Generating Trail Conditions Using User Contributed Data Through a Web Application

by

Alvin Yeung

A Thesis Presented to the
Faculty of the USC Graduate School
University of Southern California
In Partial Fulfillment of the
Requirements for the Degree
Master of Science
(Geographic Information Science and Technology)

December 2017

To my family, who have supported me throughout this process.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am grateful to my advisor, Professor Kemp, for all her assistance throughout this entire process, as well as my other committee members. I would like to thank the City of Santa Barbara for providing the data used in this project. I would also like to thank all the hikers for their participation.

List of Abbreviations

AND	Automotive Navigation Data
API	Application Programming Interface
ArcMap	ArcGIS for Desktop
CC BY-SA	Commons License Attribution-ShareAlike
CSS	Cascading Style Sheets
DWG	Data Working Group
GIS	Geographic information system
HTML	Hyper Text Markup Language
LAFD	Los Angeles Fire Department
ODbL	Open Data Commons Open Database License
OSM	OpenStreetMap
PGI	Professional Geographic Information
RDBMS	Relational Database Management System
RMSE	Root-mean-square-error
USC	University of Southern California
VGI	Volunteered Geographic Information

Abstract

Hiking trails are subject to change over time due to several factors, manmade and natural. Changes in trail condition pose a danger to unprepared hikers. Collecting data about trail conditions in rough and remote terrain can be difficult and expensive. Gathering information through volunteers can be a cheap and effective alternative. In this project, hiking trail locations and features were obtained from the City of Santa Barbara. A web application and website were developed as a proof of concept to show users existing data for hiking trails in Santa Barbara and to collect volunteered geographic information (VGI) about trail conditions from users onsite. Three different analysis methodologies were performed on the VGI to determine its validity. A positional analysis of the features submitted presented clustering of the VGI towards the original feature point, a thematic analysis revealed a general consensus on the condition of a feature, and a comparison of the VGI data against control points using photographs taken with a mobile device exhibited variances in the distance from the original feature. The VGI data was then modified into the same format as the City of Santa Barbara's dataset to demonstrate how it could be prepared for submission to the city. The results of this project show that a website and web application can be used to collect VGI on hiking trails. The VGI data collected can be used to update the City of Santa Barbara's hiking trails dataset.

Chapter 1 Introduction

Providing information on trail conditions allows hikers to prepare before they start on their journey. Accurate, updated trail information can be generated by users who are or were physically present at the site. The objective of this project was to create a prototype web application to collect volunteered geographic information (VGI) to that could be used to update the current data set of an administrative area, in this case, the City of Santa Barbara. The web application developed in this project is a proof of concept, demonstrating that it is possible to apply this web application in order to collect information about trails.

1.1. Motivation

Because hiking can be a dangerous activity, the motivation behind developing this application is a desire to keep people safe on hiking trails by providing the most up to date information regarding trail conditions from people who have traversed the trail recently. Such an application can improve the overall safety of hiking and allow more people to go onto a trail better prepared. The benefit such a web application can provide for a hiker is up-to-date information. Rather than being limited to a specific route with a limited number of features as shown on a traditional trail map, with the flexibility that geographic information systems (GIS) provide, hikers can interactively view several trails in the same interface, update trails, and plot new features that are not specified in an application's legend.

The objective of this project is to provide a tool in which VGI about hiking trails and their features can be crowd-sourced to update the City of Santa Barbara's trails dataset. While crowd sourcing is one method of VGI data collection and it is quickly becoming a viable alternative to more costly, traditional methods, there are concerns with regards to the quality of such data. If the methodology for gathering and validating the VGI is not up to par, then the results will not be acceptable. Thus, this project also makes some initial attempts at building a framework for assessing the quality of collected data.

1.2. Objective

As stated above, the objective of this project was to create a prototype web application to collect hiking trail feature VGI that can be used to update the current official trail data held by the City of Santa Barbara. The project has two sub-objectives:

1. Create a prototype web application to display official trail data and collect related VGI,
2. Demonstrate how this VGI can be validated.

1.3. Application Overview

This section provides an overview of the intended users for this project, this project's study area, and the application design used to develop the web application.

1.3.1. Intended Users

As this project designed a web application to collect VGI on hiking trails, the intended users of this application are hikers. Since the application is developed to be accessible to all users, the amount of hiking experience of the intended users ranges from experienced to beginner. Allowing a greater number of users the ability to use this web application is beneficial to the VGI collection since the more users engaging with the same feature, the more accurate the data collected can be (Goodchild and Li 2012).

1.3.2. Study Area

The study area of this project is the Santa Barbara front country area. The trail data used in the development of this application contains trail location and trail feature data for eleven different trails in the front country of Santa Barbara.

1.3.3. Application Development

The web application used for this project was developed using Esri's ArcGIS Online, specifically using the Web AppBuilder tool. However, first, the hiking trail data was input into Esri's ArcMap to more easily create appropriate symbology for the trails and trail features. Once the design of the map was satisfactory in ArcMap, the hiking trail data was published and hosted on ArcGIS Online. Next, a web map was created from the published hiking trail data. Then the web application was created based on the web map using Esri's Web AppBuilder. The web application was then embedded on a website designed and created by the author, hosted on the University of Southern California's (USC) servers. The user can interact with the application through the website. Figure 1 below visualizes this process.

The user is able to view and query all features and trails displayed in the web application through the website. The user is able to see other user submissions. The main functionality of this web application is to allow the user to submit their VGI through the web application. Validation of the VGI was done through accessing the data in ArcMap after the users submitted their VGI through the web application.

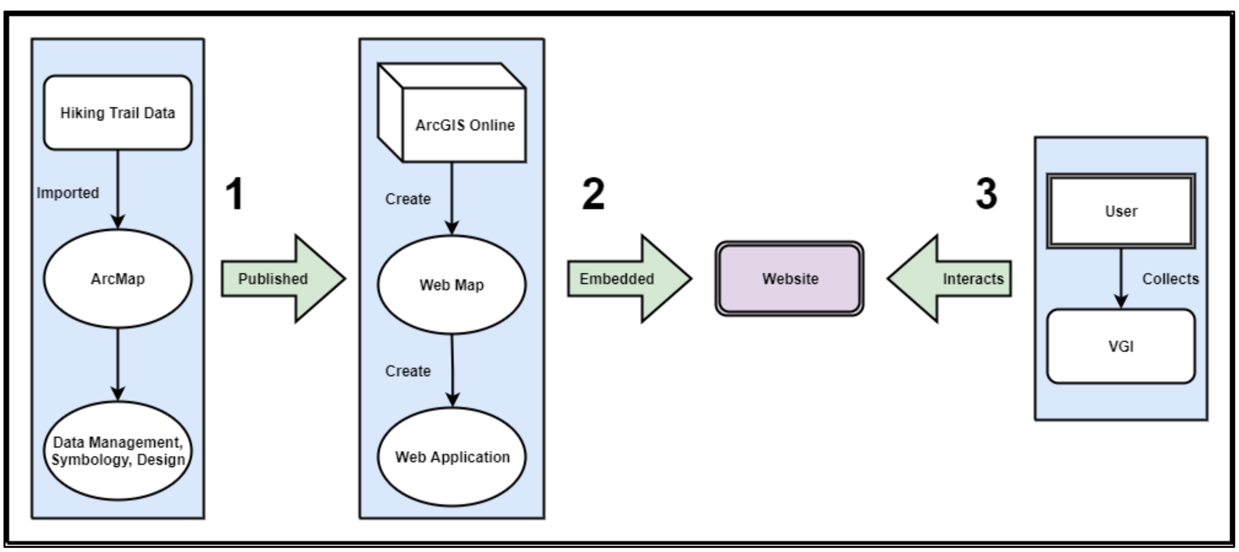


Figure 1 Workflow diagram

1.4. Thesis Organization

This thesis is divided into six chapters. Chapter 2 describes the research related to this project. In particular, Chapter 2 delves into VGI and how it has been used in other, similar studies. Chapter 3 states the functional requirements and design principles of this project. Chapter 4 discusses how the tools used in this project were developed. Chapter 5 evaluates how well the web application achieved this project's objectives. Chapter 6 discusses how this project performed relative to its objectives, any issues encountered during this project's duration, any future improvements that can be made, and the conclusions of this project.

Chapter 2 Related Work

Crowdsourcing data is a prominent method for data collection in recent times. Popular websites such as OpenStreetMap (OSM) and Wikipedia rely on volunteers for their information. Crowdsourcing is an alternative to more expensive traditional data collection methods. OSM is a platform that is similar to this project, where volunteers can submit their VGI through their website. OSM, however, focuses their efforts mainly on streets, roads, and areas that are more easily accessible than the hiking trails that are the focus of this project. While there are various web and mobile applications available related to hiking, most do not allow users to submit VGI about a specific trail.

This chapter provides background on VGI and various methods for validating it, and reviews some web applications that collect and use VGI. It also provides information on design principles for web pages and web applications.

2.1. Volunteered Geographic Information (VGI)

"VGI is the harnessing of tools to create, assemble, and disseminate geographic data provided voluntarily by individuals" (Wikipedia 2017). VGI uses public volunteers to gather information and contribute facts that are synthesized into databases (Goodchild and Li 2012). The use of VGI can improve data accuracy and accessibility, as discussed in various sections in this chapter. However, various laws and policies, either organizational or governmental, affect how the spatial data is integrated into existing infrastructures and how the public accesses such data (Elwood 2008).

VGI can be seen as a special case of the web phenomenon of user generated content, where users submit geographic information through a web interface (Goodchild 2007). The rise of Web 2.0, the emergence of user generated content and interoperability among platforms, enabled greater user generated spatial content (Lin 2013). The volunteers may be untrained but can make reliable contributions. Flickr allows users to upload and geotag their photographs with latitude and longitude coordinates. At the time of this writing, there are over 100 million Flickr users who have shared over 10 billion photographs (Smith 2016).

Collaborative mapping is another popular Web 2.0 activity. Two well-known sites that collect VGI from volunteers are OpenStreetMap (OSM) and Wikimapia. Similar to Wikipedia, Wikimapia is a collaborative map of the world. According to their website, the goal of Wikimapia is "to describe the whole world by compiling as much useful information about all geographical objects as possible." Like all collaborative compendia, a key feature about Wikimapia is that it is constantly changing, their data being updated by the users. The data is geo-located and traced over Google Maps imagery. Wikimapia is licensed under a Creative Commons License Attribution-ShareAlike (CC BY-SA). Due to the use of Google Maps imagery, the data can be seen as derived work, which has raised legal concerns in the past (OpenStreetMap 2017). OSM is also a free, editable map of the world built by volunteers, licensed under the Open Data Commons Open Database License (ODbL). Due to potential legal issues with Wikimapia, this paper focuses in this chapter on OSM. Section 2.2 goes in depth on OSM.

By design, VGI may be collected repeatedly or continuously. Multiple users can contribute information for the same object under differing circumstances (Jiang and Thill 2015). Goodchild (2007) reports on a geographic feature, Ayer's Rock, from Flickr, with more than 2,500 photos submitted at the time of his writing. Eleven years later, at the time of writing this paper, there are more than 6,000 photos found when searching for "Ayers Rock" on Flickr.

VGI can impact the way spatial data is produced and shared (Elwood 2008). Services such as OSM and Wikimapia provide new ways for the public to not only access but also to produce spatial data. By making it possible for more users to contribute VGI, the volume of spatial data increases. However, Elwood argues that while the volume of spatial data increases, there may be fundamental differences between VGI and expert created data. In particular, VGI may show more heterogeneity, due to the diversity participants and their knowledge.

2.1.1. VGI for Use by Public Agencies

Traditionally, geographic information production has been prepared and published by public sector agencies (Goodchild 2008). This phenomenon can be broadly defined as professional geographic information (PGI). Goodchild states that in the past, acquiring, compiling, and publishing geographic information has been expensive. Only certain publishing agencies and centralized governments have been able and willing to absorb the costs.

VGI can provide a cost-effective alternative to traditional data collection methods due to its use of volunteers. However, there are challenges for governments adopting VGI. VGI, data collected by volunteers, fundamentally differs from PGI, data collected by experts in the subject matter, that is often used by governments (Goodchild 2007). The difference is that VGI is voluntarily submitted by individuals without formal qualifications in that field whereas PGI is generally collected within a formalized framework, often by an expert in the topic, as part of their paid work. One major concern of government agencies in utilizing VGI is establishing the source of the data (Johnson and Sieber 2013). Because VGI is submitted by individuals, there is often an anonymous nature in the data. Government agencies may not be able to fully verify who submitted the data and thus may be reluctant to assure its validity and quality. Compared to PGI, where the source is undoubted, VGI data quality is a key concern in its use. The next section explores the question of spatial data quality in general.

2.1.2. Elements of Spatial Data Quality

Quality can be defined by the product's ability to satisfy stated and implied needs (van Oort 2006). By examining several sources of information about spatial data quality, van Oort outlined the key elements of spatial data quality. These sources included the 1992 US Spatial Data Transfer Standards, the 1998 European standards, and the 2002 ISO standards as well as publications by Aronoff (1989) and Guptill and Morrison (1995).

The US Spatial Data Transfer Standard (USA-SDTS) contains a section on spatial data quality elements. These standards were accepted by the Department of Commerce in 1992 and incorporated into the USA metadata standard. Aronoff (1989) provides an interpretation of a draft of the USA-SDTS from a management perspective. Guptill and Morrison (1995) edited a book on behalf of the ICA, the International Cartographic Association, titled *Elements of Spatial Quality*, with contributions from several authors. In 1998, a European Union technical committee (CEN/TC 287) published standards for spatial data quality to be used by European national mapping agencies. Finally, following these earlier efforts, an ISO technical committee (ISO/TC211) published the international metadata standards for geographic information in 2002.

According to van Oort, there are 11 spatial data qualities found in these five sources. These are listed in Table 1 below. An "S" or "M" value indicates the source explicitly recognizes the element as part of spatial data quality while an "I" value implicitly recognizes the element. Of the 11, three are particularly relevant to this project. These are positional accuracy, temporal quality, and completeness. Each of these is discussed in the following paragraphs.

Table 1 Elements of spatial data quality in five sources (Source: van Oort 2006)

Element	Aronoff (1989)	USA-SDTS (1992)	ICA (1995)	CEN/TC 287 (1998)	ISO/TC211 (2002)
Lineage	S	S	S	S	S
Positional Accuracy	S	S	S	S	S
Attribute Accuracy	S	S	S	I	S
Logical Consistency	S	S	S	S	S
Completeness	S	S	S	S	S
Semantic Accuracy			S	S	
Usage, purpose, constraints	S	M		S	M
Temporal quality	S	M	S	S	S
Variation in quality		I	I	S	I
Meta-quality		I	I	S	I
Resolution	S	I	I	I	M

The positional accuracy of spatial data is defined by van Oort as the accuracy of its coordinate values. There are two different types of positional accuracy, relative and absolute. Relative accuracy refers to the accuracy in relation to other data in the same dataset while absolute accuracy refers to the accuracy of the coordinate values in relation to a reference dataset. If datasets are combined, van Oort comments that the absolute positional accuracy is necessary.

Temporal quality is the quality of spatial data in respect to time. There are many elements to temporal quality. One element of temporal quality is the accuracy of time measurements, defined by CEN/TC287 and ISO/TC211 as the summary of errors in time measurements. The rate of change in the phenomenon is also a key element of temporal quality. The rate of change is defined by CEN/TC287 as "…an estimate in the rate of change in the phenomenon represented in the data" (van Oort 2006, 17). Knowing the rate of change of an object can inform a user about when should have been the latest update on an object as well as the validity of the information recorded. Temporal validity is another key element in temporal quality, defined by

CEN/TC287 through three values: “out_of_date,” “valid,” or “not_yet_valid.” Together, the elements of temporal quality can inform a user about the expected change over time in a phenomenon in a data set.

Completeness is another key element in spatial data quality, defined by van Oort as “…the measure of the absence of data and presence of excess data” (van Oort 2006, 15). An important factor in determining the completeness of data is knowing what the producer intended to collect in the dataset. To determine the completeness of data, according to the USA-SDTS, the producer must provide “…selection information about selection criteria, definitions used and other relevant mapping rules. For example geometric thresholds such as minimum area or minimum width must be reported” (van Oort 2006, 15).

2.1.3. Assuring the Quality of VGI

Due to potentially less stringent methods of data collection and the frequent use of untrained non-specialists, there are concerns with the quality of VGI data. Quality assurance methods are needed in order to ensure that data collected are useful. Goodchild and Li (2012) identified three approaches to VGI quality assurance: the crowdsourcing approach, the social approach, and the geographic approach.

2.1.3.1. Crowdsourcing Approach

According to Goodchild and Li, crowdsourcing has two distinct meanings. The first meaning comes from the origin of the term which combines crowd and outsourcing. Crowdsourcing provides an opportunity to find a solution to a problem by referring the problem to any number of people, regardless of their credentials (Goodchild and Li 2012). The idea is that this approach may yield a solution to a problem that even experts may not see.

