A STUDY OF DATA COMPONENTS AND KNOWLEDGE COMPONENTS FOR PLANT BIODIVERSITY AREA BY USING SPATIO-TEMPORAL DATA MODEL TECHNIQUES

KAJIAN KOMPONEN DATA DAN KOMPONEN PENGETAHUAN UNTUK KAWASAN TUMBUHAN BIODIVERSITI DENGAN MENGGUNAKAN TEKNIK MODEL SPATIO-TEMPORAL DATA

Mohd Taib Wahid

PM Abd Mannan

PM Dr. Harihodin Selamat

Pn Norhidayati Zakaria

Research Vote No: 75164

Jabatan Sistem Maklumat
Fakulti Sains Komputer Dan Sistem Maklumat
Universiti Teknologi Malaysia

2006

ACKNOWLEDGEMENT

In the name of Allah, the Most Gracious, the Most Merciful, for giving me the determination and will to complete this study.

We would like to express our sincere thanks and appreciation to researcher for their advisors PM Abd Mannan Ahmad, Associate Professor Dr. Harihodin Selamat, and Pn Norhidayati Zakaria for their continuous advice, guidance, support, and have patiently guided me through the steps of this research. We thank them for being co-researchers.

We would like to thank to my research assistant, practical students who work in database lab, colleagues, friends, staff, and lecturers in the Faculty of Computer Science and Information Systems for their help, facilities and for providing who have extended their fullest cooperation and encouragement.

Last not least, We would like to thank to Research Management Centre (RMC) for providing us research grand to carry on the research.

.

Abstract

The economic importance and uses of the large number of biodiversity plant species of Malaysia make it essential for their biodiversity to be conserved. The complexity of natural history collection information and similar information within the scope of biodiversity informatics poses significant challenges for effective management of plant biodiversity data. In order to undertake the steps for biodiversity conservation such as identification of species, monitoring climate conditions, it is essential to efficiently manage the vast amount of biodiversity related data. This project discussed about the object-relational conceptual design of biodiversity data model for long term stewardship of biodiversity information. In this research design a biodiversity data model that better facilitates the exploration and analysis of spatio-temporal biodiversity data by using object relational data model techniques. The conceptual model that is efficient for data management and able to integrate diverse set of spatio-temporal data for analysis, and monitoring biodiversity data. Moreover, forest department, land and agricultural management and other related research organization could be benefited from this research.

Abstrak

Kepentingan ekonomi dan penggunaan spesis tumbuhan biodiversity yang terdapat dalam Malaysia dalam kuantiti yang besar membuatkan ia perlu diperlihara. Sejarah maklumat dikumpul yang kompleks dan maklumat yang sama dalam skop maklumat *biodiversity* merupakan cabaran untuk membuat pengurusan yang efektif. Sejajar dengan itu, langkah pemuliharaan perlu di ambil seperti butiran pengenalan spesis, mengawal keadaan cuaca kerana ia penting untuk mengurusakan jumlah data biodiversity yang besar. Tesis ini membincangkan tentang rekabentuk konseptual model data *object-relational* untuk memenuhi keperluan maklumat biodiversity. Kajian ini juga merekabentuk model data untuk biodiversity mempunyai fasiliti yang baik untuk pencarian data dan analisis untuk data biodiversity spatio-temporal menggunakan model data object-relational. Model konseptual yang efisyen untuk pengurusan data dan boleh diintergrasikan dengan set data spatio-temporal dan melihat data biodiversity. Tambahan, organisasi perhilitan, tanah dan pengurusan agrikultur dan beberapa organisasi yang terlibat boleh menerima kebaikan daripada kajian ini.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ADT	Abstract Data Type
APASD	Asian-Pacific Alien Species Database
APMIS	
ASC	Association for Survey Computing
ASEAN	Association South East Asian Nation
ASP	Active Server Page
BCIS	Biodiversity Conversation Information System
BD	Biodiversity Feature
BDBMS	Biodiversity Database Management System
BF	Biodiversity Feature
BODHI	Biodiversity Object Database System
CBD	Convention on Biological Diversity
DBMS	Database Management System
DNA	Deoxyribonucleic acid
ESTDM	An Event-Based Spatio Temporal Data Model
GIS	Geographical Information System
GLOBIO	Global Methodology for Mapping Human Impacts on the Biosphere
IOPI	International Organization for Plant Information
IT IS	Intergrated Taxonomic Information System
IUCN	International Union of Conservation of Nature and Natural Resources

MACRES	Malaysian Centre for Remote Sensing
MARDI	Malaysian Agricultural Research and Development Institute
MONRE	Ministry of Natural Resources and Environment
MOSTI	Ministry of Science, Technology and Innovation
MS ACCESS	Microsoft Access
MS dot net	Microsoft dot Net
NGO	Non Government Organization
OBM	Object Behavior Model
OID	Object Identity
OODBMS	Object Oriented Database Management System
OR	Object Relational
ORDBMS	Object-Relational Database Management System
ORM	Object Relational Model
OSHADHI	
RDB	Relational Database
RDBMS	Relational Database Management System
SD	Spatial Data
SQL	Structured Query Language
Stop-BDM	
Stop-BIODI	Spatio Temporal Object Relational Plant Biodiversity
TD	Temporal Data
UKM	Universiti Kebangsaan Malaysia
UM	Upper Model
VB.net	Visual Basic dot Net
WWF	World Wildlife Fund
XML	Extensiblel Markup Languange
XP	Microsoft XP

INTRODUCTION

1.0 Introduction

Over the last couple of years, there has been phenomenal growth in the area of bio-diversity studies, largely motivated by the economic and humanitarian potential that underlines the understanding of biodiversity dynamics. New area has been arisen which is called Bio-prospecting, which focus is solely on shifting through biodiversity data to locate potentially profitable biological sources (Srikant *et al*., 2000).

The major difficulty was faced by biodiversity scientist, is the effective management and the access of large amounts and varied types of data arise in their studies, ranging from micro-level biological information (genetic makeup of organism and plants) to macro-level information (including ecological and habitat characteristics of species).

Biological diversity plays an important role in our lives. The interaction of species within highly diverse ecosystems performs ecological functions that are extremely important to many human activities. Ecosystems function to maintain hydrological cycles, regulate climate and recycle essential nutrients in the soil to maintain its fertility.

Researchers have been given various definitions on biodiversity data. Perhaps, the best definition of biodiversity is the following (adapted from the Keystone Dialogue on

Biodiversity in Federal Lands by Noss and Cooperrider (1994): Simply put, biodiversity is the assortment of different types of organisms that co-occur in time and space. Biodiversity should facilitate abstract representations of real world objects which are understandable and easy to use.

Despite having this importance, humanity imposes serious negative impacts on biodiversity at multiple scales of time and space. The popular media regales us with stories and reports of the ongoing destruction of natural habitats around the world, impending climate change disasters and other ecological travesties all having potentially dire consequences upon biodiversity. After more than 150 years of scientific study, it is still not clear what factors or process lead to observe patterns of biodiversity. If we are able to mitigate the cost and develop a proper data model with the real or/and potential loss of biodiversity, we must understand the factors or process to design data model of biodiversity.

1.2 Problem Background

1.2.1 Technical Problem I: Large Amount of data

In studies of biodiversity data modeling, few problems have been categorized about biodiversity data modeling. First, large amount of data, biodiversity data is too large because of the variety of different types of plants species that growth in various landscape also co-occurs in time and space. Facilitate focus attention on specific area of plan biodiversity need to conserve. Second, biodiversity data are complex because biodiversity includes the variety of genes within one species through the complex interconnection of all life within an environment. There is no stable biodiversity database system and biodiversity analysis tool to maintain infrastructure and monitoring current and future data of biodiversity. Third, effective data management system means storing, retrieve and accessing data from the system. There is no such model designed yet which can support data management and time consuming accessing data from the system.

Fourth, most of the data models are relational data model. Rational data model cannot support temporal data, only support spatial data. Fifth, there is no proper biodiversity retrieval system yet to exist. Existed system only provided general information about biodiversity.

1.2.2 Technical Problem II: Biodiversity Data Models

Biodiversity data models or databases have been developed either locally or internationally since 1992.

Table 1: Biodiversity Data Models/ Databases developed since 1992

Model / Database	**Year**
ASC	(Bolton *et al.*, 1992; Sinnot, 1993; Wilson, 1993).
Data model for Botanical Garden and Botanical Museum IOPI's Global Plant Checklist project	(C. McMahon and Berendsohn, 1993; Wilson, 1994).
OSHADHI	Vidya and Haritsa (1995)
UKM Wildlife DNA database	(1996) (name of author)
Database of Malaysian medicinal and aromatic plants	Mat Salleh and Latif (1997)
ITIS	Roy McDiarmid (1997)
BODHI	Srikanta and Haritsa (2000)
Alian Plant Management Information System (APMIS)	Joseph Maada Korsu Kendeh (March, 2002) (not clear)
Biodiversity Database for Malaysian Flora and Fauna	Suhami *et al.*, (2001)
GLOBIO	Tekelenburg *et al.*, (2003)

Relational data models lack the modeling power and extensibility needed for complex objects and operations used in geographic applications. Most current models do not have benefit of fully-fledged DBMS features. Some packages in Banking, commercials GIS marketing use an interface to a relational DBMS to handle spatial and non-spatial data, but temporal data are not in the DBMS. There is a need for DBMS and modeling power better than that of relational DBMS. A few object-relational and object-oriented GISs come to the market based upon spatial analysis. Recently spatio-temporal object relational data model has been developed based on hydrological data. In relational DBMS, a table is created for each entity type. An entity is something that has a physical or conceptual existence. A row or column in a table column corresponds to an entity and attributes. An attributes is restricted to being a built-in type relation schema ends with many additional or extra tables, making less easy to depict the relationship of complex entities in the real world (Weigand, 1994)

Another problem is temporal pattern of Biodiversity data. There are various aspects of data values with respect to time. Current DBMSs is lack extensibility in providing for special and temporal application needs. There are no provisions in traditional systems for adding new data types and methods, add user-defined code or design new storage methods. Another drawback is that difficult to obtain historical data, often collected at varied intervals and diverse scales, for temporal events can be considered as the smallest entity, not easily detected. As a result, a great need for spatio-temporal data has led to the advent of object-relational spatio-temporal Biodiversity data, as most existing data models provides only static historical information in a discrete time structure separated from attribute data. The time at which data values effective is called valid time. On the other hand, transaction time represents the time a data value is recorded in the database (R. Snodgrass, 2000). So the time varying data is commonly represented by time-stamping values. The time-stamps can be time points (Adnan Younas *et al.*, 2004; Dangermond and J., 1990), time interval (Bellamy and S.P., 1996; Benoit, *et al*., 1993), temporal sets (Claramunt and Thériault, 1995) or temporal elements (Armstrong, and M. P., 1988). Further more, these time-stamps can be added to tuples or attributes, the two different approaches for handling temporal data in a relational data model. Time points, intervals, temporal elements, and temporal sets are used as attribute timestamps. Generally, the end points of an interval are added as two separate attributes to a relation. Time points, time intervals, temporal sets and temporal elements are all needed in temporal query languages. They all added to the expressive power of a temporal data model and its query languages (Tansel A. U, 1997)

Biodiversity data have been collected from different department such as Mardi forest department and also analyzed three databases such as *MARDI KUINI* database (contains with 86 tables), *Arthopod Database* (contains with 41 tables) and *MARDI PLANT GENEBANK* database (contains with 69 tables). We also have analyzed Pahang and different area from Johor state's data and their database system. In the Pahang forest database systems are currently storing biodiversity data such as flora and fauna, area, weather and water condition's data. This system has been storing data with simple designing table. This system contains with 20 tables which are hard to manage and make relationship. Beside that, system is only storing general information about those mentioned data, does not support monitoring, and analysis data for prediction or making decision. From the Johor state, we have collected different year's data including floara and fauna from Johor Park Corporation, Tanjung Piai, Pulao Kukup national park, Endau-Rompin National park. They have huge amount of biodiversity data which are collected traditional way from different places and different times. After the suggestion from those departments and analyzing biodiversity data, decision has been made which consist of designing of biodiversity data modeling to manage, to store, to analyse and to monitor biodiversity data including spatial and temporal factor.

1.3.3 Technical problem III: Spatio-temporal Data Models

So far, only a few data models for spatio-temporal data have been proposed. Some of them are based on the raster approach, others on the vector approach and yet other concepts can be applied to both vector and raster data. Some models are based on the object-oriented paradigm, an approach that is capable of integrating both vector and raster data into one data model. Most of the spatio-temporal data models have been developed for GIS such as snapshot model (Armstrong, 1988), ESTDM (Peuquet, and Duan, 1995), Spatio-temporal object model with ST-atoms (Worboys, 1992). One of the problems in adopting the proposed models is that they are based on existing commercial GIS. Most of the commercial systems are closed systems, which cannot be extended or modified (Sylvia de Hoop, *et al*., 1993 as cited by Mioc, et al., 1998).

Research in spatio-temporal data modeling is scattered. Few integrated approaches exist that can treat the spatio-temporal data in a unified manner. Past research provides a basis but not a complete design of a spatio-temporal model (Flor Eugenia, 1999). A spatial data model has been generalized to become spatio-temporal (M. Worboys, 1994). On the other hand, temporal data models have been generalized to become spatio-temporal and include variants of Gadia's temporal model (A. U. Tansel, *et al*., 1993) which are described by (M. H. *et al*., 1998; T. S. Cheng and S. K. Gadia, 1994). The main drawback of all these approaches is that they are incapable of modeling *continuous* changes of spatial objects over time. The definition of a temporal object in general is based on the observation that anything that changes over time can be expressed as a function over time.

Spatio-temporal datasets have some unique characteristics that make them different from traditional relational and transactional datasets. One important difference is that changes can be continuous (for example, the position of a moving object). Traditional database systems assume that data changes through and explicit update. Thus, to avoid continuous database updates on a spatio-temporal database, we need to store description of the changes as a function of time (Peuquet, 2001). Spatio-temporal biodiversity data require complex concepts to describe objects and specify relationships between them. So pure relational data model is not suitable for spatial data, information concerning one object is spread over many relations due to the normalization performed upon the relations, and many join operations have to be performed in order to recreate complex spatial objects. The relational data model alone cannot provide appropriate indexing and retrieval operations for spatial data (Leonid Stoimenov, et al., 1996). Because of continuously growing demands for temporal Biodiversity Information System (BIS) and GIS, Frank explains any object-oriented approach to a property registration system (1996) and spatial and temporal reasoning (Frank, 1997). More progressive study is contributed by Al-Taha (1992) based on a relational model associated with temporal legal changes. Worboys (1992, 1994) emphasizes spatio-temporal elements using object-oriented techniques. Several geographers (eg, Peuquet, 1994; Yuan, 1994) concentrate on features (what), event times (when) and locations (where) based on relational environments.

1.3 Specific Objectives

The main objective of this research is to design and to develop a data model for efficient management for biodiversity data. This research also aims to achieve the following objectives:

a) To study about data and knowledge component of plant biodiversity and spatiotemporal data model techniques.
b) To find out data conservation and data retrieval issues on plant biodiversity.
c) To find out appropriate techniques such as pyramid framework to support spatio-temporal data for plant biodiversity data model.

