Developing a Data Collection and Data Management Methodology Using The Esri Collector App, AWS EC2, PostgreSQL, PostGIS, and QGIS

Luis Saenz

Master of Science in Geographic Information Science (MSGISci)

Department of Geography, California State University, Long Beach

Introduction

Geospatial data collection, storage, and management are vital components for any organization that utilizes geographic information systems (GIS). Ground surveying data collection can be a very expensive and time consuming process, however it is the best method to capture accurate and updated point locations with corresponding detailed attributes. There is great value in the development of an efficient data collection process, however of perhaps greater value is the development of a proper storage and management practice that supplements the data captured. The value of the data increases when it is readily accessible to practitioners who can then assess, analyze, visualize and disseminate the data for a variety of purposes on a variety of platforms.

An effective and efficient data collection process is important and necessary for any GIS practicing organization. Many of the components necessary to develop a process for data collection, storage and management can be considerably expensive yet invaluable to an organization. The GIS community has observed a recent growth in the availability, compatibility and popularity of free and open source software that has been revolutionizing GIS data collection, storage and management processes in many organizations. Open source software has been providing competing options to proprietary software and challenging the way the GIS community is thinking about data processes. Open source software offers software that is more affordable and more customizable to fit the needs of the project at hand, however sacrificing a thorough customer support that is characterized with proprietary software. Developing a system in which an organization can collect accurate data with time and economic efficiency can reduce significant costs and increase productivity in a GIS practicing organization.

In this applied research project I examine the extent to which free and open source software can contribute to the GIS community through the development of a system to collect, store and manage the data of the emergency phones on the California State University, Long Beach campus.

Data and Data Sources

The project used two sources of geospatial data. The CSULB campus Aerial Imagery was captured by LARIAC 3 and the data was available for this project through the California State University, Long Beach. The the Emergency Phones on the CSULB campus file was created through field work as part of this project's outcomes. Data was collected by Luis Saenz (author) and two volunteers via the use of Esri's ArcGIS online and the "Collector App using IPhone 4 mobile devices

Table 1. List of data and data sources used in the project

Dataset	Source
Emergency Phones on Campus	Collected through field work by Luis Saenz (author) and volunteers
Campus Aerial Imagery	LARIAC 3 via California State University, Long Beach

Timeline

Activity	Time
Learn about data acquisition, Servers, Databases, Open Source, Configurations	June 15 - June 30th
Develop the Framework	July 1st - July 22nd
Data Preparation and Collection	July 20th - July23rd
Data Management	July 24th - August 4th
Data Visualization	August 5th - August 10th

Methodology

Database Development

Amazon Web Service (AWS) EC2 cloud server space was rented to host the database for this project. The open source Relational Database Management System (RDBMS) PostgreSQL with the PostGIS extension was configured on the server to manage the database and store geospatial data. The open source QGIS desktop GIS software was configured to communicate with the server to upload and retrieve the data from the PostgreSQL/PostGIS database.

Field Data Collection:

ArcMap was used to prepare for the data collection process. Using ArcMap, a new feature class was created for the CSULB emergency phones with attribute fields to store the desired attributes. Upon creating the fields, domains were assigned to the fields in order to provide a drop down menu which would simplify the data collection process and increase efficiency when collecting information in the field. I created fields for the following attributes: phone ID, condition of the phone, color of the phone, a picture of the phone, and the name of the person who collected the point. The prepared point feature class was then uploaded to ArcGIS Online. After data collection was completed the data was then downloaded as a geodatabase, managed with QGIS and through this software uploaded to the PostgreSQL/PostGIS database on the AWS EC2 server.

Figure 1. *This diagram depicts the process of the methodology used for this project.*