One example of a service for outsourcing tasks to a crowd is Amazon's Mechanical Turk (www.mturk.com). Amazon's mturk program allows users to complete tasks that are outsourced by businesses for monetary compensation. Through this service, businesses or individuals outsource work that requires human intelligence.

The second meaning of crowdsourcing given by Goodchild and Li is more related to quality assurance. It refers to the ability of a crowd to validate and correct errors an individual might make. This is demonstrated by a study by Latonero (2010) that examines how social media platforms, specifically Twitter, can be used by emergency service professionals. Working in conjunction with the Los Angeles Fire Department (LAFD), Latonero analyzed Twitter content for reports on fire. Goodchild and Li observed that Latonero's study was relevant to their interpretation of crowdsourcing as a means of quality assurance. They noted that Latonero's study gave greater weight to a spatial clustering of similar reports rather than to a single report. The public crowd of the Twitter users was able to validate other users' reports on a fire emergency for the LAFD.

The second meaning of crowdsourcing is also defined by Goodchild and Li as "the ability of the crowd to converge on the truth." If an individual contributes an error, others can be expected to correct the error. This scales with the number of users per contribution.

A study by Wilkinson and Huberman (2007) demonstrates how others correct an error by analyzing Wikipedia articles and edits. Wikipedia presents a great platform to study the ability of a crowd converging on the truth, as the site contains a large number of users. The average number of edits on the most popular articles on Wikipedia labelled as "featured" articles, were compared to all other articles. Featured articles are reviewed by Wikipedia's editors for their accuracy, neutrality, completeness, and style. At the time of Wilkinson and Huberman's study, there were 1,211 featured articles. Article age and visibility were taken into account, as older, more popular articles have more edits. To control for the article's visibility through Google, the article's Google PageRank, Google's algorithm to rank web pages in their search results, was used. Similar Google PageRanks were grouped together before comparison. To control for visibility through Wikipedia, as some featured articles gain visibility by appearing on Wikipedia's front page, edits during that period were discounted. To control for age, the number of edits to an article was normalized by the mean and variance for all articles of that age.

Though the authors do not explicitly state how it was measured, they concluded that as articles on Wikipedia increase their number of edits and distinct editors, the article's quality increases. While a small number of articles accumulates a disproportionately large number of edits, on average, the more edits an article accumulates, the higher quality it will be.

Crowdsourcing presents a unique approach to data collection that requires the participants' assistance. If people are interested in the problem, they are more willing to assist in the resolution. It may be to get people on board with the task, it is only a matter of presenting the right question to them (Corney et al. 2010). While crowdsourcing presents a method of quality assurance, Goodchild and Li conclude that it can be less effective for geographic facts compared to other types of information. This is because errors in geographic facts can persist, even in a large crowd, due to the variance of knowledge.

2.1.3.2. Social Approach

A second approach to VGI quality assurance is the social approach, sometimes called the hierarchy approach. The hierarchy approach requires a few trusted high level users making the bulk of the contributions and many other lower level users making one or a few contributions (Goodchild and Li 2012). The few high-level users are generally moderators or administrators with certain privileges over general users. In the case of Wikimapia, a high-level user, called an "advanced user" is given additional tools and responsibilities. The tools and responsibilities for high-level users of other systems can vary based on their needs. However, the privileges of the high-level users allow the users to supervise the contributions made by lower level users. This process has been integrated into existing websites, such as Wikipedia and OSM, as it allows the data to maintain a certain level of quality.

Goodchild and Li report that Wikipedia has a group of high-level users called "sysops" who have privileges normal users do not. Those privileges include deleting articles, protecting pages from editing, and blocking users (Goodchild and Li 2012). Goodchild and Li also report on the hierarchy approach used by OSM that incorporates ordinary users and the Data Working Group (DWG). Ordinary users can add and edit geographic features, however, if any issues arise, such as vandalism, copyright violation, disputes, etc., the DWG are the people to resolve the issue. Both systems of Wikipedia and OSM maintain a level of quality, based on the efforts of the high-level users.

2.1.3.3. Geographic Approach

The third approach to VGI quality assurance is the geographic approach. It is described by Goodchild and Li as a comparison of geographic facts within the rules of geography. The rules of geography define what can and cannot occur at a given location, defined by either scientific research or regulations of a governing body. The rules can be very simple, such as basic geometry defining what is possible, or abstract, such as the first law of geography by Tobler (1970). Due to a large number and varying type of rules about geographic locations, both from a scientific and regulatory agency standpoint, there is currently no framework or conceptual idea to house all the rules. Thus, Goodchild and Li conclude that while the geographic approach has promise, there is currently no effective method to implement the geographic approach.

2.2. Assessing VGI Data Quality

Because many VGI data collectors are untrained in specific collection protocols, there are concerns with the quality of data gathered. However, Goodchild and Li (2012) suggest that data quality of user generated data is quickly becoming a viable alternative to costly data collection methods. This section discusses how VGI quality is assessed, the positional accuracy of VGI, how VGI can be combined with existing data and potential issues with the quality of VGI.

2.2.1. Completeness of VGI

One method of assuring data quality is by determining the completeness of the data. Zielstra and Zipf (2010) conducted a study where user generated street data from OSM was compared to the Tele Atlas MultiNet proprietary data to determine the completeness of OSM's VGI data. Tele Atlas MultiNet is a commercial information provider of digital street networks who state that their data is "unrivaled in its depth, richness, and accuracy" (TeleMart 2017). The study compared the amount and accuracy of the data generated by OSM against Tele Atlas MultiNet's dataset. This was done by calculating and comparing the overall lengths of streets from each dataset. The results showed that it is possible to collect detailed data from user input, as long as the user base is large enough. They concluded that in densely populated urban areas, where there is a large potential user base, VGI could be a cost-efficient alternative to proprietary data sets.

In a similar vein to proprietary data, control data can be used to determine the quality of VGI. Comber et al. (2013) conducted a study on land cover classification comparing user submitted classification to control points to determine the reliability of VGI. The control points that the VGI was compared to were created by experts. The project asked its participants to classify land cover into ten different classes. The data was filtered to exclude participants who did not submit more than twenty data points, to ensure sufficient data to for reliable data. The results of this method show an overall accuracy of the VGI compared to the control dataset to be 62%, with a caveat that two land cover classes were missing from the control dataset.

Comber et al.'s study also compared the VGI to three existing land cover databases, the GLC-2000, GlobCover, and MODIS. The GLC-2000 is a land cover database developed by the European Union, GlobCover is a land cover database developed by the European Space Agency, and MODIS is a land cover database developed by NASA. Comber et al. concluded that the spatial variations between each database affected the accuracy of the VGI. One dataset may be more accurate over another, depending on the type of data and location of interest.

Based on the results of their study, the reliability of the assessment of VGI in this scenario is dependent on the quality of its control data. If the control data is inaccurate or of poor quality, then the VGI assessment can be unreliable.

2.2.2. VGI Positional Accuracy

The positional accuracy of VGI is also a concern, as the accuracy of the general population's GPS devices (most commonly smart phones) will not be as accurate as professional equipment. As such, Haklay (2010) conducted a study to compare the positional accuracy of volunteered OSM data to the data from the Ordnance Survey of Great Britain, the national mapping agency. The study notes a couple of caveats, such as the dataset cannot be more accurate than the quality of the GPS receiver, with a margin of error between six to ten meters, and the resolution of Yahoo! aerial imagery, which provides a fifteen-meter resolution. The Yahoo! aerial imagery dataset was used due to the Ordinance Survey dataset used in this study only covering London. Based on those parameters, Haklay concludes that the expected accuracy of each dataset, the VGI submissions, and the Yahoo! aerial imagery, to be within twenty meters from its true location, under ideal conditions.

Haklay's study found an average overlay of 80% between the OSM VGI dataset and the Ordinance Survey and Yahoo! aerial imagery dataset. The results showed some variance in the percentage overlay, ranging from 60% to 89%. Haklay concludes that there is variance between the OSM dataset and the London Ordinance Survey and Yahoo! aerial imagery dataset. However, Haklay also concludes that there is a guarantee that the feature will be within the given range of displacement stated by the map provider. The guarantee is based on the trust of the provider and its quality assurance practices.

Goodchild and Li (2012) also commented on Haklay's study that there was a variation in positional accuracy and completeness of the data, based on geographic location. However, they note that even with the variation, the data gathered by the public today is more accurate than many authoritative data as they are becoming increasingly out of date as they were acquired using older technologies that were less accurate than those available today.

2.2.3. Using VGI to Enhance Datasets

With newer technologies, VGI can be used to update and enhance existing datasets. See et al. (2015) conducted a study where non-expert VGI data was used to enhance an expert dataset on land cover. The three expert global land cover sources are the GLC-2000, MODIS, and GlobCover mentioned previously. The study used Geo-Wiki, a visualization, crowdsourcing, and validation tool for improving land cover data to create a hybrid land cover map as its non-expert dataset.

See et al. (2015) ran crowdsourcing campaigns in order to gather VGI data for their study. The campaigns produced several VGI products. One of the products from the crowdsourcing campaign was the development of a hybrid land cover map. See et al. took the VGI gathered from one of its crowdsourcing campaigns and used geographically weighted regression to determine the best land cover product at each grid cell, using both the VGI and the expert sources. The resulting map was compared to an external dataset developed specifically for the validation of the Chinese 30m global land cover map. The Chinese land cover map is at a much higher resolution than the three expert sources listed above. The validation method used was an equal-area stratified random sampling. This method divided an area into 7000 equal-area hexagons, where five random samples were taken from each hexagon. The comparison between the hybrid land cover map and the individual maps against the Chinese 30m validation dataset reveals that the hybrid land cover map outperformed the individual land cover maps based on the performance measures used in this study.

2.2.4. Issues with VGI Quality

Even with the improvements in technologies and the validation methods listed in the studies above, there still can be issues with the quality of VGI. One issue with VGI can be the number of users that can contribute. Zielstra and Zipf’s 2010 study discussed in Section 2.2.1 shows that in medium sized towns, such as Zwickau and Marburg, the amount of VGI collected was less than proprietary sources. This can be attributed to a lower overall population compared to the large towns used in this study. It can be inferred that in smaller towns than Zwickau and Marburg, the two towns that collected less than 10% more data than Tele Atlas MultiNet’s dataset, there may not be enough participants to collect sufficient VGI for comparison to a proprietary source.

Another potential issue with VGI quality assessment can be the quality of control data used for validation. As seen in the 2013 land cover classification study done by Comber et al. discussed in Section 2.2.1, the VGI achieved an accuracy of only 62%. This may be due to the control data used in the study missing two of the possible ten land cover classes. The accuracy of the VGI might have been improved with the inclusion of control points for the two missing classifications.

The accuracy of the equipment is also a concern when dealing with VGI. The accuracy of VGI cannot be more accurate than the accuracy of the equipment used. As Haklay notes in his comparison of OSM to the London Ordinance Survey dataset, the accuracy of GPS units has a tolerance of six to nine meters. While GPS units may have improved since the study, as the study took place in 2010, there is still a margin of error in GPS units. The inaccuracy of the equipment used can lead to an inaccurate dataset in VGI.

There is also the issue that VGI collected by volunteers generally does not adhere to stringent data-collection standards (Haklay 2010). The quality of the VGI heavily relies on the quality of the equipment and the quality assurance methods. If the equipment or validation methods used are subpar, then the VGI quality can be subpar.

2.3. OpenStreetMap (OSM)

OSM is a major international project that provides user generated street maps, focusing mainly on urban areas and roadways (Haklay and Weber 2008). It has been critically important in several major relief efforts including helping after the earthquakes in Haiti and Nepal (Humanitarian OpenStreetMap Team 2017). The goal of the OSM project is to create a free to use and edit set of map data. Originally conceived at the University College London in 2004, the motivation for the project is to provide free data in locations where data collection would be expensive or otherwise prohibitive.

An analysis of OSM data by Heipke (2010) shows that OSM has 80% of the road coverage in London mapped out, with an error tolerance of six meters. On a national scale, 30% of England had been mapped by volunteers in a four-year span from 2004 to 2008, with a noticeable difference in road coverage between cities and rural areas, due to a lower user base in rural areas. OSM and similar major projects have been very successful in collecting lots of high quality data due to its many members who can offer a complex diversity of spatial data (Zielstra and Zipf 2010).

OSM only allows registered users to edit data. However, users can view all VGI submitted, similar to Wikimapia. According to Haklay and Weber (2008), this is because the project leads at OSM wanted the ability to trace the information source should any copyright issue arise.

OSM provides a JavaScript editor as well as an older, flash editor for users to submit contributions. Users are able to add, update, or delete geographic features, as well as upload GPX files from GPS units. The user interface is kept deliberately simple, with more complex features available through keyboard shortcuts (Haklay and Weber 2008). OSM uses predefined tagging schemas for the more frequently occurring features. For example, a primary road is a road classification in the United Kingdom that refers to the major roadway to and from a major urban center.

For more advanced contributors of OSM, a java based editor more akin to traditional GIS packages is available (Haklay and Weber 2008). This application provides the features of the online flash editor offline. Users can import, edit, and tag data offline and bulk upload their contributions to OSM through the OSM application programming interface (API). This java based editor is one of the main tools used for data validation and is the tool OSM lists in their wiki.

OSM uses a two-stage validation method that relies on more advanced users utilizing the offline, java based editor. All steps are done by a higher-level user using the java based editor. A detailed instruction guide can be found on OSM's wiki. The following is a summary of their process.

There are three main steps in the first stage. The first step in Stage 1 is to do a cursory scan of the validation area, a small square area within OSM. The size of the square area is defined by OSM to be about 2.5 kilometers. If it is obvious that the data needs major improvements, such as the majority of the validation area is not mapped, the area is invalidated, and a comment left for the contributor with explanations of what needs to be improved. If the initial scan of the area is acceptable, then a tool within the java based editor called the "validator" is run. The validator will show everything that can be corrected, such as duplicated points or overlapping areas, and the corrections are done by the advanced user in the java based editor. A full list of validations performed by the validator can be found in the validator wiki.

The second step of validation in Stage 1 is to check for buildings. The main focus of this step is to ensure buildings are not missing and are traced properly. The third step of validation in Stage 1 is to ensure the highways are correct. This step entails checking the tags and geometry of the highways, determining the start and end points of highways, and determining if the highway is correctly classified.

Stage 2 of the validation process in OSM occurs over a larger area, usually an entire town or city. Larger cities may be validated in stages, as the dataset is much larger. The validator tool used in Stage 1 is used again here, however, if the errors from the validator in Stage 1 were corrected, there should be minimal errors. After running the validator tool, the VGI is then compared to satellite imagery, to ensure the geometries and start and stop points of roads are correct. The contributor can also use satellite imagery to validate the correct tags for each road. The OSM wiki also lists other validation tools in addition to the validator that contributors can use during the validation process.

While OSM relies heavily on user contributions for their data and validation, it is not their only source of data. A major source of road data for OSM originated from satellite imagery and out-of-copyright maps (Haklay and Weber 2008). Towards the end of 2006, Yahoo allowed OSM to use its satellite imagery to trace roads into their road network. OSM also takes advantage of free geographic information, such as the public domain TIGER information from the US Census Bureau. OSM has also received contributions towards their street map from a commercial entity, Automotive Navigation Data (AND), as well as local governments, such as the Isle of Man's Department for Local Government and the Environment (Haklay and Weber 2008). AND's contributions provided street information on the Netherlands and the Isle of Man's contributions provided geographical information for the entire island of the Isle of Man. In more recent times, several new data sources have contributed data to the OSM network, ranging from the United States Geological Survey to various countries within Europe.

2.4. VGI Text Analysis

Text analysis is an interpretation of a text (McKee 2003). As VGI can contain text, analysis of the text allows the user to better understand how the VGI fits within the context of its submission. There are two general methods used in text analysis, quantitative and qualitative text analysis. Qualitative analysis can provide higher quality data through interviews and focus groups. However a limitation to that method is the smaller sample size due to the availability of subjects to interview or engage (Lai, Li, and Harrill 2013). Quantitative analysis provides its data through statistical inferences from text through pre-defined properties. There are two main types of quantitative analysis: thematic and semantic analysis (Roberts 2000).

2.4.1. Thematic Analysis

Thematic analysis can be seen as finding common themes among text. The methodology used in a thematic analysis counts specific words, phrases, or themes that appear in a text. The key word(s) that define a theme is designated by the user to reflect the meanings in the text. The resulting data provide limited information on the text as this method does not take into account the context of the text (Roberts 2000). However, a thematic analysis can provide insight into the data and the narrative it tells. A semantic analysis, discussed in the next section, considers the context of the text in its analysis.

To perform a thematic analysis, there are six steps: familiarize yourself with the data, generate the initial codes, search for themes, review the themes, define and name the themes, and produce the report (Braun and Clarke 2006). The six steps by Braun and Clarke are described in detail below.

The first step is to familiarize yourself with the data. This is done through reading and re-reading any data collected. If the data is verbal, then the data is transcribed. This step is also where any ideas for potential thematic codes are noted.

The second step is to generate the codes from the data. A code identifies a feature of the data that appears interesting to the analyst that may form a repeated pattern. There can be two different approaches to this method, a theory driven approach or a data driven approach. A data driven approach gathers its codes, and potential themes directly from the data whereas a theory driven approach gathers its codes and potential themes by approaching the data with specific questions in mind.

