DEVELOPING AN ECOLOGICAL VISUALIZATION SYSTEM FOR BIODIVERSITY AND WATER QUALITY DATA

CHAM HUI

FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

DEVELOPING AN ECOLOGICAL VISUALIZATION SYSTEM FOR BIODIVERSITY AND WATER QUALITY DATA

CHAM HUI

THESIS SUBMITTED IN FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

INSTITUTE OF BIOLOGICAL SCIENCES
FACULTY OF SCIENCE
UNIVERSITI MALAYA
KUALA LUMPUR

2022

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **CHAM HUI**
Matric No: **SVA180048**
Name of Degree: Doctor of philosophy
Title of Thesis: **DEVELOPING AN ECOLOGICAL VISUALIZATION SYSTEM FOR BIODIVERSITY AND WATER QUALITY DATA**

Field of Study: **BIOINFORMATICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature Date: **06/02/2022**

Subscribed and solemnly declared before,

Witness's Signature Date: **06/02/2022**

Name:
Designation:

DEVELOPING AN ECOLOGICAL VISUALIZATION SYSTEM FOR BIODIVERSITY AND WATER QUALITY DATA

ABSTRACT

Information visualization is an essential tool for effective communication and interpretation. When opposed to numerical quantities, comprehending information through graphics is easier. Google Earth can be used to create a useful monitoring site visualization system. However, research in web-based visualization systems is still lacking, and it deserves to be addressed. The research aims to create a web-based system that depicts the Water Quality Index (WQI) and the biodiversity index on Google Earth. For easy interpretation of water quality conditions, as well as species richness and biodiversity indexes, the system utilized graphical representation through various forms and colors. pH, dissolved oxygen, chemical oxygen demand, biochemical oxygen demand, total suspended solids, and ammonia nitrogen are all used in the stand and manage WQI for water quality visualization. For the biodiversity element, the program includes common biodiversity indexes such as Species Richness, Shannon Index, Simpson Index, and Hill Numbers. Citizen science has contributed to the collection of biodiversity data that is useful for biodiversity site conservation and preservation. By combining graphical components, the applicability of site biodiversity indexes can be improved. Previous research that attempted to integrate the calculation of biodiversity indices into biodiversity monitoring systems lacked the visual element. This unique study aims to create an online module that integrates citizen science biodiversity data with a visualization component. A graphical depiction of the system output is shown. It uses a variety of forms and colors to make water quality and diversity status easier to understand. It promotes environmental protection and improved interpretation of data, aiding biologists and environmentalists to better und the environment. The web-based system prototype can be accessed at the following website: http://www.umlivinglabsystem.com/.

Keywords: Water Quality Index, Shannon Index, Simpson Index, Hill Numbers, Citizen Science

MEMBANGUNKAN SISTEM PENGLIHATAN EKOLOGI UNTUK BIODIVERSITI DAN DATA KUALITI AIR

ABSTRAK

Visualisasi maklumat adalah alat penting untuk komunikasi dan tafsiran yang berkesan. Apabila bertentangan dengan kuantiti berangka, memahami maklumat melalui grafik adalah lebih mudah. Google Earth boleh digunakan untuk mencipta sistem visualisasi tapak pemantauan yang berguna. Walau bagaimanapun, penyelidikan dalam sistem visualisasi berasaskan web masih kurang, dan ia patut ditangani. Penyelidikan bertujuan untuk mencipta sistem berasaskan web yang menggambarkan Indeks Kualiti Air (WQI) dan indeks biodiversiti di Google Earth. Untuk tafsiran mudah tentang keadaan kualiti air, serta kekayaan spesies dan indeks biodiversiti, sistem ini menggunakan perwakilan grafik melalui pelbagai bentuk dan warna. pH, oksigen terlarut, permintaan oksigen kimia, permintaan oksigen biokimia, jumlah pepejal terampai dan nitrogen ammonia semuanya digunakan dalam dirian dan mengurus WQI untuk visualisasi kualiti air. Untuk elemen biodiversiti, program ini termasuk indeks biodiversiti biasa seperti Kekayaan Spesies, Indeks Shannon, Indeks Simpson dan Nombor Bukit. Sains warganegara telah menyumbang kepada pengumpulan data biodiversiti yang berguna untuk pemuliharaan dan pemuliharaan tapak biodiversiti. Dengan menggabungkan komponen grafik, kebolehgunaan indeks biodiversiti tapak boleh dipertingkatkan. Penyelidikan terdahulu yang cuba mengintegrasikan pengiraan indeks biodiversiti ke dalam sistem pemantauan biodiversiti tidak mempunyai unsur visual. Kajian unik ini bertujuan untuk mencipta modul dalam talian yang mengintegrasikan data biodiversiti sains warganegara dengan komponen visualisasi. Gambaran grafik output sistem ditunjukkan. Ia menggunakan pelbagai bentuk dan warna untuk menjadikan kualiti air dan status kepelbagaian lebih mudah difahami. Ia menggalakkan perlindungan alam sekitar dan penafsiran data yang lebih baik, membantu ahli biologi dan alam sekitar untuk memperbaiki alam sekitar. Prototaip sistem berasaskan web boleh diakses di laman web berikut: http://www.umlivinglabsystem.com/.

Kata Kunci: Indeks Kualiti Air, Indeks Shannon, Indeks Simpson dan Nombor Hills, Sains Warganegara

ACKNOWLEDGEMENTS

`I would like to take advantage of this wonderful opportunity to express my gratitude to everyone who has helped me with my studies and research. Many thanks to my supervisor, Dr. Sorayya Malek, for her consistent support, motivation, guidance, and full support on documentation and administration work. I am grateful for her advice, patience, and support in developing the system as well as in writing my thesis. Dr. Pozi Milow deserves special recognition for her consistent assistance with my journal writing and knowledge sharing throughout my research.

`Last but not least, I would like to share the success of this project with my friends and family, particularly my wife, Kong Po Lian, and my parents. Thanks to Song Cheen, in assisting me in the project as well. My friends, Lau Chia Fong, Tang Chee Kuang, and Oh Jin Heng, also helped me on this journey. I couldn't have finished this research without their understanding, assistance, and support.

TABLE OF CONTENT

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

AN	:	Ammoniacal nitrogen
API	:	Application Programming Interface
BOD	:	Biological Oxygen Demand
COD	:	Chemical Oxygen Demand
CSS	:	Cascading Style Sheets
DFD	:	Data Flow Diagram
DO	:	Dissolved Oxygen
DOE	:	Department of Environment
DSDM	:	Dynamic systems development method
ERD	:	Entity Relationship Diagram
FAQ	:	Frequently Asked Questions
GIS	:	Geographic Information System
HTML	:	HyperText Markup Language
JSON	:	JavaScript Object Notation
KML	:	Keyhole Markup Language
NWQI	:	National Water Quality Index
SUS	:	System Usability Scale
TSS	:	Total Suspension Solid
WQI	:	Water Quality Index
XML	:	Extensible Markup Language

CHAPTER 1: INTRODUCTION

1.1 Overview

This study aims to develop an Ecological Visualization System for Data on Biodiversity and Water Quality. It is a web application that can calculate biodiversity indices while also visualizing species topology. The water quality index (WQI) value is also included in the ecological visualization system. According to (McGranaghan, 1993), Considering different data, users will have different requirements for data quality visualization, graphic visual representation is essential when establishing communication objectives. Map users rely on the graphic quality of the map to determine the accuracy and quality of the data. Colors are fundamentally exquisite and often indicate quantitative changes in three distinct dimensions: hue, value (brightness), and saturation. Thus, colour is a critical part of graphic visualization for conveying information, as it is perceptible to the human visual system.

Data visualisation is a crucial component of good information sharing and interpretation. When comparing to numerical quantities, comprehending information via visuals is easier. Google Earth provides a framework for developing a realistic system for visualizing aquatic data monitoring sites. The background satellite and road datasets on Google Earth are up-to-date as compared to those of other organizations.

Over time, conservation biology has become an emerging issue. The term 'conservation biology' was introduced by (Dasmann & R. F., 1985) and (EHRENFELD & W., 1970) together with (Soulé & E., 1985) which address the loss of biological diversity. Species richness is one technique to quantify the biodiversity in a particular place; in locati ons with a high species richness and a high concentration of rare species, scientists might as sert that the area is a biodiversity hotspot and advocate for its protection.

Biodiversity indexes such as the species richness, Shannon-Wiener Index, Simpson's Index, and Hill Numbers are used as common measurements for biodiversity. The indexes mentioned above are commonly used in ecology. These indexes are mathematical functions that take account of species richness and evenness and calculate in a single measure (Colwell & K., 2009).

In recent years, there is an increase in volunteer participation within the conservation and ecology community and can be known as the term "Citizen Science" (Wiggins & Crowston, 2011); (Cohn, 2008); (Oberhauser & Prysby, 2008); (Droege, 2007); (Silvertown, 2009); (Catlin-Groves, 2012). It is a form of research collaboration or data gathering performed by unskilled or non-expert individuals. According to (Toogood, 2013), involving the public to participate in biodiversity monitoring and research are regarded as having significant or potential for biodiversity conservation, and often held to achieve a win-win benefit, whereby allowing the public to engage and help in monitoring the environment, collecting research data, eventually raise the concern and awareness of conserving the environment. Besides, it can boost scientific education, and most prominently, the researchers can utilize various publics as data gatherers and obtain vast biological data within a short period.

The available biodiversity systems only concentrate on one element, which is either a biodiversity index calculator or a visualization element. The existing system is not linked to data from citizen science, which is more easily accessible.

With rising population rates, water conservation and management have become critical for environmental preservation, as only a small percentage of the water on Earth is drinkable. Effective water management enables the control of water resources, the minimization of damage to people and property, and the maximization of beneficial use.

The water quality index (WQI) is a measure of the water's quality. WQI is a standard scale single number that relates a set of water quality parameters (Leščešen I, 2015). WQI is one of the most effective instruments for assessing the quality of groundwater, lakes, and rivers (Anyachebelu, 2015). According to (Yogendra, 2008), policymakers and environmentalists consider WQI as a valuable instrument for monitoring water quality.

Researchers have pointed out the existence of several versions of water quality indexes in literature (Avvannavar, 2008) (Boyacioglu, 2007); (Gandaseca, 2011); (Said, 2004); (Sarkar, 2006). This is primarily due to the indexes being developed using a variety of different values or a combination of water quality criteria. A water quality index based on fourteen parameters (pH, electrical conductivity, dissolved oxygen, total dissolved solids, turbidity, salinity, major cations, major anions, and alkalinity) was developed by (Meher, 2015) for evaluating Gange's river water quality. A water quality index for monitoring the water quality of Tigris and Euphrates rivers based only on seven parameters (total dissolved solids, total hardness, pH, dissolved oxygen, biological oxygen demand, nitrate, and phosphate) was developed by (Al-Shujairi, 2013).

However, only a limited amount of research on online ecological visualization systems has been published. Expert evaluation of data and measurements contained in technical reports is required for previously developed systems for biodiversity and water quality. Additionally, there is a shortage of graphical representations of water quality status, which makes it difficult for water quality managers, policymakers, environmentalists, and the general public to evaluate and use data on river water quality monitoring. Additionally, proper biodiversity indexes are required to make sense of the massive volume of data collected through Citizen Science. As a result, the goal of this study is to propose a visualization system deployed on Google Earth and using data gathered from Citizen

Science to provide a clearer picture for biologists and scientists in understanding the biodiversity structures and patterns, as well as water quality management, which will ultimately provide better conservation management.

1.2 Problem Statement

Over time, conservation biology has become an emerging issue. Species richness is considered one of a way to measure the biodiversity in a specific area, areas with high species richness and containing a lot of rare species, scientist can claim that the respective area is a biodiversity hotspot and prote that respective area. Normally, biologist makes assumptions and decision based on the species richness and the value of diversity indexes, without knowing the distribution and the patterns of the diversity. Hence, visualizing the patterns of the landscapes and ecosystem can also provide better insight for conserving biodiversity (Murray & Williamson, 2002).

Meanwhile, the water quality index has been known as the yardstick of measuring the health of water resources (Berlemann, 2013). According to (Adelagun, Etim, & Godwin, 2021), the WQI is meant to express complex water quality information in a simple and understandable by the public. The numbers do provide meaning regarding the water quality, but without a conception towards the health of the water resources, thus a visualization system of water quality index is created to portray the information, aiding the public to gain a better insight into the water condition.

This study is conducted is because based on several findings, few empirical studies present this information, which is utilizing ecological information and visualizing it through Google Map. Most of the online system does not include visualization elements and their main focus is on creating an online calculator, measuring the biodiversity indexes and water quality index.

- Conservation biology has become an emerging issue, this research is conducted most research has been focusing on calculating the water quality index and biodiversity index but does not include graphical representation and vice-versa.
- Unavailability of online web system that measures water quality index and biodiversity index that consists visualization element, mainly focus on the calculation.
- Biologists and environmentalists usually make assumptions and decisions based on calculation without visualization interpretation.
- The lack of graphical illustration of water quality status makes it challenging for water quality managers, regulators, environmentalists, and the general public to evaluate and use data for river water quality monitoring.

1.3 Research Questions

RQ1 : Which are the common biodiversity indexes used to measure species diversity?

RQ2 : How data collected from citizen science can be included in biodiversity modeling?

RQ3 : What are the parameters used to measure the water quality index in Malaysia?

RQ4 : How to develop a visualization system that allows public participation and understanding towards conservation biology?

1.4 Research Objectives

The goal of this study is to create a web-based system that can visualize both the water quality and biodiversity indexes, which can assist environmentalists and biologists to gain insight from the geographic patterns and make a better decision in conserving the environment through visualization.

1. To develop a water quality visualization monitoring system on Google Earth which is based on the National Water Quality Index of Malaysia (NWQI).
2. To develop a web application that allows users to calculate diversity index and visualize the species available using data from citizen science.
3. To produce and publish water quality reports and biodiversity data related to the user interest.

1.5 Scope of Study

The research primarily focuses on establishing an online web module that allows users to quantify and visualize plant species diversity, richness, and indices, with the University of Malay serving as a test location. The system also includes a water quality system that intends to store, organize, and disseminate hydrological and ecological data via an interactive user interface, report generator, water quality management, learning tool, and data sharing across water research and management institutions. The Langat River was chosen as the trial location for the water quality system due to a lack of data on University Malaya water bodies.

1.6 Outline of Study

Chapter One: Introduction. In this chapter, we explained the general research framework, which introduces the ecological visualization system which mainly focuses on the biodiversity indexes and water quality index parameters, besides presenting the research questions, objectives, and scope of this study.

Chapter Two: Literature Review. This chapter embraces the detailed information obtained through the fact-finding process. It reviews the past studies and researches done for the biodiversity indexes, water quality index parameters, and environment systems available online. It includes a whole picture from a general view to a specific idea about the study field in this research.

Chapter Three: Materials and Methods. This chapter will discuss and analyze the requirements of the system such as fact-finding, hardware and software requirement, database management system, Entity Relationship Diagram (ERD), Data Flow Diagram (DFD), and system design architecture. Besides that, it reviews the methodology to develop the web module and illustrates the task and function of the web module together with the application of programming language. Besides it also includes the development of the web system, and described the user interface and database implementation design that will be applied in the system. The system testing will be discussed and carried out in this chapter.

Chapter Four: Results and Discussions. This chapter is mainly considered and discussed the outcome of the ecological visualization system. This chapter also includes the system usability evaluation results.

Chapter Five: Conclusion. The whole research will be summarized in this chapter together with the research strengths, limitations, and future enhancement. Finally, the overall conclusion is also given in this section.

CHAPTER 2: LITERATURE REVIEW

2.1 Biodiversity

The term biodiversity is an abbreviation of biological diversity. According to the Convention on Biological Diversity, they had defined biodiversity as "the variability among living organisms from all sources including terrestrial, marine and other aquatic ecosystems and the ecological complexes of which they are part; this includes diversity within species, between species and of ecosystems" (Murray, 2002).

Based on Reece et al. (Campbell, 2010), biodiversity can be studied at three main levels, which are genetic diversity, species diversity, and ecosystem diversity.

I. Genetic Diversity

It is the genetic variation within a population that also includes individual genetic variation. Microevolution is possible when one population is extinct and lost some genetic diversity. When the genetic diversity is decreasing, eventually it will lead to the reduction of the adaptation of the species (Reece, et al., 2014).

Genetic diversity is assessed at the molecular level and analyzed through laboratory techniques to understand the polymorphism and variation between species.

II. Species Diversity

The second level of diversity is the number of different species that are represented in a community it is assessed as species richness. Where species that give the higher number in a site, it is considered the following species has a greater impact due to its abundances.

Generally, the most commonly used indexes that are used to measure the diversity of an area are species richness, Shannon-Wiener Index, and Simpson's Index (Hill1973). For the biodiversity component, this study will focus on species diversity.

III. Ecosystem Diversity

The third level of biodiversity consists of the mixture of biosphere ecosystems, which include the interaction between the population of different species. If extinction occurs of one species, it can cause a huge impact on the environment (Reece, et al., 2014). Unfortunately, many ecosystems have been altered and affected by human activities at a rapid pace.

2.2 Conservation Biology

Conservation biology was introduced by (Dasmann R. F., 1968) and (EHRENFELD D. W., 1970) together (Wilcox, 1980). Conservation biology is a multidisciplinary science that has been developed to address the loss of biological diversity, it differs from other biological science whereby is often known as a "mission-oriented crisis discipline" (Soulé M. E., 1985).

Conservation biology exists due to the need to protect, manage and restore the environment inhabited by flora and fauna communities. Conservation biologists had put efforts into addressing and discovering potential threats to the natural environment that can be caused biodiversity to be threatened, factors such as global climate change, deforestation, and various pollution that can lead to extinction and habitat fragmentation (Rinkesh, 2016).

The emergence of conservation biology consists of mainly two objectives, which are (Short, Bradshaw, Giles, Prince, & Wilson, 1992) (Soulé M. E., 1985):

1. To evaluate human impacts on biological diversity.

2. To develop practical approaches to prevent the extinction of species.

2.3 Importance of Conservation Biology

We all should be aware that biodiversity provides the foundation for life on earth and brings us many practical benefits, thus it is every mankind's responsibility to protect and conserve the wealth of the world not only now, but also for the future generation.

(Kellert, 1995) stated the reason why we all should raise our awareness on conserving biodiversity, a term known as biophilia, which is our sense of connection to nature and all life. It is an innate and genetically determined affinity of human beings with the natural world. Hence, when biodiversity deteriorates, eventually it will affect our life.

By safeguarding biodiversity, we can ensure that our quality of life will be elevated, surrounded by an environment that is richer and more diverse. Besides, we will have plenty of resources that can supply foods, fibers, medicine, and various natural resources, eventually enhancing our economy in agriculture, health, and tourism.

Furthermore, if we did not protect wild populations of plants that are closely related to agricultural species, we might lose precious genetic resources that could be used to improve crop qualities, such as disease resistance, better use of herbicides, through cross-breeding, genetic modification. People will not notice species extinct because the rate of extinction occurs rapidly, eventually, people will lose valuable species which may have unique genes and medicinal benefits (Ford-Lloyd, et al., 2011).

All in all, conserving biodiversity is very important to the well-being of our planet. The most important is to conserve ecological stability. Besides, biodiversity also has economic benefits to humans as for the manufacture of food, pharmaceutical, and cosmetic products, indirectly it can serve as a tourist hotspot. On the other hand, we all know that extinction is a natural phenomenon, however, various human activities have threatened biodiversity and had caused extinction to several species. Humans are the ones who should be responsible for today's biodiversity crisis (Reece, et al., 2014).

2.4 Biodiversity Indexes

Biodiversity is an indicator to measure ecosystem health and is generally used in environmental assessment studies. Biodiversity is increasingly recognized as one of the cornerstones of healthy ecosystems (Laurila-Pant, Lehikoinen, Uusitalo, & Venesjärvi, 2015).

According to (Pyron & Burbrink, 2010), biodiversity communities are associated with several attributes, which are species richness, species diversity, and species evenness.

Species Richness is expressed as the number of individuals of species represented in an ecological community. Besides measuring the species richness, species diversity also measures the relative abundance, it is more complex, and generally is described using a diversity index, such as the Shannon-Wiener index. As for species evenness, it is a measure of the homogeneity of abundances in a sample or a community, the value of species evenness is high is when where the community has the same abundance (Colwell R. K., 2009).

The reason why we should measure biodiversity is that it can give information regarding ecosystem health. A diverse ecosystem is considered to have higher ecological stability, productivity and ensure greater sustainability. Furthermore, measuring diversity able to give information about the common and rare species in a community, it is an important measurement that allows us to understand the community structure (Beals, Gross, & Harrell, 2000).

In this study, biodiversity indexes such as the species richness, Shannon-Wiener Index Simpson's Index, and Hill number are used as common measurements for biodiversity. The indexes mentioned above are commonly used in ecology. These indexes are mathematical functions that take account of species richness and evenness and calculate in a single measure (Colwell R. K., 2009). Besides, there are no assumptions made about species distribution, since the indexes are just calculated based on the sample given.

2.4.1 Shannon-Wiener index (H')

A study has been conducted by (Spellerberg & Fedor, 2003), regarding the use of species index, especially the 'Shannon-Wiener Index'. In this literature, the researchers are providing clarification on the terminology used for the Shannon-Wiener index. In some cases, it is called the Shannon index, Shannon-Weaver Index due to 'mislabeling' of the index. Hence, in this study, they had suggested the preferred name for this index is the 'Shannon-Wiener Index' is because Shannon had built the work together with Wiener (Spellerberg & Fedor, 2003).

$$H = -\sum_{i=1}^{s} p_i \ln p_i \tag{2.1}$$

The Shannon-Wiener index accounts for both the abundance and evenness of the species present. It is an information statistic index; it represents all the species in the community and is commonly used to characterize species diversity in a community.

The value H increases, when the number of species in the sample increases. Normally this index is used to compare the diversity of a series of samples. Hence, a higher value of H indicates high species diversity in the sample (Magurran, 2005).

The Shannon-Wiener index is useful to determine whether an area consists of valuable and rare species. Besides, it is sensitive to changes and relatively easy to measure.

2.4.2 Simpson's Index (D)

(Simpson E. H., 1949) introduced the Simpson's diversity index which is used to describe the probability that a second individual drawn from a population would be of the same species as the first. Simpson's index is known as the measure of "dominance concentration", since it is sensitive to the abundance only of the more plentiful species in a sample (Whittaker, 1965), unlike Shannon-Wiener index, Simpson's diversity index is sensitive and gives more weight to the most common and dominant species and rare species will not affect the calculation.

$$D = 1 - \sum_{i=1}^{s} p_i^2 \tag{2.2}$$

The value of D has the maximum value of 1, the higher the value, the higher the diversity.

2.4.3 Hill Numbers

Hill numbers are a combination of taxonomic, functional, and phylogenetic diversity indexes (Hill, 1973). Hill numbers can be used to quantify and compute the effective number of species using the following formula (Clarke, Gorley, Somerfield, & & Warwick, 2014):

$$
\begin{aligned}
{}_0D &= S \\
{}_1D &= \exp\,(H') \\
{}_2D &= \frac{1}{\sum p^2} \qquad (2.3) \\
{}_nD &= \frac{1}{max\{p_i\}}
\end{aligned}
$$

${}_0D = S$ where often implies the total number of species, ${}_1D$ is the exponential of Shannon index. As for ${}_2D$ is the reciprocal of Simpson's index and if ${}_nD$ is another possible evenness index (Clarke, Gorley, Somerfield, & & Warwick, 2014).

Table 2.1 below shows the variables involved in the calculation of biodiversity indexes:

Table 2.1: Variables Involved in the Calculation of Biodiversity Indexes

Variables	Statements
N	Number of species
N	Total number of species
p = n/N	The proportion of the individuals

Numerous biodiversity proposed and reported are described in table 2.2 below:

Table 2.2: Summary of Biodiversity Indexes

Biodiversity indexes	Proposed / Introduced by	Type of Index	Formula	Descriptions
Number of Species (n)	(Gotelli & Colwell, 2001)	Species Richness	n n = number of species	The number of species in each community.
Shannon-Wiener Index	(Shannon, 1949)	Abundance-based diversity index	$H = -\sum_{i=1}^{s} p_i \ln p_i$ p = Proportion that the species indexes are specific forms of the total number of organisms found in the community	Characterize species diversity in a community. When H increases, indicated a more diverse community. Rarely exceeds 4 in practice.
Simpson's Diversity Index	(Simpson E. H., 1949)	Dominance-based diversity index	$D = 1 - \sum_{i=1}^{s} p_i^2$ p = Proportion that the specific forms of the total number of organisms found in the community	The probability of finding a second individual in a sample will the same as the first. D value ranges from 0 to 1. The closer the value to 1, indicates the more diverse the community.

Table 2.2, CONTINUED

Hill Numbers	(Hill, 1973)	Assessing changes in species variety of ecological communities over time, which measures the effective number of species.	${}_0D = S$ ${}_1D = \exp(H')$ ${}_2D = \frac{1}{\sum p^2}$ ${}_nD = \frac{1}{max\{p_i\}}$ ${}_0D = S$,the total number of species. ${}_1$D is the exponential of the Shannon index. ${}_2$D is the reciprocal of Simpson's index. ${}_n$D is another possible evenness index. (Clarke, Gorley, Somerfield, & & Warwick, 2014).	Estimates of the effective number of species present. Incorporate relative abundance and species richness.

2.4.3 Advantages and Disadvantages of Biodiversity Indexes

Table 2.3 depicts the benefits and drawbacks associated with each biodiversity index summarized in table 2.2.

Table 2.3: Advantages and Disadvantages of Biodiversity Indexes

Biodiversity Indexes	**Advantages**	**Disadvantages**
Species Richness	- Simplistic. - Easy to calculate. - Basis comparisons among the site.	- Sensitive to the sample size. - Takes no account of relative abundance. - Does not give weight to species evenness.
Shannon-Wiener Index	- Most widely use. - Easy to calculate. - Represent all species, including rare species. - The mathematical model combines species richness and evenness in a single measure. - Gives more weight for species abundance.	- Insensitive towards species evenness.

Table 2.3, CONTINUED

Simpson's Index	- Widely use. - Easy to calculate. - The mathematical model combines species richness and evenness in a single measure. - Does not require all species represented. - Alternative to Shannon-Wiener index. - Depends on the most abundant species in the sample.	- Heavily affected by changes in common species.
Hill Numbers	- Incorporate biodiversity indexes into one expression - Values are more comparable between two assemblages - Able to calculate absence-presence data - Provides more meaning in the calculation	- Difficult to interpret - Requires ecological interpretation - Not commonly used, however, it gives more insight.

There are many other biodiversity indexes, such as the Berger-Parker dominance index, McIntosh diversity measure, Margalef index, and many more that are not widely used in biodiversity studies. Shannon-Wiener index and Simpson's Diversity Index is the most cited and widely used biodiversity indexes by biologist and ecologist, and hence it has been chosen to be embedded into the online biodiversity module. Both Shannon and Simpson diversities increase as richness increases, for a given pattern of evenness (Colwell & K., 2009)

In this study, the Shannon-Wiener index is chosen as the index for calculating and measuring the species diversity and evenness because it is a useful index for the community in which two or three species are dominant. The indexes also represent a value of maximum diversity, dominance, and evenness as well as a diversity index at once (Yeom & Kim, 2011).

2.5 Citizen Science

Volunteer participation within conservation and ecology has always been an important aspect of research. The use of volunteers has begun to increase and evolve into a term known as "Citizen Science" within the past two decades (Wiggins & Crowston, 2011) (Cohn, 2008) (Oberhauser & Prysby, 2008) (Droege, 2007) (Silvertown, 2009) (Catlin-Groves, 2012). Citizen science is described as a form of research collaboration or data gathering that is performed by unskilled or "non-expert" individuals. It allows the public to participate and collaborate in scientific research and at the same time, it may ease data monitoring and collection since a large amount of data is needed to be collected. Eventually, the collected data can be put up with the study of species occurrence, large-scale patterns in nature, and distribution around the world (Geoghegan, Dyke, Pateman, West, & Everett, 2016).

(de Sherbinin, et al., 2021) stated that citizen science is a significant vehicle for democratizing science and promoting the goal of universal and equitable access to scientific data and information. Citizen science serves multiple needs, both in research and in education. For instance, it allows gathering data across a wider range, while compared to traditional scientific teams and methods. It also presents a new model for delivering educational information, while also offering engagement and inquiry opportunities for the public to learn and gain awareness advocated by a learning specialist. Besides, citizen scientists are learning science at the same time they are challenging scientific orthodoxies and making claims on the governance of science (Lewenstein, 2016).

Citizen science has the potential to transform environmental protection by involving the public in working with agencies to generate knowledge and find solutions (Ballard, Phillips, & Robinson, 2018). Environmental protection agencies are increasingly turning to citizen science to assist in the achievement of environmental protection (Owen & Parker, 2018) and innovative technology is enabling more citizen-agency dialogue and feedback on the use of evidence (Wiggins & Crowston, Surveying the citizen science landscape, 2015) (Gray, et al., 2017). Environmental protection agencies rely heavily on good-quality evidence about the health of ecosystems, pressures on these natural resources, and the effectiveness of regulatory and other interventions. citizen science has an increasingly important role in providing evidence, raising environmental awareness, and empowering the public (Smallman, 2018) (Nascimento, Rubio Iglesias, Owen, Schade, & Shanley, 2018).

According to (Toogood, 2013), involving the public to participate in biodiversity monitoring and research are regarded as having significant or potential for biodiversity conservation, and often held to achieve a win-win benefit, whereby allowing the public to engage and help in monitoring the environment, collecting research data, eventually raise the concern and awareness of conserving the environment. Furthermore, it is capable of enhancing scientific education and, most importantly, researchers are capable of utilizing various publics as data gatherers and obtaining vast biological data in a short period. The figure 2.1 below depicts a straightforward illustration of the definition of citizen science by (Haklay, et al., 2021).

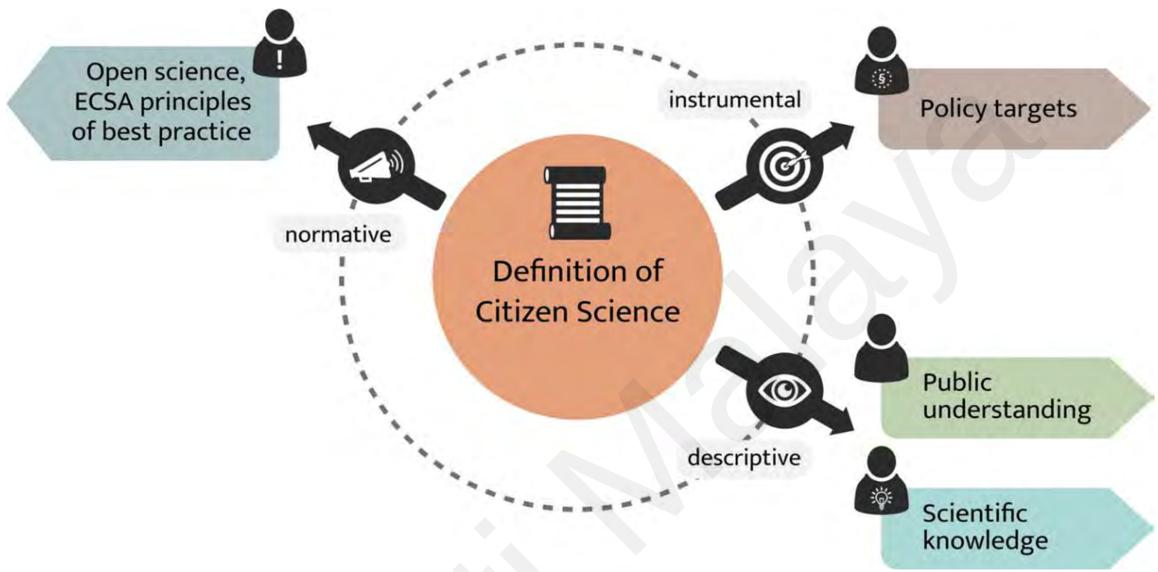


Figure 2.1: Citizen Science Definition adapted from (Haklay, et al., 2021).

2.5.1 Citizen Science Projects

Various Citizen Science Projects have been established and have remarkable success especially with the technology available, such as mobile phones. Citizen science projects cover a wide variety of topics, from astronomy to zoology. Citizen-science projects may include wildlife-monitoring programs, online databases, visualization, sharing technologies, or other community efforts.

The National Geographic Society established the BioBlitz citizen science project, which focuses on finding and identifying as many species as possible in a specific area in a short period. It is also known as a biological database, and the project's goal is to get an overall count of the plants, animals, fungi, and other organisms that live in a given area (Parker et al., 2018), as well as to raise public awareness about biodiversity. The Seek is a tool that focuses on younger people in identifying plants and animals in a few quick steps, whereas iNaturalist collects actual science data and shares these observations with a network of enthusiasts and experts (Geographic, 2019).

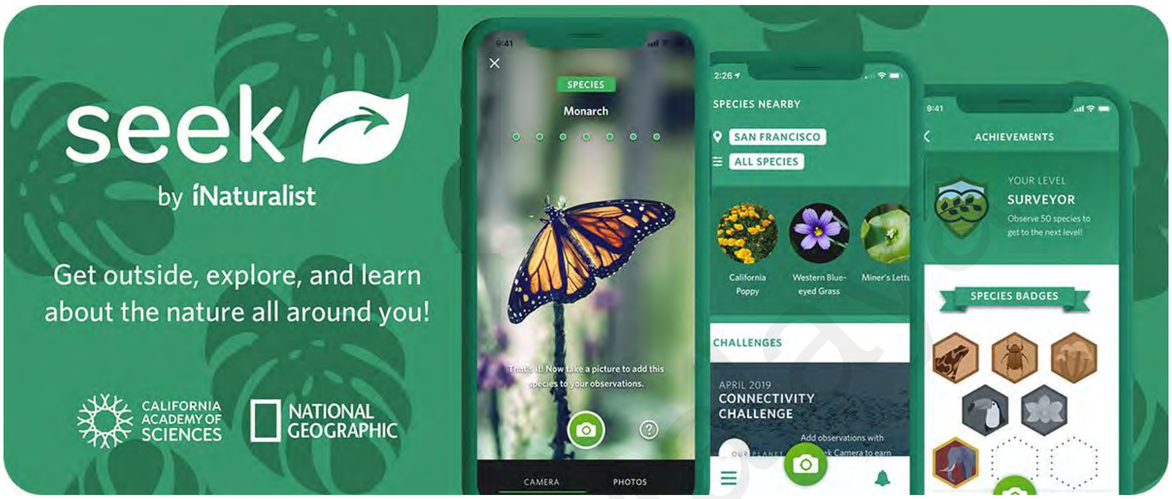


Figure 2.2: Citizen Science Project Seek Application Developed by iNaturalist that assists users in identifying natural species (Casimiro, 2022)

iNaturalist is a platform that was used as the recording tool, which built on the concept of mapping and sharing observations of biodiversity across the globe. iNaturalist is an online social network of people sharing biodiversity information to help each other learn about nature, focusing on connecting people to nature (Matheson, 2014). iNaturalist is an application and consider as one of the Citizen Science projects based on crowdsourcing of observation and identification of plants and animals. Besides, iNaturalist can create high-quality research data for scientists, whereby the data is verified by experts (iNaturalist, iNaturalist, 2020). Figure 2.3 below shows the print screen of the iNaturalist webpage.

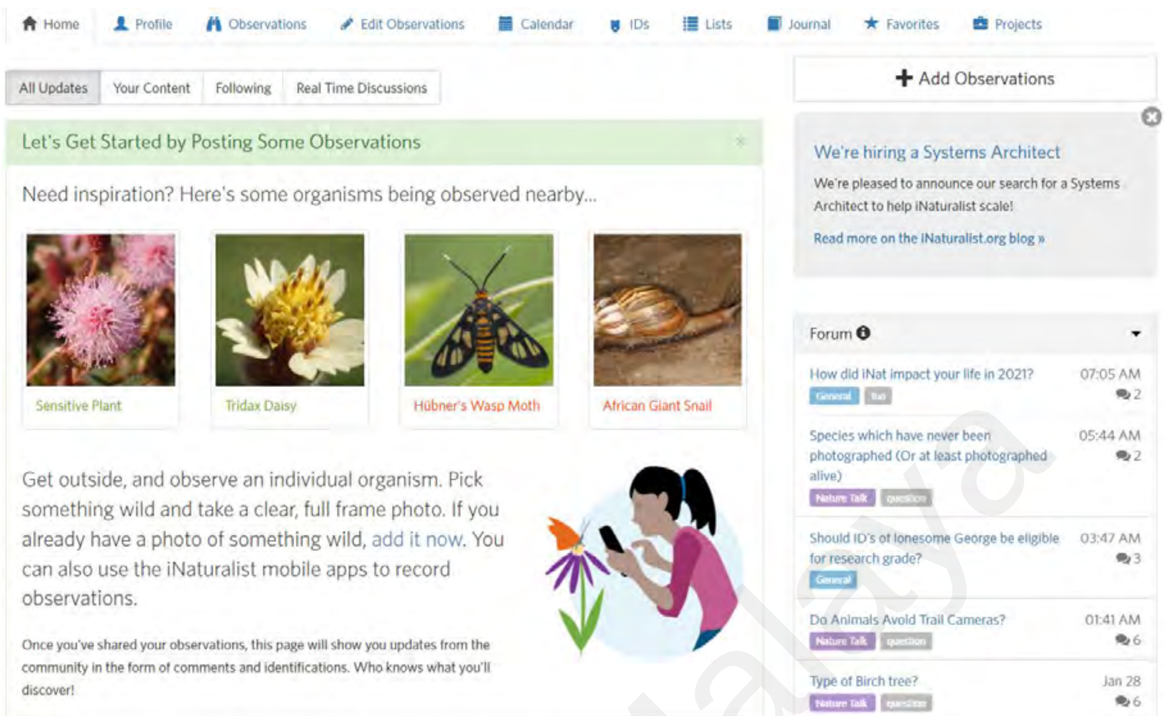


Figure 2.3: Citizen Science Project iNaturalist Log in Homepage (iNaturalist, iNaturalist, 2020)

Project BudBurst is a citizen science project that focuses on climate and climate change by encouraging the public to monitor leafing, flowering, and fruiting of plants. This project fosters collaboration among gardeners, hikers, botanists, ecologists, government agencies, and educators to monitor climate change and its impacts on plants. It is a national field campaign that is designed to engage the public in collecting data about plant phenology (Henderson, Ward, Meymaris, Alaback, & Havens, 2012).

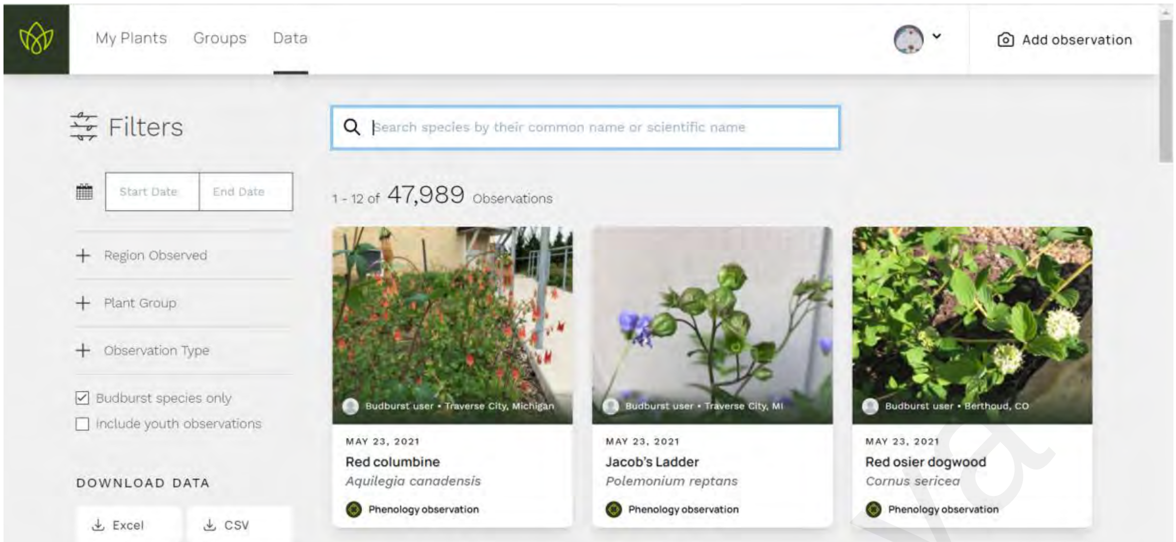


Figure 2.4: Citizen Science Project Budburst Homepage (Garden, 2021)

There have been a lot of Citizen science projects as newer technologies increased the options for collecting data for research and experiments (Drollette, 2012). Projects such as *BugGuide* is another platform where naturalist share their observations about bug creatures, including spiders and other insects. The information is collected by the in-house expertise of scientists and amateur experts, and the information is used to study the diversity of bug species, however, the project is more focused on the USA and Canada (Acorn, 2017).

