



# **Nature's Palette: An open-access digital repository of spectral data**

## **Comp 6905-001 – Software Engineering Assignment 01**

Fall 2019

By

Lavaanantha Thayaparendran	[201990251]
Sourav Barua	[201991158]
Shiplu Das	[201891201]
Sourav Sarker	[201892985]
Vivekkumar Dharmendrakumar Patel	[201991314]

Faculty of Computer Science

## Table of Contents

<b>1. Introduction.....</b>	<b>3</b>
<b>1.1. Vision .....</b>	<b>3</b>
<b>2. System Modeling.....</b>	<b>4</b>
<b>2.1. Use Case Description and Diagram.....</b>	<b>4</b>
<b>2.2. Essential Use Cases - Detailed Description.....</b>	<b>6</b>
<b>2.3. Class Diagram (Includes Entity Boundary and Control Objects).....</b>	<b>9</b>
<b>2.4. Sequence Diagram .....</b>	<b>10</b>
<b>2.5. Design Goals .....</b>	<b>12</b>
<b>2.6. System Decomposition .....</b>	<b>15</b>
<b>2.7. Logical Architecture.....</b>	<b>18</b>
<b>2.8. Web Link .....</b>	<b>19</b>
<b>Annexure 01 .....</b>	<b>20</b>

# 1. Introduction

## 1.1. Vision

In the world of free information, it is now very easy for anyone to get any information they need. Just a few clicks open the world of information a user need. However, the question arises are all the information we get over are authentic? The word “free” give access to many fake users to upload and mis-lead people. Dealing with wrong information may not only be the reason of losing someone’s motivation over the work, it might also cause a violation of the society or even worse. For someone doing research on a topic, it is always important to him to assure that the data he is using came from an authentic source.

The scope of the system is to develop a repository that enable the researcher worldwide to upload their analysis data (spectrometer reading). In that case the researchers should be authenticated by admin (login) to upload the data. The data can be large set of files (readings) with raw data and metadata. While uploading system should check raw data along with metadata to ensure the validity.

Once it uploaded, the data should be cleaned by the given script of filtering mechanism and matrix file will be created. After that all the files should be stored in backend database.

Any interested user can access the repositories (as a guest) and search for any required data and will be able to download raw data and metadata file.

Searching will be based on the metadata file column descriptions. For example, a person can search for a specific species based on Darwin code or can search based on geo coordinates. Optionally, it is expected to provide a solution to search for data based on geo area referencing where a user will have the option to point over the map to retrieve the respective data.

At the end of the day the users or the researcher who are interested about the spectral data will have a common digital online repository that they can easily access over the internet. It helps the researcher to easily share their spectral data. Also, it helps other interested people to access those data for their research purpose or any other analysis. As only the authenticated (authenticated by administrator or ORCID) users will be allowed for uploading, it’s fully reliable to use those data; The in build advance search option will help them to easily target for the interested data and can visually model the data from system itself for any statistical analysis.

