Pers Ubiquit Comput (article available online as of February 19, 2015)
DOI 10.1007/s00779-015-0837-0

Applying Geocaching Principles to Site-Based Citizen Science and Eliciting Reactions via a Technology Probe

Matthew A. Dunlap · Anthony Hoi Tin Tang · Saul Greenberg

Corresponding Author:
Saul Greenberg · saul.greenberg@ucalgary.ca · +1 (403) 220 6087

Cite as: Dunlap, M., Tang, A.H.T, and Greenberg, S. (2015) Applying Geocaching Principles to Site-Based Citizen Science and Eliciting Reactions via a Technology Probe. Personal and Ubiquitous Computing, Available online February 2015 at DOI 10.1007/s00779-015-0837-0, Springer.

Note: This is the submitted version. It differs in minor ways (mostly its formatting) from the final published version, which is available at DOI above.

Abstract Site-based 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 difficulties exist. We developed SCIENCECACHING, a prototype citizen science aid designed to mitigate four specific problems by applying aspects from another thriving location-based activity: geocaching as enabled by mobile devices. Specifically, to ease problems in *data collection*, SCIENCECACHING treats sites as geocaches: volunteers find sites opportunistically via geocaching methods, and use equipment and other materials pre-stored in cache containers. To ease problems in *data validation*, SCIENCECACHING flags outlier data as it is entered so that on-site volunteers can be immediately check and correct data. Additionally, other volunteers are directed to that site at a later time for further readings that provide data redundancy. To ease *volunteer training*, SCIENCECACHING directs volunteers to training sites on an as-needed basis, where they are taught and tested against known measures. To ease *volunteer coordination*, SCIENCECACHING automatically directs volunteers to particular sites of interest, Real-time communication between volunteers and scientist is enabled as needed. We developed SCIENCECACHING primarily as a technology probe—a working but quite limited system—primarily to embody these ideas and to evaluate their worthiness by eliciting reactions from scientists involved in citizen science. Scientists saw many opportunities in using fixed location caches and geocaching techniques to aid citizen science. Yet they expanded the discussion. Amongst these, they emphasized practical concerns that must be addressed, and they argued that future systems should carefully consider the role of the social experience—both the "online" experience, and the shared physical experience of visiting sites.

Matthew A. Dunlap
Department of Computer Science, University of Calgary, Calgary, AB Canada
+1 (403) 220-6087
matthew.a.dunlap@gmail.com

Anthony Hoi Tin Tang
Department of Computer Science, University of Calgary, Calgary, AB Canada
+1 (403) 210-6912
tonyt@ucalgary.ca

Saul Greenberg
Department of Computer Science, University of Calgary, Calgary, AB Canada
+1 (403) 220-6087
saul.greenberg@ucalgary.ca

Keywords Citizen science, location-dependent applications; geocaching; pervasive computing.

1 Introduction

Citizen science connects non-expert volunteers (citizen scientists) with the scientific process and the natural world [3]. Traditionally, citizen scientists can help in two ways: they can collect data in the physical world (e.g. [1, 16]), or assist in data analysis (e.g. [6]). Both provide scientists with an affordable mechanism to perform geographically broad data collection and subsequent analysis. Citizen science also acts as a mechanism for public outreach and education (e.g. [10, 16]), and is an activity that many enthusiasts enjoy (e.g. [1]).

Considerable recent work has explored the design and use of mobile technologies to support how non-expert volunteers can collect data in the field for citizen science purposes [8, 10, 16, 21, 26], and more broadly for collective data gathering by a community via what is now often called participatory sensing [4]. Generally, citizen science projects focus on the use of mobile devices as electronic note-taking devices, leveraging sensor capabilities such as GPS, camera, as additional recording devices. Perhaps the most well-known project in this space is ProjectBudburst [16], where volunteers, using their own cell phones, take photos of specific plants during different stages of their lifecycle as a means to collect mass amounts of data to identify phenological changes in the environment [8]. Similarly, the ButterflyNet project makes use of, and exploits the time-keeping and multimedia capture capabilities of a smart camera combined with digital pen [26]. Similarly, Common Sense Community leverages the distribution of people and smartphones to collect air quality information by scaffolding the reporting task for novices through a series of mini-applications [23].

Our own work focuses on technological support for site-based citizen science, where volunteers work with scientists to collect data at particular field locations. We are motivated by increasing calls to view citizen scientists using technology as more than "button-pressing non-experts" [14], where we want to take fuller advantage of their abilities as people, and the technical capabilities of their mobile devices. Specifically, we want to design citizen science technology that can take advantage of modern pervasive activity ideas (e.g. [18, 19]), where those ideas have potential to enable more useful data collection for scientists (e.g. [7], to make the process even more enjoyable for citizen scientists (e.g. [9]), and also to scaffold their on-the-ground experiences to train expertise [8, 9, 14]). We are particularly interested in applying geocaching to citizen science. We know that the geocaching community makes use of physical caches at the sites, providing assurance to participants about when they have successfully navigated to a site, as well as allowing them to log this achievement [18]. Similarly, the community is structured in such a way that, by virtue of each participant's logging of his own achievements, the community can ensure that containers hidden in the world are present, and otherwise maintained without active, meaningful oversight [18]. It is likely that citizen science can benefit from this, and similar design ideas.

To focus our inquiry, we first identified common citizen science issues that appear amenable to technology support. In particular, we focused on four fundamental citizen science problems centered on data collection, data validation, volunteer training and volunteer coordination. Second, we developed a technology probe [12]—a partially implemented system—that functioned as a working, malleable sketch of how mobile technologies, married with basic ideas from geocaching and crowdsourcing, might help address these four problems. Third, we used this probe to solicit feedback from nine natural scientists who use, or have used citizen science extensively in their work. Our sessions helped us understand the nuances and practicalities of technology use in real-world citizen science, and also allowed us to critically evaluate particular design ideas. For example, they revealed both limitations in the use of caches, as well as how caches could be used to address several practical problems beyond what we had originally considered. The sessions also helped us understand how technological ideas beyond citizen science can be incorporated in our design. For instance, these discussions elucidated the importance of the social experience in citizen science, and consequently how social networks should be incorporated into our technologies.

We are not the first to apply geocaching notions to citizen science. Even so, we make two contributions that extend work in this area. First, we articulate and demonstrate (via a partially implemented system) how we can apply particular concepts from geocaching to the design of tools that mitigate four known citizen science problems. Second, we used the system as a technology probe to discuss and seek feedback on these design ideas from natural scientists, where their views refine and extend our perspective on this problem space and potential solutions to it.

2 Four Site-Based Citizen Science Problems

We approach site-based citizen science as designers interested in how technology can help support or change basic citizen science practices or tenets. To provide focus, we considered four fundamental problems concerning citizen science as articulated by previous researchers [3, 5, 11, 20, 25].

1. *Data collection*. How can collection (performed primarily by untrained volunteers) be made to

ensure that the data collection is easy, that it is correct (it is often error-prone), that it is structured (thus affording later data interpretation), and that it allows for flexibility as necessary (e.g. for unusual observations)?

2. *Data validation*. How can scientists ensure that the collected data can be trusted, since physically validating the data (i.e. revisiting a site) can be impossible or time-consuming for the scientist?

3. *Volunteer training*. How can volunteers be trained both initially and on an ongoing basis? Can scientists track the training level and competency of particular volunteers in order to know what tasks they can do correctly and judge the likelihood that their data is correct?

4. *Volunteer coordination*. Larger projects need to coordinate actions between unpredictable involvements of large numbers of volunteers. Scientists have limited time to do this coordination, and costs limit their ability to hire a coordinator.

3 Applying Geocaching Concepts to Citizen Science Problems

Except for a few notable instances, most site-based citizen science practice has not changed meaningfully since its inception. When technology is leveraged, it mostly used as an electronic replacement for the paper notepad. Yet various technologies already promote site-based activities, albeit in different domains. This begs the question: can other existing methods assist (albeit in a modified form) the citizen science process? In particular, we speculate (along with others) that citizen science could benefit from ideas coming out of pervasive, location-based gaming—not only the specific the "game aspects" of these games (e.g. [9]), but also the structure of the interaction between people, devices, and infrastructure. Thus we turned to geocaching.

