THE UNIVERSITY OF CALGARY

Science Caching: Applying Geocaching to Mobile Citizen Science

by

Matthew Alan Dunlap

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF COMPUTER SCIENCE

CALGARY, ALBERTA

MAY 2013

THE UNIVERSITY OF CALGARY

FACULTY OF GRADUATE STUDIES

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies for acceptance, a thesis entitled "Science Caching: Applying Geocaching to Mobile Citizen Science" submitted by Matthew Alan Dunlap in partial fulfillment of the requirements for the degree Master of Science.

____

Supervisor, Saul Greenberg
Department of Computer Science

____

Supervisor, Anthony Tang
Department of Computer Science

____

Jonathan Sillito
Department of Computer Science

____

External Examiner, Shelley Alexander
Department of Geography

____

Date

Abstract

Citizen science occurs in part when scientists work with volunteers to collect science data in particular field locations. The benefit is that citizen science eases and lessens the cost of collecting such information. Yet it has a variety of known problems. This document focuses on four specific citizen science problems concerning difficulties in *data collection, data validation, volunteer training* and *volunteer coordination*. The thesis is that these problems can be mitigated by applying aspects from another thriving location-based activity: the geocaching treasure hunt as enabled by mobile devices. Citizen science can exploit geocaching's location-based design, its use of physical objects, and its user maintained content. To explore and critique this thesis, a prototype mobile system called Science Caching was developed, along with various scenarios that describe how it addresses issues in collection, validation, training and coordination. The system and scenarios – which serve as a working sketch – were shown to citizen science experts via an interview-based design critique. In particular, they provided feedback on the choice of the problems addressed by the system, the approach to the problems as realized by Science Caching, how those approaches could be extended, and what other areas in citizen science they could be applied to. The results were analyzed via affinity diagramming, which uncovered various overarching themes. Generally, the combination of geocaching and mobility was received quite positively, where participants indicated various areas where it would be applicable. Problems and improvements were also suggested, giving insight into future iterations of the method and the system.

Acknowledgments

To Saul, my advisor, who provided me the opportunity to continuously improve myself. You pushed me to be sharper and gave me the confidence to succeed in this world. Your guidance and commitment was vital in pursuit of a research topic new to us both.

To Tony, my advisor, who has remained a positive force throughout my studies. You helped me hone my research when it was at its roughest. Our brainstorming sessions were the source of so much inspiration for this work.

To Tedd, who introduced me to Calgary, helped me settle here and has been a giving friend. Our learning together in undergrad allowed me to excel as a computer scientist and our times in grad school have retained my sanity in the face of intense strain.

To Rachel, who has never stopped loving or believing in me. You understood when my studies left me without time and energy, yet you pushed me to find time and energy I did not believe I had. Our times together will always remain with me.

To Andy, my brother and best friend, who continues to remind me how awesome every moment in life can be.

To all the iLab, past and present. Our ever-changing community is a sight to behold.

To my Grandma Mucci, whose beauty, grace and relentless pursuit of knowledge has me hoping to grow old, not fearing it.

To Calgary and all my friends in it, who have made my stay here kind, safe and beautiful.

To my parents Annemarie and Milton, who have always been there for me. Your love is something I have never doubted, and continues to give me strength when times are rough. Thank you for making me who I am.

To my whole family, who I have missed immensely during this beautiful but trying time.

Dedication

I dedicate this thesis my Grandfather Taton. You were amazing, strong, loving, and I wish you were still with us. Hours before passing from cancer, you took my brother and me aside, telling us to push as hard as possible for what we want in life. This has never left me as I continue to discover who I am.

Table of Contents

List of Tables

List of Figures

Chapter 1. **Introduction**

1.1 Motivation

Scientists always need more information to discover the patterns and rules that define our world. One way that scientists get this information is by working with volunteers to collect or process scientific data, a method known as citizen science (Silvertown 2009). Citizen science connects generally non-expert volunteers (citizen scientists) with the scientific process and natural world. One example of citizen science is the Christmas Bird Count, in which 60,000 to 80,000 bird-watching citizen scientists record sightings around Christmas. These sightings are sent to the Audubon Society, compiled, and then used to inform bird research (Cohn 2008; Audubon Society 2012). Compared to using a small pool of scientists, citizen science provides an affordable means for collecting and processing large quantities of science data.

Citizen science has always taken advantage of new technologies to improve data collection. In the past, this has meant the use of desktop computers and databases to analyze collected data. Recently, technologies such as the Internet, GPS, and mobile devices are becoming an important part of data collection and data analysis. They allow citizen scientists to collect data more accurately and create less work for scientists processing the collected data (a large problem with using paper forms) (Cohn 2008).

One project that illustrates how new technologies are incorporated is Project BudBurst[1], where citizen scientists collect data on *plant phenology*: the timing of yearly

[1] http://neoninc.org/budburst/

plant biological events. This project is accessible and shared over the Internet, exposing new groups to citizen science. Participating citizens use GPS to quickly and accurately record the locations of plants, allowing repeat visits and improved geography-based analysis. Citizens submit collected data instantly from any internet-connected computer. This process can also be done through a smartphone application. The application provides project resources on demand in the field (e.g., taxonomic information), simple digital data collection forms, the annotation of recordings with location data, descriptions, pictures, audio and video, and the instant submission of data to scientists.

Figure 1.1 illustrates Project BudBurst's use of mobile technology via screen snapshots of selected application pages. The citizen scientist who wish to report a reading begins on the left page and chooses to do a single report. She is taken to the middle page, where she is provided with a taxonomy for plant identification. After selecting the species of plant and its phenological phase (screen not shown), she is taken to the right page to annotate the recording with a photograph and description (GPS data is added automatically).

Figure 1.1: Project Budburst mobile pages. From the main page (left), a citizen scientist can browse through a taxonomy and select a plant to record (middle), and annotate the recording with photos and text (right).

In this example, the benefit of technology on citizen science can be seen in specific ways. Foremost, it is very easy for a citizen scientist to collect new information; everything needed to record is built into the mobile device. The recording of location data is automatically performed when data is recorded. Information to assist identification is easily accessed through the recording device. Photographs are easily attached to the collected information, improving trust in the information. The power of this technology allows citizen science to be done by more people with less effort.

Even with these benefits, citizen science still has problems, most arising from working with disparate groups of volunteers who do not have the same level of experience as the coordinating scientist (Silvertown 2009; Cohn 2008). In this research, I look consider four specific problems:

1. **Data Collection.** Citizen science is primarily about data collection. The act of collecting data is complicated and contains many tasks, including: finding the collection site, using tools, recording data and submitting that data to scientists (Cohn 2008; Kim et al. 2011). Some of these are supported by current technologies, but there is room to design new and better ways of performing these tasks as well as supporting new tasks (Silvertown 2009).
2. **Data Validation.** Collected data needs to be validated in order to be trustworthy. Citizen scientists can make mistakes, misunderstand collection needs, or purposefully submit malicious information (Silvertown 2009). To check for these errors, scientists and automated systems validate information. Yet manual validation by scientists is time consuming and automated systems can only catch some errors (Wiggins et al. 2011).
3. **Volunteer Training.** Training is important for citizen scientists to be able to collect quality data, as well as perform more complicated tasks. Direct training by scientists

provides feedback and immersive education, but this can be time consuming and impossible with large or widely distributed projects (Cohn 2008). Training materials such as videos and pamphlets can be used, but hands-on, real world education has been shown to improve retention (Haury and Rillero 1994).

4. **Volunteer Coordination.** Citizen scientists need to be coordinated to best achieve project goals. Communication between participants (scientists and citizen scientists) is needed for efficiency and validity (Silvertown 2009; Bendell 2011). Information about citizen scientists' knowledge and activities is needed to correctly assign tasks. There is potential to support communication and information through new technologies (Greg Newman et al. 2012).

In general, my thesis concerns how we can mitigate the four problems. My approach specifically considers how particular technologies (GPS-enabled mobile devices) and particular technological approaches (geocaching, CSCW and Mobile HCI) can provide new opportunities for easing these problems in Citizen Science.

1.2 Background

The context of my research is outlined in Figure 1.2. My work is firmly placed within the discipline of Human Computer Interaction, drawing mostly from the fields of Computer Supported Cooperative Work (CSCW) and Mobile Device Interaction (Mobile HCI). My work also draws from geocaching, a location-based treasure hunt. I discuss aspects of these areas that are important to my research.

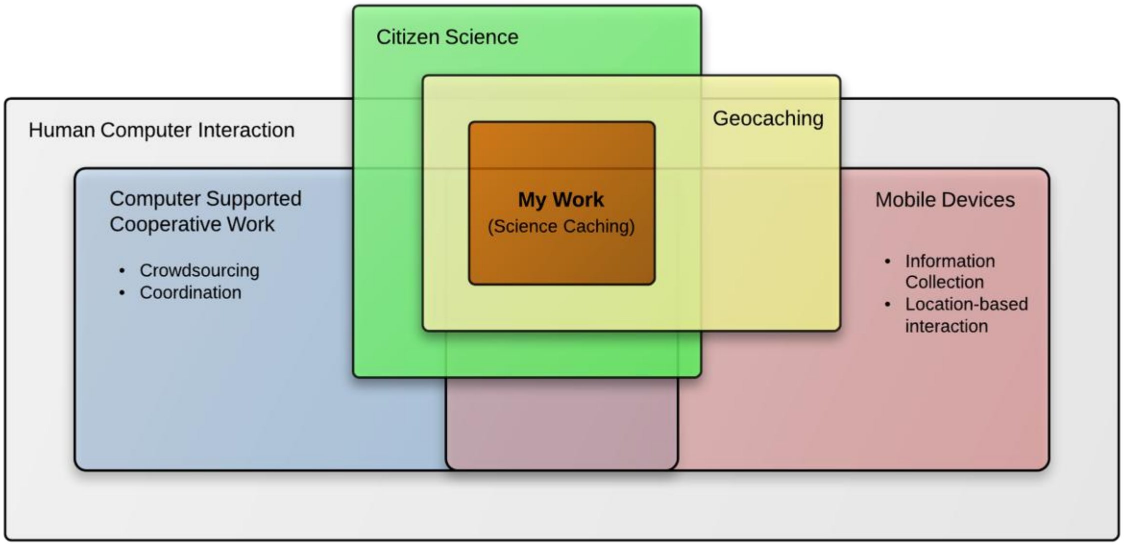

Figure 1.2: Research context

From CSCW, I consider how scientists can use crowdsourcing and coordination to mitigate problems in different citizen science areas, for example volunteer coordination and data validation.

From Mobile HCI, I consider how we can take advantage of the many capabilities of mobile devices to afford effective mobile information collection. This entails the use of on-board sensors (including GPS), the availability of content on demand, the ability to collect and transmit data at any time and particularly how mobile devices can present information about a location.

Finally, my research is centered largely on important concepts from *geocaching*. Geocaching is a location-based game, where participants hide physical containers (caches) in known physical world sites for others to find via GPS. The social nature and rules of use keeps the geocaching crowd creating and hiding new caches (O'Hara 2008; Neustaedter et al. 2010). I consider how various aspects of geocaching – such as its approach to cache discovery, cache site information, use of physical containers and user generated content – can be applied to mitigate the aforementioned problems in citizen science. My approach is to apply and modify aspects of geocaching to citizen science, rather than have scientists and citizen scientist fit their work into the current Geocaching[2] game structure.

In sum, geocaching – and the surrounding literature on Mobile HCI, CSCW and HCI – is key to how I aim to use technology and technological methods to support citizen science. I will specifically look at how scientists and citizen scientists can interact with fixed, known sites through mobile devices for: collecting various kinds of data, validating the data of others, on-site training, and coordinating their activities. I have dubbed this area of study Science Caching.

1.3 Research Goals

My overarching research question is:

> **How can we use mobile technology and aspects of geocaching to inspire and design new solutions for citizen science's problems with data collection, data validation, volunteer training, and volunteer coordination?**

I have four thesis goals; all centered on pursuing techniques that target the four areas outlined in my research question.

1) Understand how *data collection* can be performed by leveraging known prepared sites and mobile devices.
2) Explore how data can be *validated* through use of known, repeatedly visited sites.
3) Investigate how to *train volunteers* in collecting new forms of data by interacting with mobile devices and real world training sites.

[2] In this thesis I distinguish common noun geocaching, the activity and concepts surrounding the location-based game, from proper noun Geocaching, the prevalent avenue for performing geocaching, which is run by Groundspeak (www.geocaching.com)

4) Examine how to *coordinate volunteers* around science sites through direct communication with scientists and better understanding of individual talents and abilities.

To achieve these goals, I will design a citizen science system and garner feedback about it. First I will implement the goals stated above in a prototype. With this done, I will hold design critiques with scientists, citizen scientists others knowledgeable in citizen science to present this prototype, the ideas it represents, and ask questions about them. Feedback is sought to test the validity of my ideas and to extend those ideas.

1.4 Contributions

The major contribution of this thesis is the *creation, refinement,* and *critique from experts* of *Science Caching;* a way of using mobile technology and aspects of geocaching to mitigate citizen science's problems with data collection, data validation, volunteer training and volunteer coordination.

Lesser contributions are also made to citizen science research:

- In-depth discussion of prior solutions related to citizen science's problems with data collection, data validation, volunteer training and volunteer coordination.
- Design and implementation of the Science Caching prototype, including non-geocaching focused aspects, which informs future design of mobile citizen science.

1.5 Thesis Outline

The thesis is structured as follows.

Chapter 2 provides background on citizen science, detailing the problems of *data collection, data validation, volunteer training* and *volunteer coordination*. It also reviews geocaching concepts that are relevant to citizen science.

Chapter 3 discusses how geocaching, mobile devices and other research can be applied to citizen science, focusing on to solutions for problems in *data collection*, *data validation, volunteer training* and *volunteer coordination*.

Chapter 4 dives into the design and implementation of the Science Caching system. Using a series of story-based scenarios, it illustrates how Science Caching addresses the four issues of *data collection*, *data validation, volunteer training* and *volunteer coordination,* in that order.

Chapter 5 analyzes the ideas presented in Chapter 3 through a design-critique involving experienced scientists, coordinators, and citizen scientists. It then presents the themes that came from the analysis of my critique results: *discussion on targeted problems, citizen science as a social experience* and *practical deployment of physical caches*.

Chapter 6 concludes the thesis. It discusses the overall contributions and implications of this work, along with future directions.

Chapter 2. **Background**

My overall research goal, set out in Chapter 1, asks how we can use mobile technology and aspects of geocaching to inspire and design new solutions for four citizen science problems. This chapter sets the scene behind this goal. First, I further describe citizen science and how projects are structured. Second, I isolate and discuss targeted sub-problems from my four citizen science problem areas: *data collection, data validation, volunteer training* and *volunteer coordination*. Finally, I introduce geocaching, describing how it works as an outdoor treasure hunting activity and introducing aspects important for citizen science.

2.1 Citizen Science

This section explores citizen science. I define citizen science, describe how it manifests itself over various types of projects, and then detail four particular problem areas within it that are addressed in this thesis.

There is no standard definition of citizen science, and indeed there is debate over the roles that citizen scientists should play in projects. This includes factors such as the degree that volunteers[3] participate in influencing the scientific process, such as being involved in hypothesis formation (Fitzpatrick 2007; OpenScientist 2011). Silvertown (2009) provides a concise and narrow definition of a citizen scientist as: “a volunteer who collects and/or processes data as part of a scientific enquiry.” With the focus of this thesis

[3] In this document, citizen scientist and volunteer are used interchangeably.

in mind, I use this definition of a citizen scientist to inform my definition of citizen science: *scientists working with volunteers to collect or process scientific data*.

As discussed in Chapter 1, the citizen science partnership benefits both scientists and citizen scientists. Scientists faced with tight budgets and the need for larger bodies of data gain help in performing collection, processing and analysis tasks. In turn, citizen scientists are educated on the domain being explored and on the scientific process; they also make connections to the scientists themselves (Silvertown 2009; Cohn 2008).

Citizen science projects are highly varied, each with its own particular nuances and different ways that it can be structured to perform needed tasks. For this thesis, it is important to understand whether a project focuses on the collection or analysis of data.

- ***Collection projects*** focus on the collection of new science data. Citizen scientists generally go outdoors to collect information for scientists. Collectors may use a variety of tools such as GPS and measuring devices (discussed in more depth in 2.1.1 Collection). This data is recorded on paper forms, desktop computers or mobile devices and sent to scientists. These projects can be small or large. The research in this thesis is focused on the collection form of citizen science.
- ***Analysis projects*** are projects where citizen scientists analyze already collected data. Most projects of this kind use a networked application to provide this information. Via these applications, citizen scientists are trained (mostly on how to process collected data), and then provided with the actual raw data to be processed. These projects are generally large. Data is typically made accessible to anyone from anywhere. For example, Galaxy Zoo is a web system where citizen scientists classify galaxy photographs taken by a robotic telescope (Lintott et al. 2008). Analysis projects such as Galaxy Zoo show how citizen science has gained from the related field of *crowdsourcing*, a term coined by Jeff Howe (Howe 2006) as "the act of taking a job traditionally performed by a designated agent (usually an employee) and

outsourcing it to an undefined, generally large group of people in the form of an open call." Specifically, analysis projects draw from the crowdsourcing concept of *microtasks*, which are small, easy human tasks that can be performed with minimal training, but that are challenging for computers. In the case of citizen science analysis projects, these tasks are ones that help solve science problems. While this thesis is not focused on analysis citizen science, it does apply some aspects of crowdsourcing both to data collection and data validation within collection projects.

From this point on, this thesis will only discuss citizen science in terms of collection projects, unless stated otherwise.

2.2 Four Problem Areas in Citizen Science

Citizen science has clear benefits, as it provides scientists with the ability to collect and process large bodies of information. Yet citizen science has many issues associated with it. Some issues revolve around the inter-personal aspects of citizen science, such as how to motivate citizen scientists, and how to provide them with feedback about how their collected data is being used and what it has found (Silvertown 2009; Cohn 2008). This thesis does not attempt to address all citizen science issues. Instead, I concentrate on four specific problems (introduced in Chapter 1) related to core aspects of 'in the field' citizen projects: *data collection*, *data validation*, *volunteer training* and *volunteer coordination*. In particular, I focus on these four problems because I hypothesise that they can be mitigated through mobile devices and elements of geocaching.

2.2.1 Data Collection

The core of collection-based citizen science is how data is collected. Two methods are used to organize data collection in citizen science. In *systematic data collection*, a *site* is specified where phenomena (e.g. animal sighting, soil sample, temperature reading) is to be recorded or gathered. In contrast, *opportunistic data collection* does not specify the

Parts of a Collection Task	Example Tools Used
Finding / identifying the location	GPS, photograph or description
Collecting phenomena through…	
• **identification**	Guides, taxonomies
• **sensing**	Thermometer, barometer, air quality sensor
• **physical sample gathering**	Airtight container, shovel, pickaxe
Recording data	Paper, computer, mobile device

Table 2.1: The different parts of a collection task and the tools used in performing them. Different tools are used to collect different types of phenomena

site where data is to be collected, and in many cases collection is instantiated upon the volunteer witnessing a phenomena (Teton Science Schools 2009; Colorado State University 2013). These methods differ in the type of information provided to scientists. Yet both types of task share some common elements. First, the site of collection must be known (ahead of time in systematic data collection, and at the time of collection in opportunistic data collection). Second, the phenomena must be identified, read, and/or sampled correctly. Third, the information on the phenomena must be recorded (ideally) without error.

Another facet of citizen science projects is the *tools* used to perform the above three task elements during the data collection task. Table 2.1 shows a sampling of tools used for each part of a collection task. For example, a citizen science requires tools to find and/or identify a location; tools to identify, sense and perhaps physically gather samples of the phenomenon of interest; and tools to record that information. Citizen scientists currently use these tools, where they are able to produce usable data for scientists. Yet current methods could be improved upon. To improve data collection, my work targets three collection-specific problems.

1. ***Finding Sites.*** Systematic collection is hampered by ability to find the site and the specific location of phenomena within it (e.g. a specific tree to be recorded). Currently, if a scientist creates a site to be visited, markers (e.g. marking tape, paint) may be used to identify the site and phenomena to be collected (e.g. University of Alaska Fairbanks 2012). Yet communicating site information to citizen scientists is complicated. Problems with site finding in systematic collection is further exacerbated when it is the citizen scientists themselves who create a site or mark new phenomena, as this now has to be communicated to the scientist, who in turn must relay this to other citizen scientists (Kim et al. 2011).
2. ***Creating Sites.*** The creation of collection sites by citizen scientists could save scientists' time and effort, but is limited by a lack of communication and tools. Few projects allow citizen scientists to create their own sites that can be accessed by others; those that do usually require scientists to take the site information provided by the citizens and enter it themselves (Han et al. 2011).
3. ***Transferring Tools, Data, and Samples.*** Procuring a tool for data collection or providing a sample to a scientist often requires that the involved parties meet, which requires coordination and is time-consuming. Data collection tools can be owned by the citizen scientist, but this is an expensive barrier to participation. Samples require meeting with the scientist afterwards to transfer them, although in some case physical mail may be used (e.g. Paleontological Research Institution 2013). Tools may also need to be transferred before and after a collection if they are in limited supply, adding to the time spent.

2.2.2 Data Validation

The value of data collected by citizen scientists is impacted by the validity of that data, that is, by the extent to which that data actually measures what it claims to measure. There are a variety of factors that can impact the quality of collected data. Some sources of error center around the instruments used to collect that data, where problems can arise from (for example) sensor malfunctions and tool contamination. Another source of error results from problems that could occur with the collection protocol itself. The citizen scientist himself can introduce yet another source of error. This includes errors made when collecting data due to the citizen scientist's lack of skill in performing a collection task, and date entry errors made when the citizen scientist enters the collected data into a data record (Wiggins et al 2011).

Wiggins et al. (2011) report several commonly used techniques to detect poor quality data, with the most common ones listed below. While a few techniques are done automatically by a system, most currently require considerable manual effort.

- *Expert review*, where scientists manually review collected data to see if there are outliers; some projects also allow for peer review of that data (Wiggins et al. 2011).
- *Paper data sheets submitted along with online entry*. Wiggins et al. (2011) report that one third of all citizen science projects they surveyed require people to submit their data both electronically and on paper. The rationale is that citizen scientists often capture more details on paper than is allowed on the electronic form, or that errors arise from transcription.
- *Replication or rating by multiple participants*, where multiple data points are redundantly collected by multiple people of the same phenomenon. A lesser used method asks the same participant to replicate their readings. This duplication and overlap allows differences and errors to be detected and dealt with.
- *Training programs* involve quality assurance and quality control in an attempt to minimize errors.
- *Automated filtering of unusual reports* (e.g., for possible outliers: Bonter and Cooper 2012). The system may include sets of reasonableness or completeness bounds for

that data, which could have been previously established (for example) by experts who had piloted ground-truth measures (Wiggins et al. 2011).

- *Rating of citizen scientists on established control items*, to test and verify that the performance of the citizen scientist and their equipment returns the expected result.

These are not the only methods. For example, some projects try to minimize errors by ensuring that instruments are regularly calibrated, by evaluating participant reliability, and in turn by using only known and qualified participants. Other projects use participants only for collecting samples, where those samples are later identified by experts (Wiggins et al. 2011). Some projects have citizen scientists take part in the review of collected data. To illustrate this, LeafSnap made the identification of photographed leaves into a matching game for citizen scientists to play (LeafSnap 2012). Sightings recorded for iSpot's nature sighting social network are posted online for other iSpot members to review. A photograph and identification is posted and other citizen scientists are able to correct errors or flag suspicious recordings (The Open University 2013). In practice, many projects use a combinations of validation methods rather than just a single one.