## 2. System Modeling

### 2.1. Use Case Description and Diagram

#### Use Case Diagram

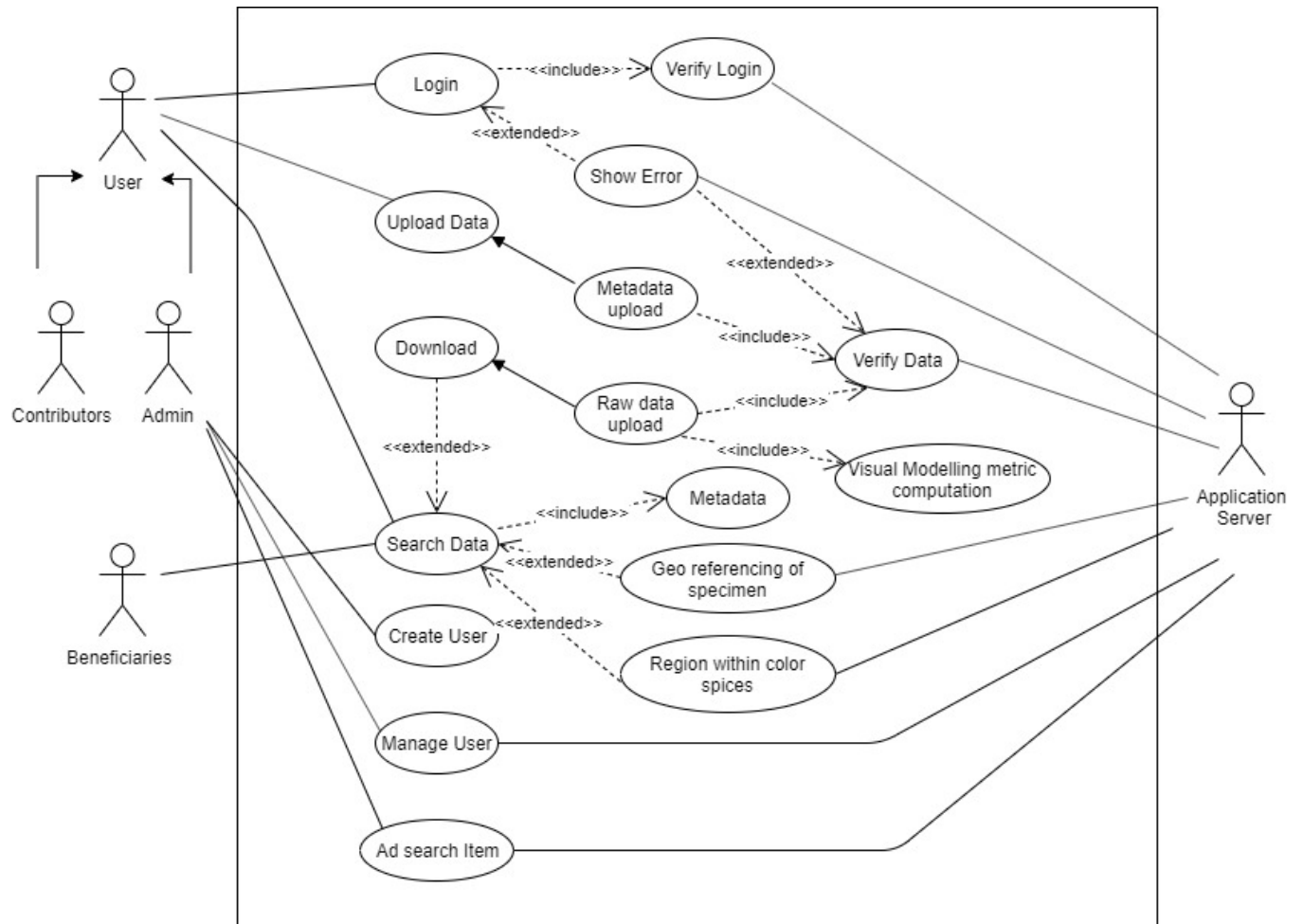


Fig 2.1.1: Use case diagram

## Use Case Description

- **Create new user:** Admin user will be the person who would be able to add a user. When a user registered into the portal, admin will check his authenticity and then approve it.
- **Manage User:** This feature will allow the admin to take over the control of any user. After a user (Contributors) has been registered, admin will have access to edit, delete that user.
- **Add new Search term:** Admin should also have the privilege to add new search term which will assist the user (beneficiaries) to search for their desired data.
- **Login:** For a user (contributors) to upload their data into the portal, they must login into the system. The system will verify the login credentials given. If the system finds the given credentials into the database, the user will be allowed to login, otherwise an error message will throw - *“Invalid Credentials. Try Again.”*
- **Upload:** An authenticated user should be able to upload their raw data and meta-data into the repositories, which can be downloaded by a guest user.
- **Search:** The search feature enables the user (beneficiaries) to search for a specific file. The search will be based on metadata file.
- **Download:** The search result a user will get will also have a download option for them. User will be able to download the required file he may need filtered from the search.

## 2.2. Essential Use Cases - Detailed Description

### File Upload:

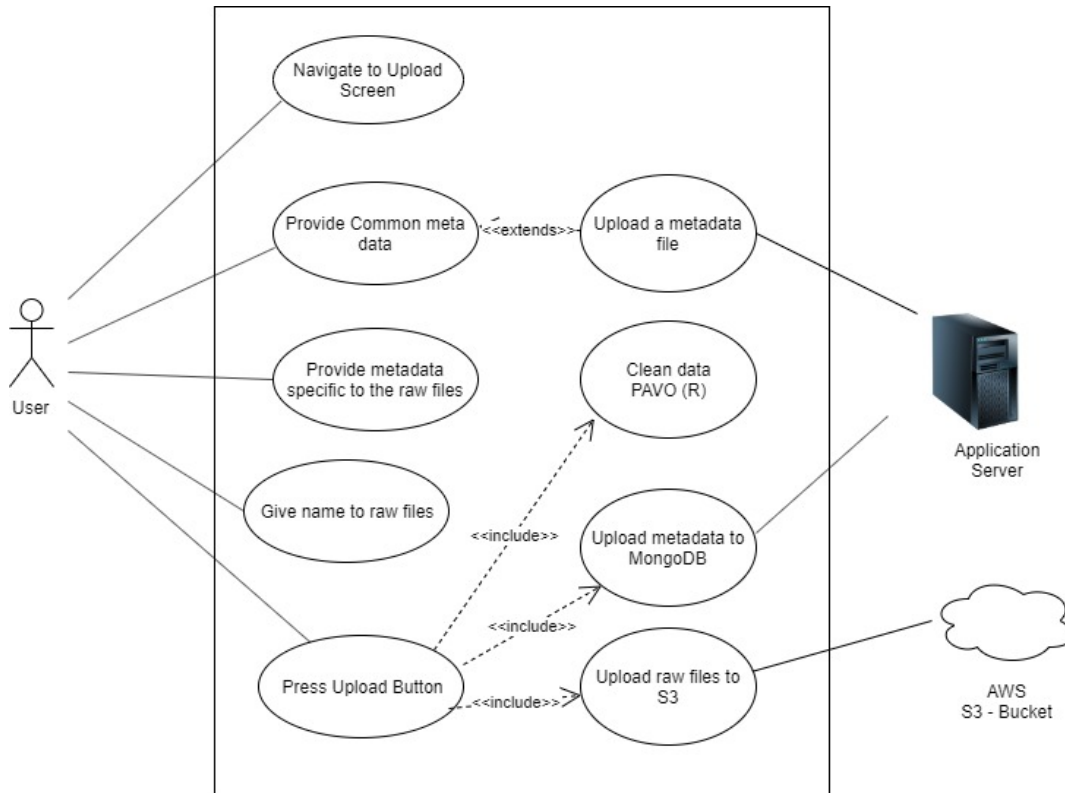


Fig 2.2.1: Use case diagram – File Upload

<b>Use case name</b>		Upload
<b>Actors</b>		User
<b>Preconditions</b>		Need to login to the system (Need a user account/ORCID account)
<b>Normal Flow</b>	<b>Descriptions</b>	User will upload the raw data file and respective meta data file; both files will go through validation and cleaning process and finally added to the database.
	<b>Postcondition</b>	Once files are successfully uploaded to the database, user will get status message.
<b>Alternative flows and exception</b>		Meta data and raw data file may mismatch, and will throw file error message
<b>Non-functional requirements</b>		During the file upload user will get the uploading progress status.