Geocaching is a location-based treasure hunting game, where participants hide physical containers (caches) in known physical world sites for others to find via GPS location. The social nature and rules of use keeps the geocaching crowd creating and hiding new caches without meaningful oversight [18, 19].

Both geocaching and site-based citizen science are location-based activities, where its actors find and pursue tasks at particular sites. Because of the popularity and success of geocaching, we speculated that we could apply geocaching concepts to help mitigate the above four problems found in site-based citizen science. In particular, we drew inspiration from several basic design ideas from geocaching, as listed in the left column of Table I. Specifically, geocaching has well-defined mechanisms to support *site discovery* (online searchable catalogue with achievement logging); *navigation to sites* (in addition to GPS coordinates, detailed site information, including photographs, descriptions, etc.); *physical containers* (signifying the actual site and containing materials of interest); *user-generated content* (participants can create new sites), *site maintenance* (players "review" sites by logging their achievements and the status of the site), and *player coordination* (where players largely coordinate themselves).

Our central research questions were thus: How can and should these ideas be translated into technological design supporting site-based citizen science activities? How would these design ideas be received by scientists who incorporate citizen science into their work practices? For our first question, Table I (right) summarizes our speculation on how particular geocaching ideas can be applied to citizen science. The following section further describes these ideas in terms of how they are translated into design through our SCIENCECACHING prototype system. For our second question, we used SCIENCECACHING as a technological probe to gather feedback from scientists, and to discuss how the core ideas could impact their citizen science projects in the future (as presented in later sections of this paper).

4 SCIENCECACHING: A Probe

We developed SCIENCECACHING, a working prototype to illustrate how the ideas in Table I can be instantiated in software supporting a site-based citizen science project. Our two-fold goal in developing this prototype was: a) to have a concrete instantiation of ideas as they might apply to a site-based citizen science project, and b) to use this prototype as a technology probe to gather feedback and insights from practicing scientists who work with citizen scientists. While our prototype is fully functional, it is not a production-quality system ready for widespread deployment. Rather, we developed it as a working sketch that highlights particular concepts through deliberately simple task scenarios (to focus scientist discussions on the design concepts of interest) rather than developing them into a production-quality for deployment. We first describe how the primitive geocaching features (as summarized in Table I) combined with mobile devices are translated and extended into

Common Need	Geocaching Approach	Use in Citizen Science
Site discovery	Geocaching employs a searchable database of caches based on various characteristics (distance, difficulty, etc.)	Through a similar mechanism, volunteers can opportunistically discover and choose between nearby citizen science sites that match their training level and that are ready for data collection.
Navigation to sites	Detailed site information is provided, including photographs, descriptions, GPS coordinates, etc.	*Pinpoint navigation.* Clear directions makes sites easier to find, while identifying information helps people find the exact locations on the site requiring inspection and data entry.
Physical containers	Physical caches (i.e., small waterproof containers) signify the end of the treasure hunt	*Markers.* The physical cache serve as a marker identifying exact site location. Other fine-grained locations can then be indicated relative to the cache. *Tools and training materials* can be placed inside containers for volunteers to use. *Storage*: Volunteers can store collected samples in the container for later retrieval.
Multiplayer / Repeat visits	Geocaching is a multi-player game, where multiple people can visit and revisit multiple sites. Continued conditions of caches are logged by participants	*Rich data collection.* Repeated visits can be used as a mechanism to gather both longitudinal and comparative data, *Site maintenance*: to maintain and update materials in the cache, and to remove samples and other materials stored there by prior citizen science visitors. *Data validation.* By making sites easier to find, multiple volunteers can be directed to a site to provide data collection redundancy. Data validation is thus easier, as data between volunteers can be checked for inconsistencies. Unusual differences can be flagged on-site, giving volunteers the opportunity to check their on data.
User-generated content	Through very simple mechanisms, participants can create new sites	*New sites.* Volunteers who uncovered phenomenon of interest elsewhere can create new data collection sites, which are added to the sites available for data collection (perhaps after verification by the scientist)
Site types	Geocaches can be of various types (e.g., traditional caches, multi-caches with multiple locations, event caches, etc.)	Various types of sites are also possible. *Training sites* are set up to train volunteers. *Multi-cache sites* are those where volunteers are directed to multiple locations around the primary cache location.
Player coordination and communication	Geocaching web sites require little explicit coordination of players. They provide terrain and finding difficulty levels so players can choose activities that match their skill level. Players can leave comments concerning particular sites (e.g., the cache state, helpful information) that assist other players interested in that site.	*Guided site selection.* Scientists can specify attributes of particular sites as well as the type of volunteer (e.g., training level, particular groups) that should collect data of that site. Volunteers searching for sites will see those that match their capabilities, or will see how they can be trained up to that capability by being first directed to training sites. They can then choose to visit that site, thus minimizing explicit volunteer/scientist coordination. *Messaging* between volunteers and scientists are supported on an as-needed basis. Real-time messaging facilitates directions scientists may give to on-site volunteers, or questions volunteers may have for scientists.

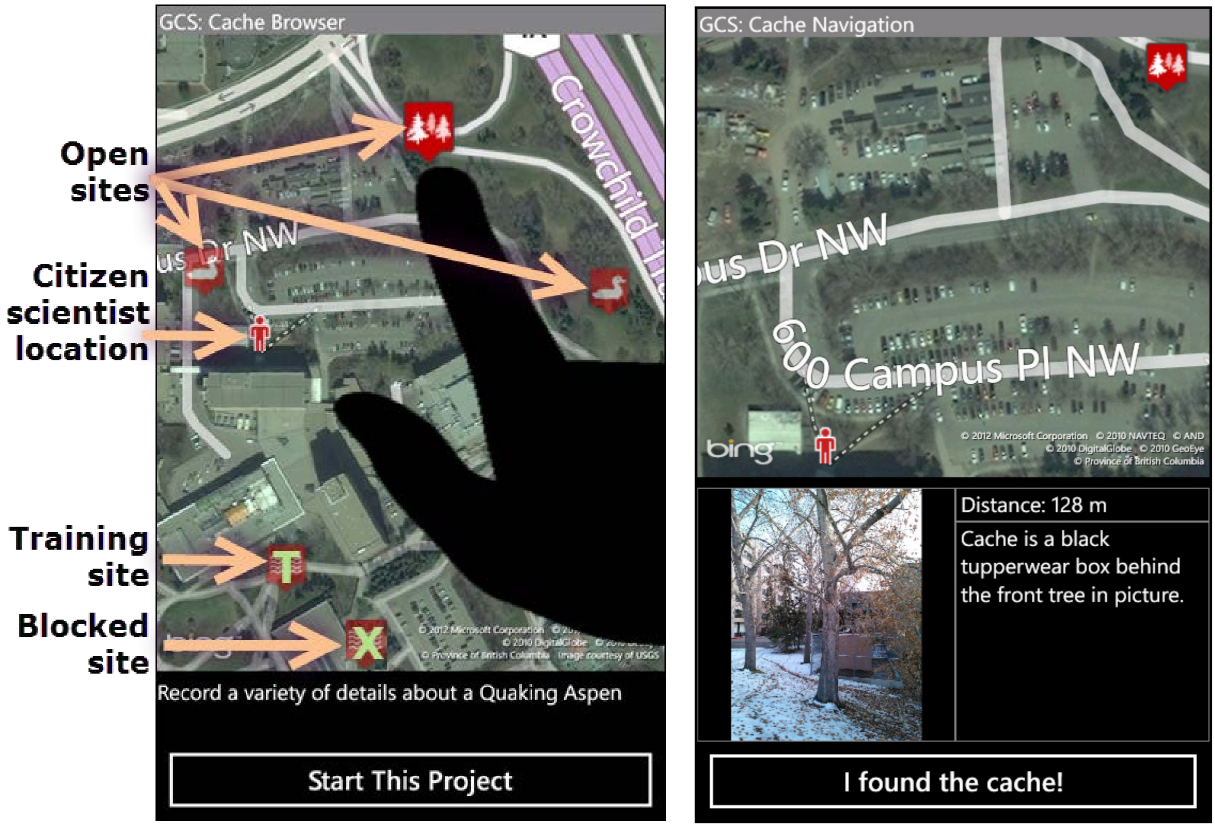
Table I Needs as addressed in Geocaching, and our initial speculations on how they can support site-based Citizen Science

SCIENCECACHING design concepts. We then briefly describe scenarios of use that applies these design concepts to mitigate the four previously discussed citizen science problems.

