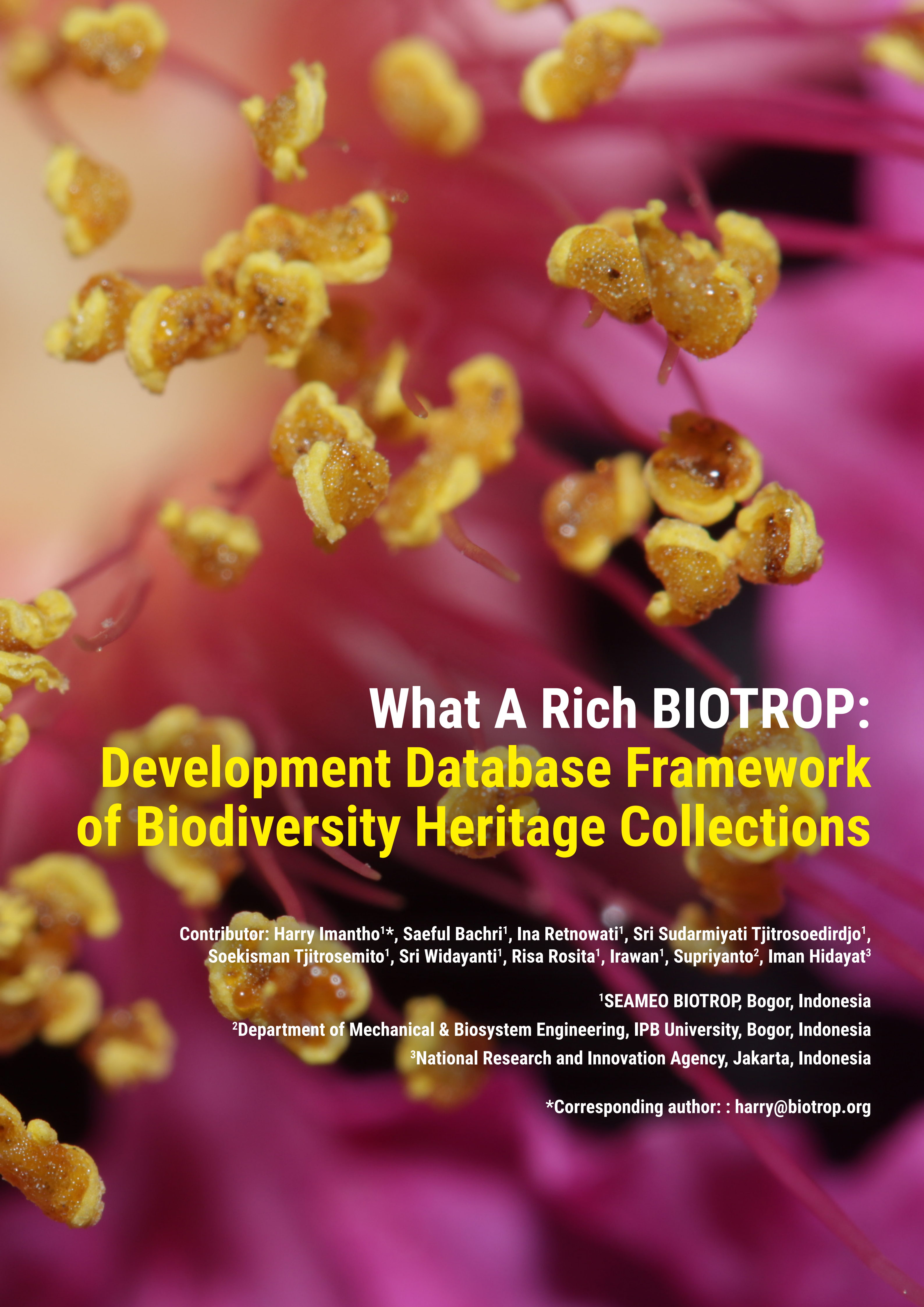


# What A Rich BIOTROP: Development Database Framework of Biodiversity Heritage Collections

Contributor: Harry Imantho[1]*, Saeful Bachri[1], Ina Retnowati[1], Sri Sudarmiyati Tjitrosoedirdjo[1], Soekisman Tjitrosemito[1], Sri Widayanti[1], Risa Rosita[1], Irawan[1], Supriyanto[2], Iman Hidayat[3]

[1]SEAMEO BIOTROP, Bogor, Indonesia
[2]Department of Mechanical & Biosystem Engineering, IPB University, Bogor, Indonesia
[3]National Research and Innovation Agency, Jakarta, Indonesia

