OBJECT-RELATION DATA MODEL (ORDM) FOR MALAYSIAN BIODIVERSITY DATA

Mohd Taib Wahid[1], A.Z.M Kamruzzaman[2], Abdul Rashid Mohamed Shariff[3], Harihodin Selamat[4]
[1]taib@fsksm.utm.my, [2]akzaman@siswa.utm.my, [3] rashid@eng.upm.edu.my, [4]harihodn@itp.utm.my

Department of Information Systems, Faculty of Computer Science & Information Systems
University Technology of Malaysia, K.B. 791
81310 Skudai, Johor, Malaysia
Tel: (607)-5532422, Fax: (607) 5565044

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 paper discusses about the object-relational conceptual design of biodiversity data model for long term stewardship of biodiversity information. A plenty of data models have been developed (such as BODHI, Oshadhi) which only support spatial data. In addition to this, most of the data models are relational model. The purpose of this research is to 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 goal of this study is to design a biodiversity model that is efficient for data management and able to integrate diverse set of spatio-temporal data that maybe required for analysis, and monitoring biodiversity data.

Keyword: Biodiversity, Data Model, Object Relational, Event Based, GIS, BODHI

1. Introduction

Biodiversity, or biological diversity, refers to the variety of life on earth and the most important factors influencing the stability and health of ecosystems. Information about biodiversity and ecology is vital to a wide range of scientific, educational, industrial, and governmental activities. Malaysia is one of the richest rain forests in the world with diverse floristic composition and complex ecosystem. It is not only rich for plant biodiversity data also rich in fauna and peat swamp. Malaysia has been loosing much of its natural resources such as plants and animal species through ecosystem and degeneration.

Although the extensive research has been performed in the area of plant biodiversity recent years, managing plant biodiversity data with database system still poses many challenges [1].

Nowadays data management tools and software need more sophisticated facilities to face new requirements from emerging application areas and non-traditional user interactions. In particular, better concepts and tools for manipulating spatio-temporal data are needed. Major DBMS tools are incorporating facilities for spatial or temporal data management (e.g., Oracle's Spatial Cartridges and Informix's Databases). Temporal systems are still somehow behind, with no generic products on the marketplace, there are few ad hoc systems or application-specific developments (e.g., for time series management). However, current tools do not match the user perception of and reasoning about the application data.

Modeling is an essential part of the environmental and physical science for last couple of decades. Recently, it also has become an increasingly vital part of biodiversity data and geographical information system. In traditional database there is mismatch between the logical, implementation-oriented view of data supported by the tools, and the application oriented, conceptual view that users follow their everyday task. This mismatch is similar in traditional databases management. The relation approach and conceptual approach to logical gap was filled by data model design CASE tools based on the entity-relationship (ER) approach. Since then, the conceptual approaches to the data modeling have been extensively demonstrated, in terms of user involvement and the
durability of the design specifications. Spatio-temporal data management tools to be complemented with user-oriented design CASE tools but the problem is that there is no agreed upon conceptual model on which to build such tools.

Thus data models have been proposed either spatial (e.g., BODHI) or temporal modeling (a few for spatial-temporal modeling), but they fail to show a clean conceptual underlying philosophy. It is rare to find a data model for the collection and storage of biodiversity data to be defined and implemented as a fully working database, prior to the collection of the data, in other words there is no "clean slate", and hence scientist are continuously forced into a position of migrating data from one model to a new model to a new, improved model [2]. This paper, report on the design of a conceptual spatio-temporal data model for plant biodiversity called BiDam (BIODIVERSITY DATA MODEL).

The criteria's that identified as guideline for building the conceptual model. Overall conceptual framework, also discuss spatial and temporal features in a conceptual model. This report also discusses little bit about existed BODHI system, extension time, event-based approach, object design and spatio-temporal conceptual model design.