5 From Geocaching to Design Concepts for Citizen Science

5.1 Site discovery

Geocachers decide what geocaching sites are of interest to them. Criteria may include a player's proximity to the site, and the site's difficulty rating. We apply the same thinking to site-based citizen science, where maps are tied to a back-end system that catalogues the collection sites (e.g. [9]). We extend discovery by matching sites not only to those nearby the citizen scientist's location, but also to the current need for data collection at the site, and that person's training level. Metrics generated from the back-end (e.g., last visit; flagged record; volunteer expertise and/or track record, etc.) can coordinate and direct citizen scientists to specific sites. As seen in Figure Ia, volunteers can pan and zoom a GPS-



a) viewing collection sites **b)** finding a chosen site

Fig. I ScienceCaching's mechanism to view and navigate to collection sites, and to display information indicating details of a found cache

enabled map on their mobile device, where the map shows various sites. Further information about each site is displayed when that site is touched (Figure Ia, bottom). Volunteers then choose which site they wish to visit. Sites are marked as:

- *open sites* of particular types that can be visited, e.g., those where data is currently required and that matches the volunteer's skill level (Figure Ia, icons illustrating a tree collection site and a duck site),
- *prioritized sites* where high priority sites appear as solid icons and lower priority sites semi-transparent (Figure Ia).
- *blocked sites* that should not be visited, e.g., because the scientist does not need data from that site at the moment, or because the volunteer's skill level as tracked by the system is insufficient to collect data from that site (Figure Ia, icons with an 'X'),
- *training sites* set up to teach volunteers further skills.

5.2 Navigation to site.

Most geocaching systems provide its players with information to find the site. Similarly, the volunteer is provided with directions to a chosen site using standard map / GPS navigation methods. We extend navigation to offer pinpoint site location (Figures 1a and 1b showing a person's location and the target destination on a map). When the volunteer arrives at that site, the mobile device displays a photo and text description (Figure Ib, bottom) identifying a pinpoint location at that site (e.g., usually the location of the physical cache, discussed below).

5.3 Physical containers for cache discovery and advanced collection tasks

The goal of a geocacher is to find a particular physical container, which typically contains a log book and some trinkets that can be exchanged. We extend the idea of a physical cache to support purposeful citizen science work, as illustrated in Figure II. First, the physical cache acts as a marker that further identifies a precise site location, which in turn can help orient the volunteer's data collection

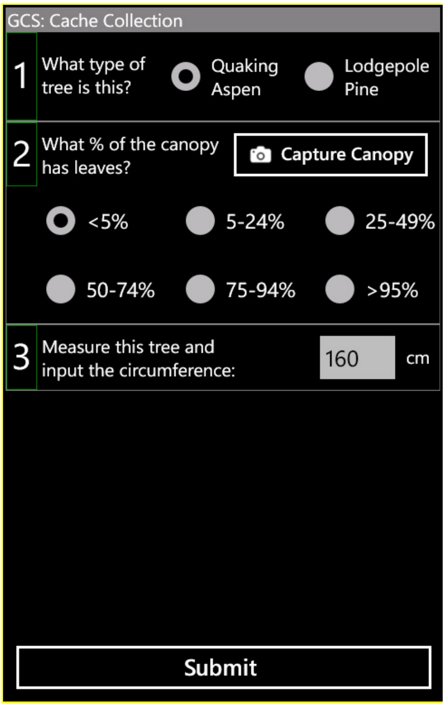
a) cache as marker identifying a precise location **b)** cache contains useful materials

Fig. II Physical caches in site-based citizen science

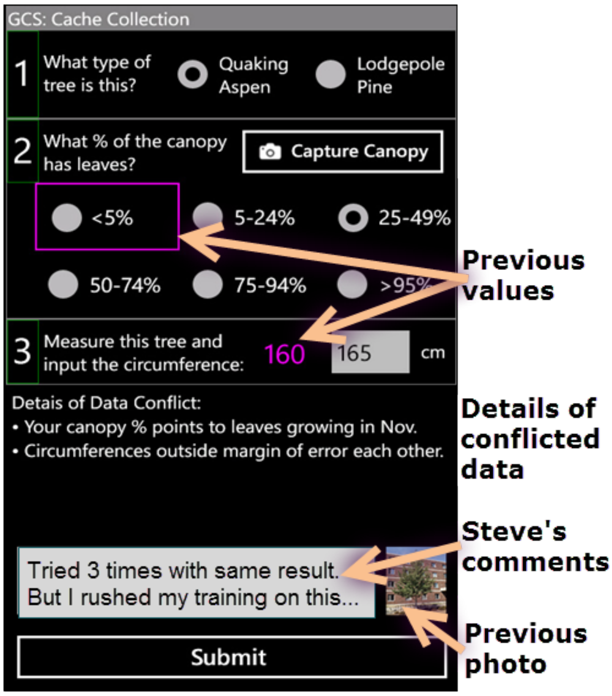
activities within the site (Figure IIa). Second, the physical containers can house useful materials (Figure IIb), such as help materials (e.g., a dichotomous key assisting identification) and physical tools supporting data collection (e.g., tape measures for measuring lengths); both allow for more sophisticated data collection tasks to be performed. Third, the cache can also be used by the volunteer to store materials (e.g., collected samples which will be retrieved at a later time) that can supplement readings entered onto the mobile device.

5.4 Repeat visits, site maintenance and data validation

As a multiplayer game, geocaching affords repeat visits, where multiple players can visit a site. Players can create logs describing the site / cache state, and what maintenance (if any) is required or has been done. This also applies to citizen science sites: volunteers can report site conditions to the primary scientist, and repeat visitors can maintain, update, and even retrieve materials in the cache as needed. Repeat visits can be leveraged to provide richer data collection, such as longitudinal and comparative data collection over time as gathered repeatedly on that site by a single volunteer, or by multiple volunteers visiting the site. Repeat visits also support data validation in two ways. First, repeat visits by different volunteers can be done for data redundancy: errors can be flagged if values differ greatly. Second, data validation can be done *in situ*—when a citizen scientist logs a record, he can perform a sanity check by comparing his record against other citizen scientist's records entered on previous visits. In cases of vast differences, the system also alerts the citizen scientist, motivating him to closely examine both his own and the previous entries for disparities (and to flag incorrect, or to correct an entry where appropriate). This data verification is supported by the fact that the volunteer does all this on site, and has access via the mobile device to the previous log and the various artefacts associated with that collection (photos, tools, samples, etc.). For example, Figure IIIa illustrates a data entry screen by a first volunteer. Figure IIIb illustrates a second volunteer who has visited the site at a later time. The system indicates what values were entered by the previous volunteer (in purple) as well as photos taken by that person (bottom right). This allows the 2nd volunteer to compare his reading against those. The system also does bounds-checking, where it provides details of data that likely conflict with expected norms (Figure IIIb, bottom half).



a) Volunteer #1's data collection screen



b) Volunteer #2's data collection screen comparing his entries against the previous values.

Fig. III. A typical data collection screen (a), and how data validation is encouraged (b) by flagging unusual data points and by showing previously recorded data.

5.5 User-generated content

Geocaching allows its players to create new geocaching sites. Similarly, citizen scientist can also generate collection sites. We designed a simple mechanism for citizen scientists to establish (and publish) entirely new collection sites. Like geocaching, which allows site creators to specify attributes of a given site (e.g. difficulty, terrain, hazards, etc.), these sites can be tagged with attributes that suggest required expertise or skills.