1.4 Scope of the project

This research focuses on exploring methods to support temporal data with biodiversity data model. Thus the scope of this research involves:

Scope of the study will focus data conservation, data modeling and data retrieval issue on plant biodiversity. Initial research would be focused on plant biodiversity to find out current research issues on plant biodiversity. The study also will enfold on pyramid frame work to support spatio-temporal temporal data with plant biodiversity data modeling.

CHAPTER 2

LITERATURE REVIEW

2.0 Introduction

This chapter presents a review of the literature covering data management issues in biodiversity informatics. The review of the literature begins with a definition of key terms used in this document, and then gives as overview of the content of the literature. It continues to review the requirements for the design of a spatio-temporal biodiversity data model. Biodiversity is the main issue in the recent world. It is very important for a wide uses, including scientific study, resources management and policy planning and education. Many researchers and developers are working to measure biodiversity data and establishing a core set or framework of biodiversity indicators. In the previous chapter, we have been discussed about the biodiversity data, biodiversity of Malaysia, problem background of biodiversity data models. This literature survey only focuses on biodiversity data model include time factor to build biodiversity data model. This chapter also elaborates data models background, classification of biodiversity data, spatio-temporal data and data models. This research comparison among data model has done to find out best data model can be suited to the biodiversity.

2.1 Malaysian Biodiversity Data

Malaysia has been endowed with vast amount of natural resources including luxuriant tropical forest which is one of the most diverse and complex ecosystems of the world. Malaysia is undoubtedly one of the richest rain forests in the world with diverse floristic composition and complex ecosystems. Malaysia is thought to have about 40,000 species, which is concludes 12,500 species of flowering plants (Mat Salleh and Latif, 1997), about 1,200 (10%) species of medicinal plants. The flora and fauna of Malaysia are exceedingly rich and unique, has mixture of the greater Malaysian elements, Asiatic elements and Australian Elements as well as the Pacific elements (Keng, 1970). This mixture of plants species itself is benefited to the region as they share many common medicinal species. Malaysia shares many common medicinal plant species with India and Indonesia. Keng (1970) estimated that over 8,019 species of speed plants occur in Peninsular Malaysia. Estimate for the whole country could be about 12,000 species. However, this estimation is still under process and that is area in Sabah and Sarawak has not been actively surveyed botanically. The above estimation has been holding since 16th century by the colonist and recently introduced by plant importers (Mohamad Bin Osman and Zaki, 2002).

More than three decades Malaysia experienced as extremely rapid socio-economic growth. Malaysia has been losing much of its natural resources including plants and animal species through ecosystem and degeneration. Several organizations all over the country are working on the implementation of the biodiversity conservation measures. One of the major hurdles faced by the conservation scientists and researchers is the management of vast amount of biodiversity related information that available. This information includes data on species characters (e.g., flora, fauna, peat swamp, and habitat), endangered species, species of medical value, climate conditions (e.g., temperature, rainfall), soils and land use, geographic distribution of species, botanical garden and herbariums, geographic regions, expert related fields and a host of others information.

University Kebangsaan Malaysia (UKM) launched first online biodiversity information system supported by Malaysia-Japan Tropical Bioresearches Project in 1996. Currently, the server hosts number of databases, which are available and targeted to different segments of biodiversity stakeholders. These include the checklist of Malayan Vascular Plants, the Herbarian UKMB Online, Plant Systematic, Conservation and Ecology Reference Online, Ethnobotany of Malaysia Plants, Native Edible Plants of Sabah, The Flora of Fraser's Hill & Bangi Permanent Forest Reserve, and UKM's Wildlife DNA Database (Suhami Napis *et al*., 2001). In Early 1999, Singapore Botanical Gardens, Forest Research Institute Malaysia and UKM hosted a joint online catalogue of more than 9,000 species of vascular plants. From this database, any one can list the species occurring in the country also possible to query species from any imaged-mapped states in the Peninsular Malaysia and Singapore. Even the basic information can be searched from the menu and others information will display next to pictures or illustration of that particular species.

Ethno-botany of Malaysia Plants online complemented all these databases to develop a comprehensive database for Malaysian (Peninsular Malaysia, Sabah, and Sarawak) medicinal, poisonous, and aromatic plants. These web base systems are having nearly 16,000 native plan names from about 400 species and being updated regularly with new information (Mat Salleh and Latif, 1997; Suhami Napis *et al*., 2001). Later on, this system was planning to enhance this category with GIS and other data, as it is available (Suhami Napis *et al*., 2001).

To achieve the objective of designing biodiversity data model, several problems are needed to overcome; these are related to biodiversity data pattern, data analysis, landscape structure and representation of space and time, estimate species richness, query language, query processing algorithms, geographical user interfaces, storage structures and query plan approach and finally architectures of Biodiversity Database Management System (BDBMS).

2.2 Biodiversity Data and Conservation

Humanity depends on biological diversity, or biodiversity for the very sustenance of life. Adequate data on species diversity, populations, location and extent of habitat, major threats to different species, etc.; and changes in these aspects over time are not available to design a proper strategy for conservation. Given our extensive diversity, ecological surveys and taxonomic investigations need to be intensified, particularly for plants and insects. For conducting ecological studies, species for such studies can be prioritized including keystone species, large mammals, migratory species, etc. The country's wildlife institutions need to network and coordinate their activities so that priority issues and areas are identified. The Ministry of Environment and Forests through the Botanical Surveys of Malaysia and the Zoological Survey of Malaysia could play a guiding role by preparing a list of priority issues and areas for circulation to relevant institutions, based on a countrywide consultation of experts. Funding for these prioritized projects could be stepped up to ensure that research focuses on these issues. They are organizing seminars, conferences to protect the biodiversity data. Recently, MACRES, MOSTI, KAS and MONRE (write full name one time) organized workshop on various issues. One of the key issues was national spatial biodiversity database development. Those departments must set up a database for the country as a whole. There are many organizations and NGO’s working on biodiversity such as Biodiversity Conservation Information Systems (BCIS) working on information security and decision making on the conservation and sustainable use of living resources, while Asia Specific Alien Species regulate invasive alien species and development database to minimize economic damage. Locally WWF (World Wildlife Fund) Malaysia is working for forest, marine, wetlands, and species education and awareness. In addition, Global Biodiversity Information Facility develops an electronic catalogue for taxonomy names. There are over 486,000 scientific names, 217 common names listed with the word with tentative position taxonomy.

Various organizations nationwide and internationally have been involved with biodiversity data and databases. Currently, World Wildlife Fund (Malaysia, China, Indochina etc.) has been working on various projects such as forest, marine costal, wetlands, species, education and awareness. ASEAN (write full name one time) regional center for Biodiversity Conservation and ASEAN Review of Biodiversity and Environmental Conservation (CBD) are also working on Conservation, Genetic Resources, Technology, and Bio-technology. CBD is also developing biodiversity smart partnership database. In addition to this, IUCN- Asia Biodiversity Program is the world's largest and most important conservation network.

After studying and analyses of biodiversity data, we can see that many organizations, NGO's have been working on biodiversity data and also have been developing database system to store biodiversity data. Those biodiversity data contains only general information about biodiversity data. Some organizations only have 2 tier web systems. Those systems have database systems which are not well designed and do not support analysis and all kind of data. Example is Asian Specific Alien Species (APASD) database system, has developed for all kinds of alien species. This database only contains general information about the species and does not support analysis and monitoring biodiversity data.

2.3 Space and Time Factors in Biodiversity Data

Time and space are important aspects of all real world phenomena. Database applications must capture the time and space varying nature of the phenomena they model. For example, biodiversity data collector needs to maintain a record of the data, collected from various areas with time. This information certainly varies over the times and the data collected during particular time may be collected same body and some where else; and also data collects should be periodically such as daily, monthly and yearly. Many applications demand for spatial temporal support (Spaccapietra, 2001), such as biodiversity applications, GIS applications, traffic management systems computing traffic network and traffic volume at different time, forecast-prediction systems recording weather change process, etc. The amount of spatial and temporal information is growing very rapidly. To satisfy the need of applications, different systems have been developed for various sub-areas related to multidimensional data. Among them, biodiversity information systems constitute a large field of interest. But existing database techniques couldn't support spatial temporal information management well (Tryfona and Jensen, 1999).

After almost two decades research, representation of space and time in databases and functional applications are still problematic (Peuquet, 2001). This paper presents a universal object-relational framework for spatial temporal data modeling. Spatial temporal data modeling aims to extend the existing data models to include space and time in order to better describe our dynamic real world.

Numbers of issues and various aspects have been identified by many researchers in the development of biodiversity data model including spatio-temporal data model, query languages and data accessing method. Various issues have been discussed in this thesis with literature review, results and discussion, conclusions and recommendations for future work. After being literature review, it was noted that, to develop a pattern bio-data; is not easy and structural model is not suitable to store the data. There are several spatial (location), temporal (time) data model exists, which are having problem with storing and retrieving data from different location beside that temporal model have time period problem (present, past and future data). According to "statement problem", it is difficult to retrieve and analyze the data. In the Geographical Information System (GIS), although many data model already existed but those data models are not related to biodiversity and cannot fit with the biodiversity data model and its characteristics.

After literature review of related to biodiversity data model. Most of the biodiversity and GIS modeling is relational data model. Those models can not support complex data, data analysis, and data retrieval. Also relational data model only can support spatial data not temporal data. After analysis of all models developed since 1993, most of the models are using to collect plant data, plant listing, and plant conservation but there is no data yet to design which can support data analysis, data retrieval, temporal data (time). Early 2000, internet based biodiversity object database system (BODHI) developed also has mentioned above paragraph to handling plant taxonomies. To support spatial and temporal of biodiversity data, one robust data model is required to developed. The purpose of this paper is to develop data model that better facilitates the exploration and analysis of biodiversity data with time factor. The goals of this research, to manage and integrates diverse sets of spatio-temporal plant diversity data together with a county-level geographic information system to build a biodiversity mapping data model and information retrieval system for plant. So that those data may be required, visualized and entered into statistical analysis. Also, allow the explicit representation of plant taxonomies, dynamic process, relationship that compose the biodiversity system in a manner that is intuitive and useful to the researcher. Meeting these goals demands a database model that not only efficiently manages large quantities of biodiversity data, but also retrieve data from data base so that researcher can make analysis.

As we point out, researchers have been trying to incorporate the time component in these data models while maintaining their original design (Peuquet & Daun, 1995). The challenge consist in the design a biodiversity data model to support spatio-temporal data, we have chosen pyramid frame work to support a biodiversity data model design. More description about pyramid framework is discussed next page:

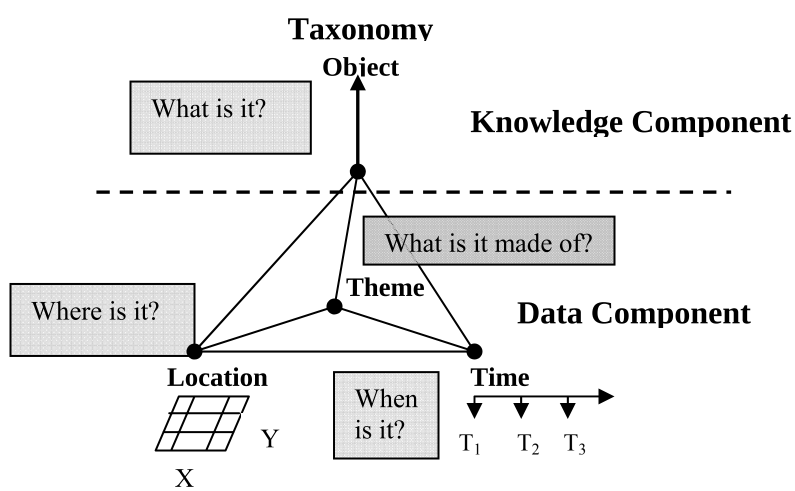


Figure 1: Pyramid Frame work: Data Component and Knowledge Component

A conceptual frame work (also know as pyramid frame work) was designed which guide the implementation of the semantic GIS database model [10]. This frame work is interrelated with two separate parts one is data component and another is knowledge component shown in Figure 1. Data component can be divided into three categories: location (position in the spatial three dimensions), time (position alone a time line) and theme (a set of observations, measurements, or attribute values associated with a particular location and time). Data components stores spatio-temporally referenced observational data such as *spectral*, *climate*, *vegetation* and *other environmental attributes*.

Data component can be computerized as a multi-dimensional, spatio-temporal referenced 'hyper-cube' of observational data that is similar to the 'feature space' concept commonly cited in analysis of remote sensing imagery. Knowledge component stores information about higher-level semantic 'objects' the geographic entities such as location and space All the objects are also placed within two hierarchical relationship structures central to cognitive knowledge representational and object-oriented modeling, plant taxonomy (generalization). The plant taxonomy structures groups similar objects within a category and stores a rule-base that describes how those objects may be identified within the data space.

Model represents, in fact, a frame work of many existing model (e.g. IMAGE 2.2, Alcamo et. al. 1998) and several new ones. In the biodiversity model we have chosen pyramid framework concept because its combination of object-oriented modeling and Artificial Intelligence (AI) knowledge representation techniques [14, 16]. Many programming languages such as Java and C++ are object oriented. Integration between the two would provide easier abstraction for developers who programs in object-oriented languages and required to 'program' in SQL [20]. Such integration would also relieve the necessity of constant translation between databases table and object-oriented data structures.

However, modeling the spatio-temporal phenomenon remains a complex task due to the large diversity of the entities and phenomena and their possible evolution in the real world. We need a data model and corresponding technologies to depict the dynamic entities and phenomena. Basically, models of this kind should treat an entity as a whole and can provide us as much more rich information as possible, not only data about its current state, but also other data about its past evolution.

Spatiotemporal data models have received much attention in the database research community because of their practical importance and the interesting technical challenges they pose. In the literature review the discussed previous research that most influenced our work on spatiotemporal data modeling. Much previous work focuses on temporal information or spatial information, rather than both.

The main reason for the lack of universal data model is that, models were not originally designed to support time. Researchers have been trying to incorporate the time component in data models while maintaining their original design. The challenge consists in the design of a spatio-temporal biodiversity data model, in parallel with the development of appropriate databases, the development of efficient spatio-temporal biodiversity query languages

2.4 Geographical Information System (GIS)

There are countless definitions for GIS, each based on the type of user and application domain. GIS definitions perhaps based on functions, toolbox, data, technology or decision support system. Roughly, GIS is a database system that using for storing georeferenced data. GIS is an information system consists of input; process and output function for georeferenced data. The term input refers to data about the storing data into the database. The term process includes capture, modeling, manipulating, retrieval and analysis of geographically referenced data, instead output were displaying and presentation of georeferenced data. Georeferenced refers to data about geographic phenomena associated with its location, spatially referenced to the Earth.

Geographical data can represent into following method; points is a single node with NO area (e.g. building), lines is a connection of nodes beginning with a “to” and ending with a “from” (e.g. road, river) and polygon is a series of arc(s) that close around a “label” point (e.g. town, a district). Geographic data component can be classified into three main categories:

2.5 Graphical Interface for Biodiversity

Biodiversity is uniquely characterized by its multi-dimensional attributes, though it has been conventionally prototyped on various plant image or map sheets. To reach a good graphical interfaces and effecting query, and the display of plant data include geographical data should be in multi-dimensional mode and be interacted preferably from various sites.