With both manual and automated validation, other readings from the same site or readings from nearby sites can help determine whether data is suspect. Several techniques are also typically used to handle suspect data. The data may be thrown out, the site may be visited again, or the citizen scientist may be contacted and queried about that data.

The above techniques for detecting and managing data validation have problems. Manual review of data is time consuming and expensive, even if done by research assistants and graduate students (Lintott et al. 2008). Automated review is useful but is limited in terms of the outliers it can identify, especially when dealing with complex data (e.g. photographs, audio recordings, video) (Bonter and Cooper 2012). To improve data validation, my work targets two specific problems associated with data validation.

1. ***How Sites are Revisited.*** As discussed in 2.1.1 Collection – Finding Sites, there is difficulty with finding the correct site for data collection. This also impacts data validation because it hampers repeated visits to a site. Improvements in how sites are found can also improve the ability to collect multiple recordings at a site.
2. ***On-Site Data Comparison.*** It could be beneficial if citizen scientists, while on location and collecting data, are told of suspect data and/or able to test the validity of that data in real time, such as by reviewing the data collected by others at that site for comparison (Wiggins et al. 2011). Currently, however, data collected by citizen scientists is typically validated away from the phenomena in question, after the fact, and usually by a scientist. Techniques for citizen scientists to take part in data validation, such as LeafSnap and iSpot, are a good starting point, but do not support validation of data on-site, where the phenomena takes place.

2.2.3 Volunteer Training

Volunteer training is typically performed in two ways, and sometimes in tandem. First, with *face-to-face training*, citizen scientists meet in person with scientists to learn about the project and how to take part in it. Citizen scientists are often taken out in the wild to experience data collection with the scientist to gain experiential knowledge about the collection task. Second, with *resource-based training*, citizen scientists may be trained through educational resources such as pamphlets, videos and websites. These resources are available on-demand and facilitate learning at a distance. With both varieties of training, the retention of knowledge by the citizen scientist is important, and tests (in person and online) are sometimes used to make sure the citizen scientist is able to do what is needed (e.g., as done by Colorado State University 2009).

These two training methods also have issues. Face-to-face training is extremely time-consuming for scientists, as they must spend time organizing training sessions, educating volunteers, and evaluating what volunteers have learned. With resource-based training, the time required of scientists is low, but the knowledge gained is weaker because the lack of real-world experience (Haury and Rillero 1994). Thus resource-based training can limit the complexity of tasks that can be taught.

To mitigate issues with volunteer training, I target two problems that can benefit from combining aspects of *face-to-face training* with *resource-based training*

1. ***Real-world Examples.*** Resource-based training does not provide citizen scientists with real-world experience. Training that can direct citizen scientists through real-world education without scientist supervision could provide some of the benefits normally associated with face-to-face training.
2. ***Testing on Known Values.*** Online resource-based education makes it easy to test citizen scientists on their training. Tests are also used in face-to-face training, but testing groups and recording results is extremely time-consuming for scientists. In conjunction with the first problem above, if real-world education can be provided through resources, testing could also be done through this experience. This would further enhance the benefit of this combination of education types.

2.2.4 Volunteer Coordination

The coordination of citizen science projects is essential to ensuring they operate smoothly. Mainly, this is done by efficiently organizing citizen scientists (Silvertown 2009; Bendell 2011). A variety of factors influence volunteer coordination, including: setting clear project goals, describing the specific collection tasks, organizing volunteer training, managing volunteer availability. Coordination of these factors requires the scientist to have knowledge about the pool of citizen scientists, and to communicate with those citizen scientists.

A variety of coordination methods are currently used in most citizen science projects. These usually involve face-to-face communication including group meetings and one-on-one meetings, as well as mass communication via email, newsletters and social networking (e.g. Facebook). One-on-one digital communications may also be used to share specific collection sites and collection protocols. Yet these methods can consume considerable time and effort, and even introduce delays over a project life cycle. Consider the simple example of a scientist who needs one or more citizen scientists to collect data at various sites. The scientist and citizen scientists may meet in person, or may communicate by email or telephone. The scientist tells the citizen scientists where, when and how the collection needs to take place. The citizen scientists in response say what sites they can visit due to time, distance and availability constraints. As time goes on, any changes must be communicated between them. If a citizen scientist encounters problems when at that site, he or she have to communicate that back to the scientist, which is often difficult (or impossible) to do. Without this sort of exchange, especially in projects requiring complicated or systematic collection, the citizen scientist would not know what to do and the scientist would not know who could do it.

There are clearly many problems with coordinating citizen science projects. In this research, I target two coordination-specific problems.

1. ***Computer-Assisted Coordination.*** Coordinating projects can consume a large quantity of a scientist's time. It is difficult to ensure that all citizen scientists are on the same page, that they know what tasks to perform, and that they are indeed performing that task. Communication with each citizen scientist individually is especially time consuming, while arrange in-person meetings with more than a few citizen scientists is a major planning hurdle. Without communication like this, it is difficult to keep track of what each citizen scientist knows and is capable of. While a solution is to offload coordination to automated systems (Lintott et al. 2008), this has not been explored deeply in the field-working context.

2. ***Enriched Real-Time Interaction.*** When a citizen scientist is in the field, it may be necessary for direct communication between scientist and citizen scientist. The citizen scientist may have problems with the task they are performing, or the scientist may want the citizen scientist to change their task. With a smartphone, this communication may be done through phone call or email, but is hampered by difficulties in communicating complex information such as location, the data itself, and the state of a site in real time. It is possible to support real-time communication of these types of information directly through a citizen science application, so that both scientist and citizen scientists have access to all the materials known in the system.

2.2.5 Summary

In this section, I briefly introduced citizen science. I identified four specific problems in citizen science, *data collection*, *data validation*, *volunteer training* and *volunteer coordination*, and particular aspects of those problems that I target in this thesis. For reference, Table 2.2 outlines the problems discussed in this section.

Problem	Sub-Problem	Description
Data Collection	Finding Sites	Systematic collection is hampered by the ability to find the site and the specific location of phenomena within it.
	Creating Sites	The creation of collection sites by citizen scientists could save scientists' time and effort, but is limited by a lack of communication and tools.
	Transferring Tools, Data and Samples	Procuring a tool for data collection or providing a sample to a scientist often requires that the involved parties meet, which is time consuming.
Data Validation	How Sites are Revisited	Difficulties finding the correct site for data collection hampers validation through repeat data collection.
	On-Site Data Comparison	It would be beneficial if citizen scientists, while on location and in the process of collecting data, could test the validity of that data in real time.
Volunteer Training	Real-world Examples	Training that directs citizen scientists through real-world education without scientist supervision could be highly effective while saving scientists time and energy.
	Testing on Known Values	Testing citizen scientists during real-world examples would improve education and enhance trust.
Volunteer Coordination	Computer-Assisted Coordination	Automating the coordination of citizen science projects would allow projects to run smoothly with less scientist effort.
	Enriched Real-Time Interaction	Supporting communicating of complex information such as location and the state of a site in real time would improve coordination in the field.

Table 2.2: The problems and sub-problems targeted in this thesis.

Figure 2.1: A geocache with a pen, instructions (top left), logbook (bottom left) and treasures (bottom right).

2.3 Geocaching

I now introduce geocaching, which is of interest to my research in citizen science as it is a successful location-based activity—one that has readily been appropriated by a broad spectrum of participants, and appears to be self-sustaining (Neustaedter et al. 2013). Specifically, I will discuss its cache discovery approach, cache site information, use of physical containers and user-generated content. Later in this thesis, I will show how aspects of geocaching in combination with mobile devices can be used to mitigate the problems described in Section 2.2.

Geocaching is an outdoor treasure hunting activity, where its players (called *geocachers)* hide and / or hunt for 'treasure' using GPS-enabled devices. As illustrated in Figure 2.1, the treasures are hidden in *geocaches*, small containers that typically contain instructions (top left), a logbook (bottom left) and trinkets (bottom right). These containers range in size from film canisters to large ammo boxes, and are hidden as to be unnoticeable by the public while still findable by those who know the cache[4] coordinates. These containers serve two functions: they tell the player that he/she has actually found the location; second, as described earlier, they serve to retain physical objects: e.g. geocaching information, log books and assorted prizes.

Geocaches are hidden all over the world, with their locations stored on various websites. These websites are central to the way geocachers pursue their activity. To illustrate, consider Groundspeak's Geocaching[5], which is perhaps the most popular geocaching website as of February 2013. Figure 2.2 left illustrates a cache listing from that web site. We see how the cache listing includes:

- general information about the cache such as a title briefly describing the cache, the cache creator, the relative cache size, and when the cache was originally hidden;
- a brief description of the route to the cache along with an easy-to-decrypt hint (i.e., so that the geocacher can select how challenging it would be to find the cache);
- the 'difficulty' of the cache and the terrain;
- the cache location in very general to very specific terms, for example, in Virginia, United States, the distance from the viewer's current location, and the exact cache coordinates in various units (latitude and longitude and UTMs);
- ways to print the GPS information, or to download it to various devices (GPS unit, phone) using several standard formats.

[4] Cache (also referred to as geocache in geocaching) is a complex term, with multiple meanings. It encompasses the physical container found in geocaching, the information used to find the container, and its web presence (reports on finding the container, photographs, hints, etc.). When using the term cache, it can mean any of these parts. In this thesis, when clarity is needed, I will use explicit terms (e.g. cache container, cache information, cache website).

[5] www.geocaching.com

Figure 2.2: Left: A cache listing from Groundspeak's Geocaching. Right: Browsing nearby caches with c:geo for Google Android.

While rich in information, the above description describes a fairly simple, single coordinate geocache. There are more complicated geocache variants, with multiple cache-stages and puzzles to find the cache. After a cache is found, geocachers publicly record their finding experience online.

Geocaching as an activity relies on people to discovery geocache information (through the website) to find them (physically). Geocachers use two methods to choose caches of interest. The first is via direct *search*: players can search for potential caches through different characteristics (size, difficulty, etc.). Once a location is found online, its cache information is quickly downloadable to a GPS or mobile device. The second method to access caches is by *proximity*, where geocachers can see what caches are nearby. For example, Figure 2.2 right illustrates a mobile application (c:geo 2013) that shows a listing of nearby caches, where a geocacher can then explore that cache's listing and decide whether or not to pursue it.

The creation of geocaching content is seen as a core part of the geocaching experience, with geocachers sharing locations, treasures and experiences with others (O'Hara 2008). Geocaching content is created almost exclusively by its users. There is a low difficulty-barrier for geocache creation, with basic geocaches being easily created by all players (Neustaedter et al. 2010). More complicated geocache variant require more experience, but are easy to create after finding and creating a few caches. Geocachers also create *travel-bugs*: cache treasures where the goal is to take a travel-bug from one location to another (e.g. from Calgary to Australia).

The content created by geocachers is also maintained by geocachers. Geocaches are maintained through three actions. Through *online logs,* geocachers report their cache finds online as a social activity, saying if a cache was found, not found, or needs to be maintained. Through *in-place maintenance,* geocachers help maintain the caches they find (e.g. by airing out a paper logbook). Through *cache adoption,* if a geocacher can no longer maintain a cache, it can be adopted by someone new (Neustaedter et al. 2010).

A primary aspect of this thesis is to use aspects of geocaching as realized on a mobile device to inspire and design new solutions for various issues and problems in citizen science. My work is not the first to apply geocaching and mobility to citizen science, as discussed in the next chapter. In contrast to these works, later chapters (particularly Chapters 4) of this thesis will reveal how I apply four core geocaching elements – cache discovery approach, cache site information, physical containers and

Geocaching Aspect	Use in Citizen Science
Cache discovery approach	Citizen science sites can be made more easily accessible. Locations can be searched for based on different characteristics (project, distance, difficulty, scientist needs).
Cache site information	Detailed site information such as photographs, descriptions, GPS coordinates and marker information can make site finding easier. By making sites easier to find, multiple people can be directed to sites to provide redundancy or check previous data collections.
Physical containers	Allows the easier finding of data collection sites. Tools and training materials can be placed inside containers, and collected samples stored there.
User-generated content	Citizen scientists can create new data collection sites. These sites can be monitored by citizen scientists when they perform data collection.

Table 2.3: Geocaching Aspects Reconsidered for Citizen Science

user-generated content − to citizen science, where the importance of these elements are summarized in Table 2.3.

2.4 Summary

In this chapter, I set the scene for my research goal, addressing how we can use mobile technology and aspects of geocaching to inspire and design new solutions for citizen science's problems with data collection, data validation, volunteer training, and volunteer coordination. To do this, I explained citizen science in-depth. Second, I looked at the four problems in citizen science at the core of this research, and the sub-problems that I seek to target through a geocaching and mobile device approach. Finally, I introduced geocaching and how it is relevant to citizen science. In the next chapter, I will show how geocaching and mobility can possibly be used to inspire and enhance citizen science.

Chapter 3. **Applying Geocaching and Mobile Devices to Citizen Science**

In this chapter, I reconsider the four citizen science problems described in Section 2.2 − *data collection, data validation, volunteer training* and *volunteer coordination* – from a geocaching and mobile device perspective. At a conceptual level, I suggest how aspects of geocaching, mobile devices and related ideas can be used – or how others have used them – to mitigate these problems. For some problems, the suggestions will directly apply geocaching methods as the primary means to solve the problem. For others, the suggestions will rely on the foundation that geocaching and/or mobile devices provide to consider related ideas. However, I begin by describing the capabilities of mobile devices in this context.

3.1 The Mobile Device

Geocaching usually relies on some form of mobile device (such as a GPS unit) to discover a location. In this thesis, I consider mobile devices more broadly. I assume that, in addition to GPS tracking, the mobile device will have capabilities typically found on a modern smart phone. For example, these would likely include telephony, network connectivity, a high quality display, text and touch input capabilities, microphones and speakers, various near-universal communication applications, numerous integrated sensors including GPS, light sensors, cameras, microphones, compasses, and accelerometers. Such mobile devices can provide many benefits for citizen scientists who are working in the field. As we will see, they can provide *interactive* experiences to the citizen scientist that leverages knowledge about the person's particular location (e.g., for on-the-spot training and for data collection). Its sensors can be used as part of scientific data collection (e.g., collecting images via its camera). Because these devices are also networked, information not on the device can be accessed and downloaded on demand, or information can be uploaded from the device to elsewhere. Because they are communication devices, a citizen scientist can potentially talk or text other citizen scientists and/or the scientist(s). As well, mobile device (such as the iPhone, the Windows phone, the Android phone and Blackberries) are increasingly common among the general public and citizen scientists, which means that people can use their own familiar devices as part of their citizen science activities (this also reduces project costs).

3.2 Data Collection

Data collection is the core of on-site citizen science projects. If citizen scientists cannot collect data, all other problems surrounding this type of citizen science is irrelevant. To make sure data collection operates, citizen scientists need to be able to find data collection sites and have the tools needed to perform data collection. Along with these needs, the ability for citizen scientists to create new sites greatly expands the possibilities for data collection by relieving scientists from performing this task. In this section I will discuss how geocaching and some mobility can support: how to find sites, how to create sites and how to transfer tools and samples between scientists and citizen scientists (Table 2.2, rows 1-3).

3.2.1 Finding Sites

Citizen science projects require people to find sites, as well as specific locations within that site. There is a plethora of ways to do this. For example, some projects use a *site description form* that records both general and specific information that could help others refind the site (e.g., the nearest town, the GPS coordinates, a description of the site relative to landmarks) and *site markers* (e.g., colored non-adhesive tape) to assure the citizen scientist that they have found the correct location (e.g., University of Alaska Fairbanks 2012). Other methods include paper maps and description of routes to the site.

Geocaching's universal approach to location finding is directly applicable to citizen science.

1. *Finding site descriptions.* Geocaches structure location information within a web site, where they are accessed via a conventional browser and (more recently) via a mobile device (perhaps running a specialized geocaching application). These sites are easy for geocaching members to browse, where no involvement is required by the geocache creator to disseminate that information. When applied to citizen science, citizen scientists can have an easy method to find sites of interest.
2. *Locating sites.* As illustrated in Figure 2.2 left, geocache site descriptions contain a plethora of information that describes and locates the site. GPS information is available in a downloadable form. If a mobile device interface for site information is available in a specialized application (e.g. c:geo 2013; Groundspeak 2013), it can take advantage of both its GPS and cellular networking capabilities, e.g., by providing detailed maps with the cache location and associated relevant information, as well as the location of the person relative to the site (as shown in Figure 2.2, right). This ease of use is important for allowing sites to be found again. Of course, some geocache descriptions are purposefully vague to create an enjoyable challenge, while other caches are made very easy to find (e.g., with easy hiding spots and verbose information). Yet within citizen science, any difficulty in finding a site will lead to less data being collected by volunteers (O'Hara 2008). Thus easy (vs. hard) geocache descriptions should be applied to citizen science, with the benefit that volunteers will have a standard description of sites, and an easy way to navigate to site location.

3. *Fine-tuning and verifying site location.* GPS coordinates are good for bringing geocachers within the vicinity of a site, but they are not accurate enough for finding an exact spot. Once within the vicinity, however, geocachers search for the cache container, which also marks the exact geocache spot. When applied to citizen science, cache containers can serve the same purpose as a citizen science site marker; both assure the finder that they have found the location. As well, additional visual information may be available on the mobile device for helping one find exact locations, such as detailed maps and text / images describing landmarks.

This research is not the first instance of applying geocaching to citizen science site finding. Project BudBurst is a citizen science project that uses aspects of cache finding (Han et al. 2011). In particular, its “Floracaches” are plant collection sites created in the vein of geocaching: sites are photographed, described and the GPS location recorded by administrators or citizen scientists for other citizen scientists to find. Somewhat like geocaching, Floracaches structure this information into challenge levels, each providing a different amount of information. Figure 3.1 left shows an easy level Floracache, with a provided map, photograph and description. The idea is to make Floracache-finding a fun challenge, although challenge levels can be changed to make sure the site is eventually located. Floracaches are also virtual, i.e., no physical cache containers are used.

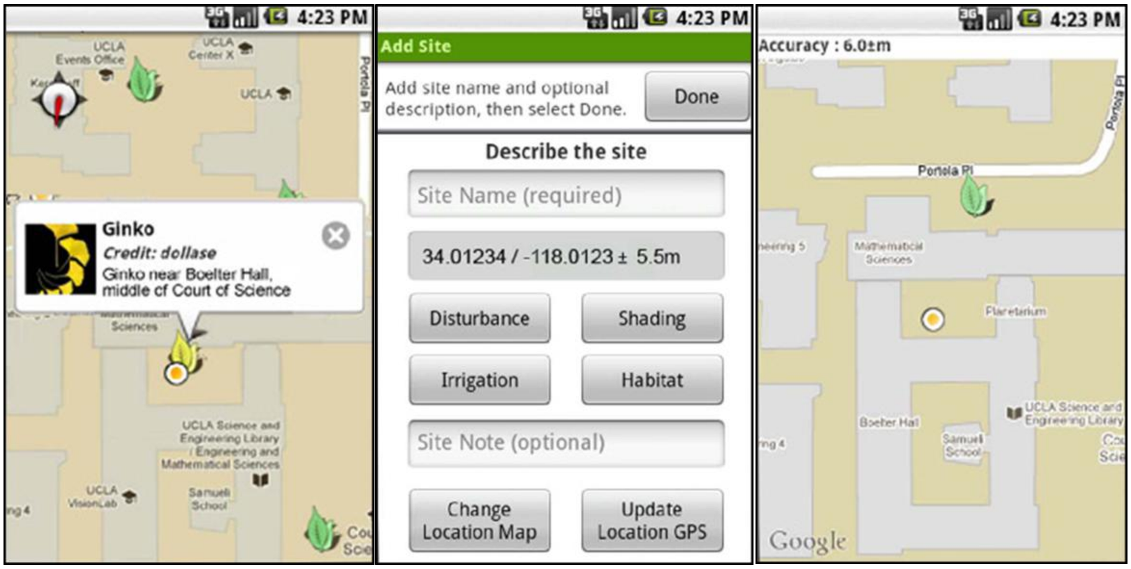

Figure 3.1: Floracaching screenshots. Left, an easy level Floracache, with a provided map, photograph and description. Middle, a Floracache cache creation screen, with automatic location information, site name and site details. Right, refining the position of a floracache by touching a map. The cache creation functionality shown was added as this thesis was being completed.

This application of geocaching to citizen science shows how verbose site descriptions and mobile device guidance can be used to find collection sites (Han et al. 2011).

3.2.2 Creating Sites

As mentioned, geocaches structure location information within a web site. The geocache site descriptions are usually easy to create (e.g., via form-filing), where any geocacher can log on and create a new site by providing a description about the cache they have deployed along with its GPS coordinates, which is then shared with the geocaching community. When this site creation mechanism is applied to citizen science, scientists can have an easy yet powerful method to create and share new citizen science data collection sites. Scientists can also use the idea of crowdsourcing to have citizen scientists find and create sites of interest.

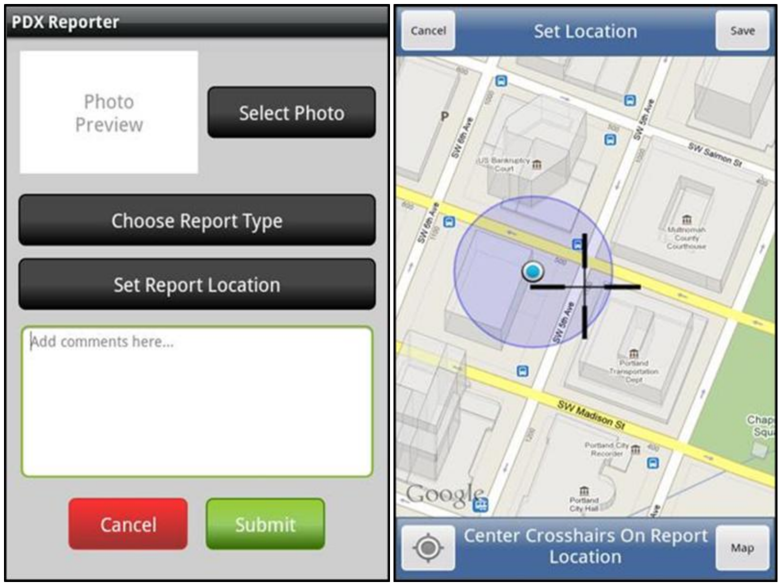

Figure 3.2: Screenshots of PDX Reporter's report creation pages. Left: the form for creating a report, with a photograph, description, report type and location. Right: The location setting page, which allows marking of the exact location of a problem on a map.

While geocaching supports site creation, this creation is not done at the site, which would be of benefit to citizen science. An example outside of geocaching suggests how this could be done. PDX Reporter (City of Portland 2013) is a mobile crowdsourcing application that illustrates how locations are marked by citizens. In particular, when a citizen comes across an infrastructure problem in the City of Portland, he or she can send a location and description of that problem to the local government. Figure 3.2 illustrates two screen snapshots of PDX Reporter. On the left screen, citizens take a photo of the problem (using the device's built-in camera), and fill in a form for describing the problem. On the right screen (accessed by the 'Set Report Location' button on the left screen), they are shown their approximate location as measured by the device's GPS unit, and can move the cross hairs to set the exact location of the problem. On completion, the information is automatically transmitted to the City using the device's networking functionality. This ease of creating a new report parallels with citizen science needs of creating and sharing sites. In this fashion, citizen scientists could create a site on their mobile device, providing photographs, a description, and GPS location to make finding the site as easy as possible.