*Corresponding author: : harry@biotrop.org

## Abstract

The need for biodiversity science with the availability and interoperability of data on the internet reinforces each other. This is driven by the increasing literacy of the world community regarding the impact of loss of biodiversity on the sustainability of human life. High-quality biodiversity data is needed to demonstrate that spatio-temporal loss of biodiversity has led to a reduction in quality of life. This facts have underpinned SEAMEO BIOTROP to increase the visibility of managed biodiversity collection data. SEAMEO BIOTROP maintains thousands of valuable herbarium collections of weeds, trees and invasive plants; fungi, insects and pests for research and testing purposes. These historic and valuable collections of biodiversity need to be preserved and better managed in digital format so that the information is available to botanists and the general public around the world. A study regarding the development of an integrated biodiversity collection database framework has been carried out by SEAMEO BIOTROP. This study aims to formulate and develop an initial framework for the BIOTROP biodiversity database that is relevant to national, regional and global needs. This study succeeded in developing an integrated database framework that brings together all digital data from the SEAMEO BIOTROP biodiversity collection into a database management system. This database management system also adopts Darwin Core metadata to ensure easy exchange and sharing of data with existing biodiversity data management systems in the world.

**Keyword:** biodiversity, database framework, herbarium specimen, insect collection, fungi collection

## Introduction

Currently, the resonance between the needs of biodiversity science and the availability and interoperability of data on the internet is getting stronger. This situation is driven by the awareness that biodiversity is essential for the sustainable development of human society (Cardinale et al., 2019). Increasing public literacy and knowledge is important in saving, preserving, and managing biodiversity wisely and sustainably.

Biodiversity refers to the various life forms that exist on earth, including animals, plants, microorganisms, and the entire ecosystem they live in. The different components of biodiversity in an environment work together to maintain the balance of the ecosystem. These interactions create several functioning systems called ecosystem services, consisting of provisioning, regulating, cultural, and supporting services (Hooper et al., 2005). However, biodiversity's sustainability, quality, and quantity continue to decline due to various pressures such as over-exploitation and utilization, waste pollution, air pollution, microplastic pollution, habitat destruction, alien species invasion, climate change, and forest fires (Reid et al., 2019; Dudgeon et al., 2005; Hooper et al., 2012).

High-quality biodiversity data is needed to show that the spatio-temporal biodiversity loss has led to a decline in ecosystem services (Sutherland et al., 2013). Furthermore, high-quality biodiversity data are also indispensable for research and management in biodiversity conservation (Jin & Yang, 2020). The growth of data digitization in the last centuries has made billions of biodiversity records from various sources (e.g., museums, herbaria, and field surveys) available to the public (Guralnick et al., 2016; Nelson & Ellis, 2019). Organizations and individuals have used this development to build databases containing massive species occurrence records. These databases have contributed to a wide range of applications, such as predicting the distribution of focal species (Lin et al. 2017), estimating the extinction risk of plants (Darrah et al., 2017), controlling plant invasions (Banerjee et al., 2021), evaluating conservation and protected areas (Pelletier, 2018), and analyzing the impacts of climate change on biodiversity (Cabrelli et al., 2014).

Over five decades, SEAMEO BIOTROP has made significant contributions to biodiversity and ecosystem conservation, such as improving plant productivity, nutrient cycling efficiency, and reforestation programs in the region. The Centre also provided knowledge and understanding of the ecology and management of tropical terrestrial and aquatic weeds, vertebrate pests, insects, and fungi that may affect stored products. One evidence of SEAMEO BIOTROP's significant contribution to biodiversity and

ecosystem conservation is its thousands of valuable herbarium collections of weeds, trees, and invasive plants. The center also maintains a high-value collection of fungi, insects, and storage pests for research and testing purposes. This valuable collection of historic biodiversity must be preserved and better managed digitally to make information available to botanists and the general public worldwide. This study aims to develop an initial framework for the SEAMEO BIOTROP biodiversity database relevant to national, regional, and global needs.

## Methods

### Time and location

The study was carried out from July until November 2022. Historic biodiversity data was collected from SEAMEO BIOTROP Herbarium, Entomology Laboratory, and Phytopathology Laboratory. The database framework formulation, analysis, development, and finalizing document of SEAMEO BIOTROP heritage collections were done in the Remote Sensing and Ecology Laboratory, SEAMEO BIOTROP, Jalan Raya Tajur Km. 6 Bogor.

### Tools and material

Hardware and software were used to develop the database framework of SEAMEO BIOTROP heritage collections, as presented in Table 1.

