Implementation of Systematic Access to Isolation Sources of Microbes in a Culture Collection System

A Special Problem
Presented to the Faculty of
The Institute of Computer Science
University of the Philippines Los Baños

Ezra Earl P. Villanueva
June 2024

Implementation of Systematic Access to Isolation Sources of Microbes in a Culture Collection System

by

Ezra Earl P. Villanueva

In partial fulfillment of the requirements for the degree of
Bachelor of Science in Computer Science

_____
Mylah Rystie Anacleto
Adviser

_____
Date Signed

_____
Maria Art Antonette D. Clariño
Director
Institute of Computer Science

_____
Date Signed

ABSTRACT

The *caves* isolation source web application was developed to provide a database for the microbial data and metadata collected from caves in CALABARZON. It implemented features like Microbial Isolation Source Ontology (MISO) tags, interactive maps, and search and filtering tools to provide systematic access to strain isolation sources and environmental information. The application received an excellent System Usability Scale (SUS) score of 82.1875 from 8 respondents, indicating acceptable usability.

ACKNOWLEDGEMENTS

I express my deepest gratitude to everyone who has supported me throughout this study.

First and foremost, my sincerest appreciation goes to my adviser, Prof. Mylah Rystie Anacleto, for her invaluable guidance, motivation, and feedback. Your expertise, patience, and belief in my potential have been instrumental in completing this study. Thank you for pushing me to achieve more than I ever thought possible.

To the NICER CAVES project headed by Dr. Marian De Leon, I am deeply grateful for the opportunity to contribute to such an incredible project, showcasing my skills and creativity. I also extend my gratitude to all the collaborating organizations and institutions that have supported this project.

A special thank you to my panelists, Prof. Arian Jacildo, Prof. Fermin Roberto Lapitan, and Prof. Joseph Antony Hermocilla. Your valuable input has significantly enriched this study and contributed to its development.

To my family, thank you for your boundless love, patience, and encouragement. Your unwavering belief in me provided the strength and motivation I needed to persevere through the challenges of this journey. I dedicate this

Lastly, I wish to acknowledge my friends and orgmates, especially my housemates and all-nighter buddies Lem, Christian, and Roel– thank you for your continuous support and encouragement. Your companionship has been a source of great comfort and motivation.

I thank you all.

TABLE OF CONTENTS

ABSTRACT....3
ACKNOWLEDGEMENTS....4
TABLE OF CONTENTS....5
I. INTRODUCTION....1
A. Background of the Study....1
B. Significance of the Study....2
C. Objectives of the Study....2
D. Scope and Limitations....3
II. REVIEW OF RELATED LITERATURE....4
III. METHODOLOGY....6
A. Development Tools....6
B. Application Design....6
C. Non-functional Requirements....8
D. Data Population....8
E. Usability Testing....9
IV. RESULTS AND DISCUSSION....10
A. Web Application....10
B. Testing and Evaluation....14
V. CONCLUSION AND FUTURE WORKS....16
APPENDIX I....17
REFERENCES....19

I. INTRODUCTION

A. Background of the Study

Caves are important biodiversity and natural heritage assets because they serve as dwelling places for many diverse and ecologically valuable organisms, including animals (e.g., bats, insects, arachnids), microorganisms (e.g., bacteria, viruses, fungi), and plants (e.g., algae, mosses). Among these, bats and microorganisms are the most prevalent and diverse organisms commonly found in caves, playing crucial roles within cave ecosystems, particularly in nutrient cycling [1]. Bat guano or the accumulated excrement of bats, characterized by its rich mineral content, is responsible for the microbial diversity in caves which may vary significantly depending on the dietary habits of the bats that produce the guano [2]. These microbial communities are essential in the decomposition of guano and other organic and inorganic materials that become the main resources for many cave-dwelling organisms [1].

In a study reported by De Leon et al., bat guano from caves was found to harbor functional and pathogenic bacteria, drawing scientific interest in the possible presence of novel bacteria, potential and sustainable sources of antibiotic- and enzyme-producing microorganisms, and species diversity differences of bat guts and guano microbes [3]. There is a need to develop a deeper understanding of cave and microbial ecology to utilize beneficial microbes and be informed about the pathogenic agents that may potentially harm humans. The collection of microbial data and metadata in caves is essential to provide easy and reliable access for future research and development. In addition, studies on Philippine cave microorganisms are necessary to comprehensively address safety and security protocols in bat guano mining and sustainable cave management and conservation.

Over the past years, microbial communities in various ecological niches have been subject to research and inventories, driven by advancements in high-throughput metagenomic technologies and approaches [4]. In search of sources of biocatalysts that can have biotechnological and industrial applications, studies expand in extreme environments, such as hot springs and chemically contaminated soils, where microbes utilize key enzymes to deal with harsh or atypical environmental conditions [5]. The information on microbial studies is indeed scattered and confined across various scientific journals, books, culture collections, and private databases worldwide [6].

Several international initiatives (e.g., NCBI databases, JGI Genome Online Database: https://gold.jgi.doe.gov/, BacDive: https://bacdive.dsmz.de/, GBIF: https://www.gbif.org/) have mobilized and integrated microbial data into metadatabases, making it easier to access microbial data and metadata. There are no or very limited accessible microbial databases available in the Philippines as well as information on microbial communities in cave environments. Hence, this study will present a local microbial database that follows the same approach as these existing metadatabases.

The ‘caves’ isolation source web application will store and structure information collected from the caves in CALABARZON. To facilitate systematic access to the isolation sources of the microbial strains, this study will adopt Microbial Isolation Source Ontology (MISO), an internationally recognized controlled vocabulary of BacDive. Instead of describing isolation sources in free text–ranging from single words like "soil" to detailed sentences– the application will use a three-level hierarchical ontology (e.g., #Environmental | #Terrestrial | #Soil) []. These keywords or tags will be incorporated into the search and filtering tools. The application will also feature maps and graphs for visualizing quantitative data and streamlining access to microbial strain origins.

