CONCEPTUAL SPATIO-TEMPORAL DATA MODEL DESIGN FOR PLANT BIODIVERSITY

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
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

Malaysia has been acclaimed as one of the mega diversity countries in the world also has been ranked the 12th richest country in terms of the number and diversity of plant species. To manage the large amount of plant species of Malaysia has become very essential with rapid growth in the number of plant species due to their rich economic importance. Although the extensive research has been performed in plant biodiversity area recent years, managing plant biodiversity data with database system still poses many challenges. 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. Geographical data have an object and location, attribute and time. 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 to this, most of the data models are relational model. However, relational model can not support temporal data (time) effectively; it is only support spatial data (space). Main purpose of this paper 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. 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

1. Introduction

Malaysia is undoubtedly on the richest rain forests in the world with diverse floristic composition and complex ecosystem. Malaysia is thought to have about 40,000 species which includes of 12,500 species flowering plants [1], about 1,200 (10%) species of medicinal plants. It is not only rich for plant biodiversity data also rich in fauna and pear swamp. Malaysia has been loosing much of its natural resources including plants and animal species through ecosystem and degeneration. Although the extensive research has been performed in plant biodiversity area recent years, managing plant biodiversity data with database system still poses many challenges.

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 a technical sense, a model is a medium to record the structure of an object in a more or less abstracted way, following predefined and documented rules. Collected data are structured which is represents a first order level of that data, for example the collection information stored on an insect specimen label. This can be redefined later to produce a formalized model. 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]. Models are the means that allows the description of real world phenomena. In this paper, we have discuss some of the issues involved in handling complex biodiversity data, spatio-temporal data and design of conceptual spatio-temporal data.

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. Figure 1 is an example of biodiversity data. More about biodiversity data describes bellow:

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.

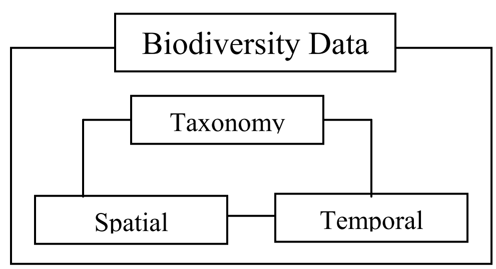

Figure 1 Biodiversity data organization

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. 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 developed.

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

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 2:

