BIODIVERSITY DATA MODEL (BiDaM) USING OBJECT RELATIONAL APPROACH: CONCEPTUAL FRAMEWORK

Mohd Taib Wahid[1], A.Z.M Kamruzzaman[2], Abdul Rashid Mohamed Shariff[3]
Harihodin Selamat[4]

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
[1]taib@fsksm.utm.my, [2]akzaman@siswa.utm.my

ABSTRACT

The extensive research has been performed on plant biodiversity and even on mangrove area recent years; managing plant biodiversity data with database system still poses many challenges. Plant biodiversity is the variety of different types of plants species that growth in various landscape. There are phenomenal growths in the area of biodiversity studies, largely motivated by its economic and humanitarian. Various data models, query languages and techniques have been proposed by many researchers. Traditional database systems (built on relational hierarchical and network models) which are widely used for commercial applications such as banking fail to meet the modeling and processing requirements of the biodiversity data. Recently developed BODHI (Biodiversity Object Database arcHItecture) data model which is designed based on Indian plant biodiversity and only support spatial data. Integration between plant Bio-diversity data (BODHI) and geographical data with event-based approach can make an enhancement for the plant biodiversity data analysis and data manipulation. In addition, most of the data models are relational model. However, relational model can not support temporal data (time) effectively; it is best for spatial data (space). The main objective of the present study is to design a conceptual data model which will be later on complete data model for plant biodiversity. Conceptual data model is the extension of BODHI data model by using object relational data modeling and event-based approach to support temporal data. Therefore, conceptual data model by using object relational and event based is an appropriate approach to design a plant biodiversity data model. Moreover, Forest department, land and agricultural management and other related research organization could be benefited from the outcome of this model.

Keyword: Biodiversity, Data Model, Object Oriented Approach, Event Based, GIS, BODHI, Mangrove

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, governmental activities. Malaysia is undoubtedly on 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 including plants and animal species through ecosystem and degeneration.

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, just a 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 approach 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).

Few sections discuss the criteria that we identified as guideline for building the conceptual model. Overall conceptual framework, discuss spatial and temporal features in a conceptual model. This report also discusses little bit about existed BODHI system, time extension, 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 given by 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 organization 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 and landscape levels, and 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 that co-occur in time and space [3]. Biodiversity data can be classified by into three groups.

a. Taxonomy
Taxonomy data is classification of plant species that reflects the similarities and differences among the species.

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 relationship.

c. Temporal data
Biodiversity data is a temporal variation (i.e. over time) of the condition. All plant species found in a geographic region, the spatial and temporal variation in a population have to be recorded.

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 related the taxonomic data and geographical data. Further, they may involve sophisticated structures such as sequences and sets. Bio-diversity scientist faced many difficulties are the effective management and access of the large amounts and varied types of data that arise in their studies, ranging from micro-level biological information such as genetic makeup of organisms and plants.

3. 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. The change of land covered by forests has been reduced from 65.9% in 1990 to 59.5% in 2002 (Thang, 2004).

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 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). Currently there are total 610,606 hectares of mangrove forest estimated in Malaysia. The largest concentration are found in Sabah (56%) followed by Sarawak (27.5%) and the remaining in Peninsular Malaysia (16.5%) [14]. However, this estimation is still on process and areas in Sabah and Sarawak have not been actively surveyed botanically. Based conservation, management, data analysis, data accessing, monitoring and data complexity of biodiversity data, data model is required to protect biodiversity data.

4. 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) is designed a data model based on Indian plant biodiversity [4, 12]. 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 [6, 7], APMIS (Alian Plant Management Information System) [8] and data model for botanical collection [9].

After study of stated above data model, most of the model of biodiversity and GIS data model is relation. Relational data model cannot support complex data, data analysis, data manipulation and time factor. 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, biodiversity object database arcHItecture (BODHI) developed to handling plant taxonomies. To support spatial and temporal of plant biodiversity data, one robust data model is required to develop.