Just as this thesis was completed, Project BudBurst's Floracaching implemented cache creation by citizen scientists (Project Budburst 2013). Their cache creation structure resembles report creation in PDX Reporter. The similarities between Floracaching in Figure 3.1 middle/right and PDX Reporter in Figure 3.2 can be seen. This concurrent development further emphasizes the relevance of cache creation by citizen scientists.

3.2.3 Transferring Tools, Data and Samples

As described in Section 2.2, geocache containers are an integral part of geocaching. They are an essential part of its 'hide and seek' gameplay, and they are used for various purposes. First, they typically hold logbooks that geocachers sign. These logbooks prove that a geocacher has been to a site, which in turn enriches the geocaching social experience (O'Hara 2008). Second, the container holds the treasure items that are hunted in geocaching. When they find a cache, they trade some of their treasure for the treasure inside the cache. A special case of this is the previously mentioned travel bugs, a cache item that has a goal to be moved by geocachers from one part of the world to another (Groundspeak 2013).

Citizen science already uses containers in some projects, such as for storing tools in the field when a site is visited regularly. However, this use is not widespread, is not used to mark the site where data is located, and is usually reserved for large items that are highly impractical to transport (e.g. shovels, tripods). Yet containers can be used more broadly. Other items can be held such as logbooks filled in by the citizen scientist during their visit, and/or smaller materials and tools required by the citizen scientist to do their data collection task. Citizen scientists can deploy these containers when creating a cache, reducing the work of scientists. The citizen scientist could also use the container as a place to store collected samples for later pickup (perhaps using a method similar to travel bugs for transportation).

3.3 Validation

As previously mentioned in Section 2.2.2, the value of data collected by citizen scientists is impacted by the validity of that data. Errors can occur for various reasons, such as typographical errors, incorrect identifications, incorrect or inaccurate measurements, problems with the collection protocol, and so on. Within citizen science, human analysis and/or automated methods are often used to detect invalid data (Section 2.2.2). Unfortunately, these methods are mostly done after the fact, i.e., after the citizen scientist has left the site and later enters data by (say) filling in a web page form. This means that there is little opportunity to repair errors by, for example, repeating and/or verifying information on the site itself. Compounding this, if another individual wants to repeat the collection at a later date, it is difficult to refind the site to do so. In this section, I will look towards how geocaching and citizen science applications can support mobile validation techniques (Table 2.2, rows four and five).

3.3.1 How Sites are Revisited

Repeat visits to a site allow multiple collections. This collected data can be validated and possibly corrected through averaging and singling out outliers for further analysis.

Within geocaching, Section 2.3 discussed how caches are visited by a large number of geocachers. When a geocacher visits a cache, he/she records the same pieces of information, both in the physical cache logbook and on the geocache's webpage. This information includes: the geocacher's name, their experience finding the cache and whether the cache was in good condition when found. The way geocachers go to the same site and record the same information can be used in site-based data collection for recording multiple readings at a site.

Geocaching's repeat visits and collections provide citizen science with different possibilities for data validation. As mentioned above, multiple recordings at the same site can be compared and averaged to gain a more trusted value. A citizen scientist can check and possibly repair data at a site where a previous record was marked (by scientists or a citizen scientist) as problematic. Also, when a citizen scientist collects data at a site, this data can be compared to previous recordings. If it is vastly different than the previous recordings, the citizen scientist can be asked to double-check their recording.

3.3.2 On-Site Data Comparison

The way that geocachers record the current state of a cache informs how citizen scientists, while on-site, may be able to reflect on data.

In geocaching, a geocacher will report if a geocache is damaged, the log book has run out of space or the cache GPS coordinates are off widely. When this happens, the geocache creator or another geocacher will fix the problem.

The way geocachers see and report problems with caches can be applied to the validation of citizen science data. If previous collections were made visible to citizen scientists while on-site, these collections could be checked and reported in a manner similar to the way geocachers maintain cache integrity.

While geocaching provides some ideas for on-site data validation, geocachers typically report cache problems off site, i.e. from their desktop computer after returning from the location. Due to this, other work must be found to inform how on-site validation can be performed with mobile devices. Outside of citizen science, Google Maps for iPhone (OS 6)(Google 2012) allows users to report map and route issues on the fly This shows how citizen scientists can respond to incorrect information in the field, but other ideas are needed to show how validation would work with science data and how a citizen scientist can change incorrect information in the field.

There is no prior work found in citizen science for performing on-site data validation. However, there is a mobile application for validating collected science information off-site. iSpot for Android (The Open University 2013) allows citizen scientists to validate nature identifications from their mobile device (Figure 3.3). The use of the mobile device is to allow validation from any location, rather than to assist validation on-site. Citizen Scientists post their wildlife identifications with photographs and descriptions to iSpot. Then, as shown in the left screenshot, other citizen scientists can access these identifications, in order to provide their own species identification for the wildlife sighting. In the middle screenshot, multiple possible species identification have been reported for the sighting (middle of the screenshot, under "Identifications") and discussion about the disagreement has taken place ending with an agreed upon identification (bottom, under "Comments"). In the right screenshot, details of the top identification are shown, which includes a comment about the choice of species and how sure the citizen scientist is about the identification.

The way that iSpot allows off-site validation is of interest to mobile on-site validation for a few reasons. First, its use of mobile devices for validation, providing photographs, maps, and other information inspires how on-site validation can be done. Second, iSpot's support of discussion and alternate identifications illustrates how validation can be performed on site. By including site-based interactions into this structure, on-site validation may be performed.

In general, the possibilities for on-site data comparison provide citizen science with techniques to improve validation. If automated techniques are used to compare a record to previous ones when it is collected on site, and the data is flagged as erroneous, the citizen scientist can check that data immediately and fix problems with the record. If the citizen scientist was provided with previous records when collecting data, the citizen scientist could review previous collections while on site, and compare them to their own.

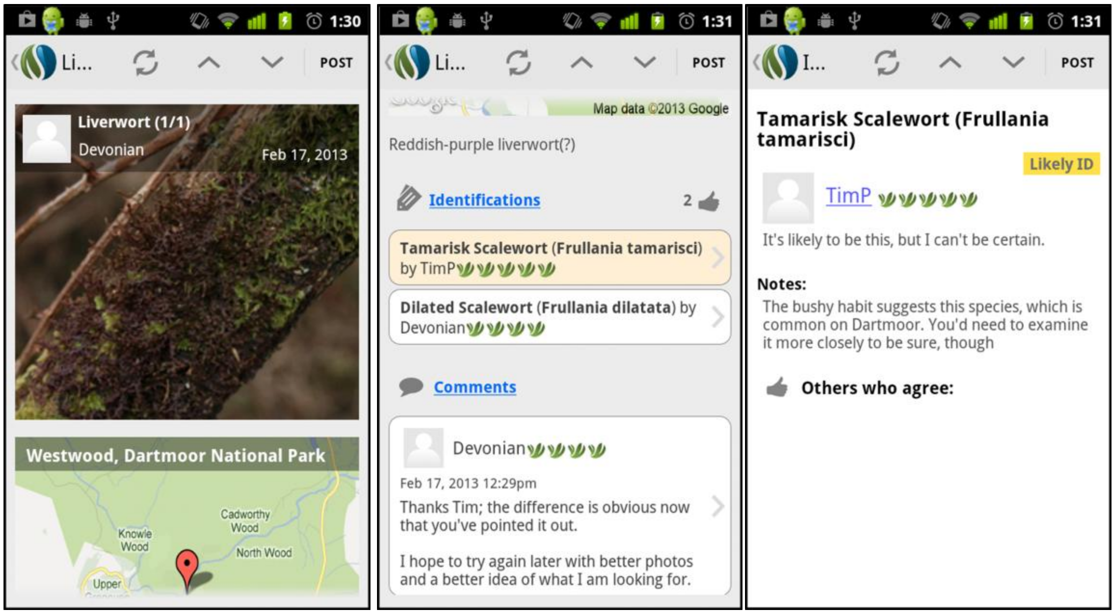

Figure 3.3: Screenshots of iSpot for Android's identification validation pages. Left: A citizen scientist can view the identification made by another citizen scientist. Center: Other citizen scientists can identify the wildlife sighting. Right: Details of an identification are show, including a comment and how sure the citizen scientist is about the identification. Species identifications can be discussed between citizen scientists through these pages.

If they see a discrepancy (e.g., in their own reading, or in the prior reading), they can try to fix the problem or manually mark the record as invalid or needing expert review.

3.4 Training

As described in Section 2.2.3, the training of citizen scientists is important for many reasons: to teach citizen scientists about the domain and to instruct them on data collection protocol. A well-trained citizen scientist minimizes data errors, which ultimately improves data quality. On-going education is also important, for example, to get citizen scientists to perform new tasks effectively and to increase their ability over time to collect different types of data. The training of citizen scientists can be performed *face-to-face* with a scientist in the field, or through *resources* such as pamphlets, videos and websites. In this section, I will look at how geocaching aspects, mobile device education and training in crowdsourcing can combine these two types of training to train citizen scientists through real-world examples, and testing on known values (Table 2.2, rows 6-7).

3.4.1 Real-world Examples

Geocaching is used in different ways to educate geocachers about the world around them. Many basic geocaches take geocachers to interesting locations and inform them about the location. *Multi-caches* and *EarthCaches* are geocache types that can require geocachers to learn about their environment to complete the cache. *Multi-caches* usually operate in stages, where the first cache must be found to gain information about the second cache, the second cache must be found to get information about the third cache, and so on. Often, multi-cache stages will have puzzles pertaining to the area around the geocacher to be answered for info on finding the next cache stage. As another example, *EarthCaches* have been specially created by Groundspeak to teach about earth features (Groundspeak 2013). Instead of finding a physical container, the cache is a geoscience feature, and to get credit for finding the "cache", the geocacher must answer questions about this earth feature. Geocaching has also been used in classrooms to educate about maps, technology and the environment (Schudiske 2012; Lo 2011). These different uses of geocaching show the potential of caching and cache sites to educate on real-world topics.

While geocaching provides interesting possibilities for real-world education, it does not generally use mobile devices to provide verbose information on the real-world experience. Still, another area that utilizes mobile devices for real-world interactive education comes from children's education with location-aware technology. Price and Rogers (2004) discuss using the physical world in education to actively engage learners. In one of their projects, Ambient Wood, children explore a digitally augmented forest via PDA with: RFID markers to trigger dissemination of plant information, probe tools to provide immediate feedback on light and moisture readings, and audio of animal sounds to teach ecology. Halpern et al.'s (2010) SunDial created three outdoor education sites at a children's science museum, where a mobile device lead families between these sites. Sites contain different activities: reading a sundial, identifying a leaf, experiencing the movement of water through dancing, and using the mobile device to ask site-related questions. For example, at the sundial, children were asked to tell the time by the sun and input it into the device. Their answers were recorded, but feedback was not provided (Halpern 2012). These two projects show the potential for real-world-integrated device interaction.

Geocaching's use of cache sites for education, in combination with the physical and mobile education ideas from Ambient Wood and Sundial, provide possibilities on how geocaching and mobility can be used to educate citizen scientists. They can be led to training caches that teach them about project-specific topics. The mobile device can lead the citizen scientist through education at the site, using the site itself. Important points at the site can be marked so the citizen scientist can better learn about them. Tools can be provided in the training caches to enhance the education experience.

3.4.2 Testing on Known Values

Along with real-world experience, it is important for the effectiveness of training to be tested and measured. Geocaching's multi-caches suggest one such approach for test-based education, where some can only be completed by solving puzzles on known information found around cache stages. While geocaching shows how learning can be required to complete real-world training, it does not account for possible benefits of having a mobile device on-site.

One non-geocaching example that suggests how a mobile device can be applied to geocache-based training is Duolingo. Duolingo is a mobile and web education system that teaches users a new language, while crowdsourcing web page translation (Hacker and Von Ahn 2011). The language lessons inside Duolingo give feedback to the user if they are having trouble with questions during a lesson. The system keeps track of the user's skill as they progress through lessons, and as they become more proficient in the language, the system provides the user with the ability to translate more complicated sentences.

Being able to track the training of citizen scientists provides more possibilities for the application of geocaching to training. Citizen scientists can be tested during training caches, helping ensure the education of the citizen scientist. Feedback can be provided to citizen scientists via mobile device during training, improving the testing process. Tools provided at training caches can be used in the quizzing process, allowing tests on tool use. Specific projects can be provided to citizen scientists upon completion of certain training caches, guaranteeing that those who take part have been trained.

3.5 Coordination

Coordination of a citizen science project is essential for the project to be successful. Scientists need to know the tasks that need to be done, what citizen scientists can do those tasks, and assign the needed tasks. Citizen scientists need to be provided with tasks to perform. This is problematic because keeping track of citizen scientists and tasks is extreme time intensive. In this section I discuss how geocaching, mobility and social technology can support: automated coordination and enriched real-time interaction (Table 2.2, rows 8 – 9).

3.5.1 Computer-Assisted Coordination

Geocaching and mobility can provide information needed to automatically assist coordination through automation in citizen science. Training and testing citizen scientists through geocaches allows for the collection of training results, which can control what tasks citizen scientists should perform at a later date. Data collected by citizen scientists at collection sites can be compared, allowing estimation of a citizen scientists' data collection accuracy and further informing what tasks they can perform. Citizen scientists can be asked to visit caches to perform different tasks (e.g. data collection, moving tools, data validation) depending on their identified skill, interests and location.

There are many non-geocaching systems that track their users and automatically tailor their experience depending on different factors. They use different attributes that define individual users, including: trust in a user, the user's ability, user location, and content preferences. I will now look at two systems that highlight the use of this knowledge.

One example is Stack Overflow, which quantifies its trust and perceived ability of a user to provide different abilities in its system. Stack Overflow is a community-based online question and answer site for programming questions (Stack Overflow 2013). Users gain reputation points by giving answers to the questions of others that are liked by the community. The more reputation a user has, the more power they have on Stack Overflow: they can ask questions with more details, and even moderate the site if they so choose. This tracking and structure allows Stack Overflow to provide and incentivize high quality responses to programming questions.

Coordinating where people should go can be customized to fit a persons location and preferences. For example, c:geo allows geocachers to search for geocaches around a location (Figure 3.4 left). This location can be the current one or a selected one of the geocacher's choice. The results of this search can be presented as geocaches sorted by one's closeness to the searched location (Figure 3.4 middle), or on a map showing the immediate area around the geocacher (Figure 3.4 right). By moving the map around, the location searched is changed, providing new content (c:geo 2013). Users can also specify the types of caches they are interested in. In Figure 3.4 right, the user has specified only puzzle caches (as can be seen by the puzzle piece icons). While c:geo does not gather information on a user, its ability to provide location and interest specific content lets

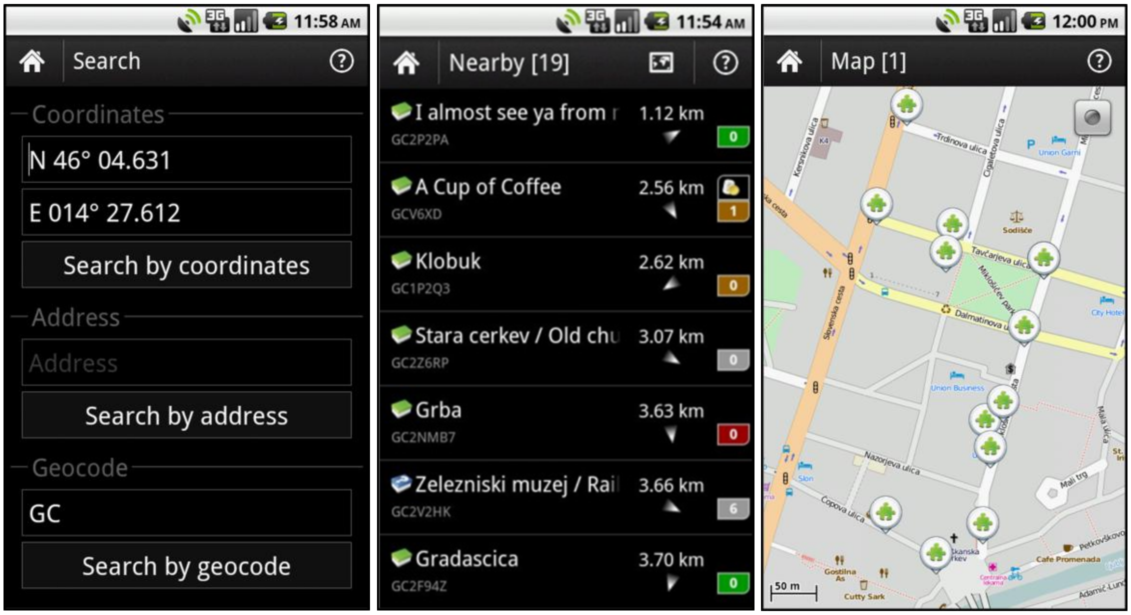

Figure 3.4: c:geo cache selection screenshots. Left: The geocacher can search by location. The current location can also be used. Middle: When searched, the caches are presented based upon how close they are to the geocacher. Right: The caches can also be viewed on a map. New caches can also be found by moving the map around.

geocachers find caches that they can reach and want to visit, increasing participation in geocaching.

The ways that these geocaching and related systems automatically track and use user information is of interest to citizen science. Citizen scientists can search for tasks they need to perform based on their location and the types of tasks they want to do. The tasks provided can be based upon the tasks the user has performed before and/or the trust the system has in the user based on training results and analysis of collections. The needs of scientists and projects can also be reflected in this map view emphasizing certain tasks. Scientists can choose the tasks that are important, or the system itself can determine this based upon programmed factors (e.g. how often the site needs to be visits, whether a record has been flagged).

3.5.2 Enriched Real-Time Interaction

It is just as important to facilitate communication between scientists and citizen scientists to allow manual coordination of citizen scientists in the field. This communication, when it takes place, is currently limited to common existing communication (e.g. phone, email). To my knowledge, no systems are designed to communicate information specific to citizen science projects, such as the state of a collection site or GPS locations. For example, consider how a map can be integrated into communication while a citizen scientist is on site. Scientists could view all the citizen scientists active in a project on the map in real-time, and send messages to particular citizen scientists as needed. Citizen scientists could also communicate back to the scientist through the map asking for help about a site or performing collection. Discussion could take place about map details and shared data, allowing both parties view information important to discussion.

3.6 Summary

In this chapter, I showed how geocaching and mobility can possibly be used to inspire new solutions for citizen science's problems with *data collection*, *data validation*, *volunteer training* and *volunteer coordination*. In the next chapter, I will discuss my application of geocaching and mobility to citizen science.

Chapter 4. **Design and Implementation**

4.1 Introduction

The previous chapter described, at a conceptual level, how aspects of geocaching can be applied to citizen science design. This chapter provides an example of how these concepts can be realized. In particular, I developed a prototype mobile device application called *Science Cacher* (to be used by citizen scientists), and a desktop application called *Scientist Controller* (to be used by the scientist). I then constructed scenarios surrounding the use of these prototypes in a citizen science context. The scenarios and prototypes serve to illustrate a specific solution to the problems mentioned in the previous chapter. They target a specific use case, that of coordinating and collecting different science data on quaking aspen trees. These scenarios and prototypes also serve as a working sketch to explain my ideas, which in turn will be used to solicit feedback and discussion from scientists, citizen scientists and project coordinators (Chapter 5).

The structure of this chapter is as follows. It begins with an overview of the scenario and prototype implementation used to extentiate the Science Caching concepts. The remainder of this chapter continues as a series of particular scenarios centered primarily on the four previously identified citizen science problems: *data collection, data validation, volunteer training* and *volunteer coordination*.

4.2 Overview

I now provide necessary background information to understand the scenarios and prototypes. I tell the basic story motivating all the scenarios, the technology used, and the way that the citizen science sites were constructed.

4.2.1 Scenario Overview

The two prototypes as well as the scenarios used in this chapter follow a specific type of citizen science project − data collection concerning quaking aspens trees located on various geographic sites. The choice of a specific (versus a generic) project was taken because it allowed for more detailed exploration of how geocaching can be applied to specific needs, and how the software could be designed to fit those needs. Quaking aspens data collection was chosen primarily for pragmatic reasons: these trees are readily available locally, and are stationary. This meant that test sites could be easily and rapidly constructed anywhere in the local region, which in turn meant that I could present and test the software with participants at locations convenient to them. Furthermore, both the sites and the actual tree data collection task were kept deliberately simple: this allowed the focus to remain on the system design and on how citizen scientists interacted with that design, rather than the particular intricacies and confounds that accompany a complex project.

The scenarios below involve several citizen science and scientist participants, various data collection sites, the Science Cacher mobile application, and the Scientist Controller desktop application. In the story, the scientist(s) have already made their quaking aspen sites available electronically, where particular sites and caches have been created ahead of time. In particular, scientist Jill has already located several physical caches within various city parks, and has marked and detailed these site locations. Jill has at her disposal several trained citizen scientists who know how to collect specific

Information	Purpose
Location (GPS coordinates)	Allowing data to be spatially plotted, refinding sites
Type of tree	Gives the tree type
Photograph of tree	Adds trust to data, refinding the tree
Description of tree	Adds trust to data, refinding the tree
Date of collection	Allows data to be temporally plotted
Living canopy cover	Allows the tracking of yearly tree cycles (tree phenology, which can be used to track environmental factors (temperature, rainfall)
Photograph of canopy	Adds trust to data, validation through photograph
Circumference of tree	Estimates age of the tree
Leaf samples	Tracking health, other uses
Photograph of samples	Allows further analysis by scientists without visiting the site

Table 4.1: Data collected in the Quaking Aspen Science Caching application.

information on trees located at those sites, as well as other volunteers who are yet to be trained. Table 4.1 lists the information her citizen scientists will collect (left column) as well as why that information is collected (right column). Using the Scientist Controller, Jill can also electronically access any data collected by her citizen scientists on their Science Cacher mobile devices.

Our focus begins with Arthur, a citizen scientist volunteer already subscribed to this project. Arthur is a 2nd year biology student and an amateur citizen scientist. He has been trained on some (but not all) quaking aspen data collection tasks, and has already used the Science Cacher to perform several data collections on his own.

4.2.2 Technology Overview and Key Replication Details

The digital interactions in this system take place through two networked custom software applications – the Science Cacher and Scientist Controller – running on particular technologies. Both are written in C# / WPF / Silverlight, and both access and store data on a centralized Microsoft SQL server.

1. The ***Science Cacher*** mobile application, designed for use by the citizen scientist, runs on a mobile device: currently a Nokia Lumina 900 Smart Phone running the Windows Phone 7.5 operating system. In this implementation, the application exploits the on-board GPS and camera sensors, as well as 3G mobile networking to send and receive interaction information with the central SQL server.

2. The ***Scientist Controller*** application, designed for use by the scientist, runs on a networked traditional desktop or laptop computer, where it is used (amongst other things) to coordinate volunteers and review collected data.

3. The ***Science Caching Server*** holds all the information used in the Science Caching system. As seen in the database diagram (Figure 4.1), this includes the site and cache description (the Cache, Figure 4.1 left), collected data about that site (Collection, middle top), descriptions about individual citizen scientists (Citizen_Scientist, right) and all text messages exchanged between citizen scientists and scientists (Message, bottom).

While the architecture of the system is relatively straightforward, a variety of development issues arose primarily because of limitations of the mobile device. While relatively routine to solve, a sampling is listed below to illustrate the kinds of issues that others may encounter when replicating these ideas[6].