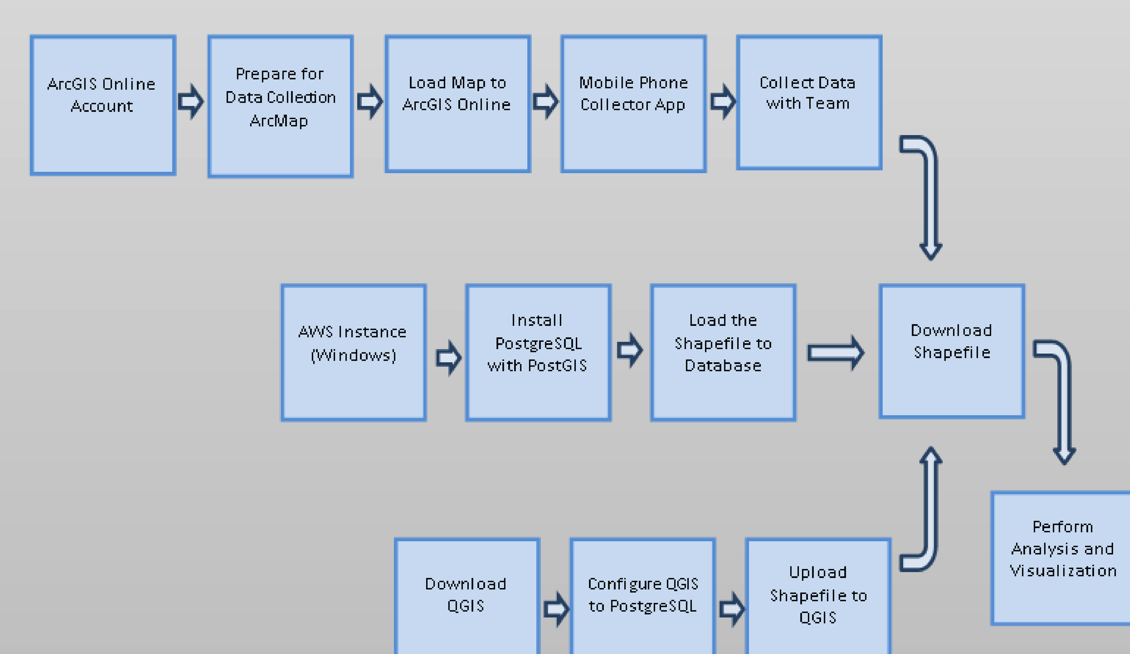
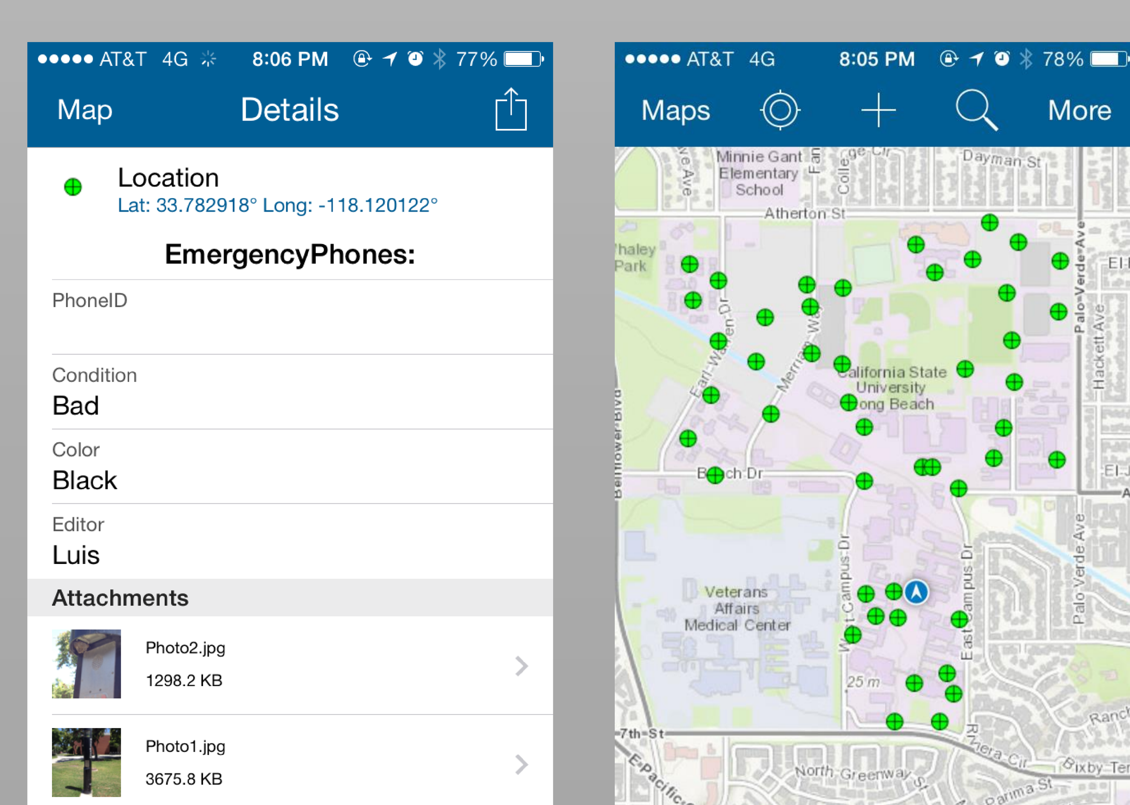


Figure 2. *The images below show the interface of ArcGIS Collector Application when collecting data. Image on the left shows the attributes collected of a given feature. The image on the right shows all the data points of the emergency phones collected on the CSULB campus.*



Results

The results for this applied research project can be divided into two categories: (1) the framework for data collection, storage, and management and (2) the CSULB emergency phone data obtained by utilizing this framework.