The third step of the process occurs when all codes have been generated. This process takes a broader view of the data and looks for themes among the codes. The codes are sorted into potential themes. There are different approaches to this step, such as using tables and spreadsheets or using a pen and paper and drawing a flow diagram of all codes and themes. This step can produce many themes and subthemes from the codes generated in the second step.

The fourth step of a thematic analysis can be seen as a continuation of the third step. This step refines the themes generated in the third step, making any changes necessary. There are two levels to this process. The first level involves reviewing all themes generated in the third step and ensuring that all themes cohere meaningfully. If any themes do not cohere, then adjustments are made, either to the theme or creating a new theme altogether for the data. The second level involves the themes in relation to the dataset. This level reviews the themes and whether the themes reflect the dataset as a whole. Themes are adjusted and amended as necessary during this process.

The fifth step defines and names the themes. Defining and refining a theme is described by Braun and Clarke as "identifying the essence of what each theme is about." Each theme should not be too complex or diverse. It is important to strike a balance for each theme. The themes should be consistent, coherent, and follow a narrative. An analysis of each theme is conducted in this process and should produce a story that each theme tells. It is important the themes fit into a narrative about the data in relation to the research question(s) to ensure the themes do not have much overlap. At the end of this stage, each theme can be clearly defined as what they are and are not.

The final step of a thematic analysis is a report on the analysis. The report should provide evidence of the themes within the data and an analytic narrative that illustrates a story within the data, making an argument of the data in relation to the research question.

2.4.2. Semantic Analysis

Similar to thematic analysis, semantic analysis can be seen as finding relationships among themes in text (Jabreel, Moreno, and Huertas 2017). Semantic analysis uses the actual meaning of the words to find themes. The syntactic components of the text are used in defining the themes and the resulting themes mapped into a template for analysis. The analysis is done at the conceptual level, rather than the syntactic level, as the syntactic level is limited to the text itself. Conceptual level analysis allows for the context of the text to be used.

2.5. Existing Trail Applications

As VGI and hiking are not new ideas, there are existing applications that have been developed for these activities. This section describes three of these applications as examples of what is available. The following subsections describe Trestima, a mobile application that utilizes crowdsourcing as its VGI collection method, Trekker, a mobile application that displays trekking trails and its features, and the Los Angeles County Trails website, the model of this project’s website design.

2.5.1. Trestima

Trestima is a smartphone application that collects VGI for forest sample plot measurements using images taken by the phone’s camera (Vastaranta et al. 2015). The application uses images taken with the smartphone’s camera of a specific tree and estimates various attributes such as the tree’s area, diameter, and height using functionality within the application. Vastaranta et al. (2015) conducted a study in which they compared the results of data collected using Trestima against measurements taken with traditional methods using calipers and clinometers. The results of the study found that the quality of data extracted from the pictures depended heavily on the quality of the images taken with the smartphone. The root-mean-square-error (RMSE), the measure of the differences between the value of the estimation based on the smartphone image and the hand measured values, varied from 19.7% to 29.3%. Increasing the number of photos to four marginally improved the RMSE. While there were variances between the VGI and hand measured values, Trestima provides an example of a platform in which users can submit their VGI collected in a non-urban terrain similar to the areas in which hiking trails exist. While Trestima is a smartphone application, the examples of VGI collection established here can be applied to a web application used on a smartphone.

2.5.2. Trekker

Trekker is a smartphone application developed for use on Android devices. Trekking is an activity that takes place on outdoor trails, though the activity is much more rigorous than day hiking. Trekking trips are typically longer, ranging from days to weeks, and in general, more demanding (Mohajeri et al. 2015). Trekker provides users with nearby trekking trails based on their smartphone's GPS location (Marita et al. 2016). It also has the ability to display and filter trails based on various parameters, such as difficulty level, length of a trail, and altitude of a trail to name a few. Trekker's design is a client/server architecture, where the user interacts with the client, such as filtering a trail based on length, and the server responds with an output within the parameters.

2.5.3. Los Angeles County Trails Web Application

The county of Los Angeles, Department of Parks and Recreation, hosts a website to allow users search for hiking trails within the greater area of Los Angeles (https://trails.lacounty.gov/). The website contains a navigation bar near the top of the webpage with several links leading to more information. Below the navigation bar is a section for a featured trail, showing several pictures of the trail in a slideshow. A web application is embedded beneath the featured trail, allowing users to search for trails within the Los Angeles area. Below the web application is an area split into three columns. The left column displays information about the County of Los Angeles Parks and Recreation department, the middle column displays a poll on how the user came across this website, and the right column shows a Twitter feed with the latest tweets that tagged the Los Angeles County Parks and Recreation department. The final section of the website shows contact information for the County of Los Angeles, Department of Parks and Recreation, along with links to its social media.

2.6. Web Application and Website Design

This main goal of this project is delivering VGI through a web application created through Esri's Web AppBuilder. The web application is being hosted on a website designed by the author of this thesis. The design of any web application and its related website is important as the users' satisfaction of both impacts their intention to visit the web page and use the web application again (Belanche, Casaló, and Guinalíu 2012). While this project is only a proof of concept, succeeding with a VGI driven project requires a large userbase, similar to OSM. User satisfaction is important in retaining a userbase. The following subsections discuss the design of the web application and website.

2.6.1. Web Application Design

One key principle in designing a web application is simplicity (Chin 2013). Simplicity in a web application can be defined as easy to use while displaying relevant information. Some applications, such as Foursquare, can attribute its success to its simplicity (Zichermann and Cunningham 2011). Chin's study is one of many that demonstrates simple design principles are important to the end users and its success (Fling 2009). In Chin's study, he asked a sample pool of thirty volunteers from a well-known cycling club of their perspective on a web application design. The results of the study show that the cyclists preferred a straight forward and practical design of the web application. Ease of use was also a major comment made by the volunteers of the study.

As shown by these and other studies, a simple and straightforward design is important in retaining a userbase. This project's web application design focused on simplicity initially with modifications to its design made based on feedback from its participants and ongoing research.

2.6.2. Website Design

Website design is also important in this project, as the web application is hosted on a custom designed website created by the author of this project. The usability of a website can be determined by considering the following aspects: (a) the ease of understanding the structure of a website, its functions, interface, and the contents that can be observed by the user; (b) simplicity of use of the website in its initial stages; (c) the speed with which the users can find what they are looking for; (d) the perceived ease of site navigation in terms of time required and action necessary in order to obtain the desired results; and (e) the ability of the user to control what they are doing, and where they are, at any given moment (Belanche, Casaló, and Guinalíu 2012).

To sum up the aspects of the usability of a website, the general user cares most about the website's ease of use, similar to comments made by the cyclists in Chin's (2013) web application study. Based on Chin's study and the general definition of website usability, this project focuses its website design on simplicity and ease of use, adjusting as needed based on its participants' feedback, as the design of both the website and web application plays a major role in userbase development and retention.

Chapter 3 Requirements

The objective of this project is to create a tool that can collect VGI to update the current data set of the City of Santa Barbara. The tool refers to the web application that is developed through Esri's Web AppBuilder. Section 3.1 describes the goal of the web application, Section 3.2 describes the user requirements for this project, Section 3.3 describes the functional requirements of the web application, and Section 3.4 describes how the website and web application was designed.

3.1. Application Objectives

The main objective of this application is to create a tool that can be used to collect VGI to update the current dataset of the City of Santa Barbara. To collect VGI, the current features from the City of Santa Barbara are displayed through the web application's initial view, showing all trails and modified feature symbology from a web map. Gathering VGI from its users is achieved through a customized widget, one that allows its users to submit their own data. This application will also allow users to query for specific trails and features within the dataset.

Demonstrating how to validate the collected VGI is also an objective of this project. This must be done to ensure the VGI is accurate and valid prior to submission back to the original data provider, the City of Santa Barbara. While this project is only a proof of concept and the data will not be submitted to the City of Santa Barbara, preparing a methodology that can be applied in a real-world submission will be critical for transitioning this project as a proof of concept into a real-world application.

3.2. User Requirements

The main requirement for the user is to have access to the web application before heading out to the trail and potentially on the trail. This requires the user to have access to a computer or a phone with a GPS. Ideally, the user will have access to both. The user would view the web application on a computer before entering the trail and would have access to the web application through a phone to submit data in-situ. If the user does not have access to the web application on site, they can submit their VGI to the web application on a computer at a later time.

3.3. Functional Requirements

A primary functional requirement of this web application is to display the City of Santa Barbara's data and gather VGI. The display of the city's data will be done through the initial extent of the application. The initial extent will show the web map created from publishing the data through ArcMap.

Gathering VGI will be achieved through the "Smart Editor" widget available through Esri's Web AppBuilder. This widget will allow users to submit their VGI data directly to the geodatabase being hosted on ArcGIS.com. This widget will also prevent users from modifying, deleting, or updating the geometry of any existing data.

Querying for trails and features is also a functional requirement for this web application. This is achieved through the "Query" widget in Esri's Web AppBuilder. Users will be able to query for specific trails and features, depending on what queries are implemented for the widget.

3.4. Design Choices

This section provides an overview of the software chosen to develop the proposed application. As the website and web application are the interactive portions of this project, the design of both tools is paramount to the end users' experience.

3.4.1. Software Choice

The main software used for this project is proprietary software, Esri's ArcGIS for Desktop (ArcMap) and ArcGIS Online. This choice was made because Esri's ArcGIS is the industry's standard, and it is the software that the City of Santa Barbara uses for its GIS needs.

For the web application, Esri's Web AppBuilder is used, due to its ease of use and current functionalities. This software was chosen over Esri's Collector App due to the Collector App requiring an organizational ArcGIS.com account for each user. This is not feasible as the participants of this project, and subsequently, most users, are not part of the organization that provides the organizational account.

The website is written in Hyper Text Markup Language (HTML) and styled with Cascading Style Sheets (CSS).

3.4.2. Platform Choice

This application is web-based, so it should allow access to anyone who has access to the website.

3.4.3. User Interface Design

The website design is intended to be minimalistic and functional. The Los Angeles County hiking trail website was used as a source of inspiration for the design of this project's website, as the Los Angeles County's hiking trail website includes a web application to view hiking trails and information regarding the hiking trails below the web application. The user interface of the web application is implemented with the same design mentality. Minimal widgets were used, and any unused features were removed. This was achieved through the built-in functionality of Esri's Web AppBuilder.

Chapter 4 Application Development

This section details how the web application and website were developed. Section 4.1 describes the data used in this project and how it was obtained, Section 4.2 describes how the web application was created, Section 4.3 describes the website creation, and Section 4.4 describes how the application was tested among this project's participants.

4.1. Data

This section details the steps involved in preparing the data for the creation of the web application. The hiking trail data for this project was obtained from the City of Santa Barbara. The data was received as a zip file, which contained a geodatabase containing two different layers. One layer is a line feature layer, which houses trail location data for ten front country trails in Santa Barbara. The second layer is a point feature layer, containing locations of all the trail features the City of Santa Barbara has gathered. This geodatabase was imported into ArcGIS for Desktop (ArcMap) to edit its features prior to publishing.

4.1.1. Preparing for VGI Submissions

In order to allow users to submit VGI directly to the geodatabase, a relational database management system (RDBMS) is required. The method used in setting up an RDBMS for this project is the creation of an enterprise geodatabase in ArcMap. This was done through the "Create Enterprise Geodatabase" tool in ArcMap, with the database platform set as a SQL_Server. This enterprise geodatabase was then connected through and registered with an ArcGIS Server.

After creating the enterprise geodatabase, all the data from the original geodatabase obtained from the City of Santa Barbara was imported.

4.1.2. Symbology

Symbology is the use of symbols to represent geographic information on a map. Importing the data directly into ArcMap defaults to a single symbol for all point features and for all line features. To differentiate each type of feature from other features, a custom symbology was designed. Table 2 shows the different features and their counts on the Rattlesnake Canyon trail contained in the geodatabase.

Table 2 Rattlesnake Canyon Trail Features

Feature Type	Feature Count
Boulder	1
Creek/Stream	6
Culvert	1
Draw	13
Entrenchment	3
Intersection	9
Meadow	1
Minimum Clearance Width	8
Other Feature	6
Parking	2
Post	2
Rock	1
Sign	5
Slide	1
Step	25
Switchback	15
Trailhead	1
Trash Disposal	1
Tree	2
Vertical Obstruction	3
Waterbar	113

Due to the amount of feature types in the original data, some features were excluded from the map symbology. Too many feature types on the map caused the map to be cluttered with many different symbols within a small space. This is partially remedied by zooming in on the map to limit the amount of features show, however, due to the small size of a smartphone screen, this method would limit the view shown on the map on a smartphone. With that in mind, certain feature types were removed based on what the general user would submit in terms of VGI. Feature types such as a “switchback” or “draw” were excluded due to the belief that the general user would have less knowledge about them.

The remaining feature types had symbols that were chosen from the default symbols available in ArcMap through the symbol selector. Each symbol was chosen to best represent the type of feature it displays. Figure 2 shows the symbols chosen for each feature type.

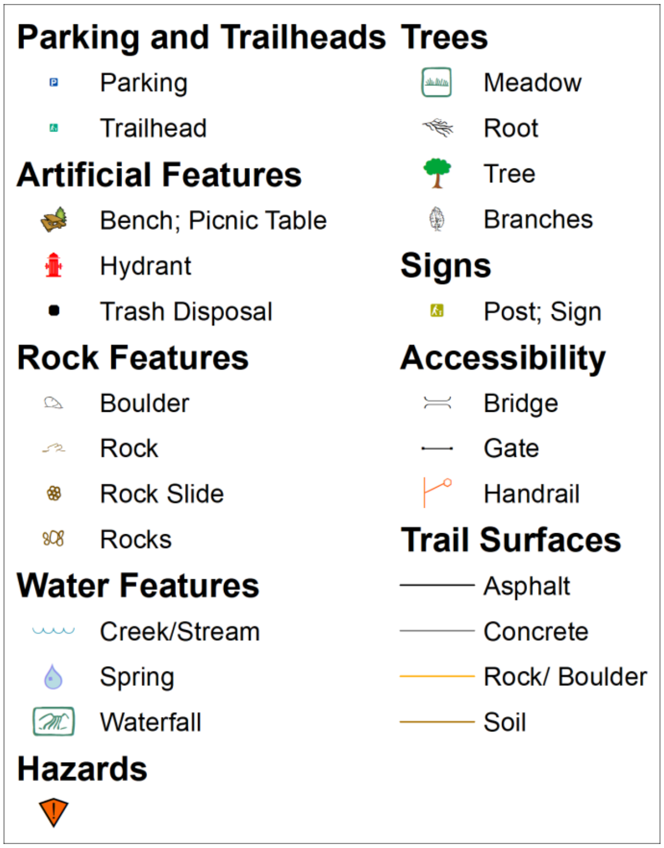


Figure 2 Symbols for each feature type

As the figure above shows, some feature types used the same symbol. Due to the web application size and the size of screens on mobile devices, too many different types of features would make the map cluttered due to too many different symbols. The features with the same symbols are grouped to condense the legend.

After determining the symbols for each feature, similar feature types were exported into layers in order to streamline the VGI submission process in the web application. This is done to separate the "Smart Editor" submission templates into different feature types, as seen in Figure 3, instead of having a single data layer. A single layer would cause the user to scroll through all feature types, as the data would not be grouped. The seven different feature layers are "Artificial Features," "Rock Features," "Water Features," "Hazards," "Trees," Signs," and "Accessibility." Refer to Figure 2 for the groupings of each feature types.

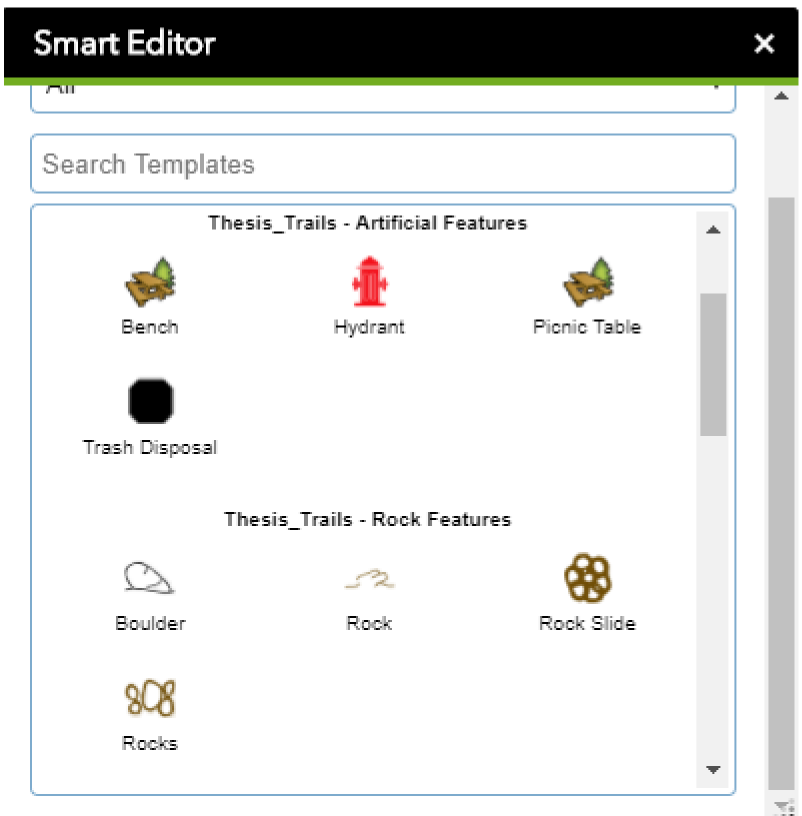


Figure 3 VGI submission groupings

4.1.3. Map Extent

Two map extents were used create the desired effect on the web app, showing only parking and trailheads in the initial view, and showing all other features when the map is zoomed in further. The "Parking and Trailheads" and "Trail Surfaces" feature layers were set to show the features at all scales. These layers are the initial view of the web application, as shown below in Figure 4. Note that, as explained below, legends, not included in this or the next image, are shown within the frame of the web page in which the web map is embedded.

