Applying Geocaching to Mobile Citizen Science through SCIENCE CACHING

Authors

Matthew Alan Dunlap, University of Calgary
Anthony Tang, University of Calgary
Saul Greenberg, University of Calgary

Abstract

Citizen science occurs when volunteers work with scientists to collect data at particular field locations. The benefit is greater data collection at lesser cost. Yet this type of citizen science has a variety of known problems. Of these, we focus on four specific problems that we believe can be mitigated by applying aspects from another thriving location-based activity: the geocaching treasure hunt as enabled by mobile devices. To flesh out this idea, we developed SCIENCE CACHING, a prototype mobile system and site preparation strategy that leverages concepts from geocaching. To ease problems in *data collection*, sites are treated as geocaches: volunteers find them opportunistically and use equipment and other materials pre-stored in caches. To ease problems in *data validation*, outlier data is flagged immediately on-site so that it can be immediately checked and corrected, and/or other volunteers are directed to that site for additional readings. To ease problems in *training,* volunteers visit training sites where they are both taught and tested against known measures. To ease problems in *volunteer coordination,* volunteers are automatically directed at particular sites of interest, and real-time communication enabled. We showed SCIENCE CACHING to citizen science experts, who confirmed the merit in applying geocaching and mobility to citizen science.

Keywords

J.9.a Location-dependent and sensitive
K.3.1 Computer Uses in Education
H.5.3.c Computer-supported cooperative work

Cite as:
Dunlap, M., Tang, A. and Greenberg, S. (2013) Applying Geocaching to Mobile Citizen Science through Science Caching. Technical Report 2013-1043-10, Department of Computer Science, University of Calgary, Calgary, AB, CANADA. T2N 1N4. June. 14 pages

Background

Scientists discover patterns and rules that define our world by collecting and analyzing large amounts of data. One way that scientists gather more data at lesser cost is through volunteers, a method known as citizen science (Silvertown 2009). Citizen science connects non-expert volunteers (citizen scientists) with the scientific process and natural world, where they do some of the work. Compared to using a small pool of scientists, citizen science provides an affordable means for collecting and processing large quantities of science data.

Yet, even with these benefits, citizen science has problems arising from working with disparate groups of volunteers without the same level of experience as the coordinating scientist (Silvertown 2009; Cohn 2008). As computer scientists, our goal was to mitigate four well-known problems in citizen science involving data collection that we believed amenable to technical solutions:

- *Data collection* - The collection of data at a site needs to be easy, methodological and well structured, yet volunteer collection – and the data produced – is haphazard.
- *Data validation* - Collected data needs to validated to be trusted, yet the effort of checking error-prone volunteer data is high (Wiggins et al, 2011).
- *Volunteer training* - Training is critical if volunteers are to perform new tasks and collect valid data, yet training of volunteers is time-intensive for scientists to do.
- *Volunteer coordination* - Scientists need to track citizen scientists and coordinate project work, which is difficult to do in practice due to unpredictable volunteer involvement (Silvertown 2009).

Our Approach: Geocaching and Mobility

In targeting these problems, we ask:

> *How can we use aspects of **geocaching** as well as **mobile technology** to inspire and design new solutions for **citizen science's** problems?*

Geocaching is an outdoor treasure hunting activity, where its players (*geocachers)* hide and/or hunt for 'treasure' using GPS-enabled devices. Treasures are hidden in *geocaches*: small containers typically containing instructions, a logbook and trinkets (prizes). Containers also tell the player that he/she has found the location. Geocaches are hidden all over the world, with their locations stored on centralized websites. Groundspeak's Geocaching (2013) is a notable example. Its webpages proves information about each cache: cache size, cache route, difficulty, GPS coordinate and notes from individuals who recently found the cache. Geocachers typically search for caches that have different characteristics (e.g. a difficulty rating) or by proximity (i.e., nearby caches). Geocachers find chosen caches by using a GPS-enabled device and information from the cache listing.

Geocaching content creation and maintenance - done almost exclusively by its users - is a core part of the geocaching experience. Geocachers share locations, treasures and experiences with others (O'Hara 2008) by reporting cache state online. They visit the cache to perform maintenance as needed, and – when they can no longer maintain it – give it up for adoption (Neustaedter et al. 2010).