## File Search:

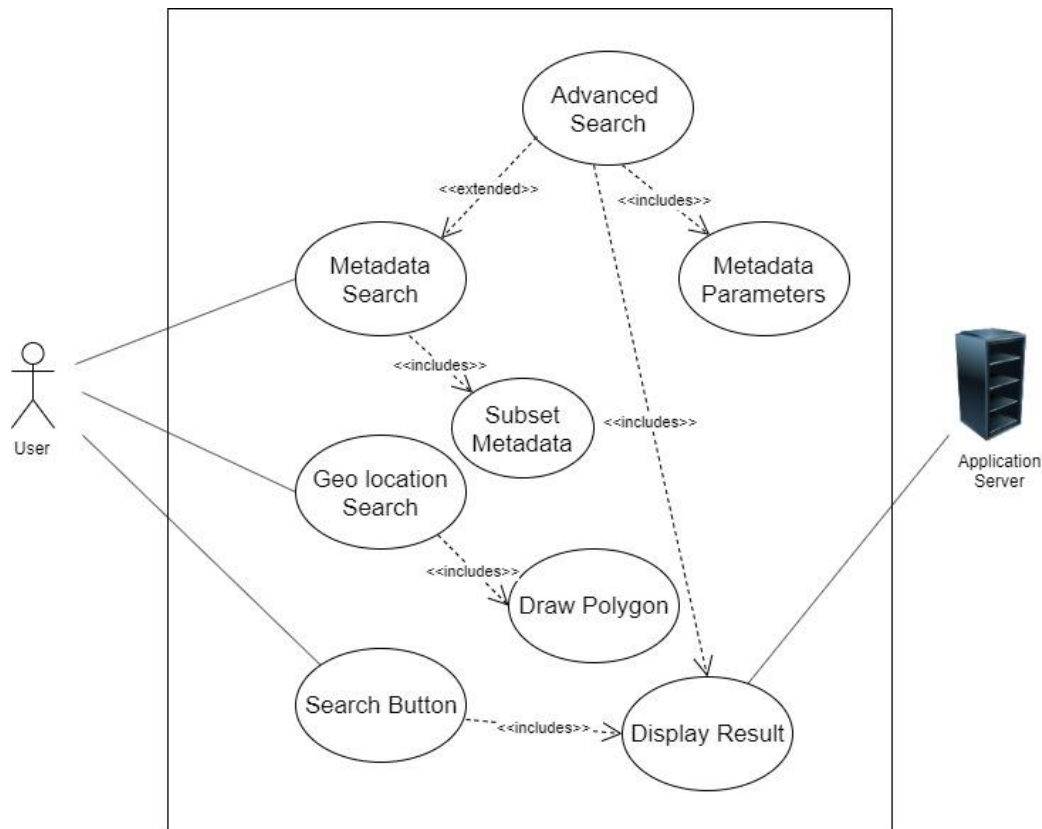


Fig 2.2.2: Use case diagram – File Search

<b>Use case name</b>		Search
<b>Actors</b>		Users
<b>Preconditions</b>		N/A, open to anyone to search for files
<b>Normal Flow</b>	<b>Descriptions</b>	User enter the system as guest, enter the search terms based on Darwin code, or geo reference and get set of possible data. User can visually model the searched data.
	<b>Postcondition</b>	User will see the possible search outputs and files to download or visual models.
<b>Alternative flows and exception</b>		In event of entering unknown search terms, will return no data message.
<b>Non-functional requirements</b>		Tolerable search response time. For geo referencing search option, there should be an instruction.

## File Download:

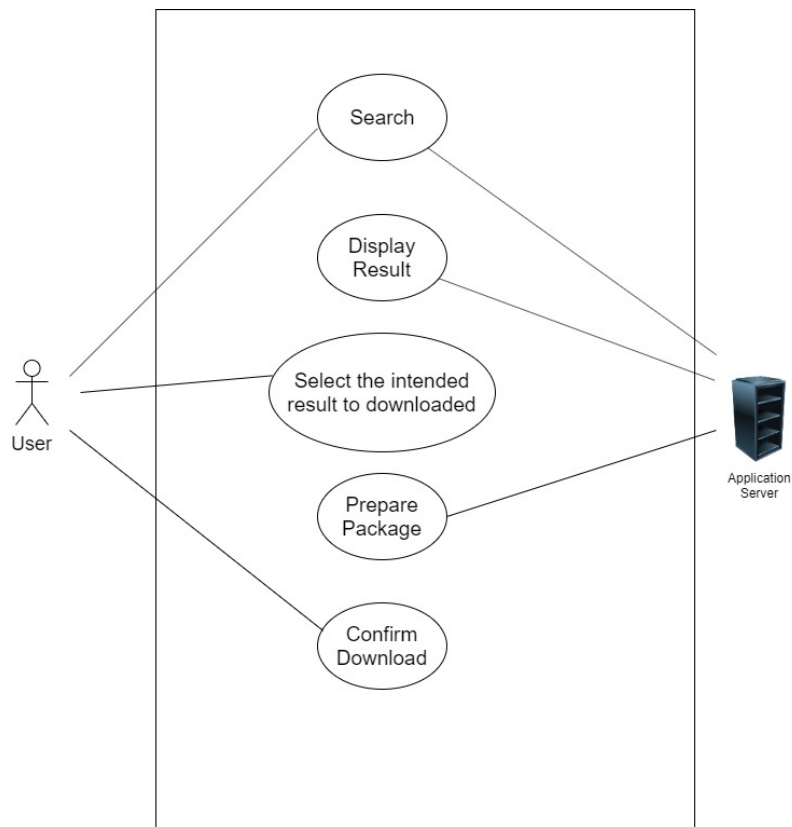
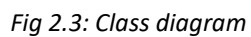