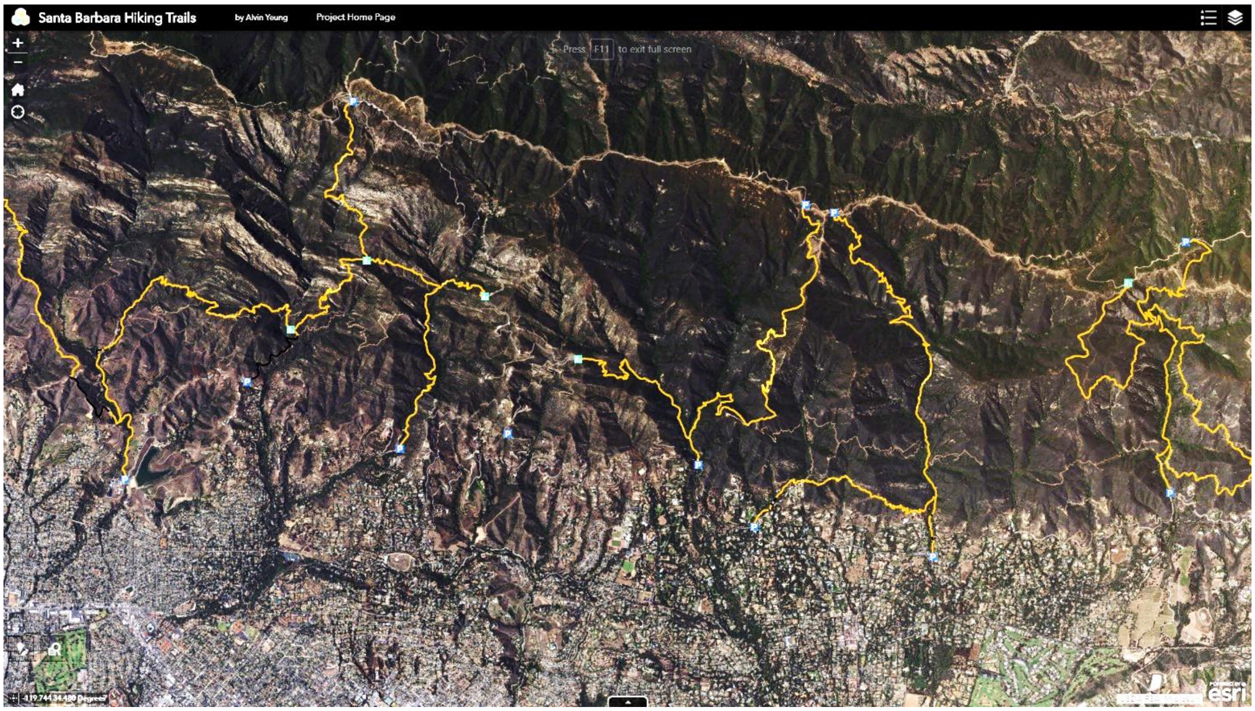


Figure 4 The initial view of the web map in the web application

The other features were set to show only when the map is zoomed in past the initial extent, as shown below in Figure 5.

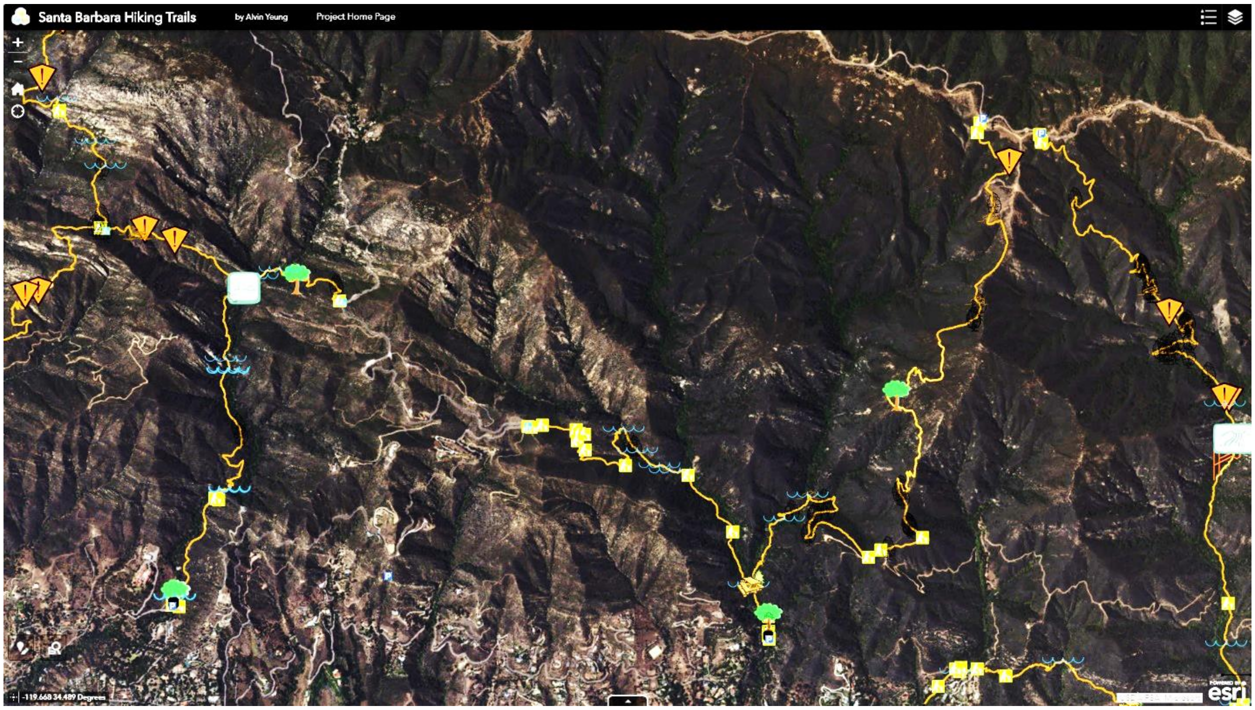


Figure 5 A zoomed in view of the web map in the web application

By setting a larger scale for the majority of features, the initial extent of the web application only shows trail locations and its parking and trailheads. This allows the user to click on an icon (parking or trailhead icon) to find out what trail it is and what parking is available at the trail. Once a trail is determined, the user can then zoom in on the map to view all the different features on the trail.

4.1.4. Attribute Fields

During the City of Santa Barbara's initial data collection, various types of information were collected about each feature. This resulted in many different fields in the attribute tables. Most of these fields are not necessary for the average user to view, as the information in those fields have no relevance for the user. Because the web application has a popup box that allows the user to view information about a specific feature, most of the fields for both the line and point layers were turned off in order to present the user with only relevant information regarding the feature.

The fields that were kept for the point feature layer are the "OBJECTID" field, for a unique identifier for each individual feature, the "FEATURE_TY" field, a classification created by the City of Santa Barbara for each feature, indicating what type of feature it is, the "DESCRIPTIO" field, a description given by the City of Santa Barbara for each feature, the "LATITUDE," "LONGITUDE," and "ELEVATION" fields, indicating the location data for each feature, and the "TRL_NAME" field, indicating what trail each feature is on. The field names were not changed to keep the fields consistent with the City of Santa Barbara's dataset. While the latitude and longitude are displayed on the web map in the web application, the elevation information is not. All location fields were kept in order to have one place in the web application that the user could look at to view location information.

For the purposes of VGI collection, two new fields were created. A "pID" field was created for the participants to enter their unique identifier, to differentiate VGI submissions from other users. A "Date_Created" field was created to identify the date which the VGI was collected.

4.1.5. Publishing to ArcGIS.com

After the enterprise geodatabase was configured with its symbology, extent, etc., the map document was published as a service. Any major errors found in the Service Editor were corrected prior to publishing. The service was published to Esri's servers through the USC organization.

4.2. Web App Creation

This section details the steps involved in creating a web application to collect VGI. Esri's Web AppBuilder was used to create the web application. Esri's Web AppBuilder was chosen over Esri Collector since use of the Collector application requires an ArcGIS.com organization named user account. The following subsections discuss how a web map was created from the service published in the previous section, how the web application was created through Esri's Web AppBuilder, and the widgets used in the web application.

4.2.1. Web Map

A web map is an interactive display of geographic information. The first step in the creation of a web application is to create a web map that shows the hiking trail data. An imagery basemap was set as the background, as it shows the most details about the hiking trails when zoomed in. A topographic basemap was also considered, as it shows elevation changes. However, after feedback from the participants, most users preferred an imagery basemap, as the participants could not interpret a topology basemap proficiently.

Using the imagery basemap as a background, the data previously published was loaded in. Popups for each feature were enabled by default in the web map, and no configurations were necessary at this stage, as the non-relevant fields for the popups were turned off in ArcMap, seen in Figure 6.

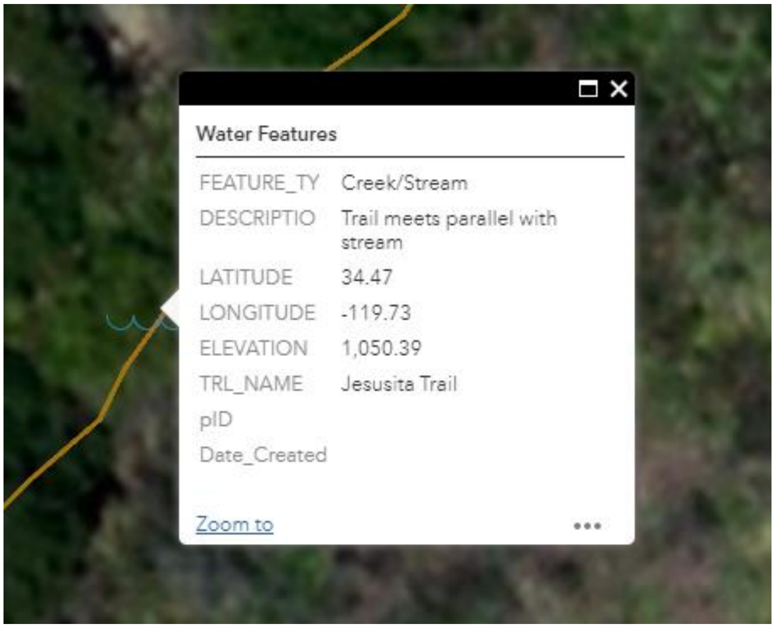


Figure 6 Popup of a Creek/Stream Feature

4.2.2. Esri's Web AppBuilder

The web application for this project was created with Esri's Web AppBuilder. Esri's Web AppBuilder is a software that allows users to customize a web application using predefined templates. The web map created in the previous section was used as the base of this web application.

A template that minimized the number of icons on the web application was chosen. Some of the default widgets and icons were also removed from the interface.

4.2.3. Widgets

Widgets are an interface that allows users to perform a specific function or access a service. Two widgets used for this project are the "Smart Editor" and the "Query" widgets built into Esri's Web AppBuilder.

4.2.3.1. Smart Editor

The "Smart Editor" widget allows a user to interact directly with the geodatabase, potentially allowing the user to edit, update, and delete existing features within the geodatabase. This widget allows the participants of this project to submit their VGI directly into the geodatabase, which could then be accessed in ArcMap. However, the settings in this widget were changed so that users could view previous submissions, but not delete, update, or modify the geometry submitted by other users or the City of Santa Barbara's dataset. This protects the original data from the City of Santa Barbara, so each user would view the same data, as well as any VGI submissions. To report on a feature, each user would be required to create a new feature for their submission instead of overriding the previous user's submission. This design choice was chosen to gather multiple submissions to perform a thematic analysis on textual VGI. Thus, the only functionality this widget provides to the end user for this project is the ability to create, edit, and delete the features created by themselves. However, each user could view other user-submitted VGI, similar to OSM and Wikimapia.

4.2.3.2. Query

A query widget allows users to query for specific features or trails within the web application. The query widget runs SQL queries to return its results. Six different query options were preset in this web application to provide users several different query options.

A "Parking and Trailheads" query allows users to query for all parking and trailheads shown on the map. This query requires a criterion entered by the user to function. The criterion is the "TRL_NAME" field in the attribute tables of the data. Users enter a trail name (or a partial trail name) or select from a drop-down list to find the location of the specific trail head or nearby parking for that trail. A "Trail Search" query allows users to search for a specific trail on the web map. This function also requires a user input in the form of a trail name, using the "TRL_NAME" field of the attribute table.

The following queries do not require any user input. An "All Parking Locations" query allows users to search for all parking locations near hiking trails as listed by the City of Santa Barbara. An "All Trailheads" query displays all trailheads on the web application. A "Benches and Tables" query displays all benches and tables on the trails on the web application, and a "Water Feature" queries displays all features categorized as water features.

Queries that do not require a user input display a list of features within the popup box and the web application adjusts the view to show all features. The user is then allowed to click on a specific feature to zoom it.

4.3. Website Creation

This section describes the development of a website to host the web application. The website was created through hypertext markup language (HTML) and cascading style sheets (CSS).

4.3.1. Design

A minimalistic design was chosen for this website. The website's only requirement was to host the web application and the instructions for the participants of this project.

The design of the website is based on the Los Angeles County's hiking trail website. This website was chosen as a model since it hosts a web application for hiking trail information in the middle of its web page and information about the trails below the web application.

A structure similar to the Los Angeles County's website was used for the website for this project. A simple title was placed at the top of the web page, and a navigation bar was placed just below the title. The web application is embedded below the navigation bar and centered with the web page. A direct link to the web application on ArcGIS.com is listed below the embedded web application in small text. Two sections are divided right below the link. The left section contains text and a title called “Data Submission,” which provides the user with instructions on how to submit their VGI through this web application. The instruction is also written in Section 4.4. The right section is titled “Feature Legend” and is contains an image of a legend of all features in the web application, as seen in Figure 7. Each item in the legend was given a symbol representative of the feature, and for this experiment, more complex symbols are not required.

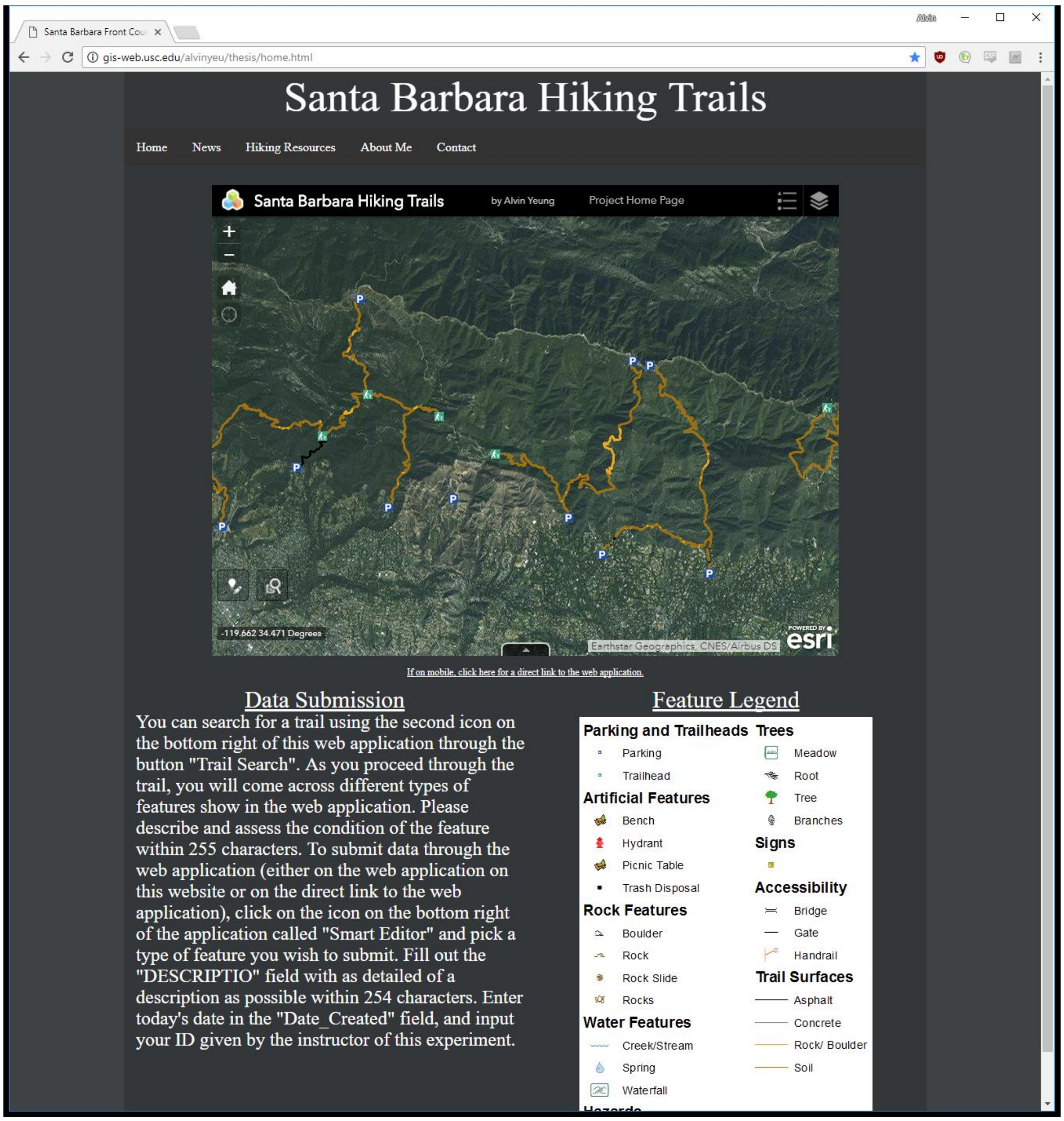


Figure 7 Instructions and a legend below the web application

4.3.2. Publishing website to USC servers

The website itself is contained in an HTML file, with the designs and styles contained in a separate CSS file. Both files are located in a single folder, along with another folder containing the picture for the legend. The HTML file for the website references both the CSS file and the legend picture inside the picture folder. The folder and all its contents were then uploaded to USC's servers through FileZilla to be accessed by the users of this study. This website is only hosted for the duration of this project, and will not be available after.