- The Microsoft SQL server uses Windows Communication Foundation (WCF) to send and receive information from the mobile application. This was done because Windows Phone 7.5 does not support direct connections to an SQL server.

[6] Several off the shelf code fragments and images were also used during development, e.g., EXIF image rotation from http://timheuer.com/blog/archive/2010/09/23/working-with-pictures-in-camera-tasks-in-windows-phone-7-orientation-rotation.aspx, and the map icons from http://mapicons.nicolasmollet.com/

Figure 4.1: Implementation Diagram of the Science Caching Server.

- To normally serve WCF content, a full ASP.NET/WCF server is needed, as the testing server does not allow traffic outside the local network. For prototyping the mobile application, the local traffic restriction was bypassed via TcpTrace[7], which tunnelled the traffic from an externally accessible port to the WCF server (this page was used as a guide[8]).

[7] http://www.pocketsoap.com/tcptrace/
[8] http://stackoverflow.com/q/18918

4.2.3 Site Overview

In our scenarios, citizen science interactions primarily occur (and were demonstrated as a walkthrough in Chapter 5) around a series of real physical cache sites created specifically to collect data on quaking aspens. Each site has a cache container containing different objects, including: tape measures, phenological guides, leaf sampling logbooks, a pen, scissors and clear tape. These objects are used by the citizen scientist both during data collection and training. Caches are sandwich-sized plastic food storage containers (e.g. Tupperware). Some sites use other markers made from post-it notes and thumbtacks in lieu of caches to designate multiple locations, as will be discussed shortly.

4.3 Scenario 1: Collecting Tree Data

Our first scenario shows the most basic element of Science Caching, the 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 data collection. This takes advantage of elements of geocaching, namely the process geocachers use to 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 Cacher, which assists in the choosing, finding and collection process.

We follow Arthur, the trained citizen scientist. At this particular moment, Arthur has some extra time on his way home from school that he can devote to the Quaking Aspen project. To see if there are any nearby quaking aspen sites that he can visit, Arthur starts the Science Cacher on his personal Windows Phone.

Figure 4.2: Screenshots for choosing a collection site. Left: Arthur sees himself and nearby sites on the map. Right: after Arthur clicks on a site, information is provided about its project.

4.3.1 Finding Sites

Arthur's first task is to find a suitable site, and then go to it. To do this, Arthur must first *choose a collection site*, then *locate and move towards the site*, and finally *fine-tune and verify the site location*. This site-finding scenario mirrors the steps outlined in Section 2.2.1.

First, Arthur has to *choose a collection site*. Upon loading, the Science Cacher shows Arthur a composite satellite and schematic map, as well as various icons (Figure 4.2 left). He can see his location on the map (the icon of the human figure at the middle left), as well as nearby collection sites represented as an icon that indicates the project type. In Figure 4.2, Arthur sees one water-monitoring site (the icon of water waves), two duck monitoring sites (the duck icons) and a single tree-monitoring site (the tree icon)(some icons are semi-transparent, this will be discussed later). If Arthur wanted to see a broader area (possibly containing more sites) he could zoom out, or pan to other regions. In this case, because he is trained in tree monitoring and because the site is nearby, Arthur taps the tree icon (Figure 4.2 left, the tapping finger) and more information is revealed (the task description below the map in Figure 4.2 right). He verifies from the task description that this site is for “recording a variety of details about a quaking aspen”. Arthur already knows how to do this, so he starts the project by tapping the “Start this Project’ button on the bottom of the screen.

Arthur’s next step is to *locate and move to the site.* To guide himself to the cache container, Arthur uses the map as well as the navigation page that appeared immediately after he started the project, (Figure 4.3 left). Similar to GPS systems, this page provides him with a somewhat zoomed in map for fine-grained navigation (although he can pan and zoom as desired). The map automatically updates his position (the figure icon) as he moves around, as well where he is relative to the cache GPS location (the icon at the top right). It also adds extra information in the navigation pane (Figure 4.3 left, bottom half). The text on the bottom right informs him of his distance from the cache (currently 128 meters), as well as the appearance of the physical cache (i.e., a black Tupperware container behind a particular tree). The photo at the bottom left is a picture of the actual cache location which he can enlarge via double-clicking, where the display would then appear as in Figure 4.3 right. Using all this information makes it easy for Arthur to locate and navigate to the site.

When Arthur has reached the site, he must find the cache container to *fine-tune and verify the site location*. By using the information supplied in the navigation pane and by visually (and physically) searching the area, Arthur finds the cache container (Figure 4.4 left), which verifies that he is at the exact correct spot for collection. He presses the “I

Figure 4.3: Screenshots for finding the collection cache. Left, the citizen scientist is provided with a navigation page to find the cache that marks the site. This page gives him details to find the cache, a map and his distance from the cache. Right, the photograph of the cache site is enlarged by double-clicking for a better view.

found the cache" button at the bottom of the navigation screen (Figure 4.3 left) to continue. The screen changes to a data collection screen, illustrated in Figure 4.5.

4.3.2 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 clear (Figure 4.4 right). His mobile device presents him with the data collection form for the site (Figure 4.5 left; also see Table 4.1). Jill (the scientist) had already set up the cache at a quaking aspen tree, where she had entered the tree type into the system. Even so, Question 1 asks Arthur to verify

Figure 4.4: Left, Arthur finding a cache container. Right, the inside of the cache, containing collection tools: a tape measure, scissors, a pen, tape and a leaf sample book.

the type of tree, which is asked to make sure that Arthur both remembers his basic knowledge of tree types, and that he is in fact at the right tree type (e.g., if the cache was inadvertently moved). Question 2 asks Arthur what percentage of the tree canopy has leaves. This data will be used by the project to help understand the yearly cycle of the tree, and to track environmental trends affecting the tree over time (e.g. global warming, particular weather events, local conditions such as air pollutants). Arthur looks up at the canopy and sees that there is no cover, and records that as <5%. Along with the canopy recording, he uses the Science Cacher and the camera on the mobile device to take a photograph of the canopy (Figure 4.6 left). This image can be reviewed later by Jill to both verify Arthur's reading, and to increase her trust in his data. Question 3 asks Arthur to measure the tree's circumference. He takes the tape measure out of the cache and uses

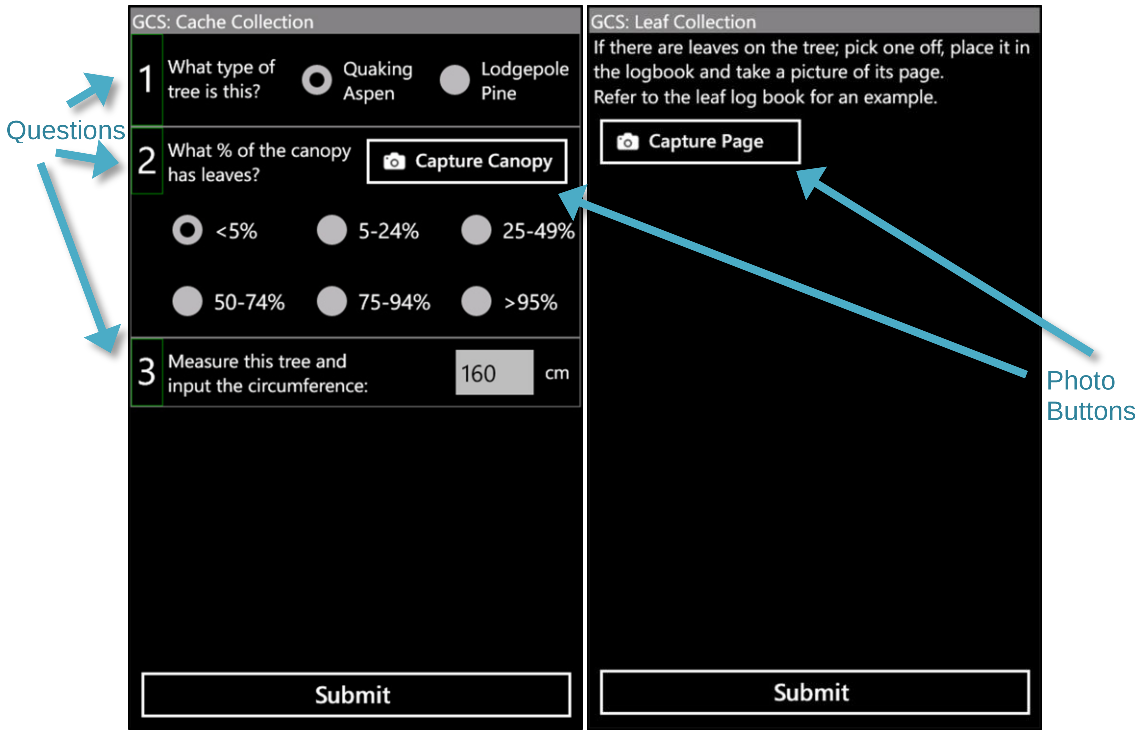

Figure 4.5: Data collection screenshots. Left, Arthur's filled out collection form, which asked him for the tree type, canopy cover, canopy photograph and tree circumference. Right, the sampling page, asking for a leaf sample to be put in the leaf logbook, along with a photograph.

it for a tree measurement (Figure 4.6 right). 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 4.5 left).

After submitting that data, Arthur is then asked to place a leaf sample in the leaf book and take a picture of it (Figure 4.5 right). He takes a leaf off the tree, takes the book out of the cache and tapes the leaf into it, and writes his name by the collected sample. He then takes a picture of the leaf and submits it.

In all these stages, whenever Arthur clicks the submit button (such as the one in Figure 4.5 left), the mobile device will try to send data to the Science Caching server (i.e., the centralized database), where it will update the records shown in Table 4.1. While the

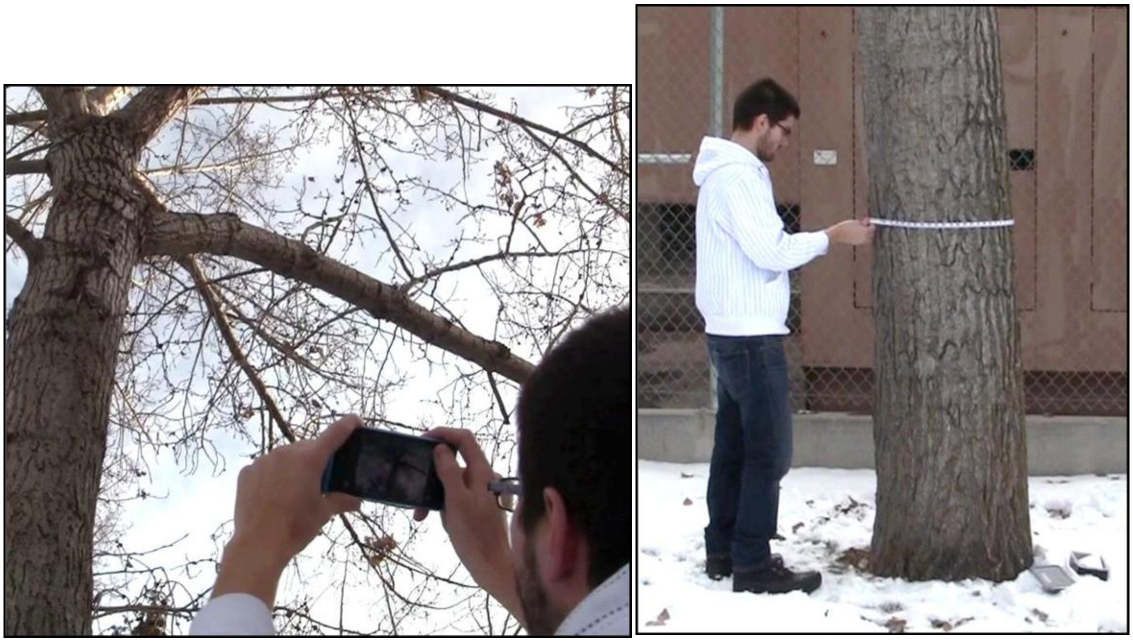

Figure 4.6: Left, Arthur takes a photo of the canopy. Right, Arthur measures the tree circumference.

prototype assumes that the network connection is always available (e.g., through cellular data), I would expect a robust system would work offline as well, where it would store the data and submit it when the network becomes available.

Arthur closes the cache container, places it back under the tree, and heads home. The entire sequence only took about 15 minutes, and Arthur is pleased that he was able to contribute to this project.

4.4 Scenario 2: Creating Collection Points

In the second scenario, we show how a citizen scientist can create *collection points* – physically marked locations around a cache container – and collect data around those points. These collection locations are very similar to sites, the only difference being that:

a) sites have cache containers marking them while collection points have labels, and

b) collection points are located in a small area around a site, where data is collected easily at multiple points during a single visit, (in contrast, caches are typically too far apart to allow this).

By locating these collection points in the vicinity of a cache container, the citizen scientist can still access needed tools and other resources. Collection point creation is highly similar to site creation. For pragmatic reasons, I demonstrated collection point creation to the participants (see Chapter 4) because collection point creation is much quicker than building a cache and moving to a distant location for deployment.

Our scenario continues on the next day. Arthur is heading home from school, and again starts the Science Cacher. A new tree data site appears on his cache browser (see Figure 4.7 left). He clicks it, and the text at the bottom of the figure informs him that this site is for collecting at different quaking aspens around a central cache (which he knows means that it should have multiple collection points). He decides to participate and heads towards it.

Similar to the first scenario, Arthur locates the site, then finds and opens the cache. In addition to the normal tools (as in Figure 4.4 right), the cache also contains thumbtacks and post-it notes. Arthur will find other quaking aspen trees nearby the cache, and mark them by thumbtacking a numbered post-it note to the tree (Figure 4.8). These marked trees will then become new *collection points*, where data will be collected on those marked trees. Arthur’s role as a citizen scientist is thus to develop the site further, where he – instead of Scientist Jill – will create numerous collection points (i.e., trees to capture data on) around the central cache. In this case, while Jill the Scientist marked the site, she saves time by leaving it to her citizen scientists to do the actual site development. Additionally, if other collection points had been created previously (say by other citizen scientists), Arthur can also collect data on those tree at those particular points.

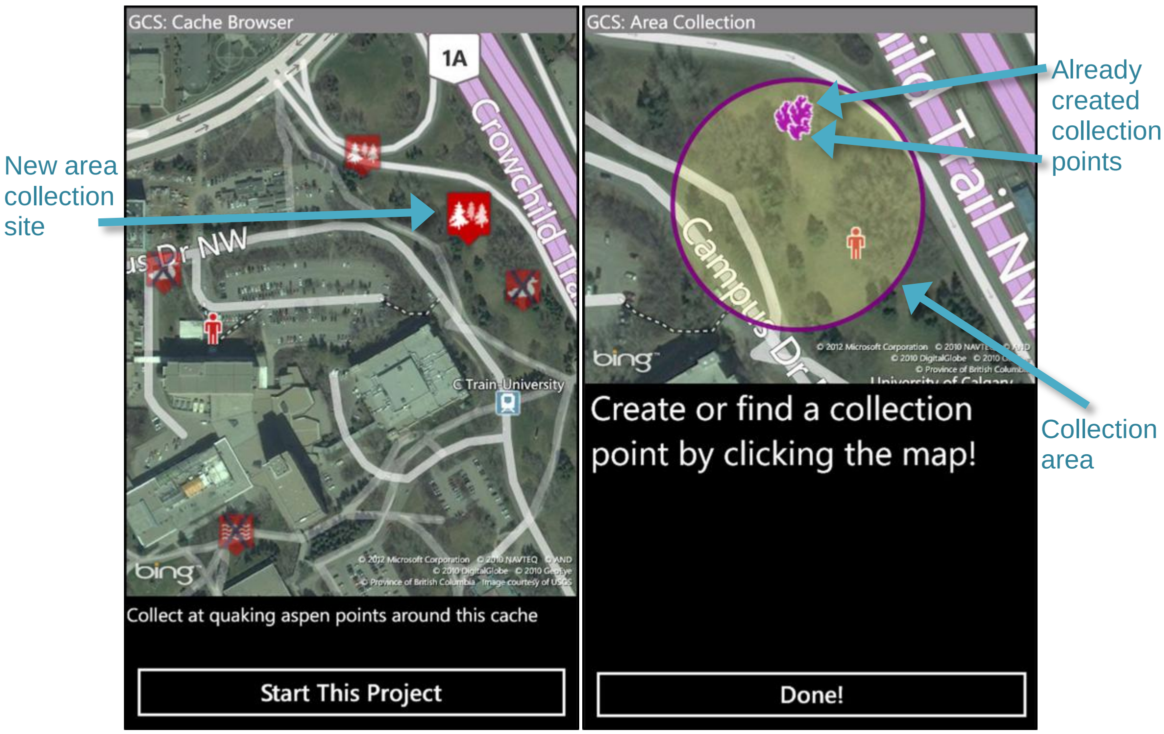

Figure 4.7: Left, Arthur has access to a new collection site, finding and collecting data on quaking aspens around a central cache. Right, the area collection point page (zoomed out), showing previous area collections. On this page, Arthur can use the map to create a new collection or click a previous one to find it.

The mechanics of creating these collection points work as follows. The map on Arthur's mobile device shows a circle, whose center is the central cache's location. Arthur knows that this circle specifies the area defining where he can place collections points and/or see existing collection points within it (see Figure 4.7 right). The application informs Arthur that he can either create a new collection point or find and collect data at one made by another citizen scientist to collect data at. Because finding and collecting data on a previously made collection point is highly similar to what was described in Scenario 1 – except that the citizen scientist looks for a marker like as in Figure 4.8 instead of a cache container – this will not be discussed further.

Figure 4.8: A tree collection point marked with a thumbtack and post-it note

Arthur decides to create a new collection point. He wanders around the area shown on his device (where he sees by his icon location that he is still within it) and spots an unmarked quaking aspen tree. The application guides Arthur through the steps he must take: mark the exact location on a map, take a picture of the location, use a physical marker to tag the tree, briefly describe that tree, and provide its marker number (Figure 4.9 left). He walks to the chosen tree (Figure 4.10 left) and tags it with a thumbtack and numbered post-it note (Figure 4.10 middle). While his GPS position is known to the device, it is approximate (i.e. with a typical error of 5-10 meters). Thus he specifies his exact position as he sees it on the map. Using the Science Cacher, he zooms in to the marked collection area and taps the exact point on the map where the tree is located (Figure 4.9 left & Figure 4.10 right). He then takes a photograph of the tree and describes it in the text field (Figure 4.9 right). With this done, he submits the location. Once it is submitted, he can then perform data collection as in Scenario 1 (see Figure 4.5 left).

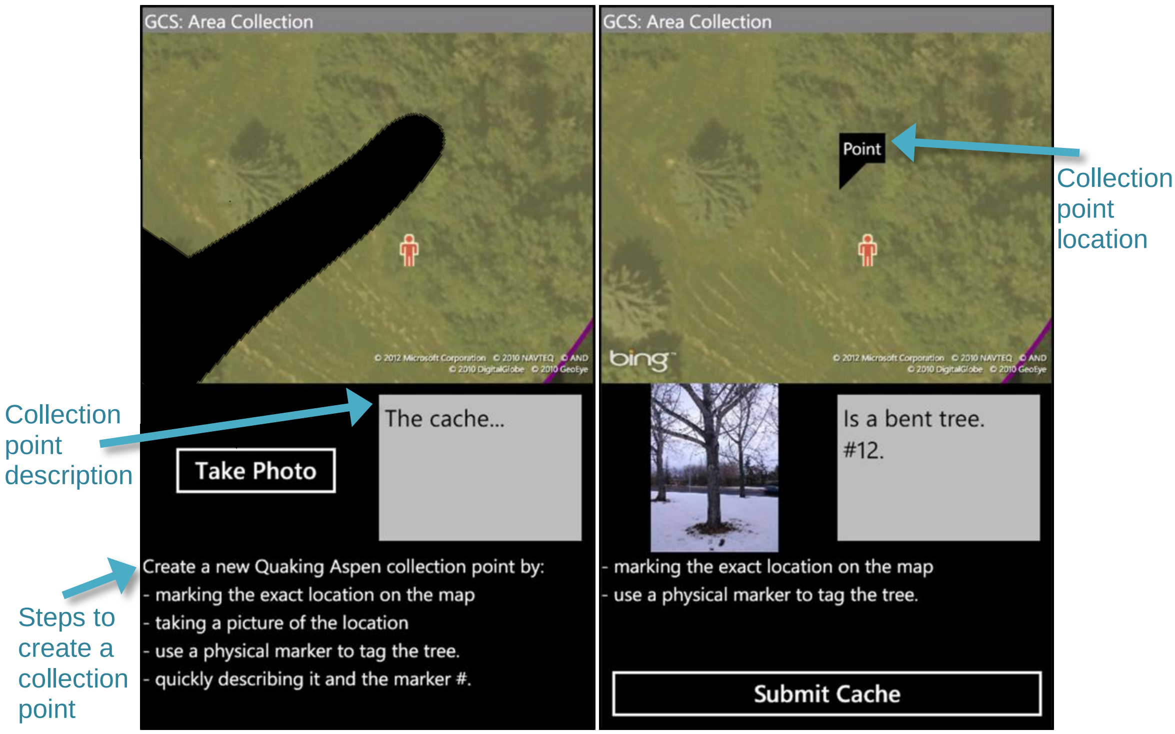

Figure 4.9: Collection point creation screenshots. Left, Arthur taps the map to create a collection point. Right, Arthur has gone through all the steps to create a collection point.

Arthur could also create a cache site using a method (and interaction sequence) highly similar to the above scenario. The primary difference is that Jill would have specified a broad set of areas as potential cache sites, and would have provided full cache containers to Arthur for deployment. Alternately, Arthur could have created the cache container himself (a task that geocachers commonly perform), and – depending on the project – have found potential sites entirely on his own.

4.5 Scenario 3: Data Validation

This next scenario illustrates how data validation is afforded by the system. We follow Steve, another citizen scientist, who happened to choose the same tree collection cache that Arthur has recently collected at in Scenario 1. Steve is new to this process, and

Figure 4.10: Arthur tagging the collection location. Left, Arthur walks towards the tree. Middle, he tags the tree with a post-it and thumbtack. Right, he marks the tree on his mobile device.

unfortunately makes several errors during data collection. Fortunately, the Science Cacher is able to use Arthur's previous collection data to detect these problems, and to guide Steve to check his measures and (ideally) correct any mistakes while still on site. Later, Jill uses the Scientist Controller to examine the collected data for errors or discrepancies, where she will decide what data to keep and what to discard.

4.5.1 Erroneous Data Collection

Steve chooses and navigates to the site as in Scenario 1. He then performs the same data collection as Arthur. Unlike Arthur, he has let his tape measure sag (Figure 4.11 top left), leading to an incorrect measurement of the circumference as 165cm (Figure 4.11 top right). He also incorrectly includes the dead leaves in the canopy count as canopy cover, and thus records the count as 25-49% (Figure 4.11 top right).