Particular geocaching structures have facilitated its growth and health as a community:

- *Cache discovery* – The way geocachers learn about new caches to find
- *Cache site information* – How geocachers find the cache, using GPS coordinates, descriptions and sometimes photographs
- *Physical containers* – How these containers ensure the geocacher finds the correct location, and how they can be used to hold different objects
- *User-generated content* – How geocachers' create and maintain the same caches they visit.
- *On-line coordination* – How geocachers self-coordinate their activities.

We consider how these strategies can enhance citizen science data collection and mitigate the problems outlined earlier. In particular, we apply these geocaching ideas to citizen science through the use of mobile devices. For instance, these devices provide citizen scientists with many useful features, including cellular networking, on-screen interactions, real-time updates, GPS geo-location and various sensors including a camera.

SCIENCE CACHING

To explore and express our ideas on the targeted problems, we created SCIENCE CACHING. The system itself runs on mobile devices loosely coupled to a central server, where it serves to provide information about particular collection sites, collect data about those sites, and leverages physical caches contained within those sites. While SCIENCE CACHING is a running system, we believe it is best considered as a working sketch that illustrates the potential of geocaching as applied to citizen science, and for gathering feedback and critique about its design from domain experts.

We crafted a set of interaction scenarios based around the collection of data on quaking aspen trees at known sites. We chose a specific (*vs.* generic) project to afford a detailed exploration of how geocaching can be applied to specific needs, and how the software could be designed to fit those needs. Quaking aspen data collection was chosen primarily for pragmatic reasons: these trees were readily available locally, and are stationary. The data collection task was made deliberately simple so the focus could remain on the system design concepts and on how citizen scientists interacted with that design, rather than the particular intricacies and confounds that accompany a complex project. SCIENCE CACHING was then developed around those scenarios, with the caveat that many of its functions are hard-wired, and a few of those functions are realized only as stubs.

Scenarios involve several citizen science and scientist participants, various data collection sites and the SCIENCE CACHING mobile application. Each scenario illustrates the system and how it mitigates one of the four particular problems previously mentioned. In our story, scientist Jill is working with several citizen scientists trained at collecting specific information on trees located at specific sites, as well as other volunteers who are yet to be trained. Jill has already created several physical caches within various city parks, marked and detailed these site locations, and made quaking aspen sites available electronically.

Scenario 1: Data Collection

Our first scenario shows the most basic element of SCIENCE CACHING: collection of data by a citizen scientist at a particular site. The different steps taken by the citizen scientist are considered in turn: choosing a site, physically finding the site via the cache container, and then using tools from the cache container to perform and / or store data collection information. This takes advantage of various geocaching elements, namely the process geocachers use to opportunistically choose a site to go to, the way geocachers find the physical cache marking the site, and the ability for a cache to hold different objects. These elements are enhanced through the use of the SCIENCE CACHING, which assists in the choosing, finding and collection process. Our approach to finding a collection site was influenced by Han et. al.'s FloraCaching (2011), where citizen scientists find trees for data collection via photographs and GPS coordinates.

Our focus begins with Arthur, an amateur citizen scientist volunteer already trained on quaking aspen data collection, who has already used the SCIENCE CACHING to perform several data collections. At this particular moment, Arthur has some free time and checks to see if there are any nearby quaking aspen sites. Arthur starts SCIENCE CACHING on his personal mobile phone.

Finding Sites

SCIENCE CACHING initially displays the cache selection screen, which is a navigable map with various icons (Figure 1, left). Arthur sees his location on the map as the red figure icon (via GPS tracking), as well as all nearby collection sites. Arthur currently sees two duck monitoring sites, a tree monitoring site, and two water-monitoring site (which he is not yet

Left: Arthur sees himself and nearby sites on the map.

Right: after Arthur clicks on a site, information is provided about its project.

Figure 1: Screenshots for choosing a collection site.

trained for: one is blocked but the other is an open training site). Arthur is trained in tree monitoring and – because the site is on his way home – he selects a tree monitoring site. More information is revealed about the task (bottom of Figure 1, left). Since Arthur knows how to "record a variety of details about a quaking aspen", he accepts the task tapping the "Start this Project" button.