5.6 Site types

While the majority of geocaching sites just provide a cache at a given location, other types of sites allow for a broad range of activities that deviate from this model. We build on this by envisioning various citizen science site types. Some sites may be particular to the data being collected and the skill set required (e.g., a duck monitoring site and a tree collections site as shown by the site icons in Figure Ia). We also see high value in creating special-purpose real-world training sites with known values, where learning occurs *in situ* on an actual site. For example, the screens in Figure IV all concern descriptions, activities, materials and entities (trees) that are actually located at that site. After the volunteer selects the 'Let's test my knowledge' button, subsequent screens will check the values entered against known values, and tutor the volunteer accordingly if errors are detected. Some training sites would be introductory, where they would provide scaffolding for volunteers to learn about basic data collection activities. Other sites would build upon existing skills, where they could teach more complex data collection (e.g., with tools, for different tasks). This progressive approach would encourage the development of expertise, and engaging volunteers with more deeply with the scientific process [9, 14]. Together, these caches train participants to perform complex collection tasks on real-world examples—wherein their performance could easily be checked against known values for collected data. The real-world nature of these training sites provides a substantial advance over most current approaches (e.g. [3]), which typically rely on text-based instructions (perhaps augmented with video). Since these operated from fixed, known sites, the mobile device could provide verbose "hints" and substantial site detail for volunteers as they progress through the training course (e.g. "The collected value should be within the range of X and Y").

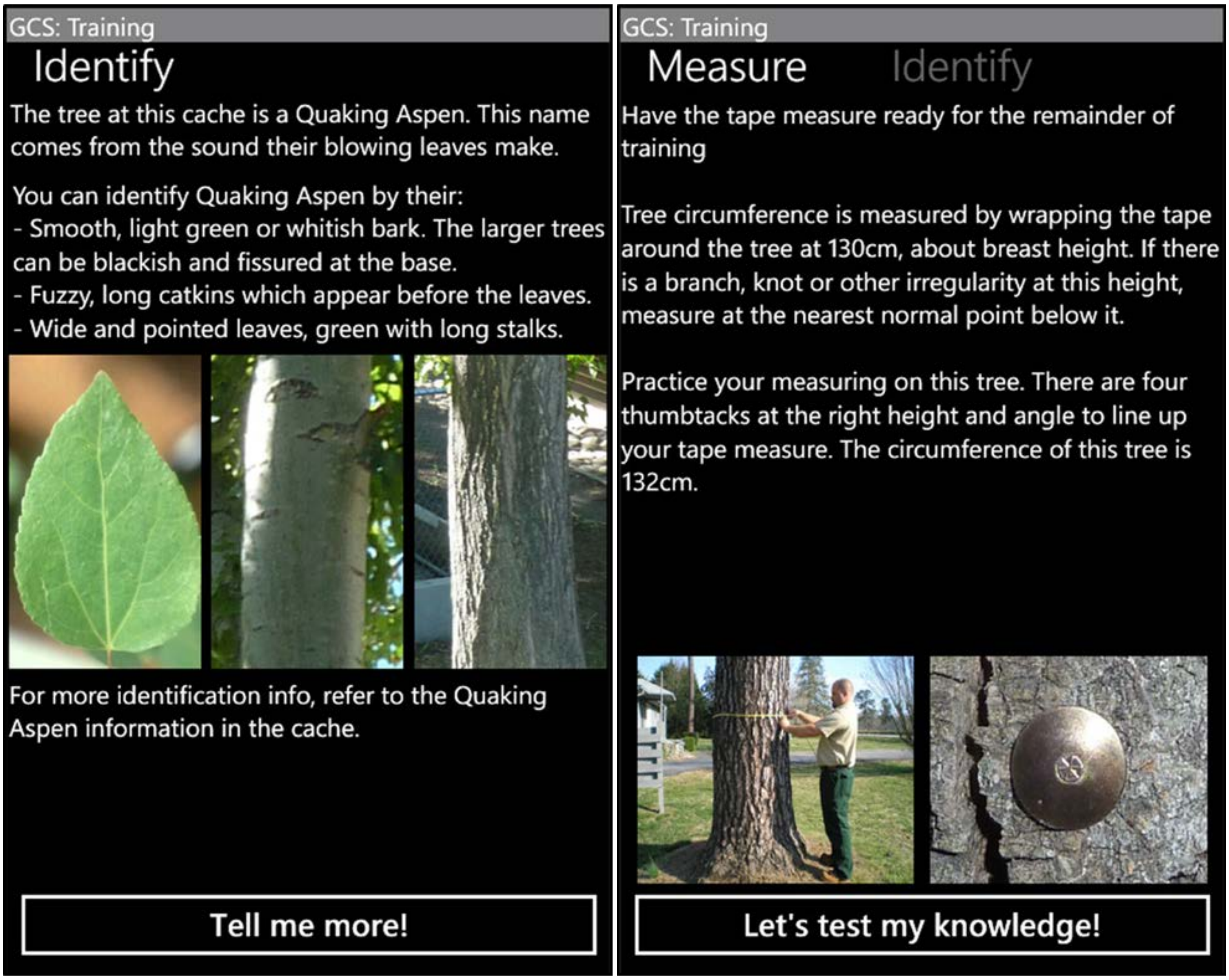


Fig. 4. Special training sites give volunteers expertise. The mobile interface includes site-specific training. Images, activities, and data-checking are based on on-site entities.

5.7 Player coordination.

Geocaching is largely as self-coordinated activity. Citizen science, on the other hand, usually requires a large degree of coordination by either the scientist in charge or a volunteer coordinator. We extend geocaching self-coordination to citizen science. As already discussed, the system can guide the volunteer to particular sites (and block sites that they shouldn't visit), where the volunteer can then self-select which sites to visit. Additionally, we enabled live communication between the mobile interface and the scientist - voice, messaging, email, etc. - that allows the citizen scientist to ask questions if needed, and for the scientist to provide immediate, live feedback.

6 Scenarios

We developed a set of scenarios based on a fictional citizen science project focused on gathering tree data, where each scenario is used to explore the four previously mentioned citizen science problems. Physical collection sites were prepared (both data collection sites and training sites). Each site was equipped with a stocked cache. Each scenario also highlights particular geocaching design concepts as described above and as realized by our SCIENCECACHING design probe as illustrated in Figures 1-4. These scenarios were used to present our design concepts to the citizen science community during our interview study.

6.1 Scenario 1 - Data Collection

The first scenario affords the most basic citizen science task: going to a particular site and collecting data from it. 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 players use to opportunistically choose a site to go to, the way players 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 SCIENCECACHING, which assists in the choosing, finding and collection process.

Figure Ia illustrates the cache selection screen, which is a navigable map with various icons. Nearby collection sites are shown, including two duck monitoring sites, a tree monitoring site, and two water-monitoring site (which the volunteer is not yet trained for: one is blocked but the other is an open training site). Here, we see the personalized coordination taking place—since the participant is "trained" in tree monitoring, this is the main one revealed, along with additional information.

GPS assists navigation to the site, showing one's relative location to the site. Additionally, it describes the nature and location of the physical cache (i.e., a black Tupperware container behind a particular tree as illustrated in Figure IIa), and provides a photograph of that location (bottom of Figure Ib). After finding the cache container, one can verify and signal to the system by pressing the "I found the cache" button (Figure Ib).

The container itself contains several tools to facilitate data collection, including a tape measure, a phenological guide, a leaf sampling logbook, a pen, scissors and other materials (Figure IIb). The device presents the data collection form for this specific site, which is different depending on the type of site (e.g. duck monitoring vs. tree monitoring). As illustrated in Figure IIc, this form asks him to perform a number of both simple tasks (identify the type of tree, taking a photo of canopy cover) and complex tasks involving tools (e.g. estimating canopy cover and measuring tree circumference). Other screens ask him to collect a leaf sample in the container.

In this scenario, we also asked participants about site creation. There is considerable overhead in creating sites and deploying physical caches. SCIENCECACHING 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 (not shown). Depending on the project, the scientist can moderate these sites and offer them to other citizen scientists for data collection.

6.2 Scenario 2 - Data Validation

This next scenario illustrates several ways that data validation is afforded by geocaching as realized by the ScienceCaching system in terms of repeat collections, and on-site data validation.