When Steve submits his data, the values that conflict with Arthur's previous record are shown (Figure 4.11 bottom left). As part of his training, he was told to double check his measurements whenever a conflict is noted. He also knows to look at the photograph taken by the previous citizen scientist (Figure 4.11 bottom left) to see if the photograph matches with the site (in case Steve or the previous citizen scientist measured the wrong tree). Steve begins by retrieving the collecting guide manual located inside the

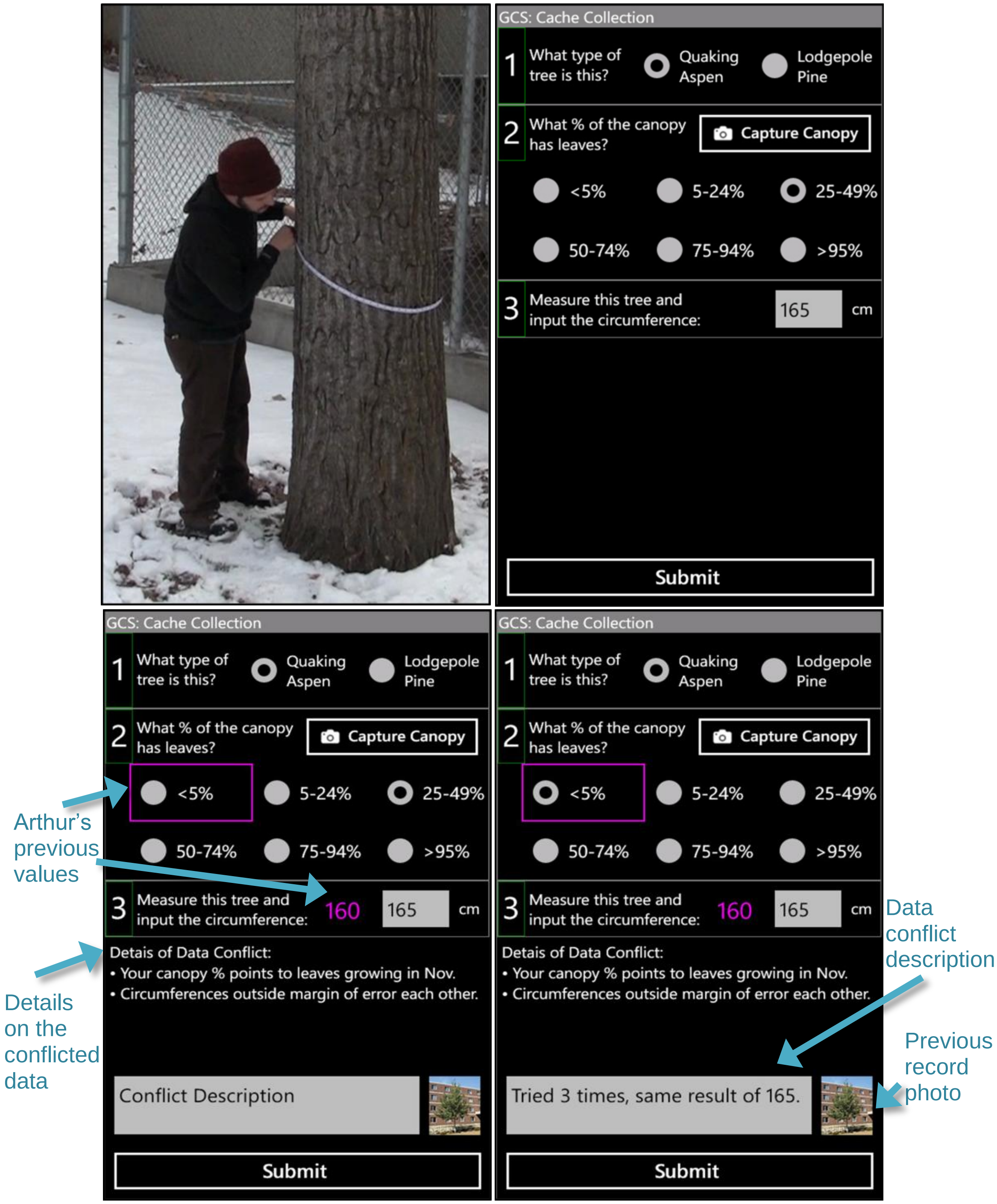

Figure 4.11: The collection and validation of conflicting data. Top left, Steve measures a tree incorrectly. Top right, Steve records his collection at the quaking aspen site. Bottom left, Steve's record conflicts with the data Arthur previously entered. Bottom right, after Steve double checks his recordings, he can't figure out if his tree circumference measurement is incorrect, or if the previous record is, so he leaves the value as is and writes a comment.

Figure 4.12: Left, citizen scientist Steve pulls the collecting guidebook out of the cache. Right, Steve refers to the guidebook to see why his canopy cover measurement is possibly invalid.

cache (Figure 4.12), and reads to make sure he is performing the collection correctly. He reads that dead leaves should not be recorded as canopy, and thus changes his answer to question 2 (Figure 4.11 bottom right). 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 (Figure 4.11 bottom right), saying “Tried this 3 times, same result of 165”. 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 that needs to be checked (perhaps by the scientist, or a third citizen scientist).

4.5.2 How the Scientist Deals with Errors

As described in Chapter 2, there are known methods for the Scientist to detect and deal with errors. In this case, there are redundant measures of the same tree (i.e., the data collected by Arthur and perhaps others). Scientist Jill can, for example, average the records (which somewhat minimizes the error effect). Alternately, she can inspect the data and throw out obvious outliers.

In addition to these known methods, she has redundant information about the site that was supplied by the citizen scientists, in this case photos associated with the data, where those photos were taken by the citizen scientists through the Science Cacher. If she is suspicious of the data (or perhaps she just checks as a matter of routine), she can view the photos supplied by the citizen scientists against the data. In the previous example, if Steve had not figured out the problem with his canopy measurement, Jill could compare that measurement to the taken photograph. Based on that photo, she could correct the data. In a more complex situation, she could also have used her Scientist Controller Application to mark the site as one requiring a further visit by a different citizen scientist, which eventually would have given her more data to examine.

Returning to the scenario, several days later, Arthur, Steve and other citizen scientists have now collected data on the site mentioned above. Jill loads the Scientist Controller to inspect that site's data. She goes to the collections window (Figure 4.13) to view the graph of the circumference measurements. She immediately sees the larger red point, meaning that it has been flagged as a possible error. She clicks the record to view its details: its data appears in the bottom right panes. Looking at the graph, she sees that three citizen scientists recorded the tree with the same measurement of 160cm. Steve's explanation is not that helpful, but she knows (from looking up his description as held by the Science Caching Server) that as a new citizen scientist he has likely made a process error when measuring the tree. She decides to discard Steve's data by pushing the "delete record" button on the right detail pane. She also sends Steve a message asking him to redo his training on tree measurement.

In summary, the above scenarios illustrate how a citizen scientist can potentially catch and repair collected data while on site, and how a scientist can deal with any errors left in the record. There are ways to improve this example, which were not shown. First, the Science Caching system could automatically analyze multiple previous records when comparing collections, improving the ability to detect outliers in collections. Second, the trust the system has for the citizen scientist can be taken into account when deciding which data are likely outliers (discussed in depth in section 4.7). Citizen scientists who have performed more trustworthy collections in the past could be weighed higher when evaluating which data point is likely correct or incorrect. Finally, if there is a data conflict at a site, the system could prioritize that site over others when presenting sites to citizen scientists, in order to get more readings. This is also discussed in depth in section 4.7.

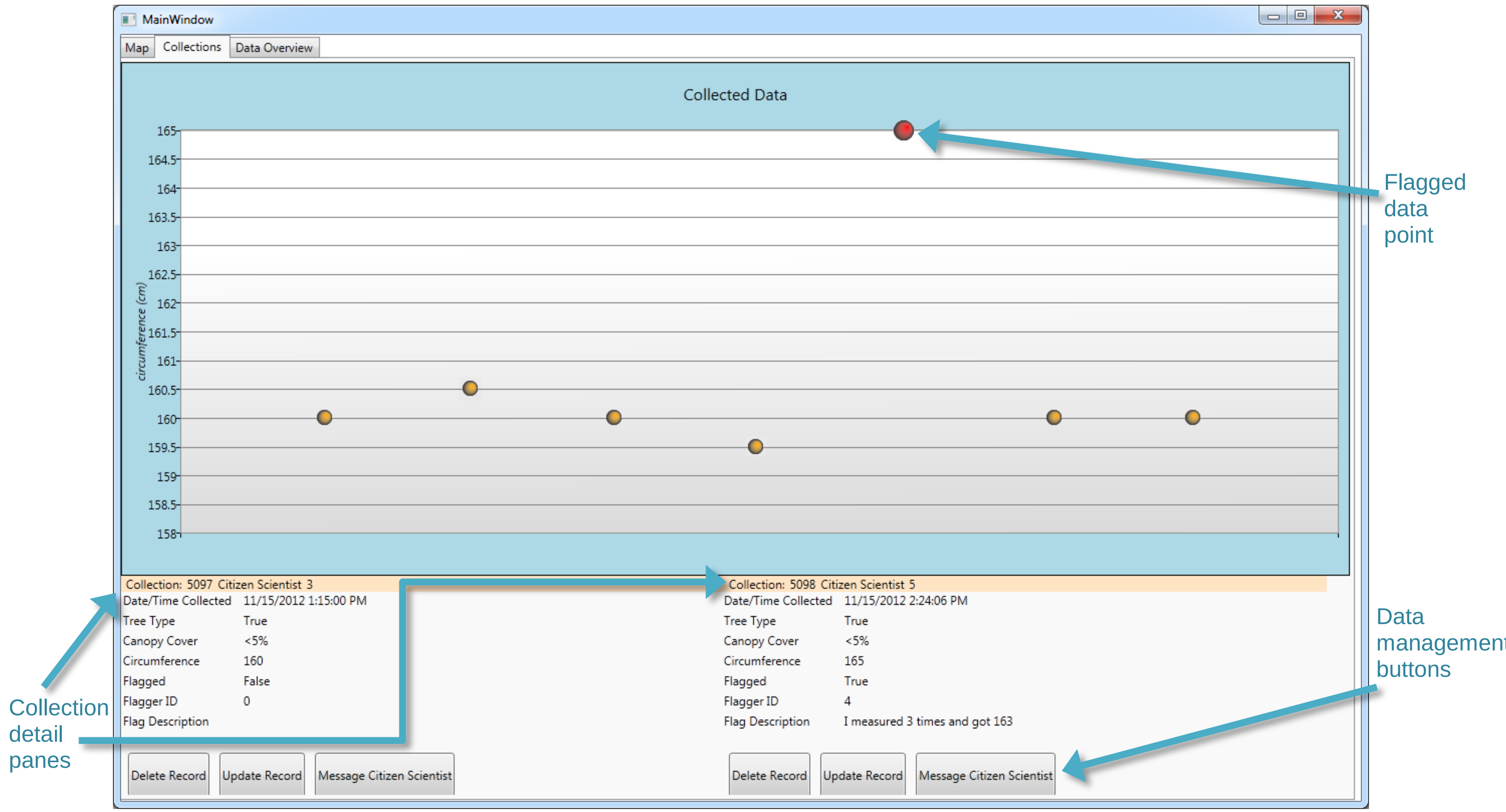

Figure 4.13: Scientist controller data collection viewing page. On this page, data collected by different citizen scientist can be viewed, details of that data inspected. The data can be deleted, changed, or the citizen scientist who collected it contacted

4.6 Scenario 4: Volunteer Training

The next scenario shows how the system can help 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 experience makes the training immersive, potentially more effective, while still at a much lower cost (especially in terms of a Scientist's time for teaching) than in a classroom setting. We follow Nolan, a brand new citizen scientist, who has just signed up but has minimal training in any data collection tasks. He has reviewed the project's web site, which explains what the project is about. He has also gone through several basic tutorials common to all sites (e.g., downloading software, how one navigates to sites, etc.). However, these tutorials did not go into detail about actual data collection methods. Nolan happens to be Arthur's twin, which is why he resembles him in the figures below.

Nolan downloads and then starts the Science Caching application on his phone. Since he is a new citizen scientist, the system only allows him to visit training caches. While other sites are visible, their icons are faded and marked with an X and are not selectable: only the nearest selectable training icon is shown (Figure 4.14 left). Steve clicks on this training cache, where it is described as a site for learning how to collect information on quaking aspens (Figure 4.14 right, text at the bottom). He decides to take part and clicks the "Start This Project" button.

4.6.1 Training in the Real-World

When Nolan arrives at the training site (navigating there is the same as in Scenario 1), he finds the cache container, opens it, and sees that it contains a tape measure and a booklet on quaking aspens (not shown). His mobile device then provides step by step training on how to identify the tree at that location (Figure 4.15 left). It informs his that the tree next to the cache is a quaking aspen (text at top), and points out the aspects of the tree that are unique and usable for identification. Nolan looks at the quaking aspen by the cache and verifies its identifying characteristics. The mobile device also provides him with

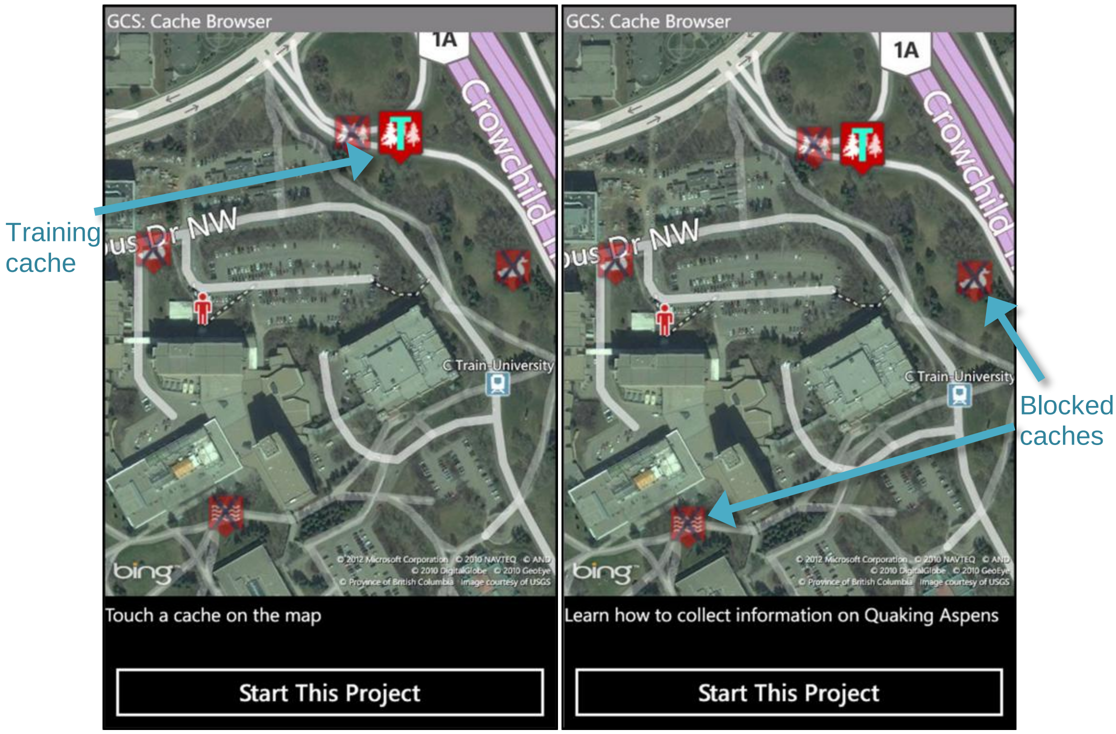

Figure 4.14: Training site selection screenshots. Left, Nolan the new citizen scientist and only has access to the training site. Right, he clicks that training site and sees it is to learn about collecting on quaking aspens.

photographs of the quaking aspen's leaves and bark, showing him characteristics that might not be present today (e.g. as the leaves are dead in winter, or that the tree is young and hasn't yet grown its fissured bark).

After Nolan has read about identifying quaking aspens, he clicks the "Tell me more!" button. This next step teaches him how to measure tree circumferences. He is asked to get his tape measure ready (Figure 4.15 right, text at top), which he does. His mobile device tells him what height to measure the tree at (130cm), and provides him with a picture of how to do it (bottom left). The site has been previously prepared for teaching, where four thumbtacks have been placed on that tree at 130cm for him to get a feel of the correct height.

Nolan is then asked to practice measuring on the cache tree. Using the tape measure, he measures the tree at the correct height, using the thumbtacks as a guide (Figure 4.16 left). He measures 132 cm, and checks the measurement with his mobile

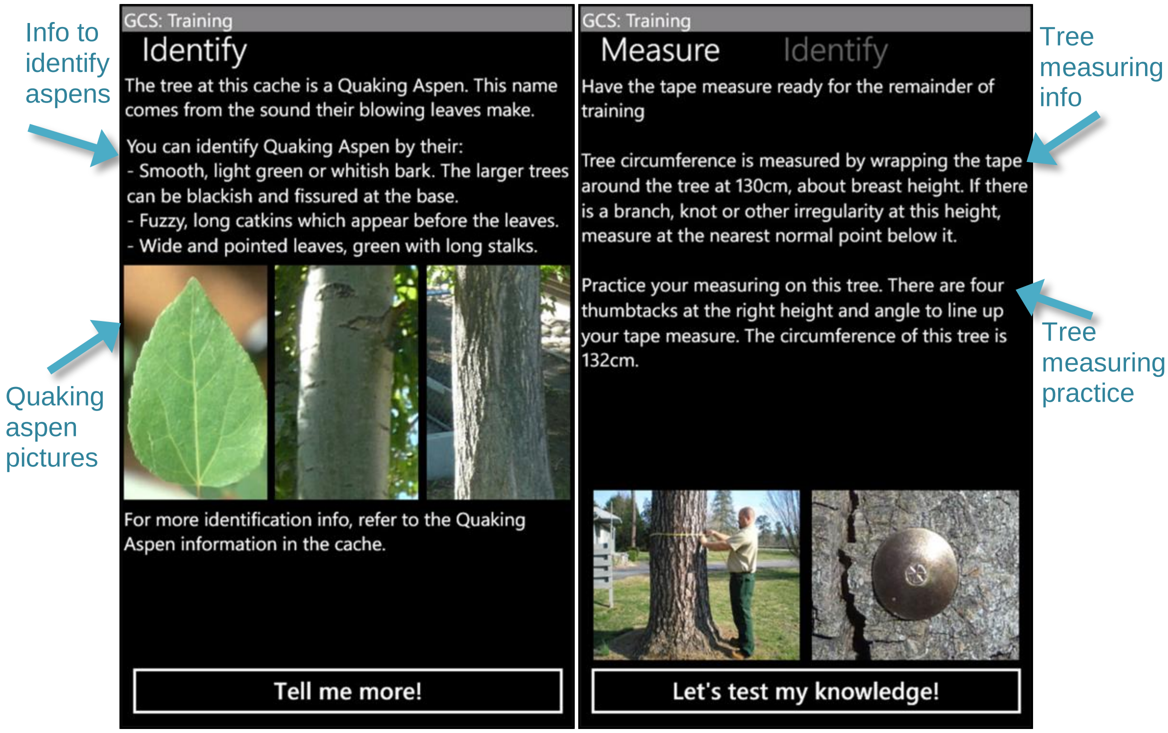

Figure 4.15: Screenshots on learning to identify and measure quaking aspens. Left, information is provided to identify quaking aspens, including descriptions and photographs. Right, information is provided on tree measurement, giving steps on how to measure. The citizen scientist is asked to practice on the training tree.

device (Figure 4.16 right). The values match, so he proceeds to the next stage of the training.

4.6.2 Testing on Known Values

The next part of the training tests Nolan on-site about the information he has just learned. First, the application asks him to look to the right of the cache, where there are three trees. It asks him to identify which one is a quaking aspen (Figure 4.17 bottom). He examines the trees and refers to the repeated training information on the question page. With this information, he notices that the tree to the right (Tree 3) has white, papery bark, unlike quaking aspens. The other trees have bark similar to the tree he was trained on, so on his mobile device he clicks trees 1 and 2 (Figure 4.17 top right, green squares) as quaking aspens. He then selects “Check Answer”. The application informs his that he is correct, and that the other tree was a paper birch (Figure 4.17 top right). If he had

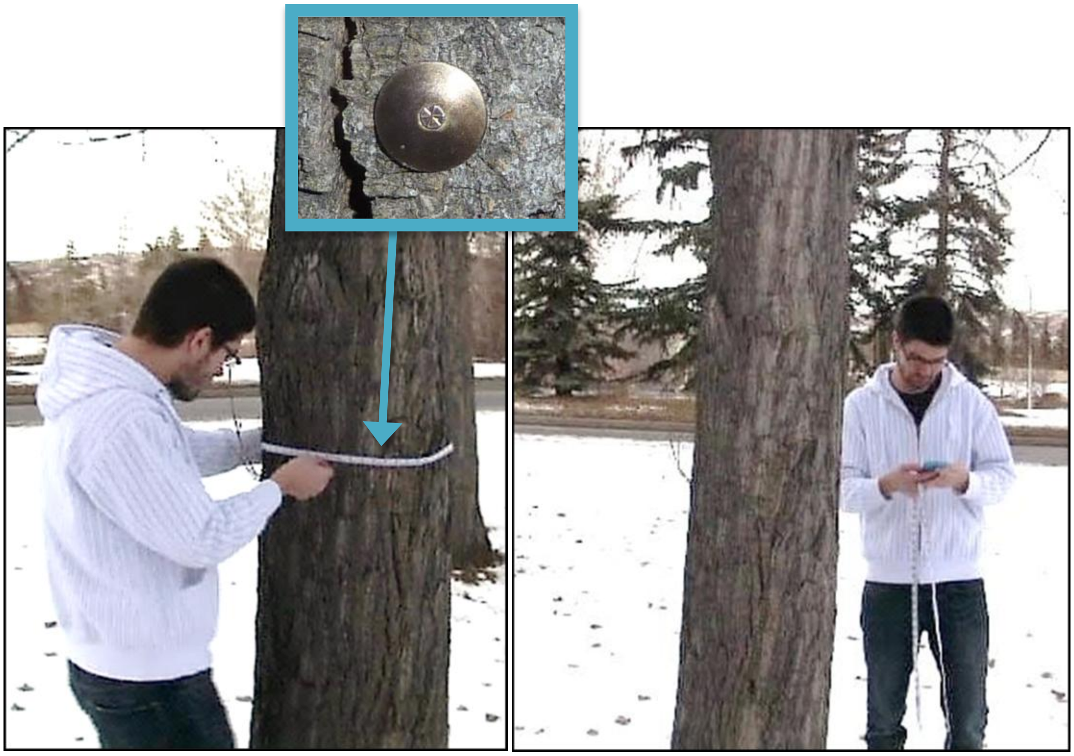

Figure 4.16: Nolan practicing tree measurement. Left, Nolan uses the thumbtacks get the correct tape measure height and alignment. Right, Nolan checks his measurement with the application.

answered the question wrongly, the application would have provided him with further hints about identifying the tree type.

After being tested on his tree identification skills, Nolan is now ready for the measurement testing. Of the three trees used in the identification question, Nolan is asked measure the left one's circumference (Figure 4.18 left). Nolan refers to the abridged training information on the device and walks to the tree. He measures 130 centimetres off the ground (Figure 4.18 middle), and measures the circumference (Figure 4.18 right). He measures 134 cm and enters it into the application (Figure 4.18 left). The correct answer is 133 +/-1cm, so he is informed his that the diameter is correct, and that it is close to the training tree's circumference because they were planted at the same time. If Nolan had entered the wrong answer, the application would have told him that the measurement was too high or low and guided him on how to redo his measures.

Nolan has finished answering the quiz questions and has completed the quaking aspen training. He is ready to take part in real quaking aspen project measurements, as in Scenario 1.