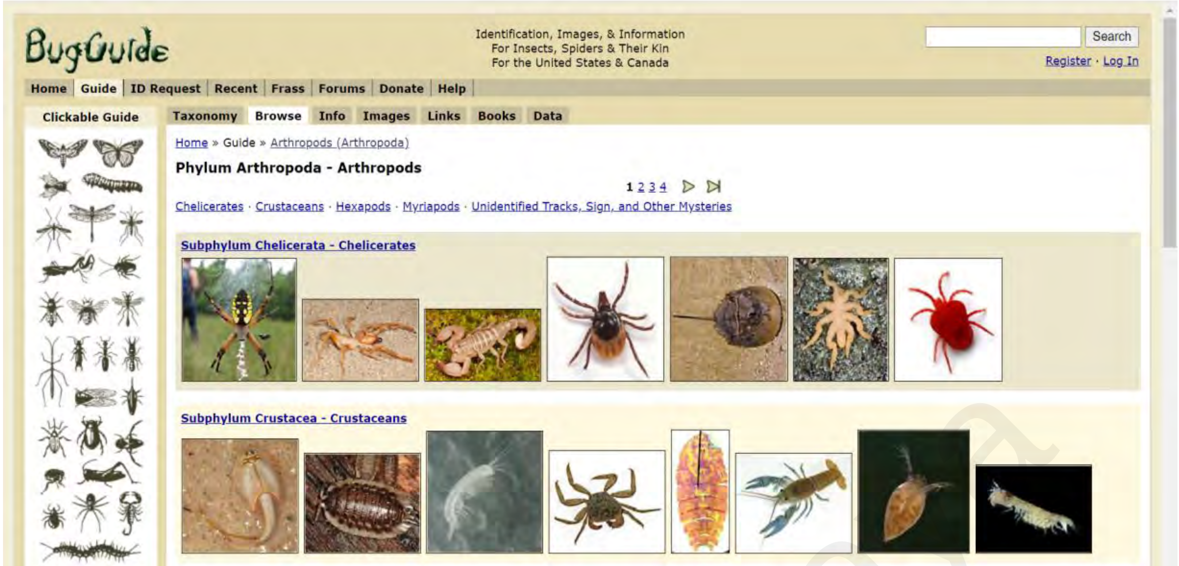


Figure 2.5: Citizen Science Project BugGuide Homepage (University, 2021)

FrogWatch USA, this program allows groups and individuals to learn about wetlands in their communities by reporting mating calls of local frogs and toads and collecting data in community wetlands (Flagg, 2016).

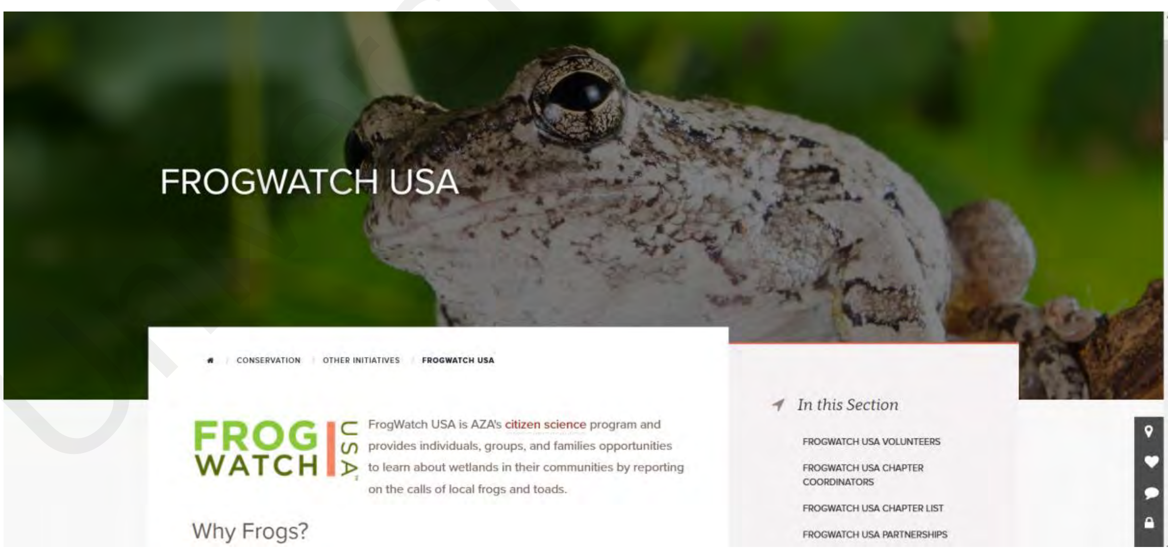


Figure 2.6: Citizen Science Project Frogwatch USA Homepage (AZA, Frog Watch, 2021)

The Cornell Lab of Ornithology plays host to many citizen-science programs. It has a variety of bird programs, including *NestWatch*, which teaches people how to monitor nests and collect breeding information to track reproduction among North American birds (Evans, et al., 2005).

Figure 2.7: Citizen Science Project NestWatch (AZA, Nest Watch, 2021)

Citizen science also includes astronomy projects – *Zooniverse* is an online aggregator that supports a wide variety of citizen scientists. With Zooniverse's Galaxy Zoo, Moon Zoo, and Solar Stormwatch projects, armchair astronauts virtually explore distant galaxies, study the surface of the Moon, and investigate solar explosions—all just a click away on the computer (Simpson, Page, & De Roure, 2014).

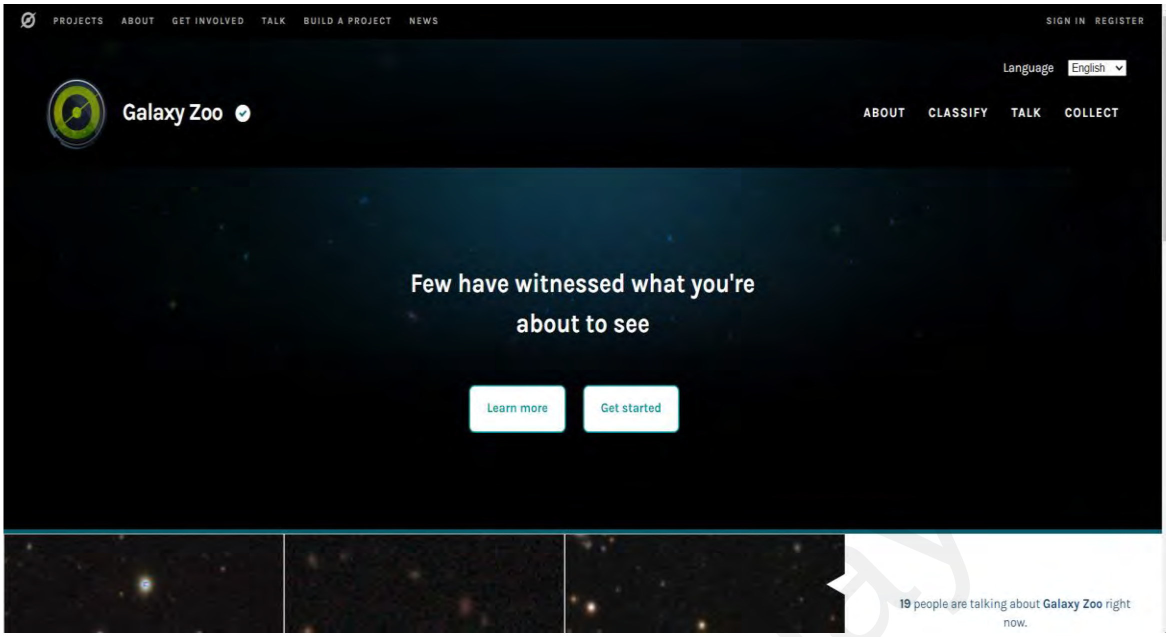


Figure 2.8: Citizen Science Project Galaxy Zoo (Zooniverse, 2021)

The EarthEcho Water Challenge, formally known as the World Water Monitoring Challenge, is a citizen science project run by Earth Echo International. This program runs annually beginning March 22 (the United Nations World Water Day), with raising public awareness and involving the public in protecting water resources around the world by conducting basic monitoring of their local waterbodies. The program begins with a simple test kit capable of testing the state of water quality, followed by online sharing of observation information to the Earth Echo Water Challenge database, and with the test results, the local community can take action and protect vital water resources (Nugent, 2020).

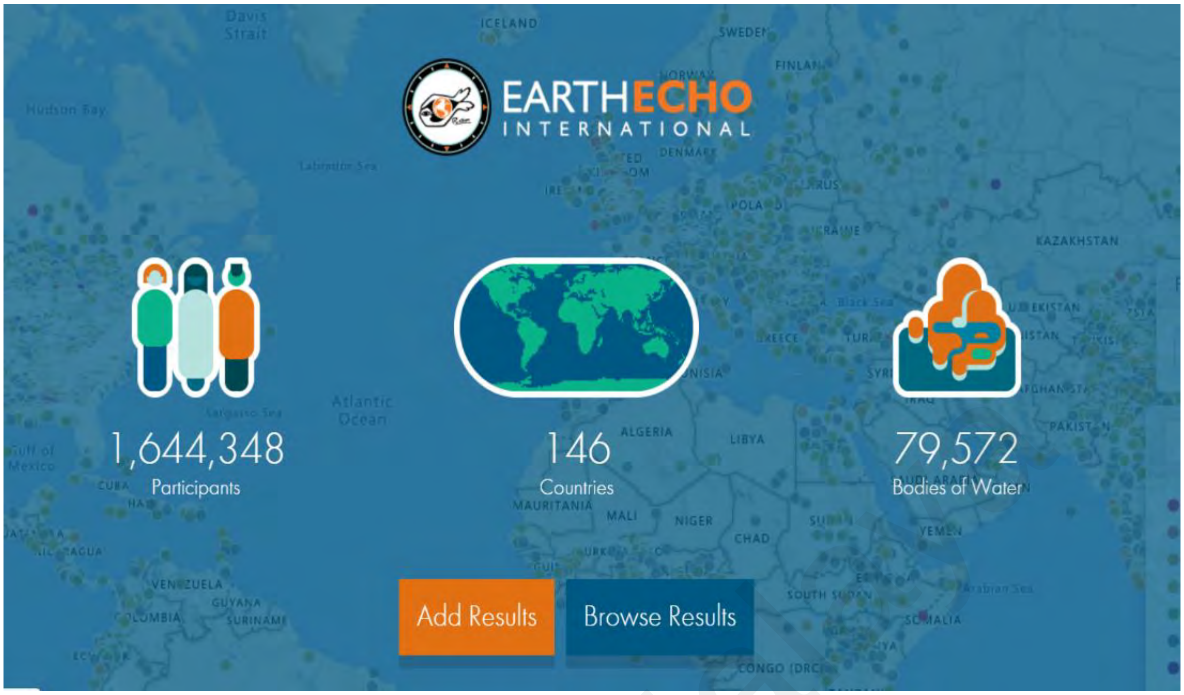


Figure 2.9: Citizen Science EarthEcho Project Dashboard (Echo, 2021)

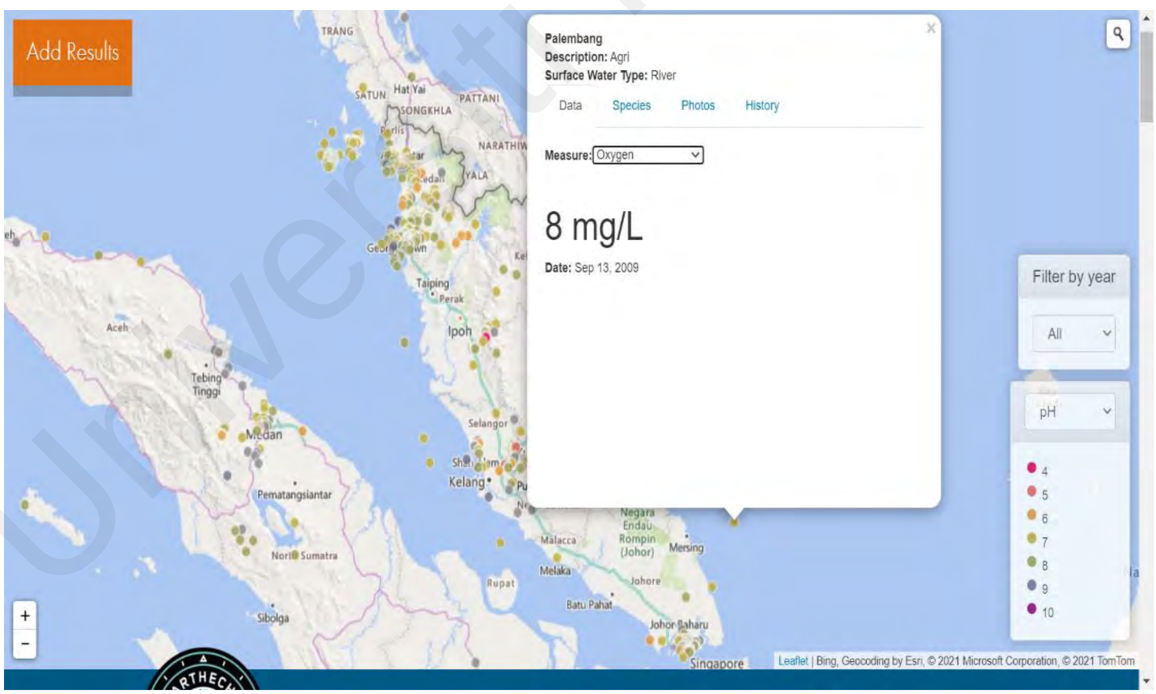


Figure 2.10: EarthEcho Browse Results Page (Echo, 2021)

The National Optical Astronomy Observatory (NOAO) runs the Globe at Night program, and the GLOBE project has been running since March 2006. This citizen science project allows "citizen scientists" all over the world to contribute simple unaided-eye observations of skyglow (sky brightness) to a growing global database (Walker, Pompea, & Isbell, 2008). Skyglow is a type of light pollution caused by the scattering of artificial light in the atmosphere, and it is a critical global environmental issue due to its potential ecological effects (Moore, Pierce, Walsh, Kvalvik, & Lim, 2000), (Rich & Longcore, 2006), (Kyba, Ruhtz, Fischer, & Hölker, 2011), and it also generates a large amount of electrical consumption (Co-operation & (IEA), 2006). Furthermore, data on the luminance of the night sky at specific locations is very poorly known on the Earth's surface, and the goal of Globe at Night is to improve understanding of the biosphere (Kyba, et al., 2013).

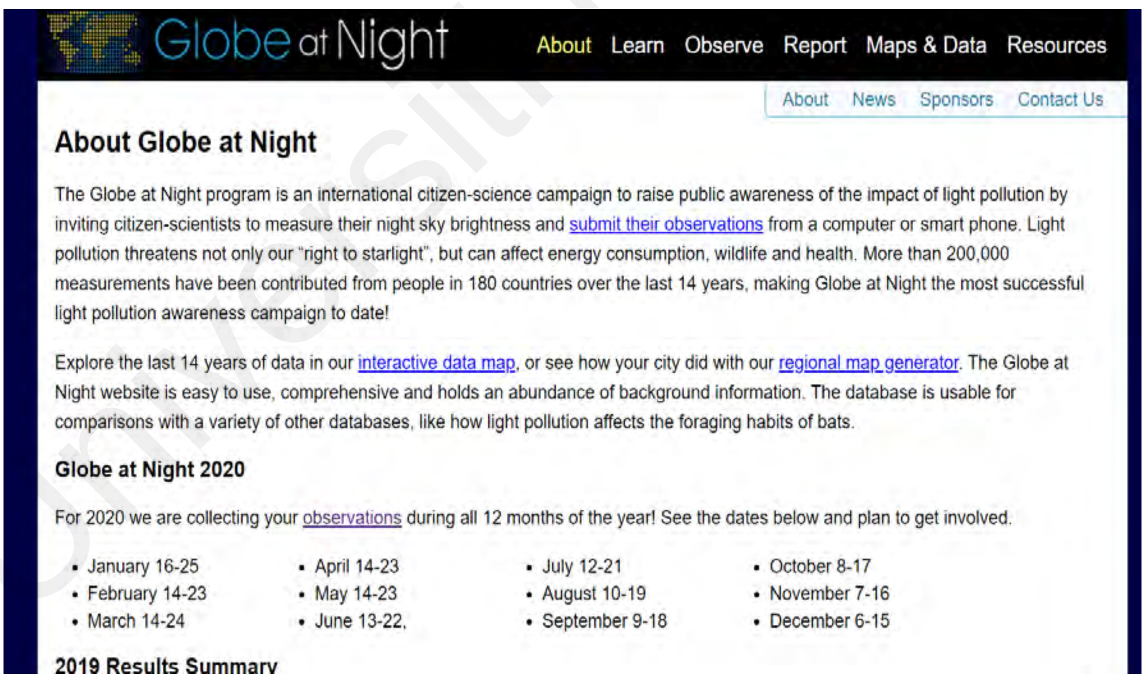


Figure 2.11: Citizen Science Project Globe at Night 'About' Page (NOIRLab, 2021)

Figure 2.12: Globe at Night Results Infographic (NOIRLab, 2021)

The University of Florida's Florida LAKEWATCH program began in 1986. Florida LAKEWATCH's main purpose is to get volunteers to collect data on total phosphorus, total nitrogen, chlorophyll a, and water clarity in Florida lakes. To assure data reliability, LAKEWATCH employees performed comprehensive quality assurance and quality control. Canfield, C. D. Brown, Bachmann, & Hoyer, 2002) found that data obtained by LAKEWATCH volunteers correlated well with data collected by professionals

(Gommerman & Monroe, 2012).

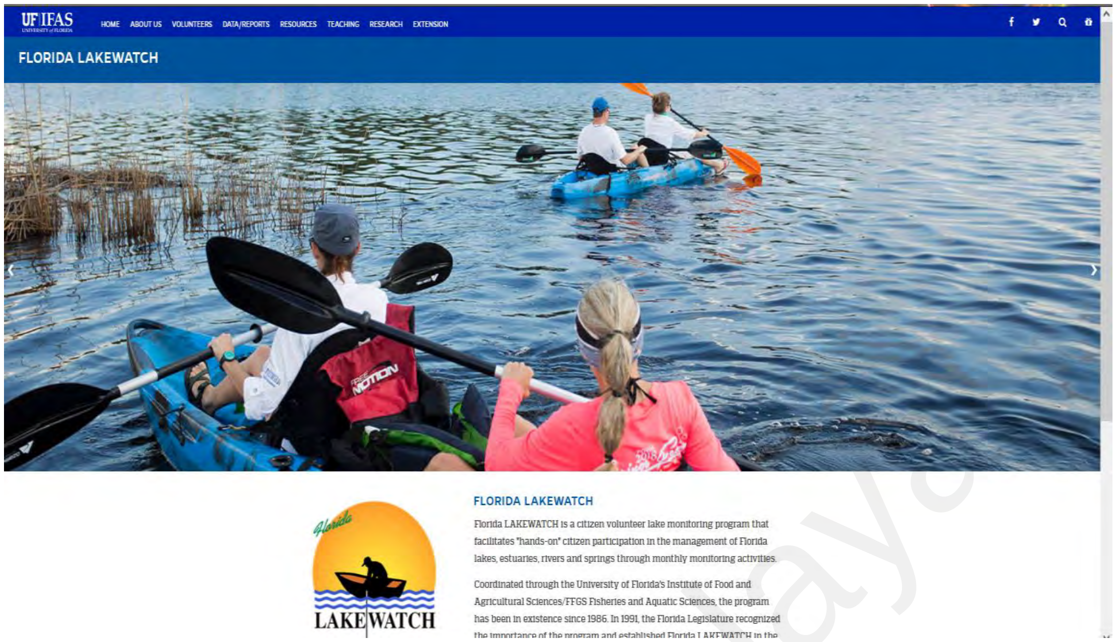


Figure 2.13: Citizen Science Project LAKEWATCH Homepage (Florida, 2021)

(Stevenson, Suomela, Kim, & He, 2021) found seven key data kinds in Citizen Science. "Carry Instrument Packages" (CIP) or "Pilot Vehicles that CIP" do not actively monitor or record data after the instrument is fitted. The second data category (IMA) is participation in algorithm invention or modification. This data type is mostly a game or a contest. The third data category is "Sorting and Classifying of Physical Objects" (SCPO), where researchers already have data but need help arranging it. Because researchers and scientists work closely together, problems about data quality are swiftly handled.

The fourth category is "Sort and Classify Digital Objects" (SCDO), where scientists collect and organize photos, audio recordings, and videos. The public sorts and categorizes the collected digital data. To find and collect CPOs, citizens assist scientists to find them over time and space. The data obtained by citizen scientists will be sent to scientists for further processing and validation, resolving the data quality issue.

The sixth category, called "Collect Digital Objects," is the same as the fifth category (CDO). Smartphones with built-in clocks and GPS units enhance the observer and observation process easier. The benefit of this category is that the data obtained may be easily shared and reviewed by the community. The seventh data type is called “Report Observations” (RO), which includes quantitative measurements, counts, and data sampling. However, most observations are made up of simple number or text data. It is easy to manage collected data. Table 2.4 summarized the citizen science project discussed in Section 2.5.1 along with the data category.

Table 2.3: Citizen Sciences Projects

Citizen Science Project (website)	Primary Research Organization Responsible	Topic	Description	Data Category	Data Contribution
BioBlitz	The National Geographic Society	Biodiversity	To get an overall count of the plants, animals, fungi, and other organisms that live in a place.	Record observation	Text from instrument readings, counts, classifications, and/or descriptions
iNaturalist (https://www.inaturalist.org/)	California Academy of Sciences and the National Geographic Society	Biodiversity	A social network of naturalists, citizen scientists, and biologists built on the concept of mapping and sharing observations of biodiversity across the globe.	Collect digital objects	Digital object obtained and submitted; collection process documented.

Table 2.5, CONTINUED

BudBurst (https://budburst.org/)	Chicago Botanic Garden	Phenology	Budburst apposes researchers, horticulturists, and citizen scientists to study plants affected by human impacts on the environment through data collection, data sharing, education, and personal connections.	Collect digital objects	Digital object obtained and submitted, collection process documented.

Table 2.5, CONTINUED

FrogWatch USA (https://www.aza.org/)	The Association of Zoos and Aquariums (AZA) the USA	Herpetology	FrogWatch USA invites citizen scientists to learn about their local wetlands and help conserve amphibians by reporting data on the calls of local frogs and toads during the monitoring season of February to August.	Report observation	Text from instrument readings, counts, classifications, and/or descriptions

Table 2.5, CONTINUED

NestWatch (https://nestwatch.org/)	Cornell Lab of Ornithology	Ornithology	NestWatch is a nationwide monitoring program designed to track status and trends in the reproductive biology of birds, including when nesting occurs, the number of eggs laid, how many eggs hatch, and how many hatchlings survive.	Report observations	Text from instrument readings, counts, classifications, and/or descriptions
Zooniverse (Galaxy Zoo) (https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/)	Citizen Science Alliance (CSA)	Astronomy	A scientific project that has invited members of the public to help classify a million galaxies	Sort and Classify digital objects	Digital object categorized

Table 2.5, CONTINUED

Earth Echo Monitor Water Challenge (https://www.monitorwater.org/)	America's Clean Water Foundation	Environment Science	Engaging citizens around the world to conduct basic monitoring of their local waterbodies.	Collect physical objects and report observations	Sample obtained and submitted, collection process documented, and Text from instrument readings, counts, classifications, and/or descriptions

Table 2.5, CONTINUED

Globe At Night (https://www.globeatnight.org/)	Global Learning and Observations to Benefit the Environment (GLOBE) Program	Environment Science	Globe at Night is an international citizen-science campaign to raise public awareness of the impact of skyglow by involving the public through measures & submitting their night sky brightness observations.	Report observations	Text from instrument readings, counts, classifications, and/or descriptions

Table 2.5, CONTINUED

Florida Lake Watch (https://lakewatch.ifas.ufl.edu/)	The University of Florida's Institute of Food and Agricultural Sciences	Environment Science	A citizen volunteer lake monitoring program that facilitates "hands-on" citizen participation in the management of Florida lakes, estuaries, rivers, and springs through monthly monitoring activities.	Collect physical objects	Sample obtained and submitted; collection process documented.

However, “Citizen Science” projects have raised questions about the quality of data collected by non-professional scientists and volunteers (Elliott, Kevin, & Jon, 2019). Some studies may not be acceptable for volunteers due to sophisticated research methods and lack of proper training in research and monitoring protocols. However, correctly managing the input data, such as researcher verification, input data is adequately supplemented with graphics and sound, can overcome the data quality issue. Various Citizen Science Projects have been launched with great success, especially with modern technologies like mobile phones (Thelen & Thiet, 2008).

2.6 Existing Biodiversity System

The available biodiversity systems only focus on one element, either a calculation or a visualization. The existing technology does not connect to more freely accessible citizen science data. Our system uses data from iNaturalist, a citizen science organization that collects biodiversity data. No other online system linked to citizen science data has been found, and their data is largely imported from local sources. The next section examines the most often used biodiversity measurement.

2.6.1 Biodiversity Calculator

Tanner M. Young (https://www.alyoung.com/labs/biodiversity_calculator.html) has developed an online biodiversity calculator mainly for biologists, ecologists, researchers, lecturers, and students that allows calculating biodiversity index immediately. Besides, this online biodiversity calculator is free to use.

Users are required to enter the number of species in a sample, then it will proceed to the next page - Enter Data, users are required to enter the name of the species and population, there are many other fields but it is optional, which are the region, meta-region, the coordinates (latitude and longitude) and date.

Next, the script will return the calculated biodiversity index values together with graphs. The biodiversity index calculators calculate various biodiversity indexes, comprises 22 formulas and methods, such as Shannon-Wiener index, Simpson index, Dominance index, Reciprocal Index, Berger-Parker Dominance, Margalef Richness, etc. Figure 2.14 – Figure 2.17 illustrates the usage of the online calculator that can be found on the URL below.

URL: https://www.alyoung.com/labs/biodiversity_calculator.html

Figure 2.14 shows the homepage of the biodiversity calculator, first, the users need to enter the number of species/taxa in a sample. For demonstration purposes, '3' is entered for the number of species and the 'SUBMIT' button was clicked.

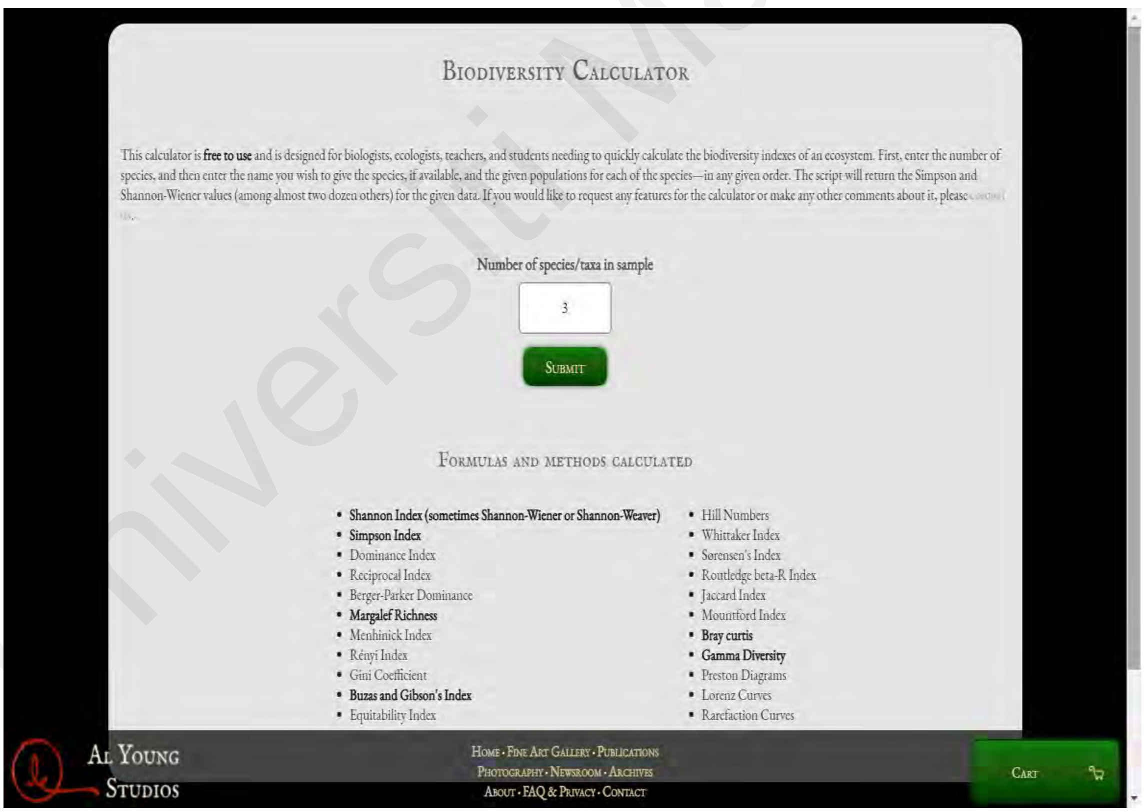


Figure 2.14: Screen Capture of Biodiversity Calculator Homepage

Figure 2.15 illustrates after entering value '3' in the biodiversity calculator homepage, it will direct into the "Enter Data" page, which allows the user to enter the plant species name and the populations. As a demonstration, rubber, palm, and bamboo with the numbers 23, 130, and 5 are entered respectively. By clicking on the 'SUBMIT' button, the results of the calculations and graphs will be displayed on the new page as shown in Figure 2.16 and Figure 2.17.

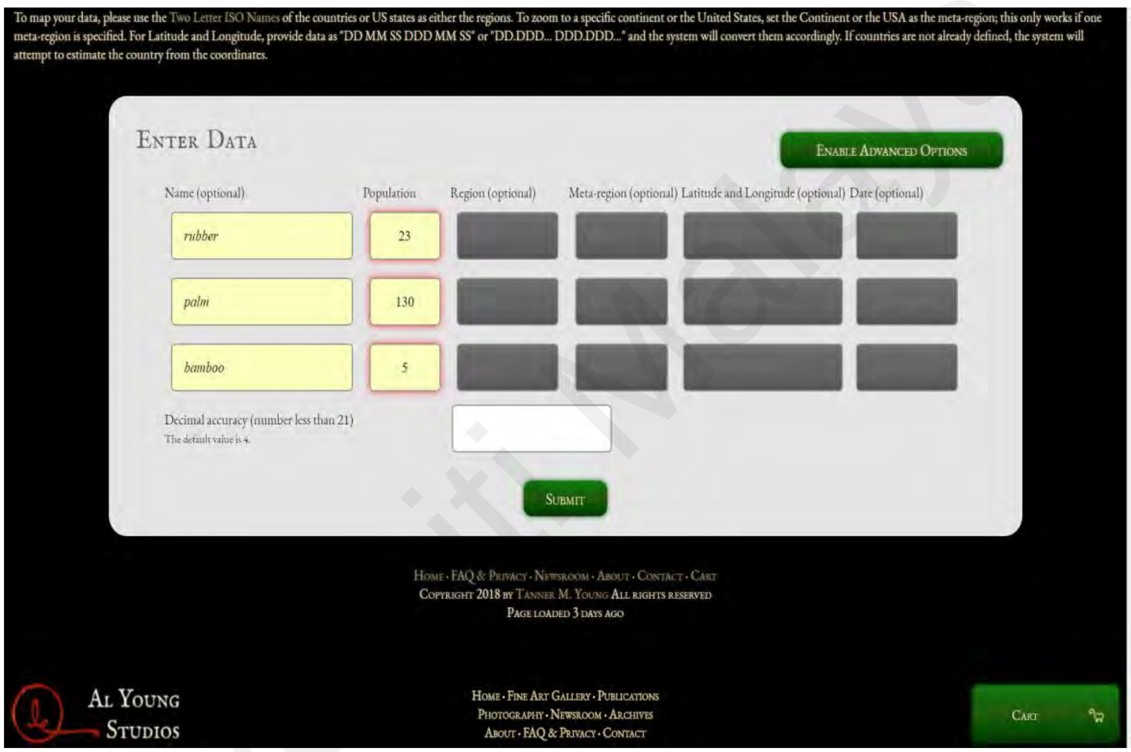


Figure 2.15: Screen Capture of entering Data Page.

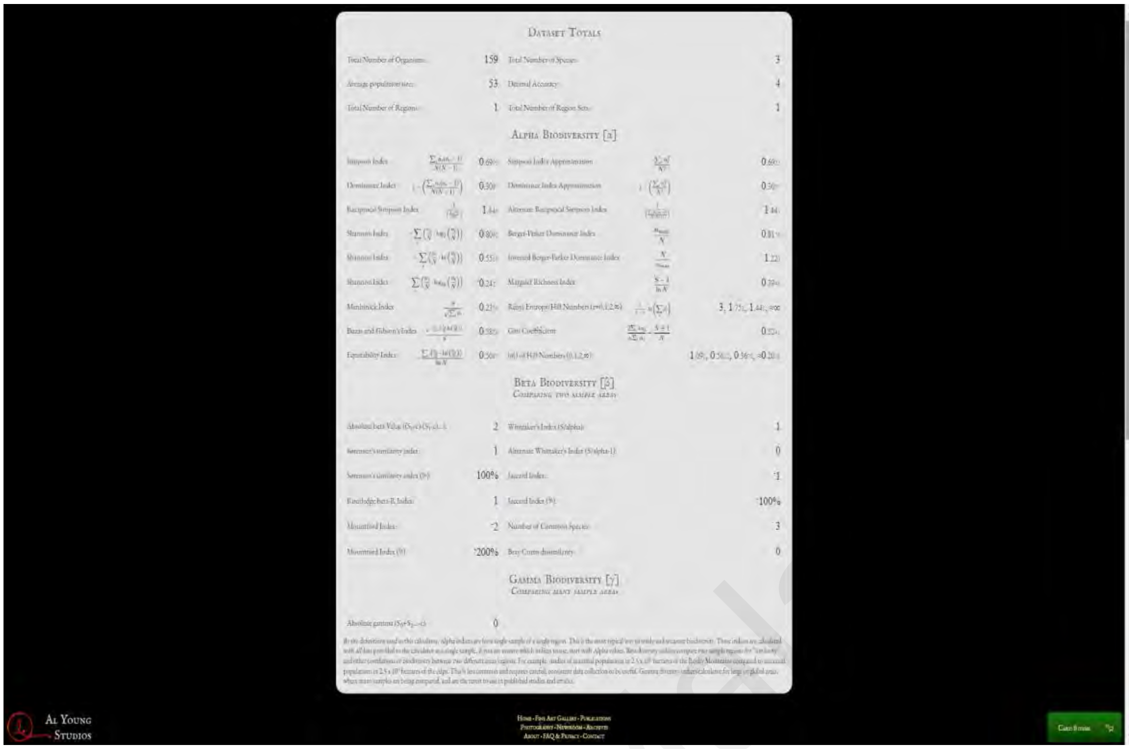


Figure 2.16: Screen Capture of the Results after Calculated the Diversity Indexes.

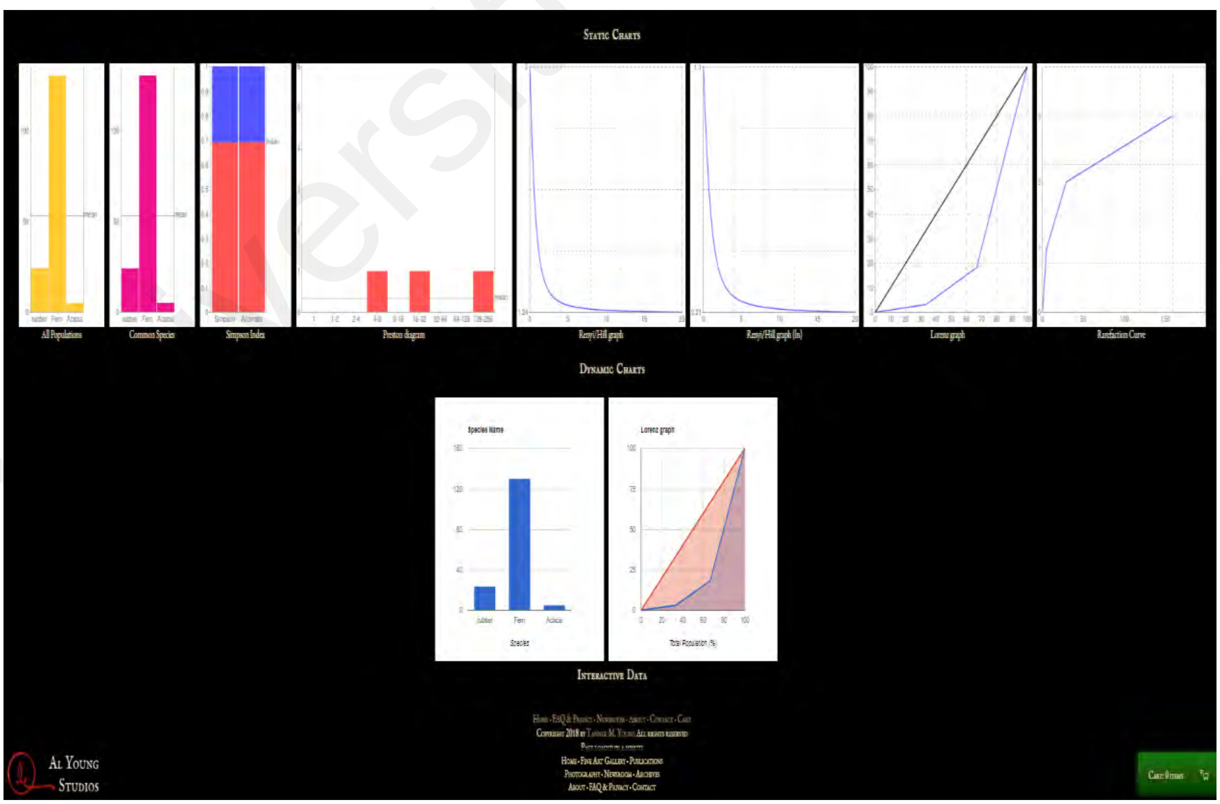


Figure 2.17: Screen Capture of the graphs of diversity indexes calculated.

2.6.2 BPMSG Diversity Online Calculator

Business Performance Management Singapore (BPMSG) developed a Diversity online calculator that has a similar function as the Al Young studio biodiversity index calculator, the difference is the online diversity calculator, users can download an excel template for this calculation.

The results of this diversity online calculator will give the Shannon entropy, Shannon equitability, true diversity, Berger-Parker index, Simpson dominance, Gini-Simpson index, and Simpson equitability. URL: https://bpmsg.com/academic/div-calc.php

Figure 2.18 shows the homepage of the BPMSG Diversity online calculator. First, enter the number of categories/classes Categories, it is like the Al Young Biodiversity Calculator, which simply means the number of species. Then, depending on the number of 'NEW INPUT', in the 'DATA INPUT' section, require users to enter the biodiversity data. As a demonstration, again value 3 is the 'NEW INPUT', bamboo, fern, and acacia with the value of 35, 150, and 24 are entered respectively, and then SUBMIT button is clicked.