B. Significance of the Study

This study will provide a database of standardized and digitized data and metadata of the collected microbial strains from caves in CALABARZON, focusing on isolation source and environmental information. The database will mobilize and harmonize the scattered information to constitute structured database content with easy and reliable user access. This will contribute to various sectors and institutions by providing valuable resources and knowledge. For academic, research, and biotechnological institutions, this will offer microbial strains for instructional materials, thesis references, potential sources of novel enzymes, and standardized protocols and assays for bacterial ecology and taxonomy. The study will also enrich culture collections with new strains and metadata that will add documentation on the cave microbial diversity in the country. Furthermore, the output database will promote collaboration among researchers and inspire further study in cave and microbial ecology and related fields.

C. Objectives of the Study

The general objective of this study was to develop a database management system for the microbial data and metadata collected from the caves in CALABARZON. Specifically, this study aimed to:

- Provide systematic access to the origin and environmental information of microbial strains through a search feature focused on isolation sources;
- Design a location-based database for the storage of microbial metadata; and
- Utilize internationally accepted controlled vocabularies such as Microbial Isolation Source Ontology (MISO) to categorize data in the database.

D. Scope and Limitations

This study covered the design, development, and testing of the web application and database based only on the information and instructions provided by the NICER CAVES study. The web application focused on isolation sources and environmental information of microbial strains. Technically, the web application would only be accessible on any devices with web browsers and active internet connections. The testing was also limited to the System Usability Scale (SUS) questionnaire, along with the feedback and suggestions provided by participants.

II. REVIEW OF RELATED LITERATURE

Microbial data and metadata are scattered in various sources, from databases to scientific literature. Large aggregations of microbial metadata are not published but stored in laboratory notes or private bacterial culture collections [6][7]. Over the past years, dramatic progress has been made in the research on the diversity and function of microbial communities in the environment as a consequence of the advancement in high-throughput metagenomic sequencing technologies and methodological approaches [4]. Despite the increasing information from various sources, it remains underutilized for many reasons: data are found in a wide range of sources; data and metadata have different formats that prevent interoperability; much information is expressed in free text even in databases, and therefore difficult to access [7].

Present contributions have been made to centralize and standardize microbial metadata with structured databases. BacDive– the Bacterial Diversity Metadatabase of Leibniz Institute DSMZ is regarded as one of the most extensive databases for standardized prokaryotic information on strain level in the world with over 80,000 strains at present [8]. BacDive advanced web service offers RESTful API for large-scale and easy data retrieval [8]. The Genomes OnLine Database (GOLD) of the Department of Energy Joint Genome Institute (DOE-JGI) is a curated collection of genome projects and metadata, featuring over 1.17 million entries across studies, organisms, biosamples, sequencing projects, and analysis projects [9]. Aside from maintaining comprehensive databases like GenBank and PubMed, the National Center for Biotechnology Information (NCBI) also manages minor databases like the NCBI Taxonomy and NCBI Genome containing a vast amount of data and metadata covering a wide range of organisms including bacteria, viruses, and fungi [10]. The majority of the NCBI databases are interlinked with each other and are publicly available.

The standardization efforts on the databases extend to microbe isolation classification to describe the geographic location and environment context from which the biological sample originated. The Genomic Standards Consortium (GSC) introduced minimum information about any (x) sequence (MIxS) and the 'environmental packages', offering a framework for microbial metadata providers and resource centers. This includes the geographic location and environmental information: biome, environmental feature, and environmental material [11]. BacDive developed Microbial Isolation Source Ontology (MISO), a hierarchical controlled vocabulary of 387 terms in three levels (e.g. #Environmental | #Terrestial | #Soil). BacDive isolation sources search implemented these terms into a tagging system to manually index strains [8]. Another, GOLD utilizes a five-level ecosystem classification system of nearly 800 distinct combinations, including microbial ecosystem, ecosystem category, ecosystem type, ecosystem subtype, and specific ecosystem (e.g. Host-associated | Mammals | Digestive System | Large Intestines | Fecal) [9].

Visualization is a vital aspect in the field of understanding microbial communities. Maps are the popular way of visualizing quantitative and geospatial data. Web mapping and the use of geospatial information have evolved over the years creating many tools, libraries, and

applications for visualization. Leaflet.js is a widely adopted open-source JavaScript library for creating interactive maps on web pages. It demonstrates efficient cross-platform compatibility that functions seamlessly across major desktop and mobile platforms, providing various mapping functionalities, including zooming, panning, adding markers, polygons, and overlays, as well as supporting tile layers from different map providers like OpenStreetMap [12]. Both the GOLD and BacDive databases utilize Leaflet.js for displaying distribution maps.

III. METHODOLOGY

A. Development Tools

MongoDB, Express.js, React, and Node.js collectively form the MERN development stack. It is a popular technology stack based on JavaScript technologies. MongoDB is a NoSQL database that stores data in flexible and JSON-like documents. Express.js handles the server-side web framework. React is a front-end JavaScript library for building user interfaces. Node.js serves as an asynchronous event-driven JavaScript runtime environment.

This study chose the MERN stack to develop the web application for its simplicity, flexibility, and scalability. The web application was built utilizing Visual Studio Code (VS Code), an integrated development environment. The backend was deployed on Render, while the frontend was hosted on Netlify.