This aim can be reached by integrating advance in computer graphics and web technology. Advance graphical tools such as ArcInfo, Manifold, ArcView, GIS software digitalize the image and later on stored in database as different image layers makes it possible to develop effective model and increase capability to operate large database on biodiversity. In addition, Andrei Ju, Korolyuk used ArcInfo to evaluate biodiversity in Southern Siberia. Most of the graphical tools (ArcView, ArcSDE) support various RDBMS platforms like DB2, SQL Server, and Oracle 9i. Acquisition of image of plant will be collecting from various places also the temporal data. These images provide digital mosaic of spatial attributes and plant types liable to computer processing.

As stated in (de By, 2000), a database is a repository capable of storing large amounts of data, equipped to facilitate concurrent use, storage optimization, data integrity, query facility and query optimization. GIS systems are not very good at these. However, the strength of GIS lies in its ability to operate on spatial data, and allows analyses that are inherently geographic in nature. A GIS software package like ArcView has powerful but highly flexible tools for map production of both hard and digital types. ArcView combines in various ways the attributes of plant invasion, the management practices in place, with spatial data layers, undertake spatial analysis and produce the results for visualization mostly through maps.

On the other hand, web technology is widely used as a platform to distribute, query, visualize and process taxonomy and geographical data. Integrating graphical software with web technology enables a simultaneous interaction with the display in global and local level.

2.6 Storage Structure for Biodiversity Data

A model for a database system, that provides a standardized environment for submission, storage, and retrieval environmental data such as geographical and biodiversity data (Scott, J.M, Davis, F.W, Scuti, 1993). Adopting a relational view, such databases are collections of entities which have either spatial attributes (e.g., geographic databases) or temporal attributes (e.g., medical databases) or combinations thereof (e.g., multimedia databases). Spatial attributes can be viewed as 1D, 2D, or 3D positions in a "space," either the physical one (e.g., map objects) or an artificial one such as a computer screen (e.g., multimedia objects). Temporal attributes capture the temporal existence of entities and in the general case can be represented as time points or time intervals [12].
The preliminary model is based on spatio-temporal data model techniques and is suitable for not only storing data, digital images, landscape map, but also for modeling domain knowledge associated with plant based data include spatial and temporal data.

Significant researches have been done to study of storage structure. Existed database only support spatial data not temporal data. Database should efficiently support retrieval of data based on the spatio-temporal extents data. Multi-dimensional access method is required to achieve this database. Many researchers has already initiated their work in this area such as approaches that extended R-trees and Quad tress (Scott, J.M, Davis, F.W, Scuti, 1993) alone with extensive experiments on a variety of synthetic data sets.

Spatial structures and efficient spatial access methods based on hierarchical regular decomposition of the space for images containing multiple non-overlapping or overlapping features were proposed. Regarding temporal structures, overlapping B^+-trees were used for transaction-time in implementation of time access method. B-tree supporting transactions were proposed in optimal multiversion B-tree (] B. Becker, S. Gschwind, T. Ohler, B. Seeger, P. Windnayer, 1996). Spatio-temporal indexing method appeared in: 3D-trees, MR-trees and RT-trees, and HR-trees (Mohamad Bin Osman and A.H Zaki, 2002). All these methods are suitable for spatial objects and not suitable for representing regional data. Indexing method also employed in large biodiversity and multimedia database (Vazirgiannis, *et al*., (1998), (] M. Miick and M. Polaschek, (1997). M., Theodoridis, Y., And Sellis, T., (1998)used query for (1) search in the complete inheritance hierarchy rooted at PlantSpecies and return object of type PlantSpecies and MedicinalPlants associated with the GeoRejion, (ii) search object of only PlantSpecies type associated with given GeoRegion, without searching for MedicinalPlants. To efficiently support both types of queries, the Multi-key Type Index (MT-index) (M. Miick and M. Polaschek, 1997) approach used in BODHI. The basic idea behind TM-index is a mapping algorithm, called Linearization Algorithm. TM-index over type hierarchy, turn out to be a multi-dimensional index with type as a dimension, after the linearization of the inheritance hierarchy as proposed in (M. Miick and M. Polaschek, 1997). SHORE storage manager already provides a multi-dimensional index, R^+-tree.

From the above discussion, application-oriented discipline of biodiversity mapping and geographic information systems, in particular algorithms, spatial and temporal data structures will have significant practical impact in the near future.

2.7 Query Processing for Biodiversity Databases

There are few processes and optimization has focused on (i) development of efficient strategies for processing biodiversity, spatial, temporal data, (ii) efficient modeling cost for query processing optimization purpose and (iii) study about biodiversity data include spatial and temporal constraint databases.

The query processor of BODHI (Srikant B.J, Jayant R. Haritsa, 2000) integrates the features provided by the service modules through extended ODL/OQL for modeling and querying the database. In addition, it optimizes the queries using the metadata and indexed information, and as part of the Client Index Framework it facilitates transformation of query results into an interchangeable format as request by the user. Example of query is written in the BODHI system, using OQL, as follows:

```
Select * from species1 in PlantSpecies,
        species2 in PlantSpecies,
where species1.flowerchar.inflochar=species2.flowchar.inflochar
and
species1.georegion overlaps species2.georegion
```

For query optimization purpose, analytical models that estimate the cost (Theodoris, *et al.*, 1998), nearest neighbor (Papadopoulos A. and Y. Monolopoulos, (1997), and similarly (Papadopoulos A. and Y. Monolopoulos, 1998) queries involving R-tree indexed spatial data were introduced. Analytical models also used as a platform to support direction (e.g. north, north-east) between two dimension objects. Problem studied about *parallel and distributed similarity* search in a shared –nothing multi-computer architecture, where an R-tree is dispersed among the site of the network (Papadopoulos A. and Y. Monolopoulos, (1996). Similar query processing techniques retaining low response times during query processing. DeokHwan Kim and ChinWan Chung, 2003) proposed content based image retrieval method using adaptive classification and cluster merging to find multiple cluster of a complex image query. In addition, this proposed technique provides significant improvement over the query point movement and the query expansion in terms of retrieval quality.

Query processing is lacking in biodiversity applications, where the data and query objects changes their locations. To support biodiversity data for high retrieval quality, accurate data and without losing time need a robust query processing techniques

2.8 Summary

This chapter discusses details issues related to the undergone research of biodiversity data management, biodiversity data models, spatial and temporal application, spatio-temporal data models and query languages. After analysis proposed models based on biodiversity data, models are designed either spatial or temporal. But most of the models are designed for commercial applications.

Problem with previous data models is that spatial and temporal aspects of databases are modeled separately (Sellis, T. (1999), (Martin Erwig, 1999). Spatial database focuses on supporting geometries (Guting, R. H., 1994), while temporal databases focus on the past state (Tansel, A. U. and J. Clifford, 1993). But in many circumstances, such as environmental monitoring, resource management, transportation scheduling, etc, spatial and temporal attributes should be connected together. Many current systems can handle only one aspect of space and time. Spatial systems always fail to cater for many temporal aspects in a dynamic environment (Abraham, T. and J. F.

Roddick, 1999). Though many researchers have found the necessity of integration of space and time in one environment, by far, little such work has been done.

The structures of space and time are identified as essential for the realization of cognitive systems (Freksa, C., 1998). According to Donna J. Peuquet and her group (Peuquet, D. J., 2001), models of spatial temporal data in geographical database representations must incorporate human cognitive principles. Human knowledge of the dynamic geographical world comprises of three different (and interrelated) subsystems that handle *what*, *where* and *when* aspects of object properties (Mennis, Peuquet et al. 2000), (Sinto 1978). Theme-based model, location-based model and time-based model separately describe one subsystem. From this view, these 3 kinds of model are all single semantic models. According to human cognitive principle, the model we build should cover all of these 3 subsystems. Donna J. Peuquet and her group's pyramid framework show how to incorporate three subsystems of human cognition into data modeling. From above analysis, we should build a unified infrastructure that integrate space and time, mediate discrete and continuous representation, describe when/where/what systematically, and be extensible to various applications. A multi-semantic model cannot ultimately be generated from extension of current single semantic models such as ER model and location/time/theme-based model. Object relational approach with its characteristic of inheritance and aggregation is capable of capturing the various notions of space and time and reflecting them into a single framework extensible to different applications.

Many database systems concentrate on the definition of a particular spatial temporal model that is related to certain application. The result is that more and more different models appeared. Each model focuses on a specific set of spatial temporal features [9]. When encountering other features and applications, the model doesn't work. So what we should do is to build an overall framework that can be extended to various applications, not driven by various applications generating different models (Mennis, J. L., D. J. Peuquet, et al., 2000). Furthermore, application specific modeling will be more efficient if it is based on a generic model (Raza, A. and W. Kainz, 1999).

In particular, we propose a conceptual object-relational spatial temporal data model based on Donna J. Peuquet's pyramid framework. However, even in the latest generation of OR (Object Relational) systems, extensibility comes with many limitations and requires significant expertise and programming effort. Indeed, the main extensibility, mechanism of OR (Object Relational) systems is allow SQL queries to call external functions coded in a procedural language (such as C or Java). Spatio-temporal queries are expressed by user defined aggregates, such as duration and during temporal relationships. Therefore, query language (such as SQL^{ST}) or other relational languages are required to support spatio-temporal queries for biodiversity data. In addition, spatio-temporal primitives and SQL^{ST} allows user to introduce additional extensions to the data model and query language.

Research Methodology

3.0 Introduction

This research framework contains with several phases, which have been identified upon completing this research. Framework also divided into two major phases. In the first phase, consist of literature review on biodiversity data model designing, problem definition can be done after analysis and study behavior of biodiversity data. Data behavior is divided into data pattern, data element and elements which are effective to the data. In the second phase, data model development will be done according to problem statement stated in the first chapter.

3.1 Operational Framework

The purpose of this operational framework is to develop a data model for biodiversity include temporal data. The operational framework has been divided into two main phases. The following subsections describe the methods, steps and procedure for proposed system. The detail steps involve in the system development will be discussed in the next section.

3.1.1 Problem Formulation (Phase 1)

In this phase, main contribution has been done such as investigation of biodiversity data, behaviors of data pattern, data elements and its effectiveness. Beside the study about biodiversity data, study about spatio-temporal database and architecture of temporal data model.

At this phase, literature review has been done two separate ways. Firstly, researchers need to study about the Biodiversity especially plant data behavior, Biodiversity is the assortment of different types of organisms that co-occur in time and space (spatial and temporal). Time element is very important to analysis in Biodiversity data such as predicting and historical analysis. On the other hand, another issue need to study is complexity of spatial and temporal data. Secondly, researchers need to study about the existed data model if there is any. In this phase, data collects from Pahang forest department, Mardi, Johor Park according to our research requirements.

3.1.2 Data Model Development (Phase 2)

In order to develop a data model a systematic approach needed. This is where we need to determine which data model development methodology to use. In this phase, firstly we have to analyze the existed model. In this part, researchers need to study user requirements in Biodiversity Applications. The investigation process from user requirement will focus on the Biodiversity systems in the market and specifically in the Biodiversity applications. The purposed of the study, is to cater requirement need in the data storage and retrieval and biodiversity analysis requirements. After completion of the investigation and study review on biodiversity, researchers have come out with a resolution to design a new data model to achieve the objective, design a data model for Biodiversity. This phase also identify data elements and data components of the model. After identified all the components a new data model has been developed.

3.2 Method Used

3.2.1 Pyramid Framework and Event-based Approach

In the design of object-relational biodiversity data model to support spatio-temporal data, pyramid frame can make enhancement to support a biodiversity data model design. A conceptual framework (also known as pyramid framework) was designed which guide the implementation of the semantic geographical information system (GIS) data model (M. Bouzeghoub, 1994). Conceptual frame work is interrelated with two separate parts one is data component and another is knowledge component (Figure 2). Data component can be divided into three categories: location (position in the spatial three dimensions), time (position alone a time line) and theme (a set of observations, measurements, or attribute values associated with a particular location and time). Data components stores spatio-temporally referenced observational data such as *spectral, climate*, *vegetation* and *other environmental attributes* that maybe queried and visualized to reveal embedded spatio-temporal pattern and relationships. Data component can be computerized as a multi-dimensional, spatio-temporal referenced 'hyper-cube' of observational data that is similar to the 'feature space' concept commonly cited in analysis of remote sensing imagery.

Knowledge component stores information about higher-level semantic 'objects' the geographic entities or process that are describe by the data. Information concerning on object's location, time and composition. All the objects are also placed within two hierarchical relationship structures central to cognitive knowledge representational and object-oriented modeling, plant taxonomy (generalization) and partonomy (aggregation). The taxonomy structures groups similar objects within a category and stores a rule-base that describes how those objects may be identified within the data space. These rules may be derived from expert knowledge or from inductive analysis of the observation data. In an object-relational data model consists of a set of *object classes* (of different *types* or *schemas*). Each object class has an associated set of *objects*; each object has a number of *attributes* with values drawn from certain *domains* or *atomic data types*. Of course, there are additional features, such as object valued attributes (Oid), methods, object class hierarchies, etc. Besides objects, attributes describing *geometries* including *tim*e are of particular interest. Hence we would like to define collections of *abstract data types*, or in fact *many-sorted algebras* containing several related types and their operations, for spatial values changing over time. This section presents a simple and expressive system of abstract data types, comprising data types and encapsulating operations, which may be integrated into a query language, to yield a powerful language for querying spatio-temporal data.

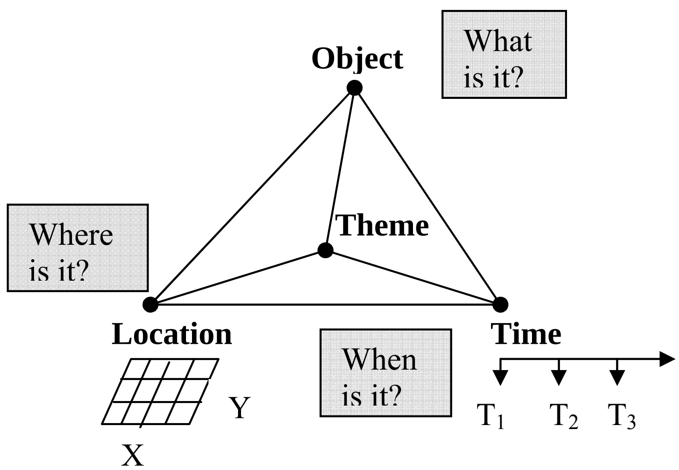


Figure 2 Pyramid Frame work with three main components; object,location, theme and time

To support temporal data, system require additional model such as event-based model. Pyramid system is a model to support multi-dimensional, spatio-temporal and geographic objects (such as location and space). All the objects are also placed within two hierarchical relationship structures central to cognitive knowledge representational and object-oriented modeling, plant taxonomy (generalization). The plant taxonomy structures groups of similar objects within a category and stores a rule-base that describes how those objects maybe identified within the data space. Pyramid framework consists of main three components; objects, location, theme and time. These features will be referred to accessed data from the database. This framework can be converted to table in stated Figure 3.