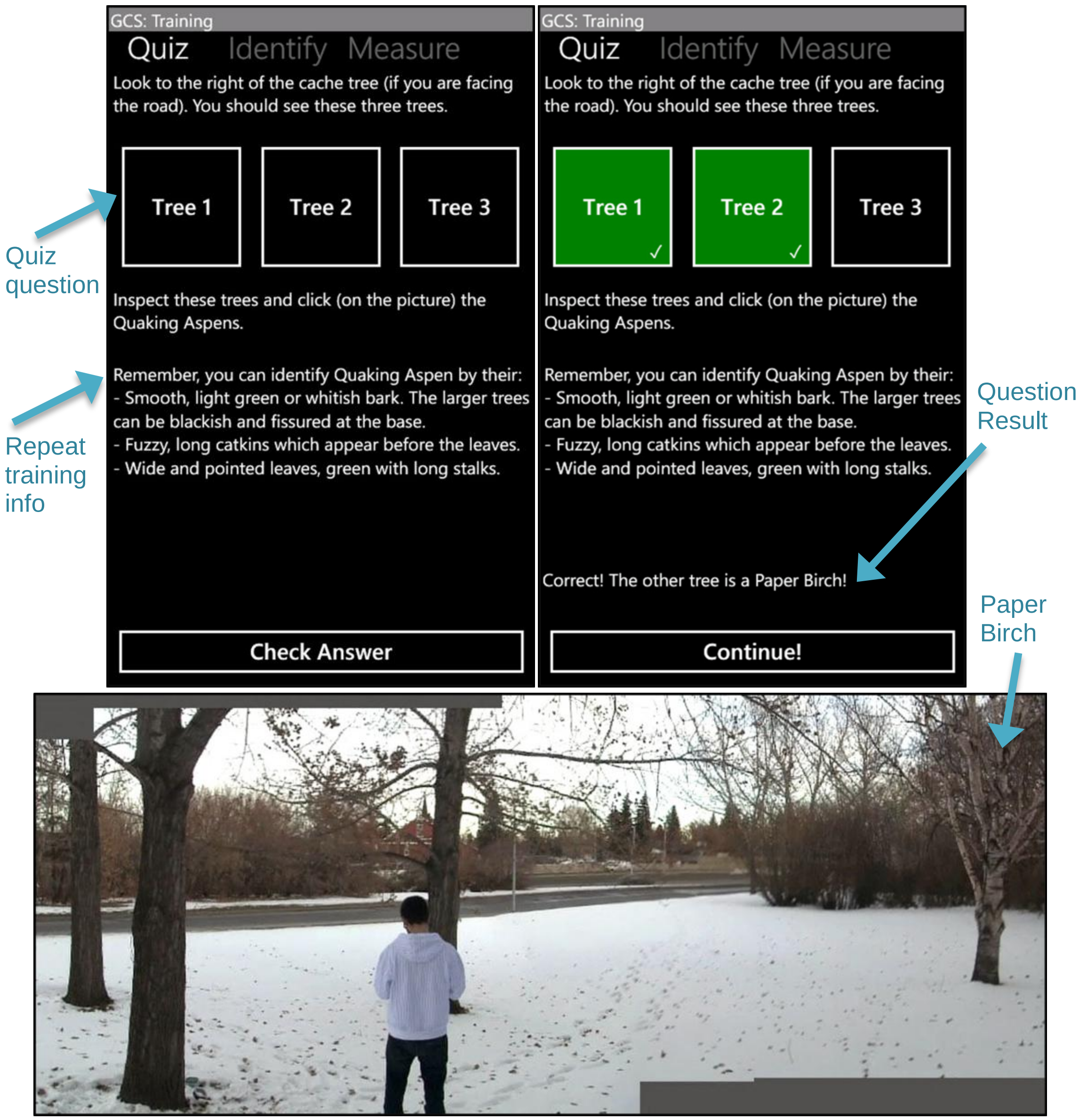

Figure 4.17: Quaking aspen identification quiz question. Top left, Nolan is asked to look at three trees and decide which one is a quaking aspen. Top right, he chooses trees 1 and 2 as quaking aspens, and is correct. Bottom, Nolan looking at the tree trees.

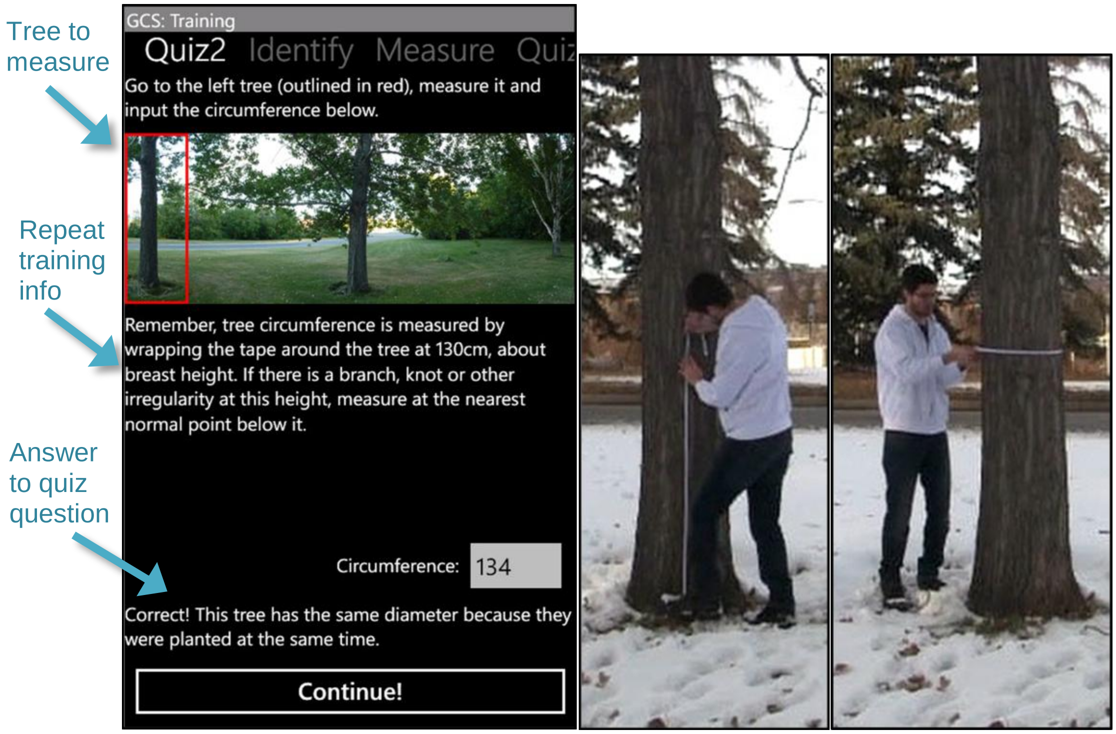

Figure 4.18: Tree measurement quiz question. Left, Nolan is asked to measure a specific tree's circumference and enter the answer. Middle, Nolan measures the height (130cm) at which to measure the tree. Right, Nolan measures the tree, getting 134cm.

4.7 Scenario 5: Computer-Assisted Coordination

This scenario turns to the scientist, where it illustrates how the Science Caching system can support scientists' needs for coordinating work across various sites. The scientist needs to indicate and track what sites should be visited, see what her citizen scientists are able to do (e.g., in terms of their training), and ultimately to guide particular citizen scientists to sites of interest that match the citizen scientist's training capabilities. As we will see in this scenario, the Science Caching system automates parts of these needs. In particular, the scenario considers how sites are given priority, how citizen scientist skills are tracked, and then how these two forms of information are used to automatically coordinate which sites are emphasized to a citizen scientist.

Citizen scientists can be tracked automatically as they perform tasks, allowing the system to build user profiles that are held in the Science Caching Server. For example, the first citizen scientist Arthur has taken part in Science Caching for a while. He has completed training for multiple data collection projects, and has collected at many sites. His data has rarely been flagged as incorrect. From this information, the system infers that Arthur is a trustworthy and experienced citizen scientist, and records it as part of his record. Information such as this can be used by the scientist later on, for example, to decide which citizen scientist's data should be trusted if an abnormal reading is made (see Scenario 3). It can also be used to decide what data collection tasks best match the particular abilities of a given citizen scientist.

Automation can also help Jill keep track of the sites that need be visited the most. When creating a site, Jill gives it a priority rating, indicating how important it is for a data collection to take place. She can change this rating at any time, for example if she thinks a site needs more visits, or because the amount of data collected so far suffices. This site priority rating can also be set to change automatically based upon different factors, such as the number of collections taken so far, a schedule for periodic collection at a site, or whether erroneous data was flagged at the site requiring a repeat visit (not shown). This allows the system to react to the changing situations of a citizen science project.

From a citizen scientist perspective, site priority and citizen scientist trust can influence what nearby sites are emphasized to a particular citizen scientist on his or her display. In the first example, Arthur is highly skilled in all the collection tasks available through the system. When he looks at his cache browser (Figure 4.19 left), the quaking aspen site is brighter than all the others due to it being the top priority cache on his map. The system's rating of his ability to do this task also contributes to how this site is emphasized. However, Arthur decides he wants to look for caches in a different area: he pans the map, and the mobile device calculates other top priority caches within the new area that matches his skills (Figure 4.19 right), where it finds one concerning duck-sighting. The cache icon becomes brighter to reflect its importance

Figure 4.19: Left, Arthur loads up his map view and sees that the tree collection cache is emphasized by being bright red, meaning it is the most important site for him to visit. Right, when Arthur moves the map so the quaking aspen site is no longer shown, his mobile device re-evaluates the most important site, and instead emphasizes the duck counting site.

In the second example, trust influences the emphasized site. In the training scenario (Scenario 4), Nolan is trained on collecting quaking aspens data. At first, Nolan is untrained in all tasks, and therefore untrusted, and is only given access to training sites. Since the only training site in his area concerns quaking aspens, Nolan is presented its icon in bright red (Figure 4.20 left). When Nolan completes that training, the system records this, and unlocks the quaking aspen collection sites (Figure 4.20 right). Because there is only one quaking aspen collection site in the map area (which happens to be near the test site), it is shown in bright red, suggesting to Nolan that he should participate.

Figure 4.20: Left, citizen scientist Nolan only has access to training caches because the system needs to build information on his ability. The training cache is highlighted because it is the only one he is trusted to perform. Right, when Nolan finishes the quaking aspen training, he is given access to quaking aspen collection caches. He is now trusted to perform this type of cache, so the one on his screen is highlighted.

4.8 Scenario 6: Enriched Real-Time Interaction

In this final scenario I discuss how a scientist is able to monitor caches and citizen scientists in real time, which assists in both coordination and communication.

Jill loads up the Science Caching desktop application to work with her project. The map view (Figure 4.21) provides her with information about her sites and citizen scientists who are active in the project. Jill can see the actual sites on a map (Figure 4.21), and a list of these sites. Through these, she can access details about the cache, including whether the cache has been flagged for possibly invalid data (Figure 4.21 top). When Jill clicks a site in one view (list or map), the site is selected in the other view. The selected site's photograph is presented, and Jill can switch to the collections tab inspect the data collected at that site (as previously shown in Figure 4.13). While not yet implemented, it would be a routine to modify the interface to allow the scientist to create a site interactively and set its properties, e.g., by dragging a 'new site' icon to a particular location, and then specifying its attributes via (say) form filling. This form filling interaction could also be used to modify already created sites, for example when site priority needs to be changed manually.

In an alternate version of validation scenario 3, Steve has sent Jill a question asking why his tree circumference measurements seem to be too large. Through the map view, Jill is able to view and interact with Steve while he performs his tasks. She sees Steve's message and clicks on his name. This reveals the citizen scientist pane appears, giving Jill information about Steve. She suspects the problem is one of training. She then sends Steve a message saying, "Are you holding the tape measure tightly? If you can't figure it out, try going through the training cache over here." She then right-clicks on training cache nearby (the one in Scenario 4) to drop a marker there. On Steve's chat window (Figure 4.22 right) he sees the marker placed by the cache and Jill's message. Steve is now able to continue with his task, where Jill was able to help him in real-time.

4.9 Summary

In this chapter I detailed an example of how Science Caching concepts can be applied to citizen science through scenarios surrounding a prototype mobile device application. I first provided an overview of the scenarios and implementation. I then detailed the scenarios which primarily centered around the previously identified four citizen science problems: *data collection, data validation, volunteer training* and *volunteer coordination*. In the next chapter, I present these scenarios to scientists, coordinators, and citizen scientists, where they were asked to provide a critique.

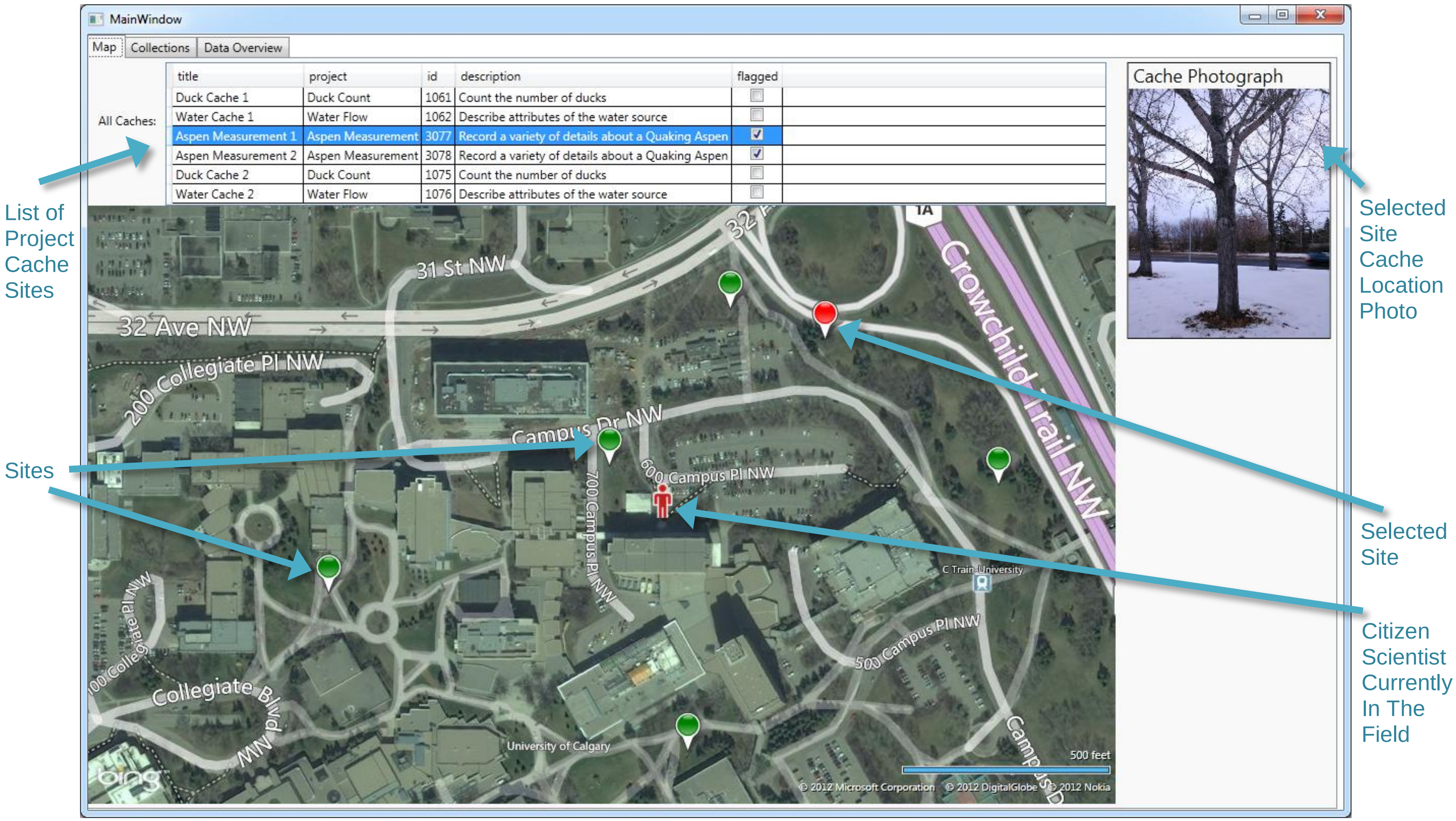

Figure 4.21: Scientist controller map view. Jill the scientist is able to see the sites in her project, view details on those sites, as well as see citizen scientists in the field.

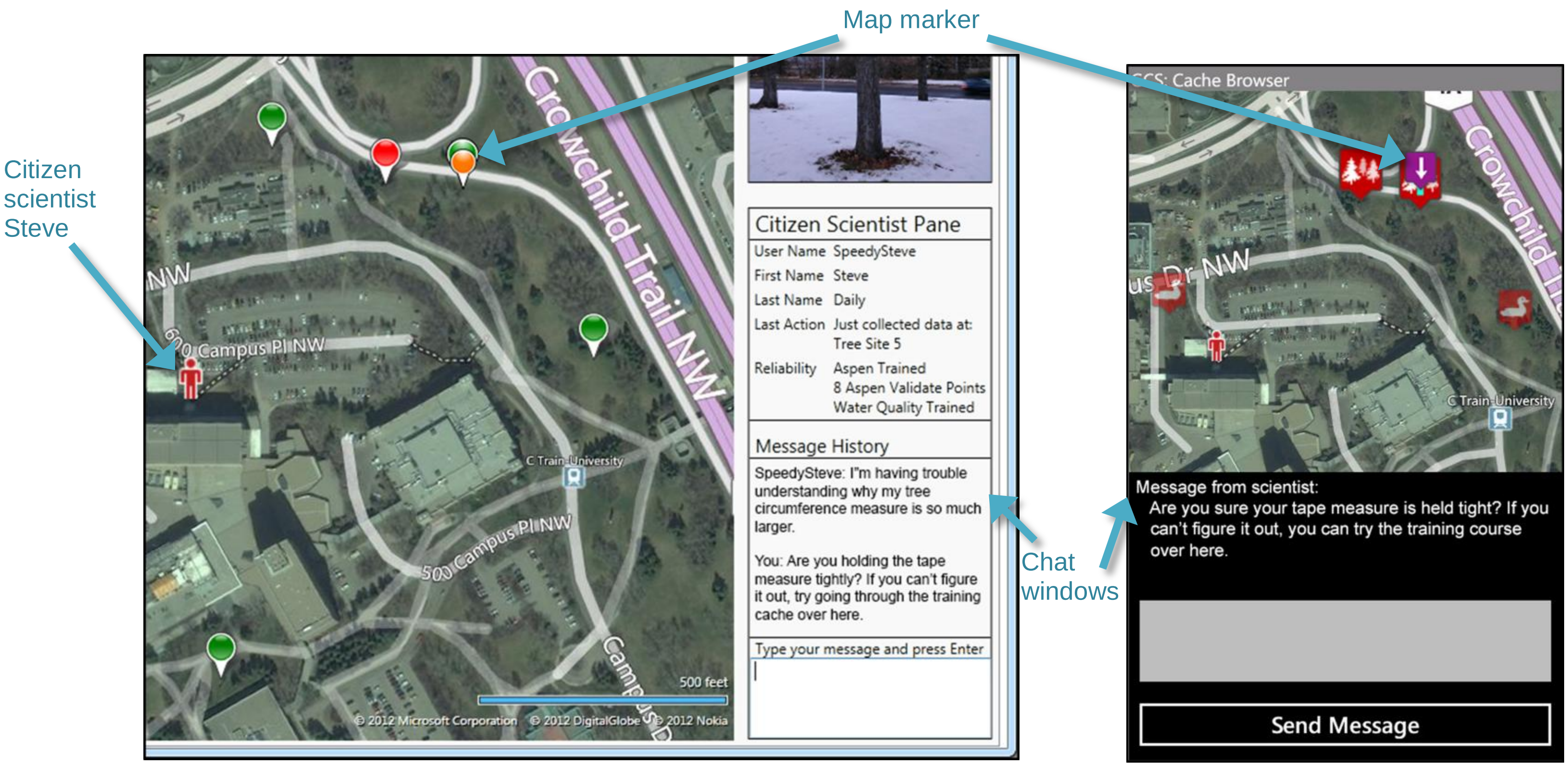

Figure 4.22: Chat functionality between scientist and citizen scientist. Left, when Jill clicks on the citizen scientist in her map view, she is given details about citizen scientist Steve. She sees that he needs help with tree measurement, so she replies with some advice and points him in the direction of a training cache if he needs it. Right, Nolan sees the marker and the message.

Chapter 5. **Design Critique**

In this chapter I reconsider the concepts illustrated in Science Caching, using a series of design critiques performed with individuals active in citizen science. First, I discuss my choice of design critique as an appropriate methodological approach that helps address particular research questions. Second, I discuss the elements that made up the critique: the prototype system presented, the different scenarios that were discussed with participants in the critique, the participants themselves, and my analysis method. Third, I detail the findings of the critique in terms of three themes that arose from the analysis of the results: *discussion on targeted problems*, *social interactions in citizen science*, and *practical deployment of physical caches*. Finally, I briefly conclude on how I validated these findings using a set of follow-up interviews with a smaller set of key participants.

5.1 Choice of Design Critique

In this section I discuss several questions that arose from the work presented in the prior chapters, and why I chose a design critique methodology to help explore those questions.

1. ***Did I target appropriate citizen science problems?*** The subset of citizen science problems targeted in the designs presented in Chapter 4 were drawn mainly from literature as discussed in Chapter 2 and 3. In particular, I focused on four particular problem areas, as these 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.
2. ***Is my design approach to these problems reasonable?*** The solutions developed for the four problem areas in citizen science, as presented in Chapter 4, have neither been tested in the lab nor deployed in a real-world citizen science project. Assuming that

the problems they attempt to solve are valid ones, I do not know if the particular designs as presented via the scenarios are reasonable solutions to those problems.

3. ***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, the designs presented in Chapter 4 are just the first working sketch. As in any user interface project, I foresee a large number of iterations over this design space before getting the design right. Related to question 2 above, I 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 I have missed design opportunities. Knowing the answer to this would be valuable to see where I should focus my efforts on future iterations.
4. ***How can these ideas be applied?*** The Science Caching system as presented in Chapter 4 was designed around several abstract ideas in Chapter 3, implemented in a deliberately simple citizen science application. While I 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).

I chose a form of a design critique as an appropriate method to answer these questions. Generally speaking, a design critique is an open-ended evaluation approach, typically involving focused conversation with either an individual or small group. Sketches or samples are shown to the group, with the intention of gathering high-level feedback about a small set of questions (as above). More generally, feedback from this kind of approach is gathered by showing an end user's flow through a design (e.g., by a walkthrough), exploring different potential ideas as thought experiments, and by requesting specific feedback about particular design ideas. As we will see, I used this form of design critique to explore the questions outlined above, where I walked experienced citizen scientists, citizen science coordinators, and scientist through the prototype system on-site and gathered their feedback and discussion, i.e., as a one on one design critique.

Other approaches besides design critique were considered, such as a *usability study* or *field deployment*. A usability study was deemed inappropriate because the ideas in this thesis are in their early stages (Greenberg et al., 2008). As well, user study results tend to return very specific feedback about the existing implementation (which as mentioned is just a working sketch) as opposed to the potential of applying the ideas of that system which I felt was more appropriate to my exploration. Field deployment on a real project would be both premature (due to the prototype nature of both the system and the design) and excessively time consuming.

5.2 Method

In this section, I now discuss the design critique methodology, detailing: who participated, how ideas were presented and discussed, what data was collected, and how the results were analyzed.

