected data. We designed the system as two parts: a data collecting iPhone application, and a data sharing and viewing website, creekwatch.org.

Data Format: Conforming to Standards

To ensure that the collected data would be usable by existing programs, we drew on the California Environmental Protection Agency’s Surface Water Ambient Monitoring Program Bioassessment Standard Operating Procedures [2] and the Rapid Trash Assessment Protocol [3]. These manuals define procedures for reporting on water flow (composed of water level and flow rate), and trash levels. With consultation from the California State Water Resources Control Board, we adopted the definitions in Table 1 for Creek Watch.

These observations, combined with a photo of the creek, GPS location and timestamp, provide the data that water monitoring organizations most requested.

Data Collection Application

We developed the data collection tool as a mobile application for the iPhone (see Figure 1). The application contains five views divided into tabs: *Report, My data, Browse, Instructions, and About*. Central to the application is the data reporting view, where users can Report on a creek they are looking at using the data format described above. From the report view users can access a set of photo examples defining the different choices for each data item.



Figure 1: Creek Watch iPhone application screenshots

While the application attempts to immediately post reported data to the server, it can also store reported data locally on the phone when the network is unavailable (as is often the case for creeks alongside hiking trails). Users can browse through data they have uploaded as well as data that still needs to be uploaded in the *My Data* section. Users can initiate another upload attempt for any data that the application has not yet loaded. Users can also manually specify the location for a report if necessary (if, for example, the GPS on the phone could not determine the user's position) before uploading a report.

The *Browse* view provides a map visualization that enables users to see recent data points collected by others in their area, including photos. The *About and Instructions* views provide general information about and detailed instructions for using the application, respectively.

Data Sharing Website

The server we implemented at creekwatch.org collects data submitted by iPhones running the Creek Watch application, and presents this data in two interactive formats – a map (see Figure 2) and a table (see Figure 3).

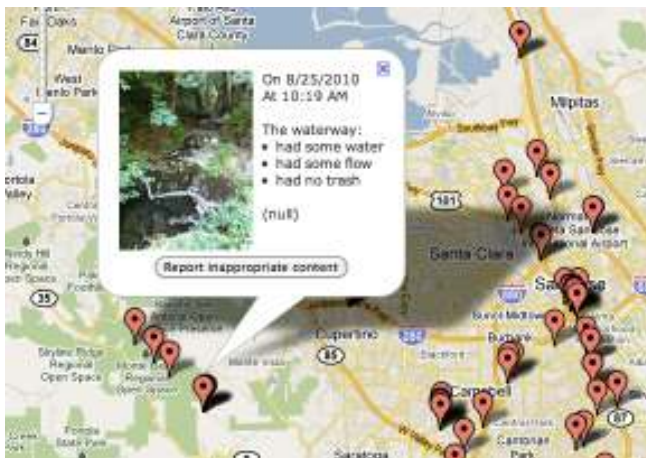


Figure 2. Map view of the collected data at creekwatch.org

Map-based data visualization

A map of all collected data points is the central visualization on creekwatch.org. We designed the map to enable browsing of the data by data contributors who wish to “poke through” their own and others’ data. We built the visualization as a Google Maps mash-up with the Creek Watch database that displays the location of every data point as map pins, centered on the user’s current location. The detailed data collected at each location is available as a popup window with a photo of the creek, or as hover text over each pin (see Figure 2).

Table based data visualization

We designed the table view for scientists to easily work with the collected data (see Figure 3). Each column provides filters (time, location, data values), and scientists can export the filtered data to a csv (comma separated values) file for inclusion and integration with existing projects. The table view is publically available, such that anyone can manipulate and download the data in this fashion; however the intended audience is data consumers in environmental management organizations.