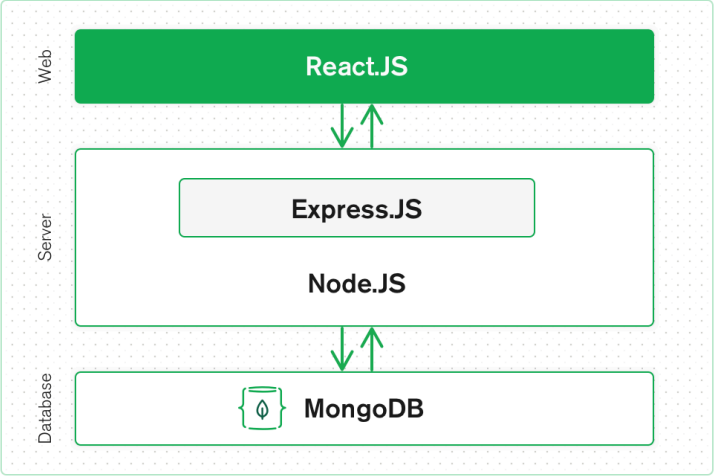

Fig. 1. The MERN Stack

B. Application Design

1. System Architecture

The web application was designed using a client-server architecture, a traditional web development model that divides the application into two parts: the client and the server. The client is the frontend or user-facing part of the web application while the server is the backend that handles database interactions, business logic, and client requests (see Fig. 1). Communication between the client and server is handled through RESTful APIs, following a resource-based approach with clear conventions to perform operations on data.

2. Database Design and Models

MongoDB serves as the primary database of the web application. Mongoose, an Object Data Modeling (ODM) library for MongoDB and Node.js, facilitates interaction with the database by defining schemas and models. The database contains collections for users and strains, each with models designed to handle CRUD (Create, Read, Update, Delete) operations on these collections.

   a. User Model: This model is designed to store user information, including name, email, password, institution, address, and user level. In addition to CRUD functionalities, it provides validation to ensure each user has a unique email address and a secure password.
   b. Strain Model: This model is designed to store strain information, including identifiers, taxonomic details, and isolation source data. It provides methods for CRUD operations for strain items.

3. User Interface Design

React, Tailwind CSS, and shadcn/ui are the complementary stacks used to build the modern and responsive interface of the web application. React served as the core framework for the frontend structure and functionality of the web application. Tailwind CSS, a utility-first CSS framework, was used for component styling with ready-to-use classes that created visually consistent UIs with minimal custom CSS. shadcn/ui provided a collection of accessible and customizable UI components, further accelerating the development process. The resulting design is minimalistic, modern, and intuitive, prioritizing user-friendliness and positive user experience.

4. Isolation Sources Modules

The systematic access to the isolation sources of the strains is a key feature of the web application. It offers several isolation source modules that align with the objectives of the study:

   a. Isolation Source Map - displays strains' count and isolation sources based on location information.
   b. Isolation Source Table - displays strains' isolation source information and its corresponding MISO tags
   c. Isolation Source Statistics - displays charts based on strains' isolation information field.

5. User Levels

The system has a dedicated user interface and privileges for each user level as seen in Figure 2.

   - Guest user: Can view and search for strain information without the need to create an account.
   - Admin user: Can add, update, and delete strain information, in addition to all the privileges available to Guest users.

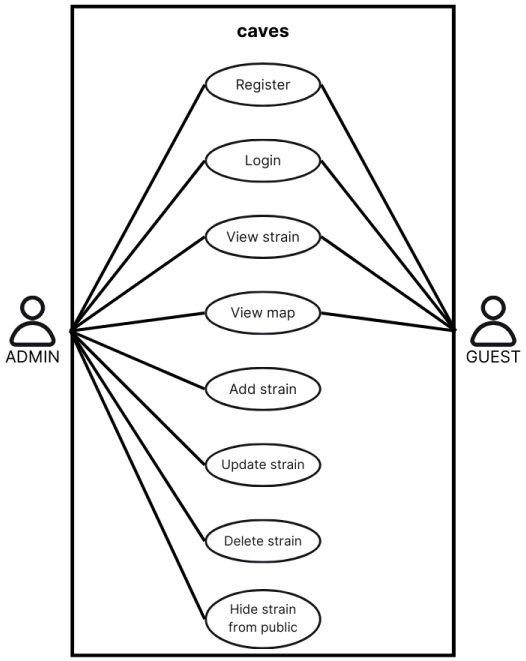

Fig. 2. User Case Diagram

C. Non-functional Requirements

Non-functional requirements (NFR) describe the capabilities and limitations of the services and features of the system. The system has the following NFRs:

- The system is web-based and should operate only in browsers like Chrome and Firefox.
- The system should be available at any time except for system maintenance or an unavailable Internet connection.
- The system has a minimal user interface that is easy to learn and use by users of any technical background.
- Personnel knowledgeable of the MERN stack is necessary to fix, update, and maintain the system.
- The system is highly maintainable for future development.

D. Data Population

NICER CAVES Project 4 provided the data that populated the web application. Isolation source information was patterned for each isolate produced on a single source or sample. For additional demo data to be displayed on public or guest view, some strain data from BacDive was inputted into the system.

Item Number	Statement
1	I think that I would like to use this system frequently.
2	I found the system unnecessarily complex.
3	I thought the system was easy to use.
4	I think that I would need the support of a technical person to be able to use this system.
5	I found the various functions in this system were well integrated.
6	I thought there was too much inconsistency in this system.
7	I would imagine that most people would learn to use this system very quickly.
8	I found the system very cumbersome/difficult to use.
9	I felt very confident using the system.
10	I needed to learn a lot of things before I could get going with this system.

Table 1. System Usability Scale Statements

E. Usability Testing

A test was conducted to test the usability of the web application. This study utilized John Brooke's System Usability Scale (SUS), a widely recognized method for measuring website usability from the user perspective. It consisted of a 10-item questionnaire (see Table 1) where each item is rated on a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5).

The 8 respondents, comprising members of NICER CAVES as well as UPLB BS Biology undergraduates and graduates, were recruited for user testing. Respondents were given access to the web application and user guide and were asked to use and explore the application. Afterward, they evaluated the user experience using the SUS survey and provided additional feedback through Google Forms.

IV. RESULTS AND DISCUSSION

A. Web Application

1. Home page

The *caves* homepage (https://isolationsources-caves.netlify.app/) provides a simple overview of the purpose and description of the web application. Navigation links to subpages, as well as login and signup options, are easily accessible via the navigation bar. Additionally, the homepage includes prominent buttons that link to the signup and isolation source pages for emphasis.