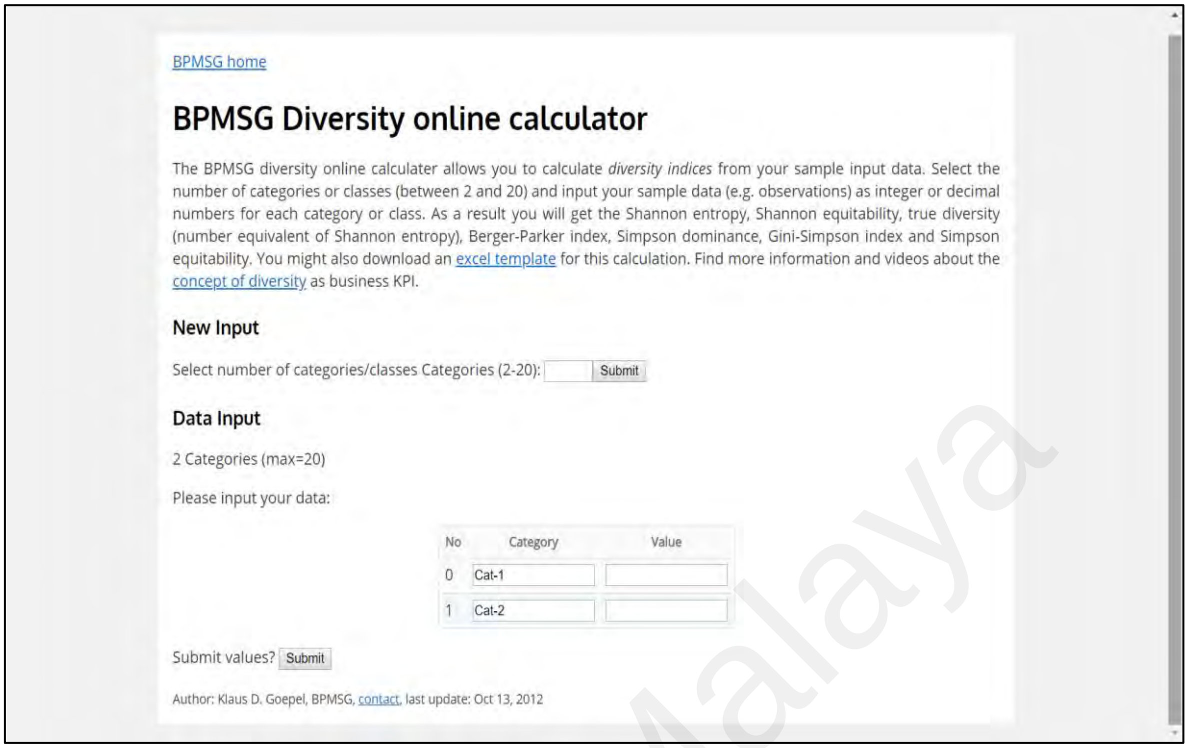


Figure 2.18: Screen capture of BPMSG Diversity online calculator.

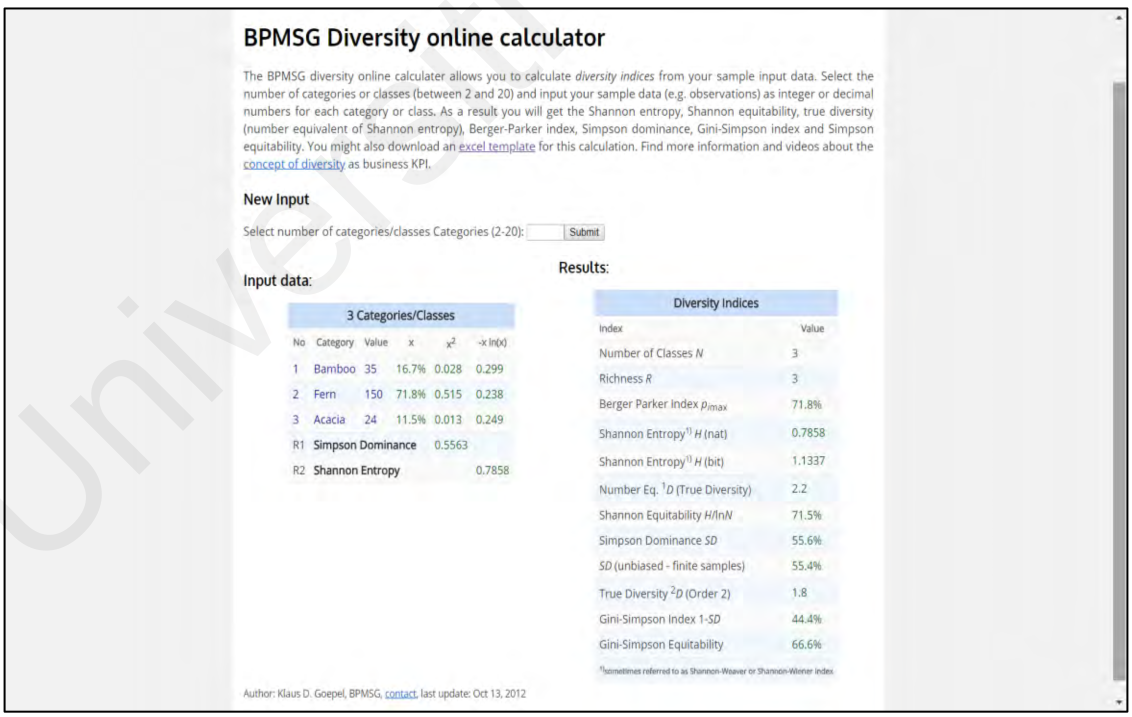
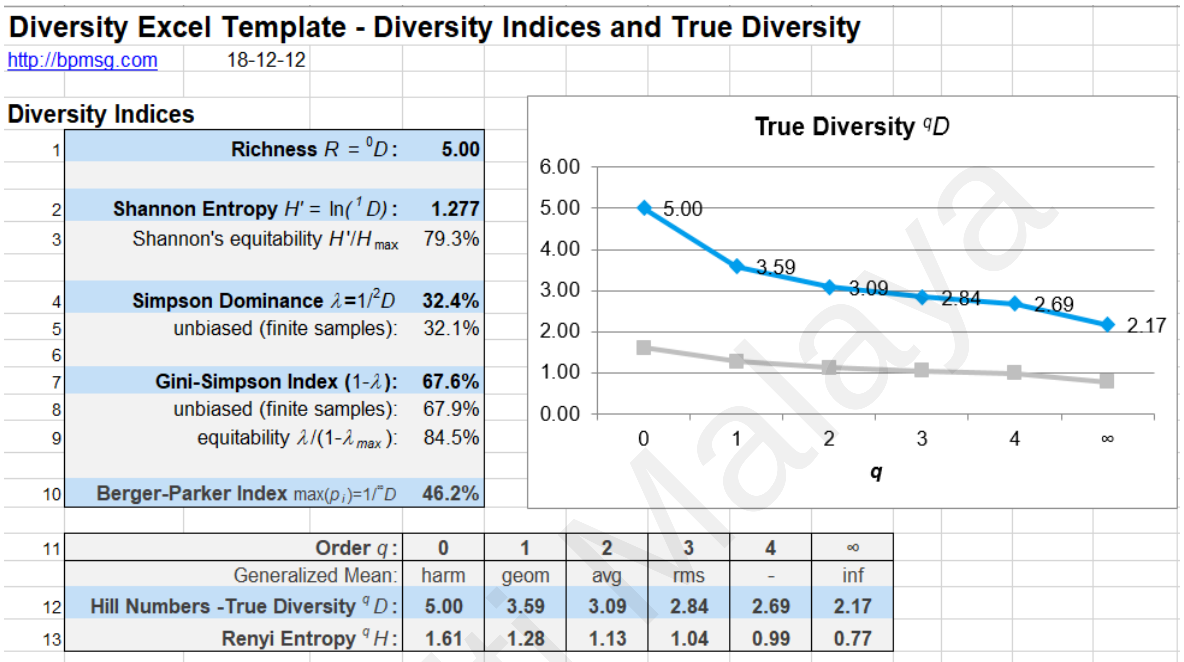


Figure 2.19: Screen Capture BPMSG Diversity Online Calculator Biodiversity Indexes Results.

Besides, the BPMSG Diversity Online Calculator also provides a template to calculate the diversity, where users just download the template and enter the desired data, the diversity indexes are automatically calculated. It is shown in Figure 2.19 – Figure 2.22.



Diversity Excel Template - Diversity Indices and True Diversity

http://bpmsg.com 18-12-12

Diversity Indices

1	**Richness** $R = {}^0D$:	**5.00**
2	**Shannon Entropy** $H' = \ln({}^1D)$:	**1.277**
3	Shannon's equitability H'/H_{max}	79.3%
4	**Simpson Dominance** $\lambda = 1/{}^2D$	**32.4%**
5	unbiased (finite samples):	32.1%
6		
7	**Gini-Simpson Index** (1-λ):	**67.6%**
8	unbiased (finite samples):	67.9%
9	equitability $\lambda/(1-\lambda_{max})$:	84.5%
10	**Berger-Parker Index** $\max(p_i) = 1/{}^{\infty}D$	**46.2%**

11	Order q:	0	1	2	3	4	∞
	Generalized Mean:	harm	geom	avg	rms	-	inf
12	Hill Numbers -True Diversity ${}^qD$:	5.00	3.59	3.09	2.84	2.69	2.17
13	Renyi Entropy ${}^qH$:	1.61	1.28	1.13	1.04	0.99	0.77

Figure 2.20: Screen capture of downloaded diversity excel template (1).

Samples/Observations/Values													
No i	Class Category Species	value n_i	prop. p_i	0	1	2	3	4					
1	A	10	5.1%	1	0.1523	0.0026	0.0001	7E-06	4	2	B	90	46.2%
2	B	90	46.2%	1	0.3569	0.213	0.0983	0.0454	1	5	E	50	25.6%
3	C	5	2.6%	1	0.0939	0.0007	2E-05	4E-07	5	4	D	40	20.5%
4	D	40	20.5%	1	0.3249	0.0421	0.0086	0.0018	3	1	A	10	5.1%
5	E	50	25.6%	1	0.349	0.0657	0.0169	0.0043	2	3	C	5	2.6%
6									6	6	0	0	
7									7	7	0	0	
8									8	8	0	0	
9									9	9	0	0	
10									10	10	0	0	
11									11	11	0	0	
12									12	12	0	0	
13									13	13	0	0	
14									14	14	0	0	
15									15	15	0	0	
16									16	16	0	0	
17									17	17	0	0	
18									18	18	0	0	
19									19	19	0	0	
20									20	20	0	0	
Total	5	195	100.0%	5	1.277	0.3241	0.124	0.0515					

Figure2.21: Screen capture of downloaded diversity excel template (2)

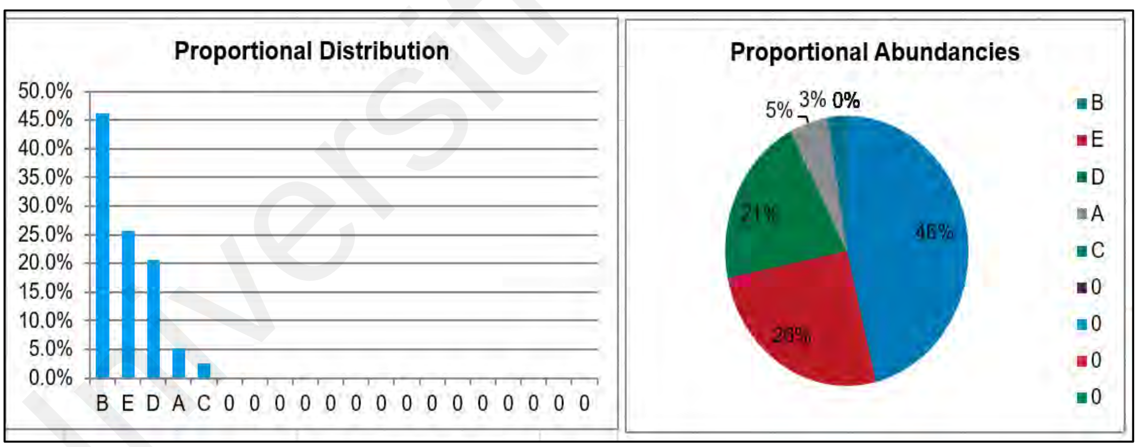


Figure 2.22: Screen capture of downloaded diversity excel template (3).

2.6.3 Pericopsis.org

Pericopsis.org is a free and collaborative database for the localization and identification of trees. The objective of Pericopsis.org is that knowledge of biodiversity is vital for living in harmony with nature. Pericopsis.org has been written by Jean Weber, Institute of Tropical Forestry and Forest Products (INTEROP), Universiti Putra Malaysia (UPM).

Thus, it allows users to contribute to the database by signing up an account and contributing to the page, it is free and the information can manage after registration. Non-contributors can access the consultation page to locate and obtain the tree species location. Pericopsis.org is open to the contributions of citizens as well as to the databases of urban communities.

Besides, Pericopsis.org has various functions, including showing trees on the map by drawing a polygon of the area of interest on the map. It provides filtering functions, users can filter the selection from the database by area, elevation range, names, and evaluation. Furthermore, the users can download 3 various files, which are the .csv file that contains the tree information, the KML file which is the polygon shape together with the marker and shapefile for GIS applications.

URL: http://pericopsis.org/index.php

Figure 2.23 shows the Contribution page of pericopsis.org. Where the users draw a polygon of area-of-interest and then click on 'SHOW TREES' it will show the trees cover by the polygon as shown in Figure 2.24.

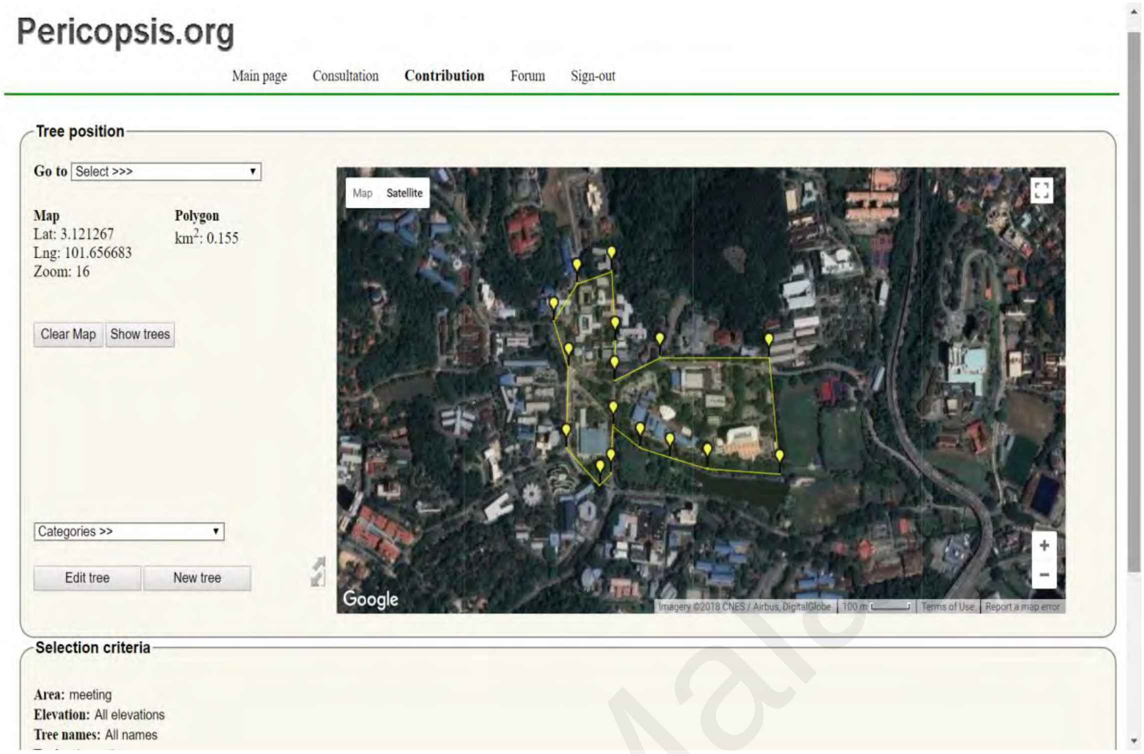


Figure 2.23: Screen Capture of Pericopsis.org Contribution Page.

Figure 2.24 shows the trees within the polygon. The users can download the file in CSV (Excel) that contains the trees information, KML file, and Shapefile that can be used in ArcGIS. The number of trees that the polygon covers are also shown beside the map.

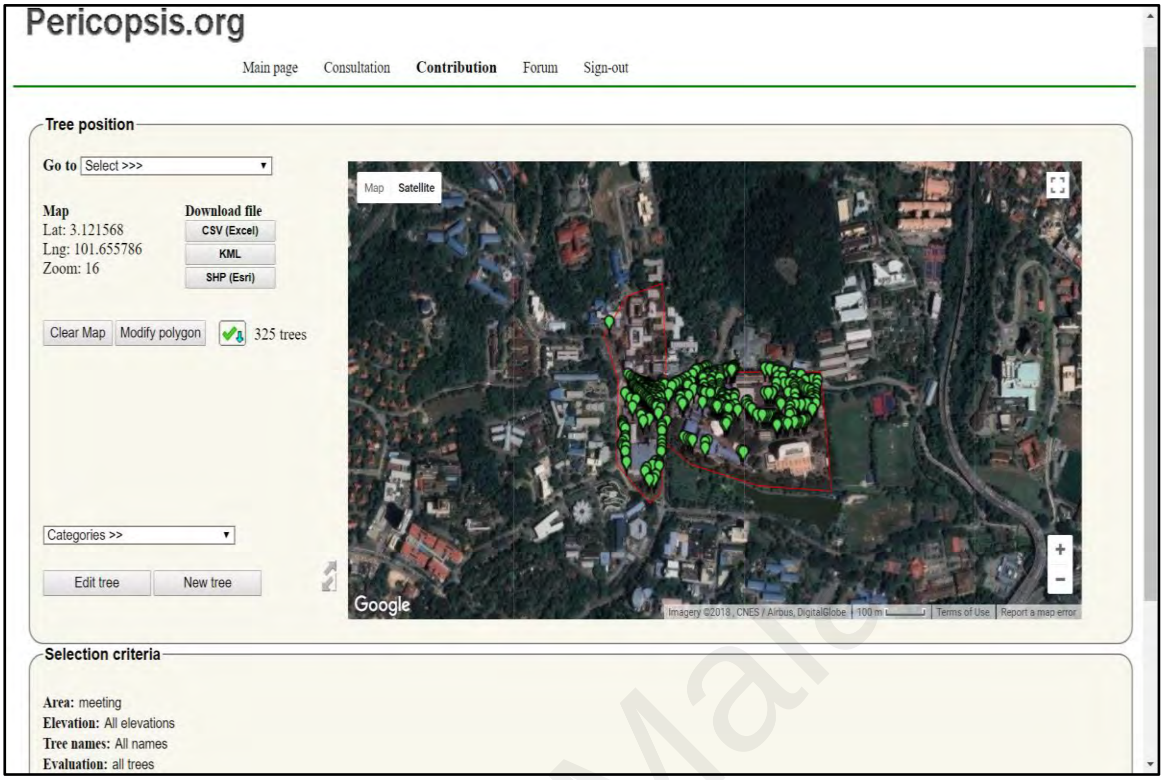


Figure 2.24: Screen Capture of Pericopsis.org Contribution Page (result).

2.6.4 Comparison of Online Biodiversity System

Table 2.5 shows the summary of the advantages and disadvantages of each available biodiversity online system.

Table 2.5: Comparison of Online Biodiversity System

Online Biodiversity System	Advantages	Disadvantages
Al Young Biodiversity Calculator	- Simple - Free - Calculates many diversities indexes - Charts and Graph are shown - Advance option available for entering data - No registration is required	- Does not include visualization element. - Unable to export the data calculated. - Data needs to be entered manually.
BPMSG Diversity Online Calculator	- Simple - Free - Excel Template for calculation	- Does not include visualization element. - Fewer diversity indexes calculated. - Unable to export the data calculated - Data needs to be entered manually

Table 2.5, CONTINUED

Pericopsis.org	- Free - Visualize the plant species location - Able to export data in excel, KML, and SHP file - Allows contribution for the database - Interactive map	- Does not calculate the diversity index - No specific polygon to choose - Does not show the species information - Registration is required - Data needs to be entered manually

2.7 Rivers

The physical environment of aquatic biomes, often known as water bodies, is what defines them. Ecologists distinguish between freshwater and marine biomes based on physical and chemical distinctions; nevertheless, we will focus on freshwater biomes with salt concentrations less than 0.1 percent in this study. Salt concentrations in marine biomes typically range from 3% to 5% (Reece, et al., 2014).

Rivers are broad freshwater streams that cross land. Source: melting glacier, overflowing lake, or alpine spring. A river gets more water from streams, rivulets, and rainfall as it runs. Groundwater seeps into a riverbed. A river's journey finishes when it meets the sea. The river's mouth is where they meet. From little, shallow streams that dry up in the summer to enormous rivers that flow for thousands of kilometers.

Rivers are classified according to their origin and size. Rivers begin as headwaters, which are the highest of all river phases. The headwaters run down as small rills with only occasional rain. Rills unite to become brooks, which move quicker. Brooks flow into streams, which merge into the main river. The tributaries of a river are all brooks and streams that flow into it. They create the river system.

Rivers have always been essential to human settlement. A river distributes water to broad areas surrounding it, which is essential for plant and animal life. Man uses river water for agriculture and daily purposes like cooking and bathing. A river carries clay and other soil particles with it, which deposit in the riverbed, making it rich and productive. Rivers also contain plants, animals, and aquatic life that humans exploit for food and commerce (Society, 2012).

Unpolluted rivers and streams have a wide variety of fish and invertebrates dispersed vertically. Waterborne organisms feed mostly on terrestrial organic materials in temperate or tropical forests. Pollution from municipal, agricultural, and industrial sources degrades water quality and kills aquatic creatures in rivers and streams. Damming and flood control harm migratory animals such as salmon (Reece, et al., 2014).

2.8 Importance of Water Bodies in Malaysia

Malaysia, a developing country, faces several pollution threats. Heavy metals are among the principal contaminants that harm aquatic life. These elements are discharged into receiving systems like sediment, soil, and water. Metal dispersion into these receiving systems has been studied (Choe, et al., 2008) (Yang, et al., 2018). Industrialization and agriculture may contribute to their rise in the aquatic system (Tarras-Wahlberg, Flachier, Lane, & Sangfors, 2001). Some metals are required by biotics yet can be hazardous in high doses. Heavy metals are non-degradable and can harm plants, marine life, and human health (Soylak & Erdogan, 2006).

Particulates polluting aquatic habitats will create as-deposited contaminants. Clays, silicates, carbonate, sulphide, minerals, and microorganisms are all found in sediments. Sediment is a well-known metal contaminant sinks and source (Li & Thornton, 2001). Pollutants such as heavy metals are spread in sediment components and react through ion exchange, absorption, and precipitation (Yuan, et al., 2004). Heavy metals are absorbed by sediments in three ways: physiochemically, biologically, and through particle accumulation (Hart, 1982).

Fish occupies the highest trophic level in the aquatic system (APHA, 1989). It also has great economic values, making fish excellent for public water quality markers. Fish can detect environmental changes and act as pollution indicators. Indicator fish are easy to obtain in big quantities, have a long lifespan, are the ideal size for study, and are easily sampled (Kannan, et al., 2007). Heavy metal intake by fish in contaminated water varies based on biological requirements, metabolism, diet, and sediment. Metals enter fish tissues via absorption, and humans are exposed via the food web. Humans will suffer acute and persistent effects (Yılmaz & Doğan, 2008) (Fidan, et al., 2008). Thus, using fish as a bioindicator can help determine pollution levels before and after monitoring.

Heavy metal investigations have been undertaken in Malaysian lakes and rivers such as Bera Lake, Lake Chini, Kenyir Lake, and Rompin River (Shuhaimi-Othman, Lim, & Mushrifah, Water quality changes in Chini lake, Pahang, west Malaysia, 2007) (Shuhaimi-Othman, Mushrifah, Lim, & Ahmad, 2008) (Ebrahimpour & Mushrifah, 2009). Lake Chini is a popular tourist site with large wetland regions that might be considered unpolluted. This wetland can control water quality and quantity, maintain the ecological system, and serve as a habitat for a variety of flora and fauna species. Local aborigines caught a lot of fish from Lake Chini for their daily meal or to sell. As a result, research into metal concentrations in sediment and fish is critical in determining the present hazards to human health posed by heavy metal contamination.

2.9 Water Quality Index (WQI)

The water quality index (WQI) is an indicator of the quality of water. WQI relates a set of water quality parameters to a standard scale and combines them into a single number (Leščešen I, 2015). WQI is one of the best tools for evaluating the quality of water resources such as groundwater, lakes, and rivers (Anyachebelu, 2015). WQI is also a useful water quality monitoring tool to policymakers and environmentalists (Yogendra, 2008).

((Avvannavar, 2008) (Boyacioglu, 2007); (Gandaseca, 2011); (Said, 2004); (Sarkar, 2006)) have pointed out the existence of several versions of water quality indexes in literature. This is mainly due to the use of different numbers or a combination of water quality parameters for developing the indexes. (Meher, 2015) developed a water quality index for evaluating Gange's river water quality based on fourteen parameters (pH, electrical conductivity, dissolved oxygen, total dissolved solids, turbidity, salinity, major cations, major anions, and alkalinity). (Al-Shujairi, 2013) developed a water quality index for monitoring the water quality of Tigris and Euphrates rivers based only on seven parameters (total dissolved solids, total hardness, pH, dissolved oxygen, biological oxygen demand, nitrate, and phosphate).

The Department of Environment of Malaysia (DOE) developed the National Water Quality Index (NWQI) for assessing the water quality of Malaysian rivers. According to DOE, 36.6 % of rivers in Malaysia were classified as slightly polluted, and 5.2 % were classified as polluted in 2013. The NWQI index uses the following parameters: pH, dissolved oxygen (DO), chemical oxygen demand (COD), biochemical oxygen demand (BOD, total dissolved solids (TSS), and ammonical nitrogen (NH_3-N). The NWQI value ranges from 0 to 100 and is calculated based on sub-index values assigned to each parameter (DOE, 2014). A monitored river can be classified into one of the five classes: I, II, III, IV, and V (in descending order of water quality, where Class I is the "best" and Class V is the "worst" water quality). The mathematical expression for NSF WQI is given by:

$$WQI = \sum_{1=1}^{n} Q_i W_i \tag{2.4}$$

Qi = sub-index for ith water quality parameter;

Wi = weight associated with ith water quality parameter;

n = number of water quality parameters.

The table 2.6 below shows the National Water Quality Standards for Malaysia according to the Department of Environment, Malaysia (DOE, Malaysia environmental quality reports., 2014)

Table 2.6: National Water Quality Standards for Malaysia (DOE, 2014)

PARAMETER	UNIT	CLASS				
		I	II	III	IV	V
Ammoniacal Nitrogen	mg/l	< 0.1	0.1 – 0.3	0.3 – 0.9	0.9 – 2.7	> 2.7
Biochemical Oxygen Demand	mg/l	< 1	1 – 3	3 – 6	6 – 12	> 12
Chemical Oxygen Demand	mg/l	< 10	10 – 25	25 – 50	50 -100	> 100
Dissolved Oxygen	mg/l	> 7	5 – 7	3 – 5	1 – 3	< 1
pH	-	> 7	6 – 7	5 – 6	< 5	> 5
Total Suspenbed Solid	mg/l	< 25	25 – 50	50 – 150	150 – 300	> 300
Water Quality Index (WQI)	-	< 92.7	76.5 – 92.7	51.9 – 76.5	31.0 – 51.9	> 31.0

Table 2.7 shows the water classes and uses adapted from (DOE, Malaysia environmental quality reports., 2014).

Table 2.7: Water Classes and Uses adapted from (DOE, 2014)

CLASS	USES
Class I	Conservation of natural environment Water Supply I – Practically no treatment necessary. Fishery I – Very sensitive aquatic species.
Class IIA Class IIB	Water Supply II – Conventional treatment. Fishery II – Sensitive aquatic species. Recreational use body contact
Class III	Water Supply III – Extensive treatment required Fishery III – Common of economiv value and tolerant species; livestock drinking
Class IV	Irrigation
Class V	None of the above

(Azizi Abu Bakar, 2013) proposed a water quality monitoring and assessment tool called the Eco-Heart Index. The Eco-Heart Index was developed to visualise water quality by drawing a 'heart' shape using the NWQI parameters. A full heart stands for clean while a broken heart indicates polluted water. A full heart shape is formed if all the parameters are classified as clean and a broken heart will appear if any of the parameters are classified as polluted. However, it is difficult to decipher water quality based on heart shape alone due to its limited resolution. The use of parameters to construct that does not fit neatly into a particular water class may also produce a heart shape that is difficult to interpret.

2.10 Existing Water Quality Monitoring System

Because individual organizations or bodies maintain ecological databases on tropical lakes in Malaysia, they are considered scattered. Although a few worldwide lake databases offer internet services, they are insufficient and unsuitable for Malaysia. The next paragraphs will explore the advantages and disadvantages of the current arrangement.

The World Lake Database is a live database system maintained by the International Lake Environment Committee Foundation. It is a 1986 database that gathered environmental and socioeconomic data on lakes and reservoirs worldwide. The data collected has resulted in a project titled "Survey of the State of World Lakes," which was undertaken in collaboration with the United Nations Environment Programme (UNEP). This approach captured some of the most basic characteristics of the lakes. Because any user can add any type of information on the lake via uploader, the collected data are not categorized or normalized. Certain data are out of date (Kira, 1988).

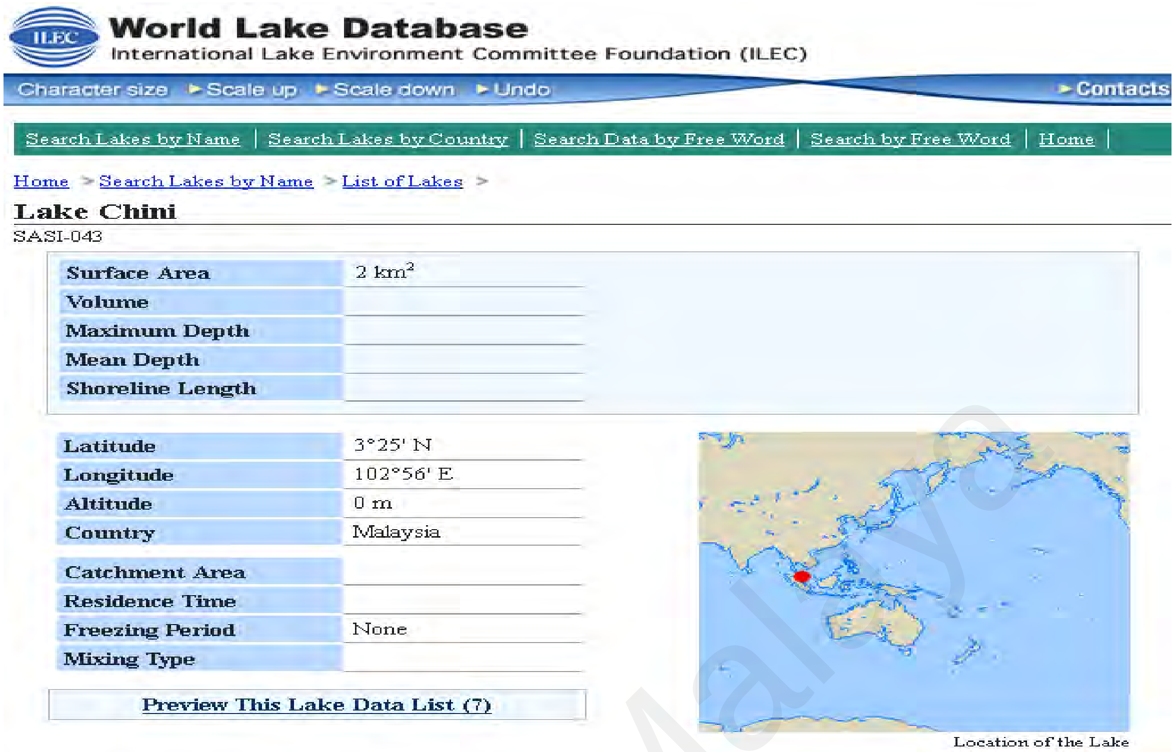


Figure 2.25: Information of Lake Chini located in Malaysia in the World Lake Database.

Figure 2.26: Data List of Lake Chini, Malaysia

LakeNet is a global information system dedicated to the conservation and sustainable development of lake ecosystems. LakeNet offers information services and exchanges online. Priority is on educating and empowering people to care for lakes and their ecosystems. They only offer basic lake data, with certain elements missing or incomplete (Seders, Shea, Lemmon, Maurice, & Talley, 2007).

Figure 2.27: LakeNet displaying the Lake Profile of Chini Lake(WorldLakes, 2021)

The Ramsar Convention is an intergovernmental agreement on natural resource conservation. Members of the Ramsar Convention select Ramsar sites that meet the criteria for Identifying Wetlands of International Importance. To join Ramsar sites simply complete the Ramsar Information Sheet. The Ramsar Information Sheet contains exact statistics, conservation parameters, and a map of the site's boundaries. These data do not include daily updates like physical, chemical, or biodiversity data. The lake data is out of date and hasn't been updated in years (Okuno, Gardner, & Beaulieu, 2015).

GIS and Global Database Data on 35000 natural lakes, 5000 reservoirs, and 220 wetlands. Geospatial data attributes include geography and chemistry. Lakes statistics in Malaysia are few and out of date (Ryanzhin, Straskraba, & Geller, 2001).

The National Hydraulic Research Institute of Malaysia has produced a National Lake Information Database for all lakes in Malaysia, natural or man-made. This database's goal is to provide useful data for effective and sustainable lake management in Malaysia. Anyone with information about any Malaysian lake may register as a "Contributor" to the Database. The Database organizes each lake's data into three topics. They are (a) lake summary data, (b) lake assessment data, and (c) lake water quality data.

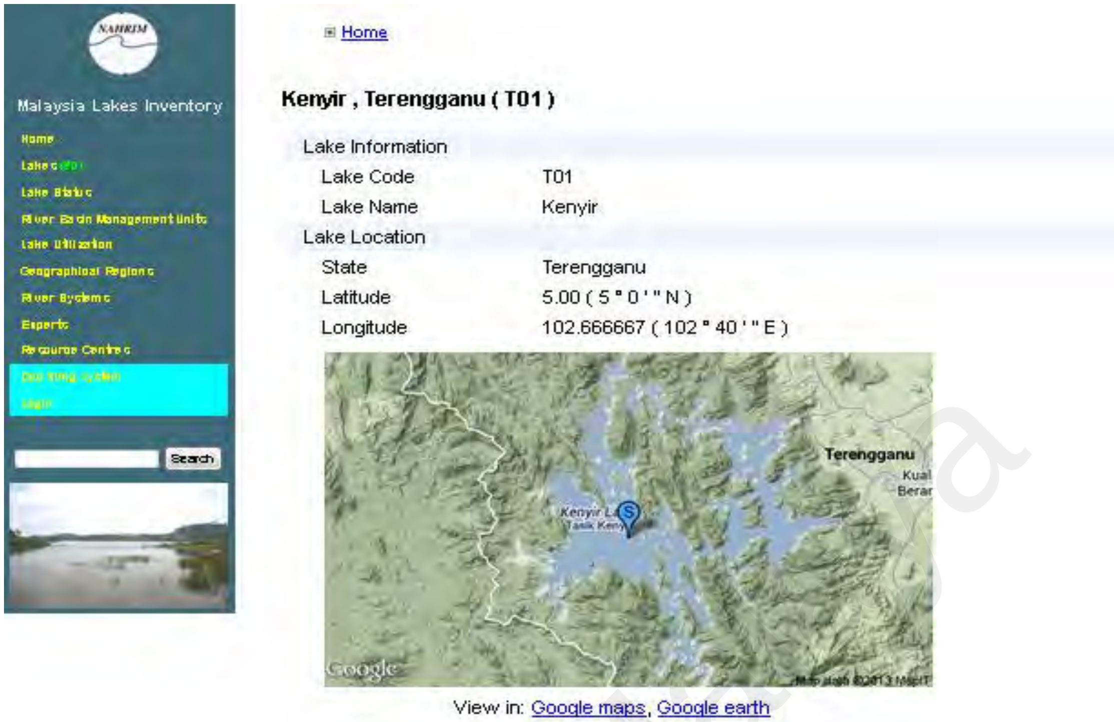


Figure 2.28: Lake Kenyir, Terengganu Lake information obtained from The National Lake Information Database of Malaysia. (NAHRIM, 2021)

The PLWMOS system is designed to be the major environmental database and spatial analysis tool for the current survey of Putrajaya Lake and wetlands parameters. The proposed PLWMOS system will allow for data entry, environmental modeling, and online report submission, and will be linked to the existing CMC system (or systems) (e.g. EMMS). PLWMOS will also be a knowledge-based decision support system meant to alert to potential environmental hazards. It saved hydrological, chemistry, and biotics data. The data and information of the system are not shared with the public and it only records 1 lake of data (Abdullah & Shobri, 2018).

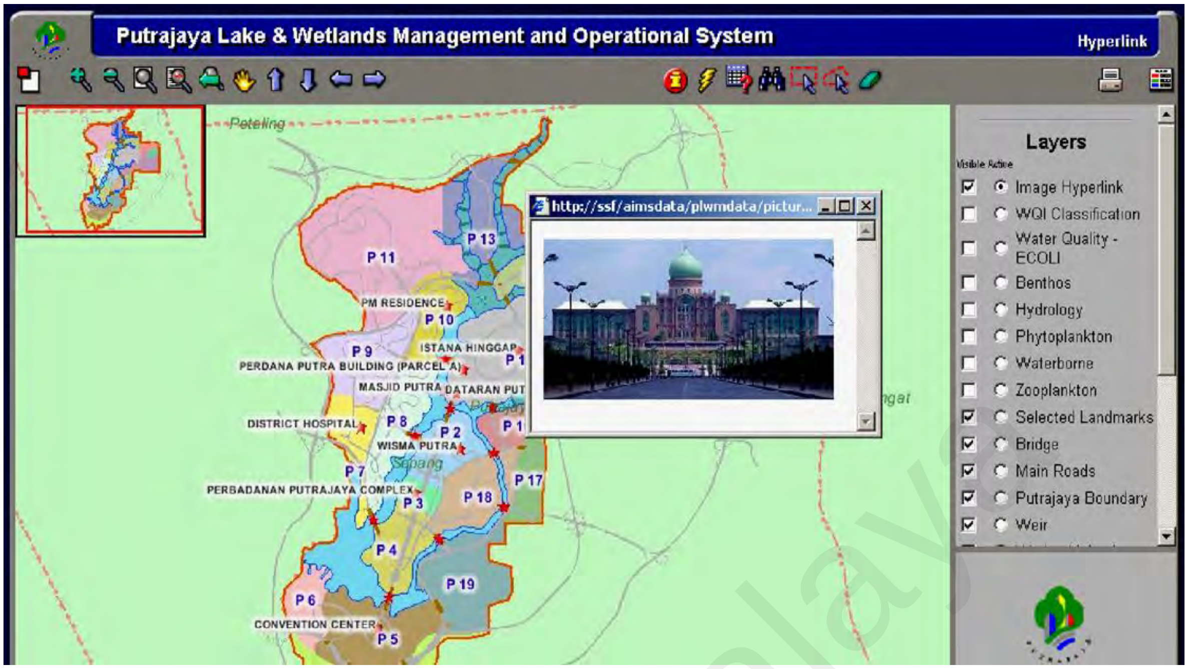


Figure 2.29: A screenshot of the PLWMOS System (Abdullah & Shobri, 2018).

Table 2.8 summarizes the water quality monitoring systems available online and some are used in the local water bodies.

Table 2.8: Summary of Water Quality Monitoring System

No	Databases	Main Roles	Comments	Reference(s)
1	LakeNet	Dedicated to establishing lake organizations, educating and inspiring people to stewardship of lake environments.	The database only contains fundamental data of lakes; contains incomplete information on Malaysian lakes.	(Ryanzhin, Straskraba, & Geller, 2001)

Table 2.8, CONTINUED

2	Ramsar Convention	Conservation and sustainable use of wetlands.	Does not store daily data; data on Tasik Bera (Malaysia) was last updated in July 1998	(Okuno, Gardner, & Beaulieu, 2015)
3	Global Database and GIS Worldlakes	Storing data on natural lakes, reservoirs, and wetlands.	Data on geography, chemistry, hydrology, ad others; only brief and out-of-date data for lakes in Malaysia	(Ryanzhin, Straskraba, & Geller, 2001)
4	National Lake Information Database of Malaysia (NAHRIM)	The database provides vital information to help effective and sustainable management of all Malaysian lakes, whether natural or man-made.	Doe does not enable information sharing and exchange and does not cover all the data available lakes in Malaysia	(Kira, 1988)

Table 2.8, CONTINUED

5	Putrajaya Lake and Wetland Management and Operational System (PLWMOS)	It is the primary environmental database and spatial analysis tool for the current study of Putrajaya Lake and wetlands hydrological, physicochemical, and biotic parameters; it is also developed as a knowledge-based decision support system to notify potential environmental problems.	Data and information of the system are not for public sharing and it caters only to Putrajaya Lake.	(Abdullah & Shobri, 2018)

2.11 Geographic Information System (GIS)

Geographic Information System (GIS) is known as a concept and slowly evolved into a science. The earliest definition is provided by (Burrough, 1986), who stated that GIS is a set of tools for collecting, storing, retrieving at will, transforming, and displaying spatial data from the real world for a set of purposes.

(Rhind, 1989) defined GIS as 'a computer system capable of storing and utilizing data representing physical locations on the Earth's surface. The Department of the Environment (1987) defined geospatial data management as 'a system for recording, storing, verifying, integrating, manipulating, analyzing, and displaying geographically referenced data'.

