COS 397: Computer Science Capstone I

System Design Document

(Adapted from Susan Mitchell)

BirdSpotter
System Design Document

Table of Contents

1. Introduction

1.1 Purpose of This Document

The purpose of this document is to define system design of this application. This information is provided to assist the developers in the creation and maintenance of the architecture of the system. The document lays out the system architecture, information organization, and information storage architecture.

1.2 References

This document references:

- System Requirement Specification document: https://drive.google.com/file/d/1gdiXsWQ8f2nXvK7SPesUauaKTBtPlzuF/view?usp=sharing
- User Interface Design Document: To be completed next

2. System Architecture

2.1 Architectural Design

The logical system architecture follows the "Model View Controller" design pattern, where the model is the stateful part of the system, the view describes what users see, and the controller describes system inputs, outputs, and system manipulation. This design pattern as applied to this system can be seen in Figure 1.

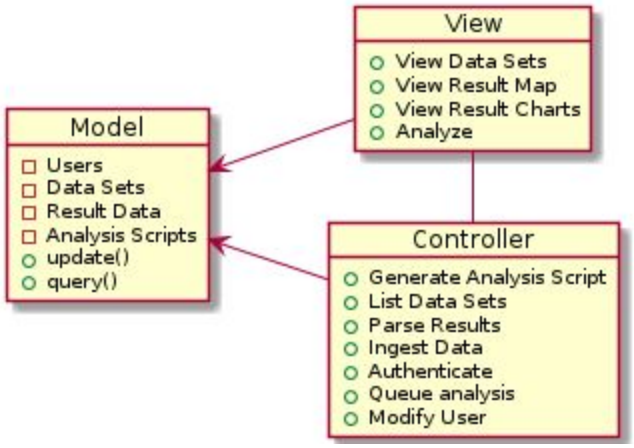

Figure 1. Logical System Architecture

As is shown in Figure 1, the system has three logical components. The first component is the view, which contains the views a user sees such as the list of data sets or the data on a map. The second is the controller, which handles inputs and outputs between the components and the user, and the model, which contains the state of the system, such as the user list, data sets, result data, and analysis scripts.

The technology architecture shows the technological layout of the system, at a software/hardware level. As can be seen in Figure 2, the system has 3 major technological components: the data storage (a relational database), the file storage, the web server, and the ACG computation system.

