Requirements Specification

## December 2nd, 2019

## Canopy - Team 11

**Team Members:**

Robert Plueger

Nicholas Lopez

Dongyu Xia

Maria Granroth

**Sponsor:** Dr. Patrick Jantz

**Mentor:** Scooter Nowak

## 

## 

## 

## 

This document has been accepted as the baseline requirements for the project.

Team Lead: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_

Sponsor: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Date: \_\_\_\_\_\_\_

Version 1

Table of Contents

Introduction ……………….……………….……………….……………….……………….……………….…………..…. 2

Problem Statement ……………….……………….……………….……………….……………….……………….…... 2

Solution Vision ……………….……………….……………….……………….……………….……………….………….. 3

Project Requirements ……………….……………….……………….……………….……………….………….. 4 - 8

Functional Requirements ……………….……………….……………….……………….………….. 4 - 6

1 Use GEDI data ……………….……………….……………….……………….…………………….. 5

1.1 User login ……………….……………….……………….……………….……………... 5

1.2 Integrity check, data cleaning ……………….……………….………………. 5

1.2.1 Perform and format analysis ……………….……………….…... 5

2 Allow users to specify an area to analyze ……………….……………….…………... 6

3 Allow users to choose a variety of statistical functions ……………….……... 6

4 Provide the analyzed data back to user ……………….……………….……………… 6

4.1 User can choose what to download ……………….……………….………. 6

4.1.1 Data guide provided alongside data ……………….…………. 6

5 Be easy to use ……………….……………….……………….……………….…………………….. 6

Performance Requirements ……………….……………….……………….……………….…………….. 7

Environmental Constraints ……………….……………….……………….……………….…………….. 8

Potential Risks ……………….……………….……………….……………….……………….……………….………….. 9

Researcher ……………….……………….……………….……………….……………….……………….……… 9

Policy Maker ……………….……………….……………….……………….……………….……………….…… 9

Project Plan ……………….……………….……………….……………….……………….……………….……………… 10

Execution Plan ……………….……………….……………….……………….……………….………………. 10

Next Phase ……………….……………….……………….……………….……………….…………………….. 10

Conclusion ……………….……………….……………….……………….……………….……………….……………….. 11

Introduction

We are Canopy, a 2019 - 2020 Northern Arizona University Computer Science capstone team. Our team members are Robert Plueger (team lead), Maria Granroth, Nicholas Lopez, and Dongyu Xia. Our project is to develop an application for characterizing 3D vegetation structure in tropical ecosystems. We will analyze the data of the three dimensional distribution of stems, branches, and leaves obtained from the Global Ecosystem Dynamics Investigation (GEDI). GEDI is funded by NASA, an Earth Ventures mission that started in December 2018. GEDI will acquire billions of vegetation profiles across the Earth’s temperate and tropical ecosystems in two years.

The sponsor of this project is Dr. Patrick Jantz. He is a member of the Vegetation Structure as an Essential Biodiversity Variable (VSEBV) project based at NAU. He aims to use GEDI data to develop a vegetation structure essential biodiversity variables (EBVs) that can be used by policy makers and scientists to improve land use decisions and guide priorities for conservation of biodiversity in tropical landscapes. The reason for launching this project is that current workflow is clumsy and requires a lot of manual work. He hopes an application can help improve processing speed and accuracy while reducing repetition and manual steps. We envision an application that supports two different end users and usage scenarios: Data Ingestion and Analysis, and Data Viewing and Access. Data ingestion and analysis can provide a set of graphical interfaces that streamline the data acquisition and integration process. Data viewing and access can provide clear interfaces for browsing, selecting, graphing, and otherwise visualizing the growing data available in the system.

Problem Statement

Our client’s current workflow is manual and is coded in R. Our client would like GEDI data analysis to be possible for anyone to work with. With its current shortcomings, which are listed below, most users would not be able or willing to analyze GEDI data.

The current system is failing to meet expectations as follows:

* The data is being processed in R, which causes accessibility issues
* Analyzing the data is currently tedious and needlessly complex
  + - The user must locate relevant NASA files by hand
    - Files must be loaded and processed one at a time, requiring operator oversight
* Output in unorganized and esoteric layout

Solution Vision

Our solution is to build a web application that allows users to see forest structure and biomass levels for a region that they choose. It will analyze the data for them and help them understand the results.