Arthur's next step is to *locate and move to the site.* The SCIENCE CACHING application modifies the navigation page to assist him in getting there (Figure 1, right) where, similar to most GPS systems, it updates his position on the map, shows where he is relative to the site location (the icon at the top right), and displays his distance from the site (currently 128 meters). Additionally, it describes the exact location of the physical cache (i.e., a black Tupperware container behind a particular tree), and provides him with a photograph of that location, which he can enlarge via double-tapping. Using all this information makes it easy for Arthur to navigate to the site and to find the cache by visually (and physically) searching the area. After finding the cache container, he verifies that he is at the exact correct spot for collection and presses the "I found the cache" button (Figure 1, right).

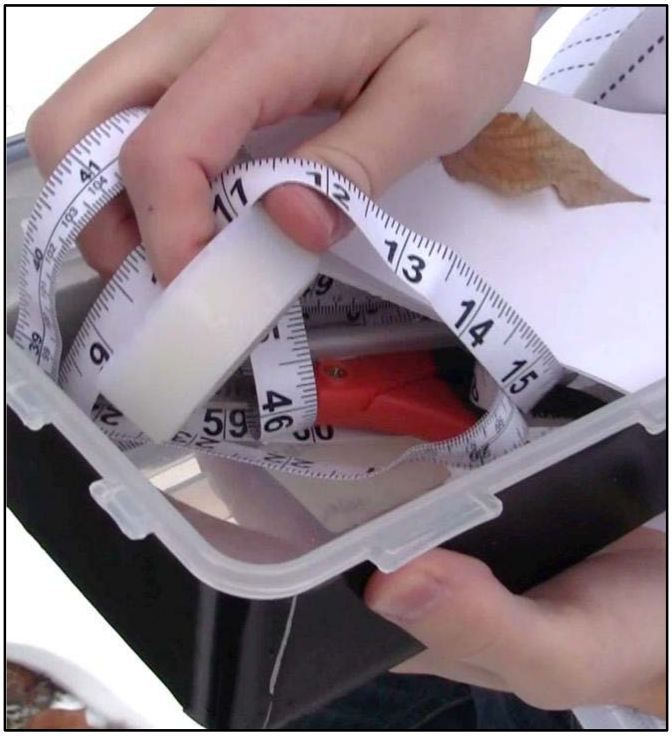


Figure 2. An example cache contains: tools needed to collect site data; informative guides; and serves as a place to store collected physical data for later pickup.

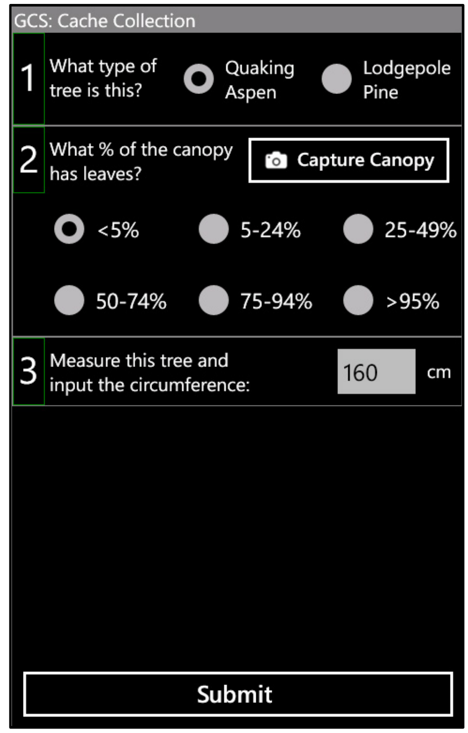


Figure 3. A filled in data collection screen.

Using Tools and Taking Samples during Data Collection

Arthur is now ready to perform data collection on the tree marked by the cache. He opens the cache container, which contains collection tools: a tape measure, a phenological guide, a leaf sampling logbook, a pen, scissors and other materials (Figure 2). His mobile device presents the data collection form for the site (Figure 3), where he is asked to answer several questions:

- *Identify the type of tree* the cache is located at. He looks at the tree and marks that it is a quaking aspen (Figure 3, first row).
- *Estimate the percentage of living canopy cover* the tree has, and take a picture of the canopy. Looking at the tree canopy, he sees that there are almost no living leaves, and records the canopy cover as 0-5%. Using the 'capture canopy' button, he takes a photograph of the canopy which allows a scientist to later review and validate his measure if necessary (Figure 3, second row).
- *Measure the tree's circumference*. He takes the tape measure out of the cache (Figure 2) and uses it for a tree measurement. Having this tool available on-site allows

> Arthur to collect data that he would be otherwise unequipped for. He then records this measure by typing in its value as 160 cm (Figure 3 third row).