A coordinating scientist can mark a cache as a "repeat collection" site, meaning that citizen scientists will continue to see the cache on their listing (as in Figure Ia) even if it has been visited before. This encourages multiple collections for the same location, allowing the coordinator to validate data. Multiple readings can be averaged (for example), or be inspected for obvious outliers, or to be explored for trends, or simply for spot-checking. Questionable data can be explored in detail: redundant information (e.g. multiple photos collected along with canopy cover estimates) can be gathered together allowing the scientist to confirm the readings.

Furthermore, data can be validated on-site through sanity checks and these repeat visits. For example, Figure IIIb illustrates an erroneous reading by a citizen scientist (who allowed his measuring tape to sag during the trunk circumference measurement). The system automatically compares his new log (Figure IIIb) with old logs (as entered by an earlier volunteer in Figure IIIa). If problems are detected, the system guides him to check his measures and to correct any mistakes while still on site (Figure IIIb). These conflicts (and corrections) are recorded by the system, alerting the coordinator to potential problems in the data.

6.3 Scenario 3 -Volunteer Training

The third scenario explores how ScienceCaching helps train new citizen scientists on prepared training sites rather than solely through digital resources. 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.

As in Figure Ia, the mobile application illustrates the various sites, including a quaking aspen training site. In this instance, an untrained participant can select this site, and navigate to it as before. Arriving at the training site, he is led through a series of containers (much like a geocaching multi-cache), where he completes a step-by-step training procedure. Figure IV illustrates a partial lesson in tree identification. Its material is site-specific. It informs him that the tree next to the cache is a quaking aspen, 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. The trainee can now look at the physical quaking aspen near the cache to verify its identifying characteristics. Other screens provide similar lessons.

The next part of the training scenario tests participants on-site about the information just learned (not shown, but somewhat similar to Figure III). 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 the participant answers correctly, 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 provides him with further hints about identifying the tree type, and would have continued testing.

When the training is successfully completed, a participant's individual map view will mark nearby Quaking Aspen sites as open (Figure Ia), where he can choose to visit those sites and perform real data collection as in Scenario 1. His training level is also available to the coordinating scientist, who can consider this participant's expertise as a factor when validating suspect data.

6.4 Scenario 4 - Volunteer Coordination

Geocaching works in part because the system is largely self-coordinating. Somewhat similarly, SCIENCECACHING 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.

The scenarios above reveal how ScienceCaching supports self-coordination by connecting the skills of citizen scientists to needed tasks. Under the covers, ScienceCaching tracks each citizen scientist. 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, ScienceCaching 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 tasks 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 Ia). 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.

ScienceCaching 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 ScienceCaching, 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.

7 Technology Probe Interview Study

The design of SCIENCECACHING was based on our understanding of the citizen science literature and our discussions with those involved in citizen science. Our next step was to capture knowledgeable reaction and critique to the concepts as presented by the design and scenarios of use.

7.1 Focus Questions

We sought broad feedback from our participants, using four very general questions. We encouraged very open-ended and general discussion: while we were interested in feedback about our approach, we

Identifier	Expertise	Project Experience
P1	Senior Scientist at National Park	Coordinator of citizen scientists in multiple projects
P2	Citizen Scientist with Geoscience PhD	Volunteer coordinator and participant in multiple projects
P3	Senior Project Manager at non-profit heavily utilizing citizen science	Helped initialize and manage many citizen science projects
P4	Coordinator of volunteer engagement at National Park	Connected scientists and citizen scientists for many projects
P5	Research Associate at non-profit heavily utilizing citizen science	Helped initialize and manage many citizen science projects
P6	Scientist at National Park	Worked with citizen scientists in a few projects
P7	Scientist at National Park	Worked with citizen scientists in a few projects
P8	Associate Professor, researcher of human/carnivore interactions	Citizen scientists has provided her with information germane to her research
P9	MSc studying youth engagement in citizen science	Volunteer for citizen science projects and researcher in citizen science

Table II The participants in the design critique including their expertise and particular project experience.

were very much interested in ideas beyond this approach.

- *Did we target appropriate citizen science problems?* The subset of citizen science problems targeted in our designs were drawn mainly from literature. We focused on four particular problem areas that seemed likely to benefit from the application of geocaching and mobile technology to citizen science. Yet solving these particular problems would have little value if they did not represent real problems encountered by real practitioners, or real problems that were in fact in need of solutions.
- *Is our design approach to these problems reasonable?* The solutions developed for the four problem areas in citizen science have not been deployed in a real-world citizen science project. Assuming that the problems they attempt to solve are valid ones, we do not know if the particular designs as presented via the scenarios are reasonable solutions to those problems.
- *How should this design approach be extended?* The design space that addresses citizen science problems through mobile devices and citizen science is very large. As mentioned, our designs are the first working sketch. As in any user interface project, we foresee a large number of iterations over this design space before getting the design right. Related to the previous question, we do not know what parts of that design space are relevant and appropriate to real citizen science problems, how existing ideas can be combined or extended, or if we have missed design opportunities. Knowing the answer to this would be valuable to see where we should focus our efforts on future iterations.
- *How can these ideas be applied?* The SCIENCECACHING system was designed around several abstract ideas implemented in a deliberately simple citizen science application. While we believe these design ideas are generalizable, they need to be revisited in terms of a concrete application of value to real scientists and citizen scientists (i.e., projects that they have worked on in the past, or working on now, or could envisage working on in the future).

7.2 Method.

We discussed SCIENCECACHING with nine people highly experienced in citizen science. Participants ranged from researchers in citizen science, expert citizen scientists, citizen science project coordinators, to scientists in charge, as summarized in Table II. Generally, we performed on-site walkthroughs (i.e. outdoor) of the SCIENCECACHING prototype through extended versions of the scenarios as described above. Participants were not confined to these scenarios. They were encouraged to discuss citizen scientist interactions as realized by the prototype, to generate ideas and /or concerns as they thought about them, to propose other scenarios, and to more generally offer their thoughts on how citizen science can be supported by such technologies. In keeping with using SCIENCECACHING as a technology probe, we stressed that the system and the scenario details were there only to seed the discussion rather than limit it.

Analysis. Analysis focused on the audio-recorded interview data and field notes, specifically on parts of the discussion that were relevant to the prototype's design and purpose. In parallel to ongoing interviews, we constructed affinity diagrams of key phrases from the transcription and field notes. As we continued to interview additional participants, we extended and/or revised groupings as needed to either accommodate new themes, to refine existing themes, or to create new groupings that better reflected what participants said. Furthermore, new data from participants was also used to evolve the questions asked of subsequent participants. At the end of the analysis process above, the categories were presented to a subset of participants (in a second set of interviews) in order to validate them. We asked whether the categories seemed appropriate and whether our interpretation was consistent with their views. The final groups or themes emerged from an on-going process of specification, organization, refinement and (to a limited extent) validation.

8 FINDINGS

Our analysis uncovered overarching themes about the nuances of applying geocaching to mobile citizen science. Here we discuss these themes, first looking at what participants saw in our approaches to our targeted problem areas: *data collection*, *data validation*, *volunteer training* and *volunteer coordination*. We then look at two additional themes that go beyond how we initially applied geocaching to citizen science: *citizen science as a social experience* and *practical deployment of physical caches*.

8.1 Data Collection

Participants were excited about how the SCIENCECACHING prototype embodied data collection, and provided several suggestions about how these ideas would be refined if they were to be deployed their own projects (P1-9). In particular, participants were interested in how the method enforces well-structured data collection while still allowing for flexibility (P1-5,8). Well-structured data (such as well-formatted data types afforded by strict item selection or numeric entry) was considered very important, as it meant that large amount of data could be collected and analyzed via a standard analytic method (*vs.,* for example, free-form text that usually had to be hand-collated and hand-analyzed). As participants were not utilizing mobile devices in their projects (cf. [16, 26]), they were excited by how the prototype SCIENCECACHING exploited basic functionality afforded by mobile technology, such as:

- easy site discovery and location (all);
- automatic recording of GPS location (P3);
- entry of structured data (e.g., numbers), and unstructured data including verbose descriptions and photographs, via mobile data collection forms (P3-5), and
- the automatic upload of data to scientists (P3-5).