We will be doing the following as key project features:

* Convert existing code to Python scripts, hosted on Microsoft Azure
* Create a region selection map
* Pull from AppEARS database API
* Process this data into a useful state
* Present data to the user in an easy to understand way

Reference *Diagram 2* on page 8 to see how we expect our system to flow.

Project Requirements

This section will detail the complete requirements for the system described. These requirements will guide the process of building the web application. Requirements were acquired through client meetings performed roughly every other week.

The purpose of this web application is to make statistical analysis of GEDI data easily accessible to a wide audience. In order to achieve this, domain-level requirements need to be established. These requirements can be described as what the user needs from the web application in order for it to fulfill its purpose. The domain-level requirements for this project are as follows:

1. Use GEDI data
2. Allow users to specify an area to analyze
3. Allow users to choose a variety of statistical functions
4. Provide the analyzed data back to the user
5. Be easy to use

These requirements will be broken down further in the following sections.

Functional Requirements

The domain-level requirements can be expanded into functional requirements. These are the functions that the team must provide in order for the project to be considered complete. Please refer to *Diagram 1*, pictured below, for a hierarchical representation of how these requirements are broken down into more detailed requirements. This section will describe each of the requirements pictured in the diagram.

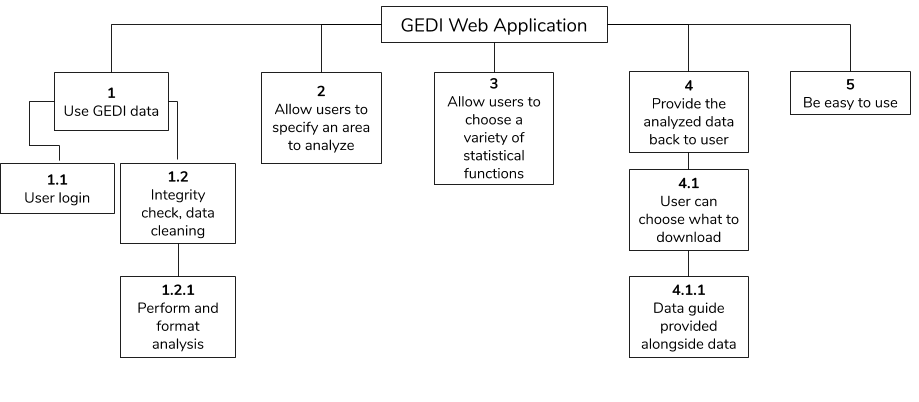


Diagram 1

1 Use GEDI data

The web application must use the most up-to-date GEDI data available to the public.

1.1 User login

In order to download and analyze GEDI data, NASA requires an Earthdata user login. There are two ways that the project can approach this. Either the web application can use Earthdata’s Login API to have the user log in, or if that fails, it can use the project client’s login to grab the GEDI data on the user’s behalf.

1.2 Integrity check, data cleaning

The system should check the GEDI data flags to see if the data is appropriate to use. If the user provided a shapefile, then the system should clean it if it is not formatted correctly. This is to ensure that the analysis is performed as correctly as possible.

1.2.1 Perform and format analysis

Once the GEDI data is received and the user has specified the area and statistical functions they wish to use, the system must perform and format the analysis appropriately.

2 Allow users to specify an area to analyze

Users must be able to specify the area that they wish to analyze. This should be done so that the user has plenty of options available to them, which allows flexibility for what they need. For example, this could include a map, menu, and an option where the user can upload their own shapefile that specifies the area.

3 Allow users to choose a variety of statistical functions

Users must be able to choose how they want their data analyzed. They may choose one or several functions to have performed on the area that they specified. Examples of the functions that a user may choose from are listed below.

* Number of Observations
* Time of Year
* Density of Observations
* Average Canopy Height
* Standard Deviation of Canopy Height
* Data Quality
* Forest Structures

4 Provide the analyzed data back to user

Once the system has analyzed the data, it must provide that data back to the user for download. The user may use the resulting data however they want.

4.1 User can choose what to download

The user must be able to specify what they want downloaded. This could include the original GEDI data or the analysis by itself.

4.1.1 Data guide provided alongside data

There must be a guide, visual or report, produced by the system and provided to the user to help explain the analysis results.

5 Be easy to use

Since the purpose of the project is to make GEDI data analysis more accessible to a wider audience, the accessibility and the ease of use of the web application is extra important. The client would prefer a minimalistic style. All functions on the application must be easy to find, use, and understand.