4.4. Application Testing

To test the functionality and success of data collection of the app, a small group of hikers was asked to use the app to collect trail data. They were given a link to the website, along with the direct link to the web application. Users were instructed to use either the embedded web application or the direct link to the web application for their VGI submissions.

Participants were instructed to submit points on the web application, either through the embedded web application on the website or through the direct link to the web application. Each participant was given a unique identifier in the form of a letter to distinguish the VGI from each participant. The participants were also instructed to enter their unique identifier at each submission.

The explicit instructions given were "As you proceed through the trail, you will come across different types of features shown in the web application. Please describe and assess the condition of the feature within 255 characters. To submit data through the web application (either on the web application on this website or on the direct link to the web application), click on the icon on the bottom right of the application called "Smart Editor" and pick a type of feature you wish to submit. Fill out the "DESCRIPTIO" field with as detailed of a description as possible within 255 characters. Enter today's date in the "Date_Created" field, and input your ID given." No further instructions were given to the participants regarding VGI submissions through the web application.

Prior to beginning their hike, users were instructed verbally to submit VGI for at least 10 features shown in the web application. The participants were asked to limit their VGI submissions to a single trail within the dataset, the "Rattlesnake Canyon" trail. This was done to gather multiple datasets from the participants for the same features. The users went out to the trail within a two-week span with no major changes in the weather.

Because the trail had a weak cellular signal (for both AT&T and Verizon carriers), if the user could not submit their VGI when they were at the feature, the user was instructed to move to a point on the trail with a cellular signal and submit their VGI.

Users were also instructed to take a picture of each feature with their smartphone camera. The photographs were used to compare the locational accuracy of the GPS units in the users' devices to the locations of the features from the City of Santa Barbara's dataset. To collect location-based data for the photos, participants were asked to make sure that the location service for the camera application was enabled. No specific instructions were given on how to take a picture of a feature. The participants were instructed to upload their pictures to a Google Drive folder. Within that Google Drive folder, each participant was given their own folder, named according to their letter designation for their participant ID.

Chapter 5 Application Trial

To determine the success of the current website and web applications, along with how the VGI submission process feels to the end user, an evaluation of these three aspects was performed. This chapter presents the evaluation process and its results. Section 5.1 describes who participated in this trial along with how they were chosen, Section 5.2 presents each phase of the testing, Section 5.3 provides an evaluation of the web application and website used in the trial phases, and Section 5.4 provides an analysis of the VGI collected. Note that the figures in this chapter have no background imagery or color, as they are zoomed in to illustrate individual features. In this case, zoomed-in imagery only shows blurry tree cover, and a topographic background does not show anything at that scale.

5.1. Hiker Subjects

Once a minimum viable product for the website and web application was achieved, a test of data collection in the field by volunteer users was necessary, as VGI relies on multiple people of their own volition to contribute. No real prerequisites were required, besides access to a computer and smartphone, along with the physicality required to hike on a trail. One restriction in the participation of this experiment was the availability of the participants within a set date. The time span targeted for the participants to collect data on a hiking trail was limited to two consecutive weekends, from June 2, 2017, to June 11, 2017. This time span was chosen to ensure the weather and temperatures within the nine days had minimal variance so there would be negligible weather based impact on the features the participants would see on a hiking trail. To participate in this experiment, the participants had to be available within the nine days. This process allowed the primary objective of this web application to be tested: to collect VGI data about features on a hiking trail in Santa Barbara. Once the subject confirmed participation, a text message was sent to their mobile phone with a shortened link to the website. Once all the participants had confirmed participation, a trail was chosen from the trails listed in the City of Santa Barbara's dataset.

Five hikers participated in this test. All five participants had access to a computer and a mobile phone. The ages of the participants ranged from thirteen to twenty-four. The participants had no previous exposure to the website or the web application. During the month of March, each participant was contacted through either a phone call, instant messaging, or in person. The participants were given a description of the project as a whole and its primary objective along with a general idea of what they would be expected to perform on the hiking trail.

The Rattlesnake Canyon Trail was chosen based on its moderate difficulty, availability of cell phone service, and a sufficient number of features on the trail as shown on the web map. Unlike other candidate trails, cell phone service for Rattlesnake Canyon was available, though limited to 1-2 bars of connectivity. Measurement of cell phone service is based on the author's previous experience with the trails. The number of features shown on Rattlesnake Canyon Trail is based on the features defined in Section 4.1.2., totaling 22 features displayed.

5.2. Phases of the Trial

The first phase of the trial involved testing the website and web application on computers. Two platforms were used, Windows 10 and Mac OSX. The participants were asked to visit the website and use the web application embedded on the website. Users accessed the website and web application through a shortened link sent in a text message, as described in Section 5.1. The participants had full access to view the data on the web application, along with any attributes shown in a popup about a specific feature. The users were asked to try out the widgets on the web application to get a sense of how the web application would work on the trail. Because there is no explicit mention of a popup box for the trail features in the web application, a specific instruction was given verbally and through a text message to the users to click on a feature. This instruction was to inform the users of the availability of additional information for the features on the trails. If a user encountered any usability issues or errors with the website or web application, they were instructed to report on the issues or errors through a text message. Given feedback from initial users, by the beginning of the first phase, all features of the web application were functional and users could query for different types of features and trails, along with preview the VGI submission form.

The second phase of the trial involved in-situ data collection on the Rattlesnake Canyon hiking trail. The data to be collected was feature points, textual VGI, and photographic VGI. The feature points were collected to determine the accuracy of user submitted points relative to the position of the original feature and to observe where the users chose to locate the positions of their VGI. As the objective of this project was to collect VGI that can be used to update the existing trail information from the City of Santa Barbara, textual VGI was collected to acquire a description and condition of the features. Photographic VGI was collected to test the accuracy of the GPS units in the hikers' mobile devices and compare the coordinates of features to those in the dataset from the City of Santa Barbara.

The in-situ data collection was scheduled to occur over nine days, during which the weather varied minimally. According to Weather Underground, the weather over the course of the collection period had a maximum temperature of 76 degrees Fahrenheit with an average temperature of 70 degrees Fahrenheit. No unusual weather events occurred during the nine-day collection period.

To collect data, the participants would venture on to the trail and at each trail feature shown on the web application, the participant would take a photo of the feature with location tags on and, using the web application, create and submit a new feature point for the feature. A new feature was created by each user for each feature shown on the web application. Figure 8 below shows the fields users had to complete to submit a feature.

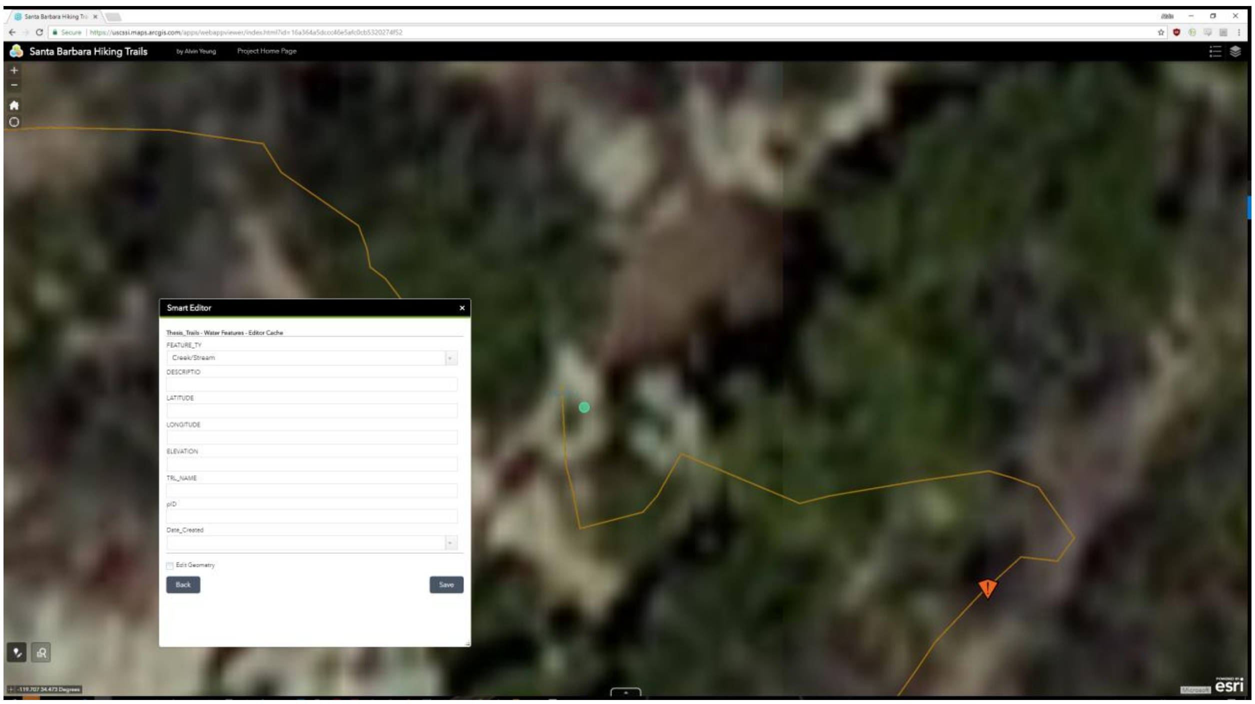


Figure 8 Feature submission fields

Each participant had access to the website and web application through their own mobile phone. The instructions, as described in Section 4.4, were followed. This was confirmed when speaking with each of the hikers after their hike, along with checking the Google Drive folder for each participants' photo submissions. It was also at this point that any verbal feedback on the web application and website were received.

5.3. Evaluation of the App from the Trial Phases

All five participants confirmed that the website and web application were functional, Via in-person or text communications, they confirmed that they could interact with the website and view instructions on how to submit VGI directly below the web application.

Regarding the feedback given on the usability of the website and web application, a few key points were noted by the participants. The first major feedback was due to the size of the mobile devices. While the screens of each device were high resolution, the total viewing area was constrained to a smaller form factor (four to five and a half inches) than on a computer monitor. This created a higher pixel density for each user, and as such, the text containing a direct link to the web application was too small and difficult to click. While the direct link to the web application viewed on a computer screen is large enough for the participants to read, on a mobile device, the text becomes too small to read or click without zooming in a large amount.

The second major feedback given by the users was with respect to the web application itself and its VGI functionality. As discussed below, the participants clustered their VGI submissions towards one another. This is due to the decision made to allow users to see other, user-submitted VGI, as well as the limited functionality of the “Smart Editor” widget provided by Esri’s Web AppBuilder. While the Smart Editor widget allows a user to update a feature, it does not allow a user to add a new field to the feature. Updating a feature would modify the previous information, and the previous VGI may be altered. To preserve any new and existing data, for the purposes of analysis, a new feature had to be created by each user. Once the first participant placed their own feature submission near the original feature, each subsequent participant also submitted their own feature near the original and previously recorded VGI feature. It became increasingly difficult to determine what was the original feature. This process also makes it more difficult to view the popup menus of each feature, original and VGI submitted, without zooming in a large amount.

Also, because cell service on this particular trail was not strong, the response time of the web application was slow at times, depending on what part of the trail the user was located. This process caused a slowdown in terms of VGI submissions and pacing for the participants, and the distance traveled by the users took noticeably longer than their normal pace. This feedback on slow response time was only given by two of this study's later users, as this issue did not occur to the earlier participants.

5.4. Analysis of Data Collected

An analysis was done on the data gathered by the users to determine its validity and quality. Three types of analysis were performed: a positional analysis, a thematic analysis, and a control data analysis.

5.4.1. Positional Analysis

A positional analysis was performed to determine patterns that might have emerged from the users' submissions. Figure 9 shows the VGI submitted by only the first participant for two features. As can be seen, the first participant submitted their VGI near the original feature. This is the case with all other features that were submitted by the first user with respect to the original feature.

Figure 9 Participant “A” VGI submissions

The later participants also submitted their VGI near the original point. Similar to the first participant, the subsequent participants’ VGI clustered their submission for the same feature to the original feature, as seen in Figure 10 and Figure 11. This may be explained by having no concrete instructions on where the participants should submit their VGI on the map, as well as the limited screen size while using the web application on a mobile phone. The positional accuracy of the features created by each user diminished in accuracy after every subsequent submission, as each subsequent submission created less room to click near the original feature. However, each subsequent submission maintained a level of precision towards the closest feature to the original. Thus, the VGI is correct in the exact feature the users were asked to report on, and other users validated the previous VGI through submission of the same feature.