According to (Chrisman, 1999), he confined that most of the definition regarding GIS has serious drawbacks, and redefined GIS as the 'organized activity by which people measure and represent geographic phenomena then transform these representations into other forms while interacting with social structures.

Environment System Research Institute (ESRI) explained that GIS is a framework for gathering, managing, and analyzing data. Mainly incorporated the science of geography, by analyzing the spatial location and rearrange the data and visualize the information using maps. Hence, GIS is a unique tool that can discover patterns and relationships, giving further insights that can solve real-world problems (Mitchell & Minami, 1999).

In concise, GIS has many advantages in solving complex problems, especially when solving problems related to geography. Besides, a picture worth a thousand words, and a map tells a thousand pictures', GIS allows us to monitor changes easily and effectively, for example, the rate of deforestation, the glacier retreat, etc. GIS also gives real-time situation awareness, such as tsunami, earthquakes, hurricanes, natural disasters. GIS is also applied in prediction and forecasting, and understanding the latest trends and patterns, which might be overlooked in an excel file (Mitchell & Minami, 1999).

(Maguire, 1991) has published a paper regarding an overview and summary of GIS. He stated that GIS is difficult to define due to the rapid advancement of GIS technology. The image below shows a separate idea of various fields and GIS is the interdisciplinary field.

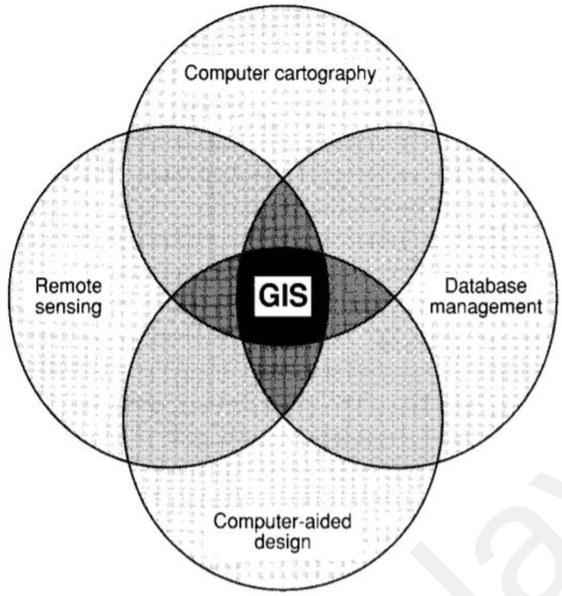


Figure 2.30: The relationship of GIS between computer-aided design, computer cartography, database management, and remote sensing information systems. (Maguire, 1991).

2.11.1 Google Maps

Google Maps is written in C++ and launched on February 8th, 2005, and is created by Jens Eilstrup Rasmussen and Lars Rasmussen (About US, Google, 2005). It is a web mapping service that provides satellite imagery, street maps, real-time traffic conditions, route suggestions, and many more functions that are explained in detail (Google Company, 2016).

After launching the desktop program, Google Maps also offers an application program interface (API) using JavaScript, so that web developers can embed maps, geocoding, places, markers, and polygons on third-party websites or applications and re-imagined the map by having a new default marker, new floating window and a brand-new style of the controls. (Google, 2013)

2.11.2 Google Earth Pro

The Google Earth Pro is a virtual globe software system which initially is released on June 11, 2001. It is a convenient and popular geographic information tool that is free and available online. Many scientists, especially GIS specialists, environmental professionals, and landscape planners use this software to visualize the environmental data due to its special features' realistic imagery of places. (Sheppard & Cizek, 2009)

Google Earth Pro has included several features that allow users to view and annotate maps, to create KML files and map layers, publicly display Content with proper attribution online, in the video, and print; and do many other things described in the Using Google Maps, Google Earth, and Street View permissions page.

By using Google Earth Pro, we can create maps with advanced tools such as, computing distances and areas using measurement tools, visualize, manipulating and exporting GIS data in KML or KMZ files and manipulate and export GIS data. Most importantly it is useful for conservation projects that require the best science and data available to design, implement, measure, and monitor the success of conservation actions.

Figure 2.31 illustrates a snapshot of Earth Pro on Desktops using Google Earth Pro, tools are provided to draw the polygon and save it as a KML file that will be used in the development of the Biodiversity Web Module.

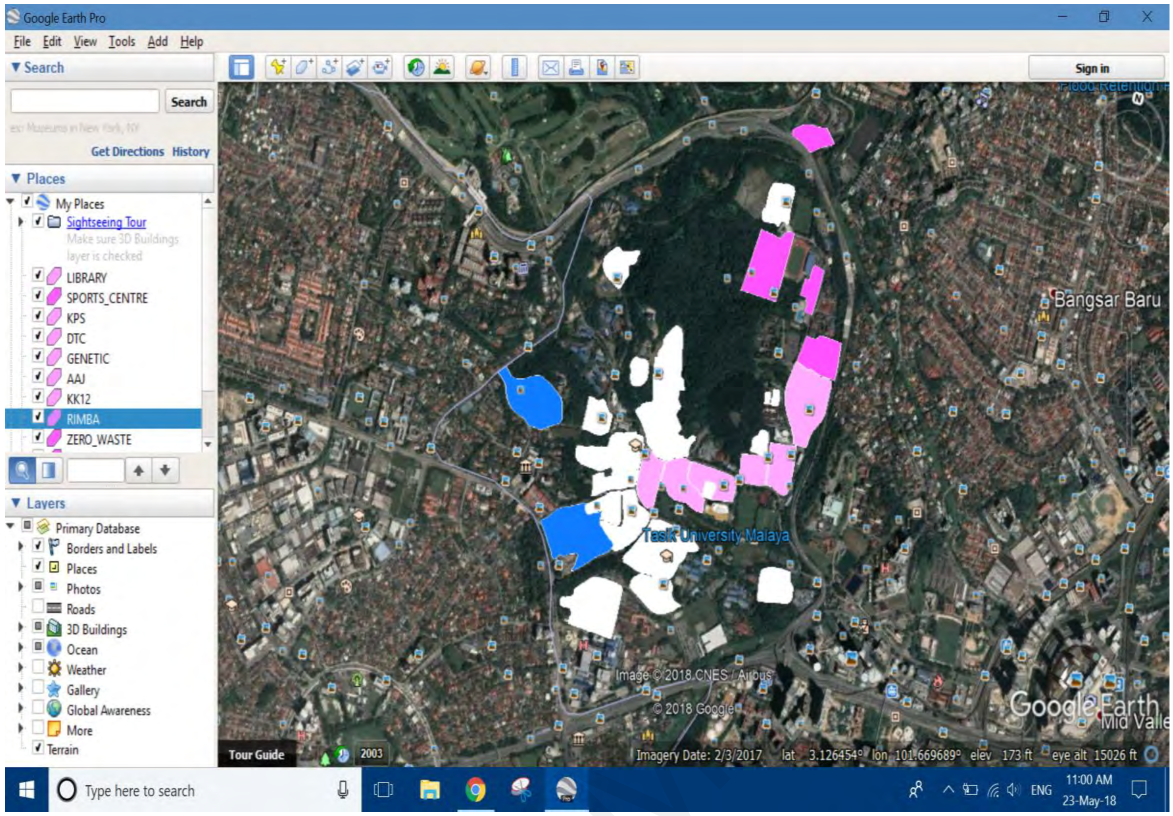


Figure 2.31: Screenshot of Google Earth Pro.

2.12 Importance of Visualization

Data visualization is an essential tool for effective communication and interpretation of information. Visualization of information in terms of the image is easier as compared to numerical values. Furthermore, the vast availability of data requires effective and efficient ways to access and communicate information (De Vries, 2010). Visualization is essential for data analysis and data representation (Xu, et al., 2010).

The emergence of Geographic Information Technology (GIS) has brought a huge advantage in managing and planning in conserving biodiversity through visualization, scientists can understand the topology and plant behavior. GIS can analyze the relationship between the plant traits and the local gradients for given communities, therefore, an explanation can be made by understanding how plants species are distributed in a community (Kattge, et al., 2011). Many scholarly works have proven the benefit of GIS and spatial analysis in understanding plant topology, ecology, and the correlation to the environment.

The reason why GIS plays a vital role in conserving biodiversity is that GIS provides maps of the distribution of important plant species, especially rare species and medicinal plants. Furthermore, GIS is also very useful in recognizing biodiversity hotspots, since conservation biologists are able to study the spatial distribution of plant species, and rare species are easy to be identified. Thus, we can identify the areas which are rich in biodiversity and should be claimed as protected habitat and preserve important genetic resources (Attorre, et al., 2011).

The application of GIS technology has become a useful tool for conserving wildlife. GIS provides a platform that connects land, people, and biodiversity. The founder of the Jane Goodall Institute, also rewarded as UN Messenger of Peace has developed one of the largest GIS animal behavior databases. The institute used the integrated data to monitor any deforestation activities and came out with a conservation plan to save chimpanzees outside the protected areas (Gilby, Eberly, Pintea, & Pusey, 2006). Hence, GIS able researchers to analyze cause and effects readily because maps created are from various discipline, and using GIS, the data can layer onto a single map, eventually, it can be used to protect wildlife and biodiversity threatened areas.

In a study of determining and mapping vegetation using the GIS approach conducted by (Usali & Ismail, 2010), the study proves that the GIS technique is useful in developing a tree mapping system and creating a geodatabase for spatial analysis. Whereby, it is recommended to integrate more data into the system for better evaluation.

In conclusion, GIS contributes to the majority of conservation biology plans by displaying and evaluating species topology and distribution, including rare species, native plant distribution, and wildlife population. Additionally, it can propose construction sites to real estate developers to safeguard biodiversity-rich environments. Finally, users may discover biodiversity hotspots and track conservation initiatives through time, allowing for a more balanced, sustainable, and healthier lifestyle.

In this study, plant species data and polygon files are mapped using GIS software and displayed on the map as a layer. This permits visualization of the plant species alongside the polygon, allowing for easy identification of plant species in a given area. Additionally, the diversity index for the specified area is calculated automatically. In terms of hydrological visualization, the map is shaped like a 'heart.'

CHAPTER 3: METHODOLOGY

3.1 Web-Based System for Visualization of Biodiversity Index

3.1.1 Study Site and Sampling Data

Our study area is the University of Malaya (UM), an urban university located in the heart of Kuala Lumpur. The University of Malaya's Environmental Secretariat initiated the RIMBA Project to protect and promote the university's biodiversity, conserve the campus's urban biodiversity and increase campus awareness (Malaya 2013). The RIMBA project is closely related to our research because they regularly host events that highlight the campus's diversity of biodiversity, including increasing green cover through the planting of diverse rainforest trees, raising seedlings of native trees, tree tagging, and several workshops on communication, writing, and media. As a result of the RIMBA strategy for conservation, the RIMBA approach to conservation has created many opportunities for volunteers through the use of citizen science software iNaturalist which contributed and give suggestions in our research study.

The University of Malaya is located in Kuala Lumpur, Malaysia, at coordinates of 3°07′15″N 101°39′23″E. The study area's principal focus is depicted in figure 3.1.

Figure 3.1: The location of the University of Malaya in Google Earth.

The species data for the University Malaya are imported from the site – iNaturalist (https://www.inaturalist.org/). Google Earth is used to identify green space area data within the university. Before integrating data from iNaturalist, the area is selected as a polygon and uploaded into the system.

iNaturalist data include the continuous integration of field images and an array of metadata associated with observations, as well as the community-driven process for taxonomic identification and record validation. In this study, two types of iNaturalist data were used: data with a Casual ID and data with a Research ID. Users can select the type of data they want to use. When Research ID is selected, the data is verified by experts and is more accurate.

iNaturalist adheres to a set of well-established taxonomic authorities, which are updated regularly by knowledgeable users ("Curators"). However, any user may add placeholder taxon names that iNaturalist does not presently recognize. There are numerous connections between these digital collections created and maintained by the iNaturalist community and actual herbaria collections maintained by botanical professionals. These photographic records are essentially digital specimens that lack reference material in the form of physical vouchers. Given the high degree of overlap between data associated with iNaturalist observation-based records and those associated with specimen-based records, iNaturalist is well-suited as a data collection tool for field plant collectors. The data types acquired by iNaturalist for the University of Malaya are presented in table 3.1.

Table 3.1: Data Type in iNaturalist

Data in iNaturalist	Description	Data Quality
Casual Grade	All observations start as a “Casual” grade. The data might be missing one or more criteria, which are the data does not contain the date where the species was observed, is not georeferenced, without photos or sound, and is a captive and cultivated organism.	Low. The observations might miss some of the data, which affects the quality.
Needs ID	When the observation has a date, it is georeferenced, contains photos or sounds of the species, and is not of a captive or cultivated organism. Verifiable observations are labeled as this grade until either the observation attain as “Research” grade status or are voted to “Casual” via the Data Quality Assessment.	Moderate. Requires verification from the identifiers. Verifiable observations

Table 3.1, CONTINUED

Research Grade	The observations meet the criteria from the "Needs ID" grade and the observations are recognized by the community which agrees on the species-level ID or lower. E.g., when more than 2/3 of identifiers agree on a taxon.	High. Accurate and well recognized by the community.

The Data Quality Assessment summarises the accuracy, completeness, and suitability of observation for sharing with data partners. The verifiable observation is the foundation of iNaturalist. A verifiable observation has a date, is geolocated, contains photos or sounds, and is not of a captive or cultivated organism. Table 3.2 shows the data characteristics for the study site imported from iNaturalist. We chose to demonstrate using research-grade qualification data in this study.

Table 3.2: iNaturalist Data Quality Assessment

Data Quality Assessment	Casual Grade Qualification	Research Grade Qualification
Date specified	✓	✓
Location specified	✓	✓
Has Photos or sound	Maybe	✓
Has ID supported by 2 or more		✓
Date is accurate		✓

Table 3.2, CONTINUED

Location is accurate		✓
Organism is wild		✓
Evidence of organism		✓
Recent evidence of an organism		✓
Community ID at Species-level or lower		✓

3.1.2 Green Space Area representation

The polygon for the University of Malaya for green space area is drawn using Google Earth Pro and Geographic Information System (GIS) mapping technology.

Each polygon drawn is saved as Keyhole Markup Language (KML) file, since Google Earth renders a KML file for displaying the virtual earth browser. The KML file is a grammar of XML for marking up spatial data, thus the coordinates could be extracted from the file and create interactive mash-ups using HTML and JavaScript (Nolan D., Land D.T., 2014). The figure 3.2 below shows the polygon drawn within the campus.

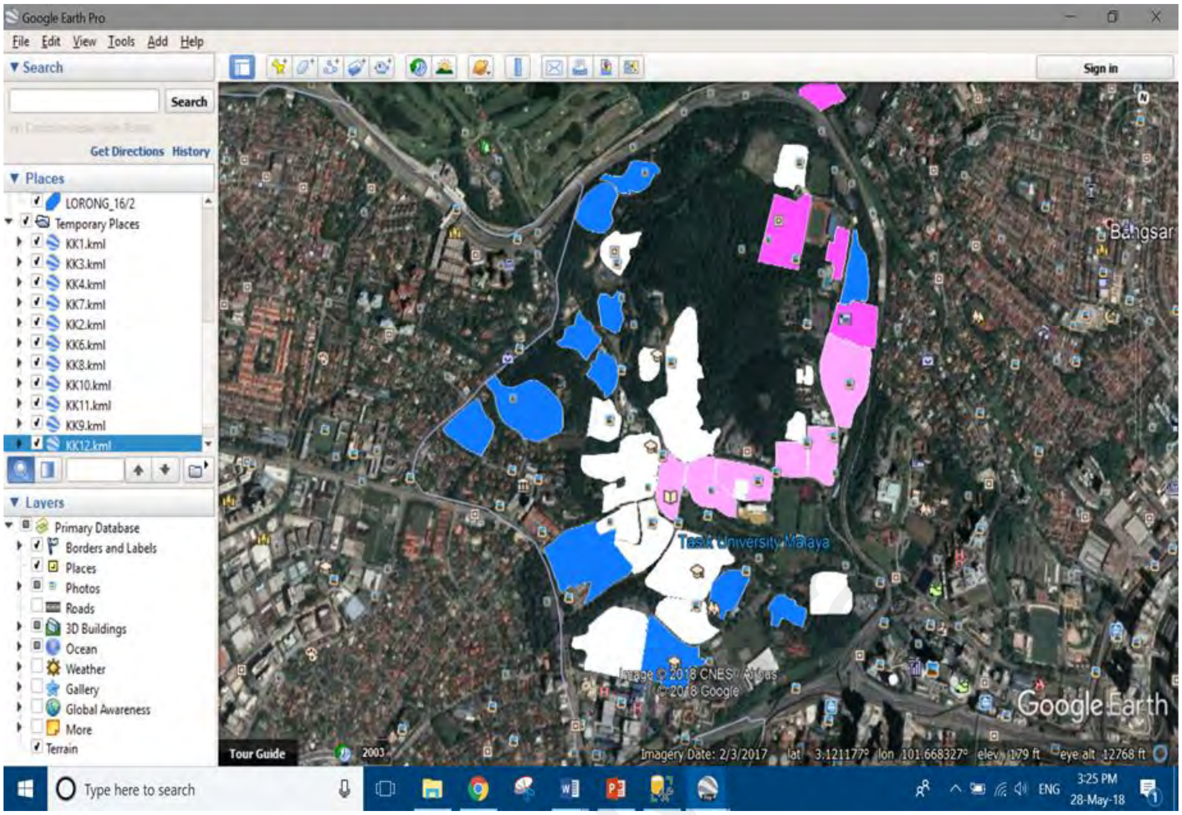


Figure 3.2: Polygon files (KML files) are drawn using Google Earth Pro.

3.1.3 Biodiversity Indexes Calculation

Biodiversity indexes such as species richness, Shannon-Wiener Index, Simpson's Index, and Hill numbers are used in the study as these indexes are common measurements for species diversity in a community. These indexes are commonly used in ecology (Buddle, et al., 2005). According to Colwell and Robert K. (2009), these indexes are mathematical functions that take account of species richness and evenness and calculate in a single measure.

(a) Species Richness

Species richness (S) is the number of species within a defined region (Moore J. C., 2013). According to (Gotelli & Colwell, 2001), community ecologists and conservation biologists regarded measuring species richness as a fundamental objective where the number of species in a local aggregation is an instinctive and natural index of a community structure. Measuring species richness and diversity using a geographic approach ables to determine what species are at risk of extinction by having the species geographic coordinates, we can assemble, store, retrieve and manipulate data on the species distribution and the geographical pattern (Scott, Csuti, Jacobi, & Estes, 1987)

$$N = Number\ of\ Species$$

(b) Shannon-Wiener Index (H')

Spellerberg and Fedor (2003) researched the use of species indexes, specifically the 'Shannon-Wiener Index.' The Shannon-Wiener index accounts for both the abundance and evenness of the species present. It is an information statistic index; it represents all the species in the community and is commonly used to characterize species diversity in a community.

$$H = -\sum_{i=1}^{s} P_i \ln P_i \tag{3.1}$$

The value H increases, when the number of species in the sample increases. Normally this index is used to compare the diversity of a series of samples. Hence, a higher value of H indicates high species diversity in the sample (Magurran, 2004). The Shannon-Wiener index is useful to determine whether an area consists of valuable and rare species. Besides it is sensitive to changes and relatively easy to measure.

(c) Simpson's Index (D)

Simpson (1949) developed the Simpson's diversity index, which describes the likelihood that a second individual drawn from a population is of the same species as the first. Simpson's index is known as a "dominance concentration" measure because it is sensitive to the abundance of only the more abundant species in a sample (Whittaker, 1965). Unlike the Shannon-Wiener index, Simpson's diversity index is sensitive and gives more weight to the most common and dominant species, while rare species do not affect the calculation.

$$D = 1 - \sum_{i=1}^{s} p_i^2 \qquad (3.2)$$

The value of D has the maximum value of 1, the higher the value, the higher the diversity.

(d) Hill Numbers

Hill numbers are a combination of taxonomic, functional, and phylogenetic diversity indexes (Hill 1973). Hill numbers can be used to quantify and compute the effective number of species using the following formula (Clarke et al. 2014):

$$_0D = S$$

$$_1D = \exp(H')$$

$$_2D = \frac{1}{\sum p^2} \qquad (3.3)$$

$$_nD = \frac{1}{max\{p_i\}}$$

$_0D = S$ where often implies as the total number of species, $_1$D is the exponential of Shannon index. As for $_2$D is the reciprocal of Simpson's index and if $_n$D is another possible evenness index (Clarke, Gorley, Somerfield, & & Warwick, 2014). Biodiversity indexes used in this study are summarized in table 3.3.

Table 3.3: Summary of Biodiversity Indexes applied in the System.

Biodiversity indexes	Proposed / Introduced by	Type of Index	Formula	Descriptions
Number of Species (S)	NJ Gotelli, RK Colwell (2001)	Species Richness	N = number of species	The number of species in each community.
Shannon-Wiener Index	Claude Shannon (1949)	Abundance-based diversity index	$H = -\sum_{i=1}^{s} p_i \ln p_i$ p = Proportion that the species forms of the total number of organisms found in the community	Characterize species diversity in a community. When H increases, indicated a more diverse community. Rarely exceeds 4 in practice.

Table 3.3, CONTINUED

Simpson's Diversity Index	Edward H. Simpson (1949)	Dominance-based diversity index	$D = 1 - \sum_{i=1}^{s} p_i^2$ p = Proportion that the species forms of the total number of organisms found in the community	Probability of finding a second individual in a sample will the same as the first. D value ranges from 0 to 1. The closer the value to 1, indicates the more diverse the community.

Table 3.3, CONTINUED

Hill Numbers	Hill, M. O. (1973).	The effective number of species	${}_0D = S$ ${}_1D = \exp(H')$ ${}_2D = \frac{1}{\sum p^2}$ ${}_nD = \frac{1}{max\{p_i\}}$ ${}_0D = S$,the total number of species. ${}_1$D is the exponential of Shannon index. ${}_2$D is the reciprocal of Simpson's index. ${}_n$D is another possible evenness index. (Clarke, Gorley, Somerfield, & & Warwick, 2014).	Estimates of the effective number of species present. Incorporate relative abundance and species richness.

3.1.4 Development of the Web-Based Biodiversity Indexes Visualization System

Visual Studio 2017 was used to develop a web-based system for measuring and visualizing biodiversity indexes. The system is developed using the C# and ASP.net programming languages. C #'s advantage is its ease of integration with components written in other languages and with a large number of Microsoft's proprietary technologies. It also runs on a variety of platforms, including Google Earth. C# has access to the.NET Framework's class libraries, which is necessary for constructing the system's graphical representation (Microsoft. n.d.). Additionally, C# is built on a network-distributed infrastructure. The.NET platform, is responsible for managing the code environment that enables C# and VB.net applications.

We also used, Hypertext Markup Language (HTML) to create web pages and web applications. HTML can embed scripting language programs, such as JavaScript, that provide the functionality of the webpage. CSS is used to design the presentation of an HTML page, including the designs, layouts, colors, and fonts.

Google Maps API V3.27 (Developers 2015) was used to view the graphical representation of species in real-time on the Google Earth platform. The Google Maps API is required for the system to display the map, and it allows for customization by displaying polygons and markers on the maps. The data for the species and diversity indexes in this study were stored in MySQL. My SQL 8.0 is a free, open-source database management system that is ideal for websites due to its high-performance query engine, fast data insert capability, and strong support for specialized web functions such as quick full-text searches (MySQL 2016). As a result, we used MySQL to store polygon and species information in the developed system.

iNaturalist is a global community of naturalists, scientists, and enthusiasts that share over a million wildlife sightings to educate and conserve the natural world. The California Academy of Sciences administers the iNaturalist technology infrastructure and open-source software to explore, explain, and preserve life on Earth (Matheson 2014).

The REST API is used to obtain specific information from iNaturalist. REST API is an architectural style for an application progr interface that uses HTTP requests to access and use data. It is used because the data can be used for a variety of functions such as GET, PUT, POST, and DELETE data types (Masse 2011), allowing the system to communicate with the iNaturalist platform and request species data from them. Furthermore, REST API consumes less bandwidth and supports a variety of data formats. In this study, we used JSON data format to 'GET' the species name, scientific name, location coordinates, observer name, and date observed (Gillis 2020).

The iNaturalist API is a collection of REST endpoints that can be used to read data from and write data back to iNaturalist on behalf of users. By appending, data can be retrieved in a variety of formats. [format] to the endpoints, such as "/observation.json" to retrieve biodiversity observation data. In our study, we used the GET/observation fields API to obtain species information, such as species location and taxon, which does not require iNaturalist authentication because we are not accessing data such as unobscured coordinates on behalf of users or writing data to iNaturalist, which requires an authenticated request (iNaturalist 2020). JSON format typically contains more information, according to iNaturalist (2020), because iNaturalist uses JSON responses internally.

Table 3.4 depicts the iNaturalist API, which enables us to obtain the necessary observation data from the iNaturalist database in JSON format, and the data acquired is updated daily and incorporated into our biodiversity visualization system. Our research is focused on Kuala Lumpur, Malaysia, and the data collected will be centered on place ids with the integer value of "97102," which represents the codes for the specific location.

Table 3.4: iNaturalist API used in the Biodiversity Visualization System Prototype

API	Usage of API	Format	Parameters	Allowed Values	Acquired Data
GET/observation	The primary endpoint for retrieving observations	JSON	place_ids ("97102" – Kuala Lumpur ID)	Integer	- Location (latitude, longitude) - Taxon

With regards to the green space area, users can upload the region they are interested in via the system using the KML file format. Then, users can build a graphical visualization of the green space region on Google Maps, with the system retrieving the coordinates based on the data to be displayed. Using species and green space area data, the system can calculate species richness, Shannon-Wiener index, Simpson index, and Hill numbers and visualize them on a Google Map.

The species data will be retrieved from the iNaturalist website via the biodiversity system that has been developed in this study. It is the iNaturalist user that collects the species data, and the biodiversity system will extract the required data from iNaturalist daily via the REST API. When the data has been retrieved, the system will process it and provide a variety of reports for the system end-user based on their specific requirements. The architecture and flowchart of the system are depicted in figures 3.3 and figure 3.4.

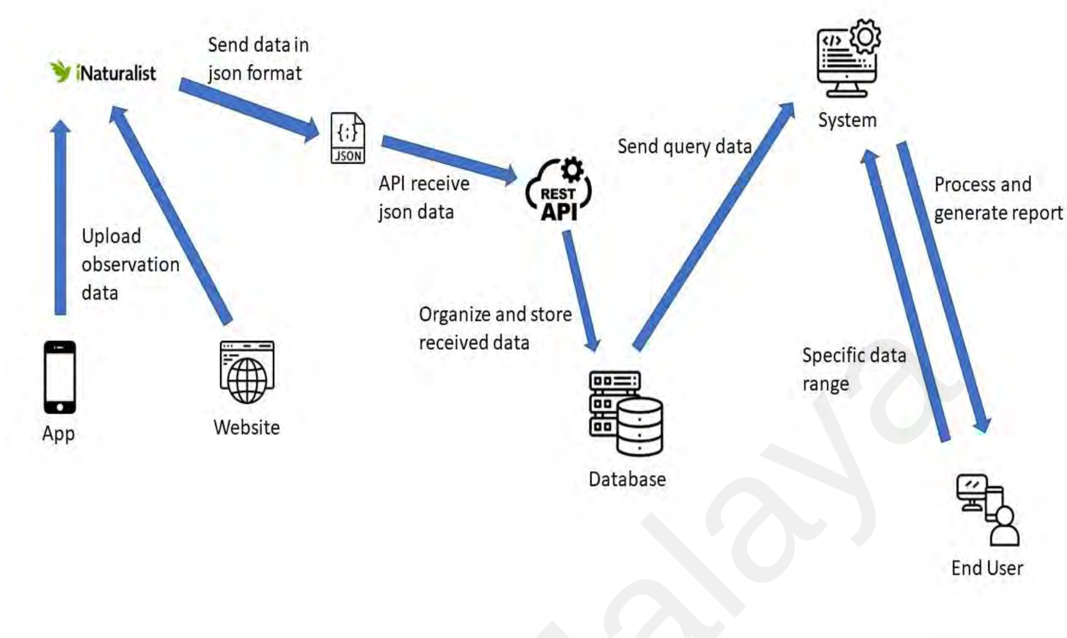


Figure 3.3: Model Architecture for the Biodiversity Visualization Module

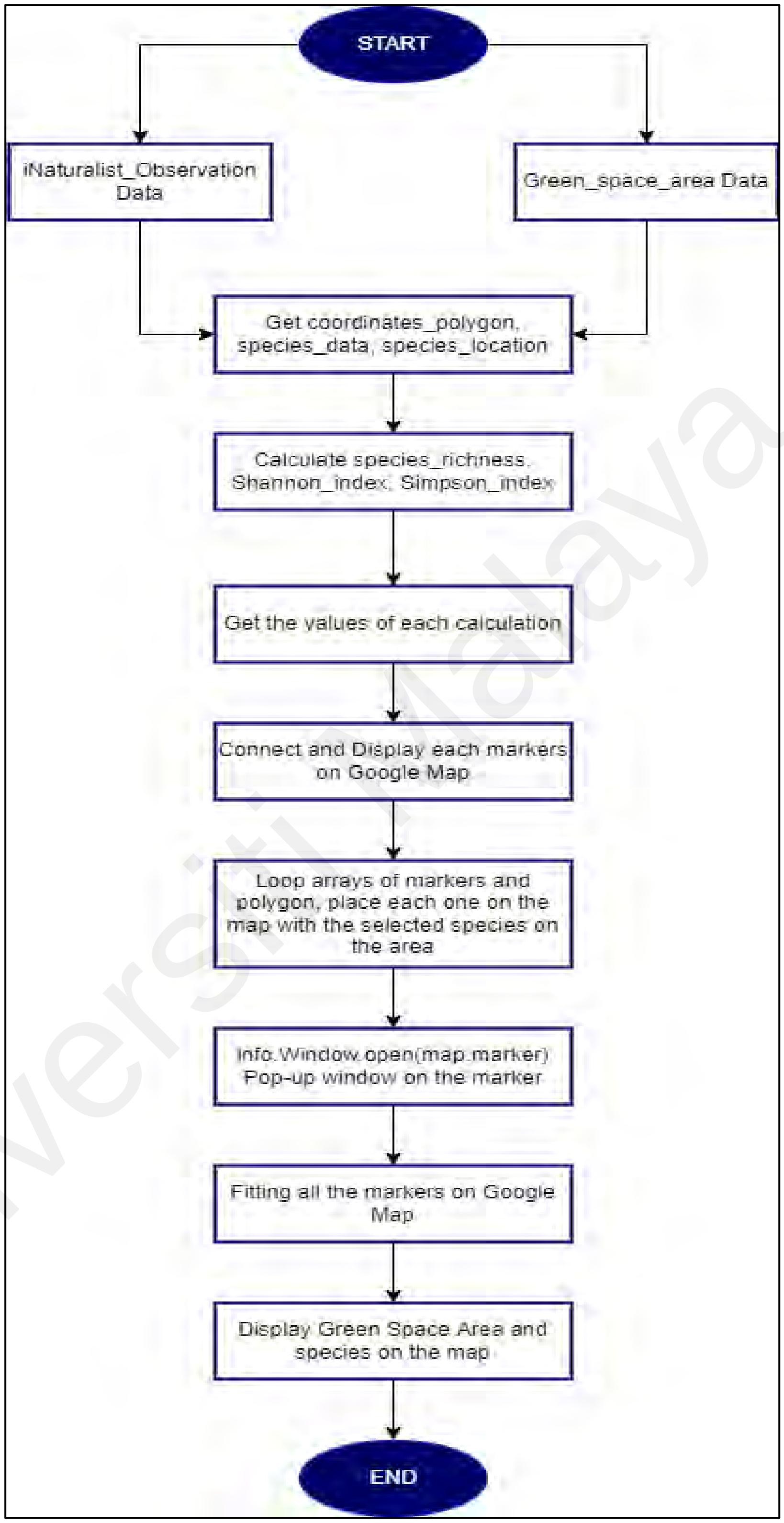


Figure 3.4: Biodiversity Indexes Visualization System Flowchart.

3.2 Web Based System for Visualization of Water Quality Index

3.2.1 Study Site and Sampling Data

The Langat River has a total catchment area of approximately 1,815 km^2. It lies within latitudes 2◦40 M 152 N to 3◦16 M 15 and longitudes 101◦19 M 20 E to 102◦1 M 10 E. The catchment is illustrated in Figure 1. The Langat River Basin comprises of 15 sub-basins as follows; Pangsoon, Hulu Lui, Hulu Langat, Cheras, Kajang, Putrajaya, Hulu Semenyih, Semenyih, Batang Benar, Batang Labu, Beranang, Bangi Lama, Rinching, Teluk Datok, and Teluk Panglima Garang. The main river course length is about 141 km and is situated around 40 km east of Kuala Lumpur. The water quality data in this study were obtained from seven stations along the main Langat River, as shown in figure 3.5. The water quality monitoring stations are determined by the Department of Environment (DOE), Ministry of Natural Resource and Environment of Malaysia. The selected stations (sampling points) are illustrated in table 3.5. All the stations were identified based on the availability of recorded data from 2005 to 2015.

Table 3.5: Sampling Points of the Langat River.

DOE Sampling Station	Longitude	Latitude	Location
1L01	02°51’896’’	101°26’251’’	Air Tawar Village
1L02	02°48’942’’	101°30’717’’	Telok Datuk, near Banting town
1L03	02°51’245’’	101°40’868’’	Bridge at Dengkil Village
1L05	02°59’523’’	101°47’151’’	Kajang Bridge
1L15	03°02’461’’	101°46’386’’	Balakong

Table 3.5, CONTINUED

1L16	03°06’675’’	101°49’005’’	Hulu Langat
1L07	03°10’045’’	101°51’034’’	Bridge at Batu 18

Figure 3.5: Sampling Point of Langat River

A total of 712 samples were used for the data analysis in this study. The six selected water quality parameters were DO, BOD, COD, AN, pH, and SS to measure the data set of 10-year as summarized in Table 3.6.

Table 3.6: Summary Statistics of Data from Langat River from 2005 - 2010.

Station	**1L01**		**1L02**		**1L03**		**1L05**		**1L15**		**1L16**		**1L07**	
Parameter	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
DO (mg/L)	1.02	0.67	4.28	1.2	5.81	0.75	5.46	1.18	6.51	1.25	7.87	0.55	8	0.55
BOD (mg/L)	4.36	1.33	5.9	3.75	5.44	2.64	8.33	8.33	7.87	6.46	2.81	1.82	239	1.81
COD (mg/L)	29.55	16.63	26.13	12.68	24.6	11.76	32.02	15.51	31.55	20.48	16.56	9.1	14.15	8.24
SS (mg/L)	159.4	194.18	120.94	114.59	177.95	197.35	284.27	357.11	340.05	499.24	43.55	49.75	7.82	9.18
pH	6.46	0.69	6.58	0.72	7	0.33	7.04	7.04	7.08	0.33	7.29	0.33	7.37	0.2
$NH^{3}NL$ (mg/L)	0.58	0.45	1.02	0.67	1.18	1.06	1.68	1.31	1.58	1.34	0.17	0.34	0.09	0.11

3.2.2 Water Quality Index Representation Development

The WQI used in this system is based on NWQI parameter value by (DOE) and employed the concept of graphical representation proposed by (Azizi Abu Bakar 2013). Figure 3.6 represents the heart-based water quality index concept. The heart shape was developed to visualize water quality by drawing a 'heart' shape using the six parameters value based on NWQI, which are monitored and marked their levels by classification data as shown in Table 3.7. The six marks are connected by a curve line in order, and water quality is evaluated based on the result drawn figure. If all the parameters are classified as clean (i.e., level I), a full heart shape appears. In contrast, if water is polluted and some of the parameters are not classified as clean (i.e., level II, III, IV, or V), a broken heart shape would appear.

Table 3.7: Classification of Water Quality NWQI Malaysia

Parameter	**Unit**	**I (clean)**	**II (moderate)**	**III (slightly polluted)**	**IV (polluted)**	**V (heavily polluted)**
pH	-	>7.0	6.00-7.00	5.00-6.00	<5	>5
DO	mg/L	>6.0	5.01-6.00	3.01-5.00	1.01-3.00	0-1.00
SS	mg/L	<25	25-50	50-150	150-300	>300.0
BOD	mg/L	>1	1-3	3-6	6-12	>12.0
NH^3-N	mg/L	0-0.50	0.51-1.00	1.01-2.00	2.01-5.0	>5.00
COD	mg/L	0-5	6-10	11-13	14-20	>20.00
WQI	-	<92.7	76.5-92.7	51.9-76.5	31.0-51.9	>31.0

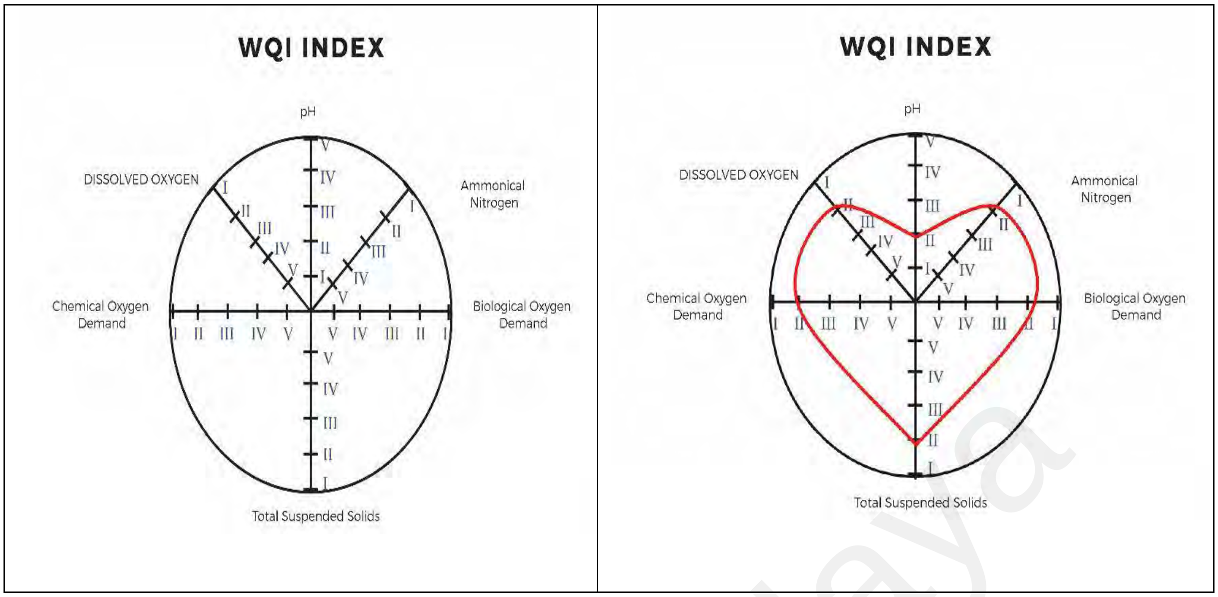


Figure 3.6: Visualization of the WQI Index proposed by (Azizi, Nik Meriem, Noor Zalina, & Azizan, 2015)

3.2.3 WQI Calculations

The six selected water quality index variables used in this study are suspended solids (SS), biochemical oxygen demand (BOD), ammoniacal nitrogen (AN), chemical oxygen demand (COD), dissolved oxygen (DO), and pH in accordance with WQI value calculation of DOE (2014). These variables were selected by a panel of experts because they reported that when calculated and used collectively will give some indication on the water quality level or water quality index of a river (DOE 1997). According to the best-fit relationship for every six parameters, the new variables of the 6 sub-indexes (SI) were determined and the overall trend for Langat River was obtained using the formula given below:

$$\mathbf{WQI = 0.22\ SIDO + 0.19\ SIBOD + 0.16\ SICOD + 0.16\ SISS + 0.15\ SIAN + 0.12\ SIpH \quad (1)}$$

Where, WQI = Water quality index; SIDO = Sub-index of Dissolved Oxygen; SIBOD = Sub-index of Biological Oxygen Demand; SICOD = Sub-index of Chemical Oxygen

Demand; SIAN = Subindex of Ammonical Nitrogen; SISS = Sub-index of Suspended Solids; SIpH = Sub-index of pH value.

Generally, WQI yields a unitless number that varies between 0 and 100. Measurements of each of these parameters are taken and compared to a classification table, where the water is identified as excellent, good, fair, poor, or very poor (DOE, Malaysia Environnmental Quality Reports, 1999). The WQI of water quality parameters was calculated for each water sample by using the best-fit WQI equation (DOE, Malaysia Environnmental Quality Reports, 1999).