2. Biodiversity Data

Biological diversity plays a very important role in our lives. There are various definitions on biodiversity has been given by many researchers. Perhaps the best definition on 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*. Plant biodiversity data is an assortment of different types of plant taxonomy or manifest itself on the genetic species, environment, landscape levels, manipulated to analyze the past, define the present and the consider possibilities of the future. Furthermore, all biodiversity data is an assortment of different types of organisms that co-occur in time and space [3]. Biodiversity data can be classified by into three groups. More about biodiversity data describes bellow:

a. Taxonomy

Taxonomy data is classification of plant species that reflects the similarities and differences among the species. In taxonomy, each unit of classification is taxon. Taxon can be classified into four major classifications of species such as order, family, genera and species. An order contains several related families which mean the families that exhibit certain similar characteristics are grouped into an order. Likewise a family is represented by a collection of genera and a genera is a collection of related species. Each taxon has a biological name [4]. The classification hierarchy of species is shown in Figure 1. Species are identified based on their leaf stem flower and fruit characteristics. Hence, these characteristics could be referred to as identifying characteristics. There are other characteristics for example the characteristic of root which do not aid in the identification of species. These characteristics could be referred to non identifying characteristics. The various characteristics which depend on these relationships may vary in time due to discover of new class characteristics, corrections to previously recorded characteristics etc. Hence every order family genera and species is associated with a set of identifying and non-identifying characteristics which differentiate it from the other orders, families, genera and species respectively. Figure 1 is an example of species hierarchy. A species may have synonyms the geographical location details and the textual in formation associated with it.

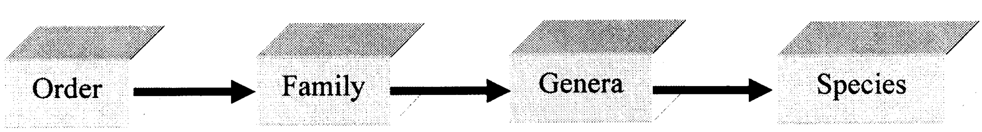

Figure 1: Taxonomy of Species Hierarchy

b. Geo-spatial data

Study of ecology of species involves recording the geographical and geological features of their habitats, water-bodies, artificial structures like highway which might affect the ecology, etc. These are represented on a map of the region and have to be handling as spatial data by the database. Geographical and taxonomy has a rich inter-relationship. The distribution of a plant species in a geographical organization is an example of this type of relationship. Another instance is that, for all species found in a geographic region, the spatial and temporal variations in population have to be recorded.

c. Temporal data

Biodiversity data is a temporal variation (i.e. over time) of the condition. Though, biodiversity itself is not a temporal data. Biodiversity itself is a spatial data but it does co-occur over time. Changes occurring in geo-referenced objects need to be captured and sufficiently represented in the temporal database. One needs to begin by identifying which are the most desirable changes to capture, as often many changes will occur to a single spatial object. Beside that, biodiversity data also captured/identified and grows in different places and different times. The types of changes captured data will vary with the type of application being designed. For example, one application might involve storing the biodiversity data of an object such as attribute data and another application might involve capturing the spatial changes of an object such as geometric data.

The above data types have complex and deeply-nested relationships within and between themselves. An important point to be noted that, all these categories are intra-related and inter-related. For example, the geographical distribution of a species relate to the taxonomic data and geographical data. Furthermore, above data types may involve sophisticated structures such as sequences and sets.

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 rated as one of the world's 12 "mega diversity" countries. It boasts over 150,000 species of invertebrates, 286 mammal species, 736 bird species and 15,000 types of flowering plant. Forest resources have been one of the major sources of revenue in the Malaysian economy. However, it is decreasing every year.

Different types of forests can be found in the Peninsular, Sabah and Sarawak. From the management perspective, forests of Peninsular Malaysia can be classified into dipterocarp, peat swamp and mangrove forests. Among forest area which approximately 95% is covered by dipterocarp forests, 3.34% by peat swamps and 1.84% by mangrove forests (UNEP, 2002). In Sabah, there is a gradual succession of forest vegetation from the coastal beach forests and mangrove forests to lowland dipterocarp forest and eventually montane forests. In Sarawak, five types of natural forest types are abundant namely: Hill Mixed Dipterocarp Forest, Peat Swamp Forest, Mangrove Forest, *Kerangas* Forest (Heat Forest) and Montane Forest (UNEP 2002). However, this estimation is still on process and area in Sabah and Sarawak have not been actively surveyed botanically.