Tabel 1. Softwares used in the study

| No | Name | Function |
|---|---|---|
| 1 | Scanner | Digitalizing of specimen and picture of collection |
| 2 | Microsoft Office | Digitalizing and develop guide book |
| 3 | Figma Interface Design Tool | Online and collaborative web interface design |
| 4 | PostgreSQL | Object Relational Database Management System |

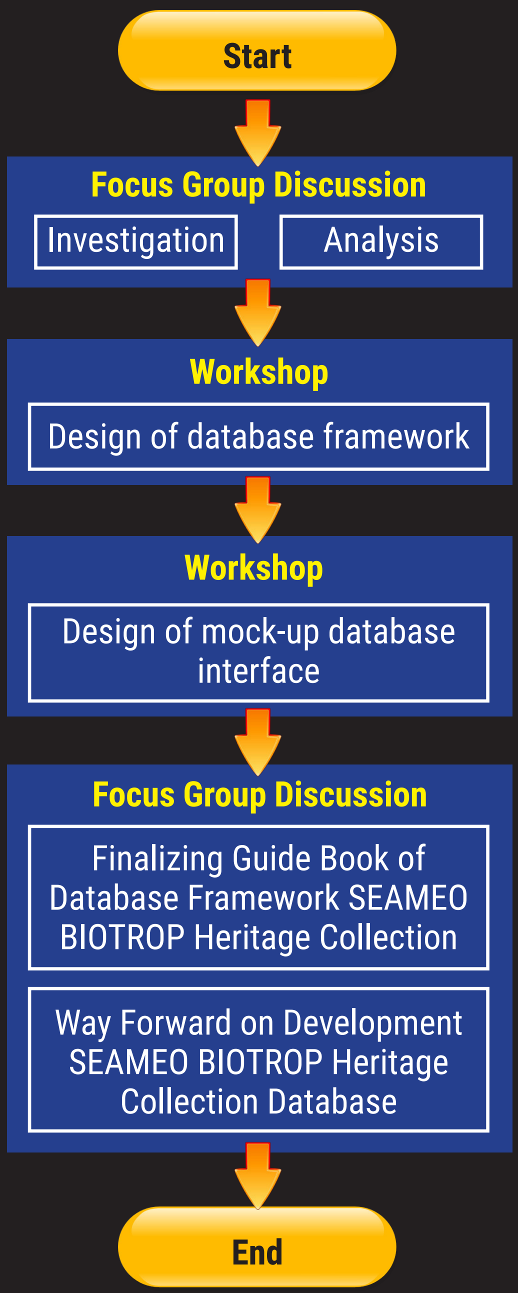



Figure 1. Stages of activity in the design of a biodiversity heritage collections' database framework.

## Method

The modified waterfall method is used as a guide in the process of designing a biodiversity database framework. The waterfall method is a systematic and sequential information system development model. Each stage of the method is carried out collaboratively by involving data owners, taxonomists and biologists, information systems experts, and database managers through FGDs and workshops, as shown in Figure 1.

### Investigation

Investigation is a process of exploring constraints and challenges in managing data and information on biodiversity. Consultation activities with stakeholders are carried out through focus group discussions (FGD). This activity also focuses on gathering information from experts regarding national, regional, and global needs for biodiversity data and information that can be handled by SEAMEO BIOTROP expertise. The results of this stage serve as a reference for the need to develop biodiversity data and information management systems, objectives, and system specifications required by stakeholders.

### Analysis

Needs analysis is a detailed process of the results of the investigation. The analysis is carried out on the system currently running/in use, standard operating procedures used, and manual, semi-manual, or automated processes using existing application systems. The analysis phase produces detailed requirements recommendations for developing existing application systems or even developing new application systems. The analysis phase also examines the resource requirements in developing an application system that fits the purpose.

### Design of database framework

The database framework design phase focuses on developing a conceptual data model from a collection of specimens of weeds and invasive plants, insects, and fungi concerning taxonomic structure. The conceptual data model is then translated into tables (entities) to build a relational database model.

### Design of mock-up database

This stage focuses on making prototypes of the database management system interface. The interface design provides features according to the needs of stakeholders, both internal users (SEAMEO BIOTROP) and users at the national, regional, and international levels.

Development of guide book and identifying way forward on development of biodiversity heritage database management system

This stage is intended to produce a very important manual for data managers, database management system developers, and general users. The guidebook is a standard reference for taking samples of collections or specimens in the field, a reference for recording collection or specimen data, and a reference for the development of a sustainable database management system.

## Result and Discussion

### Baseline Data

The International Association of Plant Taxonomy has internationally recognized the SEAMEO BIOTROP Herbarium and has been listed in the Herbarium Index with the abbreviation BIOT since 1990. The SEAMEO BIOTROP Herbaria maintains collections of specific weeds and invasive plant species specimens for research and reference. Data for 2022 show that the SEAMEO BIOTROP Herbarium (BIOT) collection consisted of 6,121 weeds and invasive alien plants, 7,720 specimens of forest herbarium, and 2,970 specimens of moss and lichen. These collections are stored as specimens, scanned images, and tabular data records. From the total collection of weeds and invasive alien plants, it is known that 1,936 species from 187 families were plant species introduced to Indonesia (Invasive Alien Species, IAS). About 17.5% of the introduced plant species were classified as weeds. Most of the SEAMEO BIOTROP herbarium (BIOT) digital collection data has been stored in a separate database.