To achieve the objective of designing spatio-temporal plant biodiversity data model, several steps (database life cycle) needed to address; these are steps are (figure 2):

a. Plan (planning, analysis, requirements collection)
b. Design (conceptual and logical design, physical design including database architecture)
c. Implementation the model legacy
d. Testing

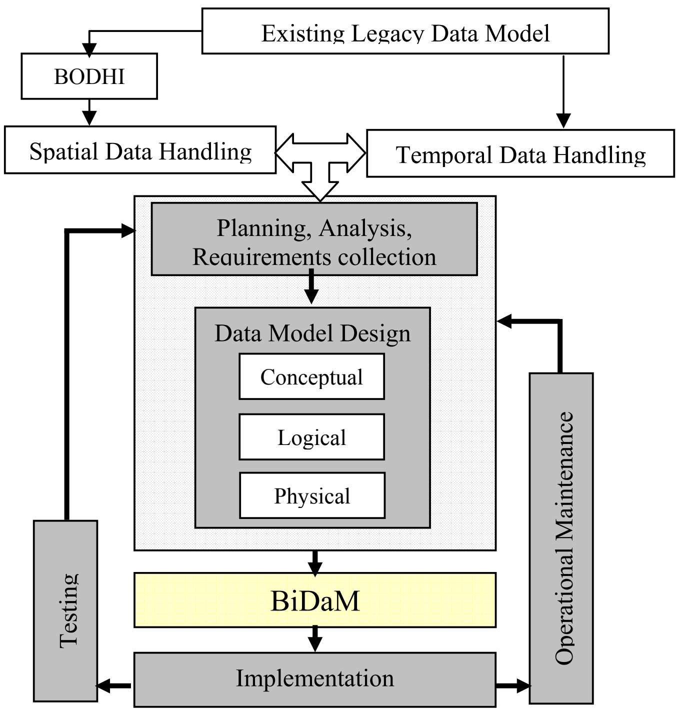

Figure 1: Overall Frame work of Plant Biodiversity Data Model

The main purpose of this research is to develop a data model that better facilitates the exploration and analysis of plant biodiversity data to support temporal data. The goals of this research is to design a conceptual plant biodiversity data model which is combination of BODHI data model and event-based techniques by using object relational approach to support temporal data. Therefore, conceptual data model by using object relational and event based is an appropriate approach to design a plant biodiversity data model. 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. In the paper we will only focus on conceptual data model design as we had done our plan stage earlier. Overall data model development life cycle has shown in Figure 1.

5. 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.

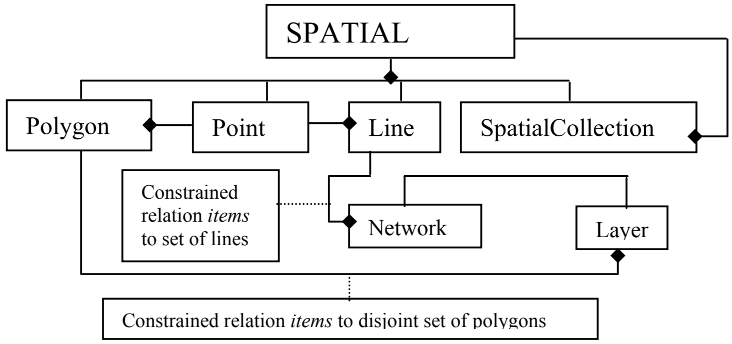

Figure 2: Class diagram of Spatial Data Model in BODHI

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:

$$f_{(SD)} = Point + Line + Polygon + SpatialCollection.$$

6. COMPONETS OF SPATIO-TEMPORAL DATA

A spatial object can be seen as a composition of three components, the spatial component, the temporal component and the non spatio-temporal or aspatial component. Thus to represent a spatial object in a spatio-temporal data model we need to define these three components for each spatial object. This would mean we would have to identify the “What”, “Where” and “When” for each spatial object. All three of these components need to be addressed in order to identify each object in a spatio-temporal data model.

The temporal component in a spatio-temporal data model is the most significant of the three components. When examining changes occurring in spatial data, we are essentially examining the non-temporal components of a spatial object such as geometry, forestry and attribute data. However, these attributes and geometric relationships in the data are influenced by time. Time defines what attributes currently exist for a spatial object, what geometric relationships are present and determines when changes have occurred to a spatial object. No spatial object is fixed with the passage of time and it is clear that when examining spatial objects and their attributes, time needs to be incorporated in the analysis if we are to correctly model the objects in their real world. If time is incorporated as a component of a geographical object, the possibilities of analyzing an object over a passage of time and keeping track of the objects' history is greatly increased. In our application, biodiversity data is underlying with spatial objects. Thus the biodiversity data is inter-related with spatial and temporal data.

The main challenge facing the design and implementation of a spatio-temporal data model for biodiversity data is the representation of the spatial and temporal component of the data model. 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.

7. 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 was true in reality. *Transaction time* records when data has 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.