After submitting that data, Arthur is then asked to place a leaf sample in the cache's leaf book and take a picture of it. He does this, using the container to store this sample (as in Figure 2) until a scientist or other volunteer can come by and pick it up. Arthur closes the cache container, places it back under the tree, and heads home. The entire sequence took about 15 minutes, and Arthur is pleased that he was able to contribute to this project.

Scenario 2: Data Validation

This next scenario illustrates several ways that data validation is afforded by geocaching as realized by the SCIENCE CACHING system.

Repeat Collections

Jill, the scientist in charge, had previous marked the cache visited by Arthur as a repeat collection site. This means that other citizen scientists will continue to see the cache on their listing (as in Figure 1, left) even if data had been recently collected from it, and can repeat the data collection. While this produces multiple collections for the same location, its major benefit is that it allows Jill to better validate her data. Jill can, for example, average the records (which somewhat minimizes the effect of errors). Alternately, she can inspect the data and throw out obvious outliers (i.e., those that differ significantly from others, or do not follow expected trends). She can also inspect questionable data: we already saw how Arthur collected redundant information, in this case photos associated with the data, which Jill can now examine to confirm the readings.

On-Site Data Validation

It is one week later, where we follow another citizen scientist Steve, who happened to choose the same tree collection cache as Arthur (as it was flagged as a repeat collection site). Steve is new to this process, and unfortunately makes two errors during data collection. We see how SCIENCE CACHING automatically compares Steve's and Arthur's data and – if a problem is detected – guides Steve to check his measures and (ideally) correct any mistakes (whether by himself or by Arthur) while still on site.

When Steve performs the data collection, he has let his tape measure sag (Figure 4), leading to enter an incorrect measurement of the circumference as 165cm (Figure 5, row 3). He also incorrectly includes the dead leaves in the canopy count as canopy cover, and thus records the count as 25-49% (Figure 5, row 2). When Steve submits his data, SCIENCE CACHING automatically compares values with (Arthur's) previous record. If there are unusual differences, SCIENCE CACHING displays the prior values in purple in context, and lists details of potential data conflict, along with previously taken photos at the mid-bottom of

Figure 4. Steve lets his tape sag, which produces erroneous readings.

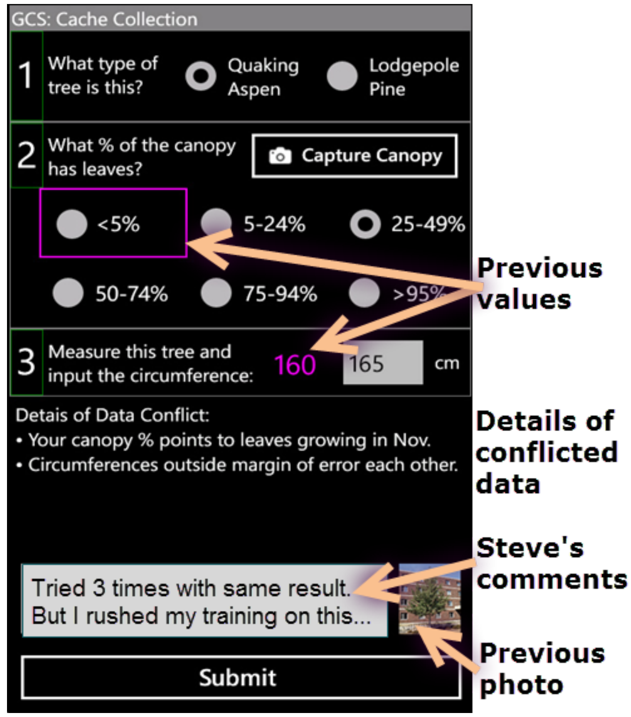


Figure 5. Steve's data collection form, showing potential conflicts and opportunities to correct or comment on that data.

the display (Figure 5). During training, Steve was told to double-check any conflicting measurements. He was also told to examine previously taken photographs to see if the photograph matches with the site (in case Steve or the previous citizen scientist measured the wrong tree).