We are confident that we will be able to meet these requirements by May 2020.

Performance Requirements

Not every web application does everything at the same speed. Some processes take longer than others naturally and vary from platform to platform. Our project has requirements about how well and how quickly some components of our system should perform.

We have two main components of our system to take into consideration regarding speed. Both the statistical analyses and user downloads will require different processing time. However, as shown in *Diagram 2*, both of the processes start with the same steps, represented by the green arrows. The process of sending a request to our web application to then request data from NASA should be able to complete in a matter of seconds.

The statistical analyses steps are represented by the blue arrows. The process of receiving data from NASA, analyzing the data, and sending it to the user takes a variable amount of time. This amount of time depends on how many files are received from NASA. The analyses should take roughly one hour if 50,000 files are received. As each request will be different, we will only be able to estimate that an analysis may take up to three days if the information is requested for an entire country / continent.

The user’s data downloading process is represented by the orange arrow. The amount of time this process takes is more difficult to control, as it is dependent on the user’s internet speed. It is difficult to determine the exact time that it will take to download data. Different download speeds will require different amounts of time to complete the data transfer.

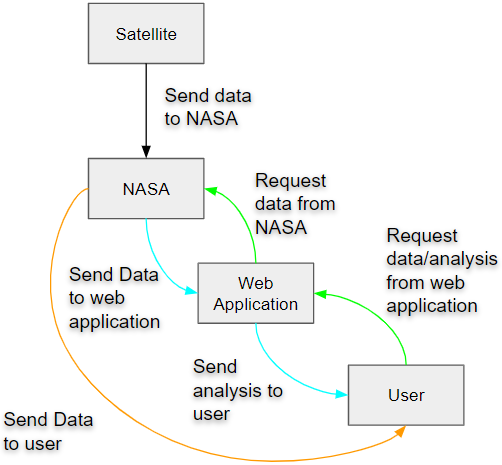


Diagram 2 - Process of User Requesting Information

Environmental Constraints

There are some things that can limit or tighten the bounds of a project. These limitations can be imposed by the client, hardware, or software compatibility.

Our client has expressed his desire to use either Python or R as our programming language. This is due to his familiarity with these languages and his desire to continue this project after our collaboration.

We are additionally constrained by the project’s budget. A portion of the budget will be devoted to hosting our web application. Hosting a web application through Microsoft Azure requires a regular payment. The price paid determines the amount of storage and connection speed our application can have.

Potential Risks

There are some challenges we may encounter that could create risks for our project and our end users. These risks differ based on which user comes across these challenges.

Researcher

Researchers may use our website and its data to gather information for a report. Such a report could be published in scientific journals, magazines, or as a standalone research paper. A challenge that we can come across is producing incorrect analyses of data. If our analyses produce incorrect information, then a researcher risks using this information to publish incorrect findings. This can result in the diminished credibility of that researcher and our website. Additionally, another researcher could use that published information and publish something else. This would create a chain of misinformation and lost credibility. Although this is moderately severe, the likelihood of this happening is quite low, as we can do extensive testing to make sure our analyses will be correct.

Policy Maker

Policy makers may use our website and its data to determine the best course of action for certain environmental areas. They may decide to increase conservation efforts of an ecosystem, costing money and resources. They may also decide to let an ecosystem be destroyed or removed if it is decided to be less valuable or less sustainable. We have the same challenge as before, potentially producing incorrect analytic information, but we also have the challenge of providing inadequate information. This stems from only knowing about the vegetation of an ecosystem but not the residing wildlife. Both of these challenges create the risk of a policy maker deciding the wrong environmental action to take. They may decide to destroy an ecosystem that is very valuable or they may decide to increase conservation efforts on an ecosystem that is not valuable or sustainable. The likelihood of providing inadequate data is moderate, as we cannot prevent policy makers from taking action aside from warning them that wildlife may be present. The severity of destroying a valuable ecosystem is severe: there could be global environmental impacts. Protecting a less valuable ecosystem has a low severity, as there will be no environmental damage, but resources will be used without profit.

Project Plan

This section will briefly discuss our project execution plans as it stands right now, and then there will be a Gantt chart to describe some milestones in our plans for the next phase.

Execution Plan

After many exchanges with our client, we have determined the client's needs. We are currently completing the Requirements Specification document based on information received from the client. At the same time, we are also preparing a Technical Prototypes Demonstration. We are making this demo according to the requirements of the client. This demo will have a basic GUI, can perform basic mathematical operations on virtual data, and upload and download files through the GUI.