Fig 2.2.3: Use case diagram – File Download

Use case name		Download
Actors		Users
Preconditions		User should select subset of search results to download
Normal Flow	Descriptions	User will search data from any search option, from the results user may download the file, interested raw data file and meta data file will be downloaded to given destination
	Postcondition	User will be notified with download package and size
Alternative flows and exception		N/A
Non-functional requirements		During the download process user should get notified with download progress status



## 9 | Page



## 2.4. Sequence Diagram

### Data Upload:

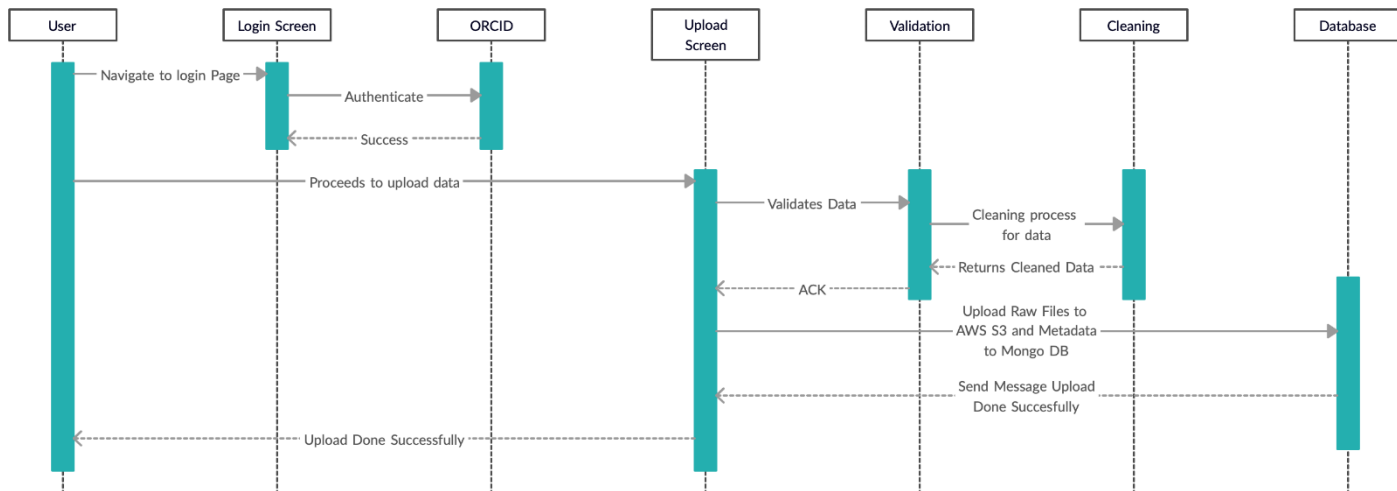


Fig 2.4.1: Sequence diagram – Data upload

### Data Search:

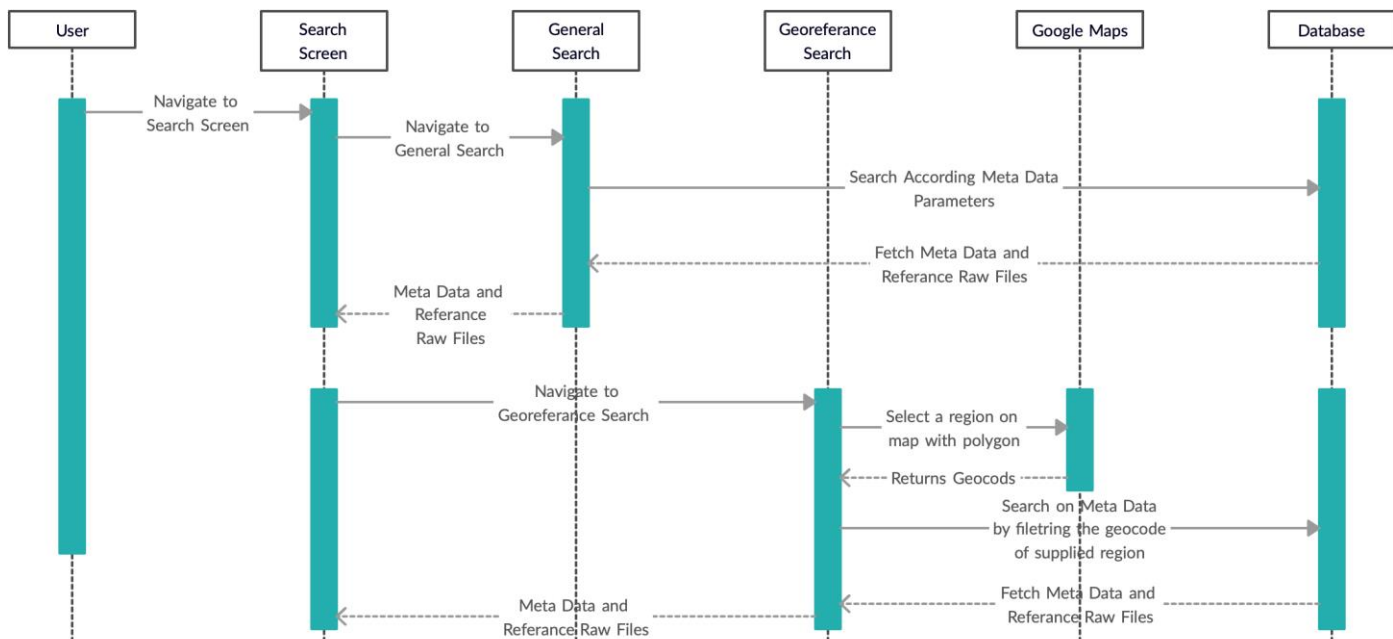
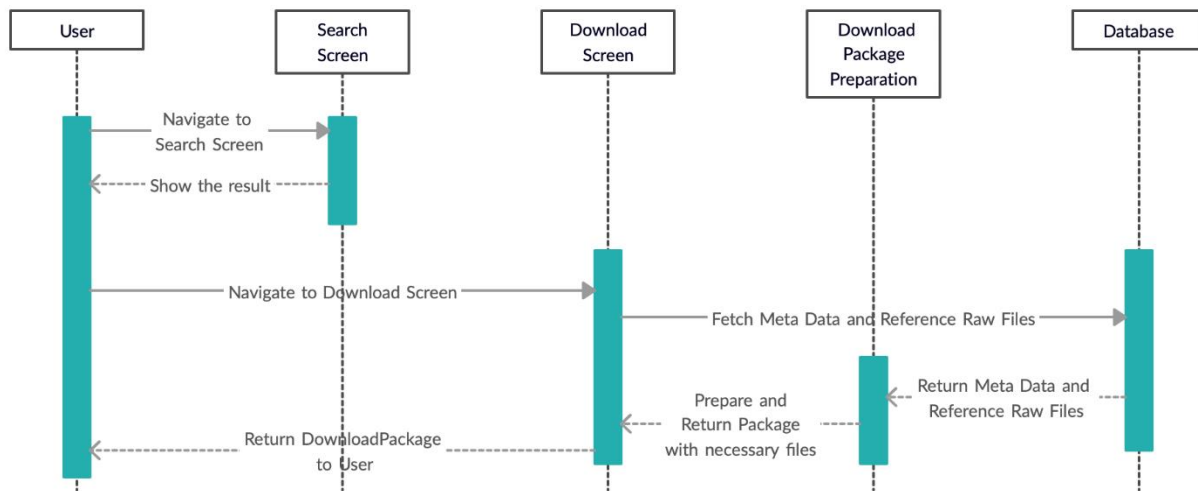


Fig 2.4.2: Sequence diagram – Data search

## Data Download:



*Fig 2.4.3: Sequence diagram – Data Download*

## 2.5. Design Goals

The main goal of this system is to provide a feasible and more convenient solution for researchers to share their spectral data in online repository and to allow any interested users to search and visualize the data and to download it as they want.

**User friendly interface/UI:** The Initial goal is to provide a user-friendly interface as it was not expected the researchers and other users to be IT geeks. So, the user interface will provide a more convenient interface to access/interact with system in terms of data uploading, searching, visualizing and downloading.

In terms of searching, it will be flexible and scalable. Based on the given requirement the search terms will be defined based on the meta data descriptions.

- By default, with the initial system, basic and important search terms (Eg: Davin codes/Species name, country) will be provided. Later, based on needs admin can add more search terms from the meta data file. Aimed to provide that level of flexibility.
- Also, the design will be scalable that would be able to add more search terms in future.

The search results will be displayed with detailed visualization for more convenience to understand the data. Visualization function also will be designed with scalability to adopt future data projection requirements.

**Access control:** There will be three main user types, *Administrator*, *Contributors*, and *beneficiaries*.

Administrators will have the full privileges through out the system. They can create new users (for contributors) and add new search terms based on new requirements.