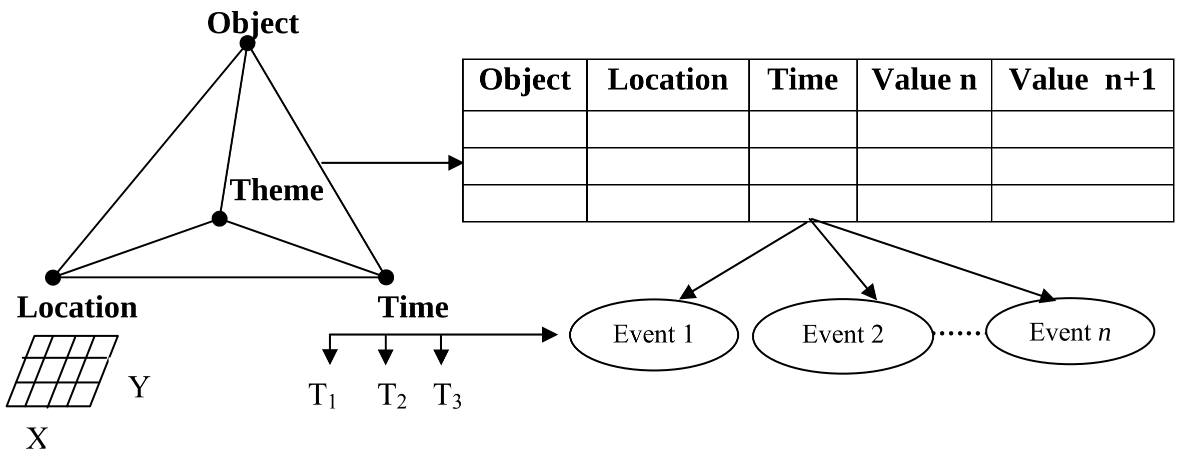


Figure 3: From the frame work data transform to the table and time modified as events

After the framework conversion as a table, we can see that each object has attributes of theme, space (location) and time. Theme attributes record the information of “what is this object” and other related property information. For example, if the theme of this object is “Forest”, a Forest must have attributes such as Area or Province it belongs to, Forest Size, and Forest Type, etc. Theme can be inherited according to certain classification hierarchy. For example, theme can be classified as natural resources (forest, earth quake, and rainfall), transportation (rail and highway), settlement (country, city), water system (river, lake, and pool), and cadastre etc.

Location attributes record space information of the object, such as whether it is a point, line, or polygon, and its location. Location attributes can be inherited into subclasses such as the point class, line class and polygon class.

3.2.2 Spatio-Temporal Object-Relational Model

Commercial GIS and Biodiversity related software support spatial object data, some times temporal objects data models based on RDBMS/OODBMS environments. However, temporal data always exists in and around daily life, in business, GIS and forest areas. Since spatio-temporal object-relational model might be new approach to biodiversity area, there are few commercial packages within the full capabilities of temporal object systems. Although there is growing demand for temporal data to be incorporated into biodiversity data, system administrators and system managers experience difficulty n responding to temporal system request because present reliability of working circumstances for temporal applications put time into secondary status.

Biodiversity data is associated with Area, forest, park, marine, flora, fauna, etc. biodiversity data can be interpreted as a digital map/picture that is equivalent to spatial data. This chapter not only examines the objects of flora playing a major role in temporal factors,

This chapter focuses fundamental aspects of the spatio-temporal domain, covering concepts such as objects, properties, relationship, attribute, method, generalization, specialization, association, and aggregation**.** Based on these, in the object-relational approach, present a small set of constructs aimed at improving the ability to conveniently design a spatio-temporal object relation model for biodiversity. Incorporate the proposed modeling constructs into the object-relation model, resulting in the semantically richer Spatio-temporal Object Relational Plant Biodiversity (STOP Biodi) Data Model.

3.3 Object Relational Model

Object relational databases provide a higher abstraction of spatial data by incorporating concepts closer to humans' perception of space. This is accomplished by incorporating the object oriented concept of a user-defined abstract data type (ADT) and associated functions. When using the Object-relational Model (ORM) for a particular system analysis, we introduce the ORM to explain generic characteristics of objects. This model is based upon a bottom up approach. We first present objects and classes and then show how groups of objects and relationships constitute object classes and relationship sets. The ORM consists of objects, object classes, relationships relationship sets, and constraints, only describing common terms of notion of object-orientation.

Object features such as abstract data type and Object relational Database Management System (ORDBMS) is a promising approach to extend the matured feature of Relational Database (RDB). Vendors such as Oracle, Informix, and Microsoft SQL Server have extended their RDBMS to support objects features (R.S. Devarakonda, 2001). Basically, the ORDBMS is the RDBMS with the object-oriented fatures. ORDBMS becomes popular because of the limitations of ODBMSs, such as lack of support for new types and lack of support for composite data values that can prevent it from taking enterprise-wide task. Therefore, by storing objects in the object side of the ORDBMS but keeping the simpler data in the relational side, user may approach the best of both worlds.

3.4 Modeling BIODI with Classical Characteristics

3.4.1 Object Characteristics

From the viewpoint of software development, the first programs were written in machine code, becoming dependent on the architecture of the computers on which they were implemented. Programming language techniques have naturally evolved towards a more clear-cut division between the concepts manipulated in programs and their internal representation.

There are many definitions of an object. Booch (1991) discusses the state, behavior, and identity of an object: the structure and behavior of similar objects defined in their common class and the interchangeable terms of instance and object. This is a classical language definition, as defined in (Coplien, 1992), where classes play a central role in the object model. Booch also insisted that the term "object" was first formally applied in the Simula language, and objects typically existed in Simula programs to simulate some aspect of reality. Martin (1992) defines an object as anything to which a concept applies and a concept is a shared idea or notion that applies to certain objects within normal awareness. Rumbaugh (1991) notes that an object is defined as a concept, abstraction or thing with crisp boundaries and meaning for the problem at hand. Shlaer and Mellor (1988) define an object as an abstraction of a set of real world things such that: all of the real-world things in the set of the instances have the same characteristics. All instances are subject to and conform to the same rules.

In the face of the increasing complexity of the problems tackled, it was apparent that software systems needed to be better structured. A program is considered to be a set of procedures together with a separate set of data on which the procedures operate. Related design methods consist of "divide and rule". The splitting up of tasks to be carried out into a set of independent modules is known as black boxes. This is called procedure-oriented programming. However, the slightest change in the data structures can lead to a complete reorganization of these procedures. The idea of information hiding (sometimes called encapsulation) obviates this difficulty because the data and the procedures manipulating it are incorporated into the same entity, namely the object. An object includes a static part, a set of data, and a dynamic part, a set of procedures manipulating this data. An object is defined by its behavior, and represented by the set of its procedures, rather than by its structure.

An object has a set of attributes (Instance variables), and relationships with other objects. An attribute models the state of an object. Procedures define its behavior. Procedures are considered to be messages. We say that procedures encapsulate the behavior of the object. They define the interface to the object, and can be used to maintain the hidden internal structures and states of the object. Consequently, procedures may define the only available way to communicate with an object and access its attributes. Procedures are sometimes referred to as functions or methods.

3.4.2 Object- relation Data Model Techniques

The object relational model (ORM) is a way of describing or representing objects, classes of objects, relationships between objects and classes, and real world memberships. The ORM can be considered the static part of object-oriented systems modeling. That is, the ORM describes the “database” of a model: what objects may exists, what objects classes they belong to and what relationships exists between objects. In the object-relational model also consists of object, object classes, relationships relationship sets and constraints.

3.4.2.1 Objects Types:

An object includes a static part, a set of data, and a dynamic part, a set of procedures manipulating this data. An object is defined by its behavior, and represented by the set of its procedures, rather than by its structure. An object has a unique and immutable identified that differentiates it from other objects. An object identity (OID) comes into being when the object is created and can never be confused with that on an other object even if the original object has been deleted. The OID never changes even if all the properties of the object change. It is depend on the object’s state. An object identifier further has the following characteristics: 1) it identifies an object not only in an extension of a class, but also in database, 2) it is constant throughout the lifetime of the object, 3) it is not visible to the user, and 4) it is created and maintained by the system.

An object has a set of attributes (Instance variables), and relationship with other objects. An attribute models the state of an object. Spatial and/or temporal information may be associated to objects, independently from the characteristics of their attributes. Consequently, an object type can be plain (neither spatial nor temporal), spatial, temporal, or spatio-temporal. The term object, however, is fundamentally different from the term entity. Entity is concerned merely with data. We typically store a record from each entity. Object concerned with data and the methods with which it is manipulated. For example, in biodiversity data model different object represents the different real world entities as shown in Figure 4.

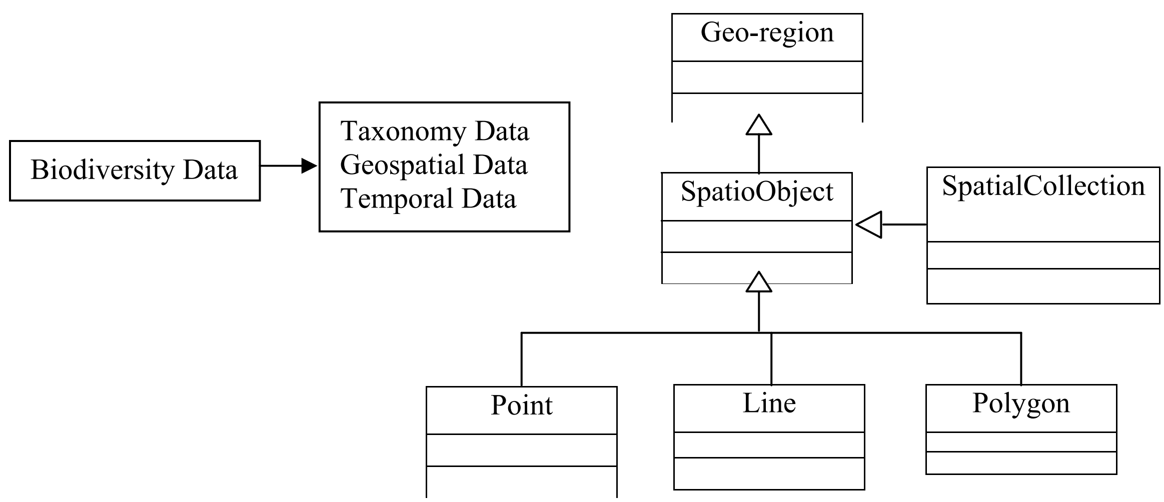


Figure 4: Object type and spatial data as object

There are two types of object diagrams: class diagrams and instance diagrams. A class diagram is a schema, pattern, or template for describing many possible instances of data. A class diagram describes object classes. An instance diagram describes how a particular set of objects relates to each other.

3.4.2.2 Abstraction mechanism

The object-relational data model is built on few basic concept of abstraction: relationship, generalization, association and aggregation. These abstraction concepts may be described by the means of relationship sets. A relationship set is a set of relationships, all of which have the same role to between objects. In an object-relational diagram, a relationship set is represented by lines connecting associated objects. As for objects, relationship may be located in space and time, via two attributes. In this case they are referred to as spatial and/or temporal relationships. Figure 5 shows identify relationship between two objects.

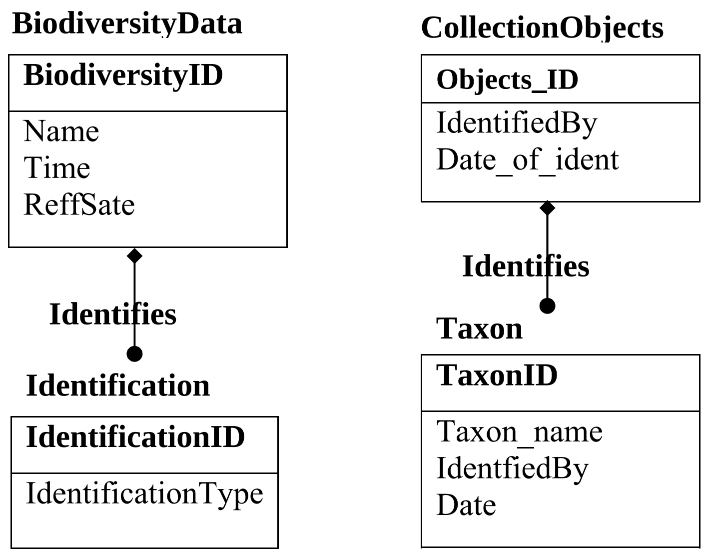


Figure 5: “Identifies” Relationship

3.4.2.3 Generalization and specialization

Generalization and specialization represents the “Is-a” relationship set, an essential element of object-relational paradigm. The main idea in generalization and specialization is that one object class (specialization) is a subset of another (the generalization). The process of generalization consists of putting classes together in what is called a superclass. Generalization enables us to perceive that all instances of more specific concept are also instance of general concept. Specialization, on the other hand, breaks apart a class by differentiation of properties or forms a set of lower class called subclass. Specialization allows us to describe each subclass of a more general class by specifying only the additional details necessary for us define. Subclass describes a specialization of the superclass.

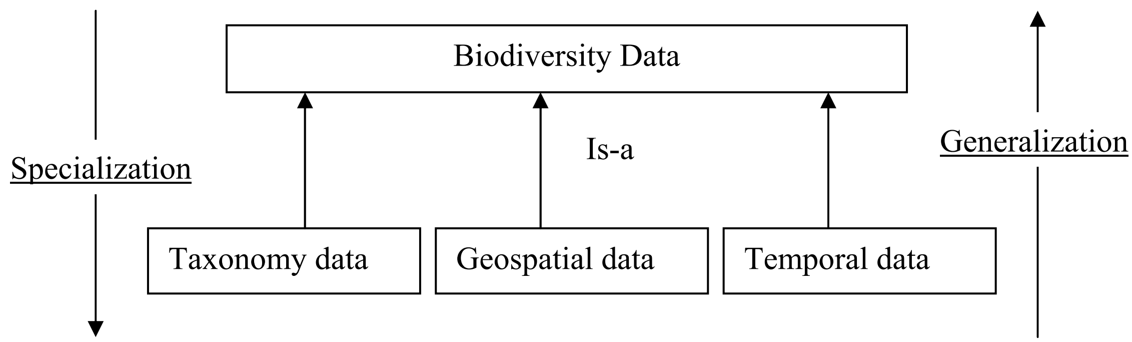


Figure 6: Generalization and Specialization

The terms superclass and subclass characterize generalization and specialization. It is important to note that superclass a subclass are abstraction of the same object, and do not describe two different objects. The direction of the "Is-a" relationship goes from the specialization to the generalization (Figure 6), that is, it may be stated as, "Specialization Class is-a Generalization Class".

3.4.2.4 Association

An association describes a group of links with common structure and common semantics. A link is a physical or conceptual connection between object classes. All the links in an association connect objects from the same classes. Association and links often appear as verbs in a problem. An association describes a set of potential links in the same way that a class depicts a set of potential objects. In fact, the notion of an association is certainly not a new concept. Associations are widely used throughout the database modeling community for years.

In an object-relational model diagram, association is also known as the “member of” relationship set (Figure 7) For example; clubs and teams are usually modeled using association. Accordingly, an object in the Taxonomy data object class is a set of objects from the PlantSpecies, IdentCharacteristic, IdentLevel. The relationship set is read “PlantSpecies is a member of Taxonomy Data”.