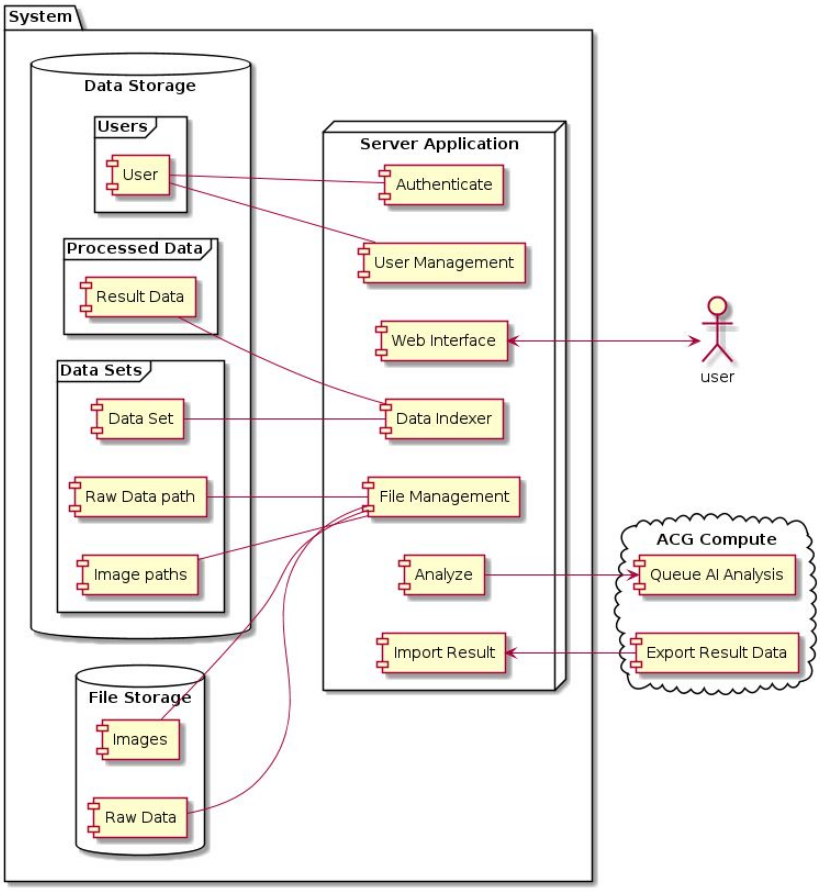

Figure 2. Technological System Architecture

As Figure 2 shows, most of the system model is contained within the storage, while the web server application is the controller, and serves the view to the client via a web interface. Analysis is also sent to an external actor, the ACG, for computation.

2.2 Decomposition Description

Model: This section is the structure of the application, and consists primarily of the users, input and output datasets, and analysis scripts, which are external to the scope of the COS397 development, but will need to be accessed. The users are split into four separate levels of access: Public, which can only view the visualizations, Registered, which can also download the data, Privileged, which can also upload data and analysis of the data, and finally Administrator, which can also modify other users' permissions as well as adding algorithms for data processing to the backend. Datasets will effectively be the input to the machine learning portion of this project; a compilation of UAV collected imagery. Result Datasets will be a combination of spreadsheets and shapefiles, similar to what is contained in the "Demariscove" folder of the COS397 Seabird

Project folder in google drive. Analysis Scripts will be the machine learning scripts provided by the machine learning portion of the research. Finally update() and query() will update and request data from the Datasets and Result datasets.

View:

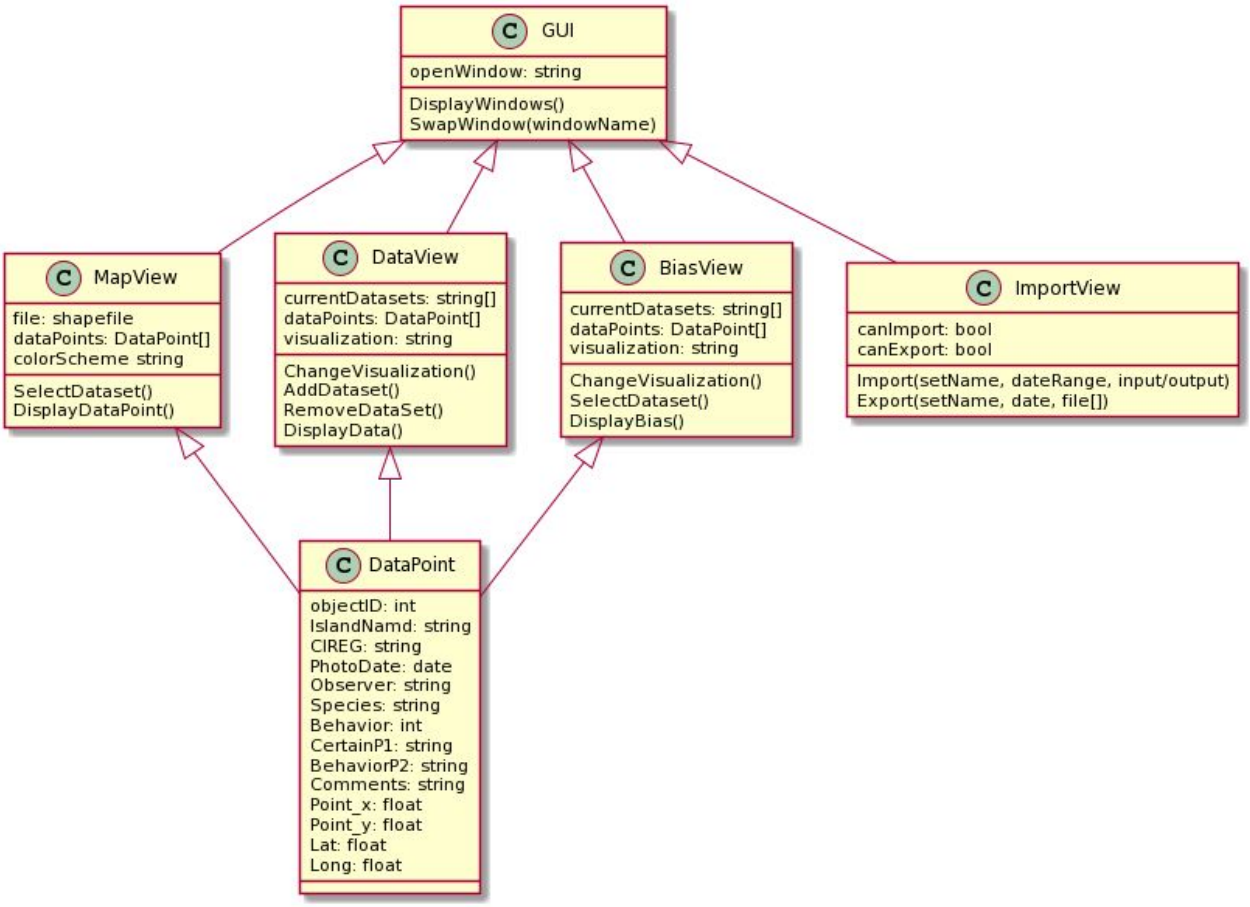

Figure 3: View Decomposition Diagram

This section is effectively the GUI of the application, and there is a breakdown in the above diagram: “View Decomposition Diagram” as well. It will be split into 4 views, and allow for various forms of data visualization. The first of these is the Map View, where given a shapefile and related spreadsheet file, will display the data points on a map determined by the shapefile. Based on the user type and Species, the map will be able to display data with appropriate levels of precision ie) less precision for a more sensitive species. This view will also have the option for the user to select the dataset(s) that they would like to display on the map. The next view is the Data View, where there will be a more standard data breakdown such as pie charts and frequency of bird types in the given region/dataset(s). The user will also be allowed to select data by attributes like Species, Behavior, and Observer. Next is the Bias View, which will work similarly to the data view, but with a different focus. Here, the focus is on the certainty of the output from the output dataset. As with the data view, the user will be able to select data by Species, Behavior, etc. Data and bias view fall under the ‘view result charts’ portion of section 2.1. Finally, there is the Import View, which will vary the most by user permissions, and is only available to someone with the permissions of a Registered User or higher. From here, a Registered User will be able to download a dataset from those that are available, and a Privileged User will be able to import data of their

own, and potentially upload corresponding analysis if it fits the output of the machine learning that is being developed. This view is also related to “ingest data” in the Controller section.

Controller: The controller will be the logic behind the application and interfacing between the ACG and machine learning that is being developed. Any user can get the system to list the available datasets, and use that to view the dataset using the application. Additionally, public users will be able to parse the results that are being displayed in various ways as is mentioned in the view section. Login/Authentication will be handled through the controller, and will ensure that correct permissions apply to the user that is being authenticated. Once authenticated, Registered Users can download data that is stored using the ACG for their own needs. On top of all previous abilities, Privileged Users may also Queue the analysis of a dataset of their own on the ACG servers. Finally, at the highest level of permissions, Administrators may modify users, changing their permissions and account type. Administrators are also the only users that will be allowed to add new analysis scripts to be used for analysis of the uploaded data from Privileged Users and above.