This project was successful in producing a viable framework for data collection, storage, and management using the Esri Collector App, ArcMap, AWS EC2, PostgreSQL, PostGIS, and QGIS. In the end, it was possible to upload geospatial data to the PostgreSQL/PostGIS database indicating that software programs selected for this framework are compatible and that this method is viable.

The framework that was developed through this project allowed for the successful collection of geospatial data for the emergency phones on the CSULB campus. In total, 46 emergency phones were located and attribute information was filled out for each one. In some cases, a phone ID could not be determined but all other attributes were successfully described.

Figure 3. *The two figures below are images that are part of the data collected using Esri's Collector App. Image on the left shows the typical image collected for each feature. The image on the right is an image of a beehive on an emergency phone that was encountered when collecting the data*

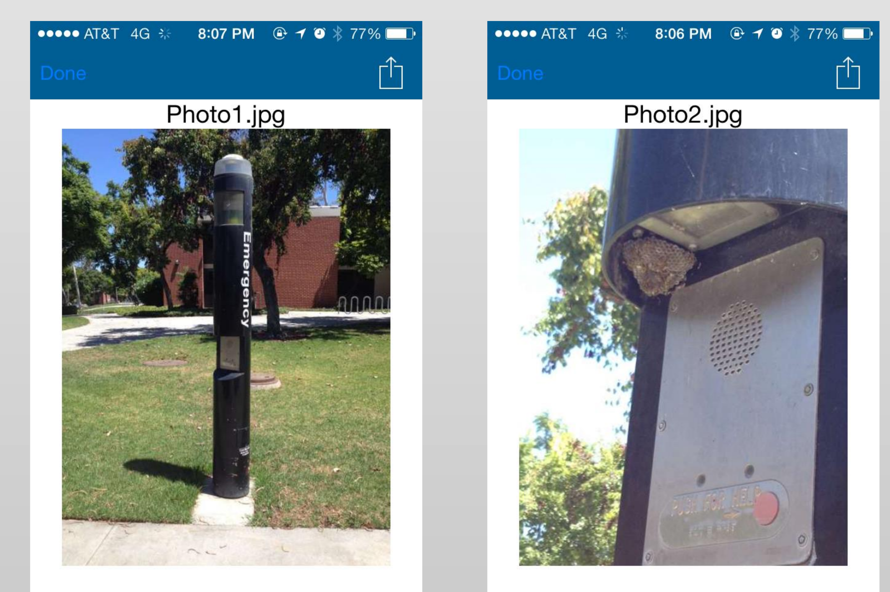


Figure 4. *The image below shows the attributes of the forty six collected point features. The left part of the image displays the table on the database. On the right side of the image, the attribute table of the same file is displayed via QGIS.*