2.2 Biodiversity Data Model

Data model for botanical collections for taxonomic databases have been developed by many researchers at various places since 1992, e.g. ASC (1992), Bolton et al. (1992), Sinnot (1993), Wilson (1993), NMNH (1994), and ITIS (1995). All represent attempts to bring order into the complex data structure which are involve when plants are named, collected, classified and investigated as to their properties. Bodhi (Biodiversity Object Database architecture) data model has designed based on Indian plant biodiversity [5, 13]. Recently the academy of natural science of America developed a relational database and
implemented for biodiversity [5]. There are few others data model developed based on Malaysian plant such as Ethnobotany of Malaysia Plants Online [7, 8], APMIS (Alian Plant Management Information System) [9] and data model for botanical collection [10].

2.3 BODHI System

In BODHI, object oriented paradigm is used to achieve following features necessary for biodiversity data:

a) Support multiple data primitives through the use of type libraries at the database layer. (The spatial data primitives are provided using facility.)
b) Representation of complex relationships such as sets, sequence and bags
c) Build new type through inheritance and aggregation of previously defined non-primitive types.

Data modeling language of BODHI extends the standard ODL by introducing new primitives for modeling spatial and sequence data. Spatial data or geographic data forms a key component of a Biodiversity data repository. BODHI provides set of spatial data types and query languages and support efficient spatial indexing and spatial joint algorithm. Primitives to represent single spatial objects like country, state, forest, river etc.

Spatial data model of BODHI provides two categories of primitives: *Simple Primitives* and *Compound Primitives*. Simple primitives enable modeling of single object in space, and includes types of *Point*, *Polyline* and *Polygon*. The compound primitives are used to model spatially-related collection of objects. Compound primitives also classified into two categories such as *Layer* and *Network*, for modeling collection of *Polygon* and *PolyLine*, respectively. The Figure 2 gives the class diagram of Spatial Data model of BODHI.

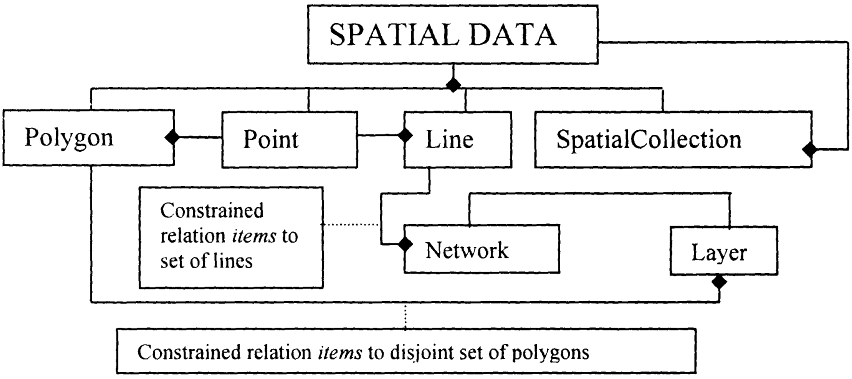

Figure 2: Class diagram of Spatial Data Model in BODHI

From the above spatial data model, we also have the spatial data function i.e. $f_{(SD)}$ which is composed of different subcomponents classes called point, line and polygon and collection. Spatial data function can be written such as:

$$f_{(SD)} = Point + Line + Polygon + SpatialCollection \ldots\ldots\ldots \text{(i)}$$

3. Design of Proposed Biodiversity data model

Based on our study on the above mentioned data models in section 2.2, biodiversity and GIS data model is relation. Relational data model cannot support complex data, data analysis, data manipulation and time factor. After analyzing of all the models developed since 1993, most of the models are used to collect plant data, plant listing, and plant conservation. There is no data yet design which can support data analysis, data retrieval, and temporal data (time). Early 2000, biodiversity object database arcHItecture (BODHI) was developed to handle plant taxonomies. To support spatial and temporal of plant biodiversity data, one robust data model is must be developed

The main purpose of this research is to design a spatio-temporal biodiversity data model by using object relational data model techniques. Two aspects that must be considered when designing an object relation data model one is identification of attributes data as an objects and another is temporal issue on biodiversity data. The goal of this study is to design a biodiversity model that is efficient management and able to integrates diverse set of spatio-temporal data that maybe required to analysis, monitoring biodiversity data

3.1 Extensions For Time In Plant Biodiversity Model

In order to implement temporal application, non-temporal database systems need to be enhanced in three ways. First, the data structures (DS) have to be extended to record the time information. Second, new operations (OP) using the additional temporal data semantics of the data have to be provided in order to queue and modify temporal data. Third, temporal constrains (C) must be expressible. So, A temporal data model $M^T= (DS^T, OP^T, C^T)$ should enhance all concepts contained in the three components of a data model with respect to time.