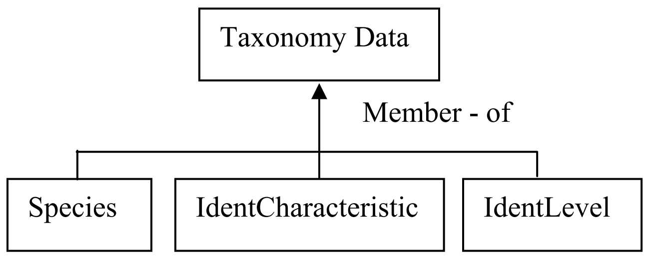


Figure 7: Association

3.4.2.5 Aggregation

An aggregation represent the “a –part-of” or “is subpart of” relationship set, in which objects representing the components of something are associated with an object representing the entire assembly. Aggregation is a strong from of association in which an aggregate object is made of components. Components are part of the aggregate. The aggregate is semantically an extended object that is treated as a unit in many operations, although physically it is made of several lesser objects. A single aggregate object may have several parts: each part-whole relationship is treated as a separate aggregation in order to emphasize its similarity to association.

However, there are may be some discussions of the differences between association and aggregation. Aggregation is a spatial form of association, not an independent concept. Aggregation adds semantic connotations in certain cases. If two objects are tightly bound by a part-whole relationship, it is an aggregation.

If the two objects are usually considered as independent, even though they are often linked, it is an association. But, the decision to use aggregation is often arbitrary a being matter of judgment. Often it is not obvious if an association should be modeled as an aggregation (James E. etl., 1991)

Most of the objects dealt with in biodiversity databases are composite, i.e., they exhibit an aggregation hierarchy. They follows are some of the aggregations in the biodiversity data models:

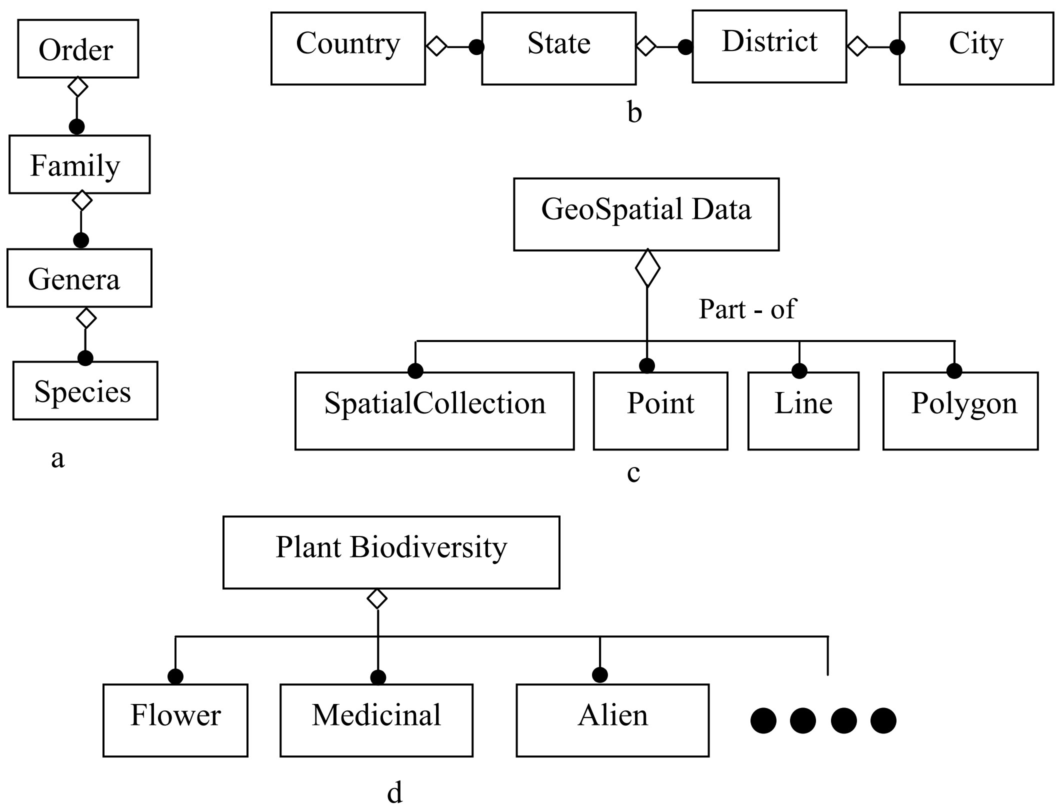


Figure 8: Aggregations in Stop-BDM data model

An order contains several families. A family in turn contains several genus and genus contains closely related species. Since a family cannot belong to two order, the order-family relationship is modeled as an aggregation rather han an association. Similar arguments hold for genera and species also. Hence, taxonomy hierarchy is modeled as a four level containment hierarchy in Stop-BDM (figure 8a)

A country contains number of states, each states is made up of districts and districts made up of cities. The relationship is also modeled as a four level containment hierarchy. Similarly, the association between a city and the forest department in the city and the association between a country and a forest are modeled as aggregations (figure 8b).

Figure 8c is an example of an aggregation. Geospatial object in as aggregation of the SpatialCollection, point, line and polygon that are the part of it.

3.4.2.6 Constraint

Most of the system can be model with objects, objects classes, and relationship sets. Constraints are functional relationships between entities of an object relationship model. Constraints provide one certain for measuring the quality of an object relational model. In object relational model, a constraint restricts the membership of one or more object classes and relationship sets. The object relational model allows analysis to express several different types of constraints. However, in the object behavior model and interaction model, it may be useful to use this constraint in order to describe some restricted condition.

3.5 Modeling Object Behavior of STOP BIODI Model

The object-relational diagram is a powerful tool for representing objects and relationships. However, it does not allow us to model the behavior of objects. We need a behavior model to document the way objects in a system function, respond or perform. We use the Object Behavior Model (OBM) to describe the behavior of the objects in a system. The OBM is a means of describing the behavior of objects. As we know that a database consists of a set of *object classes* (of different *types* or *schemas*). Each object class has an associated set of *objects*; each object has a number of *attributes* with values drawn from certain *domains* or *atomic data types.*

In this chapter, thesis describes three steps of object processes pertaining to evolution and relationships of objects, and structural changes of objects. Basic processes of spatio-temporal objects are essential to portray physical phenomena, arising from the fact that object behaviors are unpredictable. These processes prove that objects move and interact with other objects in a specific place at a given time. Object attributes are continuously changed when space interacts with time. In order to depict object process in space and time, we examine five spatio-temporal data models, ranging from a cube model to an object model.

In addition, three kinds of data domain are reviewed to understand current temporal GIS developments and technologies because there is no agreement on a spatio-temporal framework, based upon the three-legged approach to location (where), feature type (what), and event time (when). An event can be describes in two ways one in the field and another within the system. Event change in reality when transaction goes in the field by some one and event change within the system which includes the creation or deletation of object or relationships, the starting or stopping of an activity and the reception of a message from other objects.

3.6 Spatio-Temporal Abstract Data Types (ADT)

One of the basic ideas behind spatio-temporal data models is that, in addition to the normal built-in data type defined as “date”, user-defined types may also be employed.

These types may be used in the same way as built-in types. For example, columns in relational tables may be defined as taking values of user-defined types, as well as built-in types. A user-defined Abstract Data Type (ADT) definition encapsulates attributes and operations in a single entity. An abstract data type (ADT) is defined by specifying a set of declarations of the stored attributes that represent the value of the ADT, the operations that define the equality and ordering relationships of the ADT, and the operations that define the behavior (and any virtual attributes) of the ADT. ADTs can also be defined as subtypes of other ADTs. A subtype inherits the structure and behavior of its supertypes (multiple inheritance is supported). Instances of ADTs can be persistently stored in the database by storing them in columns of tables. Points, line, and polygons in the Georelational data model are examples of spatial ADT stored in table columns. Spatio-temporal ADT is a research topic that attracts many researchers, and is represented by the Event-based Spatio-temporal Data Model, the STComposite data model, and the Worboys data model.

3.6.1 Data Type "Date"

One easy way of developing spatio-temporal data types within spatial data models is to use data type "date." Applications adopting the data type "date" are mostly built on relational or object-relational DBMSs. The data type "date" specifies user-defined time attributes and timestamp attributes in spatial data model schemas, which means that timestamps are defined either as attributes of a relational table or time instants assigned as a timestamp of a relational table (Armstrong, 1988; Clifford, 1983; Hunter and Williamson, 1990; Lum et al., 1984; Snodgrass and Ahn, 1985; Stickler et al., 1992). This method is simple, but it has problems in terms of its data maintenance and efficiency. The lack of direct query support for spatio-temporal data leads to complex query operations, which in turn results in the difficulty of maintaining the data model. It is also unable to solve data redundancy from the snapshot spatial data model, which manages time-varying spatial data with separate GIS layers.

3.7 Temporal Object System

3.7.1 Object in Time

When describing the modeling of time in temporal database, there have been many efforts to model temporal data. As depicted above, time is a linear and discrete collection of time units, of which the smallest one is considered as chronons. An interval is represented by two chronons (e.g., moment, instant) indicating the beginning and end of the interval. In view of temporal issues, transaction time can be one of multiple past states of an object. Chakravarthy and Kim (1993) argue that there nay be multiple interpretations for the temporal behavior (validity) of class of data objects. Using the idea of temporal validity, they define the notion of history as being a representation of the evolution of an object along a one-dimension time. The significance of temporal validity is associated with each individual property, that an object has its own characteristics as temporal attributes (or data values) related to the temporal behaviors of data objects.

3.8 Spatio-temporal Object

Time provides a class for the general concept of a temporal object, including all points, volumes, and undivided wholes, etc with temporal dimensions. Time points, which are non-decomposable in the style of zero dimensional points in time, and a single explicit subtype, are defined here as zero dimensional time (zero-D-time). Zero dimensional points are usually referred to using a preposition such as "**at**". Time as a general undecomposable substance, is the combination of *temporality* and *substance* (Bateman et al., 1995), used as an analogous spatial concept that might support, for example, selection of the interrogative form "**when**". Time-interval is a one, two, or three dimensional, set of time points. It is also an ordered-object. Two subclasses (or subtypes) are differentiated by linguistic grammar as one or two-D-time and three-D-time. A three-

D-time is a portion of time that is being viewed on a sufficiently large scale as if it were a volume within which a thing occurred (e.g, in “1996”) rather than a plane on which a thing occurred (e.g, “on that day”). A one-or-two-D-time is a temporal object that is associated with time intervals or smaller scale successions of time instants, e.g, a day (as opposed to a year).

The spatio-temporal sub-hierarchy is organized along a number of dimensions, which combine to form specific categories of relationships between entities and locations in space and time. The Upper Model (UM) (Bateman et al., 1994) delineates that most spatio-temporal subclasses are responsible for the appearance of specific prepositions within prepositional phrases generated by the grammar to express spatio-temporal relationships. According to the UM, spatio-temporal linguistics includes relations associated with time, space, directions, and extents. In principle, there are two level distinctions between locating and extents. In English, the extent-dimension is responsible for the selection of prepositional phrases involving prepositions such as *for, along during*, as in the prepositional phrases: “for six days”, “for three miles”, “across the bridge”, “along the street”, “during the seminar”, etc.

3.9 Object-relational Features

As we know that within the classic relational database model there are only scalar but no complex data types. With the introduction of object types the definition and composition of abstract data types is possible. An abstract data type can be comprised of a multitude of scalar types and again of user-defined complex types and enhances consistency when creating database models.

```
create or replace type Address_TY as object (
	City VARCHAR2 (30),
	Sate VARCHAR2 (30),
	Area_code NUMBER (8),
```

```
        Country VARCHAR2 (30));
```

Here create or replace the Address_TY data type after that create Org_Address_TY by using Address_TY data type.

```
create or replace type Org_Address_TY as object (
        Org_name VARCHAR2 (30),
        Address  Address_TY,
        Org_url VARCHAR2 (50),
        Phone NUMBER (12),
        Email VARCHAR2 (30));
```

Data types only represent descriptions of data structures. To ensure persistency a table must be bound to the data type. A relation with column objects can be created that is a set of Observer_ID with Observer_name, Observer_org and Observer_Address as:

```
        create or replace type Observer as object (
                Observer_ID VARCHAR2 (15),
                Observer_name VARCHAR2 (30),
                Observer_org VARCHAR2 (50),
                Observer_Address Org_Address_TY );
```

Object views allow user to treat relationally data as they allow synthesizing object from data that continues to be stored in relational tables. Object views have similar functionality like object tables. They can have methods, belongs to collections, reference one another, have object identity and can be accessed from SQL. In addition tables are assigned to object views can be uploaded by using special instead of triggers. Any instance (row) of an object class contains a unique ID called Object_ID (ID). Generally OIDs are system-generated but can also be derived from a primary key column or can be user-defined.

Relationships between objects can be defined using reference types. A reference column stores OIDs of associated (row) objects since column objects do not have inherent OIDs and therefore cannot be referenced. Row objects that belong to a reference can be selected and dereferenced using the DEREF VALUE operator. Modeling object relationships with OIDs and REFs is often compared with foreign key relationships inside the relational model but implicates some benefits like the ability to distinguish between equal and identical objects. Objects are identical if they have one common OID. They are equal if they have different OIDs but coincide with their attributes and values.

In the classical Entity-Relationship-Model aggregations and compositions are modelled through master-detail-relationships. Object-relational dbms provide collection types that contain multiple elements and thus are suitable to express 1: N relationships directly. Each element or value for a collection has the same substitutable data type. The most popular collection types are varrays and nested tables.

A varying array contains a variable number of ordered elements. Varying array data types can be used as a column of a table or as an attribute of an abstract type. Named table types can be created in an Oracle database using SQL. These table types can be used as nested tables to provide the semantics of an unordered collection. As with varray a nested table type can be used as a column of a table or as an attribute of an object type.

Named table types can be created in an Oracle database using SQL. These table types can be used as nested tables to provide the semantics of an unordered collection. As with varray a nested table type can be used as a column of a table or as an attribute of an object type.

Create type Address_TY as table of Org_Address_TY;

Multi-Level-Collections that lead to multiple nested tables can be realized if useful for applications but it’s up to the user to balance – a more intuitive representation of data vs. higher complexity of accessing the data.

An object type declaration can also include methods that are defined on values of that type. When using these objects types in tables their methods are also applied to the data of these tables. The method is declared in the *create type* command and the code for the function itself (the definition of the method) is in a separate *create type body* command.

```
create or replace type Taxon_TY as object (
Taxon_Lname VARCHAR2 (20),
Tanon_Ename VARCHAR2 (20),
Taxon_Species  Species_TY,
Taxon_family Family_TY,
Taxon_area Area_TY,
Taxon_description VARCHAR2 (50),
Taxon_start_date DATE,
Taxon_end_date DATE,
Member function DURATION (start_date, end_date in Date) return Number);
```

Names the function that is a "member" of the Taxon_TY datatype. Since Duration will return a value, it is a fuction. To define DURATION function, use the create body command, whose full syntax is shown in the Alphabetic Reference.