Each class of water is represented on the web-based system with the varying color of red based on the WQI to indicate the status of the water quality as depicted in Figure 3. This color would then fill out the heart shape drawn earlier on the web-based system. The color is used to represent quantitative data on maps by employing a single hue scheme which enables the user to assess water quality status with the hue and the different quantities or magnitudes with the lightness or darkness of the hue (Brewer, 1994). Lighter hues represent lower quantities (lower WQI value and polluted water) while darker hues are for higher quantities (higher WQI value and clean water).

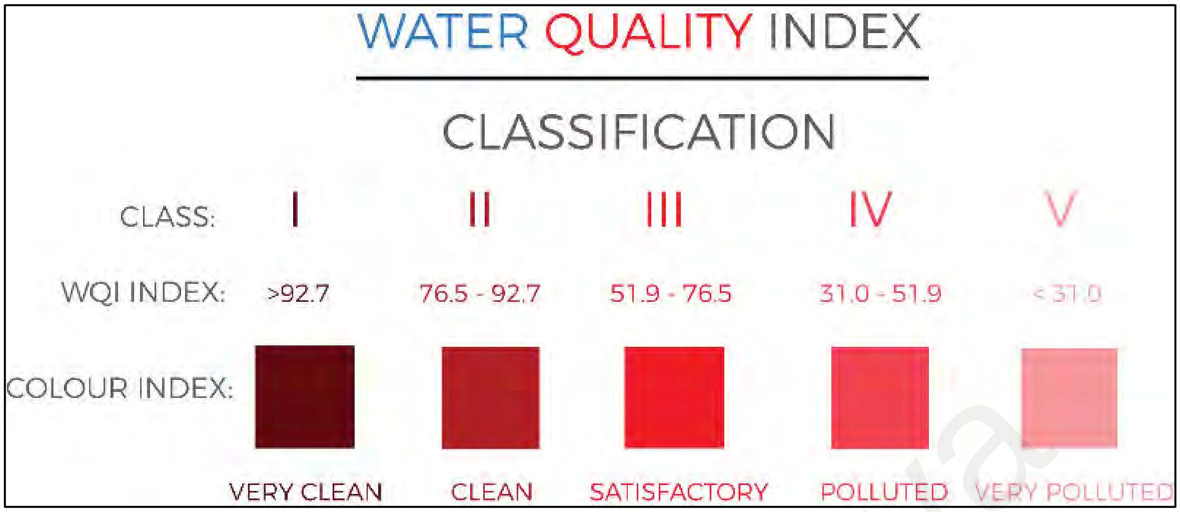


Figure 3.7: WQI Classification by color deployed in Water Quality Index Representation System.

3.2.4 Development of Web-Based WQI Visualization System

The web-based water quality system was developed using the C# programming language (Microsoft 2000). The main advantage of C# is the ease to integrate with components written in other languages and many of Microsoft's proprietary technologies and can be run on other platforms such as Google Earth. C# have access to the .NET Framework class libraries, which is essential in developing the graphical representation of the system.

Google Maps API V3.27 has been deployed in this study to view the graphical representation of WQI real-time on the Google Earth platform. The water quality data in this study have been archived using MySQL. My SQL is a free, open-source database management system that is suitable for websites because of its high-performance query engine, tremendously fast data insert capability, and strong support for specialized web functions like quick full-text searches.

To generate graphical visualization of water quality index on the web, the selected site unique ID stored in the system will be retrieved together with starting and ending date of the data that needs to be displayed. The associated parameters for the selected data and sampling site were extracted and averaged if the time frame differences were more than one day. The second step is designed to retrieve the longitude and latitude of the selected site. This important step is used to indicate the position on Google Earth where the heart that represents the WQI would be drawn. Based on the classes of the water quality parameter the heart shape would be drawn. The WQI index would then be calculated automatically and based on the WQI value, the color would be assigned as shown in figure 3.8.

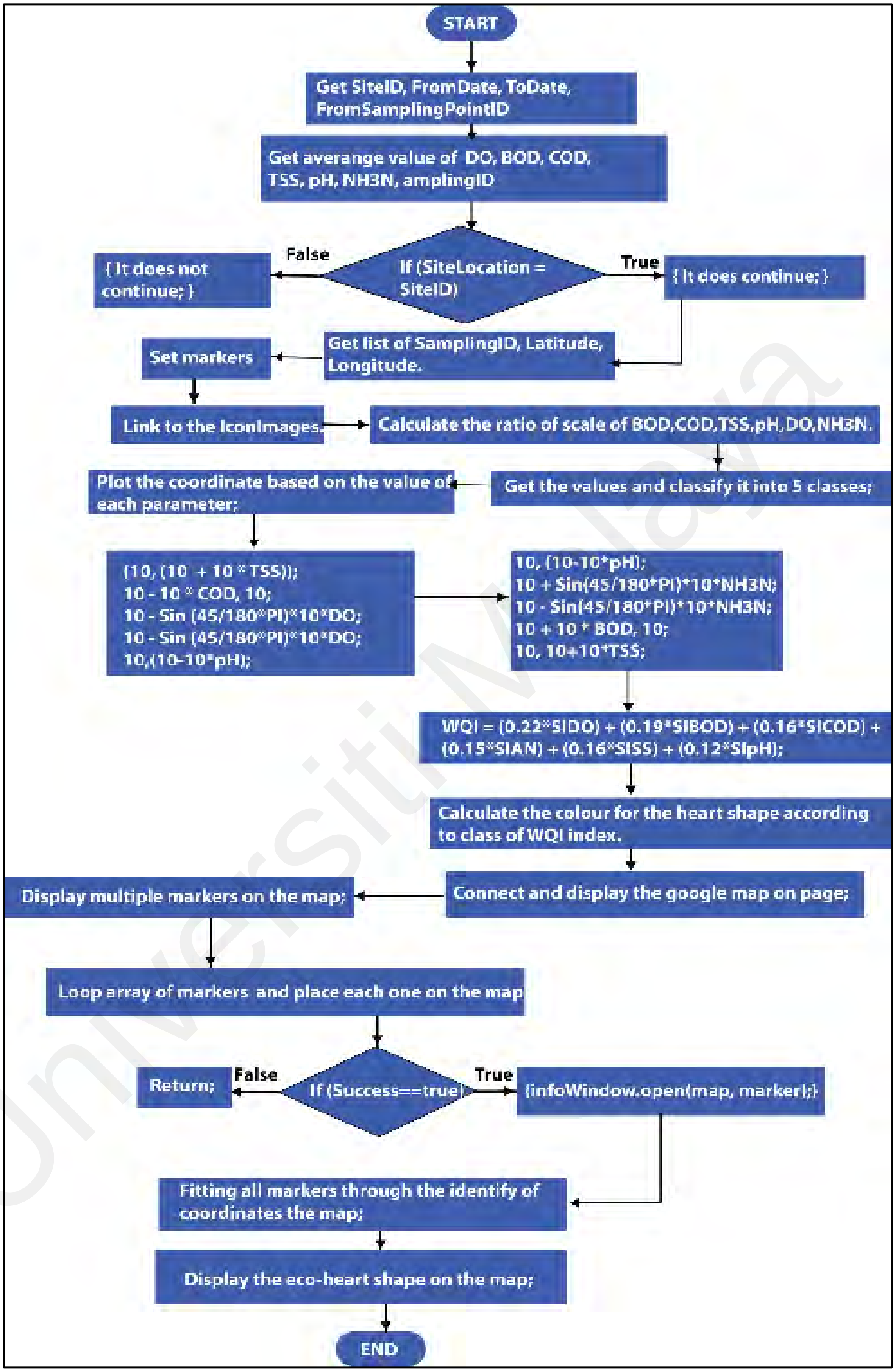


Figure 3.8: WQI Web-Based System Flowchart

The parameters to visualize the water quality index are Total Suspended Solid (TSS), Chemical Oxygen Demand (COD), pH, Dissolved Oxygen (DO), Ammonical Nitrogen (NH^3N), Biological Oxygen Demand (BOD). These parameters are then categorized into classes which are shown in table 3.8 below by using the “if-else” statement.

Table 3.8: Classification of Water Quality NWQI Malaysia.

Class	**I**	**II**	**III**	**IV**	**V**
TSS	< 25	25 – 50	50 – 150	150 – 300	> 300
COD	< 10	10 – 25	25 – 50	50 – 100	> 100
pH	> 7	6 – 7	5 – 6	< 5	> 5
DO	> 7	5 – 7	3 – 5	1 – 3	< 1
NH^3N	< 0.1	0.1 – 0.3	0.3 – 0.9	0.9 – 2.7	> 2.7
BOD	< 1	1 – 3	3 – 6	6 – 12	> 12

The obtained values are subsequently classified into 5 different classes based on the National Water Quality Standards for Malaysia (Department of Environment Malaysia, 1985.). The following steps are to plot the coordinates according to value based on each parameter. These values are calculated and converted into percentages according to the formula and the data obtained comprises an array of numbers that indicate the location of the point on the plane, based on the coordinate system. The data obtained is defined in both horizontal and vertical coordinate systems from the calculation previously to be able to identify the position on the plane.

WQI is calculated to identify the color using the “if-else” statement to classify the water quality based on its ranking according to figure 4. According to Microsoft .NET, to complete the shape by connecting the points and display the color on the plane using the C#

function “FillClosedCurve(Brush, PointF[])” to fill the interior of a closed cardinal spline curve defined by an array of the PointF structures which represents an ordered pair of floating-point x- and y- coordinates that define a point in a two-dimensional plane to display the shape. The “Brush” determines the characteristics of the fill and for the points “PointF” is the array of x- and y- coordinates structures that define the spline. The points are connected and displayed on the google map on the page, by repeating the array of makers and placing each one on the map, fitting all markers by identifying the coordinates on the map and displaying on the map. (ESRI., n.d.)

3.3 System Development

3.3.1 Dynamic System Development Method (DSDM)

Dynamic System Development Method (DSDM) is applied in this study to build the biodiversity and water quality index visualization system, where it provides a framework to produce high quality of deliverables and best practices for rapid application development since this method is both iterative and incremental (Stapleton, 1997)

DSDM is used in developing the visualization system is because there are multiple the module to be added, and it is expandable. The biodiversity indexes visualization system is planned and developed, then the water quality index visualization system is added. Thus, this system development method is versatile and the users could provide feedback on the system performance and improvement still can be done when the product is launched. Future work can be added to the system, such as waste management, electricity monitoring system, etc.

The DSDM includes the following steps, and it is repeated where there is any addition required:

1. Pre–Project
2. Feasibility
3. Foundation
4. Exploration
5. Engineering
6. Deployment
7. Post–Project

The DSDM model used in this study is depicted in figure 3.9 below. In this project, the process went through four cycles to deliver the final product. It begins with the Pre–Project phase, then moves on to the Feasibility and Foundation phases, then to the Exploration phase, then to the development and engineering of the web application, and finally to the Deployment phase, where the user testing and feedback are gathered and improvements are made. If the deliverable is accepted, the cycle will enter the Post–Project phase; otherwise, the overall project can be considered complete; however, it can still be improved or any other relevant module added, and the cycle will be repeated. Figure 3.10 depicts the development of the biodiversity indexes visualization system using the DSDM. Figure 3.11 depicts the DSDM of a water quality index visualization system. Figure 3.12 depicts the DSDM of both systems after they have been combined and launched on the internet.

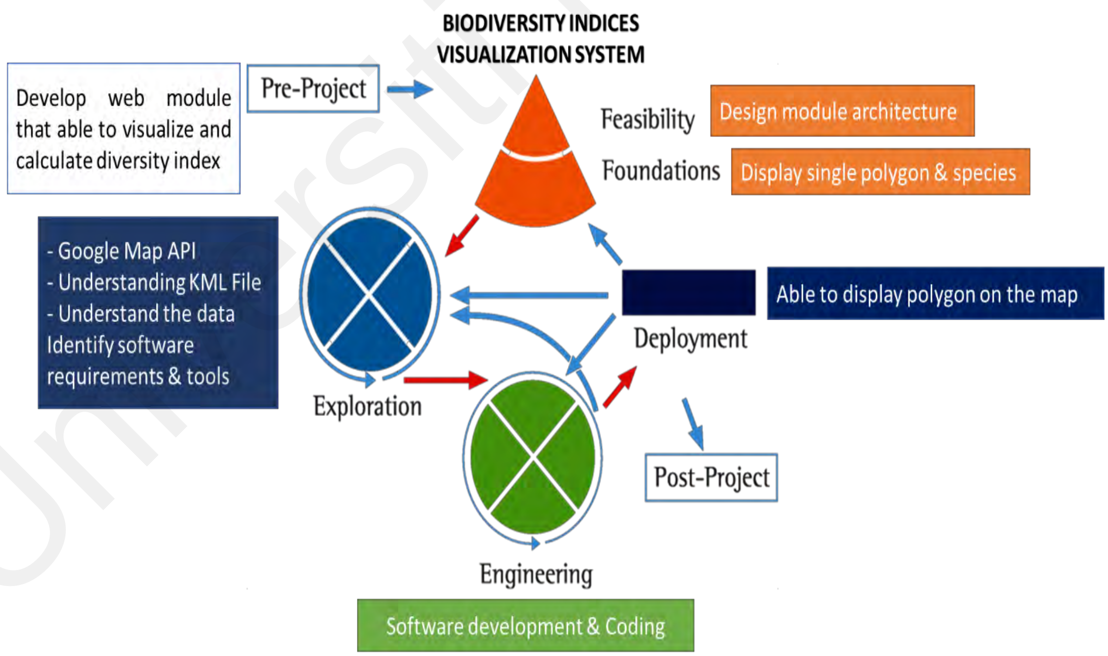


Figure 3.9: DSDM Phase1 in developing the Biodiversity Indexes Visualization system.

In phase 1, the Pre – Project is to develop a biodiversity indexes web module that can calculate and visualize biodiversity indexes and plant species on the Google Map. In the feasibility phase, the web module architecture is designed whether the project is feasible to proceed and identify the resources. A clear objective is defined at the foundation phase, where in the first phase, it can only display a single polygon and species on the map. Move on to the exploration phase, where the possible solution and research is done regarding Google Map API, understanding the map file (.KML), and fully identifying the software requirements and tools. The engineering phase included the software development and coding of the system. The expected outcome would be the polygon and species are successfully displayed on the map in the deployment stage.

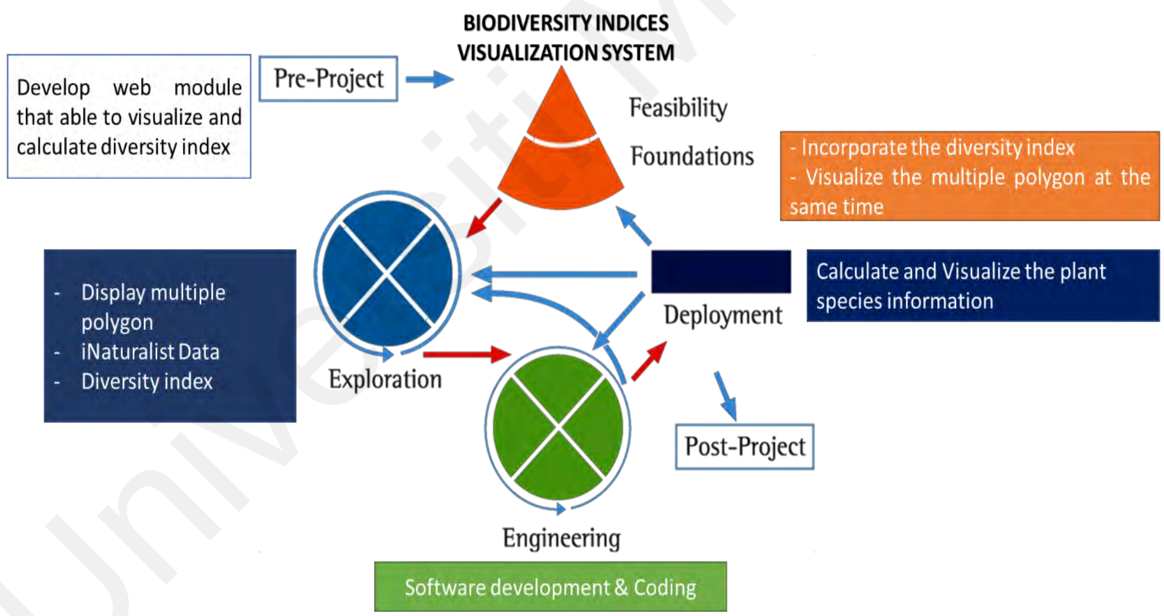


Figure 3.10: DSDM Phase 2 in enhancing the Biodiversity Indexes Visualization system.

After phase 1, the system still can be improved by incorporating the diversity indexes, displaying multiple polygons, and also adding iNaturalist species data into our system. So, in the exploration phase, the biodiversity indexes are studied and determined which indexes to be used in the system, Shannon-Wiener, Simpson Index, and Hill Diversity Number is decided after discussion with biologist and environmentalist. Besides, ways in displaying multiple polygons and the API of iNaturalist are well studied, to ensure that during engineering and coding phase runs smoothly. The expected output would be the biodiversity web module can calculate and display the biodiversity indexes together with species information.

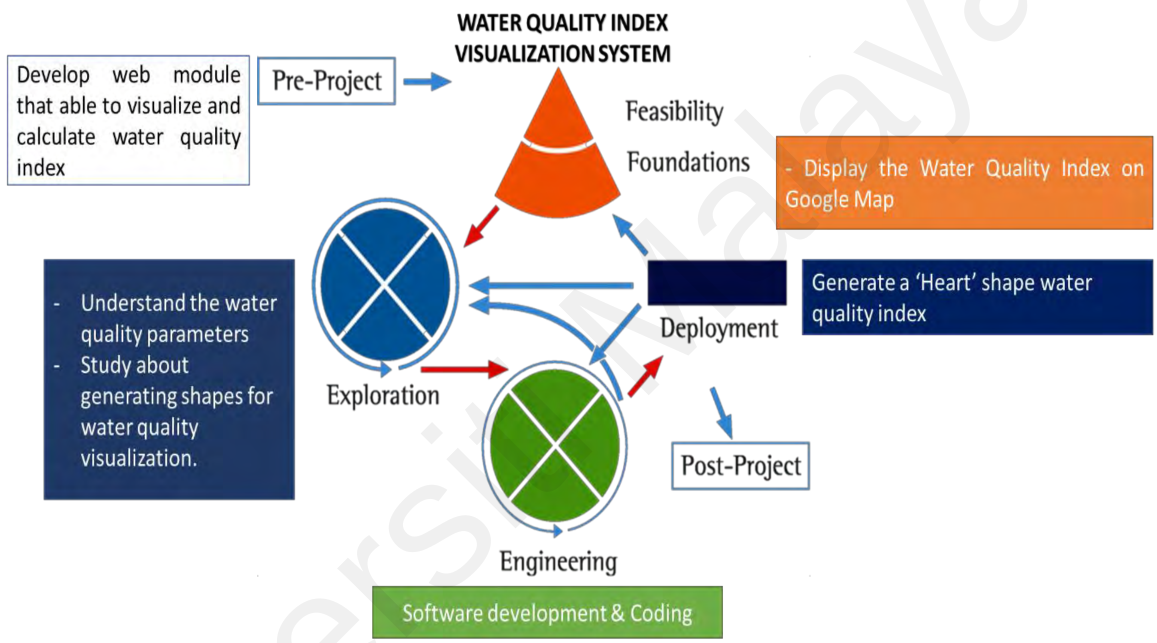


Figure 3.11: DSDM Phase 3 in developing the Water Quality Index representation system.

Besides biodiversity indexes, we also work on the water quality index representation system. In the data exploration phase, water quality parameters are well-research, and the idea was to generate a 'heart' shape representing the water quality index. Follow by the development and coding stage with the main objective in developing the visualization of the water quality index. Finally, where the objective was achieved so moving on to Phase 4, whereby incorporating both systems into one which is illustrated in Figure 3.12.

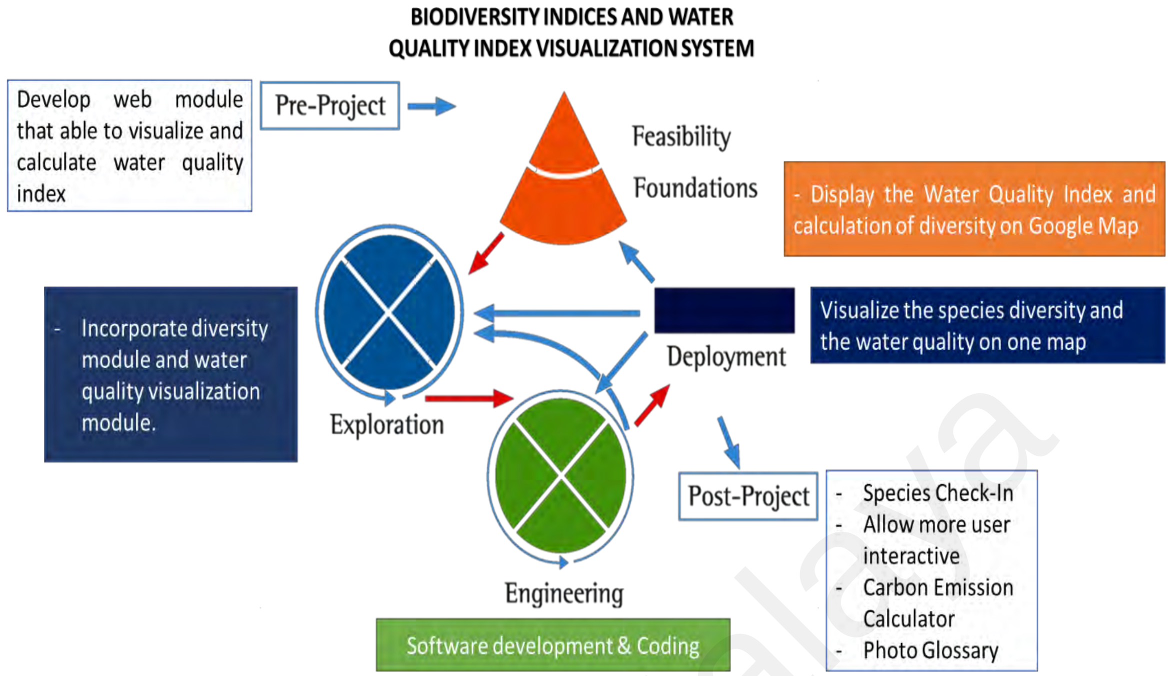


Figure 3.12: DSDM Phase 4 merging the biodiversity indexes and water quality index visualization system.

In Phase 4, both of the well-developed web modules are combined, where the Google Map is able to display both of the information simultaneously. With a well-defined objectives, moving on to the engineering phase to build the solution, and finally the web system able to visualize the species diversity and water quality index together. Thus, the overall objective of this project is achieved, in Post–project phase, where any future enhancement could be done with the system, such as Species Check-In module, enhancing in the user interactive experience, carbon emission calculator, photo glossary

3.3.2 Requirement Analysis

The requirement analysis covers the area of functional requirements and non-functional requirements for this research.

(a) Functional Requirements

A functional requirement is a description of the capabilities of the system to satisfy and to be accepted by the users. Contents of the requirements are typically the inputs of the system, expected outputs of the system, and the relationship between inputs and the outputs. The following table 3.9 discussed the functional requirements for the system.

Table 3.9: Ecological Visualization System Functional Requirements.

Functional Requirements	Explanation
Homepage	- Landing page which allows user to understand the functionality of the webpage
User Login page	- Only registered and approved users can assess the web application.
Tutorial page	- A FAQ sheet for users to understand the website functionality.
Data Upload Page	- Biodiversity species information and Water Quality Index data are required. - Users are able to upload information they are interested in.

Table 3.9, CONTINUED

Results page	- Return the results (biodiversity indexes and water quality index) from the API to the users. - Users able to download the generated results.
WQI Learning Tool	- An educational WQI learning tool, users are able to select the variables' value and visualize the water quality 'heart' index.
Report Generator Page	- Users able to generate graphs and reports of the selected information.
Area Visualization Page	- Uploaded map file and successfully capture in the database. - Users able to view the uploaded map file

(b) Non-Functional Requirements

The non – functional requirements describe the expected features and constraints of the system. This requirement is vital because it will influence the performance, security, and usability of the overall web visualization system. The non – functional requirements for this study are as table 3.10 below:

Table 3.10: Ecological Visualization System Non-functional Requirements

Non-functional Requirements	Explanation
Reliability	The web system must be reliable and can be accessed by various browsers including Google Chrome, Microsoft EDGE, Safari, Mozilla Firefox, BRAVE, and other web browsers. The system must be available and accessible always.
Security	Users are required to register and log in to access the web application, to ensure the data uploaded are secured. No third party has the right to access the data.
Understandability	Avoid the use of a long sentence in acquiring information from the users. Precise and straight-to-the-point sentences should be used. Explanation of the biodiversity indexes and water quality index is stated with the instruction given, such as video tutorial, FAQ and research template is provided for the users as guide. It will increase usability and easier, to manage, maintain and enhance the system in the future.
Performance	The system should respond fast to users' actions. Elements that need a longer pre-loading time for first-time users should be reduced.

3.3.3 Process Model

The models used to develop the proposed system are included in this section.

(a) Data Flow Diagram (DFD)

A data flow diagram (DFD) is a graphical tool that shows the flow of data in an information system. The DFD includes several components which are the data flow, process, data store, and entities. To depict the flow of data in the system, a context diagram and level 0 diagram are constructed.

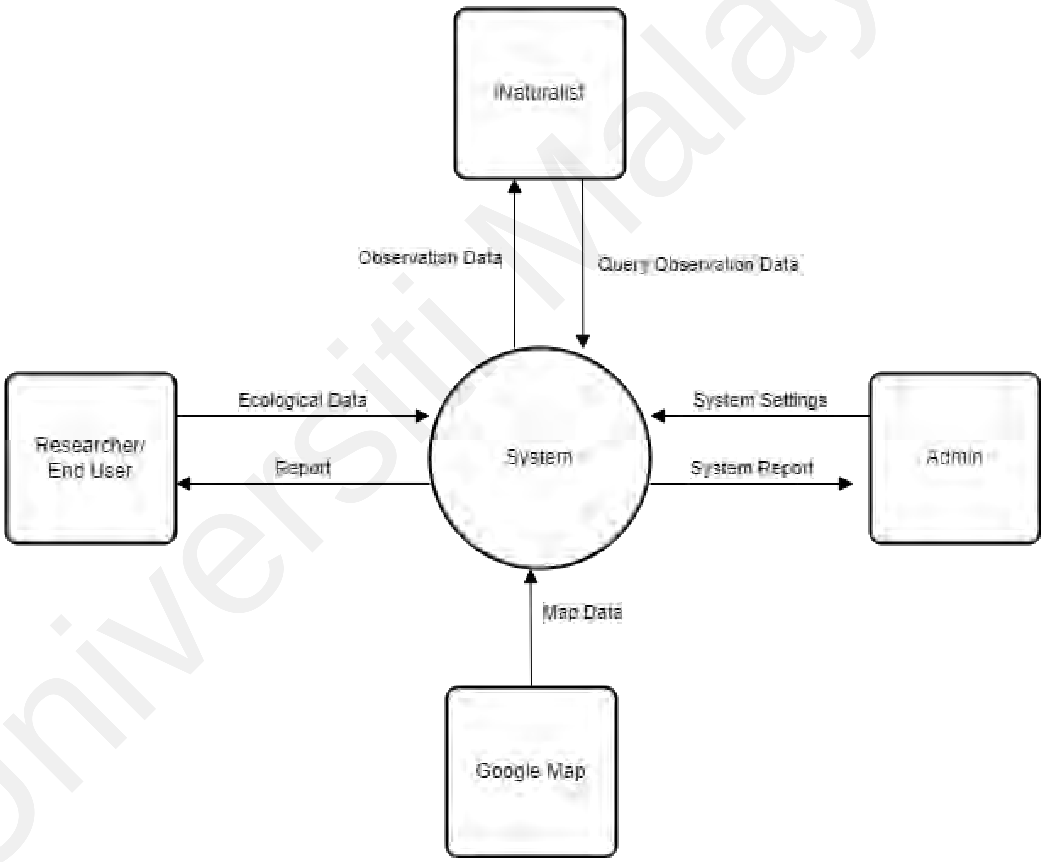


Figure 3.13: Ecology Visualization System Context Diagram.

A context diagram is the first level of the DFD, which contains the main process of the overall system. It is the most abstract view of a system, displaying the whole ecology visualization system, including the inputs and outputs, from and to the external system as shown in figure 3.13.

The ecology visualization system is the core of the research, thus it is located at the center of the diagram, where it receives inputs from Google Map, researchers, iNaturalist, and system administrator. Google Map provides the map information and the system will process the map and display the biodiversity indexes and water quality index. As for the researchers will supply ecological data into the system, then the system will process and return the information in a report format for the end-user. Follow by iNaturalist will constantly update the query observation data into the system, and the system will return the observation data to the iNaturalist server. Next, the system will receive system settings, where there is any amendment that needed to be done or requested by the user, and the system will return generated report back to the system administrator.

(b) Diagram 0

Diagram 0 zooms in the context diagram and shows the major processes, data flows, and data stores. Diagram 0 is the repeating of the entities and data flow that appearing in the context diagram. Figure 3.14 illustrates diagram 0 of this study.

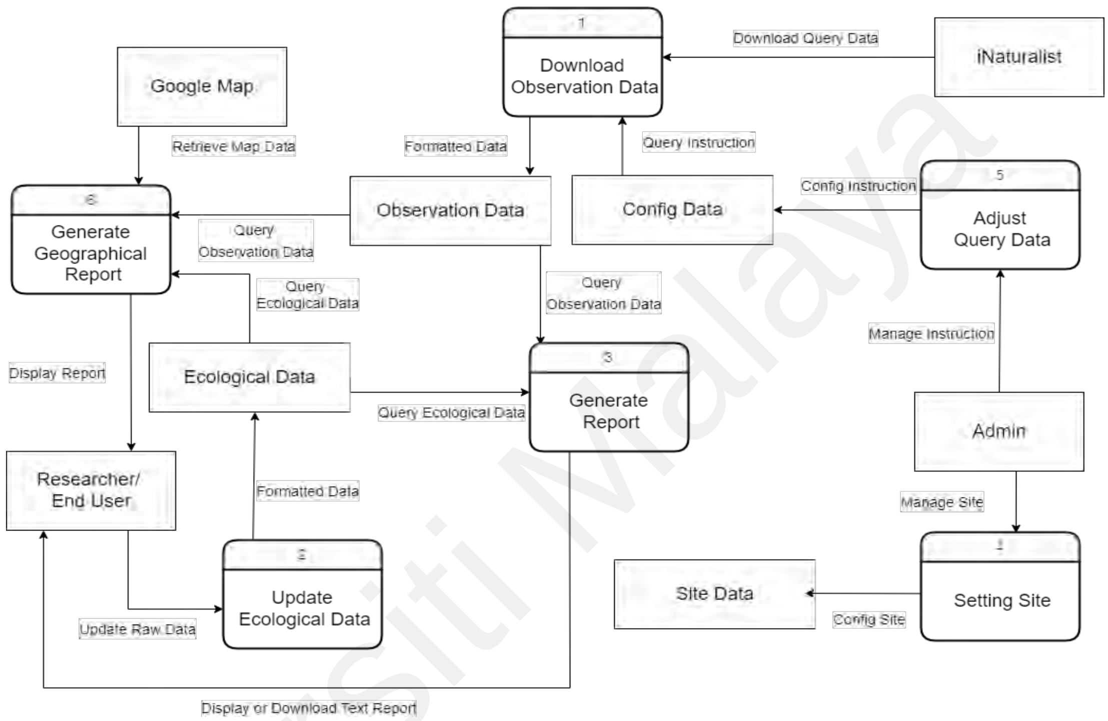


Figure 3.14: Ecology Visualization System Diagram 0.

The following walkthrough explained the DFD Diagram 0 shown in Figure 3.14:

1. 'Download Observation Data' processes the species observation information provided by iNaturalist. It processes the input data into observation data in a specific format where the information will store in the database and passed to the 'Generate Geographical Report' and 'Generate Report' processes.
2. 'Update Ecological Data' process allow the user to upload the ecological data manually from a formatted file. The file mainly included variable, collected data, data collector, sampling point, and data collected to date and time. After the ecological data is uploaded, it will be stored in a related table in the database.

3. 'Generate Report' is a process used to generate excel file type data. The user queries out the required data based on the date range and sampling point or even station. These data main used for machine learning and data prediction.
4. 'Setting Site' is a process used by the admin to set up the Site information. Each site will have its station and 1 station will have a few sampling points. This data are set up by the system admin or the site admin of a specific site.
5. 'Adjust Query Data' is the process that allows system admin to configure which data should download from iNaturalist. The system admin can add city, country, or state information into the query based on the site information. If there is a new location added, the system will download the data slowly and it takes a few days to complete the whole data. iNaturalist controlled size of downloaded data if size exceeds certain amount will treat the downloader as network attacker.
6. 'Generate Geographical Report' is a process retrieve data from a database based on the requirement of users and download maps from Google map firstly. Then it plots all the data on the sampling point as a marker based on its longitude and latitude of it. At the same, it also fills up the related contents like heart index or hill index into the dialog. Whenever the user clicked on the marker, it will show the dialog of the specific sampling point.

3.3.4 Object Model

(a) Use Case Diagram

The Use Case Diagram shows the user's interaction with the system. It shows the relationship between the users and the system, and the users is able to understand easily the system requirements.

Figure 3.15 displays the Use Case Diagram of this study, where the researchers or end-users are able to access these various functions: (1) manage ecological data; (2) generate a report, and (3) generate a geographical report. The administrator of the ecological system will be managing and maintaining the generate report function, adjusting on retrieving information from iNaturalist, and managing the site.

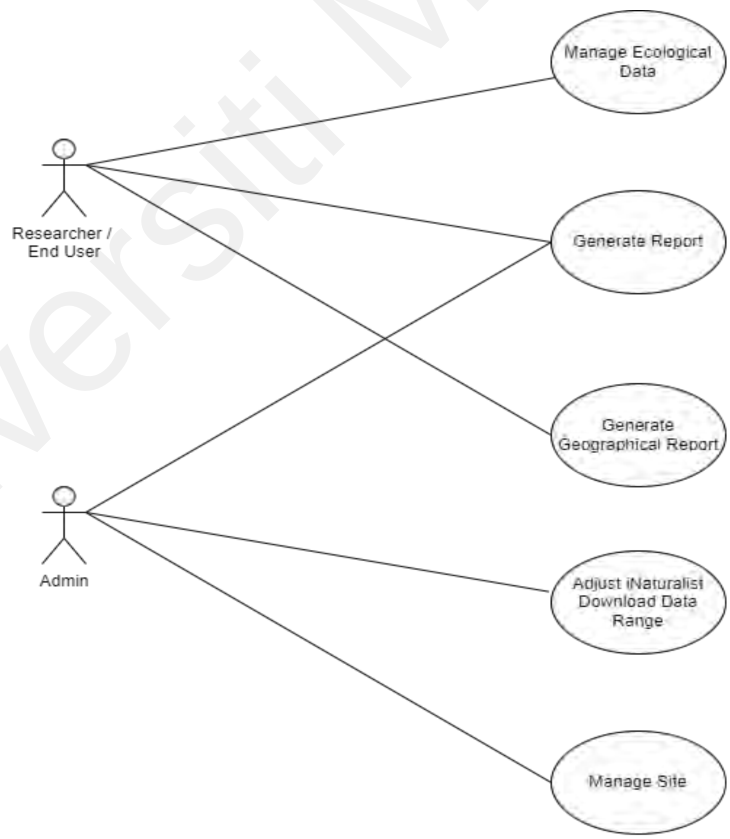


Figure 3.15: Ecological Visualization System USE Case Diagram.

(b) Class Diagram

The class diagram graphically illustrates the classes whose objects are the construction block of a system. Each class contains methods and attributes. The methods represent the processes and attributes are used to store data. The Class is the template of an object. Figure 3.16 shows the class diagram for the ecology visualization web module.

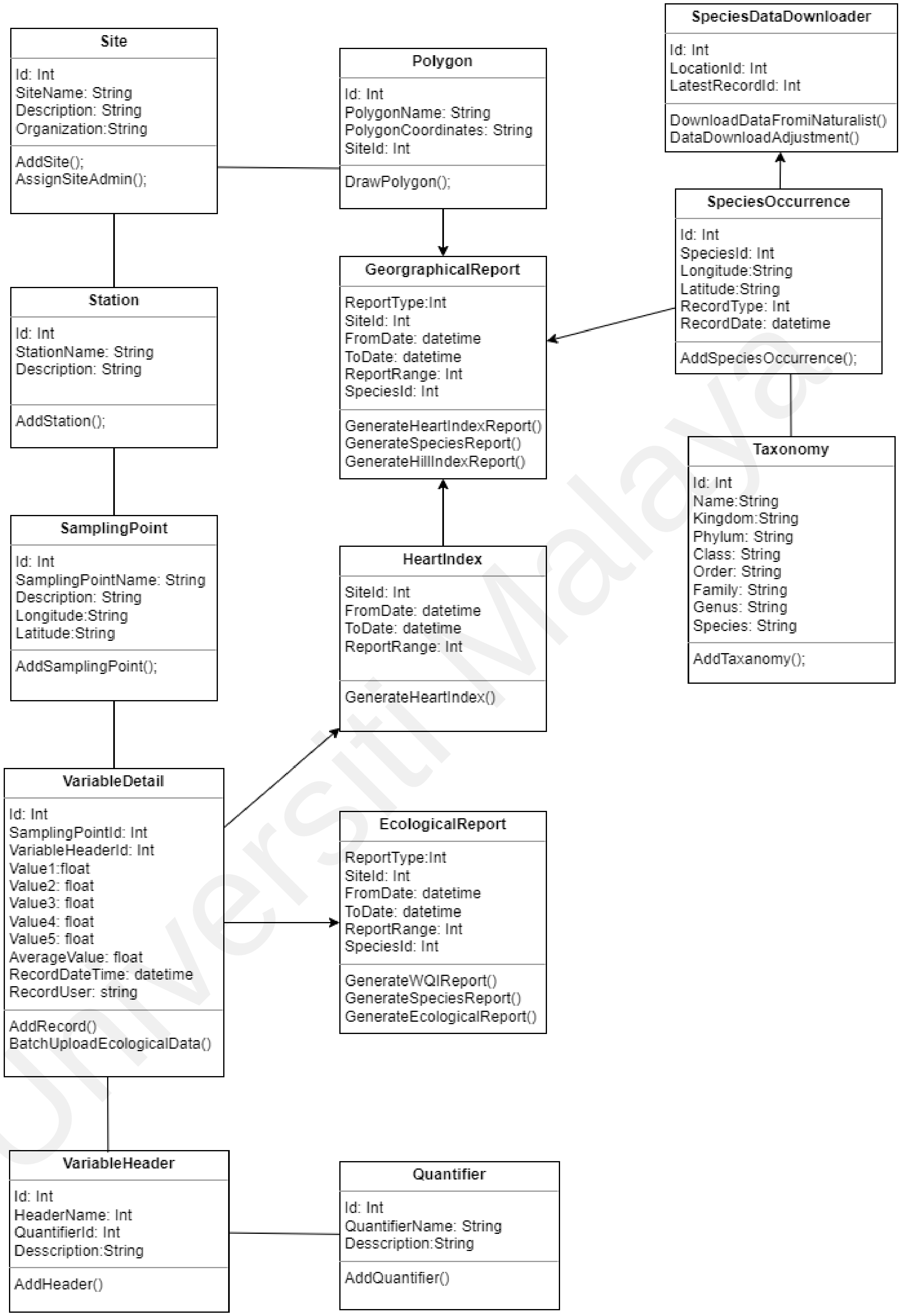


Figure 3.16: Ecological Visualization System Class Diagram

3.3.4.1 System Sequence Diagram

The sequence diagram is an interaction diagram that is used to display the scenario of USE CASE. It is often used to show the process describe in the USE Case (Kendall & Kendall, 2011).

(a) Generate Geographical Report – 'Heart' Water Quality Index Report

The process of generating a water quality index geographical report is shown in figure 3.17 below. Starting with the researcher or the end-user selecting the interested date range and site, the date range and site page will process the user request and select all stations within the site. The station data will receive the input of station data and pass the query data of the station within the data into the water quality data class. The water quality data class will process the query into a progress report and the progress report class will request the geographical map information from the Google Map API and return it with the map data. The process report class will receive input of map data and generate heart index and mark on the map, the end-user will then receive the final output, as the heart index report is displayed to the requested user.