Contributors will be allowed to upload the spectral data with meta data (If those two files match only).

Beneficiaries will have read-only level of privilege. They can search for any data and visualize it from the system itself, also can download the raw data and meta data. They can't do any level of modification. In system perspective beneficiaries will be considered as guest users.

**Flexibility and scalability:** System will have the flexibility to add more search terms and visual models as per the future requirement.

- By default, with the initial system, basic and important search terms (Eg: Davin codes/Species name, country) will be provided. Later, based on needs admin can add more search terms from the meta data file. Aimed to provide that level of flexibility.
- Also, the design will be scalable that would be able to add more search terms in future.

According to the given requirement, it was expected around one million files will be uploaded in first year and one to three million files per year in upcoming period.

- Therefor the system and the database will be structured to handle massive data sets also to adopt the gradual increment throughout the year.
- Additionally, the database backup (remote backup) option also will be considered during the design and implementation as a backup plan.

**Reliability:** Nature's Palette system will be reliable in terms of persisting data on the permanent storage system (like Flat-File, MongoDB etc.) and be able to distinct different kinds of data. The searching algorithm will be consistent and optimized with the logics for effective response time.

**Fault tolerance:** In case of any error or failure, the system will inform the user about the failure with a proper message/logs without affecting other parts of the working system.

**Target environment:** The program will be written in JavaScript to be able to run on Nodejs environment. The data will be stored in MongoDB.

**Hardware and software mapping:** In the production stage the system will be deployed in two hardware nodes (Or two virtual nodes in cloud like AWS). One node will be used to install DB and the other node will be assigned to deploy the web server. (For the POC/Demo both may installed in single node). Both nodes will be in same network therefor direct communication is possible.