Furthermore, SEAMEO BIOTROP Laboratory phytopathology maintains a collection of fungi (fungi and mold) consisting of 390 cultured fungi, including preharvest and postharvest species of fungi and 5 uncultured mycorrhizal species. This valuable collection data has not been fully managed using a database approach, so the availability and accessibility are still limited for microbial experts (mycologists) and the public. Currently, limited use of the fungi collections is for research and testing. SEAMEO BIOTROP also manages and maintains a collection of warehouse insects and pests in the form of collections of live insects, dry-preserved insects, and wet-preserved insects, which are stored in the Entomology Laboratory. Insect collecting activities started from 1979 to 1984, resulting from the research project TDRI (TPI) - BIOTROP. The recording of insect data refers to the book "Identification Records of Insects and Arachnids from Stored Products in Indonesia" compiled by C.P. Haines. Some of the digital insect collection data has been stored in a separate database but is still limited to internal use and has not been developed into institutional explicit knowledge.

Although most of the valuable collection data at SEAMEO BIOTROP has been stored in digital form, it is managed using a separate database approach and only oriented towards internal data management needs. This valuable collection of historic biodiversity must be preserved and better managed in an integrated database format to make the information available to botanists and the general public worldwide. This effort also supports the Center's program initiative to strengthen its visibility, contribution, and role in saving biodiversity by researching and developing tropical biology learning models in Southeast Asia An integrated biodiversity database system. It can be used as a reference in assessing the condition of ecosystems prior to land conversion and over-exploitation in several regions in Indonesia. An initial framework for the SEAMEO BIOTROP biodiversity database has been proposed and initiated for development, responding to national, regional, and global needs, as presented in Figure 2.

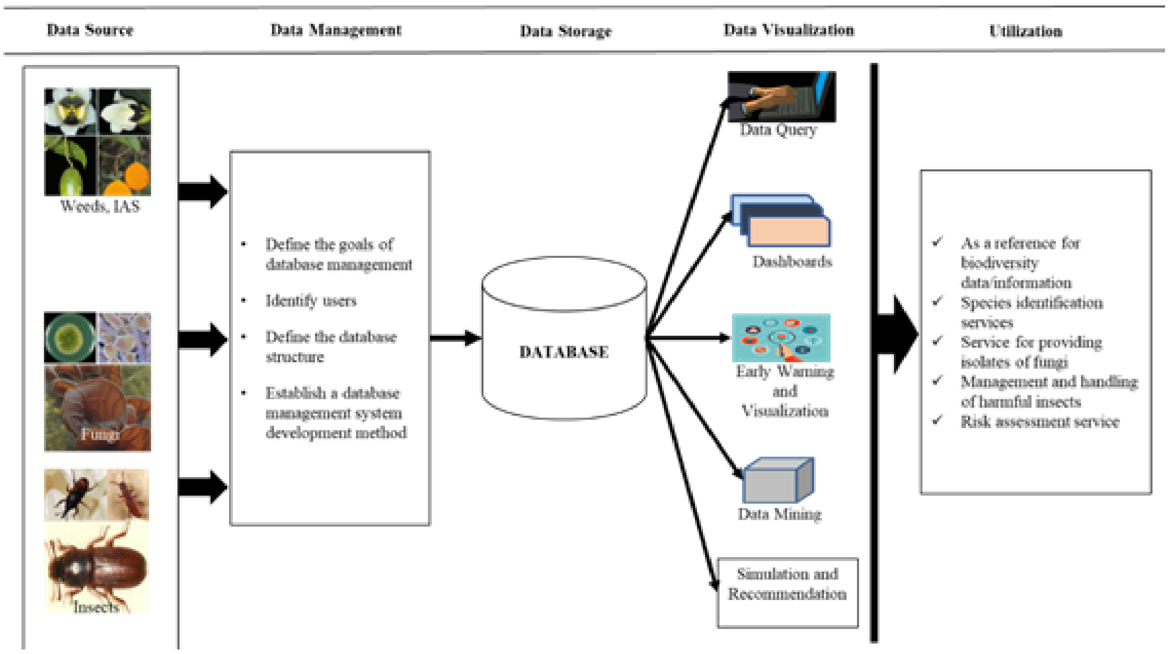



Figure 2. Biodiversity heritage collections' database framework.

## System Development

The development of the database framework has been carried out by identifying specific and general data and information in detail from each type of collection that has been recorded and is available in each laboratory, as presented in Tabel 2 - 4. Information of the same nature, contained in the three taxonomy types of data collection, has become a reference for building an integrated conceptual database. Specific information is retained as information tied to each type of collection. Some information and metadata are added with reference to Darwin Core guidelines to ensure ease of data sharing and potential for developing integration with existing database management systems in the world, such as the Global Biodiversity Information Facility (GBIF).