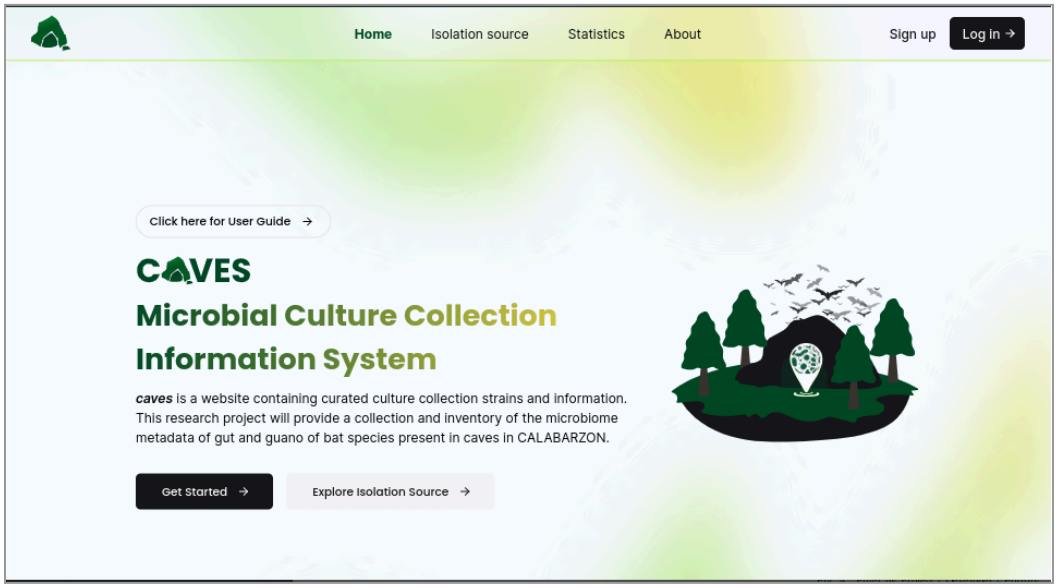

Fig. 3. Home page

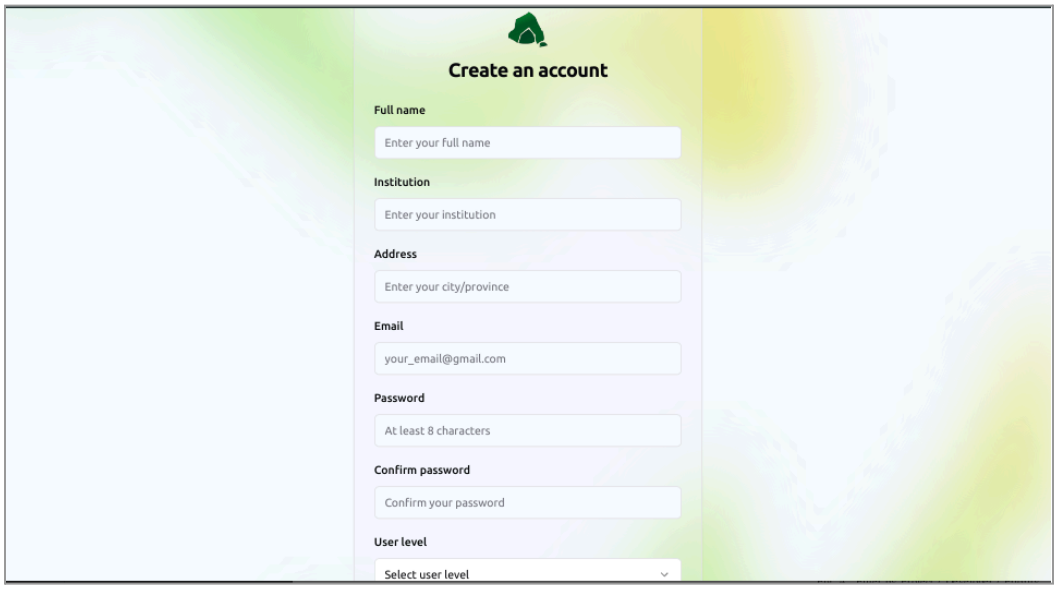

Fig. 4. Signup page

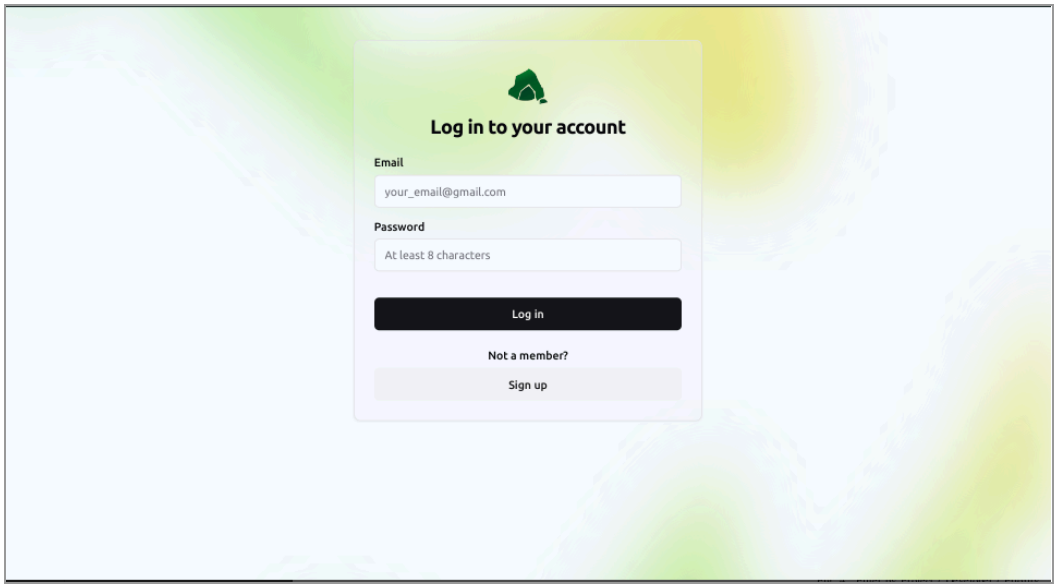

Fig. 5. Login page

2. Signup and Login pages

New users will be asked to provide their name, email, password, and desired user level, with the option to include their institution and address. Upon completion, they can access the web application features based on their specified user level. Guest users may opt not to create an account. Registered users can log in to the application by entering their email and password, with successful authentication redirecting them to the homepage.