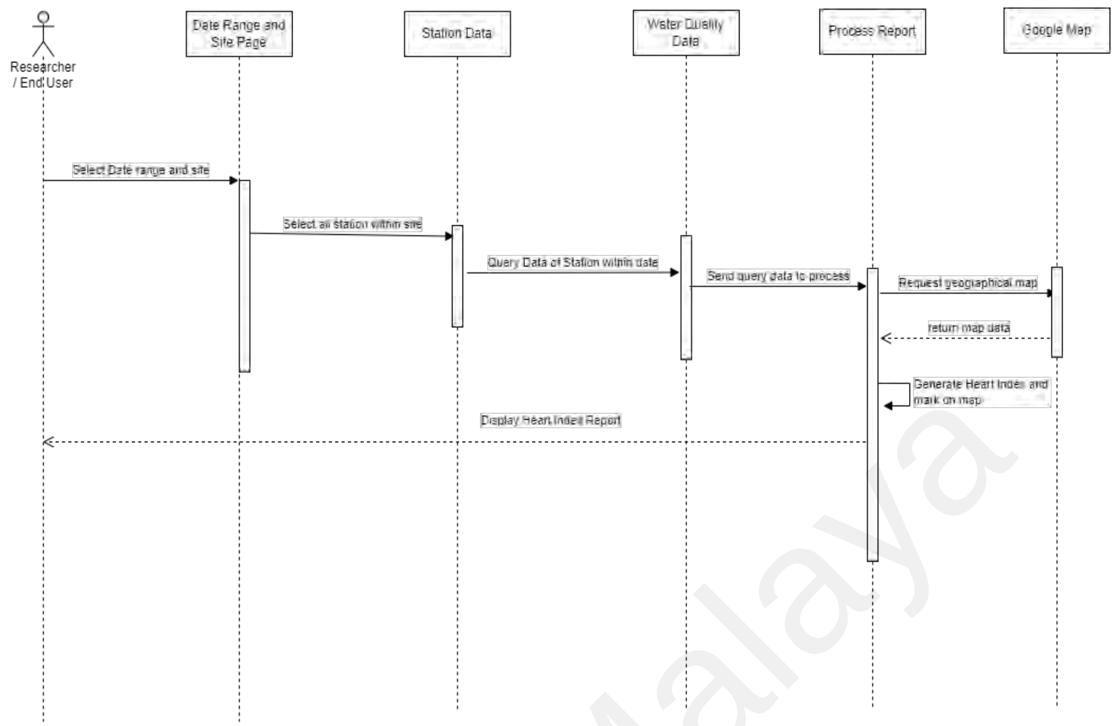


Figure 3.17: Generate Geographical Report – 'Heart' Water Quality Index Report Sequence Diagram

(b) Generate Geographical Report - Biodiversity Indexes and Hill Index

Figure 3.18 explains the sequential order of generating biodiversity indexes and hill index geographical reports. The sequence is similar to the generated a geographical report for the water quality index. First, the researcher would like to generate a biodiversity indexes geographical report, they are required to provide the input data range and interesting area. The selected site will pass the to the polygon data class and processes all the polygon within the site. The polygon data contains the required information of the interested site and the query will automatically download iNaturalist data. Then, the system will process the report by requesting a geographical map from Google Map and it will return map data to the progress report. The process report class will receive two inputs which is the return map data and the query data from iNaturalist, follow by the calculation of the diversity indexes and generate the biodiversity indexes and hill index on the map.

Lastly, the end-user will receive the geographical biodiversity indexes report as the end product.

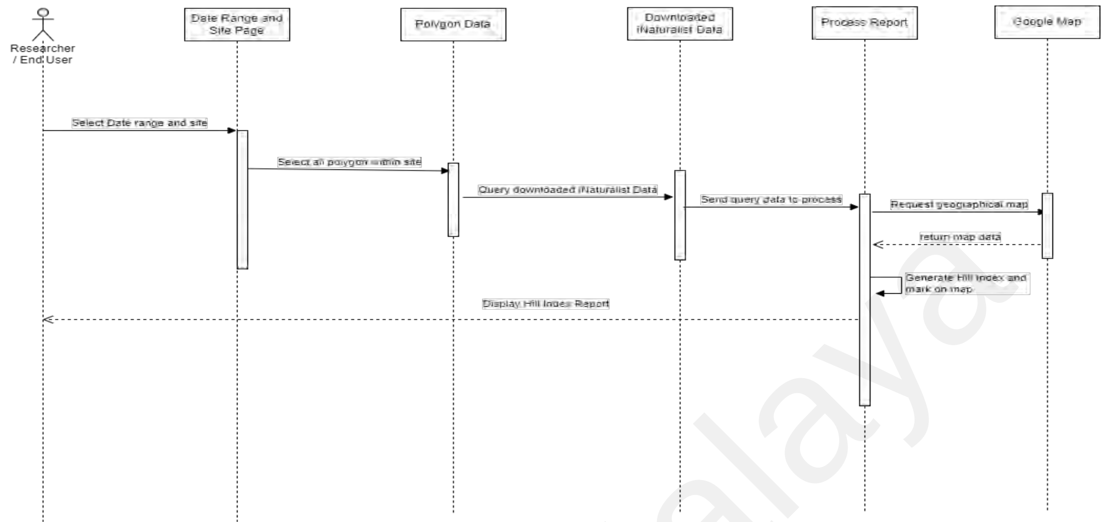


Figure 3.18: Generate Geographical Report - Biodiversity Indexes and Hill Index Sequence Diagram

(c) Generate Biodiversity indexes Graphs and Report

The users request biodiversity indexes graphs and report through the biodiversity indexes report page by selecting a date range and interesting site. The page receives the input from the user and the polygon data class will process all the polygons in the selected site, with the acquired coordinates and area, next the site will download iNaturalist data according to the requested query and pass the download data back to the system to process the data and generate graph and statistic biodiversity report. The process will export reports and display the statistical reports and graphs back to the researcher.

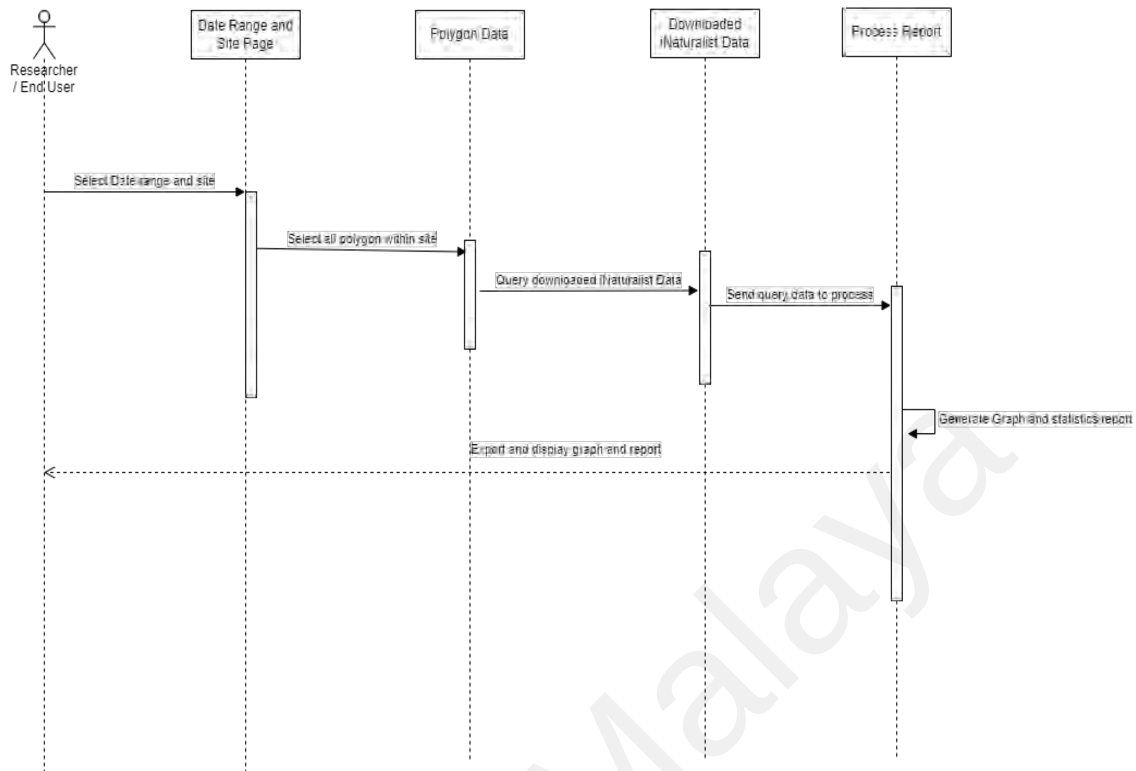


Figure 3.19: Biodiversity Indexes Generate Report Sequence Diagram.

(d) Generate Report - Water Quality Index

The researchers initiate a request in the Water Quality Index Report page by providing the selected date and location, the page will request station data and create a query from the water quality data database. Both the information will be processed into a report format where users are able to download the Water Quality graph and report.

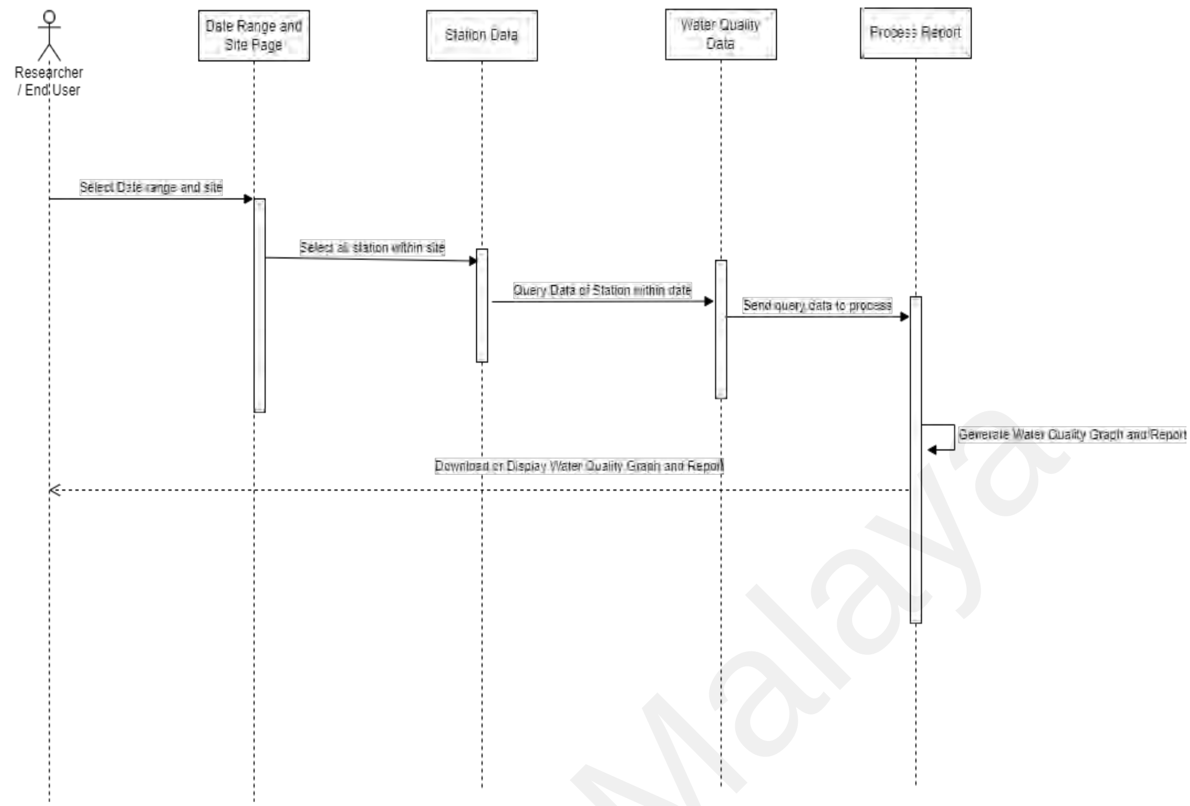


Figure 3.20: Generate Water Quality Index Sequence Diagram

(e) Manage Ecological Data – 'Add Polygon'

Users are able to upload interesting sites as 'Polygon' data into the system in the shapefile (.KML) format. The users draw the 'polygon' through the draw polygon page and the shapefile is generated. Then the generated shapefile is uploaded into the upload polygon page, which will create a new record and returns with the updated results. The updated results will then inform the users whether their file is successfully uploaded into the system.

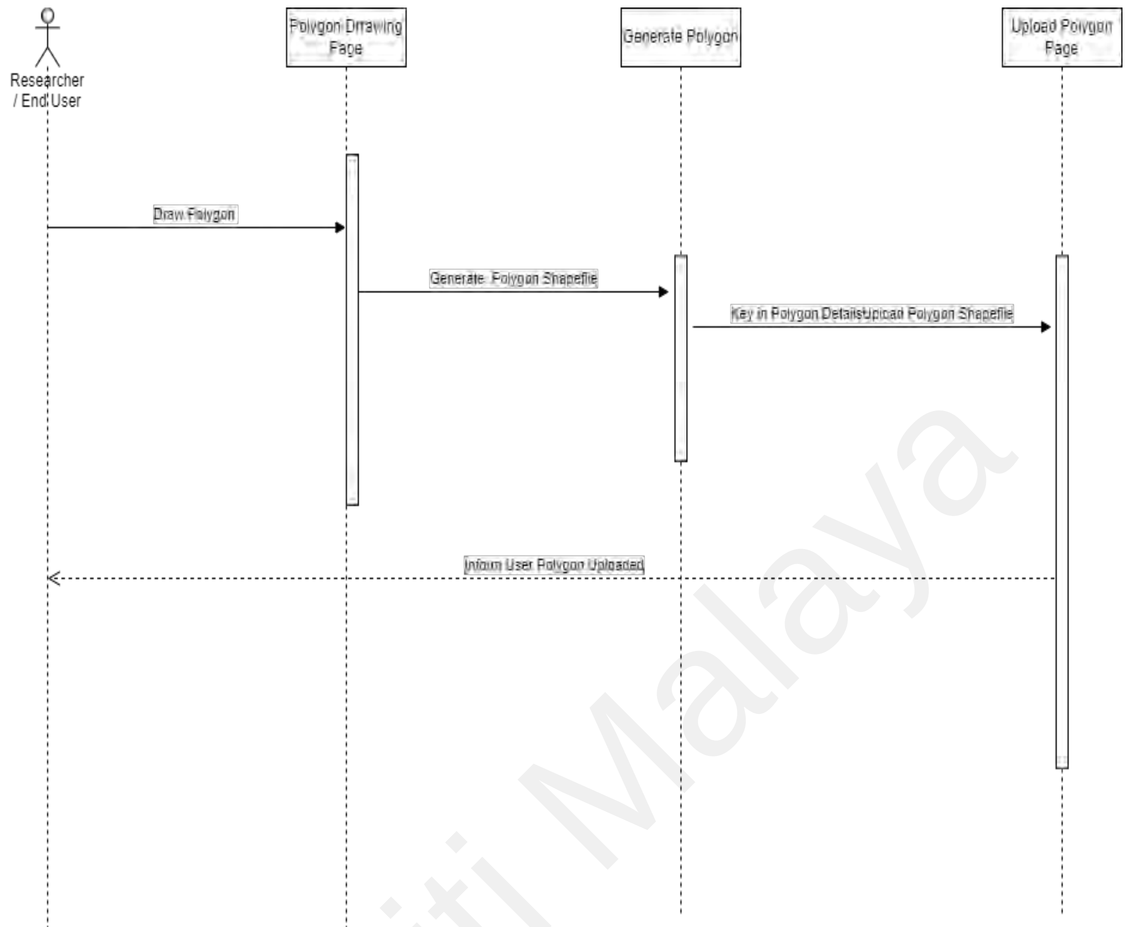


Figure 3.21: Add ' Polygon' Data Sequence Diagram

(f) Manage Ecological Data – Manage Station

Figure 3.22 illustrates the sequence diagram in managing station information. Start with researcher select the location, the page will redirect the user into to edit station page to acquire station details. Modification of the station data is done and the updated station record will replace the old record and returns the updated result to the end-users.

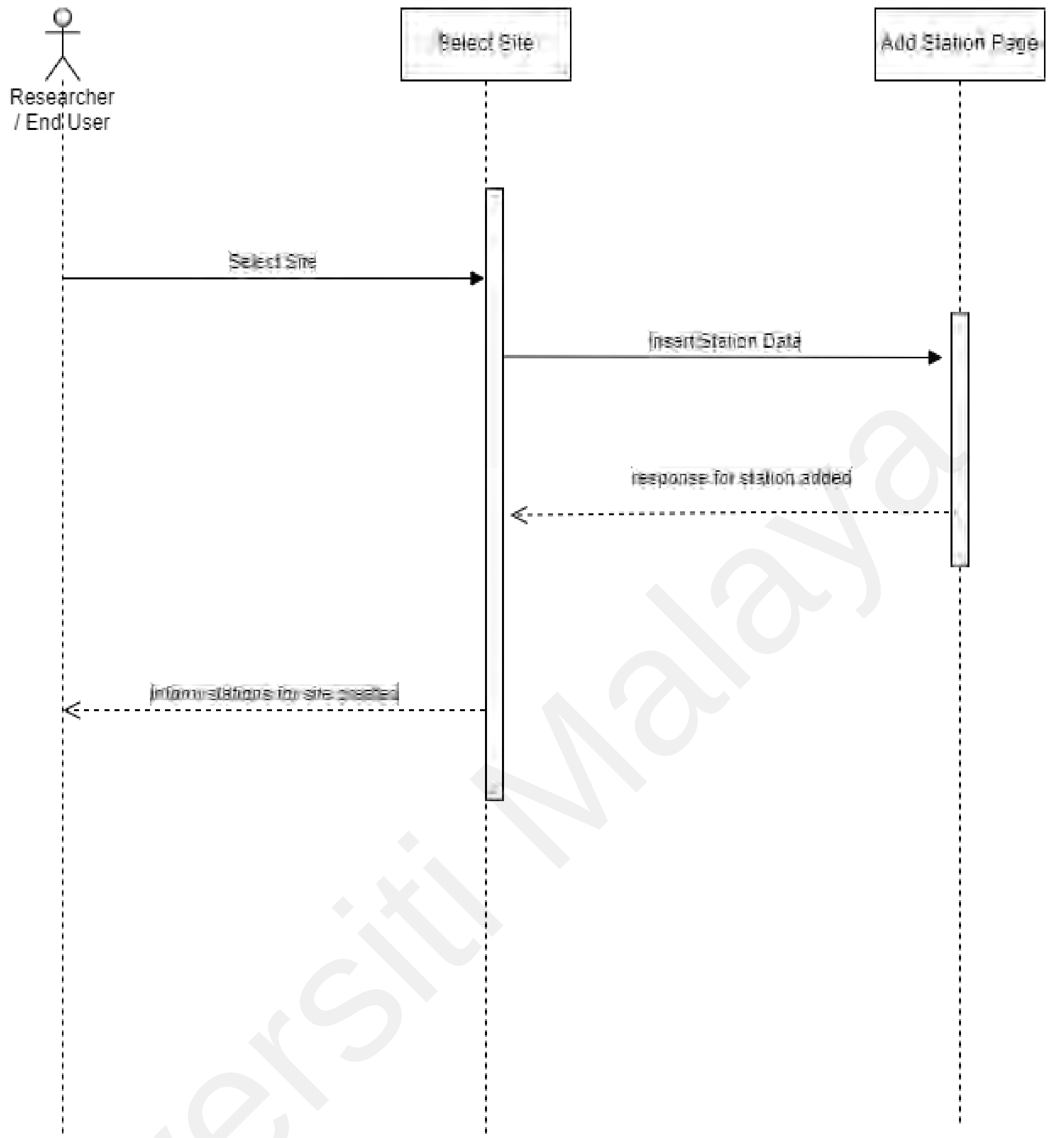


Figure 3.22: Manage Biodiversity Stations Data Sequence Diagram

(g) Manage Ecological Data – Add Water Quality Data

The sequence diagram for managing water quality data is shown in Figure 3.23. Where the user selects the location which requires modification. The page will redirect to the manage site page where the user is required to enter the required information or upload their data. The newly added and updated data will be stored in the database and returned a notification to the end-user informing whether their data has successfully been updated or not.

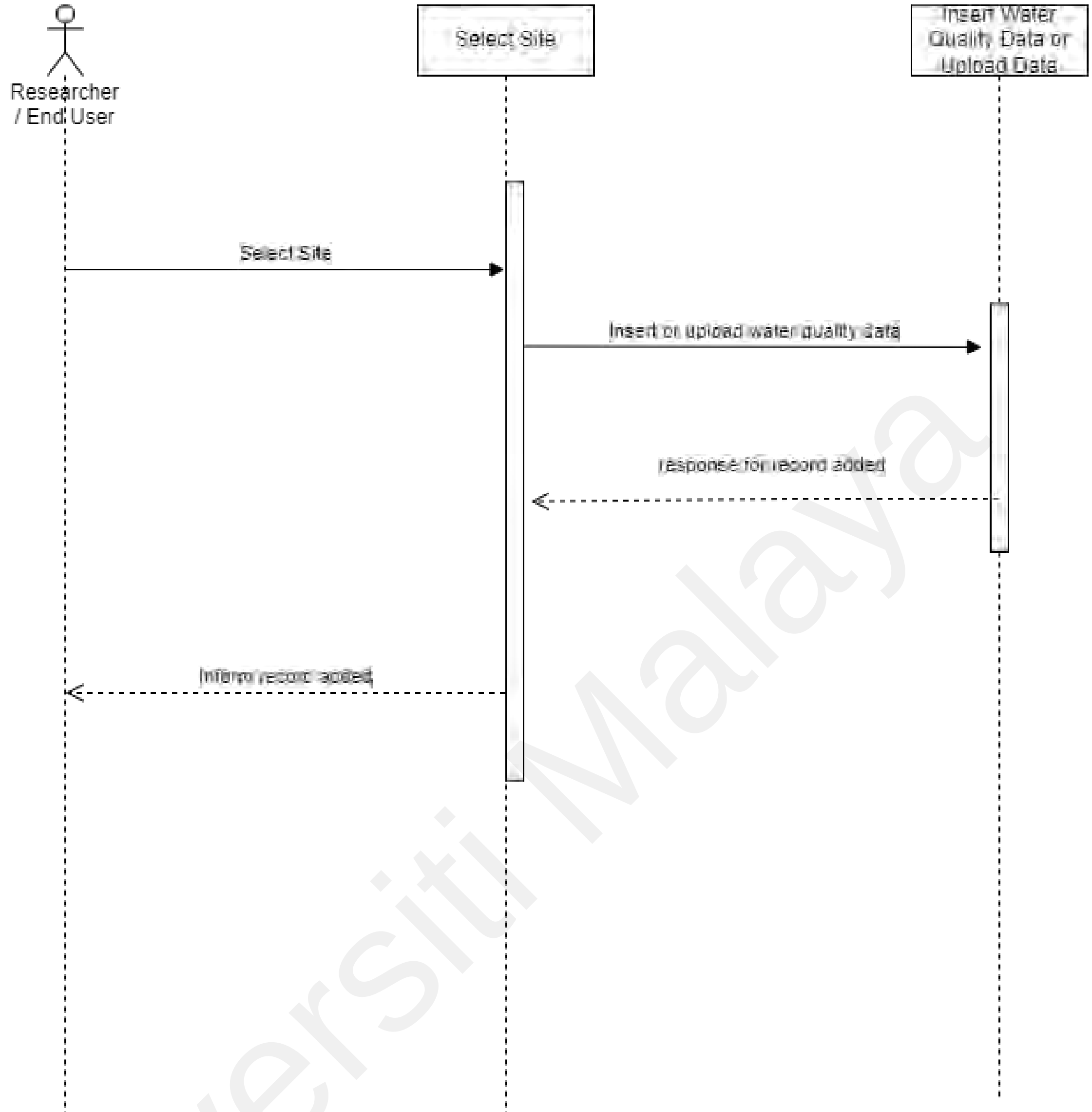


Figure 3.23: Managing Water Quality Data Sequence Diagram.

(h) Manage Site – Add Site

When the administrators are required to add new locations into the system requested by users, the administrator is required to insert site details in the 'Add Site' page. The newly added location is then assigned to the registered user who requested, then the system will respond for the user assigned and dispatch informing the requested user the new site has been created.

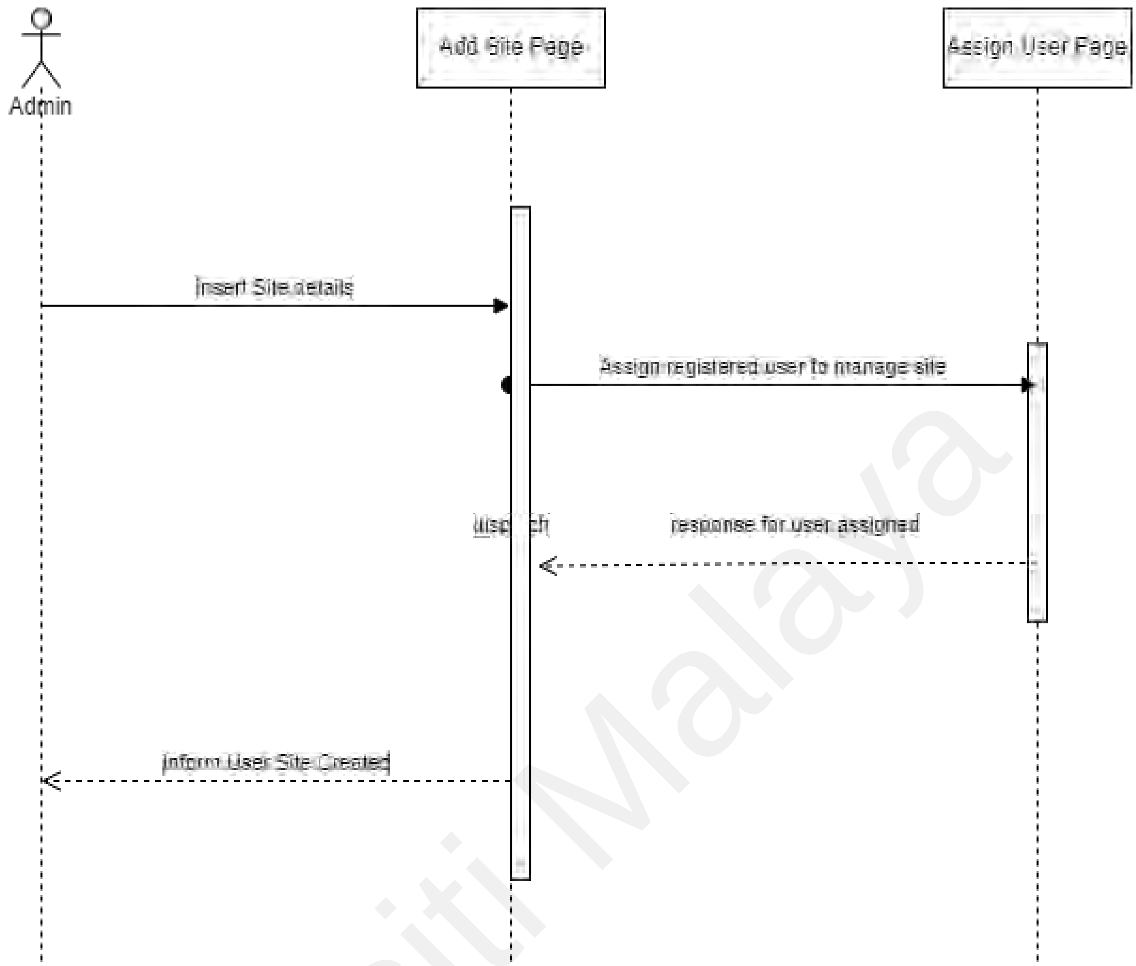


Figure 3.24: Add Site Sequence Diagram.

3.3.5 Database Design

The ecological visualization system employs the relational database model in the database implementation. The database is developed using Microsoft SQL Management Studio 17. Attributes are listed below related to the database design.

Table 3.11: Ecological Visualization System Database Design Summary.

File Name	UM Eco
Type	Microsoft SQL Management Studio 17
Usage	Maintain and keep records related to the system.
Number of tables	9

(a) Entity Relationship Diagram (ERD)

The entity-relationship diagram (ERD) depicts the interaction among system entities and provides an overall view of the system, it is a good way, to begin with modelling tasks and understands the requirements of the ecology system. Besides, it assists in the functionality of the overall system. Figure 3.24 illustrates the ERD of the ecology visualization system which is auto-generated by MySQL.

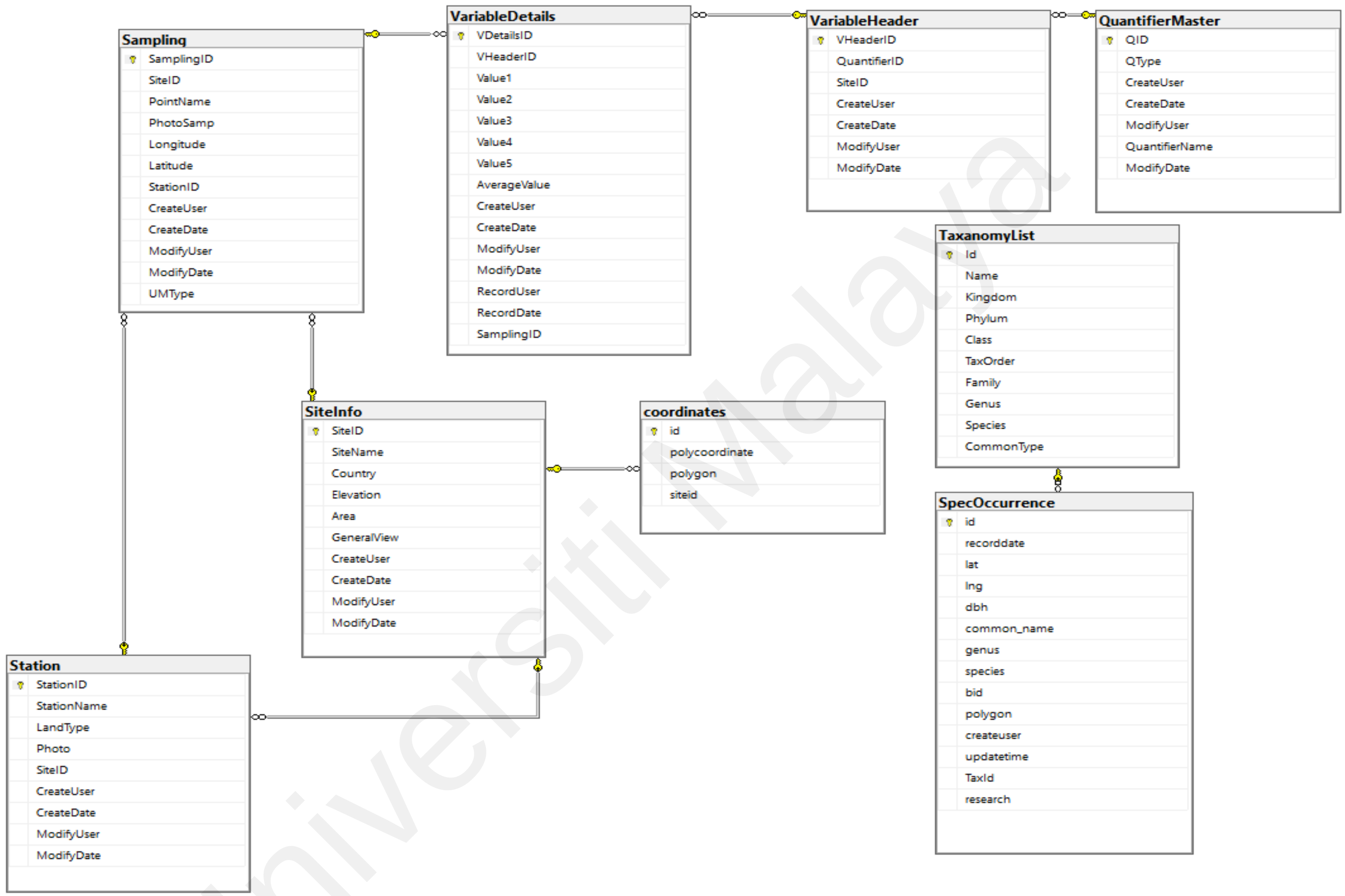


Figure 3.24: Entity Relationship Diagram

(b) Data Dictionary

Coordinates table

ColumnName	DataTypeName	MaxLength	Constraint	Description
Id	int	10	Primary Key	Auto-generated id
polycoordinate	text	16		combination of polygon coordinates
polygon	varchar	50		polygon name
siteid	int	10	Foreign Key	site unique identifier

SiteInfo table

ColumnName	DataTypeName	Max Length	Constraint	Description
SiteID	int	10	Primary Key	Auto generated id
SiteName	varchar	50		Site name
Country	varchar	50		Site country
Elevation	varchar	50		Site elevation
Area	varchar	50		Site area
GeneralView	text	16		General Description
CreateUser	varchar	50		Creator
CreateDate	datetime	8		Created date
ModifyUser	varchar	50		Last update user
ModifyDate	datetime	8		Last updated date

TaxanomyList

ColumnName	DataTypeName	Max Length	Constraint	Description
Id	int	4	Primary Key	Auto generated id
Name	nvarchar	510		genus and species
Kingdom	nvarchar	510		species kingdom
Phylum	nvarchar	510		species phylum
Class	nvarchar	510		species class
TaxOrder	nvarchar	510		species order
Family	nvarchar	510		species family
Genus	nvarchar	510		species genus
Species	nvarchar	510		Species
CommonType	nvarchar	510		Common name

Station

ColumnName	DataTypeName	Max Length	Constraint	Description
StationID	int	10	Primary Key	Auto generated id
StationName	varchar	50		Station name
LandType	varchar	50		Land type nearby station
Photo	varbinary	50		URL of station photo
SiteID	int	10	Foreign Key	Site identifier
CreateUser	varchar	50		Creator
CreateDate	datetime	8		Created date time
ModifyUser	varchar	50		Last update user
ModifyDate	datetime	8		Last update date time

Sampling

ColumnName	DataTypeName	Max Length	Constraint	Description
SamplingID	int	10	Primary Key	Auto generated id
SiteID	int	10	Foreign Key	Site identifier
PointName	varchar	50	Foreign Key	Sampling Point Name
PhotoSamp	image	16		URL of sampling point photo
Longitude	varchar	50		longitude of sampling point
Latitude	varchar	50		Latitude of sampling point
StationID	int	10	Foreign Key	Station Identifier
CreateUser	varchar	50		Creator
CreateDate	datetime	8		Created date time
ModifyUser	varchar	50		Last updated user
ModifyDate	datetime	8		Last updated date time
UMType	bool	1		UM data

QuantifierMaster

ColumnName	DataTypeName	max_length	Constraint	Description
QID	int	10	Primary Key	Auto generated id
QType	varchar	50		Quantifier data type
CreateUser	varchar	50		Creator
CreateDate	datetime	8		Created date time
ModifyUser	varchar	50		Last updated user
QuantifierName	nchar	20		Quantifier name

VariableDetails

ColumnName	DataTypeName	max_length	Constraint	Description
VDetailsID	int	10	Primary Key	Auto generated id
VHeaderID	int	10	Foreign Key	Variable header identifier
Value1	float	8		1st collected value
Value2	float	8		2nd collected value
Value3	float	8		3rd collected value
Value4	float	8		4th collected value
Value5	float	8		5th collected value
AverageValue	float	8		Average of all value
CreateUser	varchar	50		Creator
CreateDate	datetime	8		Created date time
ModifyUser	varchar	50		Last updated user
ModifyDate	datetime	8		Last updated date time
RecordUser	varchar	50		data collector name
RecordDate	datetime	8		data collected date time
SamplingID	int	10	Foreign Key	Sampling Point Identifier

VariableHeader

ColumnName	DataTypeName	max_length	Constraint	Description
VHeaderID	int	10	Primary Key	Auto generated id
QuantifierID	int	10	Foreign Key	Quantifier identifier
SiteID	int	10	Foreign Key	Site identifier
CreateUser	varchar	50		Creator
CreateDate	datetime	8		Created date time
ModifyUser	varchar	50		Last updated user
ModifyDate	datetime	8		Last updated date time

SpecOccurrence

ColumnName	DataTypeName	max_length	Constraint	Description
Id	int	10	Primary Key	Auto generated id
recorddate	date	3		Recorded date time
Lat	float	8		Latitude of recorded point
Lng	float	8		Longitude of recorded point
dbh	float	8		Diameter at breast height
common_name	nvarchar	100		Common name of species
genus	nvarchar	100		Genus of species
species	nvarchar	100		Species name
Bid	int	10		Unique record Id from iNaturalist
polygon	nvarchar	100		Polygon name
createuser	nvarchar	100		Creator
updatetime	date	3		Last updated date time
TaxId	int	10	Foreign Key	Taxonomy identifier
Research	int	3		Data from researcher or public

3.3.6 System Implementation

System implementation is the process of converting the system from the planning phase into operational mode. This is the actual implementation of the system; therefore, this phase should be carried out carefully. There is the possibility of the system design is not suitable during the actual implementation phase. Thus, the modification should be done, since there might be some features that are not suitable to be implemented or maybe overlook some important aspects of the system. In this chapter, system coding is emphasized and discussed.

(a) System Coding

Coding is the process of turning program logic with specific instructions that the computer system can execute. This section will discuss the system coding. Thus, several source codes for the system will be discussed, it includes the coding for client side, server-side, retrieving data from the database, the calculation of the biodiversity indexes and water quality index representation, please refer to Appendix III for the system coding.

I. Client-Side

The client-side prepared the Google library and google map calling scripts. Once the user submits any query for the data, the client Literal will receive JavaScript generated from the Server side. Then the browser will load the page and map based on the received JavaScript and functions called by Google map.

II. Server-Side

First, the server will retrieve data from the database and organized it into JSON format. Then it will generate the JavaScript for marker and dialog separately. After that server-side will send the generated JavaScript to the client-side. The client will call Google library functions and display the Google map and the Google library will plot the marker and dialog based on the received JavaScript from the server-side.

III. Retrieving Data from Database

The data is normally retrieved by calling the stored procedure. The figure above is 1 of the stored procedure The advantages of stored procedure are system admin can change the methods to retrieve aggression data without updating the whole system on IIS. The system admin can directly change the methods in stored procedures. Some of the fixed calculation or fixed methods will be called by using normal SQL queries generated in the system.

IV. Water Quality indexes Calculation

Firstly, the system will retrieve user query data and make calculations of WQI based on the queried data. Then the system will plot the value of each parameter based on the limit of the parameter. There are 6 points that expanded to 6 directions to form a heart shape. After 6 points are plotted, the system will link up all the points with a curvy edge. Lastly, it will paint the whole heart with red colour and based on the WQI to determine the contrast of the colour and save it in a temporary folder. Then the web client will plot the heart on a google Maps based on the recorded longitude and latitude.

V. Hill Index Calculation

The hill index will retrieve the proportions of the specific value. Based on the calculation, total richness will display to the user when q is 0. When q is 1, the Shannon Index will generate to the user. When q is larger than 1, the hill index will calculate based on the given formula and generate normally.

3.3.7 System Testing

Testing is the phase of testing the complete system. The purpose of testing is to ensure the deliverables meet will all the design requirements. To identify defects that will only surface when a complete system is assembled. Also, to ensure all the modules are well functioning and able to work correlate. A series of tests are designed to ensure the modified module interacts correctly with other system components. For the ecological visualization system, four types of testing were done, which are unit testing, system testing, integration testing, and acceptance testing.

I. **Unit Testing**

Unit testing is carried out on each module to verify the functionality is correct and bug-free. Each unit of the module are tested to ensure that the module source code is working properly. When there is any bug detected, it must be fixed.

II. **System Testing**

System testing is carried out after unit testing. The test checks if all the modules can work together properly as a system. If an error is found, affected modules will be debugged and tested all over again. The web system is fed with input to see whether the system processes information correctly into correct output and the system behaves as expected.

III. **Integration Testing**

After system testing is completed, integration testing is carried out. It is to make sure the web module can work with the existing system error-free. In the integration testing, the biodiversity indexes visualization module and water quality index representation module need to be integrated into the Ecological Web System, and able to retrieve and store data correctly on the database system.

IV. **Acceptance Testing**

Acceptance testing was carried out after integration testing. The purpose of acceptance testing is to ensure that the system meets the expectation of the user requirements, both functionally and non-functionally. Users are requested to test the completed system using the system evaluation form. Then, the feedback is collected and analysed to correct and improve the existing one.

The acceptance test for this ecology web visualization system is based on the System Usability Scale (SUS) developed by John Brooke in 1986. It is known as "quick and dirty" methods, with low-cost assessments, fast and reliable to measure the usability in the system, which only comprises 10 questions. The results of the SUS test are discussed in Chapter 4 - Results and Discussion. Please refer to the Appendix for the sample of the evaluation form.