Physical Caches. SCIENCECACHING introduced the use of physical caches to represent data collection points, and participants were generally favorable to this idea. (P3,4,9) commented that the geocaching approach of marking collection sites by physical containers could be potentially fun for citizen scientists. Similarly, several participants felt that the use of a physical cache would both aid repeat visits, and support more accurate data collection (P5-8). Indeed, (P6,7) already used similar aspects of this site-finding in their current project: they stored site locations on a GPS device, and marked sites physically with rock cairns (stacks of rocks). Nevertheless, this was not a unanimous opinion. (P1) felt that generally physical caches presented far too much work for scientists for far too little payoff. We discuss issues with the practical deployment of physical caches later.

Advanced Data Collection: Tools, Data and Samples. Storing tools, data or samples in caches was seen as a possible way to save scientist and citizen scientist time. While participants saw value in particular variations of this general approach, they did not see instrumenting *every* cache as an effective use of time (P1-9). For instance, rather than placing tools in every possible data collection point, participants suggested the use of "Area Caches", where a cache could be used to store tools used for a number of collections around the area (P2-5).

(P3) proposed another cache variant, where special caches could be designated as "sample-holding caches". These could be deployed, for instance, near a trailhead or roadway. Citizen scientists would take sample-holding equipment from the cache, go into the field, collect a sample, place it in the container and record (using a mobile device) that a sample was gathered. The scientist could then gather the sample at his or her convenience, possibly after multiple samples were collected. Alternately, the transfer of samples stored in the cache could be a task performed by the citizen scientist. Several participants thought this was a useful idea (P4-9), for example:

> *"…when we did the bear DNA stuff, the volunteers all had to bring their samples of hair in their little envelopes back to the volunteer office and then the researcher came to the office, but this (idea) cuts out that middleman." (P4)*

Citizen-Science Created Data Collection Sites. (P8) saw citizen scientist site creation as a way to extend geocaching to exploit observations of transient phenomena. For example, a citizen scientist could create a site when they encountered a rare animal (e.g. a bird that is not normally found in the area). In this way, the site could then be re-visited by other citizen scientists to see if an animal is spotted there again. Importantly, the citizen scientist would make a record whether or not an animal was seen, making what would otherwise be only opportunistic, unstructured data collection about a rare animal sighting into a more structured, systematic collection point.

8.2 Data Validation

Scientists and coordinators generally supported our use of different techniques to improve validity: multiple readings (P1,3,4,6-8), photographs (P3,9), and flagging of suspicious records on site (P3,4,9). However, our participants generally only used expert review to validate data [25]. Thus they did not have enough experiences with other methods to give substantial feedback except for general approval.

The prototype took the approach of presenting previous records for validation when the current citizen scientist's collection varied widely from the previous collection. This could allow the current citizen scientist to change his/her record, or to comment on the differences or potential problems with the previous record. Some participants thought this approach could potentially introduce bias that could influence the recorded results (P1,3,8). (P1), for example, saw bias as problematic in situations where a high degree of data quality is needed. Yet participants also viewed this bias as an acceptable risk in many citizen science projects.

8.3 Volunteer Training

Participants were generally positive about how SCIENCECACHING realized volunteer training. The training scenario explored how to lead citizen scientists through a real-world training course, providing potential citizen scientists with real-world experience, a mechanism to teach citizen scientists how to perform more involved data collection (e.g. with tools), and a mechanism for scientists to ensure that citizen scientists would be trained before going out into the field. This would have the added benefit of providing projects with more instances of correctly-collected data. Participants saw real-world guided education as part of the success of many projects (P2-4,6,7). The training scenario presented to them specifically pointed out aspects of the phenomena being trained on (e.g. the amount leaf cover on the training tree), which was seen as important for guiding digital training (P1-4,7-9). For instance, (P2) stated:

> *"(Your training is) very specific to what they are collecting. You've got the picture (of the tree), the whole bit like this."*

Participants thought the training scenarios could be applied to a broader range of contexts. Examples included training in data collection projects, education-based citizen science projects and teaching school groups. Training caches could be deployed in small numbers close to the scientist's (or teacher's) base of operations, allowing those taking part to learn while minimizing the training effort normally required of the scientists. Tools could be stored in the caches so that trainees could practice their skills. Indeed, P1's citizen science project was very similar to the aspen tree scenario used in presentations, and he saw training as being directly applicable to his project. (P3,5) saw the use of training sites to educate grade school citizen scientists about human impact on the environment. A few caches could be deployed so that students could engage with technology and guide themselves through environmental education. Similarly, (P8) saw training caches useful for teaching her university students about field sampling and data measurement.

8.4 Volunteer Coordination

In general, our specific approaches to coordination were seen as useful. However, participants indicated that our approaches overlooked the role of the social structure of citizen science, and how that facilitates coordination.

Computer-assisted Coordination. Beyond our original vision, some participants opined that automatically tracking the skill and participation of citizen scientists could be a good approach for identifying highly engaged participants (P3-5,8). (P3) described how, along with (P5), they had manually identified such participants and how they use that information:

> "*Once we started (our project) and figured out who our keeners were, we created a new component for them and they found it really rewarding… How do we figure out who these people are? Well, we start with the more opportunistic approach, with the goal of what you are trying to introduce is more of an educational thing to get people talking and aware of the wildlife in their region. Then when you find those real keeners (key participants) you introduce a little more systematic approach.*"

(P8) echoed the value of tracking, where she even attributed the failure of one of her projects partially on the inability to identify highly engaged participants. (P4) further proposed an idea for identifying key participants: citizen scientists would start on easier, digital collections, where only engaged participants would be invited only later to take part in a project that requires a scientist's direct involvement. This has a multiple benefits. Scientists would be able to spend their time with participants who they knew were motivated. Citizen scientists will have already gone through the basics of citizen science work, allowing scientists work with them on more difficult problems. Furthermore, working with a scientist is a major motivating factor for citizen scientists, and would give them a goal to work towards (P1,3-5,9) (to be discussed further in the next theme).

Opinions differed on how much automation should be part of a citizen science project. (P3-5,8) saw coordinators and scientists necessary in all projects, where they were required to train, engage and manage citizen scientists. (P1) on the other hand, was passionate about seeing citizen science projects becoming projects where each would have "a life to itself." He detailed a motivating problem with monitoring systems:

> "*If you had to pick the most classic flaw in monitoring systems is that they are all tied to the individual's expertise at the time, and they usually get very complex (because they are) tied to that person's interest. Ironically after 40 years, you have nothing because it was only collected for two years; it was too complicated for somebody to understand in year three and that other person left.*"

(P1) saw the benefit of citizen science projects that could be *"picked up off the shelf"*, e.g., were well-structured and accessible through the web or mobile applications. These would allow data to be collected with little or no scientist coordination. This would provide flexibility for citizen science to be used as a social activity by any group that was interested.

Enriched Real-Time Interaction. Participants thought it important to support a scientist's communication directly with citizen scientists, especially for coordination in more complicated projects (P2,3,5). As (P2) stated:

> "*It's not just 'when you are out walking, do this'. It's 'We have X number of sites to visit. Team A is going on (these days), Team B is going on (these days), and they are getting their training on this day.'*"

The scientist controller addressed these needs somewhat with the ability to message citizen scientists, but communication also needs support when citizen scientists are not in the field. Perhaps adding other interaction possibilities, such as interacting with the collected data through video, audio, or a map would have made communication in the SCIENCECACHING system even more valuable. Being able to send more verbose information through the chat system, e.g. map locations and walking paths, may better support the changing needs of some projects.

(P4) proposed a different use of real-time interaction during training. Specifically, it could be used by coordinators training citizen scientist in-person while in the field. Additionally, mobile quizzing could be used to facilitate and track the progress of citizen scientists through training. The success this training can be tracked, as (P4) stated:

> "*Sometimes it's hard to keep track of the kinds of training and how well people did, even when you do in-person training… (This would mean) I*

don't have to sit at a database and enter in all the info."

For instance, a scientist or coordinator could use their mobile devices during in-person training, answering questions and interacting with training sites, ensuring training has been verified and recorded. They could have different levels of interaction, either being present to answer citizen scientists' questions as they participate in the course, or actively leading citizen scientists through training, using the device for its testing functionality.