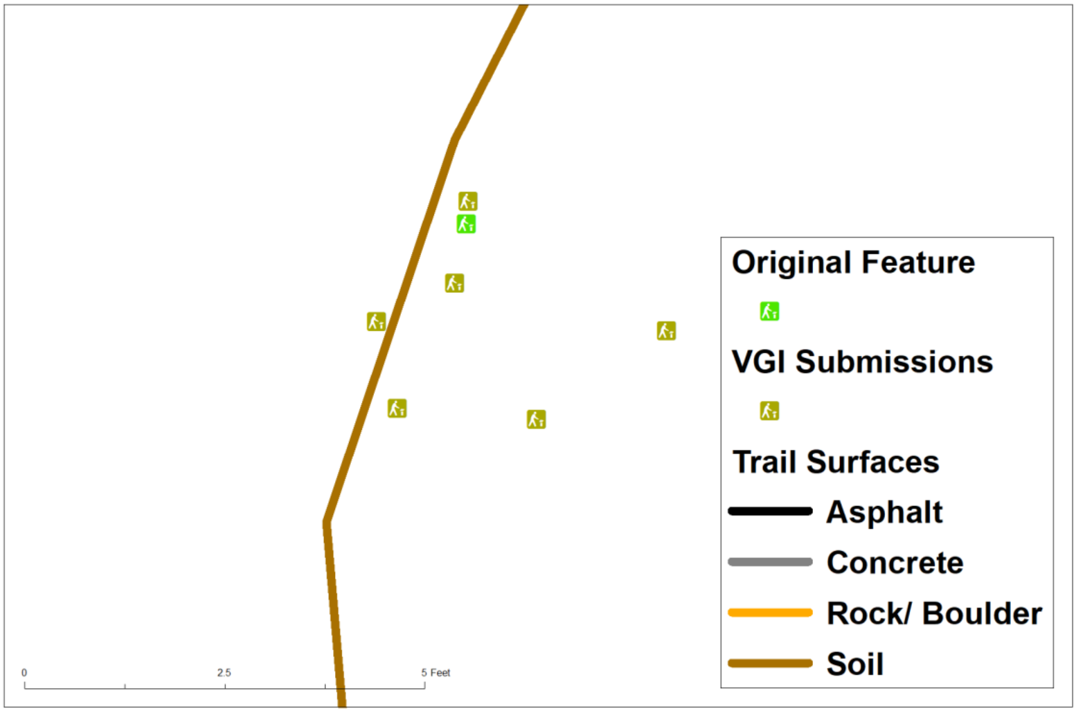


Figure 10 Clustering of VGI submissions for a sign feature

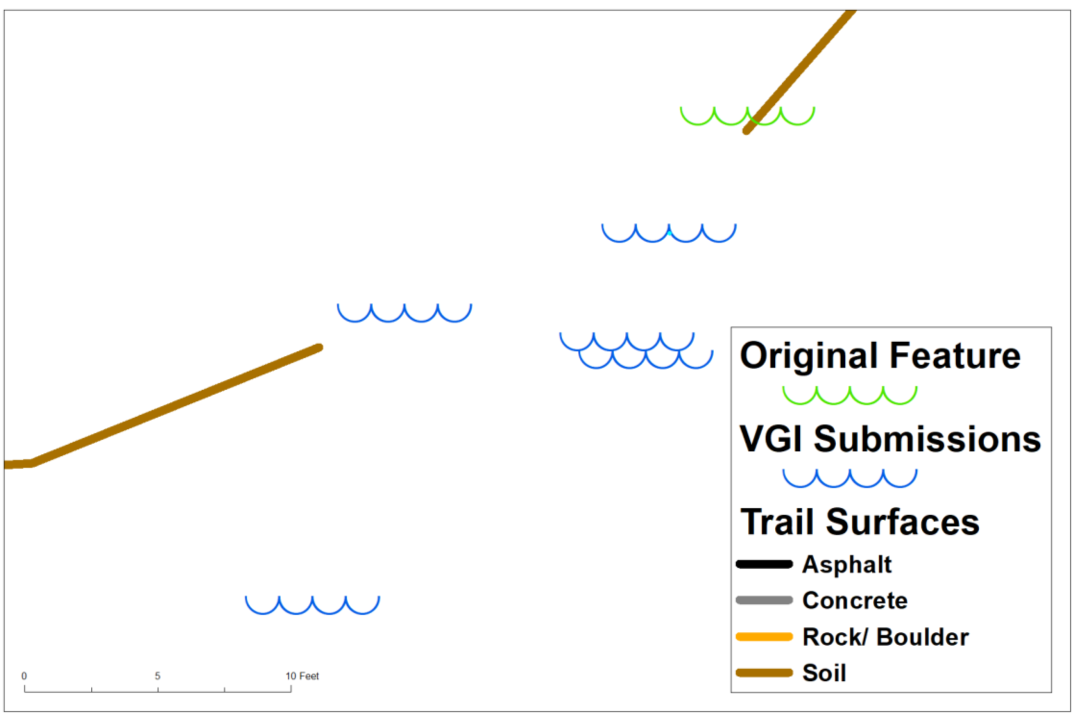


Figure 11 Clustering of VGI submissions for a river feature

5.4.2. Thematic Analysis

To determine common themes among the non-spatial VGI submitted, a thematic analysis was conducted. To become familiar with the data, all VGI submissions were first separated from the original features in the database, sorted by their participant ID, and then exported from ArcMap into Excel spreadsheets. Each participants' VGI was then reviewed carefully, two times, to get a grasp of what the participants submitted about each feature relative to other participants' submissions. Table 3 below shows a sample of the VGI text submitted by a participant.

Table 3 A single participant's VGI submissions

Object ID	Feature Type	Description
2	Trash Disposal	SB PARKS Trash can. Con. Ok
11	Creek/Stream	Rock bridge. Good condition
12	Creek/Stream	Flowing water. Excellent condition
23	Tree	Red wood. Poor condition
35	Sign	No fishing sign. Condition good
37	Sign	Rattle snake canyon sign. Con. Good
39	Sign	Prohibit sign. Con. Ok
41	Sign	Metal pole. No metal pole
42	Sign	Bicycle sign gone. Next to no fishing sign

The second step of generating the initial codes for the thematic analysis came from key words submitted by the participants. Because the participants were asked to describe and assess the conditions of the features to the best of their knowledge within 255 characters, the codes generated leaned towards ones that fit a feature's conditions and its descriptions. Figure 12 shows the overall themes that codes were assigned to.

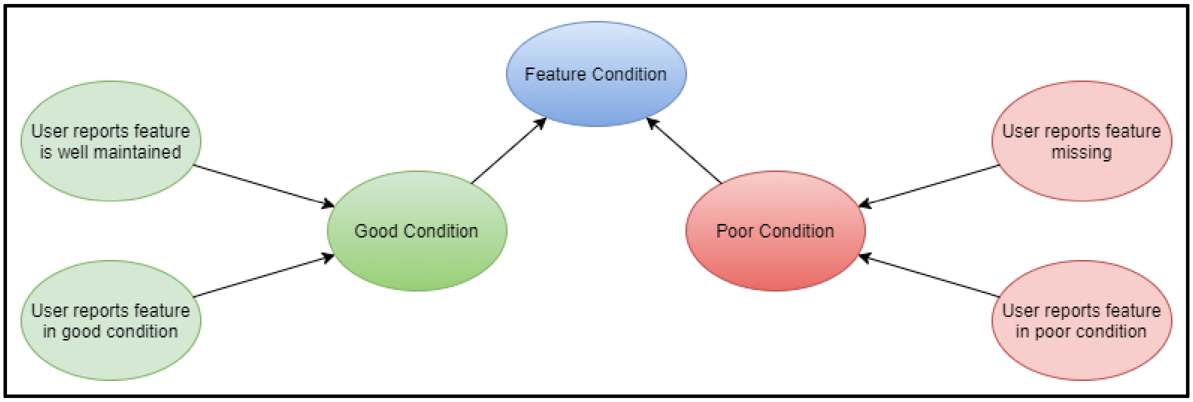


Figure 12 Theme diagram for the thematic analysis

As seen in Figure 12, the entire VGI dataset was classified under a general "Feature Condition" theme, as the participants were instructed to report on the features' condition. Two subthemes were created from the "Conditions" theme to better divide and classify each VGI submission. A "Good Condition" theme was created to group the VGI submissions that reported the features in good conditions. Conversely, a "Poor Conditions" theme was created to group the VGI submissions that reported the features in poor conditions.

Based on how the participants worded their VGI on a particular feature, a code for the entire submission of a single feature was created. Table 4 below shows the keywords used in determining if a description fell into a "Good Condition" subtheme or a "Poor Condition" subtheme. A "Good Condition" subtheme defines the feature as being in good condition. This was described in the VGI either through an explicit grading of the feature written in their text submission, with grades ranging from "Ok" to "Excellent" or by descriptions of the feature as being well-maintained. Similarly, a "Poor Condition" subtheme defines the feature as being in poor condition. This condition was also described in the VGI as "Poor" or "Missing" as well as the feature being described in poor health or woefully maintained. If necessary, two codes for a single submission were generated based on non-similar phrasing.

Table 4 Keywords to determine if a feature is in “Good Condition” or “Poor Condition”

Good Condition	Poor Condition
Ok Condition	Poor Condition
Good Condition	Missing
Excellent Condition	Dead

A review of the themes and codes generated was done after creating the main and subthemes, to ensure the VGI is correctly labeled for this analysis. This process was done through reviewing each submission manually and ensuring the descriptions submitted by the participants fit the themes it was assigned to. Figure 13 below shows how each feature was categorized. The users also unanimously agreed on the condition of most features. Based on the text of the VGI, all users reported that both the bike sign and the metal post was missing from the trail. The tree feature on the trail was either dead or in poor condition. The users graded the trail sign and “No Fishing” sign in good condition explicitly, and the trash can was graded as “OK” and “Fair” in the text submissions. One user was not as explicit on the grading of a feature. However their VGI was more descriptive. Using words such as “well maintained,” this feature was classified under the “Good Condition” theme. This process was not necessary for most submissions, as the users provided explicit grades in their descriptions.

Based on this trial, it was concluded that the quality of VGI on hiking trail features can be assured if multiple users report on the same feature. A general consensus was found on each feature by the hikers, and each submission fit into the subthemes generated by this thematic analysis.

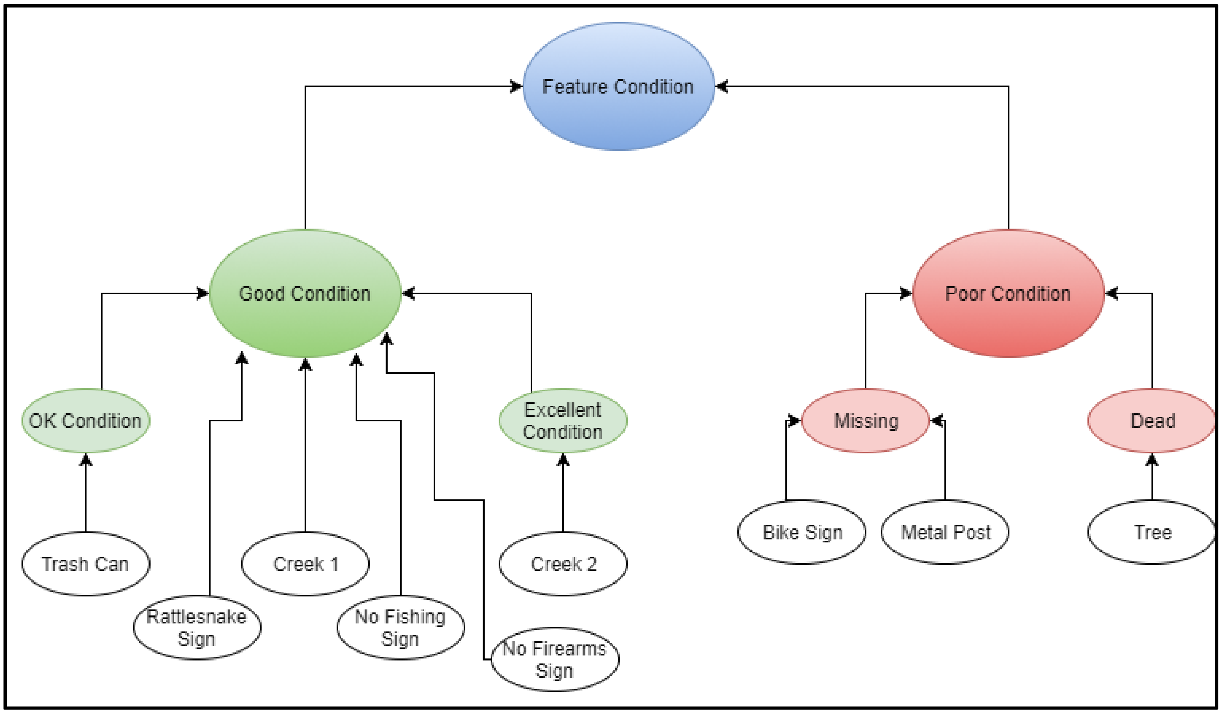


Figure 13 Categorized features

5.4.3. Control Data Analysis

In order to further compare the participants' locational VGI to the City of Santa Barbara's dataset, used here as the control data, the participants were instructed to take a photo of each feature. The participants used two different types of phones, a Samsung Galaxy S7 and an iPhone 5. Both types of phones have the ability to insert a location tag feature for each photo taken. This feature uses the GPS of the phones to geotag each photo taken with the position of the phone at the instance the photo is taken.

Each participant was asked to take a photo of every feature they submit through the web application. Some participants progressed further along the trail than others, so there are a couple of feature points submitted with only photos from a couple of the participants. Each participant however did submit photos for the features at the beginning of the trail. Figure 14 shows a map of the photographic VGI points of three different signs at the beginning of the trail from each participant. Participant B shows only two points as the participant only submitted photos for two of these signs. Participant C has three features shown on the map, with two features positioned in the same coordinates, as the participant took a photo of the two features in the same location. Participants A and E are not shown in this figure as the GPS on their devices malfunctioned when taking photos of these features.

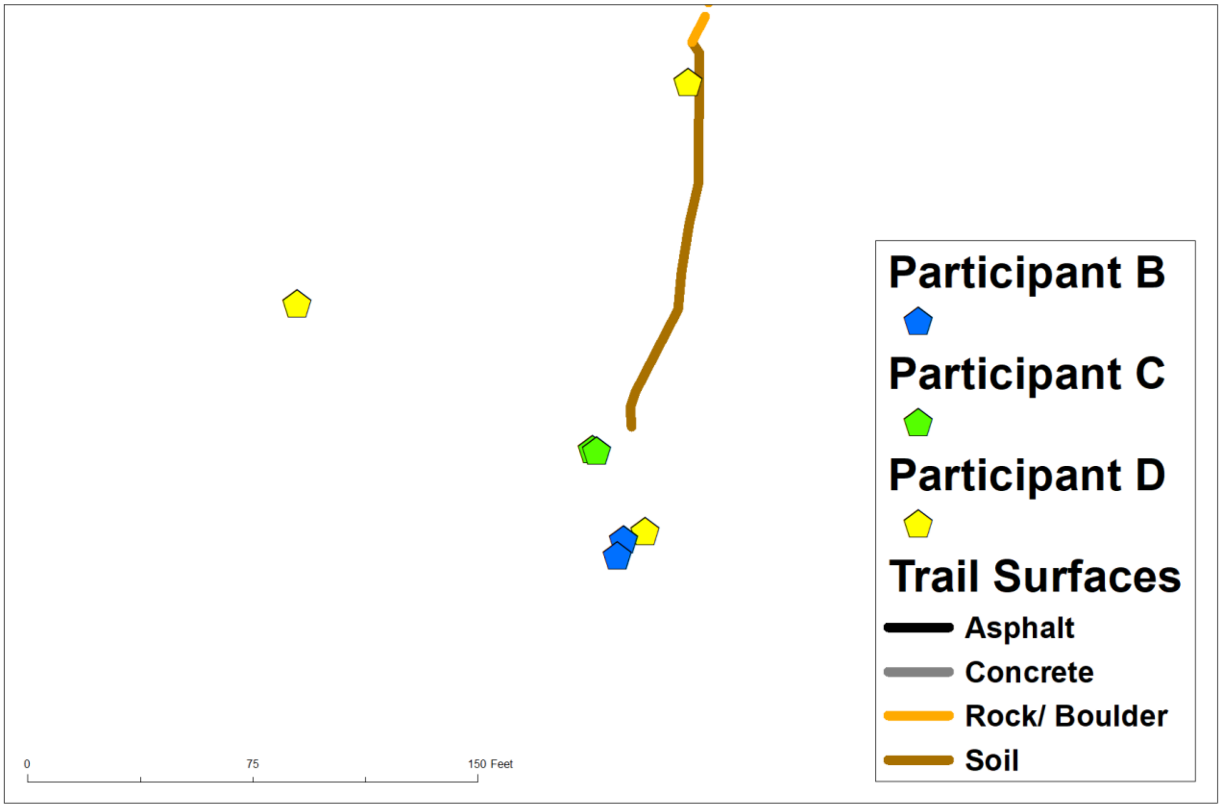


Figure 14 Photographic VGI of each participant

The tool used in ArcMap to display the location data for each geotagged photo was "GeoTagged Photos to Points." This tool takes the longitude and latitude coordinates from the photos taken by the users and creates point features for each photo in the current map document. Each individual point displays where the GPS tagged the photo taken of each feature.

To compare the locations of the participants' VGI against the locations from the City of Santa Barbara's dataset, a tool in ArcMap called "Point Distance" was used. This tool was run against all of the participants' VGI. The "Point Distance" tool calculates the distance between two points, measured in feet for this project. The measurements in feet are due to the City of Santa Barbara’s data standards, as the projected coordinate system used by the city is the NAD_1983_StatePlane_California_V_FIPS_0405_Feet projection. The results were then exported to Excel files for each of the participants and inserted into one Excel file, with different sheets for each participant. Because the “Point Distance” tool calculates the distance between two points, with one point being the desired feature and the other point being all other points on the map, excess points were removed manually due to the small size of the dataset. Matching points were kept for analysis.

Due to the location and coverage of the trail, some points taken by the phone’s GPS were far away from its intended feature, due to the GPS drifting too far away or outright malfunctioning, as shown in Figure 14 above. Figure 15 below shows all points submitted by the participants where the GPS malfunctioned. Because these points are outliers in the data, they were not included in this analysis.