(a) Ecological Visualization System Test Case

Test Case is necessary and it is a sequence of steps to verify a certain feature and capability as specified by the end-user. A test case includes test case ID, test case description, test data, test steps, expected result, actual result, and test status. The test case is able to determine software or a system's performance is functioning as per the requirements of the users (Rungta, 2021).

The test case used in Ecological Visualization System to ensure the system is competent and meets the standards as per requirement, please refer to Appendix IV for the test case.

(b) System Usability Scale (SUS)

The System Usability Scale (SUS) developed by John Brooke in 1986 is used as the acceptance test for our ecology visualization system. It is a "quick and dirty" method, low-cost assessments, fast and reliable to measure the usability in the system which only comprises 10 questions.

The SUS consists only of 10 questions as proposed by (Brooke, 1996) which are scored on a 5-point scale of the strength of agreement. The range goes from "strongly agree' to 'strongly disagree" and because the statements fluctuate between positive and negative, additional attention must be used when responding to the survey.

The SUS results are discussed in Chapter 4 – Result and Discussion. The sample questionnaire is included in the appendix section. Table 3.14 illustrates the comparison of the original SUS questionnaire from (Brooke, 1996) and the modified SUS statements which are included in the acceptance testing.

Table 3.12: The original SUS statements by (Brooke, 1996) and edited statements.

Original SUS Statements	Edited SUS Statements
I think that I would like to use this system frequently.	I think that I would like to use ecology visualization system frequently.
I found that this system is unnecessarily complex.	I found that ecology visualization system is unnecessarily complex.
I thought the system was easy to use.	I thought the ecology visualization system was easy to use.
I think that I would need the support of a technical person to be able to use this system.	I think that I would need the support of a technical person to be able to use this ecology visualization system.
I found that the various functions in this system were well integrated.	I found that the various functions in this ecology visualization system were well integrated.
I thought that there was too much inconsistency in this system.	I thought that there was too much inconsistency in ecology visualization system.
I would be imaging that most people would learn to use this system very quickly.	I would be imaging that most people would learn to use ecology visualization system very quickly.
I found the system very cumbersome to use.	I found the ecology visualization system very cumbersome (awkward) to use.
I felt very confident using the system.	I felt very confident using ecology visualization system.
I needed to learn a lot of things before I could ger along with this system.	I needed to learn a lot of things before I could ger along with ecology visualization system.

The users will rank each of the questions as the following, with the score of 1 indicating "Strongly Disagree", 2 indicates "Disagree", 3 indicates "Neutral", follow by a score of 4 indicating "Agree" and 5 indicating "Strong Agree".

The scores are then converted into numbers and calculated the usability score using SUS. The outcome of the calculated score is average at the SUS score of 70, consider as Good according to the SUS score shown below:

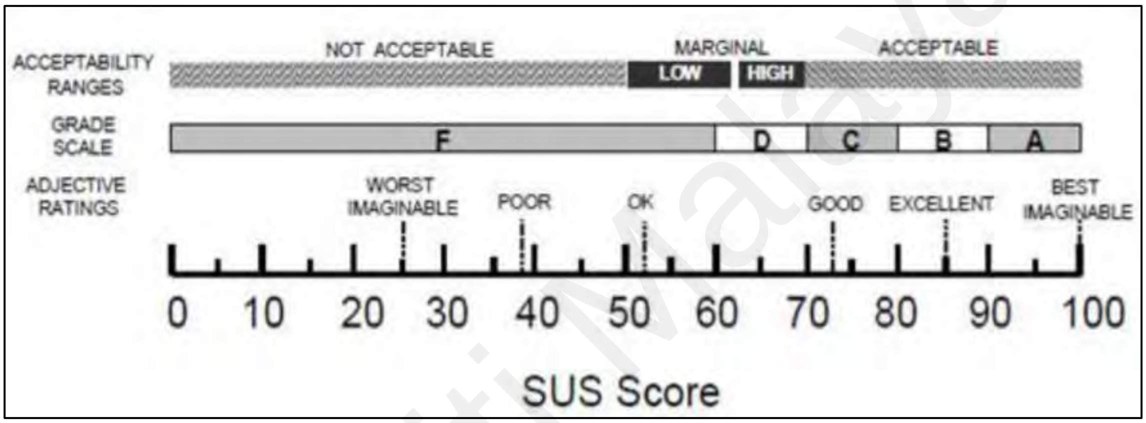


Figure 3.25: SUS scores Grade Rankings from "Determining what individual SUS scores mean: Adding an adjective rating scale." By (Bangor, Kortum, & Miller, 2009)

SUS is chosen as a usability test for our study based on its wide advocacy, its quick processing time, where the respondents can give rapid feedback and comments, as an outcome of which the information collected is processed quickly. SUS is versatile and its wide application for various programs and application systems. The SUS score can be interpreted easily and improvements can be made to improve the system's performance. (Bhat, 2018).

The SUS questionnaire is created using Google Form, it can be found in the link: https://forms.gle/1yEgxHwepcfQbqrg9 after the users have explored and test on the system.

Ecological Visualisation System: http://umlivinglabsystem.com/Default	
Username	: demo01
Password	: 123123

3.3.8 Requirements on Run Time

Hardware Requirements

The basic hardware requirements for developing the ecology visualization system are as follows”:

- Windows 10
- Intel ® Core ™ i5 – 820U Processor
- 4.00 GB Installed Memory (RAM)
- 64-bit Operating System, x64-based processor

Software Requirements

The software requirements for developing the Ecology Visualization system are as follows:

- Google Earth Pro
- Microsoft Visual Studio 2017
- Microsoft SQL Management Studio 2017 (Version 17.7)
- Google Chrome
- Mozilla Firefox
- Notepad++
- Adobe Reader
- Microsoft Office Excel

CHAPTER 4: RESULT AND DISCUSSION

4.1 Web Based System of Biodiversity Indexes Visualization System

The prototype of the biodiversity indexes visualization system is available at the following URL: http://www.umlivinglabsystem.com. The polygon representing the green space region will appear on the Google Earth Map along with the species information based on the user's selection of the species or area, as illustrated in figure 5. Additionally, users can read additional information about a species by clicking on the placemark. This information includes the species' common name, scientific name, location, date observed, and images, if available. As depicted in figure 4.1, the system enables users to select plant species based on their species, genus, or common name. After selecting a certain plant species, the system will display its availability for the polygon of interest.

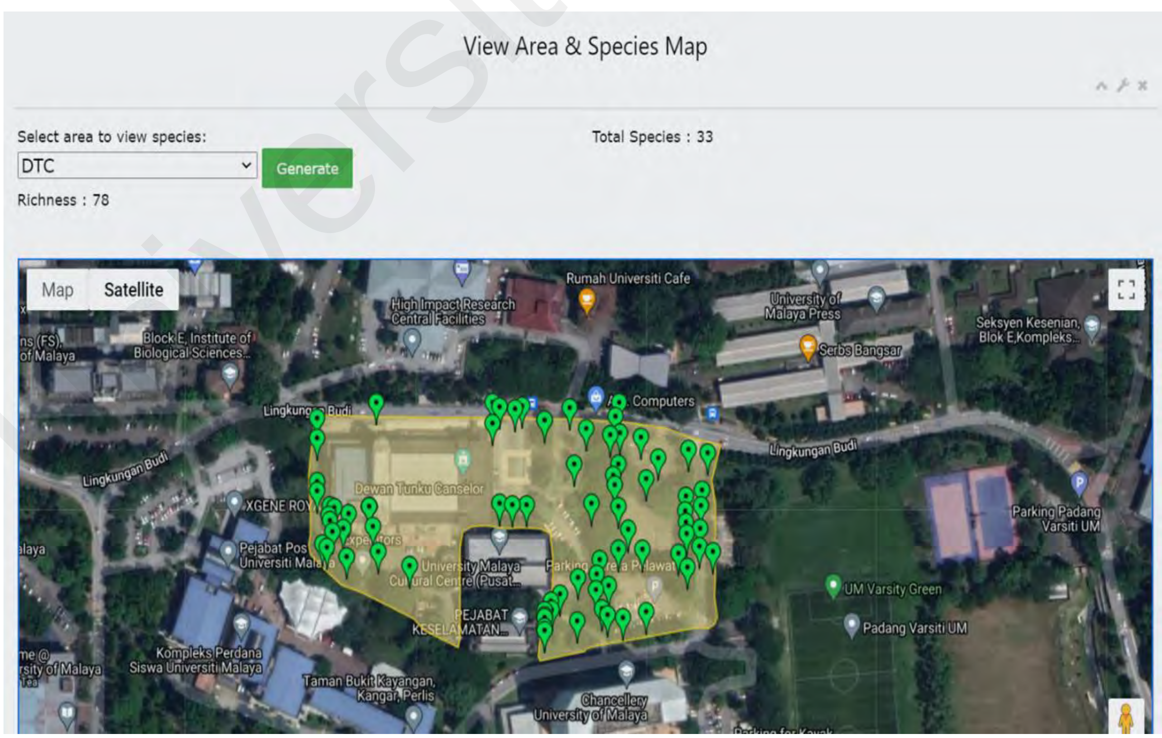


Figure 4.1: 'DTC' as the selected green space area, the species richness of the selected area is 78 with 33 different species.

Additionally, the system offers users guidance in the form of a color range for assessing biodiversity, which is dependent on the proportion of species richness, as illustrated in figure 4.2. The biodiversity range is calculated using the Species Richness (N') value, which is then converted to a percentage (percent). The polygon is colored to correspond to a given biodiversity range, which aids in the visualization of data. The color guide is capable of attracting the user's attention and instantly identifying the portion of the polygon that contains the greatest percentage of species, as illustrated in figure 4.3.

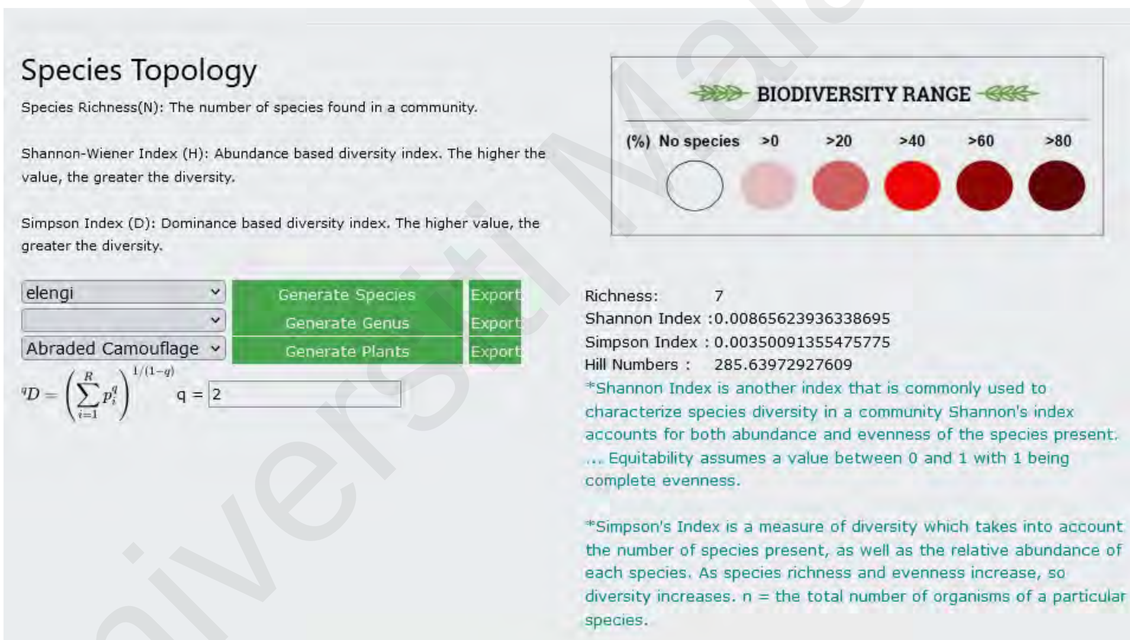


Figure 4.2: Species Topology page with the displayed Biodiversity Indexes and Hill Numbers calculations (n=0, 1, 2, … n).

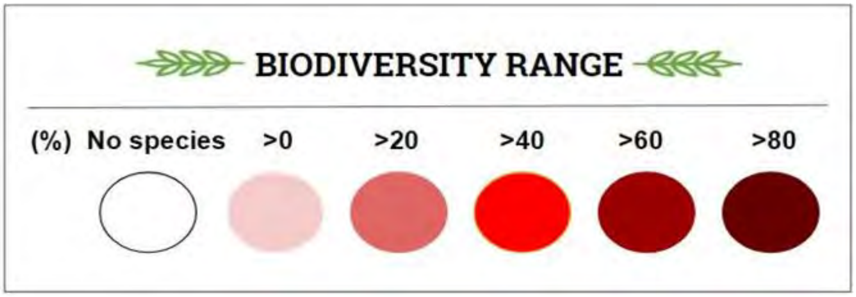


Figure 4.3: Biodiversity Range based on Species Richness (N') Percentage.

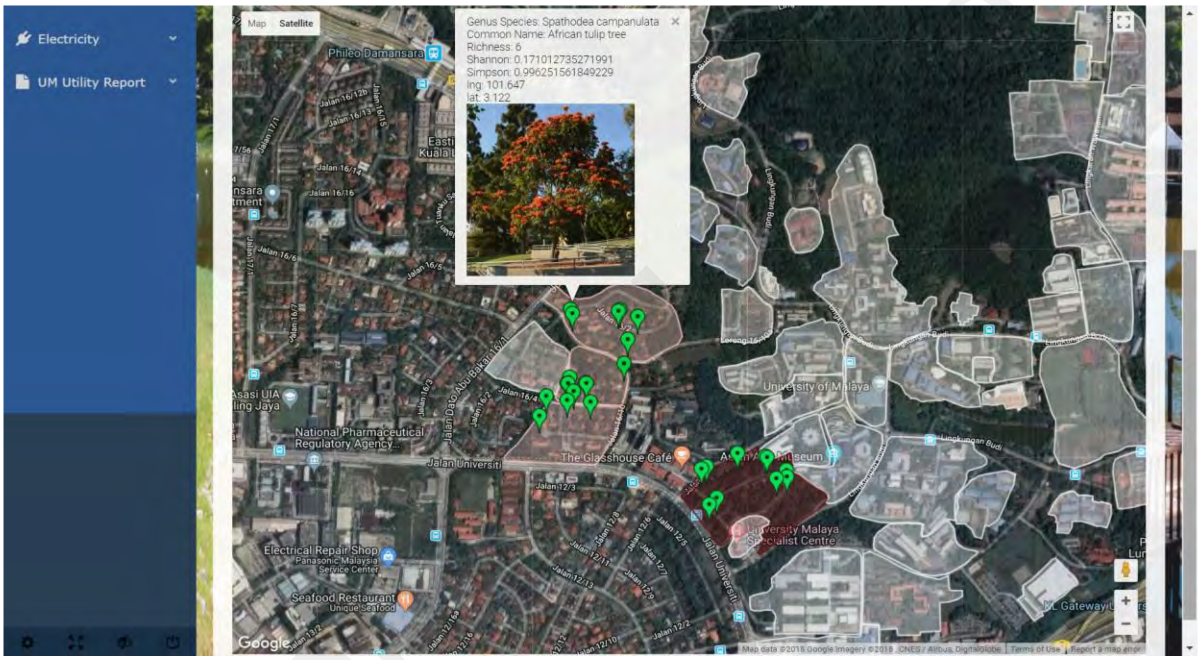


Figure 4.4: Polygon is red on the map indicates an area of with higher species density on the Google map.

When the generate option in figure 4.2 is clicked, the information about the species of interest is displayed on the Google map as a marker. When a marker is clicked, species information is displayed in a floating window, as illustrated in figure 4.4.

The floating window displays information on plant species and the diversity indexes, including the taxonomic classification, common name, coordinates, an image of the species, and the diversity indexes Species Richness, Shannon-Wiener index, Simpson's index, and

Hill numbers. Users can choose the hill number order (q = 0, 1, 2,) in a drop-down list and the system will automatically calculate it, as illustrated in figure 4.2.

Figure 4.5: Float window that shows species information and diversity indexes.

To demonstrate the biodiversity indexes, we used Geographic Information System (GIS), symbols, and various color and species information, which we deployed on the Google Earth platform. To ensure the accuracy of the information provided by the developed system, we used data from iNaturalist, a well-designed and widely used platform for species observation, with experts validating each observation. Citizen science is defined by (Burgess et al. 2017), (Cooper et al. 2007), (Cosquer, Raymond, and Prevot-Julliard 2012), (Danielsen et al. 2013), and (Theobald et al. 2015). Increased use of citizen science in ecology and conservation would be enormously beneficial due to citizen science's ability to offer data at large spatiotemporal scales. Additionally, by applying the citizen science idea, global conservation issues can be addressed (Burgess, et al. 2017).

The iNaturalist project's purpose is to enable citizens to record biodiversity data; it has been crowdsourcing biodiversity observation since 2008, collecting over 25 million recordings from over 700,000 people and involving 90,000 people in data verification

(Seltzer 2019). However, citizen science data are frequently not well welcomed by the scientific community due to worries about the rigor of data collection, which results in data quality erosion (Burgess, et al. 2017). iNaturalist's effort to ensure the data's quality and suitability for use by the scientific community includes the introduction of "Research Grade" observations, which are spatially georeferenced, contain images or recorded sound, and are identified and additionally examined by two independent reviewers. We choose to add "research quality" observation data into our system since they are more reliable and persuasive to the scientific community and can be used in future scientific publications.

As a result, the information obtained is trustworthy and can be used in the system for visualization, calculation of diversity indexes, and decision making. We chose Google Earth because the interface is simple to use and allows for quick scale changes from global to local and back. Sampling points are easily located, and Google Earth's spatial database keeps background information such as rivers and road data up to date. The lighter green space area represents fewer species observed in the area, whereas the darker green space area represents otherwise. Each placemark can be clicked, and a more detailed explanation, as well as the diversity indexes, are displayed in the web-based system's pop-up window. This allows for improved visualization of biodiversity data for use in biodiversity-related decision-making.

4.2 Web Based System of Water Quality Index Visualization System

The web-based WQI graphical visualization system is as depicted in Figure 4.6. The system has been tested on data from Langat River for the years 2005, 2010, and 2015. A pop box that contains associated water quality parameters will appear when there is a shape on the Google Earth map is clicked. The summary WQI report can be downloaded in excel format for user perusal. Figure 4.6 and figure 4.7 depict WQI generated automatically by the system for Langat River for the years 2005, 2010, and 2015. The detail of the WQI value for each figure is extracted and presented in table 4.1 for the year 2005, table 4.2 for 2010, and table 4.3 for the year 2015.

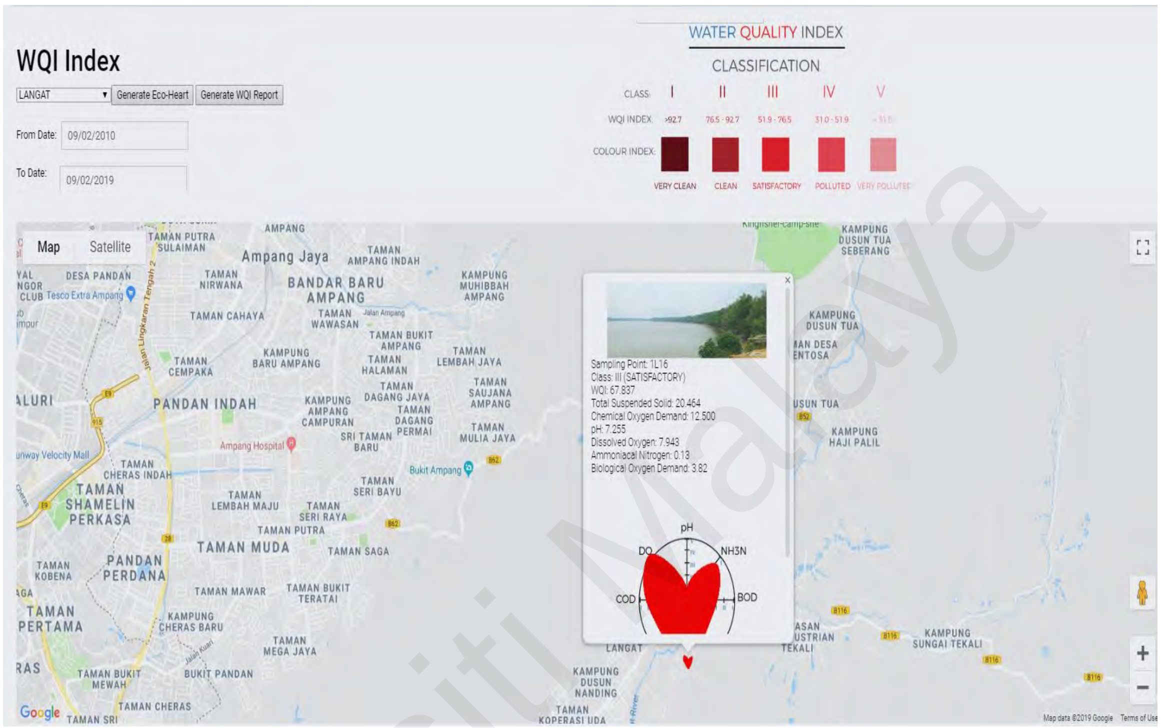


Figure 4.6: UMH2O WQI Index Graphical Visualization System.

Table 4.1: WQI parameters and water quality status at sampling points of the Langat River for the year 2005.

Sampling Point	TSS (mg/L)	COD (mg/L)	pH	DO (mg/L)	NH^3N (mg/L)	BOD (mg/L)	CLASS	WQI
1L01	130.67	25.83	6.41	4.590	0.42	2.17	III	57.408
1L02	122.75	29.58	6.428	3.958	1.33	6.50	IV	48.714
1L03	114.58	32.83	6.903	5.861	1.85	6.83	IV	46.435
1L05	258.92	45.83	6.895	5.380	2.67	11.50	IV	36.863

Table 4.1, CONTINUED

1L15	225.00	38.917	6.927	6.573	2.68	8.08	IV	40.600
1L16	83.33	27.167	7.260	8.117	0.54	2.00	III	61.202
1L07	8.17	20.67	7.46	8.19	0.04	1.50	III	71.470

Table 4.2: WQI parameters and water quality status at sampling points of the Langat River for the year 2010.

Sampling Point	TSS (mg/L)	COD (mg/L)	pH	DO (mg/L)	NH^3N (mg/L)	BOD (mg/L)	CLASS	WQI
1L01	260.67	41.50	6.11	3.51	0.60	9.00	IV	45.77
1L02	186.42	16.50	6.55	3.71	1.07	5.00	III	56.51
1L03	186.42	16.83	6.94	5.46	1.19	4.92	IV	51.16
1L05	250.17	20.08	7.01	5.21	2.12	6.67	IV	45.26
1L15	315.36	19.64	7.03	6.51	1.47	5.36	IV	47.64
1L16	85.33	9.67	7.31	7.28	0.16	3.00	III	64.31
1L07	5.67	6.33	7.35	7.70	0.18	2.17	III	70.93

Table 4.3: WQI parameters and water quality status at sampling points of Langat River for the year 2015.

Sampling Point	TSS (mg/L)	COD (mg/L)	pH	DO (mg/L)	NH^3N (mg/L)	BOD (mg/L)	CLASS	WQI
1L01	70.86	32.14	6.59	5.25	0.81	9.14	III	53.67
1L02	47.55	29.82	6.73	4.35	1.42	8.91	III	52.76
1L03	45.62	21.92	7.20	5.52	1.96	6.85	III	53.78
1L05	41.17	33.92	7.31	4.80	2.41	11.50	IV	47.90
1L15	41.69	32.15	7.32	5.11	2.49	10.85	IV	48.45
1L16	12.75	16.75	7.75	8.34	0.07	4.75	III	67.68
1L07	3.25	16.50	7.68	7.98	0.10	5.25	III	67.77

As can be seen from Figure 9 below, the WQI for Sungai Langat from 2nd September 2010 to 2nd September 2019 at the sampling point 1L07 is 71.073 under Class III, which is considered as Satisfactory. The shape generated shows a proper heart shape, which indicates the water from the 1L07 sampling point is considered clean and not polluted.

Figure 4.7: Sungai Langat WQI from 1L07 Sampling Point.

In contrast, the WQI obtained from the sampling point 1L14 is 44.651 with the Class IV, under the category of polluted. The map shows an irregular heart shape, which indicated that the water at the sampling point is polluted as shown in Figure 4.8.

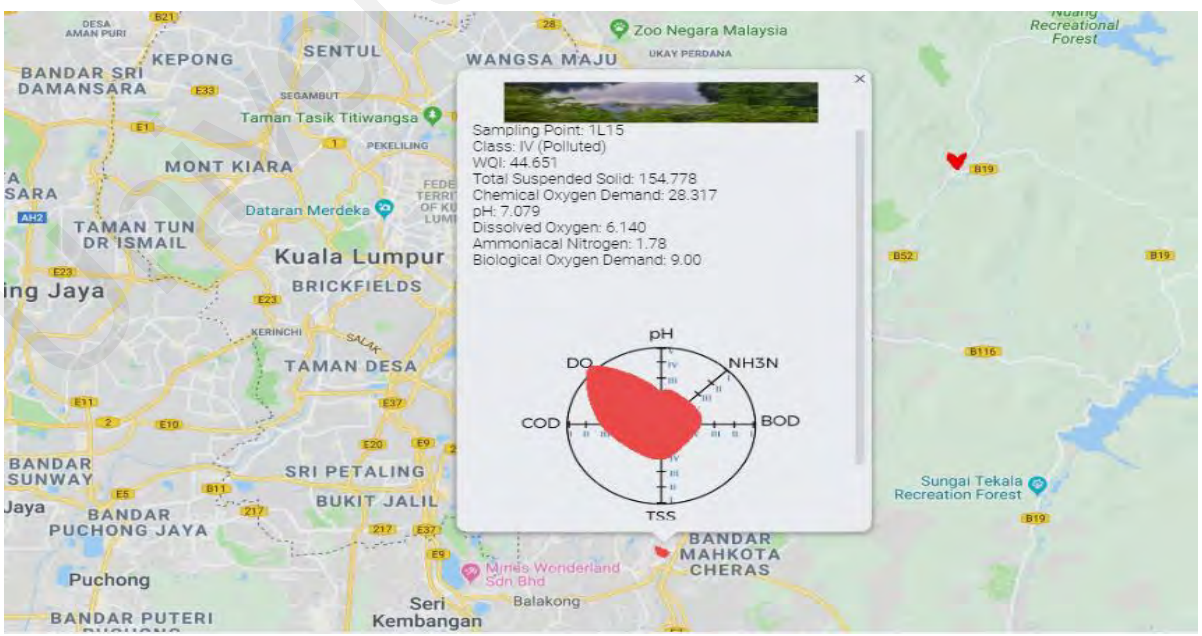


Figure 4.8: Sungai Langat WQI from 1L15 Sampling Point.

The water quality index visualization system allows users can visualize the water bodies, hydrological and ecological data with an interactive user interface easily. The interactive Google Map enables users to visualize the WQI indicated by the Eco-Heart and able to decide easily whether the river is polluted or not. From figure 4.6, figure 4.7, and figure 4.8, we can notice that the upstream of the Sungai Langat is not contaminated as the downstream.

A heart shape representing water quality status (WQI) will appear on Google Earth map based on users' selection of location and date. Users are also able to download more details related to the heart representing water quality status for the selected location and date.

This study proposes the usage of symbols and various colours to demonstrate WQI value deployed on the Google Earth platform. The Google Earth interface as it is easy to use and allows rapid changes of scale from global to local and back. Sampling points can be located easily and the spatial database of Google Earth keeps the background information of such as rivers and roads data up-to-date. Lighter hues represent lower qualities (lower WQI value and polluted water) while darker hues are for higher quantities (higher WQI value and clean water). The shape has been said to be inappropriate for encoding quantitative data because the shape does not form a natural (Cleveland, Graphical perception: Theory, experimentation, and application to the development of graphical methods., 1984). However, we believed that shape might be a useful code for conveying quantitative data if a continuum of shapes could be established in the legend and detailed explanation of the heart shape is illustrated in the pop-up window on the web-based system. The web-based WQI monitoring has been tested on data from Langat River.

The heart shape and color for all the sampling sites in this study has shown improvement in water quality from the year 2005 to 2015 except for sampling points 1L05 and 1L15. Both of the sampling stations are represented as distorted heart shape with a lighter color of red that indicates a lower WQI value. The water at this sampling point represents class IV that is polluted water. Station 1L05 and 1L15 are both located at the Hulu Langat catchment area of the Langat catchment. The Langat catchment area consists of agriculture, forest, water bodies, and commercial and residential area.

Rapid urbanization from agricultural land of the Langat River has been identified since the year 1981 (Amini, Adjustment of peak streamflows of a tropical river for urbanization., 2009) (Jaafar, 2009). Station 1L05 and 1L15 are located along Langat River which is used for drinking water, recreation, industry, fishery, and agriculture. The river flows from Gunung Nuang across Langat Basin to Kuala Langat and land use activities along the river banks contribute to the deterioration of river water quality (Charlie, Assessment of water quality characteristics during base and storm flow events on Sungai Langat basin., 2010).

Langat River deteriorating water quality is caused by the point of source pollution and non-point sources. Point of source pollution at the Langat River is from industrial and agro-based industry discharge and domestic sewage from treatment plants and animal farms. The non-point of source pollution in Langat River is contributed from rainfall, irrigation, and overland surfaces into the drainage system (Juahir, 2011).

The water quality system is also for water quality managers to utilize; anyone may participate in monitoring the health of a river, dam, estuary, or wetland near them. This is owing to the system's simplified WQI interpretation, which does not necessitate specialist expertise. By minimizing the sources of pollution to the river, the community obtaining information on the status of the watershed nearby can play a key role in protecting the catchments.

As a result, graphical analysis is critical for determining patterns, trends, and other elements that are difficult to discern from numerical summaries. Visual analytics is significant because it allows for the clear and effective transmission of information through graphical means. It's useful for analyzing data that are crucial for monitoring biodiversity species, water quality, and making decisions.

4.3 Combination of Water Quality Index and Biodiversity Indicies Representation

Figure 4.9 illustrates the integration of the water quality index and the biodiversity indices. The green marker represents the available species and the red shape appears on the map represents the water bodies collected. Researchers and ecologists able to make assumptions and inferences with the graphical presentation as shown in figure 4.9 below:

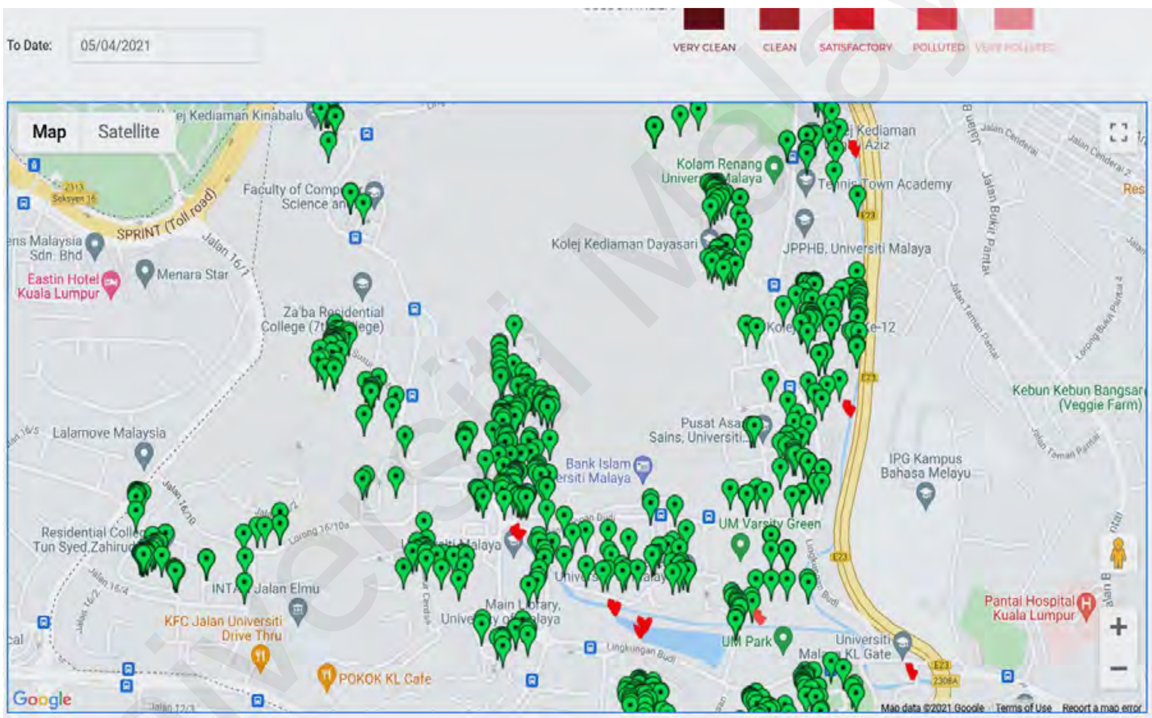


Figure 4.9: Users able to visualize the species topography and the Water Quality Index at the Water Bodies simultaneously in University of Malaya.

4.4 Reports and Graphs

4.4.1 Biodiversity Species Distribution Report

Figure 4.10 below displays the biodiversity species distribution report page. The users can select the date range and interesting species, when the ‘submit’ button is clicked, the graph that displays the species richness by the specific location is shown to the users.

Besides, users can select various graphical representations including, bar charts, line graphs, stacked charts. Users can save the generated statistical graph as an image and export it in an excel format for their studies.

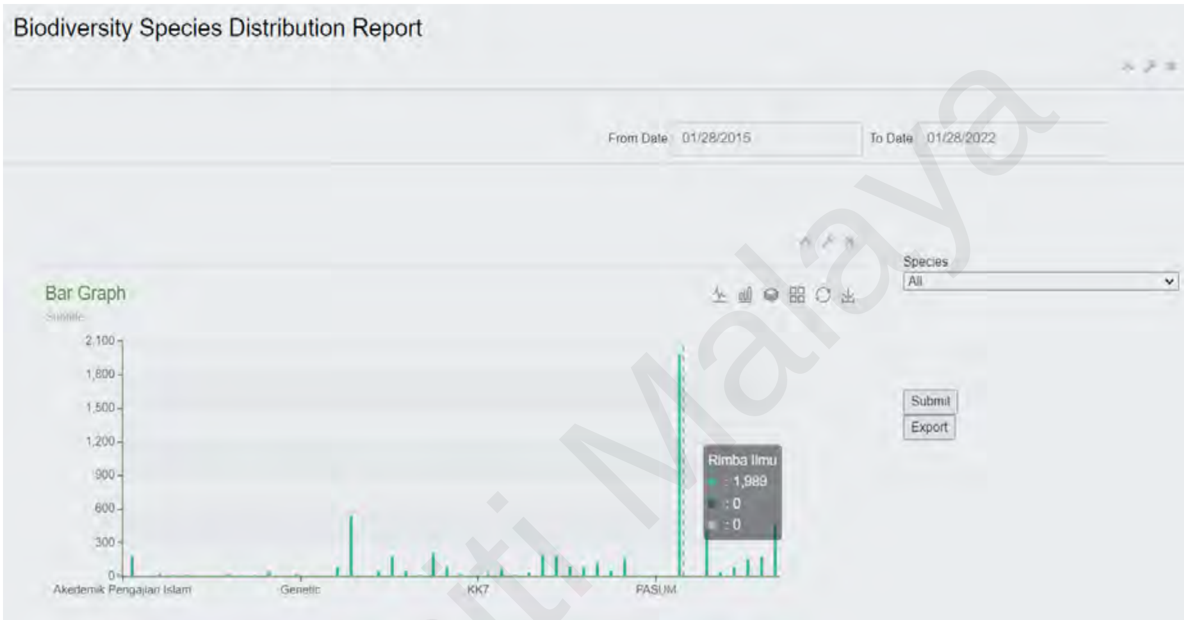


Figure 4.10: Biodiversity Species Distribution Report.

4.4.2 Water Quality Index Report

Figure 4.11 shows the generated report of the water quality index, where the users selected the date range from the year 2005 and the Langat River. The file is then automatically downloaded in .csv format.

Sampling Point	Class	WQI	Total Suspended Solid	Chemical Oxygen Demand	pH	Dissolved Oxygen	Ammoniacal Nitrogen	Biological Oxygen Demand
4EN01	III (SATISFACTORY)	70.525	8.2	9.6	6.378	6.356	0.1	2.8
4EN02	III (SATISFACTORY)	65.824	10.2	18.4	5.51	5.01	0.09	4.2

Figure 4.11: Generated Water Quality Index Report in Excel (.csv) format.

Besides, the site provides an interactive tool for the users to learn the water quality index representation, where users able to select various parameters and generate the ‘eco heart’ water quality index as shown in figure 4.12.

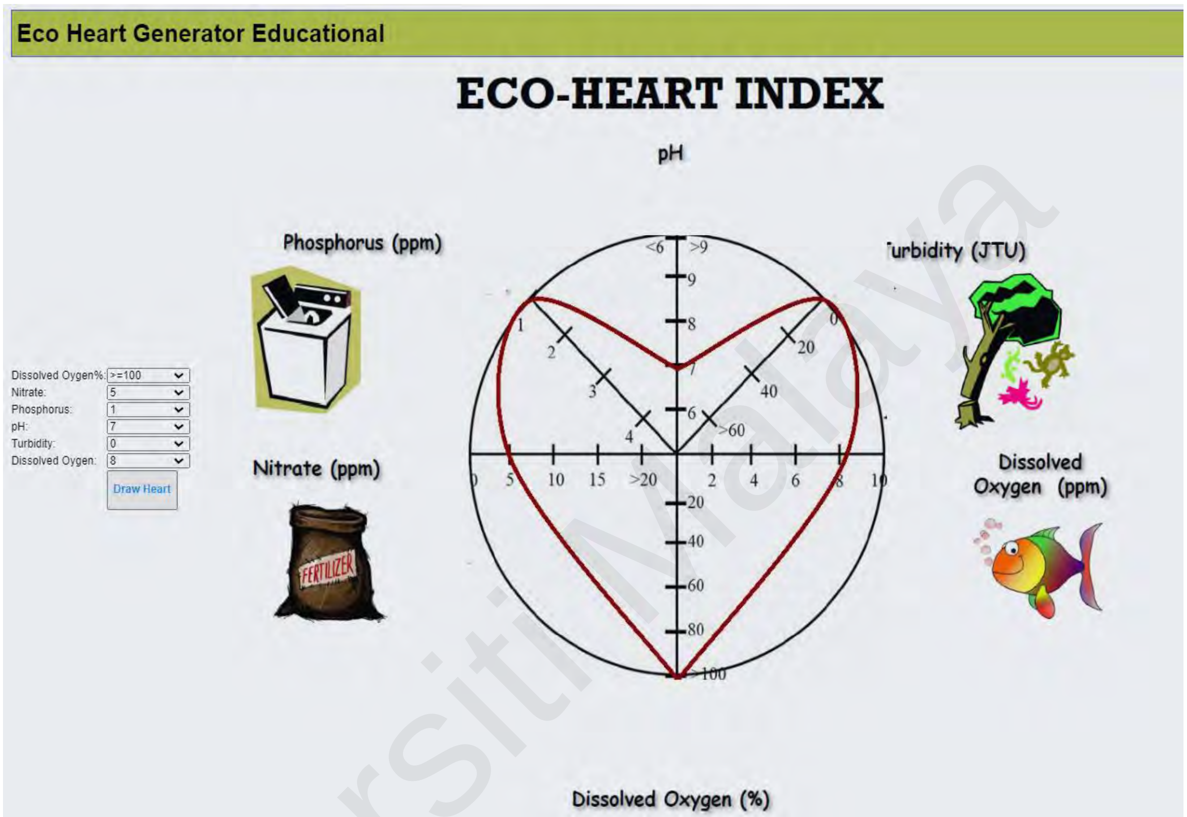


Figure 4.12: Educational Eco-Heart Generator that allows the user to have a clearer idea about the Eco-Heart Index.

A multiple station report page is developed as shown in figure 4.13. This page allows users to view the graphical statistics report of multiple sampling points and compare the values of different variables. The water quality variables included are paraquat, ammoniacal nitrogen, biochemical oxygen demand, chemical oxygen demand, dissolved oxygen, pH, color, electrical conductivity, floatables, ordour, salinity, taste, total dissolved solids, total suspended solids, temperature, turbidity, faecal coliform and total coliform. Besides, the data can present in yearly, monthly and weekly ranges, depending on the user preferences. Figure 4.13 shows the selected date from 29/01/2010 until 29/01/2022, two of the sampling points 1L01 and 1L02 from Langat River was selected. The line graph shows the variables of ammoniacal nitrogen in yearly trends.