8.5 Citizen Science as a Social Experience

Participants saw further opportunities for a tool such as SCIENCECACHING to support the various social interactions amongst the people involved in citizen science. In particular, participants suggested emphasizing the social "group" nature of data collection, and supporting that through technology, as well as providing additional mechanisms for scientists and coordinators to share and communicate knowledge.

Performing Tasks as a Group. Participants (P1,2,4,9) saw the opportunity for technology to support tasks to be performed as a group. Some projects need groups to perform tasks, as (P2) described in a water-monitoring project she participated in:

> "*You have to have three people. That's because you are measuring slopes of streams. So purely for carrying in the equipment, holding the equipment and safety [it is typically collected in groups]."*

(P2,4) also said it is more efficient for one person to collect the data while another records it. (P7,8) added that with more complicated projects, the group may comprise both scientist and citizen scientists, where the scientists would direct the citizen scientists while also monitoring their work.

Future ScienceCaching systems can support such group tasks in different ways. Multiple people could log onto the same mobile device when collecting data, so their actions are recorded. Group training could be similarly supported in this way, although individual testing might still be necessary. Scientists could require collaboration for performing collection tasks that need multiple individuals. To make group collection easier (and to allow socialization as discussed in the next section), individuals could meet via the ScienceCaching system, possibly using networked calendars, chat rooms, or social network groups.

Social Motivations of Citizen Science. Participants stressed that many citizen scientists are motivated to participate in citizen science because of its social atmosphere (P1,4,6,7,9). They reported that some of their projects needed social interaction if they were to have people perform tasks they otherwise would not be inclined to do:

> "*A bunch of people getting together sharing coffee … at 4 a.m. sitting out there in the morning light banding birds … has a social dimension to it. (P1)"*

Many citizen scientists, especially the older age group, need this motivation to perform any task (P1,4,7). As (P1) said:

> "*The folks I've dealt with beyond the age of 50 are there for the social experience and have no interest probably going out and [doing one of your caches] by themselves."*

(P1,9) also described how citizen scientists find it important to combine their volunteering with social events, such as eating a meal or going to a bar. This type of volunteerism can be accommodated in future versions of the system. A citizen scientist could, for example, create a meet-up directly in a system like SCIENCECACHING, or meet-ups could be supported in existing social networking sites like Facebook (P1). These meet-ups should flexibly support different social needs, be it meeting for breakfast before collection, using citizen science data collection as a dating activity, or ridesharing. The less a scientist or coordinator has to be part of assigning tasks, the greater potential for these to work into citizen scientist social preferences (P1). Citizen scientists could also be made aware of one another when performing tasks, so they can interact and possibly collect together.

Sharing Scientist Knowledge and Outreach. Participants emphasized that scientists need to share their knowledge with citizen scientists. This is especially important for scientists involved in Provincial/National Parks, as outreach is part of their mandate. (P4), who coordinates citizen science volunteers, discussed the need to support this dual goal of outreach and citizen science, stating:

> "*We who get to work [at these parks]… have special experiences because we work in these places. We can share them more, and more effectively."*

This connection is also important to citizen scientists. Participants pointed out that citizen scientists were highly motivated by direct connection with scientists and experts (P1,3-5,9). Indeed, getting this connection with scientists was seen as a motivating factor both in getting citizen scientists started with a project (P1,3-5,9), and with them coming back (P1,3-5). Participants also thought technology could be a vehicle to help support this connection (P3-5,9). As (P4) stated:

> "*I think you can… use technology as something that can possibly make it easier for the scientist or researcher to share. …Whether that is done in person, or if it is facilitated in another way.*"

For example, cache sites could be augmented with videos created by a scientist, accessed on-site (e.g., via a fiduciary) via a mobile device. These could provide additional information or context about the collection site and the area. These videos could also share, more generally, the scientist's work, or help answer citizen scientists' questions.

8.6 Practical Deployment of Caches

Use of physical caches/containers was seen as simultaneously exciting and problematic among the scientists. On the one hand, they present a number of new opportunities, but maintenance may be troublesome. Nevertheless, participants identified several applications where the benefits would outweigh the problems with physical caches.

Value of Physical Caches. For physical caches to be of use, a site must need multiple visits. If multiple visits were needed for a specific location, participants saw that a physical marker would make the site easier to find (P1-9). The more visits needed, the more useful a physical cache could be, whether the site is a collection or training site (P1-9).

One use of a cache was simply as a site marker. Indeed, many participants already used physical markers to support refinding: as described earlier, (P6,7) use rock cairns, and (P5) had used rebar (reinforcing steel bars) embedded in the ground..

We suggested that caches allow different tools and other materials to be stored in the field. While participants saw that using caches for this "would work" (P2,3,8), no participants identified specific ways to incorporate such tools into their own existing collection projects. This could be because participants has already developed alternate non-cache strategies. For example, (P1) said that hey preferred to meet up before and "give them their pack and they are off for the day".

Participants were more interested in using caches to store samples (P3-5,8,9). Many specific applications of caches holding samples were suggested, including: animal hair (P3), bear DNA (P4), and water samples (P3,8). These sampling ideas were especially interesting to participants in the context of a cache near a trailhead or road to hold samples, as they would be easily accessible for scientists to pick up, (similar to the Area Cache example) (P3-5,8,9).

Problems with Physical Caches. (P1,4,6,7) identified that one problem with caches is the amount of effort needed for their creation and deployment:

> "*By the time I get here [to the cache site], I could have done this five times.*" *(P1)*

While this seems to contradict the fact that some scientists deployed physical markers at their sites (P5,6,7), the difference seems to be that such markers required less effort to create (e.g., in some cases markers were created from natural elements available on-site). (P1,4) said that the citizen scientist could be given the materials to make and deploy caches themselves, which could be a good way to decrease the scientist's effort.

(P2,4) were concerned about the impact that cache containers could have on the environment, especially with remote areas. (P2) noted that scientists are always wary of the impact of geocaches, but that caches may be less visually obtrusive than putting marks on a tree. (P4) shared the same concern that geocaches conflicted with scientist's interests in preserving nature. Scientists (P5-7) did use markers in their practice, so in some cases the impact was deemed worthwhile. (P4) thought impact could be mitigated by removing caches when no longer needed, and even suggested monitoring the site to determine the impact of the caches themselves.

Project Areas Amenable to Physical Caches. Participants suggested that projects which already deploy physical tools at a site (e.g. remote wildlife cameras) could be made into caches. One suggested example was where citizen scientists could help scientists with remote wildlife cameras by performing needed tasks: downloading camera data, replacing storage (SD, Compact Flash) cards, replacing batteries, or moving cameras to new locations. (P3,5) described a project which had citizen scientists maintaining wildlife cameras for monitoring wolverines in Alberta. Because these remote camera are already deployed, they could easily be treated as caches, giving a structure for citizen to find and maintain them. Participants also discussed using caches for non-citizen science park needs, such as to maintain fences and signs (P4). The maintenance tasks needed on these objects are similar to those on wildlife cameras: checking on the park object and repairing it if needed. This makes applying geocaching concepts to these needs possible with little added effort.

Sites that need to be visited many times over a long period would be suitable for cache deployment. One example introduced by (P1,5) was the use of repeat photography to monitor land and engage the public. The location could be marked with a cache (P5 used pieces of rebar in her project) and stored digitally. Citizen scientists could access the site for several years, monitoring the location and taking repeat photographs at that exact spot (see also [7, 17]). Another example presented by (P6,7) used rock cairns to mark pika nests (a pika is a small rodent), so that they could be counted over several years. Similar to the repeat photography project, these marked locations could be stored digitally for easier access and review. With both of these projects, the exact site location could be photographed. Then, if the cache was tampered with, lost or destroyed for other reasons, citizen scientists could possibly re-find the site via these photographs. This allows the long-term data collection to continue.

9 DISCUSSION

Our research places itself inside a large body of research in participatory sensing and citizen science. Our application of geocaching and mobility is one of many possible approaches for dealing with the four problems targeted in our work. To apply our ideas it is important to see how our research fits with these other approaches. The discussion below builds upon our initial premises, the feedback provided by our participants, and other uses of geocaching and crowdsourcing in the literature.