Usually, extending the data structures with time attributes does not cause any severe problems. When timestamping data; two different time dimension can be distinguished. *Valid time* records time when data is true in reality. *Transaction time* records when data is stored in the system. To store valid time data, two additional attributes of type **Date, VTS** (Valid Time Start) and **VTE** (Valid Time End), can be added to (maybe already existed) non-temporal data structures denoting the start and the end point of valid time. The same can be done for transaction time. However, the idea presented could easily be generalized to deal also with transaction time.

Spatial data (SD) is a set of Spatial Service (combination of spatial index, spatial operation and spatial data) and Object Service (combination of object operation, object index and taxonomy data). As from the discussion valid-time as V and transaction-time as T, AD contains attribute data and TD which contains the valid time-time and transaction-time

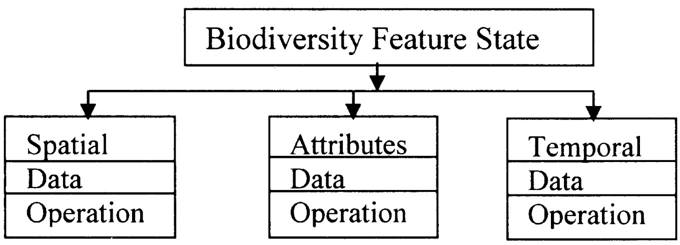

Figure 3: Biodiversity Features state objects

From the above figure 3, found that, biodiversity feature state is a combination of spatial, attribute and temporal data.

3.2 Object Extension of Biodiversity Data Model

Structural modeling, which describes the structure of similar objects in terms of classes, their similarities and differences (generalization), the associations or connections among these classes, and the structural - constrains.

First of all, beside built-in data types (also called scalar), the ability to define *user-defined data types* (also called abstract data types) is provided. There are two types of user-defined data types:

- Object types
- Collection types

An object data type is an abstraction of real world entities, representation of which must be stored in a database. An object type is a schema object with a name, set of *attributes* and *methods*. Each attribute is of either a build-in data type or a user-defined data type. This allows defining object types with complex data structure.

Methods of object types fall into one of three categories: member, static and comparison. Member methods are used for modeling behavior of objects, static methods are used for modeling of the object type as a whole, and comparison methods are used for comparing instance of given object type. In addition, every object type has a system-defined constructor method.

Similarly like a class in programming language, an object type is a temple for instances called *objects.* Objects can be instantiated by means of the constructor method of a given object type and stored in *object tables.* An object table can be viewed either as a single column table of row objects or as a multi-column table. Each row object has assigned a unique object identifier.

The last but not least major object extension concerns views. Just as a view is a virtual table, an *object view* is a virtual object table. By using object views, it is
possible to create a virtual object table with column of both built-in and user-define data types mapped to columns of relational or object based tables. Object views provides the ability to offer specialized or restricted access data stored in relational and object tables. In addition, object view definition contains information about the object type of the view objects, about the way of constructing OID from attributes and column of base table, and a mapping SELECT statement.

A simplified model of biodiversity object relational data types is shown in Figure 4 and structure class diagram in Figure 5.