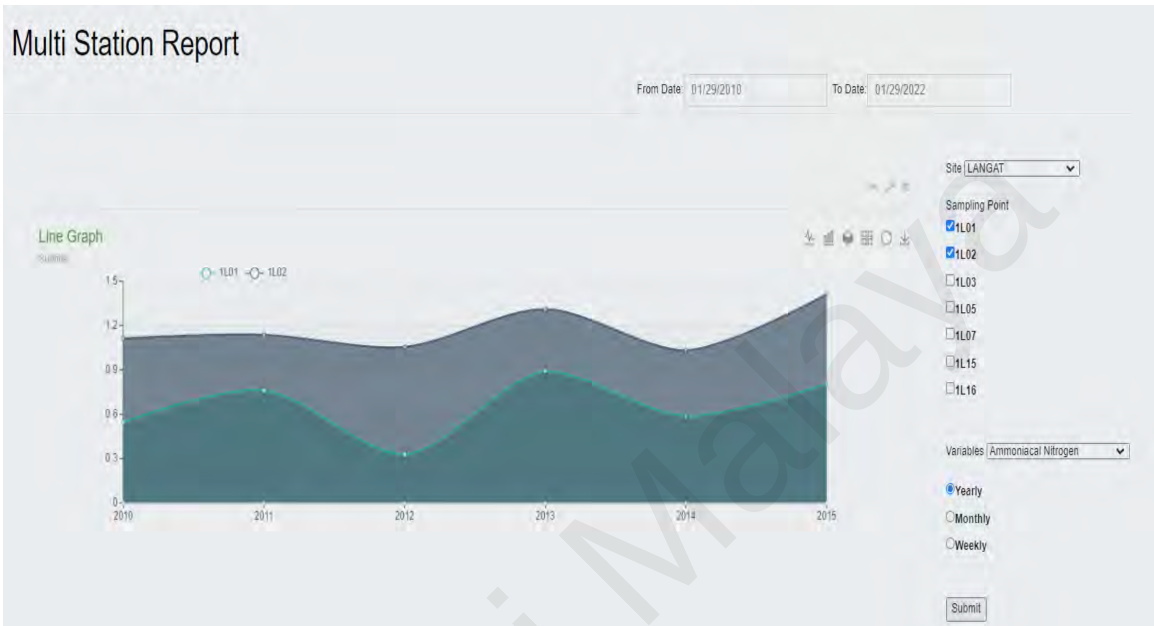


Figure 4.13: Water Quality Index Multiple Station Report Generator.

The Multi-Variable Report (figure 4.14) page allows users to select multiple water quality variables and represents in a graphical format, where it is more easier for user to study the data and the correlation between the variables.

Figure 4.14: Water Quality Index Multiple Variable Report Generator.

In this study we have introduced effective data visualization facilitates inference and decision-making, as well as the clear and effective transmission of information via graphical means that are not readily visible in numerical summaries. Data visualization is important because it permits the clear and effective conveyance of information via graphical means. It's helpful in assessing information that's important for making decisions about the community's endangered and threatened species. The graphical analysis makes environmental data more understandable and enables finding trends, patterns, and outliers easier. Patterns, trends, and other qualities that are not readily visible from numerical summaries require graphical analysis. To visualize multivariate and spatiotemporal ecological data, effective visual data mining approaches are required so that patterns and relationships may be easily observed. Geometric, glyph or icon-based, pixel-oriented, and structured ways for visualizing multivariate data seem to be the most frequent (Schroeder 2005).

In order to analyse vast volumes of biodiversity data, we also combine human vision, creativity, and general knowledge with the storage capacity and processing power of current computers. In this study, this is accomplished by showing the user graphical representations of the data, which allows the user to interact with the data to gain insight and efficiently form conclusions (Keim 2005). Additionally, the ability to depict relationships and patterns revealed in complicated ecological data is critical. To accomplish this, we used data gathered through citizen science, as well as established biodiversity indexes, and the visualization capabilities of the Google Earth platform to accomplish our research.

4.5 System Usability Testing (SUS)

The evaluation form is created based on the System Usability Scale (SUS) which comprises 10 questions to assess the usability and functionality of the website. The system usability evaluation form is given in the form of Google Forms to several users that are related to the research, such as; the Google Form question can refer to Appendix III.

- Biologist
- Environmentalist
- Researchers
- Students

The scores are then converted into numbers and calculate the usability score using SUS. The outcome of the calculated score is average at the SUS score of 80, consider as "Good" according to the SUS score shown in figure 4.48, which falls under the acceptability range. From the result indicates that the system is acceptable and could be used by environmentalist researchers.

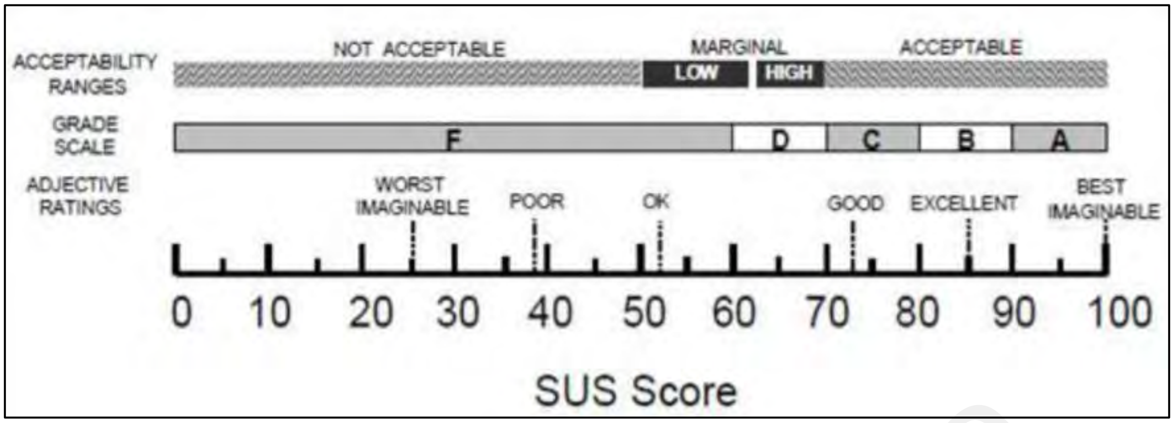


Figure 4.15: SUS scores Grade Rankings from "Determining what individual SUS scores mean: Adding an adjective rating scale." (Bangor, Kortum, & Miller, 2009)

CHAPTER 5: CONCLUSION

5.1 Problems Encountered and Recommendation Solution

The following are some of the problems encountered during the research project analysis and development.

I. Confusion in choosing biodiversity indexes

There are many biodiversity indexes available to measure the biodiversity. Confusion on which biodiversity index should be incorporated into the system. By reading pass research studies and interview some experts from this field, the most commonly used indexes are the Shannon-Wiener index and Simpson's index. In addition, Hill Numbers were added into the system as environmentalist and ecologist will able to gain more information from the calculations. Therefore, only these three indexes are incorporated into the system.

II. Limited Plant Species Images

At first, there are only 10 images provided. However, the image can be included from time to time and images can be retrieved from iNaturalist automatically.

5.2 Ecological Web Visualization System Strengths

The following features illustrate the strength of the biodiversity web module.

I. **Simple and user-friendly interface**

The website is designed to simplify the instruction and operability of the website. A side menu is included so that the users can navigate to different page at ease. Simplistic is the basic concept of designing the interface, description is added to for those that are not proficient in using the website.

II. **Manageable polygon and plant species database**

The information of polygons and plant species can be added, edited and deleted by the users including the name of polygons, and the plant species details. It encourages users' participation in managing the data from time to time and expanding the database, and more and more relevant information can be retrieved from the website making the website more comprehensive and complete.

III. **Functionality**

The system is user friendly where the website is design in a sequential fashion, start with managing the polygon and plant species data, follow by visualization of the plant data on the map. This website is designed and planned according to the users, especially biologist and ecologist to meet their requirements. Besides, this website can be access on various browsers such as Google Chrome, Microsoft EDGE, Internet Explorer, Mozilla Firefox and others.

IV. **Interactive Map**

The 'heart' shape of WQI appear on the Google Map, where user obtain more information when click the 'heart' shape, a float window contains the information will appear. Same goes to, the biodiversity indexes, where the polygon is colored according to the species density and the species is represent in a marker, when the marker is clicked, float window that contains the species information will be shown to the user.

5.3 Future enhancements

There are several enhancements can be considered in the future to increase the usability of the website.

The recommended enhancements are described in the following:

I. Picture Glossary

A picture glossary is strongly recommended to be added into the web module, since not everyone able to recognize the species name, especially the scientific names. A page with all the plant species information together with plant image, will help users in identifying the plants.

II. Increase the biodiversity indexes

There are much more biodiversity indexes that has not been included in the module, if there is any request for it, it still can be embed on the web module as well.

III. Increase the flexibility of the map

The map can be enhanced, where the users can draw the polygon on the area of interest and show the trees on the map. Besides, screen capture function can be added, where users can save the image of the map.

IV. Carbon Emission Calculator

The web module can further expand by adding a carbon emission calculator, since the data regarding the number of plants is available in the database.

V. **Increase in Green Space Area Coverage**

VI. Presently, the area of research is limited to the University of Malaya campus; however, the web application may be expanded to provide a ready-made solution for rapid assessment of biodiversity state, which may be valuable for policy decisions.Expand in Animals Species Data

The research study focuses more on plant species than on animal species; therefore, future enhancements will be able to add data on animal species into the online application. This will be very helpful in the study of animal migration and resident species.

5.4 Overall Conclusion

This research developed a web-based ecological system with an emphasis on the display of biodiversity indices and water quality indexes, utilizing Google Earth and GIS technology to visualize species topology and water quality status as determined by the WQI value. The use of colour, form, and marking is advantageous since it facilitates the interpretation of data and comprehension for non-expert users. The public can participate in environmental monitoring and management. It is beneficial because it enables a straightforward understanding of species distribution and diversity patterns and enables more informed judgments about conserving and protecting the ecosystem, implying less damage. This project focuses on the University of Malaya campus, and with the data available, this system seeks to capture and manage further hydrological and ecological data in the future, educating and empowering the community about the value of conservation. Additionally, to provide a deeper knowledge and appreciation of the various patterns and to indirectly educate the community about the importance of environmental conservation. Below is the link to the ecological visualization system - http://umlivinglabsystem.com/.

REFERENCES

Abdullah, M., & Shobri, N. I. (2018). Putrajaya Lakes management: sensitivity and conservation towards sustainable community. *3rd International Conference on Rebuilding Place (ICRP2018),* (pp.94-108). Malaysian Journal of Sustainable Environment (MySE).

Acorn, J. H. (2017). Entomological citizen science in Canada. *The Canadian Entomologist, 149*(6), 774-785.

Adelagun, R. O., Etim, E. E., & Godwin, O. E. (2021). Application of Water Quality Index for the Assessment of Water from Different Sources in Nigeria. *Promising Techniques for Wastewater Treatment and Water Quality Assessment*, *267*(1), 978-986.

Al-Shujairi, S. O. (2013). Develop and apply water quality index to evaluate water quality of Tigris and Euphrates Rivers in Iraq. *International Journal of Modern Engineering Research, 4*(1), 2119-2126.

Amini, A. T. (2009). Adjustment of peak streamflows of a tropical river for urbanization. *American Journal of Environmental Sciences, 3*(2), 285-286.

Anachebelu, T. K. (2015). Application of water quality index for pollution detection at Luton Hoo Lake. *2015 IEEE SENSORS* (pp. 1-4), South Korea: IEEE.

Andrew D. (1989). *Standard methods for the examination of water and wastewater.* Washington, US: American Public Health Association.

Atkins, D. E. (2003). *Revolutionizing science and engineering through cyberinfrastructure: Report of the National Science Foundation blue-ribbon advisory panel on cyberinfrastructure*. Arizona, US: National Science Foundation.

Attorre, F., Taleb, N., De Sanctis, M., Farcomeni, A., Guillet, A., & Vitale, M. (2011). Developing conservation strategies for endemic tree species when faced with time and data constraints: Boswellia spp. on Socotra (Yemen). *Biodiversity and Conservation , 20*(7), 1483-1499.

Avvannavar, S. M., & Shrihari, S. J. E. M. (2008). Evaluation of water quality index for drinking purposes for river Netravathi, Mangalore, South India. *Environmental monitoring and assessment, 143*(1), 279-290.

AZA. (2021). *Frog Watch.* Retrieved on 21 May 2021 from https://www.aza.org/frogwatch?locale=en

AZA. (2021). *Nest Watch*. Retrieved on 21 May 2021 from https://www.aza.org/nestwatch

Azizi Abu Bakar, A. N. (2013). *Development of Eco-Heart Index for Water Quality Awareness by Riverine Community*. Presented in University of Malaya Researchers' Conference 2013, 11-13 November 2013, Kuala Lumpur, Malaysia.

Azizi, A., Nik Meriem, N., Noor Zalina, M., & Azizan, B. (2015). *Development of eco-heart indicators for water quality status and riverine community awareness.* Presented in 18th International River Symposium, 21-23 September, Brisbane, Autralia.

Ballard, H., Phillips, T., & Robinson, L. (2018). *Conservation outcomes of citizen science,* London, UK: UCL Press.

Beals, M., Gross, L., & Harrell, S. (2000). DIVERSITY INDICES: SHANNON'S: H AND E. Tennessee, US: QUBES Educational Resources

Berlemann, A. (2013). Using a water quality index to determine and compare creek water quality. *Journal-American Water Works Association, 105*(6), 291-298.

Best, B. D., Halpin, P. N., Fujioka, E., Read, A. J., Qian, S. S., Hazen, L. J., & Schick, R. S. (2007). Geospatial web services within a scientific workflow: Predicting marine mammal habitats in a dynamic environment. *Ecological Informatics, 2*(3), 210-223.

Bhat, A. (2018). *What is System Usability Scale?* Retrieved on 13 May 2020 from https://www.questionpro.com/blog/system-usability-scale/

Boyacioglu, H. (2007). Development of a water quality index based on a European classification scheme. *Water Sa*, *33*(1), 71-72.

Brewer, C. (1994). Colour Use Guidelines for Mapping. *Visualization in Modern Cartography* (pp. 123-148). New York: Elsevier Science.

Brooke, J. (1996). Sus: a "quick and dirty'usability. *Usability evaluation in industry* (pp. 189). Florida: CRC Press.

Burbrink, F. T., Pyron, & A., R. (2010). How does ecological opportunity influence rates of speciation, extinction, and morphological diversification in New World ratsnakes (tribe Lampropeltini)? *Evolution: International Journal of Organic Evolution, 64*(4), 934-943.

Burgess, H., DeBey, L., Froehlich, H., Schmidt, N., Theobald, E., Ettinger, A., Parrish, J. (2017). The science of citizen science: Exploring barriers to use as a primary research tool. *Biological Conservation, 208*(1), 113-120.

Burrough, P. A. (1986). *Principles of geographical. Information systems for land resource assessment.* Oxford, UK: Clarendon Press.

Canfield, D. E., C. D. Brown, Bachmann, R. W., & Hoyer, M. V. (2002). Volunteer lake monitoring: Testing the reliability of data collected by the Florida LAKEWATCH program. *Lake and Reservoir Management, 18*(1), 1–9.

Casimiro, S. (2020). *The Species Identifying App 'Seek' Will Change Your Life.* Retrieved on 10 Jan 2021 from: https://www.adventure-journal.com/2020/05/the-species-identifying-app-seek-will-change-your-life

Catlin-Groves, C. L. (2012). The citizen science landscape: from volunteers to citizen sensors and beyond. *International Journal of Zoology*, *2012, 10*(1)*,* 1155-1156.

Charlie, L. S. (2010). *Assessment of water quality characteristics during base and storm flow events on Sungai Langat basin.* (PhD's thesis). Universiti Teknologi Malaysia, Johor, Malaysia.

Choe, E., van der Meer, F., van Ruitenbeek, F., van der Werff, H., de Smeth, B., & Kim, K. W. (2008). Mapping of heavy metal pollution in stream sediments using combined geochemistry, field spectroscopy, and hyperspectral remote sensing: A case study of the Rodalquilar mining area, SE Spain. *Remote Sensing of Environment, 112*(7), 3222-3233.

Chrisman, N. R. (1999). What does 'GIS'mean? *Transactions in GIS, 3*(2), 175-186.

Clarke, K. R., Gorley, R. N., Somerfield, P. J., & & Warwick, R. M. (2014). Change in marine communities: an approach to statistical analysis and interpretation. *Primer-E Ltd*, *8*(2), 1-68.

Cleveland, W. S. (1984). Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American statistical association, 79*(387), 531-554.

Cohn, J. P. (2008). Citizen science: Can volunteers do real research? *BioScience, 58*(3), 192-197.

Colwell, & K., R. (2009). Biodiversity: concepts, patterns, and measurement. *The Princeton guide to ecology, 663*(1), 257-263.

Cooper, C., Dickinson, J., Phillips, T., & Bonney, R. (2007). Citizen science as a tool for conservation in residential ecosystems. *Ecol. Soc., 12*(2), 11.

Cooper, P. F., & Findlater, B. C. (2013). *Constructed Wetlands in Water Pollution Control: Proceedings of the International Conference on the Use of Constructed Wetlands in Water Pollution Control.* Cambridge, UK: Elsevier.

Mandil, C. (2006). Light's labour's lost: policies for energy-efficient lighting. *Energy world*, *1*(343), 14-15.

Cosquer, A., Raymond, R., & Prevot-Julliard, A. (2012). Observations of everyday biodiversity: a new perspective for conservation? *Ecol. Soc. 17*(4), 2-17.

Council, N. R. (2011). *Advancing the science of climate change.* Washington, US: National Academies Press.

Cox, S., Cuthbert, A., Daisey, P., Davidson, J., Johnson, S., Keighan, E., & Vretanos, P. P. (2002). *Opengis® geography markup language (gml) implementation specification, version 1.* Arlington, US: Open Geospatial Consortium Inc.

Curbera, F., Duftler, M., Khalaf, R., Nagy, W., Mukhi, N., & Weerawarana, S. (2002). Unraveling the Web services web: an introduction to SOAP, WSDL, and UDDI. *IEEE Internet computing, 6*(2), 86-93.

Danielsen, F., Adrian, T., Brofeldt, S., Van Noordwijk, M., Poulsen, M., Rahayu, S., . . . Bang, T. (2013). Community Monitoring for REDD+ International Promises and Field Realities. *Ecol. Soc., 18*(3), 28-56.

Dasmann, & R. F. (1985). Achieving the sustainable use of species and ecosystems. *Landscape Planning, 12*(3), 211-219.

Dasmann, R. F. (1968). *Different kind of country*. New York, US: Macmillan

de Sherbinin, A., Bowser, A., Chuang, T., Cooper, C., Danielsen, F., Edmunds, R., . . . Popescu, I. (2021). The critical importance of citizen science data. *Frontiers in Climate, 3*(1), 20-21.

De Vries, P. (2011). The resilience principles: A framework for new ICT governance. *J. on Telecomm. & High Tech. L.*, *9*(3), 137-138.

Developers, G. (2015). *Introduction to the Google Maps SDK for iOS.* Retrieved on 2 Jan 2021 from https://developers.google.com/maps/documentation/ios/intro

DOE. (1999). *Malaysia Environnmental Quality Reports*. Ministry of Science, Technology and Environment, Kuala Lumpur, Malaysia.

DOE. (2014). *Malaysia Environmental Quality Reports*. Ministry of Science, Technology and Environment, Kuala Lumpur, Malaysia.

Droege, S. (2007, June). Just because you paid them doesn't mean their data are better. In *Citizen Science Toolkit Conference* (pp. 13-26), Cornell Laboratory of Ornithology.

Drollette, D. (2012, March 29). *Citizen science enters a new era*. Retrieved on 5 Jan 2021 from: http://www.bbc.com/future/story/20120329-citizen-science-enters-a-new-era?selectorSection=technology

Ebrahimpour, M., & Mushrifah, I. (2009). Variation and correlations of selected heavy metals in sediment and aquatic plants in Tasik Chini, Malaysia. *Environmental geology, 57*(4), 823-831.

Echo, E. (2021). *EARTHECHO WATER CHALLENGE*. Retrieved on 1 May 2021 from https://www.monitorwater.org/

Edwards, J. L. (2004). Research and societal benefits of the Global Biodiversity Information Facility. *BioScience, 54*(6), 485-486.

Ehrenfeld, D. W. (1970). *Biological Conservation.* New York, US: Holt, Rinehart and Winiston.

Elliott, K. C., & Rosenberg, J. (2019). Philosophical foundations for citizen science. *Citizen Science: Theory and Practice*, *4*(1), 34-38.

ESRI. (n.d.). *Coordinate systems, projections, and transformations.* Retrieved on 15 November 2019 from https://pro.arcgis.com/en/pro-app/help/mapping/properties/coordinate-systems-and-projections.htm.

Evans, C., Abrams, E., Reitsma, R., Roux, K., Salmonsen, L., & Marra, P. P. (2005). The Neighborhood Nestwatch Program: Participant outcomes of a citizen-science ecological research project. *Conservation Biology, 19*(3), 589-594.

Fidan, A. F., Ciğerci, İ. H., Konuk, M., Küçükkurt, İ., Aslan, R., & Dündar, Y. (2008). Determination of some heavy metal levels and oxidative status in Carassius

carassius L., 1758 from Eber Lake. *Environmental monitoring and assessment, 147*(1), 35-41.

Flagg, B. (2016). Contribution of multimedia to girls' experience of citizen science. *Citizen Science: Theory and Practice, 1*(2), 3-4.

Florida, U. o. (2021). *Lakewatch.* Retrieved on 10 Jan 2021 from https://lakewatch.ifas.ufl.edu/

Ford-Lloyd, B., Schmidt, M., Armstrong, S., Barazani, O., Engels, J., Hadas, R., . . . Li, Y. (2011). Crop wild relatives—undervalued, underutilized and under threat? *Bioscience, 61*(7), 559-565.

Foster, I. (2005). Service-oriented science. *Science, 308*(5723), 814-817.

Frehner, M., & Brändli, M. (2006). Virtual database: Spatial analysis in a Web-based data management system for distributed ecological data. *Environmental Modelling & Software, 21*(11), 1544-1554.

Gandaseca, S., Rosli, N., Ngayop, J., & Arianto, C. I. (2011). Status of water quality based on the physico-chemical assessment on river water at Wildlife Sanctuary Sibuti Mangrove Forest, Miri Sarawak. *American Journal of Environmental Sciences*, *7*(3), 21-22.

Garden, C. B. (2021). *Budburst*. Retrieved on 1 May 2021 from https://budburst.org/

Geoghegan, H., Dyke, A., Pateman, R., West, S., & Everett, G. (2016). Understanding motivations for citizen science. *Final report on behalf of UKEOF, University of Reading, Stockholm Environment Institute (University of York) and University of the West of England, 78*, Article#4896.

Geographic, N. (2019). *A Guide to BioBlitz: for afterschool programs.* Washington, US: National Geographic.

Gilby, I. C., Eberly, L. E., Pintea, L., & Pusey, A. E. (2006). Ecological and social influences on the hunting behaviour of wild chimpanzees, Pan troglodytes schweinfurthii. *Animal Behaviour, 72*(1), 169-180.

Gillis, A. S. (2020). *REST API (RESTful API)*. Retrieved on 12 May 2020 from https://searchapparchitecture.techtarget.com/definition/RESTful-API

Gommerman, L., & Monroe, M. C. (2012). Lessons learned from evaluations of citizen science programs. *EDIS, 201*(6), 158-159.

Goodall, J. L., Horsburgh, J. S., Whiteaker, T. L., Maidment, D. R., & Zaslavsky, I. (2008). A first approach to web services for the National Water Information System. *Environmental Modelling & Software, 23*(4), 404-411.

Gotelli, N. J., & Colwell, R. K. (2001). Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. *Ecology letters, 4*(4), 379-391.

Graham, A. J., Atkinson, P. M., & Danson, F. M. (2004). Spatial analysis for epidemiology. *Acta tropica*, *91*(3), 219-225.

Gray, S., Jordan, R., Crall, A., Newman, G., Hmelo-Silver, C., Huang, J., . . . Singer, A. (2017). Combining participatory modelling and citizen science to support volunteer conservation action. *Biological conservation, 208*(1), 76-86.

Haklay, M. M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., & Vohland, K. (2021). What is citizen science? The challenges of definition. *The Science of Citizen Science*, *13*(1), 34-51.

Hart, B. T. (1982). Uptake of trace metals by sediments and suspended particulates: a review. *Sediment/freshwater interaction*, *2*(1), 299-313.

Henderson, S., Ward, D., Meymaris, K., Alaback, P., & Havens, K. (2012). Project Budburst: Citizen science for all seasons. and Bonney R., editor.[eds.]. *Citizen science: Public participation in environmental research, 15*(3), 50-57.

Hey, T., & Trefethen, A. (2005). Cyberinfrastructure for e-science. *Science, 308*(2), 817-821.

Hill, M. O. (1973). Diversity and evenness: a unifying notation and its consequences. *Ecology*, *54*(2), 427-432.

iNaturalist. (2020). *iNaturalist*. Retrieved on 5 May 2020 from iNaturalist: https://www.inaturalist.org/

Jaafar, O., Mastura, S. A., & Sood, A. M. (2009). Land use and deforestation modelling of river catchments in Klang Valley, Malaysia. *Sains Malaysiana*, *38*(5), 655-664.

Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. I., Armi, A. S., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental monitoring and assessment*, *173*(1), 625-641.

Kannan, S. K., Batvari, B. P., Devarajan, P., Periakali, P., Krishnamoorthy, R., Rao, N. R., & Jayaprakash, M. (2007). A preliminary study on herbal resources and their contamination with toxic heavy metals around Pulicat Lake, North Chennai, South East Coast of India. *Res J Environ Tox, 1*(2), 78-84.

Kattge, J., Diaz, S., Lavorel, S., Prentice, I., Leadley, P., Bönisch, G., . . . Cornelissen, J. (2011). TRY–a global database of plant traits. *Global change biology, 17*(9), 2905-2935.

Keim, D. A. (2005). Information visualization: Scope, techniques and opportunities for geovisualization. In *Exploring geovisualization* (pp. 21-52), Elsevier.

Kellert, S. R. (1995). *The biophilia hypothesis.* Washington, US: Island Press.

Kira, T. (1988). International Lake Environment Committee (ILEC). *Environmental Conservation, 15*(3), 277-278.

Kyba, C. C., Ruhtz, T. H. O. M. A. S., Fischer, J., & Hölker, F. (2011). Lunar skylight polarization signal polluted by urban lighting. *Journal of Geophysical Research: Atmospheres*, *116*(24), 334-356.

Kyba, C., Wagner, J., Kuechly, H., Walker, C., Elvidge, C., Falchi, F., . . . Hölker, F. (2013). Citizen science provides valuable data for monitoring global night sky luminance. *Scientific reports, 3*(1), 1-6.

Lane, M. A. (2006). Information infrastructure for global biological networks. *Microbiology Australia, 27*(1), 23-25.

Laurila-Pant, M., Lehikoinen, A., Uusitalo, L., & Venesjärvi, R. (2015). How to value biodiversity in environmental management?. *Ecological indicators*, *55*(1), 1-11.

Leščešen I, P. M. (2015). Statistical Analysis of Water Quality Parameters of the Drina River (West Serbia). *Polish Journal of Environmental Studies*, *24*(3), 8-12.

Lewenstein, B. V. (2016). 'Can we understand citizen science?'. *JCOM, 15*(01), 23-26.

Li, X., & Thornton, I. (2001). Chemical partitioning of trace and major elements in soils contaminated by mining and smelting activities. *Applied geochemistry, 16*(15), 1693-1706.

Maguire, D. J. (1991). An overview and definition of GIS. *Geographical information systems: Principles and applications, 1*(2), 9-20.

Magurran, A. E. (2005). Species abundance distributions: pattern or process? *Functional Ecology, 19*(1), 177-181.

Malaya, U. o. (2013). *The RIMBA Project*. Retrieved on 4 April 2020 from https://www.um.edu.my/the-rimba-project

Masse, M. (2011). *REST API design rulebook: designing consistent RESTful web service interfaces*. California, US: O'Reilly Media, Inc.

Matheson, C. A. (2014). iNaturalist. *Reference Reviews*.

McGranaghan, M. (1993). A cartographic view of spatial data quality. *Cartographica: The International Journal for Geographic Information and Geovisualization*, *30*(2-3), 8-19.

McGuire, M., Gangopadhyay, A., Komlodi, A., & Swan, C. (2008). A user-centered design for a spatial data warehouse for data exploration in environmental research. *Ecological Informatics, 3*(4-5), 273-285.

Meher, P. K., Sharma, P., Gautam, Y. P., Kumar, A., & Mishra, K. P. (2015). Evaluation of Water Quality of Ganges River Using Water Quality Index Tool. *EnvironmentAsia*, *8*(1), 8-9.

Microsoft. (n.d.). *Graphics.FillClosedCurve Method.* Retrieved 15 November 2019 from https://docs.microsoft.com/en-us/dotnet/api/system.drawing.graphics.fillclosedcurve?view=netframework-4.8

Mineter, M. J., Jarvis, C. H., & Dowers, S. (2003). From stand-alone programs towards grid-aware services and components: a case study in agricultural modelling with interpolated climate data. *Environmental Modelling & Software, 18*(4), 379-391.

Mitchell, A. (1999). *The ESRI guide to GIS analysis: geographic patterns & relationships* (Vols. 1). California, US: ESRI, Inc.

Moore, J.C. (2013). Diversity, Taxonomic versus Functional. *Encyclopedia of Biodiversity*, *13*(14), 648-656.

Moore, M. V., Pierce, S. M., Walsh, H. M., Kvalvik, S. K., & Lim, J. D. (2000). Urban light pollution alters the diel vertical migration of Daphnia. . *Internationale Vereinigung für theoretische und angewandte Limnologie: Verhandlungen, 27*(2), 779-782.

Murray, M. G. (2002). *Current issues in biodiversity conservation.* Rome, Italy: FAO.

MySQL. (2016). *MySQL 8.0 Release Notes*. Retrieved on 9 December 2020 from https://dev.mysql.com/doc/relnotes/mysql/8.0/en/

NAHRIM. (2021). *Mylake.* Retrieved on 23 May 2021 from https://mylake.nahrim.gov.my

Nascimento, S., Rubio Iglesias, J., Owen, R., Schade, S., & Shanley, L. (2018). *Citizen science for policy formulation and implementation*. London, UK: UCL Press.

NOIRLab. (2021). *Globe at Night.* Retrieved on 5 May 2021 from https://www.globeatnight.org/

Nugent, J. (2020). EarthEcho Water Challenge. *The Science Teacher, 88*(1), 22-23.

Oberhauser, K., & Prysby, M. D. (2008). Citizen science: creating a research army for conservation. *American Entomologist, 54*(2), 103-104.

Okuno, E., Gardner, R. C., Archabal, M., Murphy, K., & Willis, J. (2017). Bibliography of 2016 Scientific Publications on the Ramsar Convention or Ramsar Sites. *Available at SSRN 3063547, 3*(2), 1-82.

Owen, R. P., & Parker, A. J. (2018). *Citizen science in environmental protection agencies.* London, UK: UCL Press.

Parker, S., Pauly, G., Moore, J., Fraga, N., Knapp, J., Principe, Z., . . . Wake, T. (2018). Adapting the bioblitz to meet conservation needs. *Conservation biology, 32*(5), 1007-1019.

Piasecki, M., & Bermudez, L. (2003). Hydroml: Conceptual development of a hydrologic markup language. *IAHR Congress.*(pp. 1-12), Thessaloniki Greece.

Pillai, S., Silventoinen, V., Kallio, K., Senger, M., Sobhany, S., Tate, J., ... & Lopez, R. (2005). SOAP-based services provided by the European Bioinformatics Institute. *Nucleic acids research*, *33*(2), 25-28.

Pyron, R. A., & Burbrink, F. T. (2010). Hard and soft allopatry: physically and ecologically mediated modes of geographic speciation. *Journal of Biogeography, 37*(10), 2005-2015.

Ramachandran, R., Graves, S., Conover, H., & Moe, K. (2004). Earth Science Markup Language (ESML):: a solution for scientific data-application interoperability problem. *Computers & Geosciences, 30*(1), 117-124.

Reece, J. B., Urry, L. A., Cain, M. L., Wasserman, S. A., Minorsky, P. V., & Jackson, R. B. (2014). *Campbell biology.* Boston, US: Pearson.

Rhind, D. (1989). Why GIS. . *Arc News, 11(3)*, 28-29.

Rich, C., & Longcore, T. (2006). *Ecological Consequences of Artificial Night Lighting* . Washington, US: Island Press.

Rinkesh K (2016) *What is a green building? Green & Sustainable Business.* Retrieved on 22 May 2020 from http://www.businessfeed.sunpower.com

Rungta, K. (2021, July 9). *Guru99*. Retrieved on 2 May 2021 from https://www.guru99.com/test-case.html

Ryanzhin, S. V., Straskraba, M., & Geller, W. (2001). Developing WORLDLAKE–Database and GIS for limnological studies. *Ninth International Conference on the Conservation and Management of Lakes, Otsu, Japan*, (pp. 10-15), Limnology Institute of RAS.

Said, A., Stevens, D. K., & Sehlke, G. (2004). An innovative index for evaluating water quality in streams. *Environmental management*, *34*(3), 406-414.

Sarkar, C., & Abbasi, S. A. (2006). QUALIDEX–a new software for generating water quality indice. *Environmental monitoring and assessment*, *119*(1), 201-231.

Schroeder, M. (2005). Intelligent information integration: from infrastructure through consistency management to information visualization. *Exploring geovisualization*, *4*(5) 477-494.

Scott, J. M., Csuti, B., Jacobi, J. D., & Estes, J. E. (1987). Species richness. *BioScience, 37*(11), 782-788.

Seders, L. A., Shea, C. A., Lemmon, M. D., Maurice, P. A., & Talley, J. W. (2007). LakeNet: an integrated sensor network for environmental sensing in lakes. *Environmental Engineering Science, 24*(2), 183-191.

Seltzer, C. (2019). Making Biodiversity Data Social, Shareable, and Scalable: Reflections on iNaturalist & citizen science. *Biodiversity Information Science and Standards*, (pp 1-8), California Academy of Sciences.

Shannon, C. E. (1949). Communication theory of secrecy systems. *The Bell system technical journal, 28*(4), 656-715.

Short, J., Bradshaw, S. D., Giles, J., Prince, R. I., & Wilson, G. R. (1992). Reintroduction of macropods (Marsupialia: Macropodoidea) in Australia—a review. *Biological conservation, 62*(3), 189-204.

Shuhaimi-Othman, M., Lim, E. C., & Mushrifah, I. (2007). Water quality changes in Chini lake, Pahang, west Malaysia. *Environmental monitoring and assessment, 131*(1), 279-292.

Shuhaimi-Othman, M., Mushrifah, I., Lim, E. C., & Ahmad, A. (2008). Trend in metals variation in tasik Chini, Pahang, Peninsular Malaysia. *Environmental monitoring and assessment, 143*(1), 345-354.

Silvertown, J. (2009). A new dawn for citizen science. *Trends in ecology & evolution, 24*(9), 467-471.

Simpson, E. H. (1949). Measurement of diversity. *nature, 163*(4148), 688-689.

Simpson, R., Page, K. R., & De Roure, D. (2014). Zooniverse: observing the world's largest citizen science platform. *23rd international conference on world wide web*, (pp. 1049-1054), WWW '14 Companion.

Smallman, M. (2018). *Citizen science and responsible research and innovation*. London, UK: UCL Press.

Society, N. G. (2012). *source*. Retrieved on 13 May 2021 from National Geographic Society: https://www.nationalgeographic.org/encyclopedia/source/

Soulé, & E., M. (1985). What is conservation biology? *BioScience 35*(11), 727-734.

Soylak, M., & Erdogan, N. D. (2006). Copper (II)–rubeanic acid coprecipitation system for separation–preconcentration of trace metal ions in environmental samples for their flame atomic absorption spectrometric determinations. *Journal of Hazardous Materials*, *137*(2), 1035-1041.

Spellerberg, I. F., & Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener'Index. *Global ecology and biogeography, 12*(3), 177-179.

Stein, B. R., & Wieczorek, J. R. (2004). Mammals of the world: MaNIS as an example of data integration in a distributed network environment. *Biodiversity Informatics 1*. (pp. 12-22) , National Science Foundation.

Stein, L. (2002). Creating a bioinformatics nation. *Nature, 417*(6885), 119-120.

Stevenson, R. D., Suomela, T., Kim, H., & He, Y. (2021). Seven Primary Data Types in Citizen Science Determine Data Quality Requirements and Methods. *Frontiers in Climate*, (pp. 49-50), Frontiers.

Sugawara, H., & Miyazaki, S. (2003). Biological SOAP servers and web services provided by the public sequence data bank. *Nucleic acids research, 31*(13), 3836-3839.

Tarras-Wahlberg, N. H., Flachier, A., Lane, S. N., & Sangfors, O. (2001). Environmental impacts and metal exposure of aquatic ecosystems in rivers contaminated by small scale gold mining: the Puyango River basin, southern Ecuador. *Science of the Total Environment, 278*(1), 239-261.

Thelen, B. A., & Thiet, R. K. (2008). Cultivating connection: Incorporating meaningful citizen science into Cape Cod National Seashore's estuarine research and monitoring program. *Park Science. 25*(1), 74-80.

Theobald, E., Ettinger, A., Burgess, H., DeBey, L., Schmidt, N., Froehlich, H., . . . Parrish, J. (2015). Global change and local solutions: Tapping the unrealized potential of citizen science for biodiversity research. *Biological Conservation, 181*(2), 236-244.

Toogood, M. (2013). Engaging publics: biodiversity data collection and the geographies of citizen science. *Geography Compass 7.9 (2013)*, *6*(2)*,* 611-621.

University, I. S. (2021). *Bug Guide* Retrieved on 13 May 2021 from BugGuide: https://bugguide.net

Usali, N., & Ismail, M. H. (2010). Use of remote sensing and GIS in monitoring water quality. *Journal of Sustainable Development, 3*(3), 228-229.

Walker, C. E., Pompea, S. M., & Isbell, D. (2008). GLOBE at night 2.0: On the road toward IYA 2009. *EPO and a Changing World: Creating Linkages and Expanding Partnerships* (pp. 423-424), Astronomical Society of the Pacific.

Whittaker, R. H. (1965). Dominance and diversity in land plant communities: numerical relations of species express the importance of competition in community function and evolution. *Science, 147*(3655), 250-260.

Wiggins, A., & Crowston, K. (2011). From conservation to crowdsourcing: A typology of citizen science. *44th Hawaii international conference on system sciences* (pp. 1-10). Hawaii: IEEE.

Wiggins, A., & Crowston, K. (2015). Surveying the citizen science landscape. *First Monday*, *20*(1)*,* 1-5.

Wilcox, B. A. (1980). Insular ecology and conservation. 6. *Insular ecology and conservation*, *1*(5), 95-117.

WorldLakes. (2021). *World Lakes.* Retrieved on 13 May 2021 from World Lakes Website: http://www.worldlakes.org/

Xu, B., Lin, H., Chiu, L., Tang, S., Cheung, J., Hu, Y., & Zeng, L. (2010). VGE-CUGrid: An Integrated Platform for Efficient Configuration, Computation, and Visualization of MM5. *Environmental Modelling & Software.*, *3*(4) 1894-1896.

Yang, Q., Li, Z., Lu, X., Duan, Q., Huang, L., & Bi, J. (2018). A review of soil heavy metal pollution from industrial and agricultural regions in China: Pollution and risk assessment. *Science of the total environment, 642*(5), 690-700.

Yeom, D. J., & Kim, J. H. (2011). Comparative evaluation of species diversity indices in the natural deciduous forest of Mt. Jeombong. *Forest Science and Technology, 7*(2), 68-74.

Yılmaz, A. B., & Doğan, M. (2008). Heavy metals in water and in tissues of himri (Carasobarbus luteus) from Orontes (Asi) River, Turkey. *Environmental monitoring and assessment, 144*(1), 437-444.

Yogendra, K. a. (2008). Determination of water quality index and suitability of an urban waterbody in Shimoga Town, Karnataka. *Proceedings of Taal2007: The 12th world lake conference*, (pp. 342-346), International Lake Environment Committee.

Yuan, C. G., Shi, J. B., He, B., Liu, J. F., Liang, L. N., & Jiang, G. B. (2004). Speciation of heavy metals in marine sediments from the East China Sea by ICP-MS with sequential extraction. *Environment International, 30*(6), 769-783.

Zooniverse. (2021). *Galaxy Zoo*. Retrieved on 13 May 2021 from Zooniverse: https://www.zooniverse.org/projects/zookeeper/galaxy-zoo/