Steve retrieves and reads the measurement manual located inside the cache, where he notes that dead leaves should not be recorded as canopy: he thus changes his answer to question 2. Unfortunately, the manual does not describe the sagging tape problem: even though he re-measures the tree three times, he gets the same result (and discrepancy). However, he is able to describe what he did in the text box at the bottom, where he notes that he rushed over his training on that aspect. He then submits the modified record to the system. Because this record still has a data conflict, it is flagged by the system as something the scientist needs to examine. Even so, the on-site feedback about possible conflicts meant that Steve was able to fix one of his errors, while still providing Jill with a bit more information that it is likely that the error was caused by Steve rather than Arthur.

Scenario 3: Volunteer Training

The next scenario shows how SCIENCE CACHING helps train new citizen scientists in real world settings, i.e., on actual sites prepared for training rather than in a classroom. The advantage is that the real-world, on-site experience makes the training far more immersive, potentially more effective, without requiring scientists to accompany and train each citizen scientist. Additionally, volunteer testing can be done against known values.

We follow Larry, where his SCIENCE CACHING application tracks what training is completed *vs.* what is left to do, where the latter includes quaking aspen data collection. Consequently, his map view (somewhat similar to Figure 1) displays nearby site icons of: open sites where he is trained to collect data; a blocked quaking aspen site where he is untrained and thus blocked from any collecting activities; and a quaking aspen training site configured for training purposes. In this instance, he decides to choose the training site, and navigates to it in the same manner as described in the first scenario.

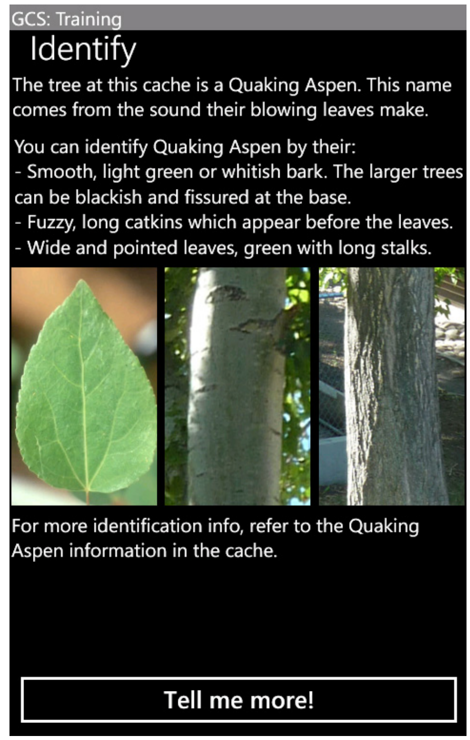


Figure 6. An example training screen.

Real-word Examples

Larry arrives at the training site and finds the cache container. His SCIENCE CACHING mobile device then provides step-by-step training. Figure 6, for example, illustrates a lesson in tree identification. Its material is site-specific. It informs him that the tree next to the cache is a quaking aspen (text at top), includes photographs to highlight features of that particular tree (such as its leaves and bark), and points out those aspects that are unique to quaking aspens and thus usable for identification. The material also includes characteristics that might not be present today, such as leaf appearance (if its winter), or how the bark appears in young trees). Larry looks at the quaking aspen by the cache and verifies its identifying characteristics. Other screens provide similar lessons.

Testing on Known Values

The next part of Larry's SCIENCE CACHING training tests him on-site about the information he has just learned. For example, the application directs him to specific nearby trees (via a photo), and asks him to identify which one is a quaking aspen. Training information is not only repeated on the screen to allow him to refer to it, but also points to particular pages and other materials in an instruction manual kept in the cache. If Larry is correct, the application confirms this, and continues the lesson (e.g., by walking him through measurement methods, and by supplying him with other information helpful for identifying tree types). If he answered incorrectly, the application would have provided him with further hints about identifying the tree type, and would have continued testing. When Larry successfully completes his training, his map view will mark nearby Quaking Aspen sites as open, where he can choose to perform real data collection as in Scenario 1. His training level is also available to Jill the scientist, where she can consider Larry's expertise as a factor when validating suspect data.