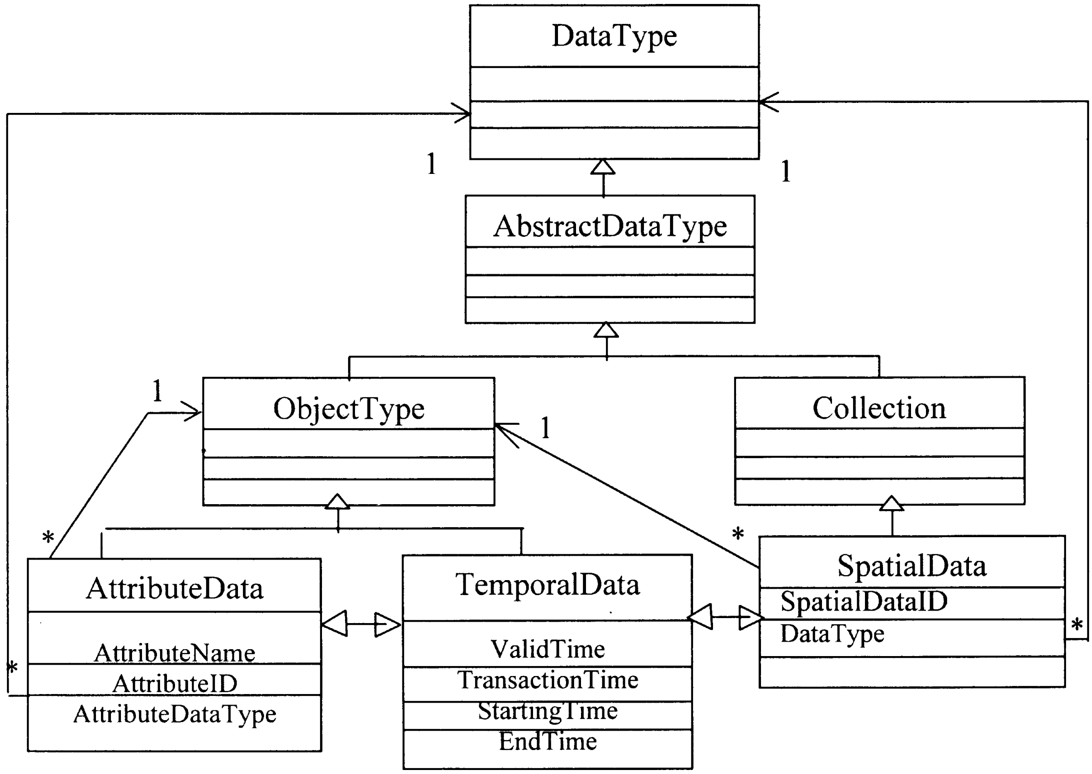

Figure 4: Data Types in Biodiversity Data Model

3.3 Spatio-Temporal Biodiversity Data Model

The design of an object-Relational Data Model generally involves two basic steps. First, structural modeling, which describes the structure of similar objects in terms of classes, their similarities and differences (generalization),
the associations or connections among these classes, and the structural constrains. The second step corresponds to behavioral modeling, which describes the behavior of different classes in terms of operations and relationships. What follows in this section is a description of the structure of the spatio-temporal data model.

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 5. It shows that five major classes were identified and incorporated into the data model: BiodiversityFeature (BF), BiodiversityData (BD), Spatial Data (SD), Attribute Data (AD), and Temporal Data (TD).