5.2.1 Participants

Nine participants were recruited: two male and seven female of various ages. Recruitment emphasized those with large amounts of citizen science experience. We targeted these participants as they would consider these design ideas from the perspective of their experiences with several citizen science projects. Experience ranged from citizen science researcher to expert citizen scientist to project coordinator, to the scientist in charge of the citizen science project. Table 5.1 identifies each participant with a participant number (P1-P9), along with their expertise and experience in citizen science projects. The identifiers in Table 5.1 will be used to discuss participants throughout the remainder of this chapter.

Participants were recruited via several means: I posted recruitment requests via email to a citizen science conference mailing list; word-of-mouth; and contact with potential participants who were experienced in the area. Participants were offered

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 5.1: The participants who took part in the design critique. Details their role in citizen science and their project experience.

compensation for travel and food for participation. Many of the participants were employed with or had worked directly with the National Parks system[9].

5.2.2 Prototype and Scenarios

Participants were walked through the scenarios and prototype detailed in Chapter 4, which included visiting actual sites. To make this convenient to participants, several sites were prepared: those in Calgary went to a pre-prepared site located on the grounds at the University of Calgary, while those who lived out near the National Parks were brought to a site prepared in a reserve located in their town of residence. After walking through and discussing citizen scientist interactions on-site, the discussion then moved indoors. In

[9] This introduces a certain bias to my findings, that participants recruited mainly had experience with small projects designed for their 'hands-on' involvement. The differences between hands-on and hands-off projects warrant future research, as discussed in Chapter 5.

some cases, this discussion structure was altered to accommodate participants' availability.[10]

The order of the scenario presentation was: *data collection*, *data validation* (without scientist validation), *volunteer training*, *collection point creation*, and finally the scientist controller interactions (*validating data* and *direct interaction with citizen scientists*). The *computer-assisted coordination* ideas were discussed when they were experienced in the interface, and were not introduced with the same level of detail (See Appendix A for the guide used in presenting, and the consent form).

5.2.3 Data Collected

Interviews were recorded with a portable audio recorded for later review and (on an as needed basis) transcription. Field notes were also collected during interviews.

5.2.4 Analysis

Analysis focused on the audio-recorded interview data and field notes. Due to the wide-ranging nature of design critique, the interview data was selectively transcribed, focusing specifically on parts of the discussion that were relevant to the prototype's design and purpose. The next step was to organize participant ideas and comments into themes using affinity diagramming. A sample can be seen in Figure 5.1. A list of all the themes from the discussions can be found in Appendix C. Affinity diagramming worked as follows.

- Affinity diagramming occurred in parallel with participants, i.e., the process below began with the data collected from the first participant with iterations performed as new data on other participants were collected.
- Key phrases from the transcription and the field notes were written on separate sticky notes.

[10] [P2]: Did not participate in outdoor demonstration. Discussion focused only on collection and training scenarios, and only with screenshots from Chapter 3. [P5]: Did not participate in outdoor demonstration. Discussion relied on video of the citizen scientist interactions and presentation of the scientist ones. [P6]: Only participated in part of the outdoor demonstration. Did not discuss site creation or scientist interactions.

- Sticky notes were related to one another and organized into groups. These groups were not predetermined, although the initial configuration was drawn from themes that emerged from the field notes.
- As new data from other participants arrived, groupings were revised as needed to either accommodate new themes, to refine existing themes, or to create new groupings that better reflected what participants said.
- New data from participants was also used to evolve the questions asked to participants. This may have affected why certain participants (especially earlier ones) did not provide feedback on some topics.
- At the end of the analysis process above, the categories were presented to a subset of participants in order to validate them. Specifically, these participants were asked whether the categories seemed appropriate and whether they aligned with the views they had presented in the original design critique.

Figure 5.1: A sample of the affinity diagramming, looking at the theme *social interactions in citizen science*.

As a consequence, the final groups or themes emerged from an on-going process of specification, organization, refinement and (to a limited extent) validation.

In the next section, I will discuss the themes that resulted from my design critique with experts in citizen science.

5.3 Theme: Discussion on Targeted Problems

In this section I look at the feedback received that relates to the four problem areas discussed in Chapters 1 to 4 of this thesis. For each problem—*data collection, data validation, volunteer training and volunteer coordination*—I begin by summarizing the feedback received and setting the scene. I then look at the sub-problems (summarized in Table 2.2) that received specific feedback, responding to the feedback when necessary. Finally, I look at any other issues that were brought forth by participants, and discuss how these could be addressed. These findings help answer questions 1-3 outlined above: *were the correct problems targeted*, *is the design approach to these problems reasonable*, and *how can this design approach be extended.*

5.3.1 Data Collection

Generally, participants saw the data collection ideas in Science Caching to be useful. The caveat was that its application would depend on the specific citizen science project [P1-9]. Details are provided below.

The *data collection* methodology embedded in the system supports well-structured data collection [P1-5,8]. This structured data approach seems to remedy a core problem that many participants were trying to address in their own projects, i.e., that many existing data collection methods are unstructured. For instance, [P3] expressed her frustration with unstructured data collection from a prior project:

> *"Right now… they're using hand written cards and someone sits and enters them. [The problem is that] it's not even entered so that you can do an analysis on it.' [For] location it says 'See the card.' [Laughter] Basically what I can [see] is that you saw five beavers, somewhere in this large study area. This is... not useless... but almost useless."*

Despite the best efforts of her citizen science volunteers, the collected data was often incomplete, or in a form that was effectively unusable. These problems were echoed by [P1,4,5]. [P1,3,4,5] thought that both the general approach to data collection as well as the approach to the data collection sub-problems illustrated in Chapter 4 helped mitigate this problem.

Participants were also quite interested in the way Science Caching exploited basic functionality afforded by mobile technology, such as:

- automatic recording of GPS location [P3];
- entry of data, including verbose descriptions and photographs, via mobile data collection forms [P3,4,5], and
- the automatic upload of data to scientists [P3,4,5].

Many participants reported that they had already foreseen the opportunity of basic mobile functionality in their projects [P3,4,5], and were excited to see it realized in the system.

Finding Sites

To perform data collection at a location, a citizen scientist must first find it. All participants saw providing site information, such as the location on a map, a site description and a photograph, as useful for finding a site.

The use of physical caches received a generally favorable reaction. [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 [P5,6,7,8]. 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, the work involved in having physical caches would present far too much work for scientists for far too little payoff. Furthermore, some participants expressed a more nuanced view of when physical caches would be appropriate. I explore these issues in more depth in Section 5.6, which discusses the practical deployment of physical caches.

Creating Sites

Reaction to site creation was mixed. [P8] saw citizen scientist site creation as a way to extend Science Caching for 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.

Other participants did not express strong feelings about allowing citizen scientists create collection sites. This could be an issue with the way site creation was presented to participants rather than any problem with the idea. In particular, while the idea of creating collection points was presented to them (to be discussed later), actual site creation was not; it was only described. This likely made the idea unclear.

Transferring 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, “Area Caches”, where a central tool cache container had many collection locations around it, were proposed as an alternative strategy. They could decrease the number of tool caches needed to be deployed [P2-5], thus decreasing time and effort needed from scientists to set up or deploy tools. Having such a sub-set of caches balances the benefits of having these resources available in the field *vs.* the costs of set-up (for more information, see Section 5.6)

[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:

> *"That's actually very cool, because 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]*

However, this possibility may apply only to certain types of samples, as some are time/environmentally sensitive.

5.3.2 Data Validation

Scientists and coordinators interviewed did not use techniques other than expert review (Wiggins et al. 2011), so they had no experience supporting and improving repeat collection through any form of mobile technology. Specifically, participants did not give substantially different feedback for the sub-problems *how sites are revisited* and *on-site data comparison*. Due to this, I will discuss all the feedback for data validation together.

Participants were generally supportive of conducting multiple readings for data validity [P1,3,4,6,7,8]. The ability for citizen scientists to flag suspicious records while on site was considered a useful tool to allow further validation by scientists or citizen scientists [P3,4,9]. Participants also emphasized the importance of requiring a photograph with collected data to improve data validity [P3,9].

The prototype system took the approach of presenting previous records for validation when the current citizen scientist's collection varied widely from the previous collection. As discussed in Chapter 4, 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], for example, saw bias as problematic in situations where a high degree of data quality is needed. Yet participants also viewed bias as something that could be managed, and even potentially turned into a learning opportunity.

Participants described several ways for solving bias problems. For instance, [P1] felt that biases were often predictable and constant, and thus could be corrected:

> *"If you consistently see outliers in the data set, you will often just download the person's data and sometimes you can correct it. [For example] the person just looks up and sees canopy cover that's twice what everyone else does."*

For learning, [P8] said that more verbose feedback – whether automatically generated or relayed by the scientist – about suspected biases could be given to the citizen scientist. [P3] stated that the citizen scientist could be shown how their data compares to the average data collected over a time period, to allow self-education on their own biases.

5.3.3 Volunteer Training

The solutions to volunteer training were received with general but unanimously positive support [P1-9]. The training scenario explored how to lead citizen scientists through a real-world training course. Participants saw real-world guided education as part of the success of many projects [P2,3,4,6,7]. 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 [P1,2,3,4,7,8,9]. 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 were interested in how these ideas could be applied to different data collection projects [P1,2] as well as educational-focused citizen science [P3,8,9]. Discussion on the possibilities for applying training to other areas is reserved for Section 5.6.3.

5.3.4 Volunteer Coordination

The two coordination solutions presented received differing responses. In this section, I first discuss the ways that *computer-assisted coordination* and *enriched real-time interaction* can be used and extended. In general, these specific approaches to coordination were seen as useful. However, participants said that these approaches missed the important social dimension essential in actual coordination (to be discussed in Section 5.5)

Computer-assisted Coordination

Automatically tracking the skill and participation of citizen scientists was seen as a good approach for identifying highly engaged participants [P3,4,5,8]. [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 [P1,3,4,5,9] (to be discussed further in Section 5.5).

Opinions differed on how much automation should be part of a citizen science project. [P3,4,5,8] 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 use as a social activity by any group that was interested (discussed more in Section 6.2.2 – Hands-on or Hands-off Citizen Science).

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 Science Caching system more usable. 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 to train 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. I have this database where it has who's completed what [training module] and you're supposed to check it off… [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.

5.4 Theme: Citizen Science as a Social Experience

While participants were generally favourable to the approaches presented to them, they saw further opportunities for a Science Caching system to support the various social interactions amongst the people involved in citizen science. In this section, I present three major social interactions that were suggested but not supported in the scenarios, and discuss how these needs can be met through future technology. Specifically, I look at *performing tasks as a group*, *citizen scientist social needs*, and *sharing scientist knowledge*. This approach will help answer questions one and three outlined in the start of the chapter: a*re the correct problems targeted* and *how can these ideas be extended?*

5.4.1 **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."*

[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 Science Caching 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. Collaboration on the same task could be supported by networking multiple mobile devices together (e.g., over ad-hoc Wi-Fi or cellular). 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 Science Caching system, possibly using networked calendar, chat room, or social network group.

5.4.2 **Social Motivations of Citizen Scientists**

Many citizen scientists are motivated to participate in citizen science because of the social atmosphere [P1,4,6,7,9]. Some projects need social interaction 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]. [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.

Yet such socialization is not supported by the Science Caching system of Chapter 4, something that can be accommodated in future versions of the system. A citizen scientist could, for example, create a meet-up in the Science Caching system, 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.

5.4.3 Sharing Scientist Knowledge

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] discussed the need to support this goal, 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."*

In many cases, scientists are not just interested in sharing knowledge, but changing public behaviour and addressing problem situations between humans and nature. [P3] discussed this with her work (during the reality check interview, Section 5.7):

> *"I think something that is really important in the social interactions is the fostering of dialog. For our projects, we are trying to address a conservation challenge or something that is a problem on the landscape. … hopefully get some more innovative solutions because you have people through their experiential learning of contributing [to the citizen science project and trying] … to understand that problem, also have discussions about it and what the solution should be. …There is stuff that they can change in their behaviour to reduce the conflict, but they have to come to that realization."*

This connection is also important to citizen scientists. Participants pointed out that citizen scientists were highly motivated by direct connection with scientists and experts [P1,3,4,5,9]. [P4] discussed this in a project she coordinated for a botanist:

> *"We did grassland plots. We were out there with this botanist, and everyone wanted to ask him things and he was the god to look up to. I think this is a critical part of people wanting to do this work is they want to connect with the experts."*

Indeed, the connection with the scientists was seen as a motivating factor both in getting citizen scientists started with a project [P1,3,4,5,9], and with them coming back [P1,3,4,5].

Participants also thought technology could be a vehicle to help support this connection [P3,4,5,9]. [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 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.

5.5 Theme: Practical Deployment of Physical Caches

The use of caches for citizen science is a core idea behind this thesis. Discussion of the use of caches came up in every conversation, specifically focusing on when physical caches should be used. In this section, I first look at various factors participants identified that make the potential use physical caches valuable. Second, I look at the problems that participants saw with physical caches. Finally, I present the areas participants identified where the benefits outweigh the problems with physical caches. This approach will help answer questions two and four: *is my design approach to these problems reasonable* and *how can these ideas be applied in practice*?

5.5.1 Value of Physical Caches

In this section I look at the factors participants identified that make the potential use of caches valuable.

Caches are Only Valuable for Multiple Visits

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]. 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..

Use of Caches as Containers

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 collection projects. This could be because participants has already developed alternate 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 did view training caches as a good opportunity, where cache containers were seen as a way to allow training with less effort by the scientist [P1-5,8,9].

Participants were more interested in using caches to store samples [P3,4,5,8,9]. Many specific applications of caches holding samples were suggested, including: animal hair [P3], bear DNA [P4], and water samples [P3,8]. Discussing hair samples, [P3] remarked:

> *"I like the idea of little hair sample envelopes in there and people [can] collect the hair right [in] them and put them back in."*

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) [P3,4,5,8,9].

5.5.2 Problems with Physical Caches

Participants identified several problems concerning physical caches as well as some potential solutions, as listed below.

Effort to Deploy 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.

Environmental Impact

[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, saying

> *"They [scientists] are thinking non-intrusive. I guess they always have problems a bit with caches, but on the other hand this means you are not marking the tree per-say."*

[P4] shared the same concern, that geocaches conflicted with scientists interests in preserving nature. Scientists [P5,6,7] 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, or even by monitor the site to determine the impact of the caches themselves.

Cache Tampering

Another concern was whether cache containers would remain intact. [P8] was concerned with humans destroying or stealing caches and their contents: *"It's important how secure it is"*. Locking the container was discussed, but [P8] saw this as problematic too, saying that

> *"if you got it locked, people are going to want to see what's inside it, it's probably nothing of interest [but] they'll just take the whole thing".*

[P8] saw possible issues with animals tampering with and destroying geocaches, but did not go into detail as to why this may happen. While animals are known to destroy geocaches, the rate of such damage is so low as to be inconsequential, thus this issue is likely minimal unless the equipment inside is expensive.

5.5.3 Project Areas Amenable to Physical Caches

Participants discussed specific projects that they thought appropriate for Science Caching, and for physical caches. From these, I extracted types of citizen science projects where they would apply.

Maintaining Physical Objects

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 Science Caching to these needs possible with little added effort.

Long-term Site Use

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.[11] 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.

Training Caches for Different Applications

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. These caches could be revisited many times, making the deployment costs worthwhile. Tools could be stored in the caches so that trainees can practice their skills. Indeed, [P1]'s citizen science project was very similar to the aspen tree scenario used in

[11] [P1] also discussed the potential of turning old landscape photographs into virtual caches, their location being the rough area where the photograph was taken. A citizen scientist could go to that area, identify and mark (as well as possible) the specific location where the photograph was taken, and take a new photograph.

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.

5.6 Reality Check

As mentioned, the analysis methodology first created a set of themes by analyzing the interview results. In order to get a reality check, these themes were then presented in a second round to two highly experienced participants, [P1] and [P3]. [P1] was interviewed in the middle of theme creation, while [P3] was interviewed as the themes were being finalized. The feedback from the reality check interviews generally supported the themes as discussed. The areas where [P1,3] agreed strongly, disagreed, or expanded upon a theme are collected in Table 5.2. Where there were differences, the themes were expanded to include those differences.

P1	• Saw physical containers to be worthless in most citizen science situations, because they require too much work. • Was very interested in citizen science projects that require little scientist input (discussed in Chapter 5), but saw the importance of scientists as motivators for citizen scientists to take part. • Discussed the ability for mobile technology to assist the citizen scientist, through apps that can perform tasks, such as bird call identification or leaf identification.
P3	• Scientists want connection with citizen scientists not just to educate, but to change behaviour and address problems groups have with nature. • Caches have risk, but if they allow you to do/collect new things, then the risk is worth it. • Was interested in putting citizen science inside of Geocaching (discussed in Chapter 5)

Table 5.2: Summaries of the two reality check interviews, with the areas where participants agreed strongly, disagreed, or expanded upon a theme.

5.7 Conclusion

In this chapter I discussed the validation of my Science Caching ideas through design critique with individuals active in citizen science. First, I explained my choice of design critique. Second, I discussed how the critique was structured. Third, I detailed the results of the critique in terms of the three themes that arose: *discussion on targeted problems*, *social interactions in citizen science,* and *practical deployment of physical caches*. Finally, I concluded with how I validated these findings using follow-up interviews with a smaller set of key participants.

Chapter 6. **Conclusion**

This thesis was motivated by the challenge of how to alleviate problems in citizen science, specifically problems with *data collection, data validation, volunteer training* and *volunteer coordination.* To address these four problems, I considered the similarities between geocaching and citizen science as well as the opportunity provided by mobile computing to inform solutions. With this information, I designed and prototyped Science Caching, a system that uses physical cache containers and mobile devices to mitigate the four problems. I created a series of scenarios illustrating Science Caching usage around these four problems. Using an interview-based design critique methodology, I then presented these scenarios to scientists, coordinators and citizen scientists. Results were gathered as feedback on the Science Caching design choices, and as discussions on how they saw using Science Caching ideas in their own (or other) projects. Using affinity diagramming to analyze this information, I discovered three overarching themes: *discussion on targeted problems, citizen science as a social experience,* and *practical deployment of physical caches.* These themes generally reflect positively on the Science Caching approach, informs future iterations of this research, and more generally suggests how to apply these ideas to various aspects of citizen science.

6.1 Contribution

The major contribution of this thesis is the *creation* and *refinement* of Science Caching, a way of using mobile technology and aspects of geocaching to approach citizen science's problems with data collection, data validation, volunteer training and volunteer coordination.

- *Creation* – In the creation of Science Caching, the four problems were targeted specifically:

1. *Data collection* is performed by leveraging known prepared sites and mobile devices.
2. Data is *validated* through use of known, repeatedly visited sites.
3. Volunteers are *trained* in collecting new forms of data by interacting with mobile devices and real world training sites.
4. Citizen scientists are *coordinated* around science sites through real-time communication with scientists and better understanding of individual talents and abilities.

- *Refinement* – These ideas were refined through design critique with those experienced in citizen science research. The Science Caching ideas were presented through a prototype and discussed with participants. The results of these discussions were analyzed with affinity diagramming to arrive at three themes that inform the future of Science Caching: *discussion on targeted problems, citizen science as a social experience,* and *practical deployment of physical caches*.

Lesser contributions were also made to citizen science research:

- In-depth discussion of prior solutions related to citizen science's problems with data collection, data validation, volunteer training and volunteer coordination.
- Design and implementation of the Science Caching prototype, including non-geocaching focused aspects, which informs future design of mobile citizen science.

6.2 Real-World Problems

The major contribution of this thesis, Science Caching, provides a possible direction for citizen science projects, but may face problems when applied in the real world. Some of these problems are shared with mobile and digital system in general, while others are unique to the application of geocaching.

There are obstacles that must be overcome when implementing a digital system, requiring time and money. First, a digital data collection system must be implemented. When implementing a mobile system like Science Caching, the mobile device platform or platforms must be chosen for implementation. Targeting multiple platforms increases potential user base, but also increases cost and system complexity. With this system in place, it needs to be tested to ensure it is will fulfill project needs. Scientists and citizen scientists need to be trained in using the system. The time needed for these steps is difficult to know at the beginning, possibly increasing the cost and time needed.

When the system is up and running, further obstacles come into play with using mobile devices. Certain demographics are uncomfortable with mobile devices, and may be alienated from using the system. If a digital system is hard-coded for the current needs of a specific project (as many are), a software developer will have change the structure of collection if project needs change. This could be avoided by creating an editor for scientists to change the collection tasks themselves, but this increases cost. Also, unlike traditional paper and web-based collections, mobile device applications may quickly become obsolete as new mobile devices and platforms enter the market.

Some obstacles are specific to Science Caching. An initial pool of collection and training caches is needed for citizen scientists to interact with. These will likely need to be set up by scientists or other experts, as citizen scientists will not be experienced with the system design. There are also obstacles faced when applying Science Caching to an already existing project. First, not all projects apply (as discussed in Section 5.4), so it must be made certain that Science Caching is a good fit for the existing project. When Science Caching becomes used for a specific project, it may alienate citizen scientists who previously collected for the project and are not interested in the new structure. To deal with this, the old collection structure may need to be retained, but this increases cost and complexity.

6.3 Future Work

In this section I discuss future directions for this research. First, I discuss how Science Caching can be extended through deployment on real projects. Second, I discuss how Science Caching can be applied to two varieties of citizen science, differentiated by the level of scientist involvement in the projects working. Finally, I look at how citizen science may be able to function inside of the real Geocaching network.

6.3.1 Extending Science Caching

The Science Caching system described is, as mentioned, a working sketch. The citizen science project in the scenarios is an imaginary one, which was kept deliberately simple to explore basic ideas. Following standard user interface practice, several additional steps are obvious. The first step is to identify several existing citizen science projects amenable to Science Caching – done in part in 4.6.3 through the design critique. The next step is to build a Science Caching system specific to those projects. The third step is to examine its use in actual practice and to iterate the system design. These and subsequent deployments will likely reveal many more opportunities for such a system, as well as additional issues and problems with it. From these experiences, the next step is to create a general mobile Science Cache platform whose content can be created and edited to fit the needs of particular citizen science projects.

6.3.2 Hands-on or Hands-off Citizen Science

Citizen science involving data collection often requires a 'hands-on' approach by scientists, e.g., for training, for coordinating volunteers, and for checking data. Alternatively, some projects have a more 'hands off' approach, that is, where scientists are not as involved in the citizen science data collection process. The design critique revealed that the application of geocaching could apply to both approaches. For busy scientists with serious time and resource constraints, a hands-off approach has clear benefits. Yet many participants in the study envisaged that the geocaching solution could also improve a scientist's hands-on interaction, that is, where the level of scientist involvement would not decrease, but would be shifted and enhanced by technology. Indeed, a few participants thought it important that scientists remain hands-on, as they had heard concerns from other scientists about the risk of losing their jobs to citizen science. The motivations and needs for hands-on and hands-off citizen science warrant future research; especially as my design critique was mainly with those experienced in the hands-on approach. As briefly described below, Science Caching can support both hands-on and hands-off citizen science.

Hands-on

In hands-on Science Caching, scientists remain at the center of data collection projects. The Science Caching structure allows scientists to offload the more tedious aspects of their citizen science projects (e.g. entering recorded data, verifying data is complete, providing introductory training), while enhancing the more critical aspects of their work. Connecting scientists to citizen scientists via technology could provide an avenue for outreach, such as connections to local communities, educating people about conflicts between people and nature, and so on. Scientists could still create mobile collection and training activities that do not need their direct involvement, which in-turn identifies keen participants they can recruit to work with directly on harder data collection projects.