Table 2. Conceptual database for weed and IAS specimens data management.

| No | Table name (entity) | Information |
|---|---|---|
| 1 | Kingdom | Contain Kingdom Plantae code |
| 2 | Family | Contain family and family name code |
| 3 | Genus | Contain genus and genus name code |
| 4 | Species | Contain species code and species name, description and references |
| 5 | Origin | Plant origin and its ecosystem |
| 6 | Collector | Contain collector code, collector name and initials Contain specimen code, collection number |
| 7 | Location | Contain the location code where the specimen was collected, the coordinate position |
| 8 | Specimen | Contain duplicate specimen information in other herbariums, drawings, references |
| 9 | Specimen photo | Contain herbarium specimen photo codes or collection photos from the field |
| 10 | Determined by | Contains the code name that performs the identification and identification date |
| 11 | Notes | Benefits, control and other important information |

Table 3. Conceptual database for fungi collection data management.

| No | Table name (entity) | Information |
|---|---|---|
| 1 | Kingdom | Contain kingdom code and kingdom name |
| 2 | Phylum | Contain phylum code and phylum name |
| 3 | Class | Contain class code and class name |
| 4 | Ordo | Contain order code and order name |
| 5 | Family | Contain family code and family name |
| 6 | Genus | Contain genus code and genus name |
| 7 | Species | Contain species code and species name, description and references |
| 8 | Collector | Contain collector code, collector name and initials |
| 9 | Isolate | Contain specimen code, collection number, duplicate number, origin, date received, date of isolation, specimen photo, micrograph photo, Sequencing information, description, symptoms, references |
| 10 | Host Plant | Contain host code plants, origin isolates (host and substrate) e.g. corn, peanuts, etc |
| 11 | Growth Condition | Contain kode dan growth condition (Medium, Cultivation Temp, Rehydration Fluid, Relative humidity, pH, Water activity, Aeration ($O_2$ and $CO_2$), light) |

Table 4. Conceptual database for insect collection data management.

| No | Table name (entity) | Information |
|---|---|---|
| 1 | Kingdom | Contain kingdom code and kingdom name |
| 2 | Filum | Contain phylum code and phylum name |
| 3 | Class | Contain class code and class name |
| 4 | Ordo | Contain order code and order name |
| 5 | Family | Contain family code and family name |
| 6 | Genus | Contain genus code and genus name |
| 7 | Species | Contain species code and species name, description and references |
| 8 | Collector | Contain collector code, collector name and initials |
| 9 | Status specimen | Status specimen: W, wet collection in fluid preservative; C, culture of live insects; P, pinned dry collection |
| 10 | Specimen | Contain specimen code, collection number, image, location, coordinate position, description, collection date |

The resulting conceptual database is translated into a more technical format as a logical framework database. An object-relational database approach is used to create relational diagrams of database management systems for weed, IAS, insect, and mushroom collections, as shown in Figure 3.