Let's create the Duration function as a member function within the Taxon_TY datatype. The DURATION function will return duration, in days, of the taxon.

```
create or replace type body Taxon_TY as
member function DURATION (start_date, end_date DATE)
return NUMBER is
begin
return totaldays (start_date + end_date );
end;
```

The following query executes the duration method in the data type Taxon_TY and returns the duration of all collection objects in relation Observation. Values of components of an object (attributes and methods) are accessed with the dot notation.

```
Select Taxon_ID, O.Taxon_Area, Duration (O.start_date, O.end_date)
```

from Observation O ;

Object types can be mapped to the prevalent OO languages like C++ and Java. Thus object type instances in the database can be accessed and modified to and from C++, Java applications (Lee, G., 2003).

In the current version Oracle database 10g native data types for storing raster and persistent topology data were added to further extend possibilities to model real world objects based on object-relational database structures. The completeness and therefore the quality of spatial features' descriptions can be further enhanced with the introduction of a new system-defined data type XMLType that can be used as the datatype for columns in tables and views to create, extract, and index XML data. A feature's position and spatial representation is then modeled by using vector resp. raster types, its characteristics are captured by (complex) attributes and any further information on the spatial object, that cannot be structured to be kept in the data types mentioned, is maintained in xml-documents, all stored natively within the database.

For modeling highly dynamic changes of real world objects the SQL type *TIMESTAMP* can be used which is a high-precision *time and date type* that allows storing fractions of a second (Qian, L., 2004). The timestamp values are basically points on a linear time axis. For slower changes in state we can use *DATE* or create own adapted abstract data types that can express discrete dates or times intervals. The data types can be flexibly extended to meet an application's accuracy requirements and provide attributes to express the fuzziness of a time record, for instance if a time interval cannot be determined as accurate as the data type could store (Plabst. S., 2001).

```
create or replace type TIME_TY as object (
    year NUMBER (4,0),
    month NUMBER (2,0),
    day NUMBER (2,0),
    hour NUMBER (2,0));
create or replace type TIME as object (
```

```
start_date TIME_TY,
end_date TIME_TY,
start_date_per NUMBER (1,0),
end_date_per NUMBER (1,0));
```

3.10 Object-relational Application Development

In terms of Biodiversity data of Malaysia, object-relational database technology can be used to model different applications, e.g. the spatio-temporal process of flora or fauna based on its influencing factors or real time deer tracking.

For the development biodiversity data model some major influencing factors have been proved to be crucial. In the model we have judged like data should be natural to human. The structures of space and time are identified as essential for the realization of cognitive systems. Models of spatial temporal data in geographical database representations must incorporate human cognitive principles. Finally model support human knowledge of the dynamic geographical world comprises of three different (and interrelated) subsystems that handle *what*, *where* and *when* aspects of object properties.

```
create or replace type Position_TY as object (
Map VARCHAR2 (30),
Grid25 NUMBER (4),
Grid100 NUMBER (4).

Create or replace type SpatialData as object (
Forest_ID NUMBER (6),
Forest_name VARCHAR2 (30),
Forest_area AREA_TY,
Forest_position Position_TY);
```

Based on this type an object view is created and OIDs are assigned to the spatial datasets of the corresponding forest area:

```
create or replace view Forestdetails_OV of Position_TY
with object identifier (ForestID) as
select Forest_name, position,
from Forest_area
where AreaName = Johor;
```

The columns of the relational base tables are now accessible as row objects through their corresponding object views. Any detail tables of the underlying relational model could now be simulated by further object views with references that could point to the row objects of the particular upper object view, to allow accessing the data in either way.

Projection operations on moving objects map either to their spatial or to their temporal aspect. Assume that we are interested in the geometric locations where the data was changed at the year January 20, 1990. These can be obtained by:

```
SELECT Flora_ID, AreaName
FROM Flora, Area
WHERE (((Forest_area =#Johor# AND #01/20/1990#));
```

This operation computes the *spatial projection* of a spatio-temporal object for the Johor forest area. For an evolving region the trajectory operation returns an object of the spatial type *region* which results from projecting the union of the region values for the Date 20th January, 1990.

The following query inquires about the largest collection of flora areas.

```
SELECT Area (max (Extent)) FROM ForestArea
WHERE Type = "Flora"
```

The query demonstrates an example of a *spatio-temporal aggregation* operation max which is an extension of the well known aggregation operator in SQL of the same name. It is here applied to a collection of evolving regions contained in a relation column and computes a new evolving region. Internally, this operator is based on a binary function $\boldsymbol{MAX_{ST}}$ applied to two evolving regions $\boldsymbol{R_1}$ and $\boldsymbol{R_2}$ 'and yielding a new evolving region in the following way:

$$MAX_{ST}(R_1, R_2) := \{(t, r) \mid t \in time \wedge r \; MAX_{geo}\,(R_1(t), (R_2(t))\}$$

This definition uses a function $\boldsymbol{MAX_{geo}}$ which is applied to two regions R_1 and R_2 and which returns larger of both regions.

$$MAX_{geo}(R_1, R_2) = \begin{cases} r_1 & \text{If area } (r_1) > \text{area } (r_2) \\ r_2 & \text{Otherwise} \end{cases}$$

Altogether this means that for 2 evolving regions $R_{1,\ldots}$, R_n we first compute the evolving region R = MAX_{st} (R_1,….MAX_{st}(R_{n-1}, R_n)..). Afterwards, we apply the raise area of R, which computes the area of R at all times as a temporal real number.

3.11 Conceptual Biodiversity Data Model Structure

After explanation of several components of object relational model, this section shows some relationships among the components of object relational model for biodiversity.

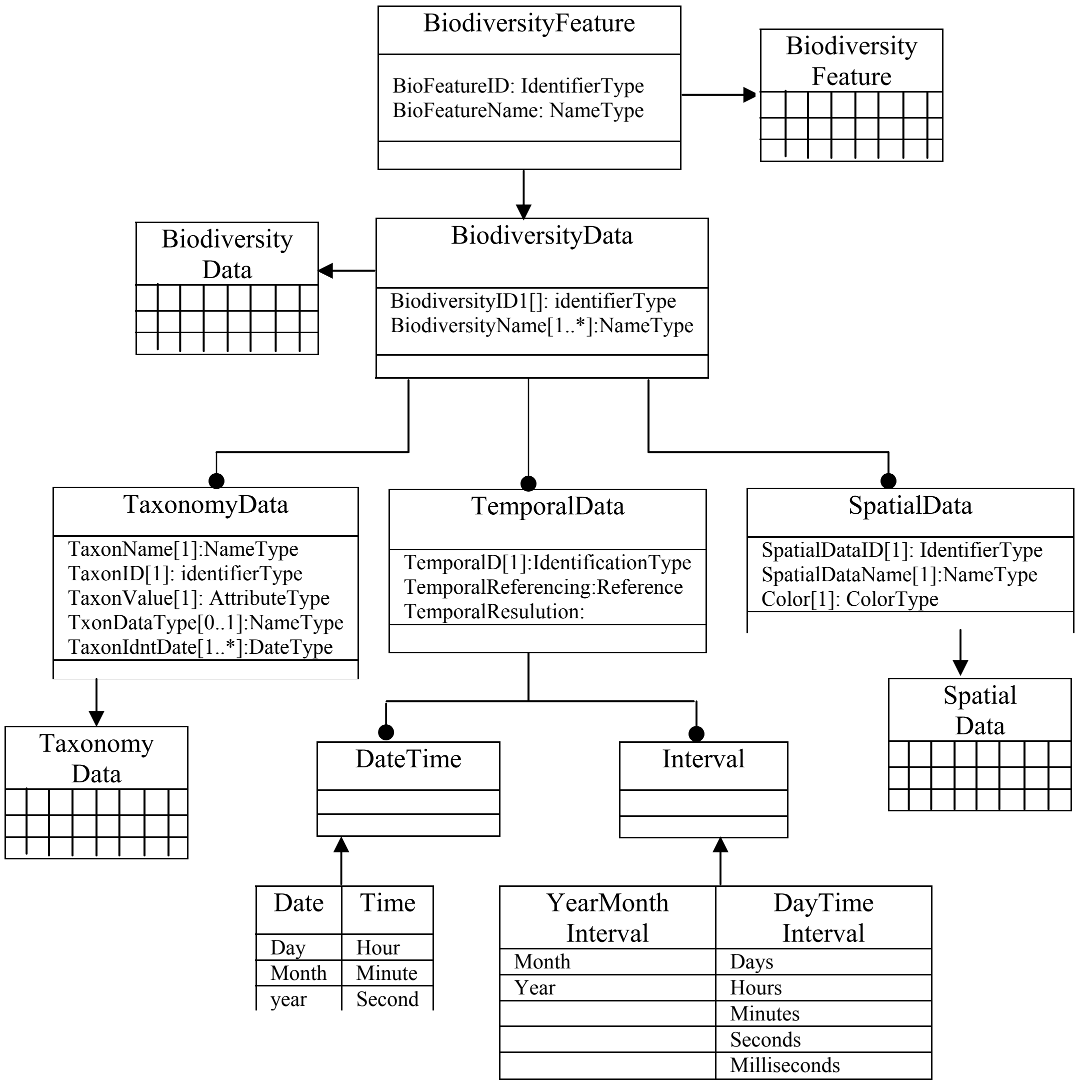


Figure 9: Detailed Structure of Spatio-temporal Object-Relational Model for Biodiversity

The proposed spatio-temporal data model is described by means of collection of one or more class diagrams that form the object model and connected with its related relational tables to form an object-relational data model.

The class diagram which describes the structural characteristics of the proposed spatio-temporal data model is presented in Figure 9. It shows that five major classes were identified and incorporated into the data model: BiodiversityFeature (BF), BiodiversityData (BD), Spatial Data (SD), Taxonomy Data, and Temporal Data (TD).

To conclude the description of the model diagram illustrated in Figure 4.9 and noted that each class has two common attributes: identifier and the name. But some classes are having more than two attributes especially in taxonomy data and temporal data objects. Each one of these attributes can have a single value and the identifier attribute is unique for every object of each class. The value of the identifierType and NameType domains will depend on the implementation of the object model. A time class carries a time value, an interval, a duration, or a collection of such values. Date and 24 hour time maybe used separately or together, in which case the object carrying both values is considered a timestamp. Two kinds of intervals are recognized as YearMonth intervals and DayTime intervals. YearMonth intervals and DayTime intervals can be viewed as objects types. In temporalData Object, we can specify the relationship among objects.

EXPERIMENTS AND RESULTS

4.0 Introduction

In this chapter, model testing is involved in the implementation of STOP-Biodi using an object-relational database management system. As a part of the testing, implementation and evaluation, a biodiversity data set covering more than 5 years data from different private and government organizations forest departments such mardi, pahang and johor national park's data loaded onto the database. This thesis describes the STOP-Biodi and its realization in an object-relational architecture, and outlines the benefits we perceived in experiments i.e. efficient retrieval, better support of space and time, complexity handling, data integrity, cost, data redundancy etc.

This chapter summarizes and discusses the performance test results of the research. A number of interesting issues are presented. The discussion of the proposed model is also presented.

The physical model is developed by defining spatial and temporal versions of datatypes. For spatial and temporal datatypes as combinations of spatio-temporal and standard datatypes, queries can be differentiated by which dimensions the desired results are restricted to, and whether they query for a range in a dimension or a point. Therefore experiments on different spatio-temporal queries were performed with several different selectivities.

4.1 Experimental Setup

In this research data model is implemented using Oracle 9i and Microsoft Active Server Page. The process of modeling the STOP-Biodi is done through defining different data types using SQL3 and ASP coding; the details of the software used in these experiments are listed in the Table 4.1.

Table 4.1: Specifications of Software's used in testing

Name	**Specification**
Operating System	Windows XP (Professional Edition.)
Database Management System 1	MS Access 2003, XP
Database Management System 2	Oracle 9i
Software Languages	MS ASP 3.0, Macromedia Dream weaver MX
Query Language	SQL, SQL3

For this experimentation we used the MS Windows XP Professional Edition operating system on which we installed the other supporting software's such as for testing the previous pahang, mardi and bodhi data model, we have used MS Access which is actually using the same system while for the implementing the STOP-Biodi we used Oracle 9i platform which is freely available from internet and also which can support spatial data and object-relational design of modeling. MS ASP3.0 and Macromedia Dream weaver MX are the software languages used for the Interface design of the testing system.

We have performed several performance tests for spatio-temporal object relational model using different set of queries including spatial, temporal and spatio-temporal queries. The model was tested on the Pentium 4 Processor system; the details of this system are listed in the following Table 4.2.

Table 4.2: Specifications of Test System

Name	**Specification**
Processor	Intel Pentium 4 1.7 GHz
System RAM	256 MB
Graphics Card	NVADIA GeForce4 MX 440

We have mentioned earlier that biodiversity data has been collected from pahang forest deopartment, mardi and johor national park. Pahang forest department developed database management system which is developed by Microsoft access.

4.2 Overview Structure of the Stop-BDM

In this section we describe the overview of structure of the STOP-BDM. This will help to understand the model at physical level in order to manipulate with the spatio-temporal data. With the description of the structure of STOP-BDM a relational schema is also prescribed, so as to briefly discuss the structural approach as shown in the Table 4.3 below

Table 4.3: Schema Structure comparison of models

Flora relation

Attribute	**Data type**
Flora_id (PK)	Varchar2
Local_name	Varchar2
Eng_name	Varchar2
Species	Varchar2
Family	Varchar2
Type_id	Varchar2
Description	Varchar2

Area	Region
File_name	Varchar2
Observer_id (FK)	Varchar2
Entry_date	Date
Update_date	Date
Entry_date_per	Varchar2
Update_date_per	Varchar2
Genus	Varchar2

Spatial Relation

Forest_name	Varchar2
Forest_id (PK)	Varchar2
Area	Region
Grid25	Number
Grid100	Number
State/region	Varchar2
Map	Varchar2
Area_code (FK)	Number
Forest_name	Varchar2

Here in the above mentioned Table 4.3 the time to collect the information from the forest area. In the STOP-BDM it is stored using the Abstract data type while in the pahang, mardi database model it used the traditional float data type. This relation is shown here for our current model. So executing the queries for the model are presented in the following sections.