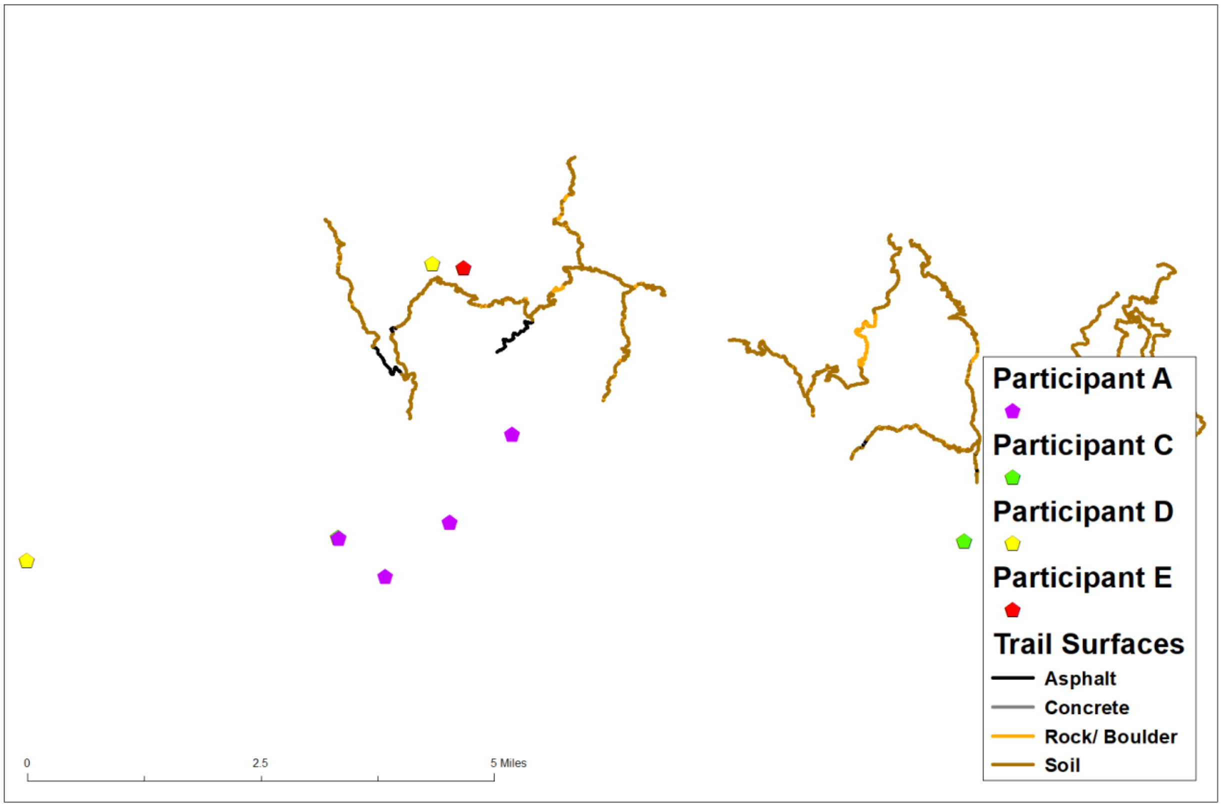


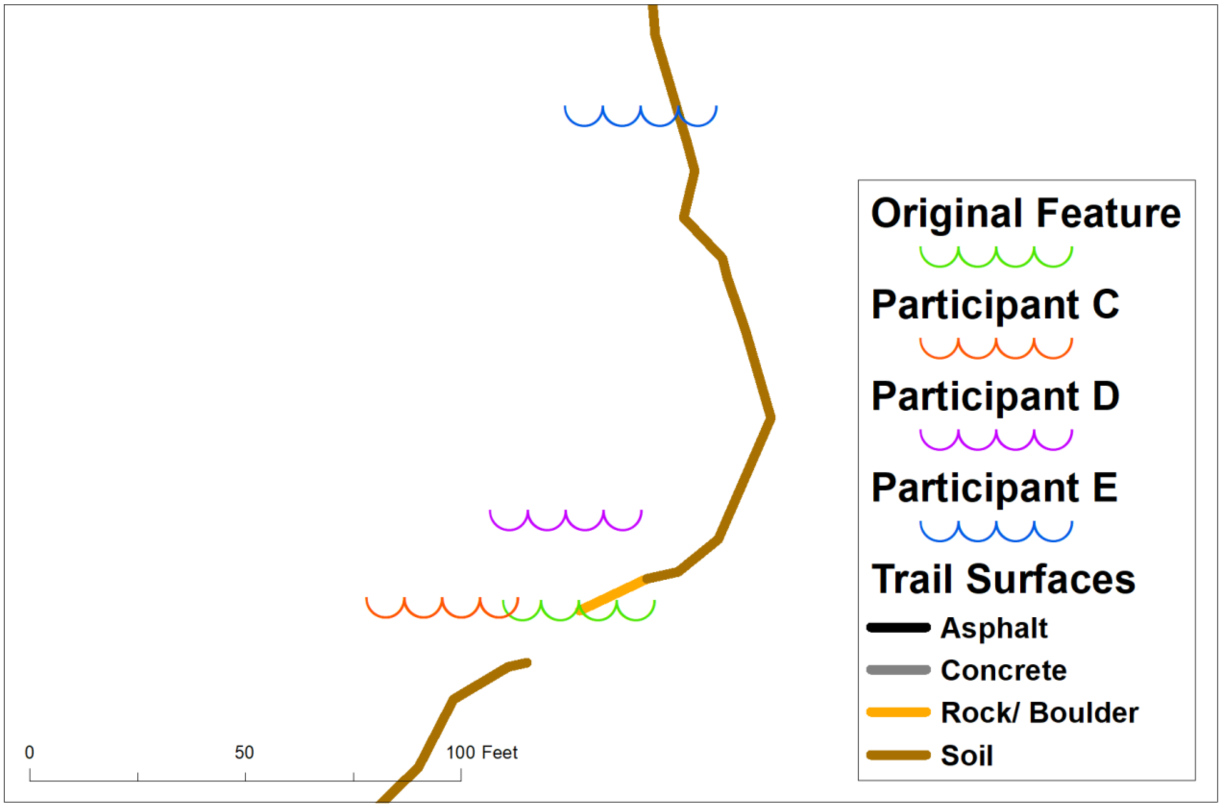
Figure 15 GPS malfunctions on the participants' mobile phones

Table 5 below shows the results of the Point Distance tool for each participant, with the outlier results removed. The Feature ID field is the unique identifier to each feature. The Distance field is the distance between the VGI submitted by the participant and the original feature, measured in feet.

Table 5 Distance in feet from photo point position to feature position

Feature ID	Feature Type	Distance by Participant ID (ft)					Mean Distance
		A	B	C	D	E	
2	Sign				22	22	22
4	Sign			118	9	35	54
5	Sign	22	68	36	114	28	54
6	Sign				19	35	27
7	Sign		50	22	47	36	39
10	Meadow				18	25	21
13	Tree	203	77	30	64	42	83
16	Creek					105	105
17	Creek			32	21	116	56
18	Creek		24	189		156	123
19	Creek	176		393	47	21	159
21	Boulder					22	22
22	Trash Can	49	102	70	35	14	54

The average distance between the photo point locations and the location recorded by the City of Santa Barbara for the associated features was 63 feet. The largest average distance from a specific City of Santa Barbara feature location was 159 feet for creeks. The top three largest average distances away from the original features were creeks. This is likely due to a creek spanning large distances. As the participants were not instructed on how to take a photo of a feature, the participants could have taken a photo of a creek at any position of the creek. The pictures submitted by the participants confirm that the participants took a picture of creeks at different locations on the same creek. Figure 16 below shows three pictures of the same creek, taken from distinctly different locations and shows the locations of the photos taken in ArcMap. The photos were taken close to the location marked by the City of Santa Barbara, as seen in Figure 17



. The positions of the photos taken for a single creek feature varies based on where the participants stood at the feature. The differences in location, as shown in Figure 16 and Figure 17 indicate that the VGI from photographs may not be precise.

Participant C Participant D Participant E

Figure 16 Three pictures of the same creek, taken from different locations

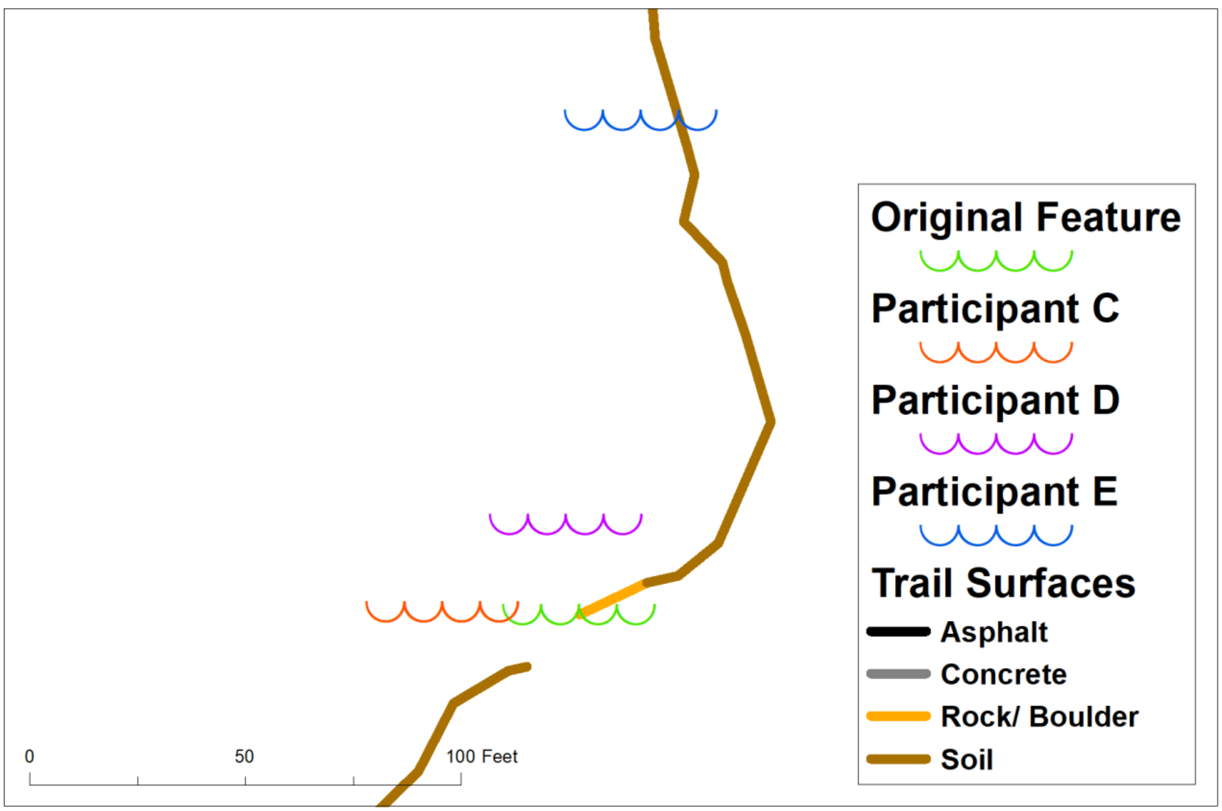


Figure 17 Locations of the photos taken in ArcMap

The lowest average distance away from the original feature was a sign. The average distances away from the original features ranged from 23 feet to 53 feet. The data shows four sign features with similar average distances away, with two signs at twenty-three feet and two signs at fifty-three feet. This is likely due to some signs being in close proximity to each other, so the participants likely took two photos of the signs at the same location. Figure 18 below illustrates two participants' photographic VGI submissions of two signs at the same location.

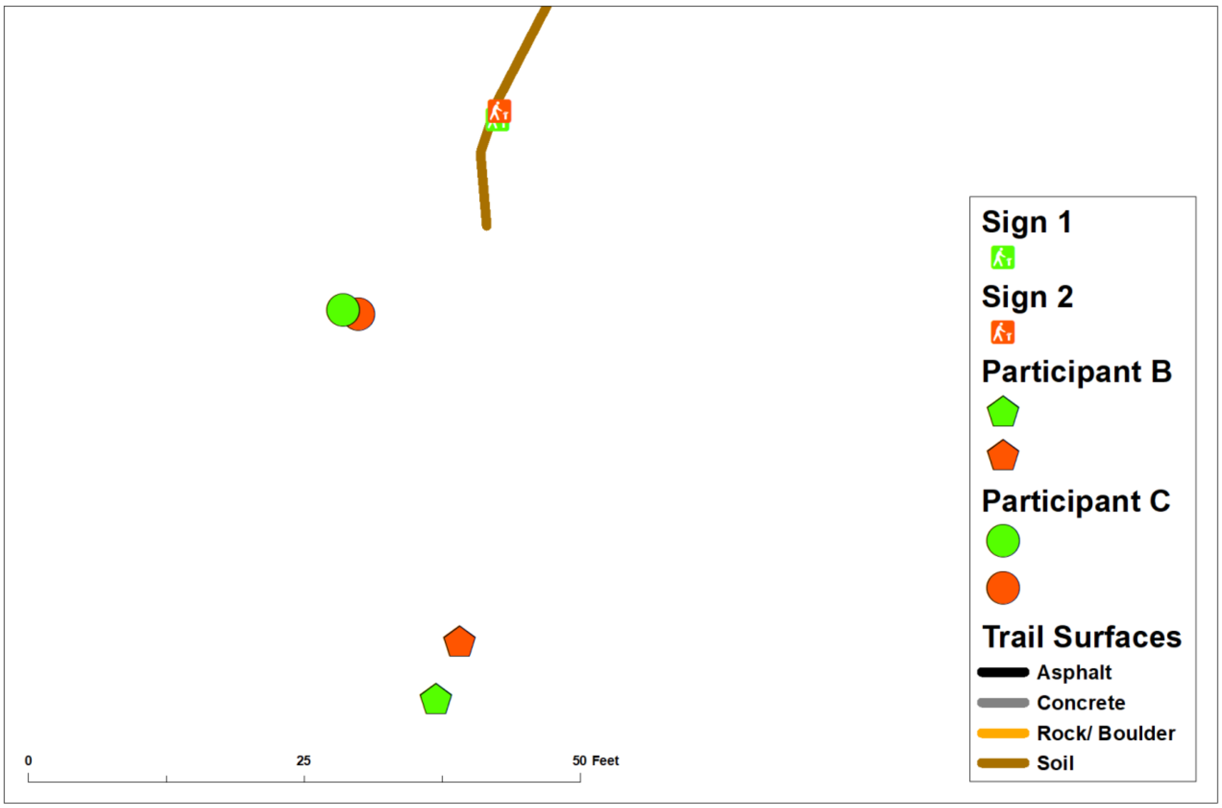


Figure 18 Participant B and C photographic VGI submissions of a sign

5.5. Summary of Application Trial

Overall, the application trial confirmed a working website and web application. While there were some limitations to the web application, the users were able to access the website and web application and submit VGI. Five separate hikers tested the web application and submitted VGI for at least ten different features on the Rattlesnake Canyon Trail. The positions of the VGI submitted were clustered around the original features. A thematic analysis of the VGI shows that while there were different descriptions of each feature, there was a consensus on the condition of the features. An analysis of the photographic VGI showed that a GPS unit can malfunction and submit inaccurate locations of photos and that the inaccuracy is visible within the photos. When the GPS was functioning properly, there was variance in the distance from a feature, as a feature could span a long distance. This could also be caused by a lack of instructions on how to take a photo of a feature, as users took photos of the same feature from multiple spots, causing variance in the GPS coordinates. Ultimately, the users were able to access the web application, submit their VGI, and an analysis was performed on the quality and validity of the VGI. The thematic analysis found that the quality and validity of written VGI can be assured if there are multiple submissions of the same feature and there is a general consensus of the feature’s condition. A positional analysis shows that as more features are submitted, the accuracy of the VGI diminishes and comparing the photographic VGI to the control data from the City of Santa Barbara shows the precision of the VGI is dependent on the type of feature.

Chapter 6 Conclusions

This chapter provides the conclusions drawn from the development of the website and web application and the process involved in analyzing the gathered VGI. This chapter also describes challenges faced during the development of the web site and web application, the limitations of the website and web application, and any future improvements that can be made to the methodology of this project as well as any future implementations or changes.

6.1. Success of Application

This project produced a website based on the Los Angeles County's hiking website. The website hosted a web application created with Esri's Web AppBuilder. The three functional requirements of the web application are to display hiking trail data from the City of Santa Barbara, allow users to query for trails and features, and to collect VGI from the participants of this project. The web application allows users to view data from the City of Santa Barbara about trails and features on the trail along the front country of Santa Barbara. A query widget added to the web application fulfills the requirement to allow users to query for trails and features. The "Smart Editor" widget allows users to submit their VGI through the web application. All three functional requirements of this project were met. The users confirmed through in person interviews after their hikes that the web application displayed the trails and features, searching for a specific trail or feature was available, and submitting VGI through the "Smart Editor" widget was functional. As such, this web application can be considered a success.

6.2. Challenges in Development

Although Esri's Web AppBuilder provided a great platform for this web application, the platform is also a limitation. Due to the platform being geared towards an easy to use focus, this software lacks several features that would have been ideal for this project. To address a specific point of feedback from a user, one feature would be the ability to create a many to one relationship between rows of data to a feature. This functionality would allow several users to contribute to a single feature instead of creating several features from several users. This would reduce the number of features displayed on the web map at any point in time. While the "Smart Editor" widget could have allowed users to edit other users' submissions, the edits would not add new rows into a single feature. The edits would overwrite the previous users' submissions.

The lack of support for a many-to-one relationship in a database using Esri's Web AppBuilder widgets is a known limitation. However, it does provide some benefits. A flat database design provides multiple point representations of the same feature. Using the work flow of this project, multiple photos over a long period of time may be collected and used for trail maintenance as well as showing how a trail has changed over time, as the photos submitted in this project are time stamped.

Another challenge encountered during the development of this web application is the ability to automatically submit a location on the web application utilizing the device's GPS. Instead of requiring the user to click on a point on the map, a point would be automatically placed based on the GPS location of the mobile device. This method could be more accurate than the user clicking on a point on a map, as it would utilize hardware rather than the user's discretion. The author could not locate a widget within Esri's AppBuilder with this functionality.