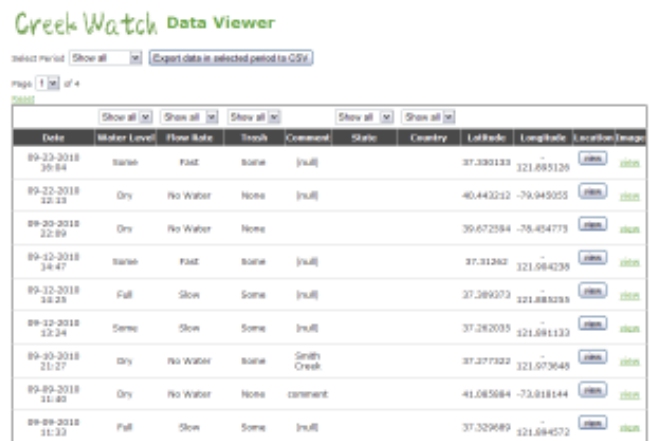


Figure 3: Table view with filters of the data collected at creekwatch.org

Design process

In building Creek Watch we employed a classic iterative design process. Following our contextual inquiry interviews, we built an early version of the application to collect flow and trash data. We demoed this first version at an open community meeting with members of the state and local water control boards, and local volunteers. The initial feedback was positive, and we made several small changes to the workflow and appearance based on this review. Examples changes include starting on the Report view instead of the Browse view, and adding an Instructions view to more clearly explain how to use the application.

We then performed an in-house usability evaluation with four iPhone users from our research lab to identify bugs and usability problems. We fixed several additional problems including using a higher contrast between interface text and the background for easier reading in sunlight, and leaving more space between buttons to reduce accidental clicking. Following this usability test and several rounds of field testing with the team, we deployed our beta version for more real world user testing.

USER STUDY & FINDINGS

We designed and conducted a field deployment study to evaluate the effectiveness of our system. We recruited 10 environmental scientists in the City of San Jose Environmental Services Water Resources Department who own iPhones. The study participants included four environmental agents, three environmental outreach or volunteer coordinators, two managers, and one environmental analyst. Six participants were male and four were female. The average age was 45 years old with a standard deviation of 9.7 years. All participants had academic degrees related to environmental management or ecology. We chose to focus on environment scientists for our study participants in order to assess both the usability of the Creek Watch system and also the utility of the data it collects.

None of our user study participants were involved in the brainstorming, early interviews, or design of Creek Watch. While we did work with other scientists from their department to develop the application, for the user study we wanted to see how unbiased scientists would react.

The user study consisted of three steps: a pre-study interview, a deployment period, and a post-study interview. The study lasted for 3 weeks.

Pre-study interview

The purpose of a pre-study interview was to understand the data needs and collection practices of the participants. Pre-study interview questions included:

- When you need data/information, how do you collect/access it?
- Do you share data you collect? How, and how is it used?
- How often do you participate in field work?

- What are the difficulties you perceive in water monitoring?

Each interview lasted about thirty minutes, followed by an introduction to the Creek Watch application and website. At the end of each interview we collected demographic and occupational information.

Deployment period

We asked participants to use both the Creek Watch application and the websites whenever they felt was convenient over a three week period. During this time they logged 65 data points in the greater San Jose area.

Post-study interview

The purpose of the post-study interviews was to understand both how useful our users felt this data would be and what their reactions were to using the app as a data collector. Post-study interview questions asked about the subjects' use of, opinions about, and recommendations for the Creek Watch application and website. Each interview lasted between thirty minutes to an hour. At the end of the interview, we gave participants a \$10 Amazon gift coupon as compensation for their time.

FINDINGS

In our pre-study interviews, participants reinforced the lessons from our initial investigations: traditional scientific methods are expensive and don't scale. The watersheds are too vast to be monitored adequately by their small staff. Several of the participants were in the process of deploying a river monitoring system in Coyote Creek, one of the largest waterways in San Jose. The system consists of nine autonomous units, each costing approximately \$8000. The units continuously measure five basic characteristics of the water but must be periodically visited to retrieve data. With units this expensive, widespread deployment is not an option; most of the creeks will remain unmonitored.

While autonomous units are not the only way to measure creek health, site visits by field agents cannot provide complete coverage of the watershed either. When describing the most important difficulty he faces in his job, one of the field agents stated:

"Access is a problem. The big thing is we can only measure in places we can get to, resulting in data gaps."

This access problem was echoed as the top problem by three out of the four field agents.

As our participants are both collectors and consumers of the data, our study provides insights into both aspects of Creek Watch. We therefore divide the findings from our user study into those relating to the consumption of data and those relating to the collection of data.