3. Persistent Data Design

3.1 Database Descriptions

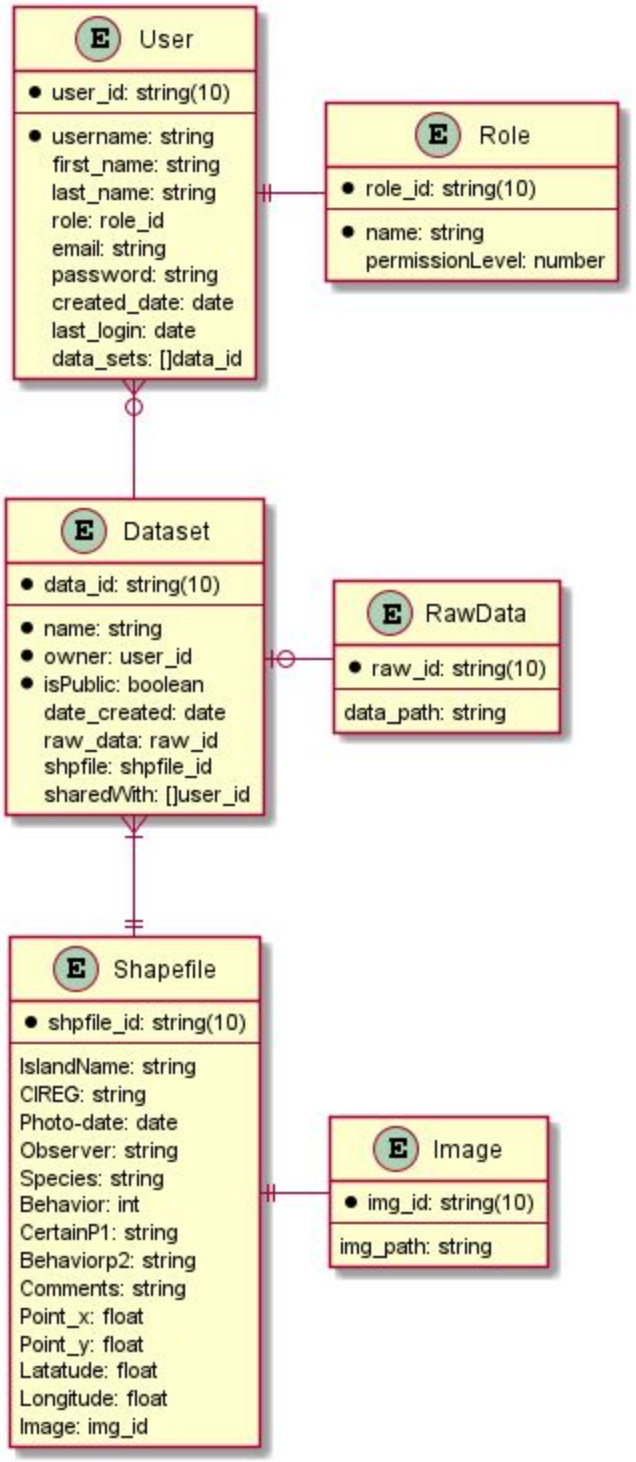

Figure 4. Database Description

User

This is any user within the system and will contain all user information including profile information and permissions within the system.

Role
A role that a user can have (Admin, user, public scientist, etc.) and the associated permissions level.

Dataset
A collection for a raw data item (geotiff) and the resulting Shapefiles that are returned from the external ACG compute server, or Shapefiles that are imported by a user. Dataset manages the visibility of a data set and provides information about who can view that data.

Shapefile
An abstraction of the shapefile format used for the application. Information from a shapefile coincides with each field in Shapefile. Shapefile also added an Image for each record which shows the bird that is present at each data point.

Image
Contains the url to obtain the image from the file server

Raw Dat
Contains the url to obtain the raw data (geotiff) from the file server

3.2 File Descriptions

Shapefile:

Shapefiles (.SHP) contain information that represents geocoded locations on a map. Each location is represented by a line in the SHP file table and all lines must contain the fields outlined below in order.

Field Name	Data Type	Description
OBJECTID	int	Unique ID for this location
IslandName	string	Name of the island the location is on
CIREG	string	Registration number for the island the location lies on
PhotoDate	date	Date of the photo of this location
Observer	string	Name of the observer who identified the bird at this location
Species	string	The species of the bird identified at this location
Behavior	int	Boolean representing whether or not the bird is nesting in this location

CertainP1	string	Certainty of identification. Y is certain, NB is unknown behaviour, and NS is unknown species.
BehaviorP2	string	Legacy field for second pass of behavior classification
Comments	string	Observer comments on the identification
POINT_X	float	X value of this location in relation to the raster image
POINT_Y	float	Y value of this location in relation to the raster image
Lat	float	Latitude value for this location
Long	float	Longitude value for this location

Example Shapefiles are located on the COS 397 Seabird Project Shared Google Drive

GEOTIFF files are image files which allow for storage of geological features as well.