3. Isolation source page

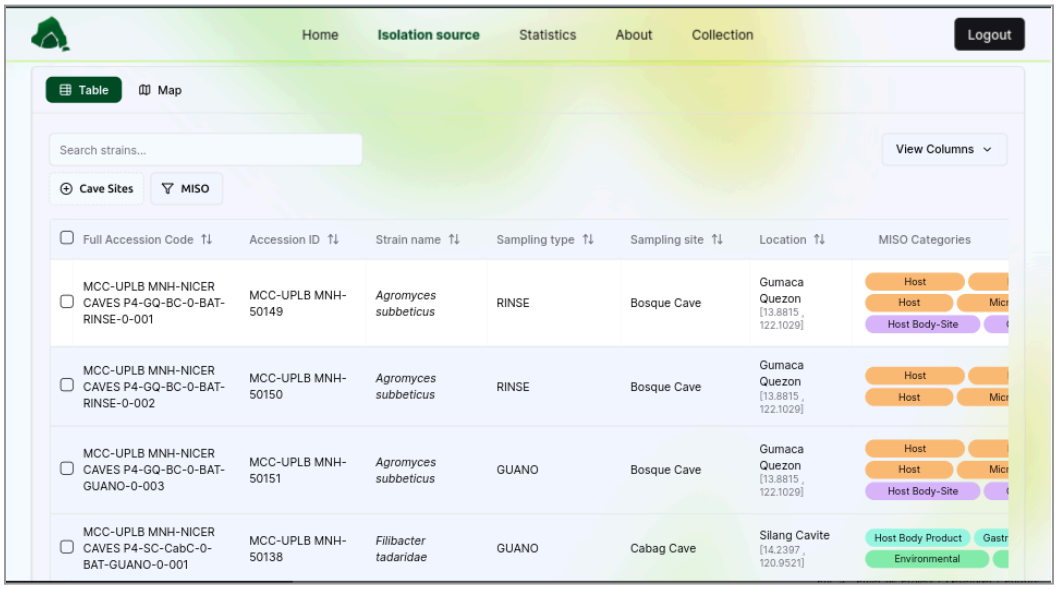

Fig. 6. Isolation source page

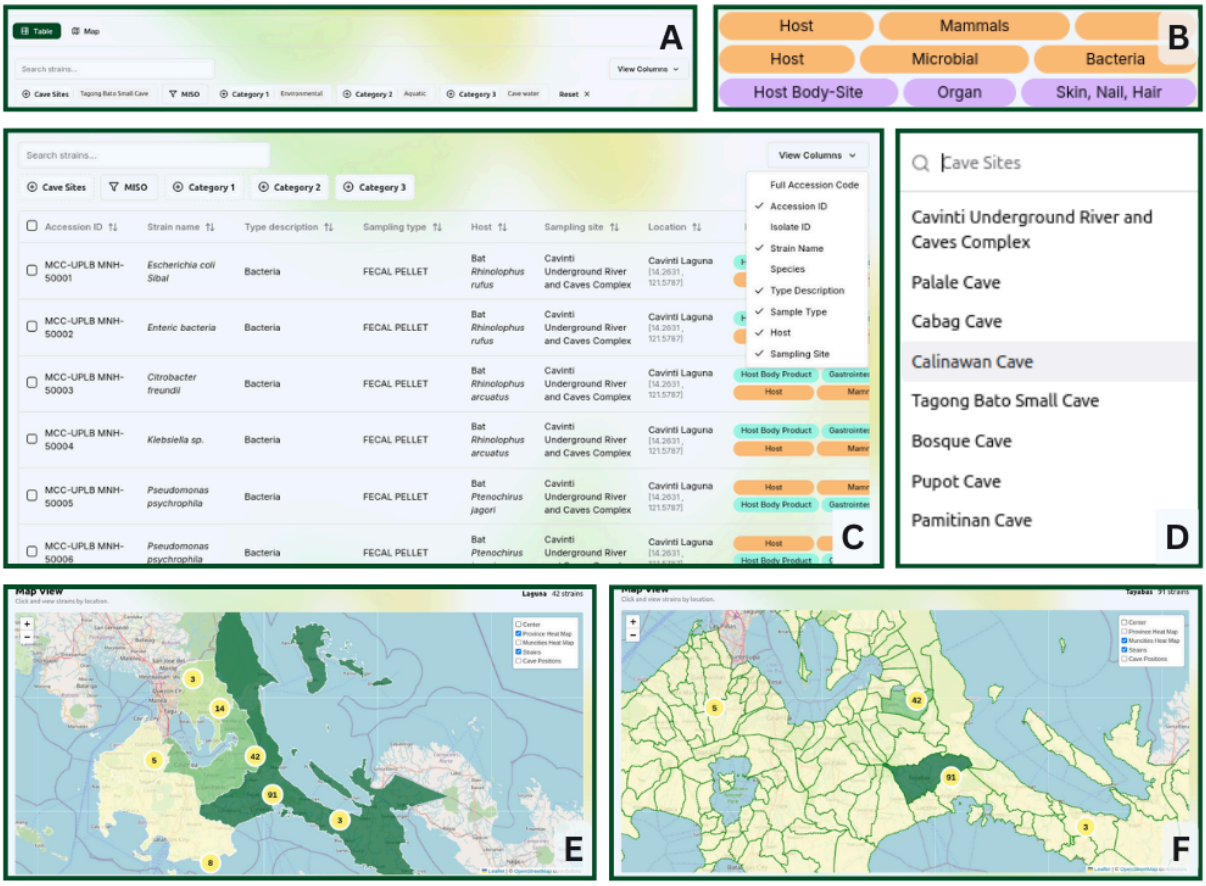

Fig. 7. Isolation Source page features. (A) Search bar and Filter and View Column buttons. (B) Sample MISO tags. (C) Isolation Source Strain Table. (D) Cave Sites Filter. (E) Isolation Source Map View with Provincial Heat Map Layer. (F) Isolation Source Map View with Municities Heat Map Layer.