Data Consumption

Our participants agreed that Creek Watch provided very useful data while also providing a low barrier to entry for

users. They found the trash data of particular interest. Participants saw opportunities to immediately make use of the data for existing programs in tracking watershed health:

“We would use this data... That’s our big focus right now- trash.” (Project manager)

“A great tool to monitor creeks and help us identify problem areas- one of my coworkers is on creek cleanup and trash keeps coming up.” (Project manager)

“For our work in particular [enforcement of dumping regulations], I would be interested in the trash information- to see where folks are finding trash along the creeks.” (Field agent)

As an example, several participants commented on the particular usefulness of this data for local trash cleanup events. One of the volunteer coordinators who manages cleanup events identified the value provided by being able to easily identify “trash clusters” in the data points contributed by other users:

“When you get a lot of data points you can see where most of the trash is in the creeks. When groups have cleanup events, now we can find these trash areas.”

We found these findings particularly interesting because during our initial contextual inquiries flow data appeared to be in greater demand than trash data. This difference in emphasis may be a reflection on the particular responsibilities of our study participants. While the scientists we interviewed during the contextual inquiry discussed longer term plans for the data’s use in environmental planning, the city water board workers seemed more concerned with immediate action items. We note that the ability for a citizen science project to yield such short-term wins may increase participation in both contributing and consuming the data.

Participants did also consider flow data to be useful. One participant commented that Creek Watch was “a good way to inventory streams and maybe even keep track of the ephemeral nature of some streams and creeks.” The general consensus among our participants was that flow data is useful in the long run for planning, trending and mapping. These results suggest that combining data that potential consumers can use immediately with data designed for longer-term use may increase the likelihood of success for a citizen science project by increasing the potential set of people interested in the data.

All of the participants emphasized their belief that Creek Watch would promote public engagement, increase awareness of watershed health, and provide informal science education for volunteers. One of their goals is to make city residents think more about their water and where it comes from, and twice about littering. Creek Watch provides a way for them to engage with their constituents over watershed management.

While participants were largely pleased with the desktop web interface for accessing data, they did make several suggestions for improvements. Several suggestions centered on new ways to filter the contributed data. For example, a number of participants wanted to be able to easily filter the data in the table by city. Others wanted the ability to filter the data presented in the map view as well so that they could use the map view for their work.

The latter suggestion was somewhat unexpected, as in our pre-study interviews most participants agreed that they would prefer to access data in table form. However, we found in our post-study interviews that the map view was more popular for data consumption. Participants described “getting a sense of the area” by browsing photographs, as well as “finding trash” using the map view. The preferred presentation mode may depend on the intended uses, with more immediate uses (such as identifying locations with lots of trash) drawing more heavily on the map while uses that involve longer-term analysis (such as changes in flow) may rely on the table format or on the ability to download and work with the data separately.

Data Collection

All participants commented how simple and easy to use they found the application. As one participant comments:

“Very easy, very quick. I pull up at the creek next to the business I am inspecting and use it.”

Participants also uniformly stated that they felt that making the application as simple and easy as possible to capture data was critical.

Five out of ten participants had friends and family members try the application out with similar results, reporting that the other people found the application “simple to use,” and “easy.” These results suggest that, in general, volunteers without a scientific background will be able to use the application without difficulty.

Furthermore, our study revealed that participants considered using the application fun. One example participant described how they turned using the application into a game:

“We played find the creek: pull over at a creek, jump out of the car, find a good vantage point, enter the data.”

Both of our participants with children of school age took their kids out to a creek to use the application and employed it as teaching tool. One participant recalled,

“We talked about recording only what you see and not what you think- a nice teaching moment.”

This participant was reacting to an interesting data collection phenomenon:

“When we went to a place that I could not see any trash, [my son] said ‘some trash’ because of what he knew from other times he’d seen in the area.”

We make two observations about this comment. This behavior emphasizes a potential problem with any user-contributed data: errors that occur to user mistakes, biases, etc. Because data consumers are unlikely to use any data that they do not trust, a citizen science application should thus ideally provide a mechanism to allow consumers to easily verify the data. We believe that the inclusion of pictures in the data helps provide this “check”. In this case, a data consumer could see if the picture actually shows any trash.

Several participants requested a comment field to write down a description of what they were seeing. This request is particularly interesting because none of these participants could think of a way that, as data consumers, they would have a use for this data. They simply “wanted to be able to add a little more data.” The disparity between their desires as data collectors and as data consumers reinforces the value in studying both aspects of a citizen science application.