**Security:** As the Nature's Palette system is for academic/research purpose the client didn't much concern about the security. Such security aspects (Eg: DB security features, HTTPS access for web server) won't be considered at this stage. But the essential security features will be implemented (passwords will be stored encrypted).

For the system access, *administrator* will be allowed through dual authentication login (password and mail verification code). Contributors need to login by using the username and credentials given by *administrator*. *Beneficiary* user can access as guest no login required.

As suggested by client, the option will be considered to integrate ORCID portal for login. So that any legitimate users (contributors) who has ORCID account, can easily login with their ORCID credentials to upload their spectral data.

*Note: ORCID is a non-proprietary alphanumeric code to uniquely identify scientific and other academic authors and contributors. (Source: Wikipedia)*

## 2.6. System Decomposition

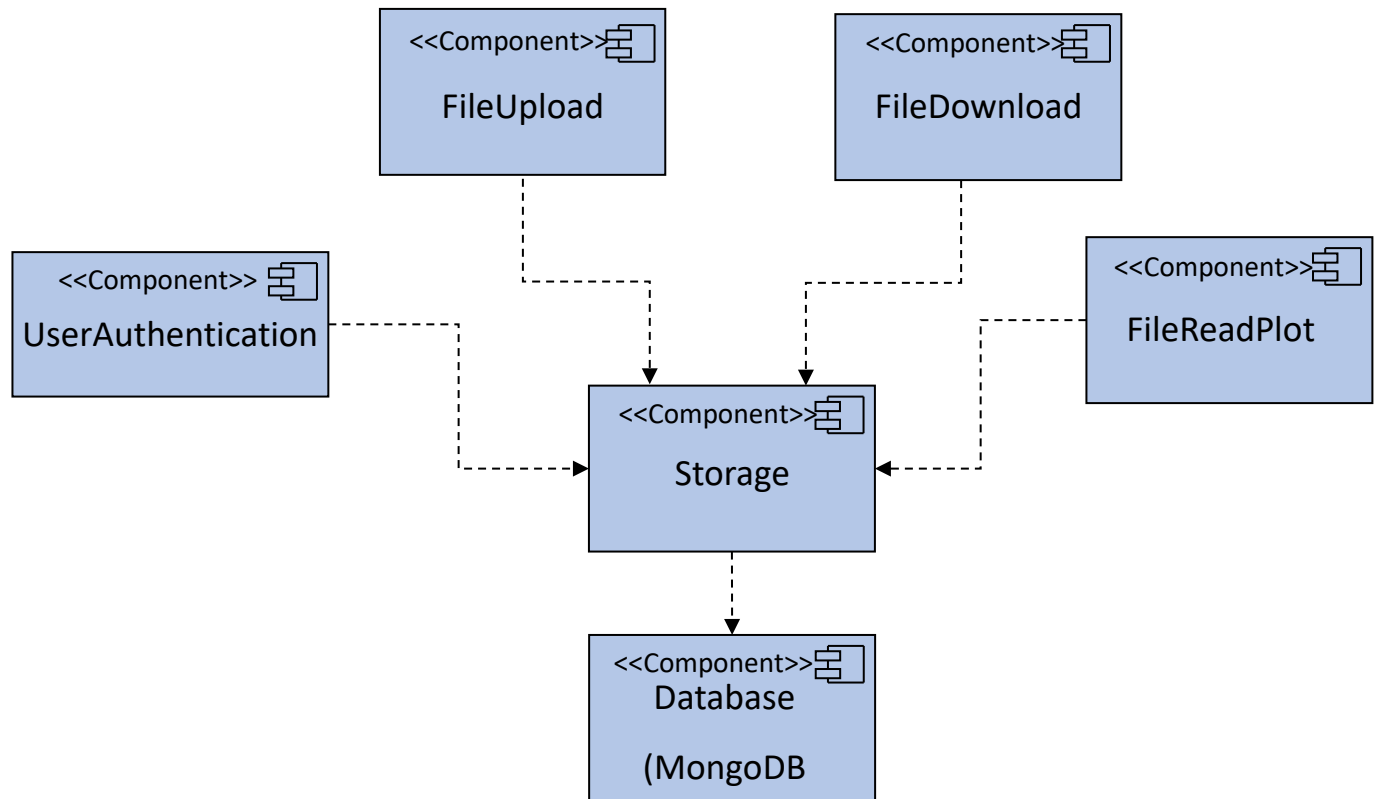
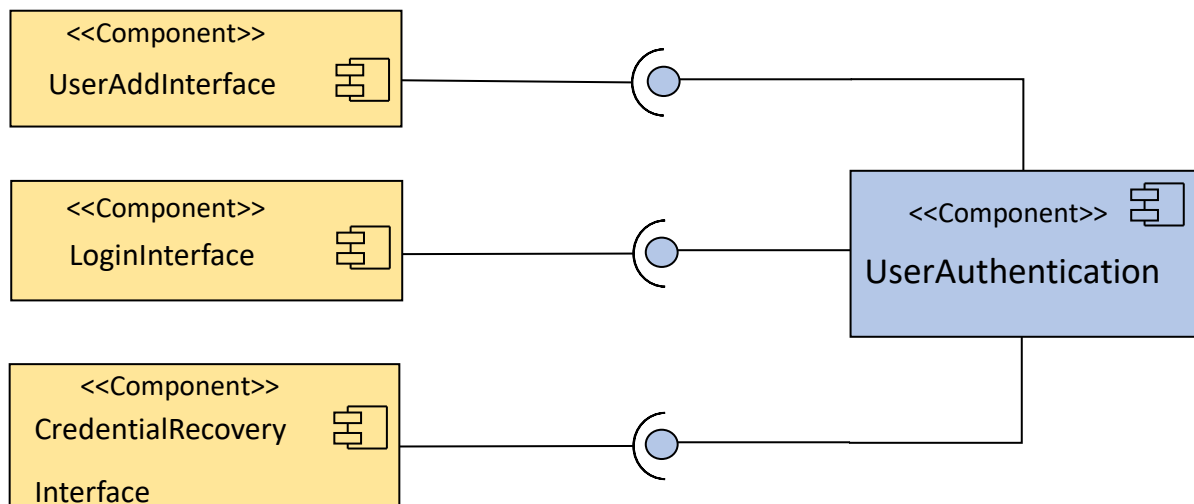