The Isolation Source page features a table of strain entries and a map view of the strain isolation sources. The table includes a global search bar that allows users to search across all fields and columns of the entries. Additionally, it offers filtering tools for cave sites and MISO categories. As shown in Fig #B, MISO tags are highly useful for searching and filtering the data users wish to display. Users can access more strain information via the "View Columns" dropdown in the upper right corner, where they can toggle which columns or fields to display. The available columns may vary depending on the user's access level.

The map view provides an interactive display featuring a heat map of data entries based on their isolation source locations, cluster markers indicating the number of strains sourced from a location or site, and clickable layers that can filter data on the table. As shown in Figure #E (Province View) and Figure #F (Municities View), the darker shades of green correspond to a higher number of strains sourced from that location.

4. Statistics page

To give users an overview of the numerical data featured in the web application, the statistics page was added to present statistical information on selected fields. At the top part, four key metrics are displayed, providing information on the total number of strains, species, sampling sites, and collaborators. Below, 4 bar charts are shown, covering only the comparison of strain quantity on fields of type description, sample type, sampling sites, and provinces. As of the moment, the system contains 167 strain entries.

Fig. 8. Statistics page

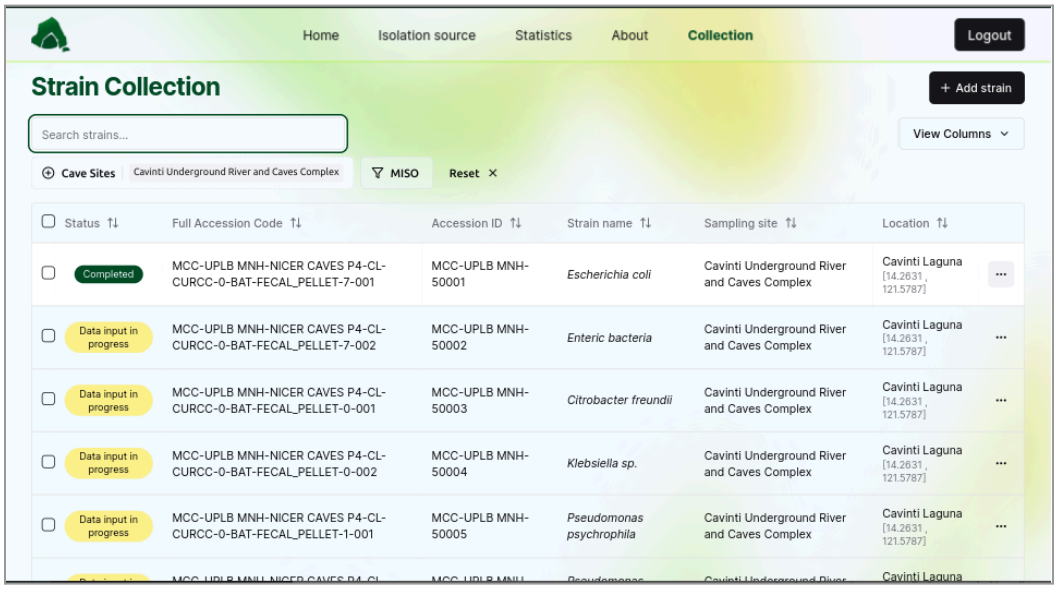

Fig. 9. Collection page

5. Collection page

The collection page is accessible only to admin users or collaborators, allowing them to perform CRUD (Create, Read, Update, Delete) operations on their strain collections. The strain collection is displayed in a table where an admin can add, update, and delete strain entries. Additional features implemented include the ability to set data input status as “Completed” and “Data input in progress” and to change the setting of whether the strain entry is visible to the public or kept private.

	ODD ITEMS	EVEN ITEMS	SCORE	GRADE
TESTER 1	22	5	92.5	A
TESTER 2	17	10	67.5	C
TESTER 3	21	7	85	A
TESTER 4	25	9	90	A
TESTER 5	24	10	85	A
TESTER 6	20	8	80	B
TESTER 7	19	9	75	B
TESTER 8	21	8	82.5	A
MEAN SCORE:			82.1875	**A**

Table 2. SUS scores

B. Testing and Evaluation

The testing was conducted in a remote setup, and the testers were provided with links to the web application, a user guide, and an evaluation form. The user guide contains details about the system’s key features and functionalities as well as instructions tailored to different user levels. The evaluation form utilized the System Usability Scale (SUS) and was administered via Google Forms.

As shown in Table 2, the web application obtained a mean score of 82.1875 across all 8 evaluation responses. Interpreting this score on the SUS evaluation framework, the web application qualifies for an A grade which indicates an acceptable score and excellent usability (see Fig. 10). This result indicates that the web application successfully met the standard of its target users, providing them an effective and user-friendly systematic access to the isolation sources of the microbial collection from the caves.