Scenario 4: Volunteer Coordination

Geocaching works in part because the system is largely self-coordinating. Somewhat similarly, SCIENCE CACHING is designed for computer-assisted coordination enriched with real-time interaction. It also supports site creation by volunteers, where those sites are communicated back to the scientist in charge for later deployment.

Computer-Assisted Coordination

The scenarios above reveal how SCIENCE CACHING supports self-coordination by connecting the skills of citizen scientists to needed tasks. Under the covers, SCIENCE CACHING learns about each citizen scientist, where it assigns him or her with a 'competency' level for particular tasks. A volunteer's level is adjusted depending upon their training, the number of times they have performed a task, and the quality of their data as determined during data validation. Furthermore, SCIENCE CACHING also tracks the tasks that are needed in a project as defined by scientists, where scientists (using their workstations) can globally create and adjust site properties as maintained in a database. When a citizen scientist browses sites on their mobile device, the most needed task sites that the citizen scientist is trained to do are prioritized on the display: high priority sites appear as solid icons, while lower priority sites are semi-transparent (Figure 1). Similarly, if more trained volunteers are needed for a particular task type, that training site's icon appears solid in the untrained volunteer's display. This helps ensure that the tasks most needed are done first, that citizen scientists are connected to tasks they are skilled in performing, and that training is encouraged for particular tasks on an as-needed basis.

Enriched Real-Time Interaction

SCIENCE CACHING also supports coordination through direct communication between scientist and citizen scientist over a map. Using traditional telephony, email, and SMS services on the mobile device, citizen scientists can ask for help with a task. Using special features in SCIENCE CACHING, scientists can immediately see and discuss the uploaded data produced by the volunteer, and can even mark locations on the volunteer's map to help direct them to particular locations.

Site Creation by Volunteers

There is considerable overhead in creating sites and deploying physical caches. SCIENCE CACHING mitigates this somewhat. Similar to geocaching systems, it allows citizen scientists themselves to create sites. Indeed, many citizen science projects do not begin with a fixed set of sites, but rather encourage citizen science to take readings whenever they encounter a phenomenon of interest. In turn, these can become sites warranting repeated visits. Site creation is done by having the volunteer record the GPS coordinates of a location, augmented by form-filling. Depending on the project, the scientist can moderate these sites and offer them to other citizen scientists for data collection.

CRITIQUE

The design of SCIENCE CACHING was largely based on our understanding of the citizen science literature and our discussions with those involved in citizen science. As a reality check of the SCIENCE CACHING ideas, we discussed the system with nine people highly experienced in citizen science, ranging from researchers in citizen science, expert citizen scientists, citizen science project coordinators, to scientists in charge. Generally, we performed on-site walkthroughs of the SCIENCE CACHING prototype through extended versions of the scenarios as described above. Participants were encouraged to discuss citizen scientist interactions as realized by the prototype, and more generally how citizen science can be supported by such technologies.

The General Consensus

Participants were generally favorable: they believed that the ideas demonstrated in SCIENCE CACHING were valuable for mitigating the four problems of data collection, validation, training and coordination. They particularly liked how SCIENCE CACHING encouraged collection of well-structured data *vs.* the somewhat unstructured data that is often the norm in citizen science. They also saw ways to leverage some of the ideas even further. For example, our participants saw citizen science as a way for scientists to both educate and change the behavior of the public. Consequently, they envisaged extending SCIENCE CACHING training into a multimedia experience as a way to imbed self-education and values beyond mere data collection.

Of course, they saw limitations. They noted that the use of such systems were highly dependent on the particular project. For example, they thought that the setup overhead (e.g., for training sites, for preparing data collection forms) may be too high for small projects with a few volunteers, but valuable for large projects with many volunteers. They were also concerned about the technical demands of such systems, which led into discussions about how such systems could be generalized and content easily added into them.