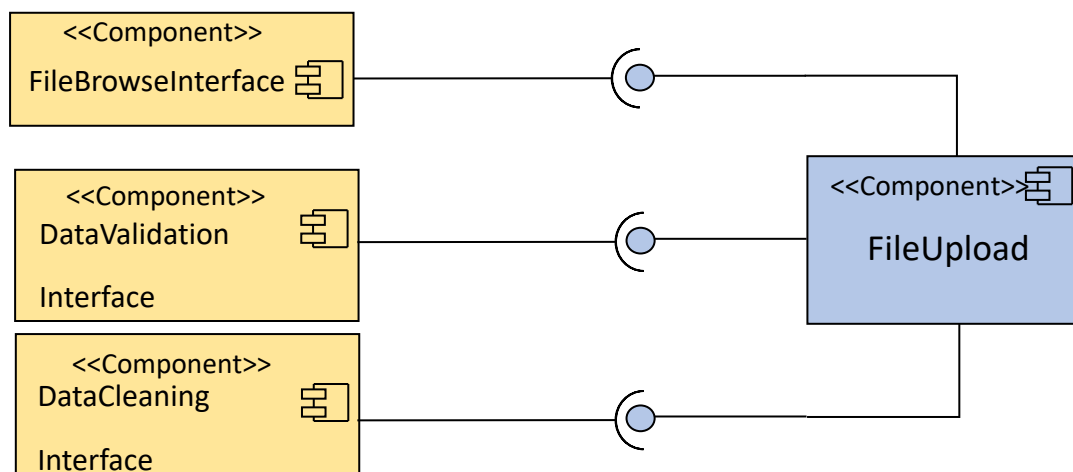
Fig 2.6.1: System Decomposition Diagram

The above decomposed subsystems will be developed in terms of combined packages as shown in class diagram as UI application package and 3<sup>rd</sup> party packages for user authentication and system interaction, Core entity packages for handle the core data types and search terms, Raw file and metadata handling packages to manage the file upload and download parts.



*Fig 2.6.2: System Decomposition Interface Diagram*

UserAddInterface, LoginInterface and CredentialRecoveryInterface requires services from UserAuthentication subsystem for admin to manage (create/delete/modify users) and for contributors to login into the system and to recover the credentials.



*Fig 2.6.3: System Decomposition File Upload*

FileBrowseInterface, DataValidationInterface and DataCleaningInterface requires services from FileUpload subsystem component in order to browse files (Raw data and spectral data) from local machine to upload, then to validate the uploaded files and to clean up the raw data.



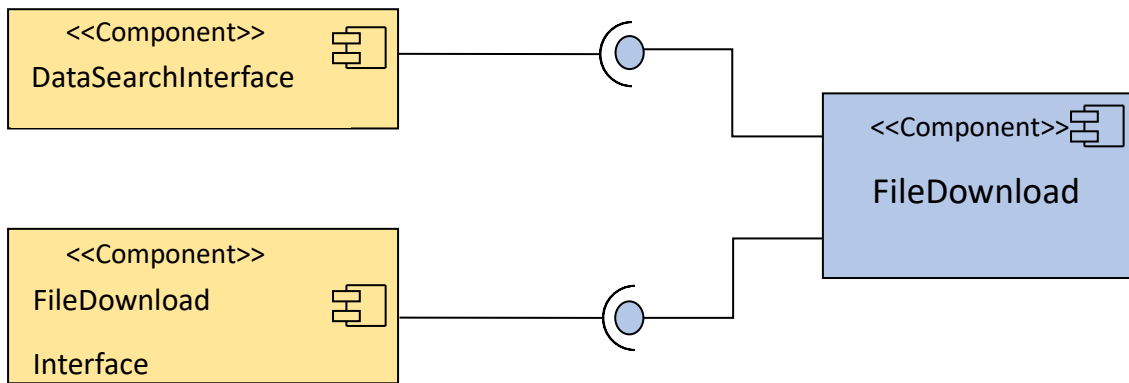


Fig 2.6.4: System Decomposition File Download

DataSearchInterface and FileDownloadInterface requires services from FileDownload subsystems to search for the interested data and to download the required files (Raw data and respective metadata).