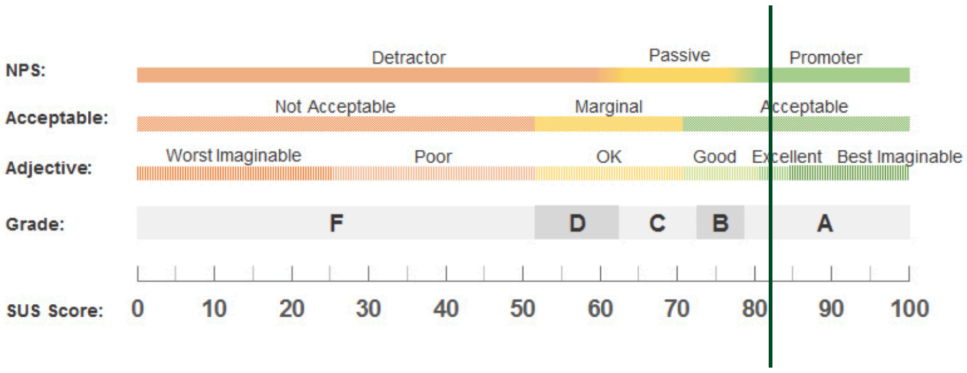

Fig. 10. SUS score interpretation

In addition to the SUS, the evaluation form also asked users for feedback and comments. Some notable comments include the web application's modern and fresh look that differs from traditional online metadatabases. Some feedback stated to include authors and citations, as well as simple descriptions and discussions on the statistical results of the data.

> Respondent 4: *"The website for the microbial culture collection system is impressive in several aspects. One, the modern design gives it a fresh and appealing look, making it stand out compared to traditional online gene bank platforms. Navigation is straightforward and intuitive as well, ensuring a user-friendly experience. The streamlined interface, free from an overwhelming number of buttons, enhances usability and efficiency. The site is both aesthetically pleasing and functionally effective, making it pleasurable to use."*
>
> Respondent 7: *"The statistical portion may also be added with separate descriptions and discussions without cramping all these data tables in one page without any explanations. This will be more helpful for biologists and students. "*

V. CONCLUSION AND FUTURE WORKS

This study focused on the development and design of the *caves* isolation source web application using the MERN stack, intending to provide a database management system for the data and metadata of strain collected from caves in CALABARZON. The web application enabled systematic access to the isolation sources and environmental information of strains by implementing various features such as utilizing MISO tags, interactive maps, and search and filtering tools. Additionally, the implementation of admin and guest user levels moderated access to the data.

The web application featured a minimalist, modern, and intuitive interface, prioritizing a positive user experience. It achieved an excellent System Usability Scale (SUS) score of 82.1875 among 8 respondents, indicating acceptable usability.

As part of the NICER CAVES project, this study has successfully met its objectives and established a solid foundation for future development and expansion. Continued alignment and integration with other studies and components within the project are recommended to utilize its potential fully. Additionally, the feedback of the respondents may be considered for future improvements to the features and functionalities of the web application. Through this future work, the *caves* isolation source web application can further reinforce its purpose as part of the resource center for cave microorganisms.

APPENDIX I

SCREENSHOTS FROM THE WEB APPLICATION

Collection Page Features and Functionalities (for Admin Users only)