4.3 Create standard query for biodiversity data

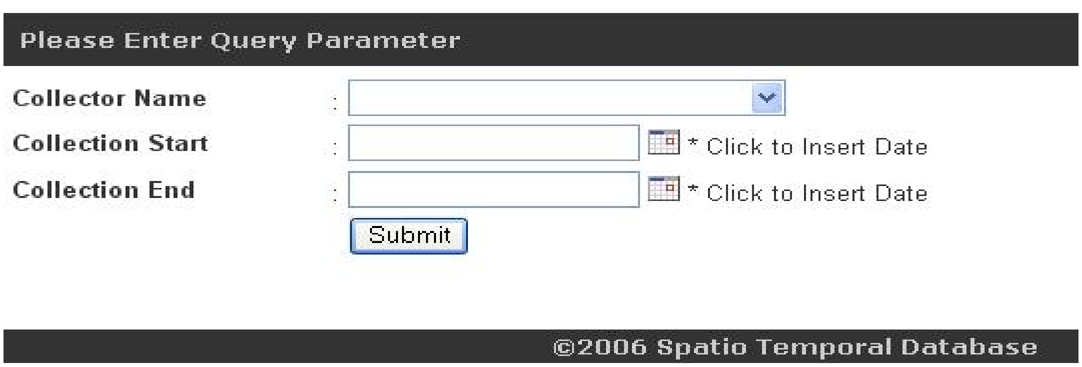
In this phase we have created standard query for biodiversity data. Some example of standard query stated bellow:

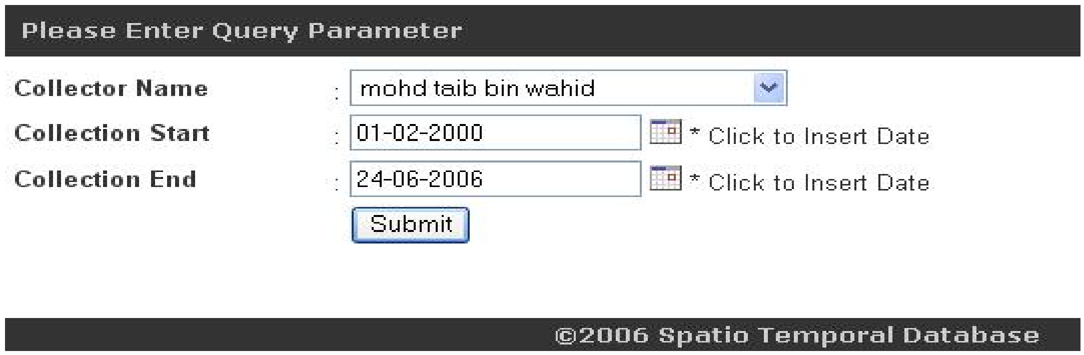
a) This query collects information about flora information and when data collected and updated

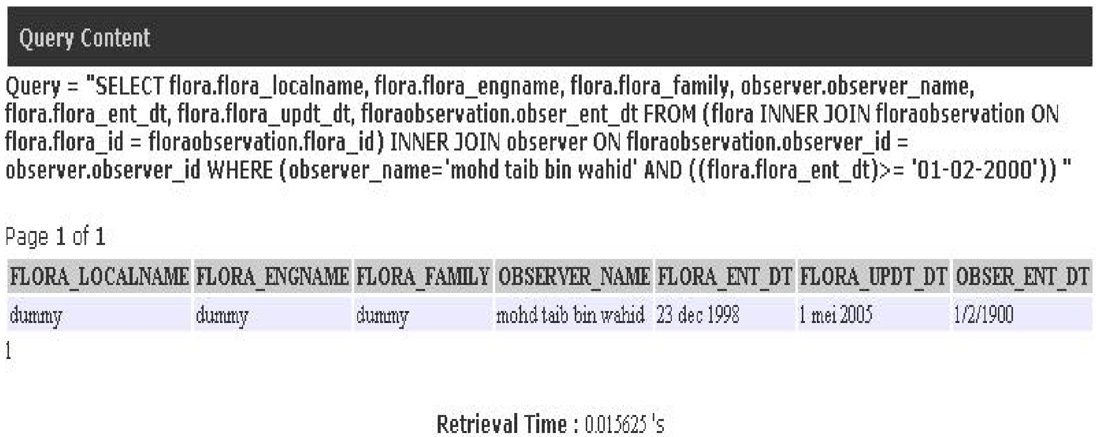
SELECT flora.flora_id, flora.flora_engname, flora.flora_family, flora.flora_species, flora.flora_entry_dt, flora.flora_typeid, flora.flora_upd_dt, floraObservation.flora_count, floraObservation.observer_id
FROM flora INNER JOIN floraObservation ON flora.flora_id = floraObservation.floraobs_id;

flora_id	flora_engname	flora_family	flora_species	flora_entry_dt	flora_typeid	flora_upd_dt	flora_count	observer_id
1	Hibiscuss	family test1	species test1	2/12/2004	type D herb	8/12/2004	2000	Zaman
2	Rose	family test2	speces test 2	2/12/2004	type id herb	8/12/2004	1500	Zaman
3	Bouganvilla	family kartas	species kartas	2/12/2004	Type kartas	8/12/2004	1550	Zaman
4	Bouganvilla	family test3	species kartas1	2/12/2003	species kartas1	8/10/2004	1000	Zaman
5	Bouganvilla	family test4	species kartas2	2/11/2003	species kartas2	8/9/2004	1000	Zaman
6	Rose	family test5	species kartas3	2/10/2003	species kartas3	8/9/2004	1200	Zaman
7	Rose	family test6	species kartas4	2/10/2002	species kartas4	2/10/2003	1000	Zaman

b) User also can find information about data collector by using web-based system. User also can see the retrieval time after retrieving data from the system.







Above queries examine the spatio-temporal data handling by our proposed model. The proposed model Stop-BDN supports well the spatio-temporal data to perform spatio-temporal queries.

In this investigation we used a Malaysian Plant Biodiversity data such as pahang forest data, mardi plane genebank data, and johor national park data. For implementation and testing of spatio-temporal object-relational plant biodiversity data model, we define a set of different relations which can contain and handle the biodiversity data.

4.4 Queries

This experiment is made for the measurement of the execution time of queries, which we have applied on the model. We used traditional queries to measure the time, using elapsed time. Elapsed time is the clock time it takes to run a piece of code. Basically, the elapsed time is the number of hours, minutes, second, and/or milliseconds it takes to run a chunk of code. Normally when we talk about how fast some code runs we are really talking about the overall elapsed time. Let's review a couple of different methods used to review the elapsed time of a query.

Prior to performing an elapsed time benchmark, the SQL queries were applied for benchmarking purposes. Moreover the queries were written using software language code on a stand-alone machine, which is where all the benchmark information was gathered.

Query 1:

```
SELECT flora.localname, flora.family, observer.name, flora.ent_date
FROM (flora INNER JOIN observation ON flora.id = observation.obser_id)
INNER JOIN observer ON observation.observer_id=observer.observer_id
WHERE (observer.observer_id='B0008' AND
(flora.ent_date BETWEEN '01-Jan-2006' AND '03-Jan-2006'));
```

```
LOCALNAME       FAMILY          NAME                      ENT_DATE
---------------  ---------------  ------------------------- ---------
brienianum      epidendroideae  Siti Aminah               01-JAN-06
pallida         epidendroideae  Siti Aminah               03-JAN-06
bicuspidatum    epidendroideae  Siti Aminah               02-JAN-06
```

Query 2:

```
SELECT f.id, f.localname, f.species, f.family, f.ent_date
FROM flora f INNER JOIN observation ON f.id = observation.obser_id
WHERE f.id>=004;
```

```
ID     LOCALNAME        SPECIES          FAMILY           ENT_DATE
------ ---------------  ---------------  ---------------  ---------
004    javanica         orchids          epidendroideae   02-JAN-06
```

008	pallida	orchids	epidendroideae	03-JAN-06
005	major	orchids	epidendroideae	02-JAN-06
006	parviflora	orchids	epidendroideae	02-JAN-06
007	bicuspidatum	orchids	epidendroideae	02-JAN-06
009	wallichi	orchids	epidendroideae	03-JAN-06
010	nuda	orchids	epidendroideae	03-JAN-06

Query 3:

```
SELECT forest_id, forest_name, grid25, grid100, state
FROM spatial
WHERE map !='perak';
```

FOREST_ID	FOREST_NAME	GRID25	GRID100	STATE
F0001	Hutan Bayu	2	1	hulu langat
F0002	Hutan Merah	4	7	kuantan
F0004	Hutan Paifik	6	77	miri
F0005	Hutan Aneh	2	4	hulu langat
F0006	Hutan Pelik	6	54	miri
F0008	Hutan Da	4	7	kuantan
F0009	Hutan Ayu	6	65	miri
F0010	Hutan Nya	6	13	miri

CONCLUSION

This research work provides an integrated conceptual model that partially overcomes some problems of spatio-temporal data model. Indeed, an analysis of existing model shows that such a basis is weakly defined. Spatial models use ad hoc ways of embedding space within data structures. Beside that temporal models tend to be poor in the supported data structures and include unnecessary constrains. Few models address both space and time, showing similar drawbacks. Based on human cognition, the model linked together the event-based space and time concepts. Such structure allows integrated operations on space and time, such as navigation, tracking and query. In this paper describes a new type of object-relational spatio-temporal data model for plant biodiversity. The data model consists of a data structure, operators and consistency rules. The conceptual schema (data structure) has been devised by the aggregating of three components of reality, i.e., space, time and attribute (each is considered as a class). By using biodiversity model forest department will be able to efficient management and integrates diverse set of spatio-temporal data that maybe required to analysis, monitoring biodiversity data. Biodiversity data model validation can be carried out to meet user requirements which are functionally created. To validate biodiversity data model few steps have been carried out.

1. We validate data model for supporting biodiversity data especially plant biodiversity collected from different biodiversity area such as Johor forest department, Pahang forest department and Mardi biodiversity data.
2. Data model is being tested for spatio-temporal object relational model using different set of queries including spatial, temporal and spatio-temporal queries. For this experimentation we will use the MS Windows XP Professional Edition

operating system on which we installed the other supporting software's such as Oracle 9i platform which is freely available from internet and also which can support spatial data and object-relational design of modeling. MS dot net and VB.net are the software languages used for the Interface design of the testing system.

3. Also set of defined queries used for performing different tests such as checking the model's Suitability for handling the spatial, temporal and spatio-temporal data

REFERENCES

M Clair Horner-Devine, “pattern of Diversity & Community Composition”, may 2004, PhD Thesis, Stanford University.

Shattri MANSOR, Mohammed Abu SHARIAH, Lawal BILLA, Iwan SETIAWAN and Fisal JABAR, “Spatial Technology for Risk Management” FIG Working Week 2004, Athens, Greece, May 22-27, 2004

DeokHwan Kim and ChinWan Chung, (2003), “Qcluster: Relevance Feedback Using Adaptive lustering for ContentBased Image Retrieval”, ACM SIGMOD 2003, June 9-12, 2003

Date, C. J. 2003. An Introduction to Database Systems. 8th Edition. Addison-Wesley, Reading, Mass.

adnan yunus, M.N. Sap, Mohd Taib, (Dec 2003), “A research Perspective Spatio-temporal Databases for Geographic Information System, Journal: Information Technolgy, Vol-15, No-2,

Momose Kuniyasu_ and Shimamura Tetsuya, June 2002, “Environments and People of Sumatran Peat Swamp Forests I:Distribution and Typology of Vegetation”, Southeast Asian Studies, Vol 40, No.1 June 2002

V. Katri, S. Ram, R.T. 2002. Snodgrass and G. O’Brien. Supporting User Defined Granularities and Indeterminacy in a Spatiotemporal Conceptual Model. Special Issues of

Annals of Mathematics and Artificial Intelligence on Spatial and Temporal Granularity, 36(1-2). 195-23

V. Katri, S. Ram, R.T. 2002. Supporting User Defined Granularities and Indeterminacy in a Spatiotemporal Conceptual Model. Special Issues of Annals of Mathematics and Artificial Intelligence on Spatial and Temporal Granularity, 36(1-2). 195-232.

Suhami Napis, Kamarudin, Khairudin & A.Latif, (2001), "Biodiversity Database for Malaysian Flora and Fauna: Update" University Putra Malaysia

Michel B. Thomas, Nan Lin and Howard W. Beck, (September 2-8, 2001), "A Database model for Integrating and Facilitating Collaborative Ethnomedicinal Research", University of Florida

Dimitris Papadia, Nikos Mamoulis, Vasilis Delis, (January 2001), "Approximate Spatio-Temporal Retrieval", ACM Transactions on Information Systems, Vol. 19, No. 1

Srikant B.J, Jayant R. Haritsa, (2000), "Design and implementation of Biodiversity Information Management System" database lab,Indian Institute of Science, Bengalor, India.

Spery L, *et al*. 2001. A spatio-temporal model for lineage metadata. Geoinformatica 5:51-70.

Gueting RH, Böhlen *et al*. 2000. A foundation for representing and querying moving objects. ACM Transactions on Database Systems 25: 1-42.

Jeremy L. Mennis, Donna J. Peuquet Diansheng Gua, (2000), "A semantic GIS Database Model for he Exploration of Spatio-temporal Environmental Data", 4th international conf. on Integrating GIS and Environmental Modeling (GIS/EMA4)

Gueting RH, Böhlen *et al*. 2000. A foundation for representing and querying moving objects. ACM Transactions on Database Systems 25: 1-42.

Tzouramanis, T., Y. Monolopoulos and N. Lorentzos, (2000), "Overlapping $_+$B tree – an Implementation of a Transaction Time Access Method. Data & Knowledge Engineering"

T. Bittner and B. Smith. 2001. "A Unified Theory of Granularity, Vagueness and approximation", In Proc. of COSIT Workshop on Spatial Vagueness, Uncertain and Granularity.

OMG. 2000. Unified Modelling Language Specification, Version 1-3. Object Management Group.

Huang B, *et al.* 2000. A spatio- temporal object model and query language (extended abstract). In: GIScience 2000: First International Conference on Geographic Information Science. 28-31 Octover, Savannah, Georgia, USA

Cattell R and Barry D. 2000. The Object Data Standard: ODMG 3.0. San Francisco, Morgan Kaufmann Publishers Inc.

Floor Eugnia Narciso, (April 1999), "A Spatiotemporal Data Model for Incorporating Time in Geographic Information Systems"

Theodoris, T., E. Stefanakis, and T. Sellis, (1998) "Cost Models for Join Queries in Spatio Database" Proc. 14th IEEE Conf. On Data Engineering, Orlando, Florida

Papadopoulos A. and Y. Monolopoulos, (1998), "Similarity Query Processing using Disk Arrays, Proc. ACM SIGMOD Conf. ACM SIGMOD Record, 27920

Tzourramanis, Vassilakopoulos, M., Monolopouslos, Y., (1998) “Overlapping Linear Quadtress: a Spatiotemporal Access Method, 6th ACM int. Workshop on Geographical Information Systems,

Vazirgiannis, M., Theodoridis, Y., And Sellis, T., (1998) “Spatio-Temporal Composition and Indexing for Large Multimedia Applications”, *Multimedia Syst. 6*, 4, 284–298

Papadopoulos A. and Y. Monolopoulos, (1997), “Performance of Nearest Neighbor Queries in R-trees” Proc. 6th int. Conf. On Database Theory

Seidl, T. And Kriegel, H.-P., (1997), “Efficient User-Adaptable Similarity Search In Large Multimedia Databases”, In *Proceedings Of The 23rd International Conference On Very Large Data Bases* (Vldb ’97, Athens, Greece, Aug.). 506–515

Mat Salleh, K. & A Latif., (1997), “Towords a comprised database of Malaysian medicinal and aromatic plants “

Andrei Ju, Korolyuk, 1997, Using ArcInfo to Evaluate Plant Biodiversity in Southern Siberia.

M. Miick and M. Polaschek, (1997) “A configurable Type Hierarchy Index for oodb, VLDB”, journal 6(4)

Michael J. Conroy And Barry R. Noon, (1996), Mapping Of Species Richness For Conservation Of Biological Diversity: Conceptual And Methodological Issues1, *Ecological Applications*, 6(3), 1996, pp. 763-773 © 1996 by the Ecological Society of America

B. Becker, S. Gschwind, T. Ohler, B. Seeger, P. Windnayer, (1996), “An Asymptotically Multiversion B-tree, The VLDB Journal 5(4)

Vazirgiannis, Theodoris, Y., M., Sellis, (1996), "Spatio-temporal Indexing for large Multimedia Applications", 3rd IEEE Conf. on Multimedia Computing and Systems, ICMCS'96, Hiroshima, Japan

Nabil, M., Ngu, A. and Shepherd, J., (August, 1996), "Picture similarity retrieval using 2D projection interval representation. *IEEE Trans. Knowl. Data Eng. 8*, 4, 533–539.