Hands-off

In hands-off Science Caching, the Science Caching tool and the communities that form around it are the center of data collection projects. Projects are simple and well-structured, allowing them to operate long term without much scientist maintenance. Ideally, this provides scientists with essentially 'free' information. Automatic coordination techniques could identify skilled citizen scientists and allow them to perform many of the roles attributed to scientists, such as creating training and data collection sites in-person training and coordination. These projects can be built into social networking (e.g. Facebook, Twitter) to connect volunteers together. This structure also allows social groups to use these projects as they see fit, as [P1] stated:

> *"People may use [a citizen science project] for something you never thought [they would], and it doesn't cost… anything. … [It's] the social dimensions that drives this, and you can make it as friendly as possible for those kinds of activities."*

This option will likely result in questionable data collected, perhaps mitigated through crowdsourcing and other validation methods.

6.3.3 Placing Citizen Science in Geocaching

This thesis has shown that aspects of geocaching could benefit citizen science. A related opportunity is to extend this research to consider how citizen science could emerge from the current Geocaching community, which would leverage the existing popularity of geocaching. That is, the idea is to turn geocachers into citizen scientists. For example, Groundspeak's Geocaching have sustained millions of interested members. In contrast, citizen science projects often suffer from low participation and waning interest ([P1,8]; Iacovides 2013; Rotman 2012). Geocachers are already trekking to remote spots to record their presence and to essentially replace trinkets in a physical cache. It seems feasible that they could also be motivated to collect data for science as part of a more purposeful (and perhaps more satisfying) geocaching effort. For example, the EarthCaching (Groundspeak 2013) variant (described in Section 3.1.4) has succeeded in getting geocachers to analyze and answer questions on natural features as an alternative to finding a cache, an activity similar to how citizen scientists are asked to answer questions on plant phenology or animal sightings. Future work in this direction would be to extend existing 'general' caches in a geocaching system (e.g. Groundspeak's Geocaching) to allow citizen science while on location, leveraging ideas already found in Science Caching. Alternately, a new variant of cache could be created targeting geocachers who seek to take part in citizen science, making their activities more purposeful and useful.

6.4 Conclusion

With the rapid proliferation of mobile devices, citizen science has the potential to provide new quantities of information to better help us understand our world. In this thesis I have concentrated on four specific problems in mobile citizen science, but these problems (and others) are far from solved. With concerns over our environment on the rise, it is important to understand and alleviate factors that limit citizens from participating in citizen science, and that provides data useful for understanding our future. One way that these factors can be understood and alleviated is by looking to other areas similar in action and structure to citizen science. My work looked at one area, geocaching, in an attempt to isolate and exploit some of its successful components to citizen science. Further research is needed in to how other areas can be practically applied to citizen science, such as social networking and other location-based activities. Even so, geocaching presents a barely-tapped potential to connect citizens to nature while using their ability to collect important data.

References

Note: Due to the project-oriented nature of citizen science, many citations are to particular projects and essays issues around citizen science are available only online.

1. Alberta Wolverine Working Group. (2013), **Alberta Wolverine: Our Research**, [Online], Available: http://www.albertawolverine.com/?page_id=13, Accessed January 30, 2013.
2. Audubon Society, (2012), **Christmas Bird Count**, [Online], Available: http://birds.audubon.org/christmas-bird-count, Accessed February 18, 2013.
3. Bendell, J., (2011), **Becoming a Citizen Scientist**, [Online], Available: http://www.frontenacnews.ca/2011/11-51_dec_22/outdoors_lol_11-51.html, Accessed February 18, 2013.
4. Bonter, D. N., and Cooper. C. B. (2012), **Data Validation in Citizen Science: a Case Study from Project FeederWatch**, *Frontiers in Ecology and the Environment* 10 (6), Ecological Society of America, pp. 305–307.
5. c:geo, (2013), **c:geo – Android GeoCaching**, [Online], Available: http://www.cgeo.org/, Accessed February 5, 2013.
6. City of Portland, (2013), **PDX Reporter**, [Online], Available: http://www.portlandoregon.gov/bts/article/419527, Accessed February 6, 2013.
7. Colorado State University. (2013), **Doing Field Surveys**, [Online], Available: http://www.citsci.org/cwis438/websites/citsci/DoingFieldSurveys.php, Accessed February 6, 2013.
8. Cohn, J. P. (2008), **Citizen Science: Can Volunteers Do Real Research?**, *BioScience* 58, American Institute of Biological Sciences, pp. 192-197.
9. Colorado State University. (2009), **Training Citizen Scientists**, [Online], *Today @ Colorado State*, Available: http://www.today.colostate.edu/story.aspx?id=1708, Accessed February 6, 2013.
10. Cornell Lab of Ornithology. (2012), **eBird**, [Online], Available: http://ebird.org/, Accessed December 12, 2012.
11. Fitzpatrick, J. W. (2007), **What We Mean by "Citizen Science",** *Citizen Science Toolkit Conference*, Cornell Lab of Ornithology, pp. 1.

12. Google. (2012), **Google Maps**, [Online], Available: https://itunes.apple.com/en/app/google-maps/id585027354?mt=8, Accessed February 18, 2013.

13. Greenberg, S., Buxton, B., (2008), **Usability Evaluation Considered Harmful (Some of the Time)**, *Proceeding of the Twenty-sixth Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*, pp. 111-120.

14. Groundspeak. (2013), **Geocaching**, [Online], Available: http://www.geocaching.com/, Accessed February 5, 2013.

15. Hacker, S., Von Ahn, L. (2012), **Duolingo**, [Online], Available: http://duolingo.com/, Accessed December 30, 2012.

16. Halpern, M. Evjen, K.M., Cosley, D., Lin, M., Tseou, S., Horowitz, E., Peesapati, T., and Gay, G. (2010), **SunDial: Embodied Informal Science Education Using GPS**, *MedieKultur. Journal of Media and Communication Research* 27 (50), Society of Media researchers In Denmark, pp. 48-65.

17. Halpern, M. (2012), **SunDial Feedback**, Personal Communication.

18. Han, K., Graham, E.A., Vassallo, D., Estrin, D. (2011), **Enhancing Motivation in a Mobile Participatory Sensing Project through Gaming**, *2011 ASE/IEEE International Conference on Privacy, Security, Risk and Trust (Passat '11)*, pp. 1443-1448.

19. Haury, D.L., and Rillero. P. (1994), **Perspectives of Hands-On Science Teaching**, *ERIC Clearinghouse for Science, Mathematics and Environmental Education*, pp. 11-16.

20. Howe, J. (2006), **The Rise of Crowdsourcing,** *Wired magazine 14.6*, pp 1-4.

21. Ioanna, I., Charlene, J., Cornish-Trestrail, C., Cox, L. (2013), **Do Games Attract or Sustain Engagement in Citizen Science? A Study of Volunteer Motivations**, In *Proceeding of the Thirty-one Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*, pp. 1101–1106.

22. Kim, S., Robson, C., Zimmerman, T., Pierce, J., and Haber, E.M. (2011), **Creek Watch: Pairing Usefulness and Usability for Successful Citizen Science**, In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI '11)*, pp. 2125–2134.

23. LeafSnap. (2012), **Leafsnap: An Electronic Field Guide**, [Online], Available: http://leafsnap.com/, Accessed March 14, 2012.

24. Lintott, C.J., Schawinski, K., Slosar, A., Land, K., Bamford, S., Thomas, D., Raddick, M.J., et al. (2008), **Galaxy Zoo: Morphologies Derived from Visual Inspection of Galaxies from the Sloan Digital Sky Survey** *Monthly Notices of the Royal Astronomical Society* 389 (3), pp. 1179–1189.

25. Lo, B. (2011), **GPS and Geocaching in Education**, *International Society for Technology in Education*.

26. Neustaedter, C., Tang, A., and Tejinder, J.K., (2010), **The Role of Community and Groupware in Geocache Creation and Maintenance**, In *Proceeding of the Twenty-eight Annual SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*, pp. 1757–1766.

27. 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.

28. 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.

29. Newman, G., Wiggins, A., Crall, A., Graham, E., Newman, S., Crowston, K. (2012), The **Future of Citizen Science: Emerging Technologies and Shifting Paradigms**, *Frontiers in Ecology and the Environment (10) 6,* Ecological Society of America, pp. 298-304.

30. OpenScientist. (2011), **Finalizing a Definition of 'Citizen Science' and 'Citizen Scientists'**, [Online], Available: http://www.openscientist.org/2011/09/finalizing-definition-of-citizen.html, Accessed February 6, 2013.

31. Paleontological Research Institution. (2013), **Mastodon Matrix Project**, [Online], Available: museumoftheearth.org/research.php?page=Mastodon_Research/Mast_Matrix, Accessed February 18, 2013

32. Price, S., and Rogers, Y. (2004), **Let's Get Physical: The Learning Benefits of Interacting in Digitally Augmented Physical Spaces**, *Computers & Education* 43 (1-2), pp. 137–151.

33. Project Budburst. (2013), **Project BudBurst – Timing is Everything**, [Online], Available: http://neoninc.org/budburst/, Accessed February 28, 2013.

34. Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., Jacobs, D. (2012), **Dynamic Changes in Motivation in Collaborative Citizen-Science Projects**, In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work (CSCW '12)*, pp. 217-226.

35. Schudiske, E. (2011), **New Teacher Tools: Geocaching and Education**, [Online], *Groundspeak*, Available: http://blog.geocaching.com/2011/12/new-teacher-tools-geocaching-and-education/, Accessed February 6, 2013.

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

37. Stack Overflow. (2013), **Stack Overflow – All Privileges**, [Online], Available: http://stackoverflow.com/privileges/, Accessed February 6, 2013.

38. Teton Science Schools. (2009), **Investigating Pikas Training Manual**, [Online], Available: http://www.naturemappingjh.org/content/doc/Project%20Pages/TSS%20Citizen%20Science%20Project%20Overview%20033012.pdf, Accessed February 6, 2013.

39. The Open University. (2013), **iSpot Android app**, [Online], Available: http://www.ispot.org.uk/help-app-android, Accessed February 18, 2013.

40. University of Alaska Fairbanks. (2012), **Citizen Science Monitoring Instructions**, [Online], *Melibee Project*. Available: https://sites.google.com/a/alaska.edu/melibee-project/citizen-science/citizen-science-monitoring-instructions, Accessed February 6, 2013.

41. Von Ahn, L., Maurer, B., McMillen, C., Abraham, D., and Blum, M. (2008), **reCAPTCHA: Human-Based Character Recognition via Web Security Measures**, *Science* 321, American Association for the Advancement of Science, pp. 1465–1468.

42. 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*.

Appendix A. Experiment Materials

A.1 Presentation "Cheat Sheet"

Below is the reference document used when presenting my prototype

Introduction:

- Who am I
    - CPSC MSc
    - Love nature / Backpacker
    - Wanted to help scientists who work in this area.
- Quick Overview
    - Supporting certain needs in citizen science.
        - Smartphones
        - Geocaching (define)
    - Specifically looking at collection at sites
    - Scalable/Lower upkeep
    - Not targeting motivation
- Ask what projects they have experience in
- Goals
    - Discover how to advance my ideas
    - Interface my ideas with your experience/reality

Simple:

- Cache Browser: Prioritized selection with areas
- Collection
    - Cache allows findability
    - Cache provides tools and information

- Leaf:
    - Cache provides storage

Verification:

- Findability allows repeat, different visitors
- Repeat visitors and mobility allows data comparison
    - Flagging
- *Is there a better way to structure this comparison?*

Training:

- Real-world examples
- Testing on known values

Complex:

- *Hypothesis*
    - *Now let me show you a more complex and realistic scenario. You are a scientist that needs to keep track of the quaking aspens in a specific area.*
- Centralized cache
- Creation of new collection points

Controller:

- *This system is especially brittle and not fully functional.*
- Map:
    - Viewing caches and citizen scientists in the system
    - Interacting with these elements
- Conflict resolution
    - Flagged data is brought to attention to be dealt with.

A.2 Consent Form

Name of Researcher, Faculty, Department, Telephone & Email:

Matthew Alan Dunlap – MSc Student
Department of Computer Science
E-mail: matthew.a.dunlap@gmail.com
Phone: (403) 999-7148

Supervisors:

Saul Greenberg – Professor
Department of Computer Science
E-mail: saul.greenberg@ucalgary.ca
Phone: (403) 222-6087

Anthony Tang – Assistant Professor
Department of Computer Science
E-mail: tonyt@ucalgary.ca
Phone: (403) 210-6912

Title of Project:

Utilizing Geocaching and Mobility in Citizen Science

This consent form, a copy of which has been given to you, is only part of the process of informed consent. If you want more details about something mentioned here, or information not included here, you should feel free to ask. Please take the time to read this carefully and to understand any accompanying information.

The University of Calgary Conjoint Faculties Research Ethics Board has approved this research study.

Purpose of the Study:

Citizen scientists are volunteers that collect data for scientists. We are investigating new strategies for coordinating, training, data collection and data validation by citizen scientists. Our designs will leverage ideas found in mobile devices and geocaching .

What Will I Be Asked To Do?

This study will be conducted in several rounds. Depending on the round you are involved in, you will be asked to do one or more of the following.

Participants will be presented with the basic ideas behind our designs strategies to support citizen science, as well as designs we have devised. The presentation will include one or more of the follwing: explanation of these techniques through slides, demonstrations and/or walk-throughs of low-fidelity prototypes, and demonstrations and walkthroughs (including interactions) with prototype and/or working mobile devices.

Depending upon the round and situation, the location will be either in a meeting room, or outdoors outdoors at a site location, or a mix of the two.

The study may be conducted individually or in small groups. Note that participation in group discussion with other participants limits the confidentiality of your individual participation.

Participants will be asked about their thoughts of the design, as well as their experiences with it. This may be through either individual or group discussion. Specific questions will be asked about different aspects of the system with the goal of sparking discussion about improving our techniques.

Participation is voluntary, where it will require approximately 2 hours of your time. You are free to withdraw from the study at any time without any kind of penalty. However, we will reserve the right to keep and use data collected until the point of withdrawal.

What Type of Personal Information Will Be Collected?

Should you agree to participate, you will be asked to provide basic demographic information, as well as your experience and knowledge of both citizen science and mobile devices (smartphones, PDAs, etc).

During the sessions, we will video tape and/or audio tape and/or take notes about your comments, and as we observe your interactions with our designs. If you interact with our designs through a mobile device, we will record mobile device location/GPS data, touch data, application 'use time' and citizen science data you collect (including photographs, inputted values, text, audio, video, etc). The video and audio tapes will only be seen by the researchers, unless you provide permissions to publish excerpt of them as described below. If extracts of your comments are published, they will be referred to by pseudonym.

We will capture your comments and interactions during interviews and discussions via audio, photographs and/or video recording. Depending upon the permissions granted, we will use extracts of these to illustrate our findings in research publications and presentations.

I grant permission for:

- video excerpts to be used in publications and / or presentations: Yes: ___ No: ___
- audio excerpts to be used in publications and / or presentations: Yes: ___ No: ___
- photographs to be used in publications and / or presentations: Yes: ___ No: ___
- written or verbal comments and answers to be used in publications and / or presentations Yes: ___ No: ___

"If you have provided permission for the public display of video-recorded or photographed images, no meaningful anonymity can be provided and you will be widely recognizable as a study participant"

Are there Risks or Benefits if I Participate?

If the study is done outdoors, you will be in a situation equivalent to walking in an urban field, woodlands, or park.

If you participate you will be given refreshments, have your parking costs paid for and be paid for longer-distance travel costs (up to $30) upon study completion. You will learn about and contribute to new developments in the conducting of citizen science projects.

What Happens to the Information I Provide?

Only the researchers will have access to the full recordings and responses that you provide. This information will be kept in a secure location (locked cabinets and password-protected discs). The information that we collect will not be associated to you personally. However, the researchers will publish the results of their analysis of your data in anonymized form in academic journals, conference papers and Master's Degree projects, depending upon the types of explicit permissions provided above.

The researchers might quote the responses in the questionnaires or any of your comments in anonymized form, and they may use still images, videos and audio taken during the interaction in research presentations and publications. Please note, that once images / videos are displayed in any public forum, the researchers will not have any control of any future use by others who may copy these images and distribute them in other formats or contexts.

All the collected data will be kept by the investigators for at least a year, where it will be destroyed after it is no longer required.

Signatures (written consent)

Your signature on this form indicates that you 1) understand to your satisfaction the information provided to you about your participation in this research project, and 2) agree to participate as a research subject.

In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. You are free to withdraw from this research project at any time. You should feel free to ask for clarification or new information throughout your participation.

Participant's Name: (please print) ____

Participant's Signature ____Date: ______________

Researcher's Name: (please print) ____

Researcher's Signature: ______________________________________Date: _______________

Questions/Concerns

If you have any further questions or want clarification regarding this research and/or your participation, please contact:

Matthew A. Dunlap Department of Computer Science 403 999 7148, matthew.a.dunlap@gmail.com	Anthony Tang Department of Computer Science 403 210 6912, tonyt@ucalgary.ca	Saul Greenberg Department of Computer Science 403 220 6087, saul.greenberg@ucalgary.ca

If you have any concerns about the way you've been treated as a participant, please contact the Senior Ethics Resource Officer, Research Services Office, University of Calgary at (403) 220-3782; email rburrows@ucalgary.ca.

A copy of this consent form has been given to you to keep for your records and reference. The investigator has kept a copy of the consent form.

Appendix B. Ethics Approval

UNIVERSITY OF CALGARY

CERTIFICATION OF INSTITUTIONAL ETHICS REVIEW

This is to certify that the Conjoint Faculties Research Ethics Board at the University of Calgary has examined the following research proposal and found the proposed research involving human subjects to be in accordance with University of Calgary Guidelines and the Tri-Council Policy Statement on *"Ethical Conduct in Research Using Human Subjects"*. This form and accompanying letter constitute the Certification of Institutional Ethics Review.

File no:	**7412**
Applicant(s):	**Saul Greenberg** Matthew A. Dunlap Anthony H. Tang
Department:	**Computer Science**
Project Title:	**Utilizing Geocaching and Mobility in Citizen Science**
Sponsor (if applicable):	**NSERC**

Restrictions:

This Certification is subject to the following conditions:

1. Approval is granted only for the project and purposes described in the application.
2. Any modifications to the authorized protocol must be submitted to the Chair, Conjoint Faculties Research Ethics Board for approval.
3. A progress report must be submitted 12 months from the date of this Certification, and should provide the expected completion date for the project.
4. Written notification must be sent to the Board when the project is complete or terminated.

John Ellard, PhD
Acting Chair
Conjoint Faculties Research Ethics Board

AUG 14 2012
Date:

Distribution: (1) Applicant, (2) Supervisor (if applicable), (3) Chair, Department/Faculty Research Ethics Committee, (4) Sponsor, (5) Conjoint Faculties Research Ethics Board (6) Research Services.

UNIVERSITY OF CALGARY

MEMO

CONJOINT FACULTIES RESEARCH ETHICS BOARD
c/o Research Services
Main Floor, Energy Resources Research Building
3512 - 33 Street N.W., Calgary, Alberta T2L 1Y7
Telephone: (403) 220-3782
Fax: (403) 289 0693
Email: rburrows@ucalgary.ca
Tuesday, August 14, 2012

To: Saul Greenberg
Computer Science

From: Dr. John Ellard, Acting Chair
Conjoint Faculties Research Ethics Board (CFREB)

***Re:* Certification of Institutional Ethics Review:** Utilizing Geocaching and Mobility in Citizen Science

The above named research protocol has been granted ethical approval by the Conjoint Faculties Research Ethics Board for the University of Calgary. Enclosed are the original, and one copy, of a signed **Certification of Institutional Ethics Review**. Please note the terms and conditions that apply to your Certification. If the research is funded, the sponsor should be notified, and the original certificate sent to them for their files. The copy is for your records. The Conjoint Faculties Research Ethics Board will retain a copy of the Certification on your file.

Please note, an annual/progress/final report must be filed with the CFREB twelve months from the date on your ethics clearance. A form for this purpose has been created, and may be found on the "Ethics" website, http://www.ucalgary.ca/research/ethics/cfreb

In closing let me take this opportunity to wish you the best of luck in your research endeavor.

Sincerely,

Russell Burrows
For:
John Ellard, PhD
Acting Chair, Conjoint Faculties Research Ethics Board

Enclosures (2): Matthew Dunlap & Dr. Anthony Tang (Co-applicants)

Appendix C. Affinity Diagramming Themes Details

Below is a detailed breakdown of the themes and subthemes that came as a result of affinity diagramming. These include themes that were not discussed due to their lack of relevance to the work (e.g. politics in citizen science).

- Discussion on Targeted Problems
    - Data Collection
        - Well Structured
            - Automatic GPS recording
            - Verbose data entry
            - Automatic data upload
        - Finding Sites
            - Physical cache use
                - Enjoyable
                - Previously used
                - Too much work
        - Creating Sites
            - New Idea: Collecting transient phenomena
        - Transferring Tools and Samples
            - Tool storage
            - New Idea: Caches specifically for sample storage
    - Data Validation
        - General support of more validation
        - Solving bias

            - Some bias is constant
            - Informing citizen scientists of their bias
    - Volunteer Training
        - "Training is awesome"
        - Real-world education important
    - Volunteer Coordination
        - Computer-Assisted Coordination
            - Identifying keen participants
            - New Idea: digital collections as recruitment for more difficult collections with scientist
            - Different levels of coordination
                - Hands-on (Scientist centered)
                - Hands-off (Interface/community centered)
                    - These have the potential to operate longterm
        - Enriched Real-Time Interaction
            - Direct communication
                - Verbose channels such as audio and video
                - New Idea: Coordinating in-person training with mobile devices
- Citizen Science as a Social Experience
    - Performing Tasks as a Group
        - Necessity
        - Efficiency
    - Social Motivations of Citizen Scientists
        - Social gets people to do activities they wouldn't otherwise
            - Older age group does citizen science specifically for social
        - Combining volunteering and social events
        - Adding citizens science to social networking
    - Sharing Scientist Knowledge

        - Scientists want to share their story
        - Changing public behaviour and addressing problem situations
        - Citizen scientists motivated by direct connection to scientists
        - Idea: Sharing scientist knowledge through media on demand at caches
- Practical Deployment of Physical Caches
    - Value of Physical Caches
        - Caches are Only Valuable for Multiple Visits
            - Already used by participants
        - Use of Caches as Containers
            - For specific projects
            - For training
            - For storing samples
    - Problems with Physical Caches
        - Effort to Deploy Caches
            - May be too much work
            - Caches may be created by citizen scientists
        - Environmental Impact
            - Non-intrusive environmental interaction important
            - Caches have impact
            - Caches aren't visually impacting like marking tape
        - Cache Tampering
            - Humans stealing caches
            - Animals destroying caches
    - Project Areas Amenable to Physical Caches
        - Maintaining Physical Objects
            - Remote cameras
            - Fences
        - Long-term Site Use
            - Repeat photography

            - Pika Nest monitoring
        - Training Caches for Different Applications
            - Data collection projects
            - Education-based citizen science
            - Teaching school groups
- Citizen Science Politics
    - Scientists don’t always want to do citizen science
    - Citizen science projects are a ploy for funding
- Basic technology issues
    - Older generation less likely to be comfortable with mobile devices
    - Critical mass needed for public application
    - Digital systems are hard for scientists to manage
    - Mobile technology can be damaged
- Other citizen science problems
    - Rewards motivate citizen scientists
    - Feedback important for citizen scientists
        - To know their time spent is worthwhile
        - To be engaged in the project