Five participants reported using the website to compare the data they had collected with other data points in the system. One reported:

“I would occasionally browse data [on the map view]-looking where other people were making observations and checking out their photography.”

Checking data on the map “as a reference” to make sure they were taking data at least as good as their peers was a common theme. While the map was designed for sharing data, it was interesting to see its frequent use as a method of self-checking. We note that this behavior presents an opportunity to seed expected use by ensuring a site has an initial set of “ideal” data points for early adopters to refer to. In addition, it suggests that a potential addition to the website would be the ability to select a small set of data points and compare them in a single view.

Overall our participants were very pleased with Creek Watch both as data collectors and data consumers. Even after the study formally concluded they continue to use the application. Subsequent conversations revealed that two of the field agent users plan to incorporate the application into their regular data collecting practices, replacing their current pen-and-paper data collection. While we did not intend Creek Watch to primarily be used by experts in their professional data collection, we regard that adoption as validation of the usefulness of the captured data.

CONCLUSIONS AND FUTURE RESEARCH

Our goal in this research was to design a citizen science application that would provide scientists useful data as well as providing a usable interface for untrained volunteers. A key part of that process was the use of HCI methods to

investigate the needs of data consumers. We worked closely with state, local, and volunteer environmental groups and employed a variety of contextual inquiry tools to determine what data they wanted and the requirements to make the data usable (e.g., format, reliability, standardization).

Those investigations revealed the following lessons that we believe generalize to other citizen science applications:

- When HCI methods are applied to the data as well as the interface, the resulting system can collect more useful data.
- Providing reliable and standardized data enables organizations beyond initial stakeholders to benefit from the data in unanticipated ways.
- Combining captured data that organizations can use immediately with data that provides long-term, aggregate value may increase the chance of success for a citizen science application.

Based on an initial study of Creek Watch, we believe that we have succeeded in designing a citizen science application that is both usable and useful. In addition, the collected data is already being used by the City of San Jose, including new uses that we did not originally foresee (e.g., the use of the data for trash clean up site identification and use as a data collection tool by field agents). Other water management organizations within California have expressed interest in using Creek Watch.

As next steps we are interested in exploring long-term usage patterns among data contributors and data consumers. For data contributors, we plan to examine factors such as adoption rates, sustained use, and data quality. For data consumers, we plan to explore how to improve the existing data presentation mechanisms and determine what new visualizations would significantly improve the ability to use the gathered data effectively. More broadly, much work remains in determining how best to design citizen science applications that are widely adopted and able to make an actual impact through their collected data.

ACKNOWLEDGMENTS

The authors would like to gratefully acknowledge the many scientists and volunteers who worked with us on the development of Creek Watch from the following organizations: the California Environmental Protection Agency, the California State Water Resources Control Board, the City of San Jose Environmental Services Water Resources, San Francisco Estuaries Institute, the Southern California Coastal Water Research Project, the Stevens & Permanente Creeks Watershed Council, the Alameda Creek Alliance, and the Water Environment Federation. We would also like to express our particular thanks to Erick Burres and Carol Boland from the state and city water control boards, respectively, to Mondy Lariz and Joanne McFarlin from SPCWC, and to our user study participants.