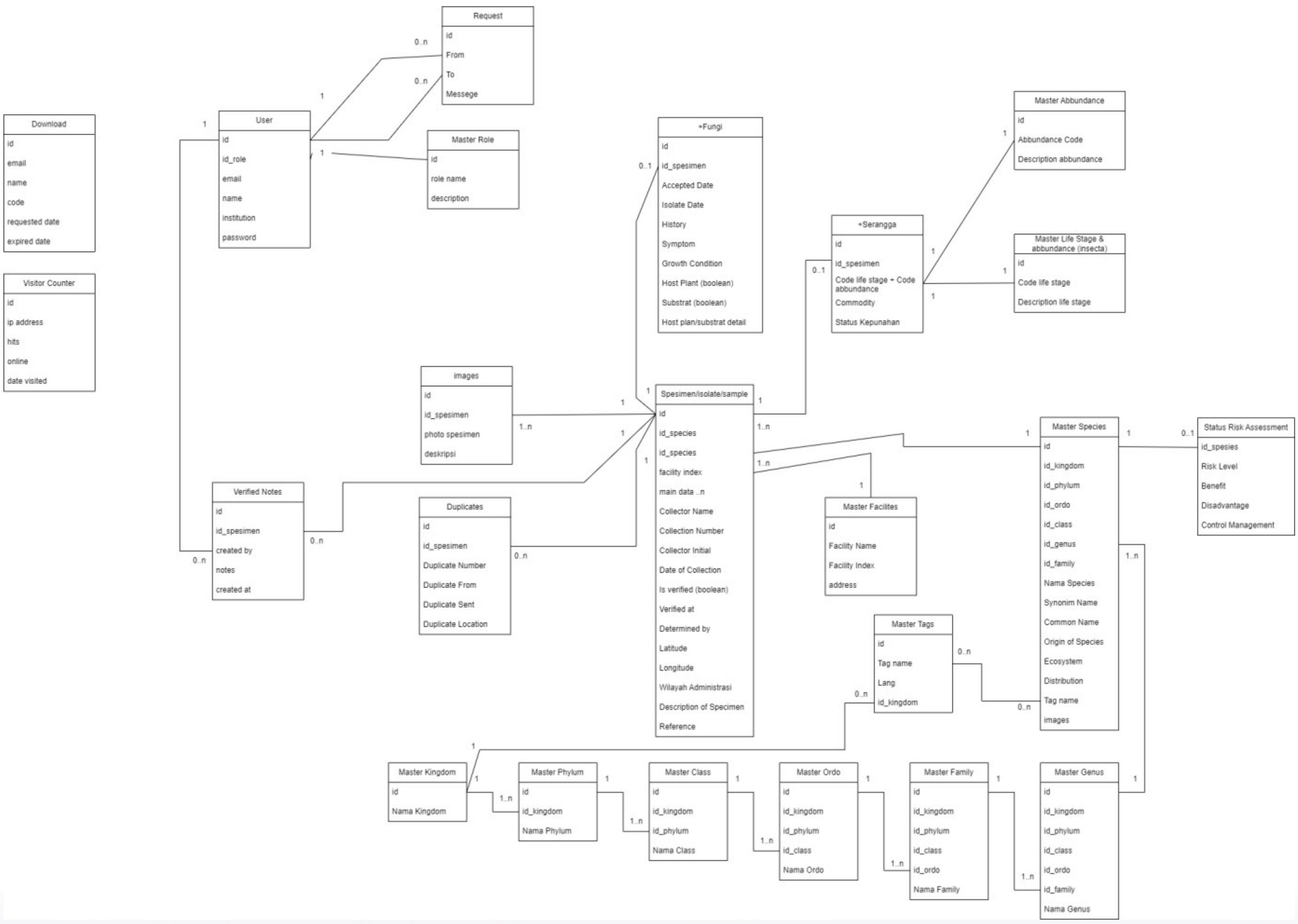



Figure 3. ER Diagram of SEAMEO BIOTROP integrated biodiversity heritage collection database.

The intensity and quality of user interaction with the SEAMEO BIOTROP collection database is the key to the success of the developed database management system. An interactive prototype of the interface has been designed, considering the identified stakeholders' needs. An attractive, interactive, responsive, easy-to-read interface and easy-to-find information that users are looking for will have a positive impact on the level of database utilization by users widely. Input from experts has been collected through FGDs, analyzed, and internalized into the features and menus available on the database management system interface prototype. Figure 4 presents an interface prototype that has been generated based on the input and opinions of experts.