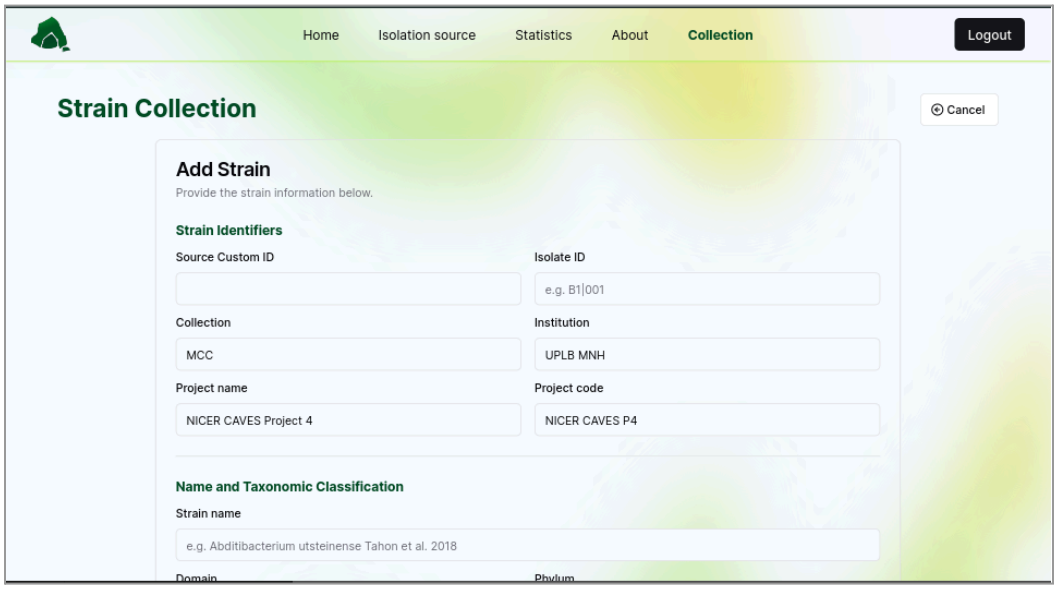

Fig. 11. Add strain

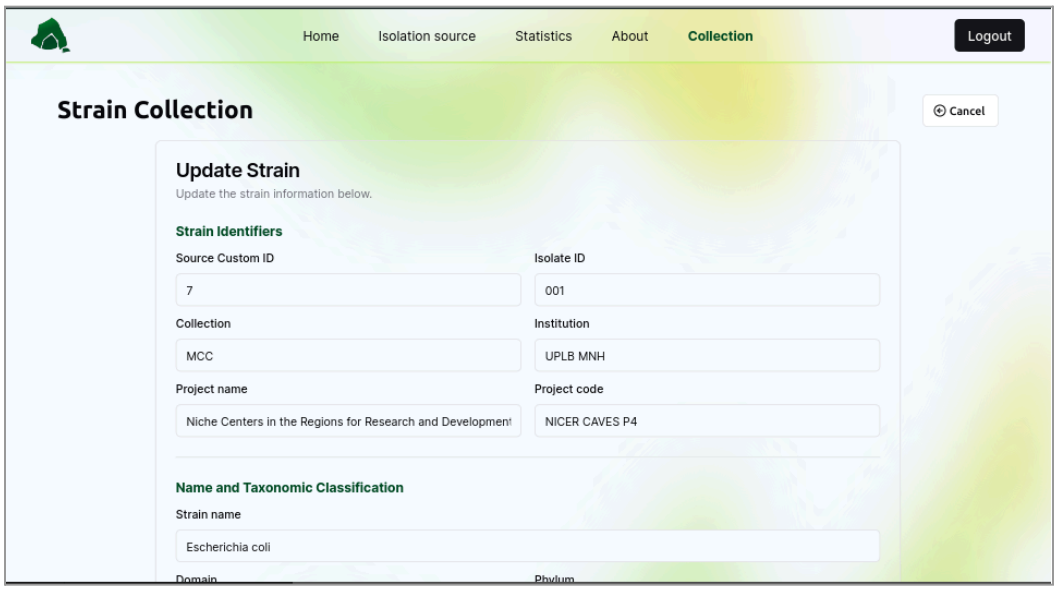

Fig. 12. Update strain

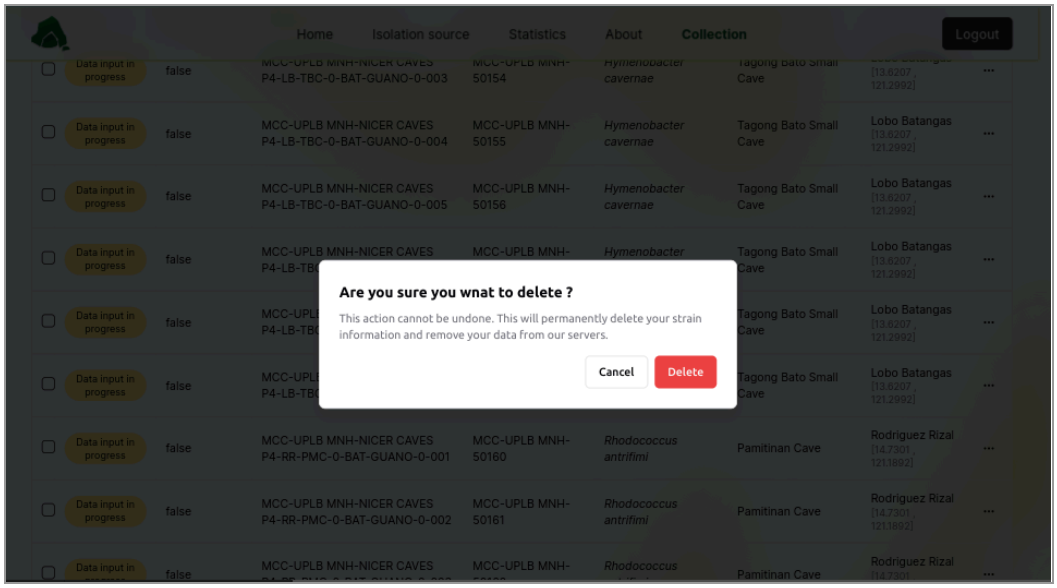

Fig. 13. Delete strain

REFERENCES

[1] *Philippine Caves: Beneath the Earth's Surface – Conservation and Management*. DENR - Biodiversity Management Bureau (BMB), 2017. Available: https://elibrary.bmb.gov.ph/elibrary/books/philippine-caves-beneath-the-earths-surface-conservation-and-management-3/

[2] S. Sakoui *et al.*, "The life hidden inside caves: Ecological and economic importance of Bat Guano," *International Journal of Ecology*, vol. 2020, pp. 1–7, 2020. doi:10.1155/2020/9872532

[3] M. P. De Leon, A. D. Montecillo, D. S. Pinili, M. A. Siringan, and D.-S. Park, "Bacterial diversity of bat guano from Cabalyorisa Cave, Mabini, Pangasinan, Philippines: A first report on the metagenome of Philippine bat guano," *PLOS ONE*, vol. 13, no. 7, 2018. doi:10.1371/journal.pone.0200095

[4] J. Imhoff, "New Dimensions in Microbial Ecology—Functional Genes in Studies to Unravel the Biodiversity and Role of Functional Microbial Groups in the Environment," *Microorganisms*, vol. 4, no. 2, p. 19, May 2016, doi: 10.3390/microorganisms4020019

[5] U. Sood, R. Kumar, and P. Hira, 'Expanding Culturomics from Gut to Extreme Environmental Settings', *mSystems*, vol. 6, no. 4, p. doi:10.1128/msystems.00848-21, 2021.

[6] L. C. Reimer, C. Söhngen, A. Vetcininova, and J. Overmann, "Mobilization and integration of bacterial phenotypic data—Enabling next generation biodiversity analysis through the Bac Dive metadatabase," *Journal of Biotechnology*, vol. 261, pp. 187–193, Nov. 2017, doi: https://doi.org/10.1016/j.jbiotec.2017.05.004.

[7] S. Dérozier et al., Omnicrobe, an open-access database of microbial habitats and phenotypes using a comprehensive text mining and data fusion approach, *2022*. doi:10.1101/2022.07.21.500958

[8] L. C. Reimer *et al.*, "Bac*dive* in 2022: The knowledge base for standardized bacterial and Archaeal Data," *Nucleic Acids Research*, vol. 50, no. D1, 2021. doi:10.1093/nar/gkab961

[9] S. Mukherjee et al., "Genomes online database (GOLD) v.8: Overview and updates," *Nucleic Acids Research*, vol. 49, no. D1, 2020. doi:10.1093/nar/gkaa983

[10] National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/.

[11] P. Yilmaz et al., “Minimum information about a marker gene sequence (MIMARKS) and minimum information about any (x) sequence (MIxS) specifications,” *Nature Biotechnology*, vol. 29, no. 5, pp. 415–420, 2011. doi:10.1038/nbt.1823

[12] V. Agafonkin, “An open-source JavaScript library for interactive maps,” Leaflet, https://leafletjs.com/ (accessed Apr. 6, 2023).