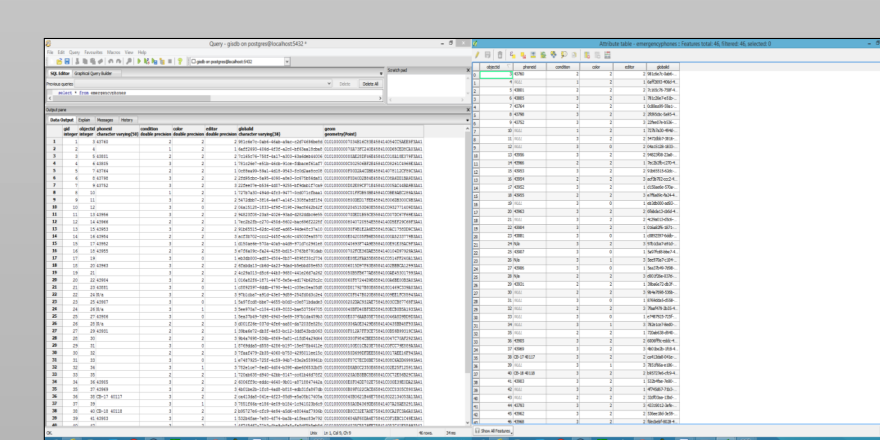
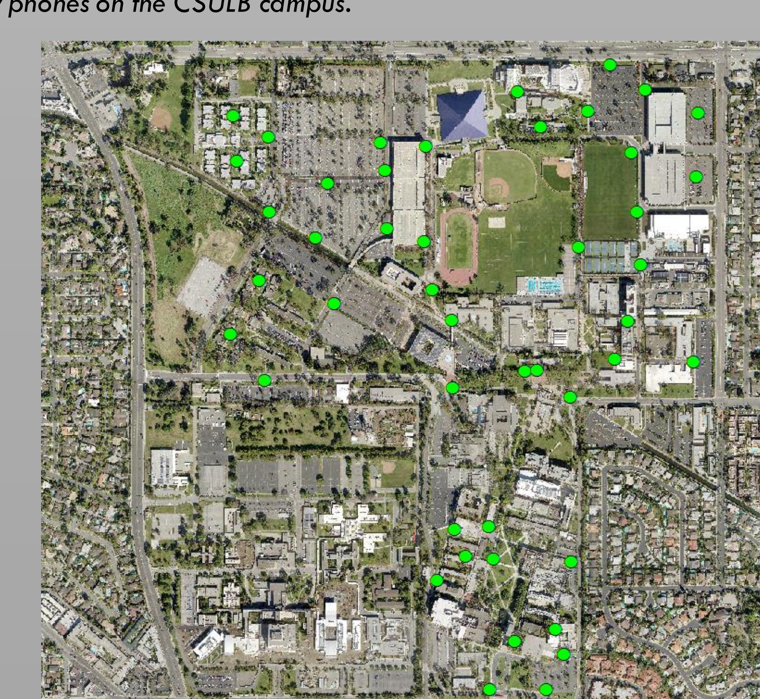


Figure 5. *The image below shows the shapefile created overlayed on the campus aerial imagery. This image shows all forty six collected point features of the emergency phones on the CSULB campus.*



Discussion

Considering the amount of time spent while data collecting, the limited amount of data collectors, and the amount of space to cover it is likely that not 100 percent of the emergency phones on the CSULB campus were collected. Emergency phones that are located in the parking structures were not collected given to the inability to reach GPS reception while being under a roof. Construction on the CSULB campus may have made emergency phones unavailable for data collection.

The accuracy of the mobile phones GPS is not as precise as anyone would like. The points collected using iPhone 4's (all three) appear to be anywhere from five to fifteen feet from their actual locations notably in locations where there are overhangs or next to large buildings. In terms of the data visualization, scale plays an important role. For example when the results and the CSULB campus are seen at a small scale (zoomed in) the inaccuracy is clear and evident, however when looked at from a large scale (zoomed out) perspective the inaccuracies are not as clear. Therefore these results can potentially be used for large scale projects where the general location of the emergency phone is sufficient, however the data may not be adept for small scale projects where the exact location of feature is necessary.

One of the greatest challenges that I faced in this project was configuring QGIS, PostGIS, and AWS to communicate with one another effectively such that files can be accessed and uploaded via QGIS on a local machine. When accessing data that is on a server, the user is essentially making a remote desktop connection via QGIS on their local computer, therefore the program and machine must have the appropriate credentials and permissions to access the server. Having QGIS configured to communicate with the database, data can then be uploaded.

Conclusion

The project outcomes are the developed system of data collection, storage and management, as well as the data collected through the execution of the system. The model utilizes a combination of open source and proprietary software to create an effective framework for data collection in the field, centrally storing data, and making the data available to multiple users. This framework can serve as a model for startup GIS companies that may not have the capital means to provide for full proprietary software. Organizations looking to reduce expenses may also consider using some of the methodologies used in this project.

The data created in this project will contribute to the growing knowledge of the CSULB campus. The location and attribute data of the emergency phones on the CSULB campus can be used by the Facilities Management department at CSULB for maintenance and accountability purposes. The location of these emergency phones will help create a safer campus, and therefore a better learning and working environment for students and faculty alike.

Future work can be focused on the difference of accuracy between mobile data collection through GPS versus collection through heads up digitizing in the field. The implementation of more open source software as it continues to be developed and compatible with GIS will be of interest for future work.

Submitted in partial fulfillment of the requirements of the Masters of Science in Geographic Information Science(MSGISci), August 15, 2015.

For additional information please contact: Luis Saenz, saenzluis54@gmail.com