Example GEOTIFF files are located on the COS 397 Seabird Project Shared Google Drive

4. Requirements Matrix

#	Name	Description	Test Case ID	Status
1.0	**Import data set**	**Upload your own data set for analysis**	**TC1.0**	**TC1.0-No Run**
2.0	**View Data**	**View list of data sets available**	**TC2.0**	**TC2.0-No Run**
2.1	**Sort/Filter Data**	**Sort, filter**	**TC2.0**	**TC2.0-No Run**
3.0	**Manage Data**	**Modify data set permissions and other meta-data**	**TC3.1 TC3.2 TC3.3**	**TC3.1-No Run TC3.2-No Run TC3.3-No Run**
3.1	**Add Metadata**	**Add additional relevant metadata to the data (collection method, notes, etc)**		
3.2	**Manage Visibility**	**Modify the visibility of data**	**TC3.1**	**TC3.1-No Run**

	of Data	**sets the owner has uploaded**	**TC3.2**	**TC3.2-No Run**
4.0	**Queue Analysis**	**Choose algorithm and run against previously specified data set**	**TC4.0**	**TC4.0-No Run**
5.0	**View Results**	**View visualizations of the results of analysis of data set**	**TC5.0**	**TC5.0-No Run**
5.1	**View Charts and Aggregates**	**View the aggregate data (cretainty, number of data points, etc)**	**TC5.0**	**TC5.0-No Run**
5.2	**View Map**	**View geodata on a map, with the ability to re-color and rescale the map**	**TC5.0**	**TC5.0-No Run**
5.3	**View Details by Data Point**	**View statistics, location, source photos, AI categorization and certainty by data point**	**TC5.0**	**TC5.0-No Run**
5.4	**Aggregate data points pere-island**	**System displays a data point per island (instead of proximity or per-bird)**	**TC5.0**	**TC5.0-No Run**
6.0	**Import Results**	**Results of analysis are associated with a data set**	**TC6.0**	**TC6.0-No Run**
7.0	**Export Data**	**Data is exported on to the users computer**	**TC7.0**	**TC7.0-No Run**
8.0	**Login**	**Registered User logs in**	**TC8.0**	**TC8.0-No Run**
9.0	**Manage Account**	**Manage Account Settings**	**TC9.1** **TC9.2**	**TC9.1-No Run** **TC9.2-No Run**
10.0	**Manage Users**	**Modify user settings**	**TC10.1** **TC10.2** **TC10.3** **TC10.4**	**TC10.1-No Run** **TC10.2-No Run** **TC10.3-No Run** **TC10.4-No Run**
10.1	**Add User**	**Adds a user to the system**	**TC10.1**	**TC10.1-No Run**
10.2	**Remove User**	**Remove a user from the system**	**TC10.2**	**TC10.2-No Run**
10.3	**List Accounts**	**View the accounts logged within the system**	**TC10.3**	**TC10.3-No Run**
10.4	**Assign Privileges**	**Assign a user privileged access to the application**	**TC10.4**	**TC10.4-No Run**

11.0	**Add Analysis Script**	**Upload an analysis script matching a predefined format to be available to users as an analysis method**		

Appendix A – Agreement Between Customer and Contractor

By signing on the provided line below, the client acknowledges and agrees that the deliverables described by the above terms in sections 1, 2, 3, and 4 are satisfactory, as well as all other clauses described in this document.

Client signature:

_____ date:_____

Client comments:

By signing on the provided lines below, all Penobscot Development Group members acknowledge and agree to meet all system design deliverables described in sections 1, 2, 3, and 4, as well as all other clauses described in this document.

Penobscot Development Team signatures:

_____ date:_____

_____ date:_____

_____ date:_____

_____ date:_____

_____ date:_____

Changes to this document may be required as the development cycle progresses. All new drafts of this document must be reviewed, agreed upon, and signed by both the Penobscot Development Group and the client. Upon the signing of a new draft, the previous draft becomes void and the new draft supersedes any obligations set by the previous draft.

Appendix B – Team Review Sign-off

By signing below, all Penobscot Development Group members acknowledge that they have reviewed all system design deliverables described in sections 1, 2, 3, and 4.

_____ date:_____

_____ date:_____

_____ date:_____

_____ date:_____

_____ date:_____

Appendix C – Document Contributions

The below table describes the Penobscot Development Group members contribution to the document for internal purposes.

20% - Jacob Morin - Section 3.2, Appendices A,B,C
20% - Devin Christianson - Section 1, Section 2.1, Appendices A,B,C
20% - Alexandre Feren - Section 2.2
20% - Kyle Walker - Section 4
20% - Nick Kania - Section 3.1