Papadopoulos A. and Y. Monolopoulos, (1996), "Parallel Processing of Nearest Neighbor Queries in Declusteres Spatial Data", Proc. 4th ACM GIS,

Smith, J. And Chang, S.,(1996) "Searching For Images And Videos On The World-Wide Web". Cu/Ctr 459-96-25

R. Vidya, Jayant R. Harisa, (1995), "Oshadhi: A Biodiversity Information System. In Proc. Of the International Conference on Management of Data", Indian Institute of Science, Bengalor, India

Peuquet, Donna J. and Niu Duan., (1995),"An Event-Based Spatio-Temporal Data Model for Geographic Information Systems," *International Journal of Geographical Information Systems*, 9(1):7-24

Claramunt C, Thériault M. 1995. Managing time in GIS: an event-oriented approach. In: Clifford J, Tuzhilin A (eds) Recent Advances in Temporal Databases. Springer-Verlag.

Peuquet D, Duan N. 1995. "An event-based spatio-temporal data model (ESTDM) for temporal analysis of geographical data", International Journal of GIS 9: 7-24.

Stonebraker, M.,(1995), "Chabot: Retrieval from a Relational Database of Images". *IEEE Computer 28*, 9, 40–48.

Worboys, M., (1994), “A unified model for spatial and temporal information”, *The Computer Journal*, 37(1): 26-34.

Faloutsos, C., Barber, R., Flickner, M., Hafner, J., Niblack, W., Petkovic, D., & Equitz, W., (July, 1994), “Efficient And Effective Querying By Image Content.” *J. Intell. Inf. Syst. 3*, 3/4, 231–262

Langran G. 1992. Time in Geographic Information Systems. Taylor & Francis, London Open GIS Consortium Inc.

Codd, E. F. 1970. A Relational Model of Data for Large Shared Data Banks. CACM 13 (6).

PUBLICATIONS

Journals

1. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohammed Shariff "Comparative Study of Plant Biodiversity Data Modeling in Research Perspective" Journal, Information Technology, Vol.16, No. 2, University Technology Malaysia, December 2004

2. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohamed Shariff, Harihodin Selamat, (2005), "Object-relation Data Model (ORDM) for Malaysian Biodiversity Data" Journal, Information Technology, University Technology Malaysia, Vol.17, No. 1, June 2005,

3. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohamed Shariff, Harihodin Selamat, (2005), "Spatio-temporal Data Model and Query Language for Biodiversity Data", Journal, Information Technology, University Technology Malaysia, Vol.17, No. 2, December 2005

International Conference / Symposium

4. Mohd Taib Wahid, A.Z.M Kamruzzaman , Abdul Rashid Mohamed Shariff and Harihodin Selamat, "Steps of Designing Biodiversity Data Model Using Object-Relational Data Model Techniques", The 4th International Multiconference on Computer Science and Information Technology (CSIT2006), Amman, Jordan, Vol. 2. Page 56, April 5-7 2006

5. Mohd Taib Wahid, Sarazmin Abd Razak, A.Z.M Kamruzzaman, Raja Baharudin Anom, “Learning Presentation Approach Using Multi-Criteria Decision Support System in A Mobile Behavior”, The 8th International Conference on Quality in Research (QIR2005), University of Indonesia, Depok, Jakarta, Indonesia, 8th to 10th August 2005
6. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohamed Shariff, Harihodin Selamat (2005), “Biodiversity Data Model (BIDaM) Using Object Relational Approach: Conceptual Framework”, An International Conference on Biogeography and Biodiversity: Wallace in Sarawak 150 years later, Sarawak Tourism Complex, Kuching, Sarawak, 13th – 15th July 2005

7. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohammed Shariff "Preliminary Plant Biodiversity data modeling and Spatio-temporal data model" International Symposium and Geographical Information System (ISG04), Hotel Istana, Kuala Lumpur, Malaysia, 21st - 3rd September 2004

8. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohamed Shariff, Harihodin Selamat, "Conceptual Spatio-temporal data model design for plant biodiversity" Malaysian Science and Technology Congress(COSTAM), CitiTel MID Valley, Kuala Lumpur, Malaysia, 18th-20th April 2005

9. Mohd Taib Wahid, A.Z.M Kamruzzaman, Abdul Rashid Mohamed Shariff, Harihodin Selamat, "Biodiversity Data Model for the Exploration of Spatio-Temporal Data”, First International Symposium on bio-inspired Computing (BIC’05) Puteri pan pacific hotel, Johor Bahru, Malaysia, September 5 - 7, 2005

University Level Conference

10. A.Z.M Kamruzzaman, Mohd Taib Wahid, Harihodin Selamat, “Conceptual Design Of Biodiversity Data Model (Bidam) Using Object Relational And Event Based Approach”, FSKSM Postgraduate Annual Research Seminar (PARS 05), D07, FSKSM, University Technology Malaysia, May 17-18, 2005

11. A.Z.M Kamruzzaman, Mohd Taib Wahid, Harihodin Selamat, "Object-relational Features for modeling and analysis of Spatio-temporal Data", FSKSM Postgraduate Annual Research Seminar (PARS 06), D07, FSKSM, University Technology Malaysia, May 24-25, 2006

Mardi Kuini Tables

- ABIOTIC STRESS SUSCEPTIBILITY
- ACCESSION
- ACCESSION CHARACTERIZATION
- ACCESSION EVALUATION
- ACCESSION REGENERATION
- adaxial leaf surface
- amount of fibre in pulp
- aroma
- arrangement of leaves
- beak
- BIOCHEMICAL MARKER
- biochemical marker reference
- biotic stress susceptibility
- BIOTIC STRESS SUSSCEPTIBILITY
- blotches surrounding stigma lobe
- branching density
- branching pattern
- collecting/acquisition source
- colour of new leaf flush
- condition of tree
- Copy of ACCESSION CHARACTERIZATION
- Copy of ACCESSION EVALUATION
- CROP
- crown shape
- CYTOLOGICAL CHARACTERS
- cytological characters name
- disease
- DISEASES
- DONOR
- dorsal shoulder
- eating quality
- ethnic group
- evaluation environment
- EVALUATION NOTES
- fertilizer type
- fibre texture
- flavour
- flesh colour
- flesh texture
- flower density in an inflorescence
- flower type
- fragrance
- frequency of plants at collecting site
- frequency of use of the plant
- FRUIT
- fruit base
- fruit bearing
- fruit color
- fruit shape
- fruiting season
- genetic erosion
- genus
- hairiness on the inflorescence
- HERBARIUM
- history of plant use
- IDENTIFIED GENES
- INFLORESCENCE
- irrigation methods
- juiciness
- leaf apex shape
- leaf base shape
- leaf colour
- leaf habit
- leaf margin
- leaf orientation
- leaf shape
- leafy bracts
- maturity period
- method of application
- method of use
- mineral toxicity
- MOLECULAR MARKER
- molecular marker name
- nature of disc
- note list
- NOTES
- number of stamens
- overall vegetation surrounding and ath the collecting site
- parts of plant used
- pest name
- PESTS
- petal colour
- petal shape
- PHOTO
- PLANT EVALUATION
- plant pesticide

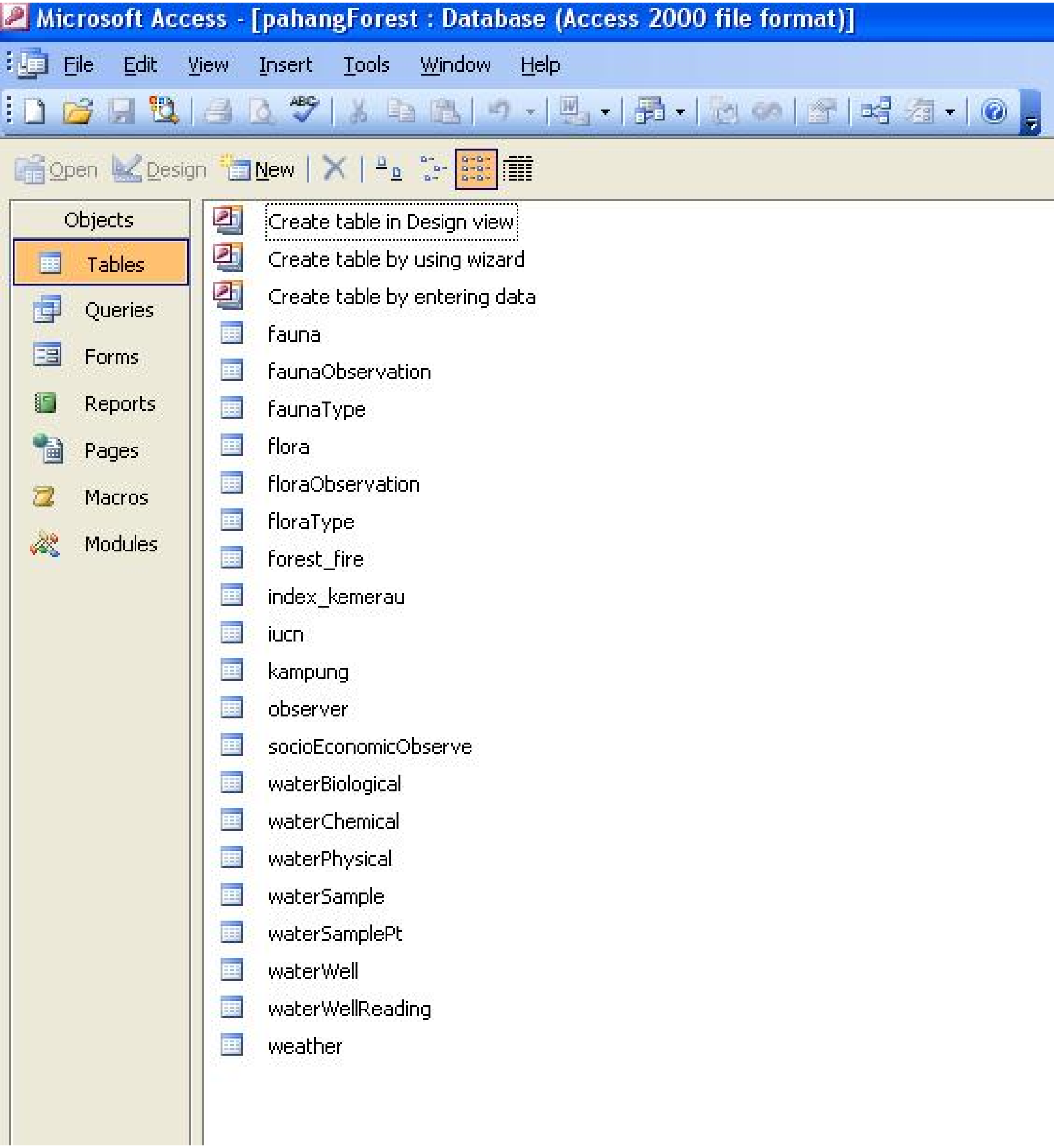
- PLANT ROOTSTOCK
- plant uses
- position of flowers
- position of inflorescence
- presence of pungent taste
- ratio trunk/rootstock diameter
- REACTION TO MINERALS
- reaction to salinity
- REGENERATION
- REGENERATION FERTILIZING
- REGENERATION IRRIGATION
- regeneration process (cryopreservation)
- regeneration process (in vitro)
- rootstock
- sampel status
- sampel type
- SAMPLING LOCATION
- scion trunk surface
- seasonality
- seed shape
- sepal colour
- shape of inflorescence
- sinus
- skin colour
- skin texture
- special uses
- species
- stalk attachment
- stress susceptibility
- sunburn susceptibility
- taste
- thickness of fruit skin
- tree vigour
- trunk surface
- type of explant
- type of germplasm storage
- type of market
- type of material for cryopreservation
- type of material received
- type of subcultured material
- type of subcultured material for recovery
- VEGETATIVE
- ventral shoulder
- waxiness on adaxial leaflet surface

Arthropod Collection System: Tables

- CLASS
- CLASS IDENTIFIER
- COLLECTION METHOD
- COMMENT
- CROP
- dataset
- FAMILY ANTHROPOD
- FAMILY HOST
- FAMILY IDENTIFIER
- GENUS
- GENUS ANTHROPOD
- GENUS HOST
- HABITAT
- HOST NAME IDENTIFIER
- Host plant
- IDENTIFIER
- INHABITANT
- INSECT HOST
- INSECT INHABITANT
- Institute
- LIFE HISTORY STAGE
- LOAN
- LOCATION
- MEMO TYPE
- METHOD OF IDENTIFICATION
- NAME QUALIFIER
- ORDER
- ORDER IDENTIFIER
- Paste Errors
- PERSONNEL
- PLANT INHABITANT
- SEX
- SPECIES
- SPECIES ANTHROPOD
- SPECIES HOST
- SPECIMEN
- Stage
- state
- storage
- STORAGE MEDIUM
- Type

MARDI PLANT GENEBANK DATABASE: Tables

- ABC
- ACCESSION
- ACCESSION CROPPING SYSTEM
- ACCESSION CULTURAL SITUATION
- ACCESSION IRRIGATION
- Accession location in orchard/crop base
- ACCESSION PROPAGATION
- ACCESSION1
- collecting/acquisition source
- condition of tree
- CROP
- CROP MANAGEMENT
- croping systems/pattern
- cultural situation
- DISTRICT
- DONOR
- ethnic group
- frequency of plants at collecting site
- frequency of typhoons/hurricane force winds
- frequency of use of the plant
- genetic erosion
- genus
- HERBARIUM
- history of plant use
- IMPORT EXPORT / MOVEMENT
- irrigation methods
- light
- method of use
- parts of plant used
- PHOTO
- plant uses
- propagation methods
- quality of the groundwater
- regeneration process (cryopreservation)
- regeneration process (in vitro)
- sampel status
- sampel type
- SAMPLING LOCATION
- seasonality
- SITE ENVIRONMENT CHAR
- SITE EVALUATOR
- SITE FERTILIZING
- SITE PLANT PROTECTION
- soil depth to groundwater table
- soil drainage
- soil erosion
- soil fertility
- soil matrix colour
- soil moisture
- soil organic matter content
- soil particle size classes
- soil pH
- soil salinity(dissolved salts)
- soil texture classes
- SOURCE ENVIRONMENT
- special uses
- species
- state
- stoniness/rockiness/hardpan/cementation
- topography
- type of explant
- type of germplasm storage
- type of market
- type of material for cryopreservation
- type of material received
- type of subcultured material
- type of subcultured material for recovery
- water availability


Microsoft Access - [pahangForest : Database (Access 2000 file format)]
File Edit View Insert Tools Window Help
Open Design New
Objects
Tables
Queries
Forms
Reports
Pages
Macros
Modules
Create table in Design view
Create table by using wizard
Create table by entering data
fauna
faunaObservation
faunaType
flora
floraObservation
floraType
forest_fire
index_kemerau
iucn
kampung
observer
socioEconomicObserve
waterBiological
waterChemical
waterPhysical
waterSample
waterSamplePt
waterWell
waterWellReading
weather