Next Phase

For next semester, we designed a rough schedule. In February, we plan to finish the function that will allow the user to specify an area and the data analysis they want. Then, in March we are going to finish the data guide and download functions. In April, we will finish the login function and communicate in depth with Dr. Jantz about the product. In May, we will improve the product, and try to develop a perfect application.

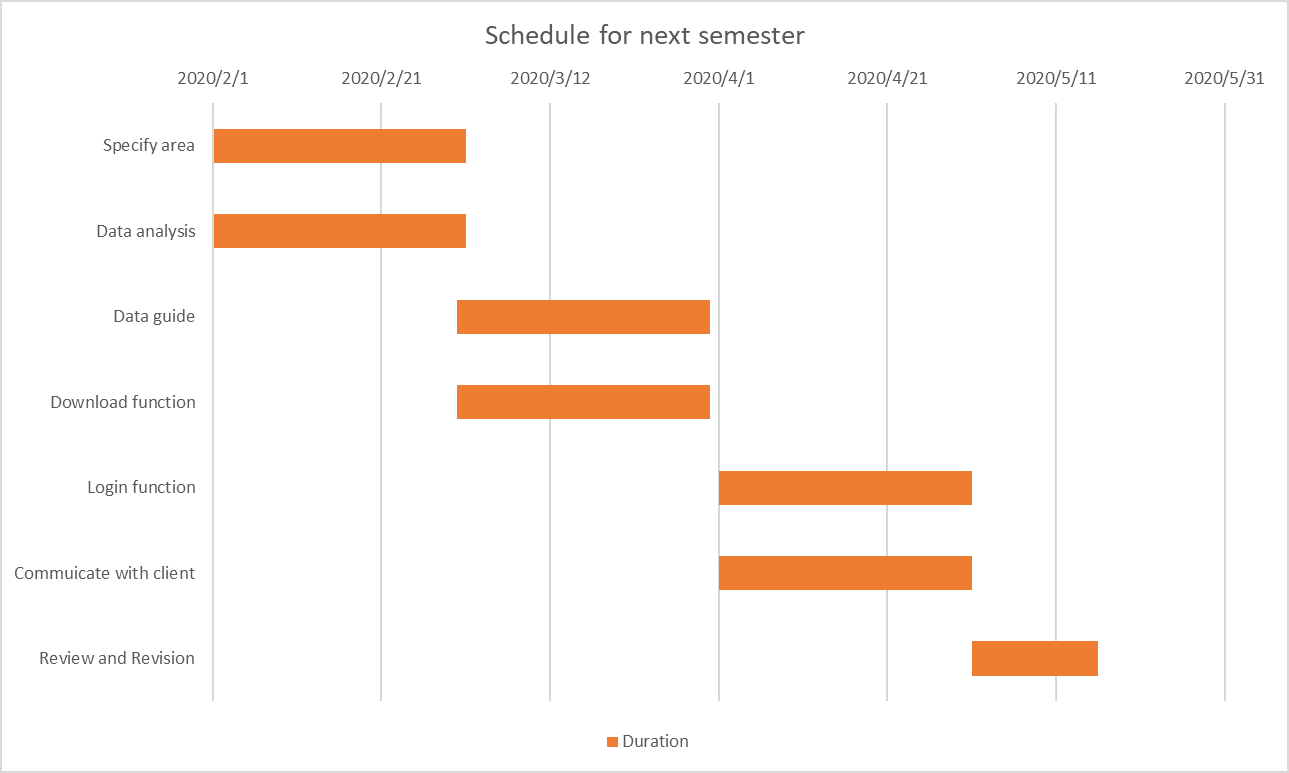


Diagram 3-schedule for next phase

Conclusion

The vegetation structure that our client Dr. Jantz studied can be used to improve land use and conserve the biodiversity of tropical landscapes. Our client’s current process is manual. To improve this, we will create a web-based system that can provide this data in an easily accessible form which will include allowing the user to specify an area and the data analysis they want, a data guide, download results and a login. We completed this document according to the client's requirements. In this document, we declared the problem and discovered the solution, then analyzed how to complete our project. After that, we put forward the expected performances and potential risks. We provided an expected project schedule to wrap up the document. This document will help us to develop our product. We trust that we will eventually be able to develop a perfect product that satisfies Dr. Jantz.