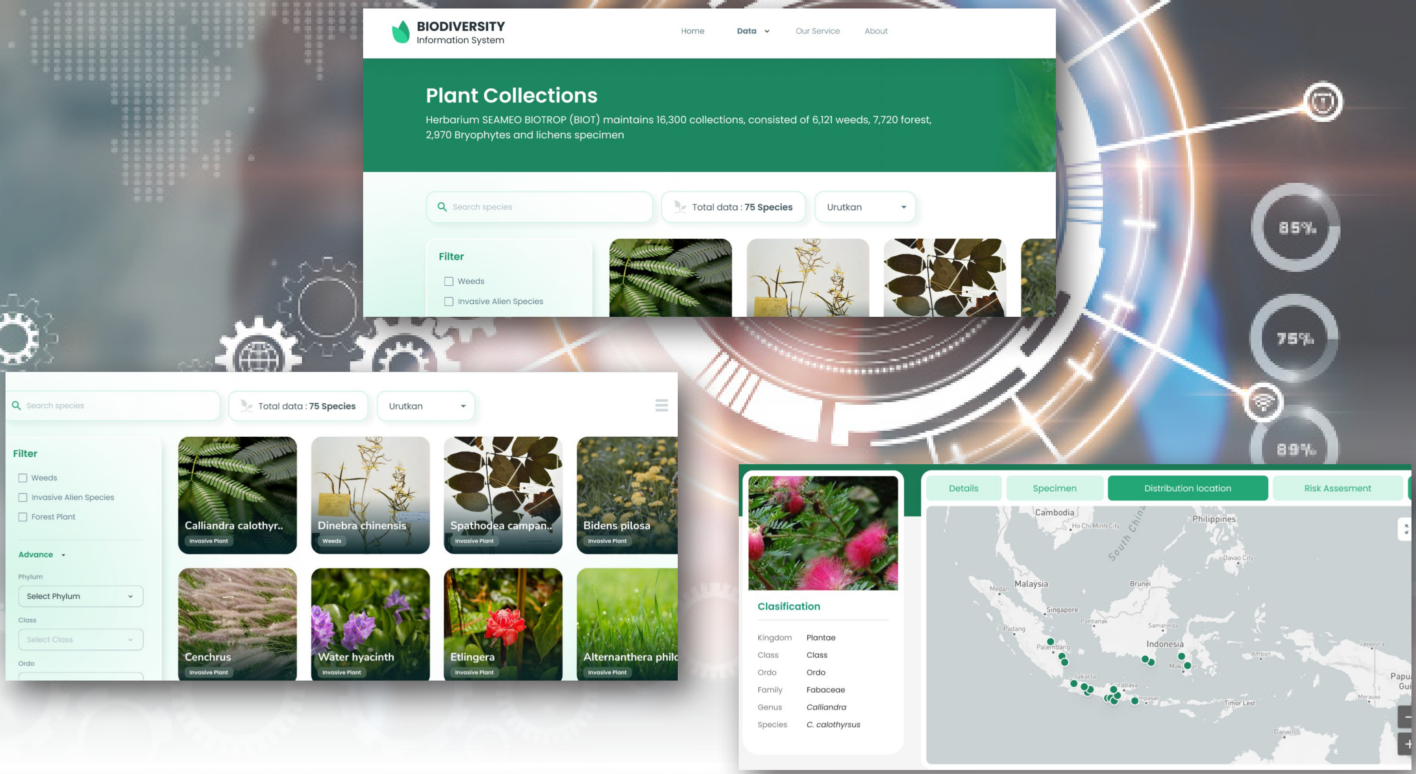
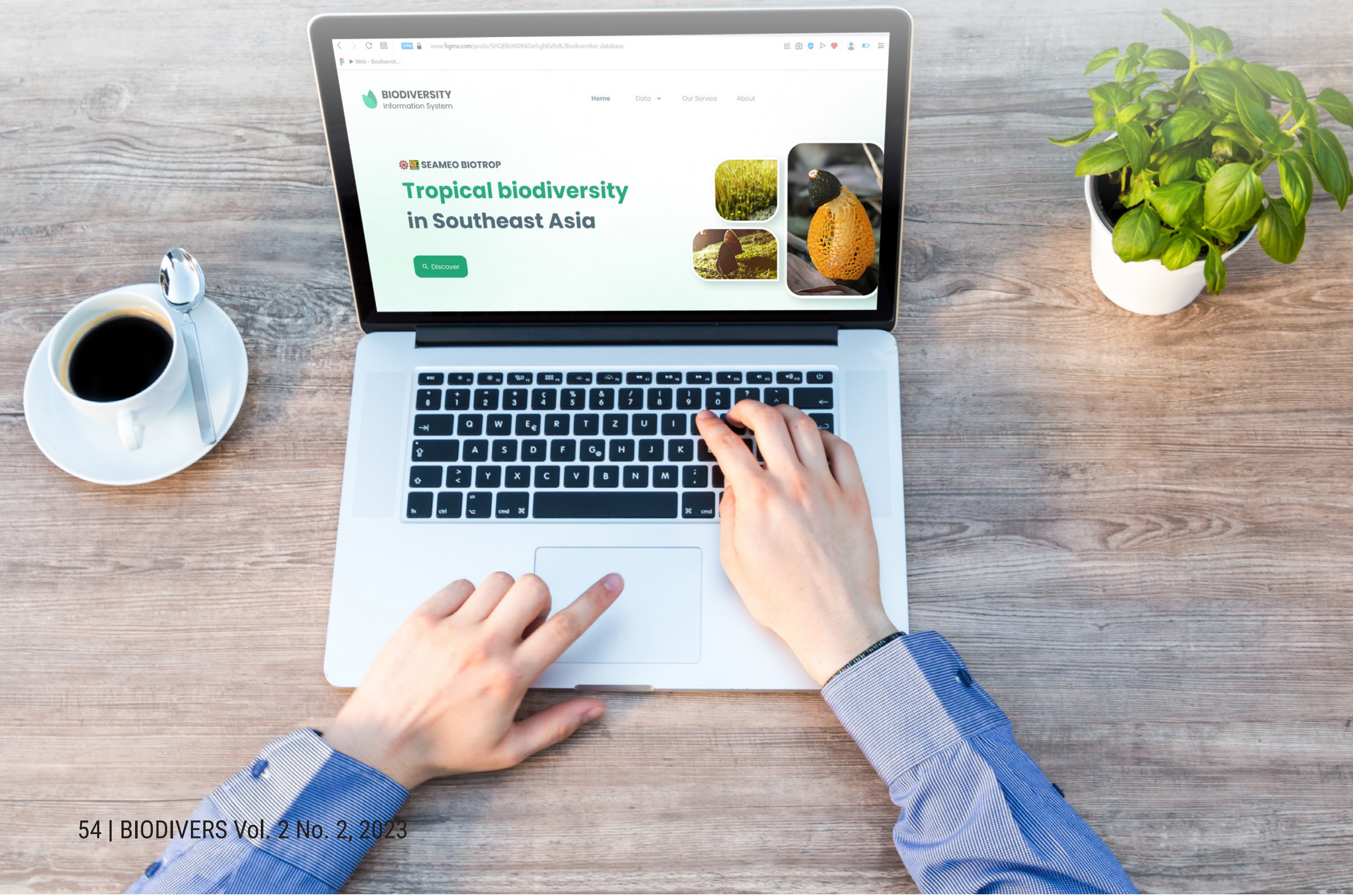



Figure 4. Prototypes interface of SEAMEO BIOTROP integrated biodiversity heritage collection database management system



## Conclusion

The study has successfully developed a framework for an integrated database of heritage biodiversity collections at SEAMEO BIOTROP. A guidebook has been written, which is expected to become a standard reference in taking samples of collections or specimens in the field, a reference in recording collection or specimen data, and a reference in developing a sustainable database management system.