6.3. Scalability

The methodology used in this web application and VGI analysis is not scalable to thousands of users. While the web application can support that number of users, the experience using the web application would not be ideal. As there is feedback on the number of icons per a feature for the VGI submissions with only five users, the icons would only increase, and the viewing experience for the web application would become cluttered with VGI icons submitted by more users.

The VGI validation and quality assessment methodology used in this project is not scalable. However the processes used can be automated. A thematic analysis can be improved through defining better themes based on new and more data. It can also be automated to include a semantic analysis with a proper lexical database and software. A script could be written to remove unnecessary words for a semantic analysis and a program can be written to develop a semantic analysis on the remaining words of the VGI submissions. A control data analysis can also be automated through ArcPy and a model in ArcMap. A positional analysis based on the photographic VGI can also be automated through a model in ArcMap for a larger number of users. While the manual analysis methods of this project are not scalable to thousands of users, the methodologies can be improved through python scripts and custom models in ArcMap to scale up to a larger number of users.

6.4. Using this Web Application as a Template for Future Work

The website and web application developed in this project can be applied to other study areas. To display hiking trail data and collect VGI for other study areas, the user would have to create a new feature service with their data and be able to host their data on ArcGIS.com. Once the feature service is created and published, the user can then recreate the web application on ArcGIS.com following the steps listed in Section 4.2.1, Section 4.2.2, and Section 4.2. with the new data. A new web application can be tweaked to the user's preferences and requirements, given that it is easy to customize websites using Esri's Web AppBuilder. The author of this project can also apply the steps taken in this project to a different study area simply by changing the enterprise geodatabase to one containing data in the new study area. The study area of this project can be expanded in a similar manner as well by adding data to the enterprise geodatabase for this project. This step would retain all existing data, including VGI submissions (though it can be removed just as easily), and expand the available data for users to view the new data. After a certain extent though, the features would be too small to view properly, though this can be somewhat remedied by changing the symbology of specific features or not showing any features until the user zooms in sufficiently.

6.5. Limitations of the Project

This web application was designed for an end user to view data regarding trail features and to submit their own VGI regarding the features. While the web application provided the functionality to submit VGI, the functionality of the VGI submissions is limited, as the users are required to click on a point on the web application’s web map to create a feature. The feature would be created at the location the participant clicked, rather than create a feature based on the participant’s mobile device’s GPS location. This is due to limitations of the software choice for this project, Esri’s Web AppBuilder. The functionality built into the software for users to submit VGI does not allow an automatic creation of a feature at the location of the user. The software only allows the creation of a feature where a user clicks on a map after selecting their desired feature.

Another limitation of this project is the number of participants who tested the web application and submitted VGI. This limitation is due to the short testing window scheduled for the trial. The short testing window was chosen to eliminate as much variance as possible between testing days, to keep the features on the trail as similar as possible between different participants. As the quality of VGI depends heavily on the number of users, more users would provide better results for this type of study. More users also mean potentially more devices for the website and web application to be tested on.

However, despite the two limitations listed above, the greatest limitation would be the web application and Esri's Web AppBuilder itself. While Esri's Web AppBuilder provides an easy to use template for building a web application, that is also its limitation with regards to this project. While a web application was created to successfully fulfill this project's objective, using Esri's Web AppBuilder cannot scale up to a much greater number of users for an ideal VGI ecosystem with the methodologies used in this project. Given the feedback from the later users of this experiment, the number of features shown from other users would become increasingly intolerable for other users to view the original feature along with what other users reported on that feature. Based on the clustering of features shown in Section 5.1, with more users, the clustering would expand and possibly encroach on other similar features, making it even more difficult to distinguish two features of the same type from each other. This issue may be remedied by a widget provided through Esri's Web AppBuilder called "Summary." The "Summary" widget would cluster like features as one and would show multiple features once the user zooms into the cluster.

6.6. Future Improvements

Due to the limitations of this project's methodologies, to properly scale this project up to an ideal VGI ecosystem, a new application would have to be created for data display and VGI submissions. Rather than be constrained by the limitations of Esri's Web AppBuilder and the tools it provides, a custom Android and iOS application could be created using an application programming interface (API) developed by Esri. Due to the time and expertise required for a standalone custom application, this option was not chosen for this project. The API would allow the Android and iOS applications to interact with Esri products, such as enterprise geodatabases and ArcGIS Online.

The custom application would allow users to create new feature points whose location reflects the mobile device's location, rather than rely on the user to manually input a point on a map. Another feature the application would possess is the ability to attach a photo of a feature to the existing feature point. Given the feedback from the later participants of this study, new feature points should not be created for existing points, as it can cause confusion among the users. The new application would allow users to edit or add new records related to existing feature points if they are trying to report on an existing feature. However, the users should only be able to add their own data. Users should not edit or delete existing data.

The new application would also provide the ability for users to edit and add features offline, a necessity where there is limited or no cellular reception on certain trails. Esri provides documentation on enabling this feature through the Collector for ArcGIS application or through their REST API. As this application was developed prior to obtaining knowledge regarding cellular reception on the specific trails listed in the City of Santa Barbara's dataset, these services were not used.

As hiking trails are not limited to Santa Barbara, this study can be expanded to other areas as long as there are hiking trails and feature data available for the new study area. However, within the current study area, the existing dataset only contains data the City of Santa Barbara set out to collect. Future improvements on the data front include gathering data about new feature types more relevant to the average hiker, e.g., poison oak locations or types of flowers blooming and its locations. The potential VGI data is only limited to what the users choose to report on and what the users wish to see on a hiking trail.

A new website could also be created with more functionality and an improved design. The new website would also host a web application to view the data. The website could be developed with more features in mind, such as a trail selector based on parameters listed by a user. For the extent of this project, however, the current website fulfills all the requirements to display necessary information as well as host the web application.

VGI analysis and validation for this project were done manually, and with an increasing number of submissions, this method is not feasible. Analysis and validation of VGI could also be improved, as the process taken for this experiment is not scalable to an ideal VGI system. There would simply be too many submissions to analyze and validate manually. This process, however, can be improved by automation. With a new, custom application for VGI submissions, the application would be able to attach a photo to a point submitted by the user. This would streamline the process in analyzing the displacement of features from user-submitted VGI to the original dataset as this process can be automated with ArcPy and a custom python tool/script.

Thematic analysis of the VGI can also be automated through a custom python script. A lexical database is required to improve the analysis of the VGI. More research is required in the creation of a custom lexical database for a thematic analysis of VGI on hiking trails. A semantic analysis could also be conducted. This would provide another form of VGI validation and quality assurance. However, similar to an improved thematic analysis methodology, more research into the semantic analysis is required.

6.6.1. VGI Preparation for City Submission

The next step after validation of VGI would be to prepare the VGI for submission to the City of Santa Barbara. There are two ways to organize the collected data for submission to the City of Santa Barbara. As the template for the VGI submission was created based on the original attribute fields, there are no major changes required to the data structure. This provides the City of Santa Barbara a reference to their original dataset and the new information without disabled fields.

The first approach is to submit the entire collection containing the original dataset along with all VGI in a single file. The second approach is to separate the VGI from the original dataset and submit only the VGI to the City of Santa Barbara. As the VGI submissions are added to the original geodatabase, separating only the VGI can be done by running a query on the attribute table and selecting the unique identifiers of all features past the highest original number. This is possible because the City of Santa Barbara's dataset is constant and their unique identifiers do not change. The VGI submissions have unique identifiers created by ArcMap that are greater than the numerical values of the original City of Santa Barbara's dataset.

Once the VGI is separated, employees at the City of Santa Barbara can filter and sort the VGI to review what was submitted. A social approach to VGI can be used, with the high-level users being comprised of employees of the city. The employees would oversee the VGI submitted by the users, similar to the approach laid out by Goodchild and Li in Section 2.1.2. Any issues with the VGI would be resolved by the employees and the level of quality of the VGI would rely on city employees.

Grouping VGI by participant would allow the employees to view how people reported on features across an entire trail much easier than grouping by individual features. Grouping by individual features, however, would provide the employees an easier time to update a single feature at a time, as all the information would be in one dataset. While there is no direct association between a specific feature and the VGI submissions, based on the clustering of VGI submissions shown in Figure 10 and Figure 11, it can be determined visually what feature a VGI submission is referring to due to its proximity and similarity of feature type to features in the original dataset. This solution does present more effort involved on the part of the employees, as the employees would have to classify all VGI submitted to their respective feature. However, if a many to one relationship was available through a widget, this process would not be necessary.

6.7. Final Words

This project developed a website and a web application to view and collect VGI about hiking trails and features. The VGI was analyzed to determine its accuracy and validity through three different analysis methods. This project fulfilled its objective through the web application.

References

Aronoff, Stan. 1989. *Geographic Information Systems: A Management Perspective*. Ottawa: WDL Publications.

Belanche, Daniel, Luis V. Casaló, and Miguel Guinalíu. 2012. “Website Usability, Consumer Satisfaction and the Intention to Use a Website: The Moderating Effect of Perceived Risk.” *Journal of Retailing and Consumer Services* 19 (1): 124–32. doi:10.1016/j.jretconser.2011.11.001.

Braun, Virginia, and Victoria Clarke. 2006. "Using Thematic Analysis in Psychology." *Qualitative Research in Psychology* 3 (2): 77-101.

Chin, Sin-Ho. 2013. “Examining the User Satisfaction on Web APP in LUI, PUI, and GUI.” *International Journal of Computer and Communication Engineering* 2 (3): 358–62. doi:10.7763/IJCCE.2013.V2.204.

Comber, Alexis, Linda See, Steffen Fritz, Marijn Van der Velde, Christoph Perger, and Giles Foody. 2013. “Using Control Data to Determine the Reliability of Volunteered Geographic Information about Land Cover.” *International Journal of Applied Earth Observation and Geoinformation* 23 (1): 37–48. doi:10.1016/j.jag.2012.11.002.

Corney, J. R., C. Torres-Sánchez, A. P. Jagadeesan, X. T. Yan, W. C. Regli, and H. Medellin. 2010. “Putting the Crowd to Work in a Knowledge-Based Factory.” *Advanced Engineering Informatics* 24 (3): 243–50. doi:10.1016/j.aei.2010.05.011.

Elwood, Sarah. 2008. “Volunteered Geographic Information: Future Research Directions Motivated by Critical, Participatory, and Feminist GIS.” *GeoJournal* 72 (3–4): 173–83. doi:10.1007/s10708-008-9186-0.

Fling, Brian. 2009. *Mobile Design and Development*: *Practical Techniques for Creating Mobile Sites and Web Apps*. Sebastopol, CA: O'Reilly Media, Inc.

Goodchild, Michael F. 2007. “Citizens As Sensors: The World of Volunteered Geography.” *GeoJournal* 69 (4): 211. https://doi.org/10.1007/s10708-007-9111-y.

Goodchild, Michael F. 2008. "Commentary: Whither VGI?" *GeoJournal* 72 (3-4): 239-44. doi:10.1007/s10708-008-9190-4.

Goodchild, Michael F., and Linna Li. 2012. "Assuring the Quality of Volunteered Geographic Information." *Spatial Statistics* 1 : 110-20. doi:10.1016/j.spasta.2012.03.002.

Guptill, Stephen C., and Joel L. Morrison. 1995. *Elements of Spatial Data Quality*. Tarrytown: Elsevier Science.

Haklay, Mordechai. 2010. “How Good Is Volunteered Geographical Information? A Comparative Study of OpenStreetMap and Ordnance Survey Datasets.” *Environment and Planning B: Planning and Design* 37 (4): 682–703. doi:10.1068/b35097.

Haklay, Mordechai, and Patrick Weber. 2008. "OpenStreet Map: User-Generated Street Maps." *IEEE Pervasive Computing* 7 (4): 12–18. doi:10.1109/MPRV.2008.80.

Heipke, Christian. 2010. "Crowdsourcing Geospatial Data." *ISPRS Journal of Photogrammetry and Remote Sensing* 65 (6): 550-57. doi:10.1016/j.isprsjprs.2010.06.005.

Humanitarian OpenStreetMap Team. 2017. "Haiti." Accessed October 23, 2017. https://www.hotosm.org/projects/haiti-2.

Jabreel, Mohammed, Antonio Moreno, and Assumpció Huertas. 2017. "Semantic Comparison of the Emotional Values Communicated by Destinations and Tourists on Social Media." *Journal of Destination Marketing & Management* 6 (3): 1–14. doi:10.1016/j.jdmm.2016.03.004.

Jiang, Bin, and Jean-Claude Thill. 2015. "Volunteered Geographic Information: Towards the Establishment of a New Paradigm." *Computers, Environment and Urban Systems* 53 : 1-3. doi:10.1016/j.compenvurbsys.2015.09.011.

Johnson, Peter A., and Renee E. Sieber. 2013. "Situating the Adoption of VGI by Government." In Sui, Daniel Z., Sarah Elwood, and Michael Goodchild, eds. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Dordrecht: Springer. 65-81.

Lai, Chengting, Xiang Robert Li, and Rich Harrill. 2013. "Chinese Outbound Tourists ' Perceived Constraints to Visiting the United States." *Tourism Management* 37 : 136–46. doi:10.1016/j.tourman.2013.01.014.

Latonero, Mark, and Irina Shklovski. 2010. "Respectfully Yours in Safety and Service - Emergency Management & Social Media Evangelism." *Proceedings of the 7th International ISCRAM Conference*, Seattle, USA. Available at SSRN: https://ssrn-com.libproxy1.usc.edu/abstract=1566423 or http://dx.doi.org.libproxy1.usc.edu/10.2139/ssrn.1566423

Lin, Wen. 2013. "When Web 2.0 Meets public Participation GIS (PPGIS): VGI and Spaces of Participatory Mapping in China." In Sui, Daniel Z., Sarah Elwood, and Michael Goodchild, eds. *Crowdsourcing Geographic Knowledge: Volunteered Geographic Information (VGI) in Theory and Practice*. Dordrecht: Springer. 84-85.

Marita, S., A. Anu, T. Makarand, Y. Supriya, and R. Vipula. 2016. "A Single App for all Your Treks." *International Journal of Computer Applications* 140 (12): 40-44. doi:10.5120/ijca2016909527.

McKee, Alan. 2003. *Textual Analysis: A Beginner's Guide*. Thousand Oaks, CA: SAGE Publications.

Mohajeri, S., B. A. Perkins, P. L. Brubaker, and M. C. Riddell. 2015. "Diabetes, Trekking and High Altitude: Recognizing and Preparing for the Risks." *Diabetic Medicine* 32 (11): 1425–37. doi:10.1111/dme.12795.

OpenStreetMap. 2017. "Wikimapia." Page laste edited February 20, 2017. Accessed October 23, 2017. http://wiki.openstreetmap.org/wiki/Wikimapia.

Roberts, Carl W. 2000. "A Conceptual Framework for Quantitative Text Analysis: On Joining Probabilities and Substantive Inferences about Texts." *Quality and Quantity* 34 (3): 259-274. doi:10.1023/a:1004780007748.

See, Linda, Dmitry Schepaschenko, Myroslava Lesiv, Ian McCallum, Steffen Fritz, Alexis Comber, Christoph Perger, et al. 2015. "Building a Hybrid Land Cover Map with Crowdsourcing and Geographically Weighted Regression." *ISPRS Journal of Photogrammetry and Remote Sensing* 103 : 48–56. doi:10.1016/j.isprsjprs.2014.06.016.

Smith, Craig. 2016. "flickr Stats (November 2016)." Page last updated November 14, 2016. Accessed October 23, 2017. https://expandedramblings.com/index.php/flickr-stats/

Tele Mart. 2017. "Tele Atlas MultiNet." Accessed October 23, 2017. http://www.tele-mart.com/teleatlas_multinet.php.

Tobler, Waldo R. 1970. "A Computer Movie Simulating Urban Growth in the Detroit Region." *Economic Geography* 46 : 234-240.

van Oort, Pepijn. 2006. *Spatial Data Quality: From Description to Application*. PhD dissertation, Wageningen, Netherlands: Wageningen University.

Vastaranta, Mikko, Eduardo Latorre, Ville Luoma, Ninni Saarinen, Markus Holopainen, and Juha Hyyppä. 2015. "Evaluation of a Smartphone App for Forest Sample Plot Measurements." *Forests* 6 (4): 1179–94. doi:10.3390/f6041179.

Wikipeida. 2017. "Volunteered Geographic Information." Page last edited August 17, 2017. Accessed October 23, 2017. https://en.wikipedia.org/wiki/Volunteered_geographic_information.

Wilkinson, Dennis M., and Bernardo A. Huberman. 2007. "Assessing the Value of Cooperation in Wikipedia." *First Monday* 12 (4). doi:10.5210/fm.v12i4.1763.

Zichermann, Gabe, and Christopher Cunningham. 2011. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. Sebastopol, CA: O'Reilly.

Zielstra, Dennis, and Alexander Zipf. 2010. "A Comparative Study of Proprietary Geodata and Volunteered Geographic Information for Germany." In *Proceedings of 13th AGILE International Conference On Geographic Information Science*, Guimarães, Portugal.