REFERENCES

1. Beyer, H. & Holtzblatt, K. Contextual Design: Defining Customer-Centered Systems. San Francisco: Morgan Kaufmann. ISBN: 1-55860-411-1. 1998.
2. California Environmental Protection Agency. Surface Water Ambient Monitoring Program: Bioassessment Standard Operating Procedures. 2010. swamp.mpsl.mlml.calstate.edu/resources-and-downloads/standard-operating-procedures
3. California Environmental Protection Agency. Surface Water Ambient Monitoring Program: Rapid Trash Assessment Protocol. 2010. www.swrcb.ca.gov/sanfranciscobay/water_issues/programs/stormwater/muni/mrp/WaterBoard%20Trash%20Assessment%20Method%20SWAMP_v8.pdf
4. City of Knoxville Tennessee. Neighborhood Codes Enforcement. www.ci.knoxville.tn.us/services/codes/. September 2010.
5. Cohn, J P. Citizen Science: Can Volunteers Do Real Research? *Bioscience* 58:192–197. 2008.
6. Cornell Lab of Ornithology. Mission: Citizen Science. www.birds.cornell.edu/netcommunity/page.aspx?pid=1664. September 2010.
7. Environmental Protection Agency. 2006 Section 303(d) List Fact Sheet for California. EPA Office of Water, Washington DC, 2006. oaspub.epa.gov/tmdl/state_rept.control?p_state=CA&p_cycle=2006#TPOL
8. Environmental Protection Agency. TMDL Program Results Analysis Fact Sheet # EPA841-F-10-002. EPA Office of Water, Washington DC, January 2010. www.epa.gov/waters/doc/factsheets/303d_impaired_waters_gis.pdf
9. Gonsalves, A. Android Phones Steal Market Share. *InformationWeek*, April 7, 2010 www.informationweek.com/news/mobility/smart_phones/showArticle.jhtml?articleID=224201881
10. Howe, J. Crowdsourcing: Why the Power of the Crowd Is Driving the Future of Business. *Crown Business*, September 15, 2009. ISBN-13: 978-0307396211.
11. Interidea. Deepwater Oil Reporter Crowdsourcing Platform. oilreporter.org. 2010.
12. Irwin, A. Citizen Science: A Study of People, Expertise and Sustainable Development. Routledge. London, UK. 1995.
13. Kim, S, and Paulos, E. inAir: Sharing Indoor Air Quality Measurements and Visualizations. *In Proceedings of the SIGCHI Conference on Human Factors in Computing System. CHI '10*. ACM, 2010.
14. Lane, N, et al. A Survey of Mobile Phone Sensing. *Ad Hoc and Sensor Networks*. IEEE Communications Magazine. September 2010.
15. Lost Ladybug Project. lostladybug.org. 2010.
16. Maisonneuve, N, et al. Citizen noise pollution monitoring. Proceedings of the 10th Annual international Conference on Digital Government Research: Social Networks: Making Connections between Citizens, Data and Government. May 2009.
17. Moore, S, Cover, M, Senter, A. A Rapid Trash Assessment Method Applied to Waters of the San Francisco Bay Region: Trash Measurement in Streams. California Environmental Protection Agency, 2007. www.swrcb.ca.gov/rwqcb2/docs/swamptrashreport.pdf
18. National Audubon Society. History of the Christmas Bird Count. birds.audubon.org/history-christmas-bird-count. 2010.
19. Open Plains. Open 311: A collaborative effort to create an open standard for 311 services. open311.org. 2010.
20. Paulos, E, Honicky, R, and Hooker, B. Citizen Science: Enabling Participatory Urbanism. *Handbook of Research on Urban Informatics: The Practice and Promise of the Real-Time City*. Information Science Reference, IGI Global, 2008.
21. Reef Environmental Education Foundation. The Great Annual Fish Count. fishcount.org. 2010.
22. Shirky, C. Cognitive Surplus: Creativity and Generosity in a Connected Age. Penguin Press HC, June 10, 2010. ISBN-13: 978-1594202537.
23. Sloan Digital Sky Survey. Galaxy Zoo. galaxyzoo.org. 2010.
24. Song, D, Goldberg, K. Networked robotic cameras for collaborative observation of natural environments. *Robotics Research*. Springer 2007.
25. Stanford University. Folding@home distributed computing. folding.stanford.edu. 2010.
26. Stream Watch: Community Based Stream Monitoring for the Rivanna Watershed. streamwatch.org. 2010.
27. Sukhatme, G. Air Visibility Monitoring. robotics.usc.edu/~mobilesensing/Projects/AirVisibilityMonitoring. September 2010.
28. University of California, Berkeley. Seti@Home. setiathome.ssl.berkeley.edu. 2010
29. University of California, Los Angeles. Garbage Watch. garbagewatch.com. 2010.
30. Weiner, J. Fishable, Swimmable, Drinkable? California Coastkeeper Alliance, May 12, 2010. www.cacoastkeeper.org/news/fishable-swimmable-drinkable
31. Whitney, L. Cell phone subscriptions to hit 5 billion globally. CNN, February 16 2010. reviews.cnet.com/8301-13970_7-10454065-78.html
32. Wikipedia. en.wikipedia.org/wiki/Clickworkers. 2010.
33. Willet, W, et al. Common Sense Community: Scaffolding Mobile Sensing and Analysis for Novice Users. *Pervasive 2010*. May 2010.