Proposed data model will be allowed to support temporal data with plant biodiversity data. From the above discussion, we can represent our model in Equation i:

$$f_{(SD)} = f_{(BFS)} + f_{(AD)} \qquad \text{-------------- (i)}$$

In the above defined equation of 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). If we further classify our model as in equation (ii) so we can get for Biodiversity Feature State (BFS) i.e.

$$f_{(BFS)} = f_{(SD)} + f_{(AD)} + f_{(TD)} \qquad \text{------- (ii)}$$

As from the discussion valid-time as V and transaction-time as T, so from the Equation (ii) we have AD contains attribute data and TD which contains the valid time-time and transaction-time so for TD we have the new equation which is:

$$F_{(TD)} = V + T \qquad \text{--------------- (iii)}$$

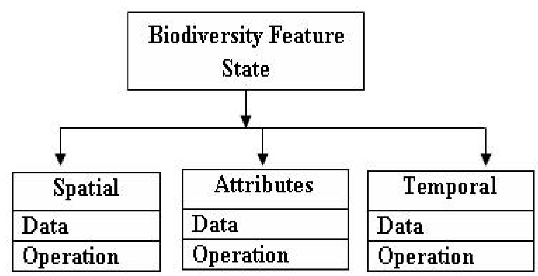

Figure 3: biodiversity feature state objects

From the above discussion and equation (ii) we have found Biodiversity Feature State is a combination of spatial, attribute and temporal data shown in figure 3.

8. STRUCTURE AND CONCEPTUAL FRAMEWORK OF PLANT BIODIVERSITY

In order to develop a well-designed model, a systematic approach is needed. This is where we need to determine that which methodology is to be used for developing model. Conceptual models allow users and analysts to concentrate on essential aspects of an application domain, without being bothered by constraints of a specific implementation platform. Thus, conceptual models provide descriptions that are closer to a human perception of events in the real world and that facilitate man-machine communication. Users need not transform their intuitive specifications to adjust to the technological constraints of specific systems.

9. SPATIO-TEMPORAL CONCEPTUAL DESIGN FRAMEWORK OF PLANT BIODIVERSITY

A conceptual design represents the structure of the system. *Conceptual Design* means the work of creating a high-level structure for the system. 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 [13]. To obtain the model, an associated visual syntax has been defined to achieve simplicity and readability. A set of well known concepts are supported as follows:

Objects Types:

An *object* represents a real-world entity. An *object type* describes a set of objects with similar structure and behavior. 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 or temporal), spatial, temporal, or spatio-temporal. For example, in plant biodiversity data model different object represents the different real world entities as shown in Figure 6.

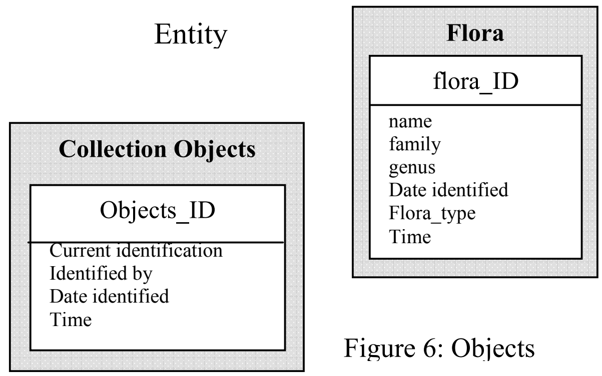

Figure 6: Objects

Relationship:

Relationships are an essential part of conceptual design. A *relationship* is a link between two or more objects, where each object plays a given role. A *relationship type* describes a set of links with similar characteristics. 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.

In the above Figure 7, two different scenarios have been shown to describe the relationships. One scenario displays a relationship named as "Identifies" that identifies an object "collection object" to another object "Identification".

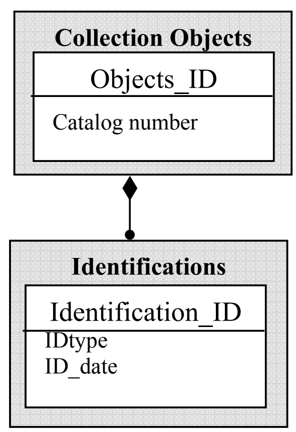

Figure 7: Identifies relationship

Attribute:

An *attribute* represents a real-world property; both object types and relationship types may have attributes. Attributes can be:

- Simple (with atomic values) or complex (i.e., composed of simple or complex attributes);
- Monovalued (with a single value) or multi-valued (with a multiset value);
- Mandatory (with a value in every instance) or optional (with a value in some instances and no value in others).

Here in the Figure 8 displays an object "flora" has three attributes i.e. flora_ID (primary key), Name, family.