Data Collection. Geocachers are motivated to perform location-based tasks [18]. The primary reason other projects have applied geocaching to citizen science and participatory sensing is to motivate users to take part and collect larger quantities of data [9]. While SCIENCECACHING did not seek to better understand motivation within this application, all our research is amplified by the potential for more volunteers and more collections, especially if these activities can be fun and integrated into daily life. Geocachers are motivated in part due to the physical nature of geocaches [19] and our research shows how this physicality can be used for the collection of data.

Data Validation. Crowdsourcing communities have developed ways of performing quality control on the information they have of these practices have arisen out of social practice in those communities, and some through the design of the site. For example, Wikipedia’s community has developed implicit forms of communication that aid editing with large numbers of users [13]. These social practices lead to certain users becoming de facto leaders of large edits. The validation structures of these different communities stand to inform citizen science on how validation can be handled socially. This can be combined with the way a system like SCIENCECACHING performs repeat collections and the viewing of recorded data in the field, in order to create new structures for data validation. These could range from validation being performed in real-time by digital users [2, 23] to the creation of validation pyramids, where leaders validate the validators; beyond simply data validation, this kind of scheme can be applied to actual analysis of data to identify new patterns altogether [24].

Advanced Collection. Much of the existing mobile citizen science work is in this space – with its primary focus to take advantage of the advanced sensors in a mobile device for new types of collections. Yet, it is possible perhaps to make use of citizens’ decision-making capabilities in the field to do more meaningful tasks beyond what a novice, untrained volunteer can do. This resonates with activity in asking-sites, such as StackOverflow, or Wikipedia, where users learn from these sites while also engaging in sophisticated text creation, editing, and research [15]. Crowdsourcing work has also seen this potential; for example, Duolingo trains users in a new language and uses those trained users to do advanced translation tasks [22]. Citizen science may be able to train volunteers immersively through games, coordinated social interactions or mobile-guided training sites to expand the type of tasks that can be performed.

Mass Coordination. Most large-scale crowdsourcing projects operate with some form of coordination, with the social nature of this coordination varying considerably across projects. Geocaching tracks those who take part, but coordination across people happens as a consequence of their own reporting. Wikipedia has various ranks/responsibilities delegated to different kinds of people (e.g. some pages are sometimes locked, etc.). StackOverflow assigns different responsibilities to use based upon points earned, but these points are awarded based upon the praise of other users on the site. This social nature influences the way projects perform different tasks such as collect data, validate that data, train their users and motivate their users to take part. These different structures of coordination can inform citizen science projects how the structure coordination, depending on the type of project and number of volunteers.

10 CONCLUSION

Our SCIENCECACHING prototype / technology probe contributes a view of how geocaching and mobility can address core problems in citizen science, namely data collection, validation, training and coordination. Our technology probe engendered discussions with expert participants, who provided us with detailed feedback that expand this view and aid in its application. While there is still much left to do, systems such as SCIENCECACHING, extended by the ideas and applications that emerged after feedback was gathered, can ease the citizen science process and potentially broaden the citizen science volunteer audience.

Acknowledgements Our participants deserve special thanks, as each spent considerable time discussing our project with us as well as their visions for how technology can help them. Funding provided by NSERC (National Science and Engineering Research Council of Canada), and AITF (Alberta Innovates Technology Futures).

References

1. Audubon Society (2012) Christmas Bird Count. Audubon Society. http://birds.audubon.org/christmas-bird-count. Accessed July 18, 2014
2. Bigham JP, Jayant C, Ji H, Little G, Miller A, Miller RC, Miller R, Tatarowicz A, White B, White S, and Yeh T (2010) VizWiz: Nearly real-time answers to visual questions. In Proceedings of the 23nd annual ACM symposium on User interface software and technology (UIST '10) ACM, New York, NY, USA, pp 333-342
3. Booney R, Cooper, CB, Dickinson J, Kelling S, Phillips T, Rosenberg KV, and Shirk J (2009) Citizen science: a developing tool for expanding science knowledge and scientific literacy BioScience. 59(11), pp 977-984
4. Burke J, Estrin D, Hansen M, Parker A, Ramanathan N, Reddy S, and Srivastava M (2006) Participatory sensing. In World Sensor Web Workshop, pp 1-5
5. Cohn JP (2008) Citizen Science: Can Volunteers Do Real Research?. BioScience 58, American Institute of Biological Sciences, pp 192-197
6. Cooper S, Khatib F, Truille A, Barbero J, Lee J, Beenan M, Leaver-Fay A, Baker D, Popović Z, and Foldit players, (2010) Predicting protein structures with a multiplayer online game. Nature 466, pp 756-760
7. Crimmins MA, Crimmins TM (2008) Monitoring Plant Phenology Using Digital Repeat Photography. Environmental Management 41 (6), pp 949-958
8. Graham EA, Henderson S, and Schloss A (2011) Using mobile phones to engage citizen scientists in research. Transactions American Geophysical Union 92(38), pp 313-315.
9. Han K, Graham E, Vassallo D, and Estrin D (2011) Enhancing Motivation in a Mobile Participatory Sensing Project through Gaming. IEEE SocialCom 2011, pp 1443-1448
10. Heggen, S (2012) Integrating participatory sensing and informal science education. In Proc UbiComp ‘12, pp. 552-555
11. Huang KL, Kanhere SS, and Hu W (2010) Are you contributing trustworthy data?: the case for a reputation system in participatory sensing. In Proc MSWIM ‘10, pp 14-22
12. Hutchinson H, Mackay W, Westerlund B, Bederson BB, Druin A, Plaisant C, Beaudouin-Lafon M, Conversy S, Evans H, Hansen H, Roussel N, and Eiderbäck B (2003) Technology probes: inspiring design for and with families. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM, New York, NY, USA, pp 17-24
13. Kittur A, Kraut RE (2008) Harnessing the wisdom of crowds in wikipedia: quality through coordination. In Proceedings of the 2008 ACM conference on Computer supported cooperative work (CSCW '08) ACM, New York, NY, USA, pp 37-46
14. Kuznetsov, S (2013) Expanding our visions of citizen science. Interactions 20 (4), pp 26-31.
15. Mamykina L, Manoim B, Mittal M, Hripcsak G, Hartmann B (2011) Design lessons from the fastest Q&A site in the west. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11), ACM, New York, NY, USA, pp 2857-2866
16. NEON and the Chicago Botanic Garden (2014) Project Budburst. http://budburst.org/en Accessed July 18, 2014

17. Nerds for Nature (2013) Monitoring ecological change with smart phones and social media. http://nerdsfornature.org/monitor-change/ Accessed July 18, 2014

18. Neustaedter C, Tang A, and Judge TK (2010) The role of community and groupware in geocache creation and maintenance. In Proc CHI 2010, pp 1757-1766

19. O'Hara K (2008) Understanding geocaching practices and motivations. In Proc CHI 2008, pp 1177-1186

20. Silvetown J (2009) A New Dawn for Citizen Science. Trends in Ecology & Evolution 24, pp 467–471

21. Valdes C, Ferreirae M, Feng T, Wang H, Tempel K, Liu S, Shaer O (2012) A collaborative environment for engaging novices in scientific inquiry. In Proceedings of the 2012 ACM international conference on Interactive tabletops and surfaces (ITS '12) ACM, New York, NY, USA, pp 109-118

22. von Ahn L (2013) Augmented intelligence: the Web and human intelligence. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences

23. Willet W, Aoki PM, Kumar N, Subramanian S, Woodruff A (2010) Common Sense Community: Scaffolding Mobile Sensing and Analysis for Novice Users. Pervasive 2010, pp 301-318

24. Willet W, Heer J, Agrawala M (2012) Strategies for crowdsourcing social data analysis. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12) ACM, New York, NY, USA, pp 227-236

25. Wiggins A, Newman G, Stevenson RD, 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.

26. Yeh RB, Liao C, Klemmer SR, Guimbrtiere F, Lee B, Kakaradov B, Stamberger JA, Paepcke A (2006) ButterflyNet: a mobile capture and access system for field biology research. In Proc CHI 2006, pp 571-580