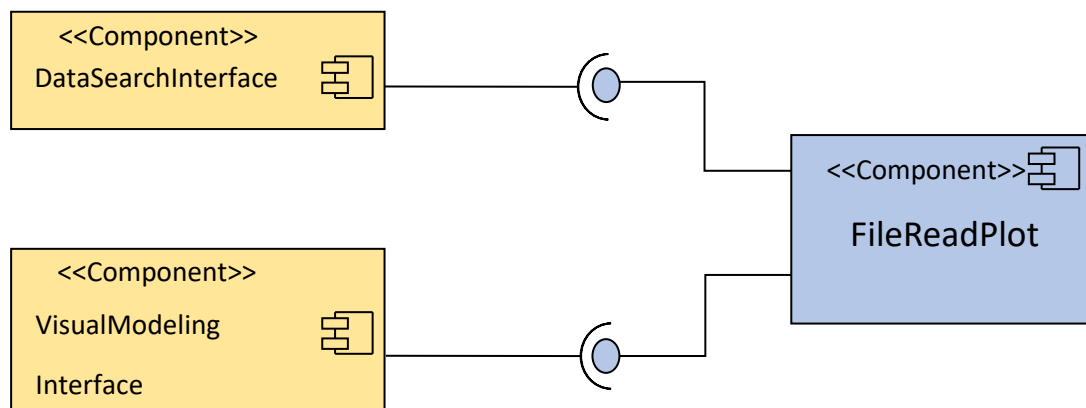


Fig 2.6.5: System Decomposition File Read & Plot

DataSearchInterface and VisualModelingInterface requires services from FileReadPlot subsystem to search for the interested data and to plot a visual model based on the search results.

## 2.7. Logical Architecture

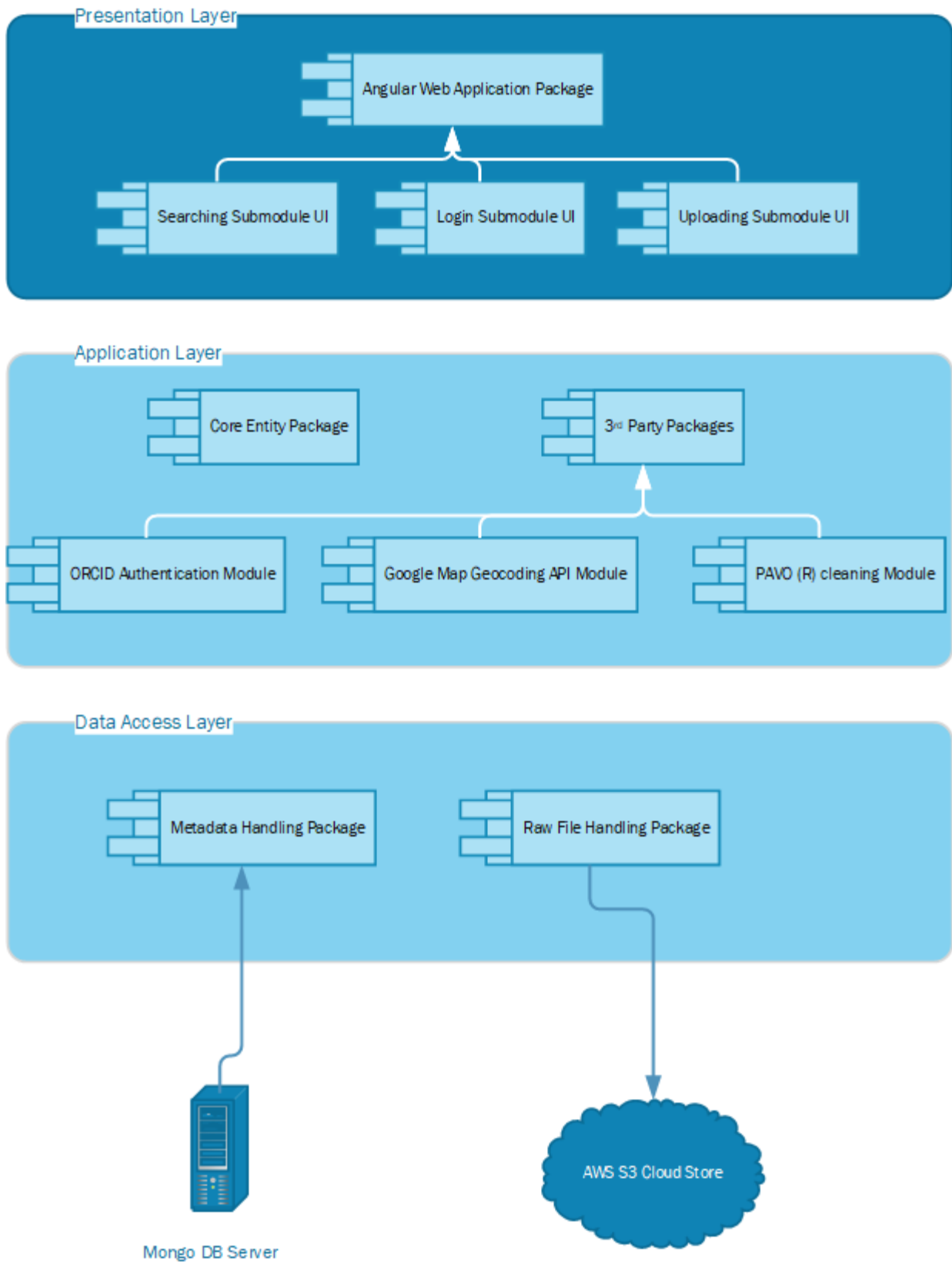


Fig 2.7.1: Logical Architecture

## 2.8. Web Link

### **Git Hub Link:**

<https://github.com/SouravBOrion/Natures-Palette>

### **Website link:**

<http://sc-5.cs.mun.ca/>

## Annexure 01

### Project Contribution:

1. Lavaanantha Thayaparendran [201990251]

Contribution 20%

2. Sourav Barua [201991158]

Contribution 20%

3. Shiplu Das [201891201]

Contribution 20%

4. Sourav Sarker [201892985]

Contribution 20%

5. Vivekkumar Dharmendrakumar Patel [201991314]

Contribution 20%