## Acknowledgements

The authors thank the Ministry of Education and Culture, Research, Technology and Higher Education, the Republic of Indonesia, for providing financial support through SEAMEO BIOTROP DIPA for the fiscal year 2022. The authors also thank PiAREA and staff for providing technical assistance in the development of the database management system prototype.

## References

Cardinale, B. J., Duffy, J. E., Gonzalez, A., Hooper, D. U., Perrings, C., Venail, P., et al. (2012). "Biodiversity loss and its impact on humanity," Nature, vol. 486, pp. 59-67, 2012/06/01.

Hooper, D. U., F. S. Chapin Iii, J. J. Ewel, A. Hector, P. Inchausti, S. Lavorel, et al. (2005) "Effects of biodiversity on ecosystem functioning: A consensus of current knowledge," Ecological Monographs, vol. 75, pp. 3-35, 2005/02/01.

Reid, A. J., A. K. Carlson, I. F. Creed, E. J. Eliason, P. A. Gell, P. T. J. Johnson, et al. (2019). "Emerging threats and persistent conservation challenges for freshwater biodiversity," Biological Reviews, vol. 94, pp. 849-873, 2019/06/01.

Dudgeon, D., A. H. Arthington, M. O. Gessner, Z.-I. Kawabata, D. J. Knowler, C. Lévêque, et al. (2005). "Freshwater biodiversity: importance, threats, status and conservation challenges," Biological Reviews, vol. 81, pp. 163-182.

Hooper, D. U., E. C. Adair, B. J. Cardinale, J. E. K. Byrnes, B. A. Hungate, K. L. Matulich, et al. (2012). "A global synthesis reveals biodiversity loss as a major driver of ecosystem change," Nature, vol. 486, pp. 105-108, 2012/06/01.

Sutherland, W. J., R. P. Freckleton, H. C. J. Godfray, S. R. Beissinger, T. Benton, D. D. Cameron, et al. (2013) ."Identification of 100 fundamental ecological questions," Journal of Ecology, vol. 101, pp. 58-67, 2013/01/01.

Jin, J and Yang, J. (2020). (2020). "BDcleaner: A workflow for cleaning taxonomic and geographic errors in occurrence data archived in biodiversity databases," Global Ecology and Conservation, vol. 21, p. e00852, 2020/03/01/ 2020.

Guralnick, R. P., P. F. Zermoglio, J. Wieczorek, R. LaFrance, D. Bloom, and L. Russell. (2016). "The importance of digitized biocollections as a source of trait data and a new VertNet resource," Database, vol. 2016, p. baw158.

Nelson, G. & Ellis, S. (2019). "The history and impact of digitization and digital data mobilization on biodiversity research," Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 374, p. 20170391, 2019/01/07 2019.

Lin, Y.-P., W.-C. Lin, W.-Y. Lien, J. Anthony, and J. R. Petway. (2017). Identifying Reliable Opportunistic Data for Species Distribution Modeling: A Benchmark Data Optimization Approach. Environments 4(4).

Darrah, S. E., L. M. Bland, S. P. Bachman, C. P. Clubbe, and A. Trias-Blasi (2017). "Using coarse-scale species distribution data to predict extinction risk in plants," Diversity and Distributions, vol. 23, pp. 435-447, 2017/04/01 2017.

Banerjee, A. K., A. A. Khuroo, K. Dehnen-Schmutz, V. Pant, C. Patwardhan, A. R. Bhowmick, et al. (2021) "An integrated policy framework and plan of action to prevent and control plant invasions in India," Environmental Science & Policy, vol. 124, pp. 64-72, 2021/10/01/ 2021.

Pelletier, T. A., B. C. Carstens, D. C. Tank, J. Sullivan, and A. Espíndola (2018). "Predicting plant conservation priorities on a global scale," Proceedings of the National Academy of Sciences, vol. 115, pp. 13027-13032, 2018/12/18 2018.

Cabrelli, A. L., A. J. Stow, and L. Hughes (2014). "A framework for assessing the vulnerability of species to climate change: a case study of the Australian elapid snakes," Biodiversity and Conservation, vol. 23, pp. 3019-3034, 2014/11/01 2014.