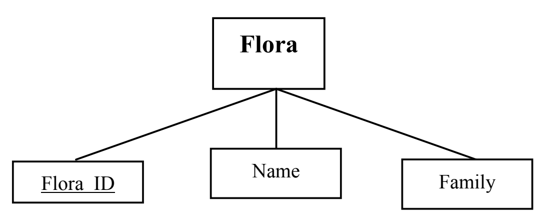

Figure 8: Entity and its Attributes

The proposed conceptual 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 four major classes were identified and incorporated into the data model: Biodiversity Feature State (BFS), Spatial Data (SD), Attribute Data (AD), and Temporal Data (TD). In this Class diagram, the

Biodiversity Feature State class represents the highest level of data abstraction and describes Biodiversity Phenomena which are composed of one or more Biodiversity features. Examples of biodiversity phenomena which is related with geographic are such as location name, where and what types of biodiversity data interrelate with that region.

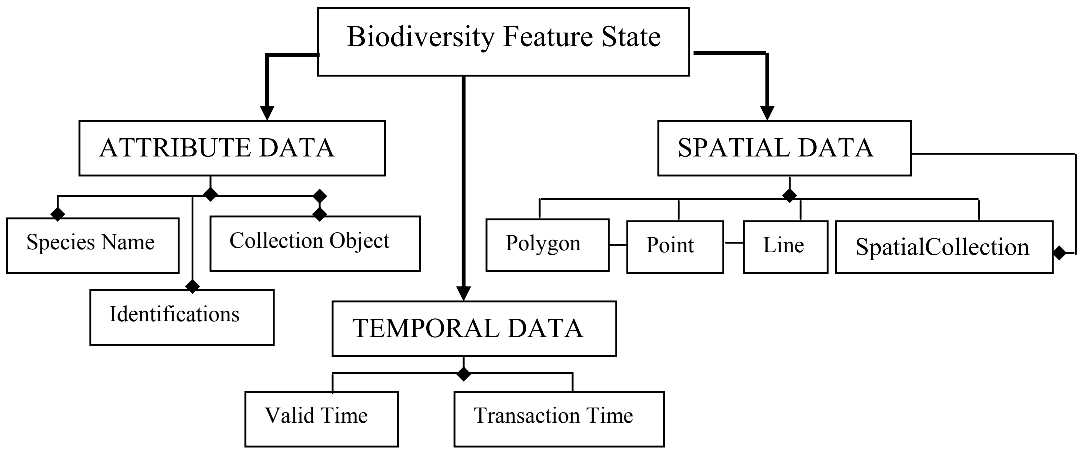

Figure 9: Overall conceptual data model of plant biodiversity

10. CONCLUSION

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. 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 /or 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 conceptual spatio-temporal data model for plant biodiversity. Unlike overall design of spatio-temporal plant biodiversity model will be designed to explicitly represent change over space relative to time. 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). To complete data model logical and physical design work will be design later.

REFERENCES

[1] Mat Salleh, K. & A Latif, 1997, "Towards a Comprised Database of Malaysian Medicinal and Aromatic Plants.
[2] Berendsohn, et al., W.G. A comprehensive reference model for biological collections. Taxon, 48:511-562, 1999.
[3] Noss, R. F. and A. Y. Cooperrider. 1994. *Saving Nature's Legacy*. Island Press, Wash. D.C. 416 pp.
[4] Srikant B.J, Jayant R. Haritsa, Uday S. Sen ,2000 "The Building of BODHI, a Biodiversity Database System"
[5] Paul J. Morris, 2005, "Relational Database Design and Implementation for Biodiversity Informatics", The Academy of Natural Sciences1900 Ben Franklin Parkway, Philadelphia, PA 19103 USA
[6] Suhami Napis, Kamarudin, Khairudin & A.Latif, (2001),"Biodiversity Database for Malaysian Flora and Fauna: Update" University Putra Malaysia
[7] Mat Salleh, K. & A Latif. 1997. Towords a comprised database of Malaysian medicinal and aromatic plants

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

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

[10] 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

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

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

[13] Adnan Yunus, (2005), "Spatio-temporal object Relational Model for GIS", Dissertation, Universiti Teknologi Malaysia, Skudai Johor

[14] Gan Boon Keong, (2004), "The Matang Mangrove Working Plan- A Case of Systematic and Planned Approach in the Management and Conservation of Mangrove Resource and Biodiversity", Treewise Forest Consu;tancy Services, Perak, Malaysia.

[15] Cristine P., Stefano S., Esteban Z, (1999), "Spatio-Temporal Conceptual Models: Data Structures + Space + Time", ACM GIS 99, Kansas City, MO USA