Participants also identified several problems and potential solutions concerning physical caches. They thought the effort needed for cache creation and deployment would be high, but said this could be alleviated somewhat if citizen scientists themselves created and deployed those caches. They were also concerned with possible environmental impact of the caches themselves, which could be mitigated by having SCIENCE CACHING ask volunteers to remove caches when no longer needed. They also thought that SCIENCE CACHING could allow for a few central caches, where people could pick up and bring equipment to the collection site, and then return it.

Social Experience

Participants stressed that SCIENCE CACHING systems should support various social interactions in citizen science projects. First, some projects require groups of citizen scientists working together, e.g., one person does a complex measurement while another records it. Second, many citizen scientists are motivated to participate because they want to interact socially with other citizen scientists, especially for more strenuous tasks involving (for example) hiking deep into the backcountry. Third, a citizen scientist motivation is their desire to interact with experts. They thought SCIENCE CACHING could support such interactions, such as by including social media that allowing citizen scientists and scientists to communicate and coordinate with one another.

Conclusion

SCIENCE CACHING is not the first system to include geocaching and mobility into citizen science, e.g., [Han et. al., 2011; Christin et. al. 2012]. However, most have concerned themselves with site finding and data collection as identified in scenario 1. We contribute a broader view: SCIENCE CACHING demonstrates how geocaching and mobility can also address data validation, training and coordination. Additionally, our expert participants saw further opportunities for such technologies, such as including social networking to support the real social and motivational needs of citizen scientists. While there is still much left to do, systems such as SCIENCE CACHING can both ease the citizen science process, and potentially broaden the citizen science volunteer audience.

References

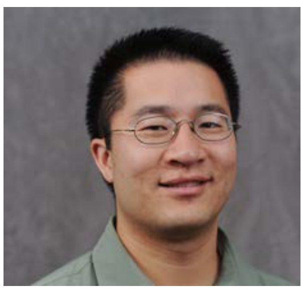
1. Cohn, J. P. (2008), Citizen Science: Can Volunteers Do Real Research?, BioScience 58, American Institute of Biological Sciences, pp. 192-197.

2. Christin, D., Buttner, C. and Repp, N. (2012) CachedSensing: Exploring and Documenting the Environment as a Treasure Hunt. IEEE Conf Local Computer Networks Workshop, pp. 973-981
3. Groundspeak. (2013), Geocaching, [Online], Available: http://www.geocaching.com/, Accessed February 5, 2013.
4. Han, K., Graham, E., Vassallo, D, and Estrin, D. (2011) Enhancing Motivation in a Mobile Participatory Sensing Project through Gaming. IEEE Conf on Social Computing, pp.1443-1448.
5. Neustaedter, C., Tang, A., and Judge, T. (2013). Creating Scalable Location-Based Games: Lessons from Geocaching, Personal and Ubiquitous Computing 17 (2), pp. 335-349.
6. O'Hara, K. (2008), Understanding Geocaching Practices and Motivations, In Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '08), pp. 1177–1186.
7. Silvertown, J. (2009), A New Dawn for Citizen Science, Trends in Ecology & Evolution 24 (September), pp. 467–471
8. Wiggins, A., Newman, G., Stevenson, R.D., and Crowston., K. (2011), A Review of Mechanisms for Data Quality and Validation in Citizen Science, Computing for Citizen Science workshop at the IEEE eScience Conference.

Author Bios

Matthew Alan Dunlap

Matthew Dunlap is an MSc Graduate in Computer Science from the University of Calgary. His interests are in citizen science, mobility and the importance of visceral experiences. Contact him at matthew.a.dunlap@gmail.com



Anthony Tang

Anthony Tang is an Assistant Professor in the Department of Computer Science at the University of Calgary. He leads the RICELab (Research in Interaction, Collaboration and Engagement) research group. His current research examines the role of technology in supporting group activity in collocated and distributed settings, and how ubiquitous technologies will aid the development of health and wellness solutions. For more information, see: http://hcitang.org/.

Saul Greenberg

Saul Greenberg is a professor in the Department of Computer Science at the University of Calgary and holds an NSERC/ AITF/Smart Industrial Chair in Interactive Technologies. He was made an ACM Fellow in 2012, and was also elected to the prestigious ACM CHI Academy in April 2005 for his overall contributions to the field of human-computer interaction. http://saul.cpsc.ucalgary.ca