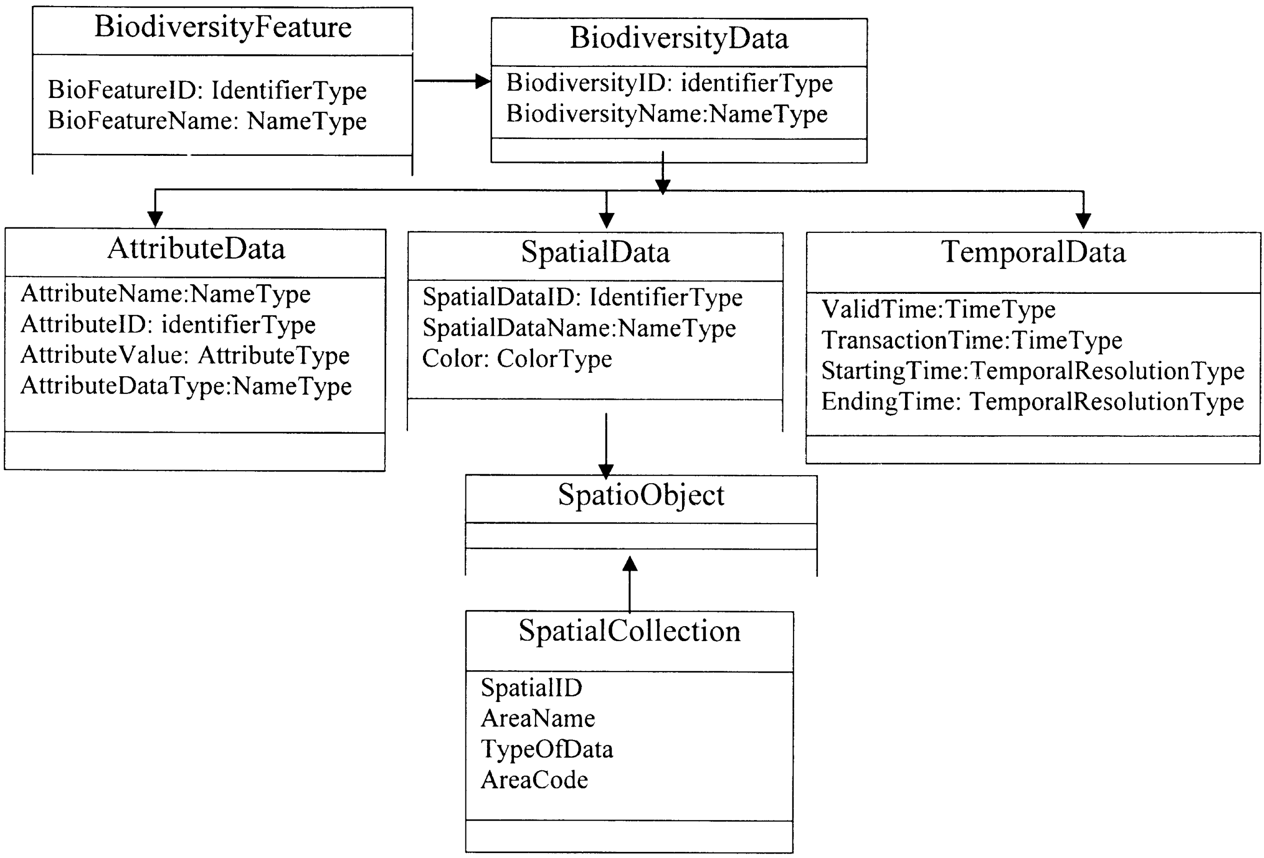

Figure 5: Conceptual class diagram of object-relational schema structure

In Figure 5 is tried figure out to show how the basic object extensions of biodiversity data model. The BiodiversityData class has three subclasses contain of complex attributes which are AttributeData, SpatialData and TemporalData. Subclass AttributeData containing of AttributeName, AttributeID, AttributeDataType and AttributeValue. Spatil data contains of SpatialObject which are containing of SpacialCollection.

4. Conclusion and Future Work

This paper 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 [14]. 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 to be carried out.

1. We will validate data model for supporting biodiversity data especially plant biodiversity collected from different biodiversity area such as Johor forest department, Pahang forest department.
2. Data model will be 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

[1] Mat Salleh, K. & A Latif, "Towards a Comprised Database of Malaysian Medicinal and Aromatic Plants. 1997

[2] Berendsohn, et al., W.G, "A comprehensive reference model for biological collections. Taxon", 1999, 48:511-562

[3] Noss, R. F. and A. Y. Cooperrider, "*Saving Nature's Legacy"*. Island Press, Wash. D.C., 1994, 416 pp.

[4] P C Vasishta, (1974), "Taxonomy of Angiosperms", S Chandand Co. Delhi

[5] Srikant B.J, Jayant R. Haritsa, Uday S. Sen "The Building of BODHI, a Biodiversity Database System", 2000

[6] Paul J. Morris, "Relational Database Design and Implementation for Biodiversity Informatics", The Academy of Natural Sciences1900 Ben Franklin Parkway, Philadelphia, 2005, PA 19103 USA

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

[8] Mat Salleh, K. & A Latif., "Towords a comprised database of Malaysian medicinal and aromatic plants", 1997.

[9] Joseph Maada Korsu Kendeh, "Geographical Information Systems for Rural Applications", international institute for geo-information science and earth observation, enschede, the Netherlands, March 2002,

[10] Berendsohn, W. G., "The concept of "potential texa" in databases", Taxon 44: 207-212. 1995. ISSN 0040-0262

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

[12] Kertesz, J.T, "A synonymized checklist of the vascular flora of the United states", Canada and Greenland 2nd edition, 1993, pp 622. Timber Press, Portland

[13] Srikant B.J, Jayant R. Haritsa, "Design and implementation of Biodiversity Information Management System", 2000

[14] 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 December2004, University Technology Malaysia