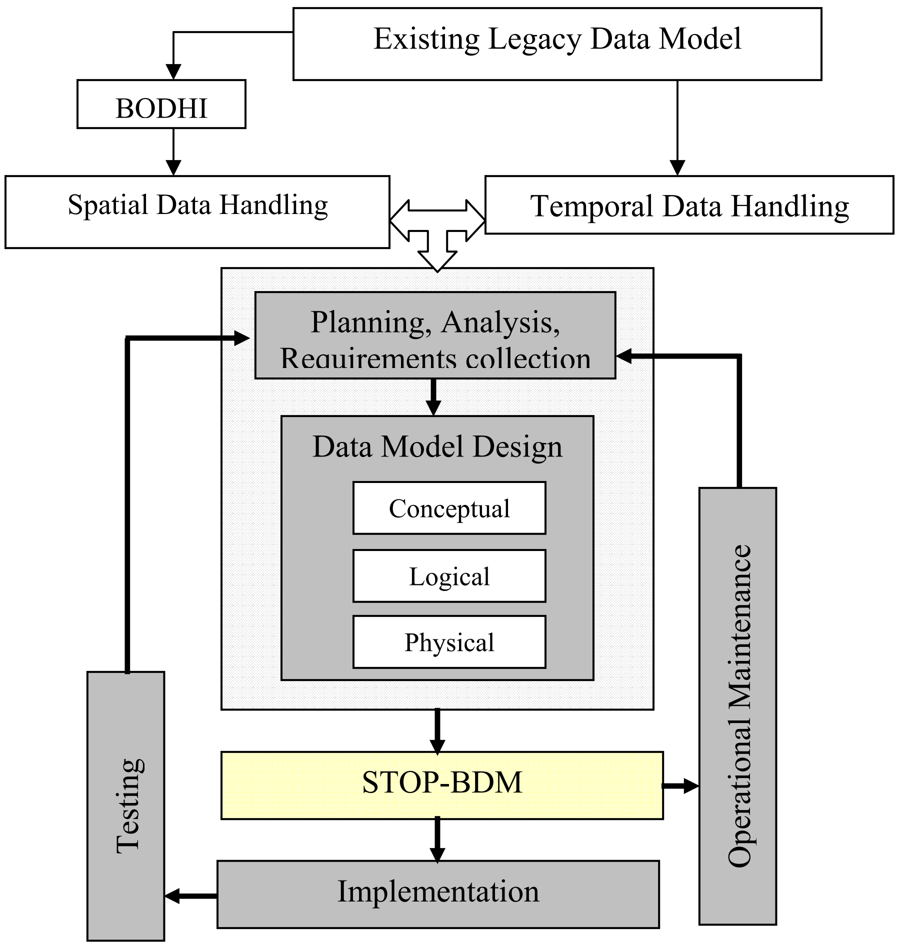

Figure 2 Data Model Life Cycle

4. 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 3 gives the class diagram of Spatial Data model of BODHI.

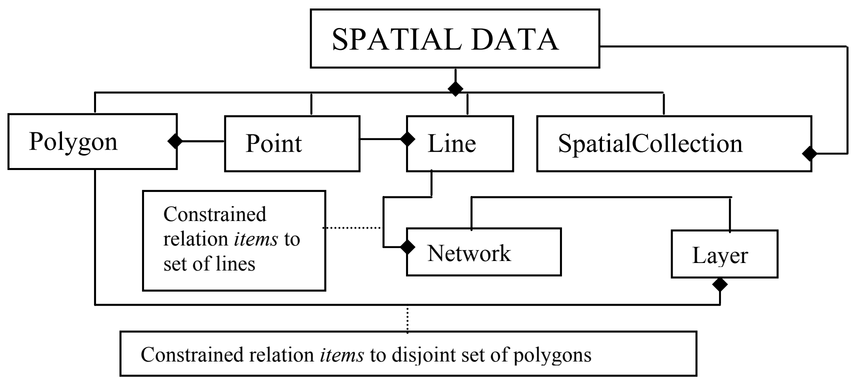

Figure 3: Class diagram of Spatial Data Model in BODHI

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:

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

5. Event Based Approach

The capability to maintain biodiversity data is a crucial requirement for biodiversity database management system. However, biodiversity data contains of spatial, temporal and attribute data. A temporal database has a time dimension and maintains time-varying data in contrast to a conventional database that carries only the current data. To support temporal data, system needs another model such as event-based model [10]. Pyramid system is a model to support multi-dimensional, spatio-temporal and geo-graphic 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 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 and time [11]. These features will be referred to accessed data from the database. This framework can be converted to table in stated figure: 4.

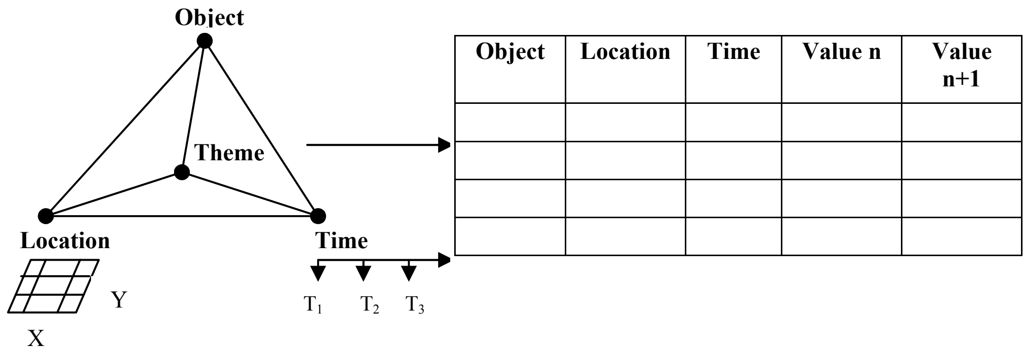

Figure 4: Transaction Pyramid Frame work to Table

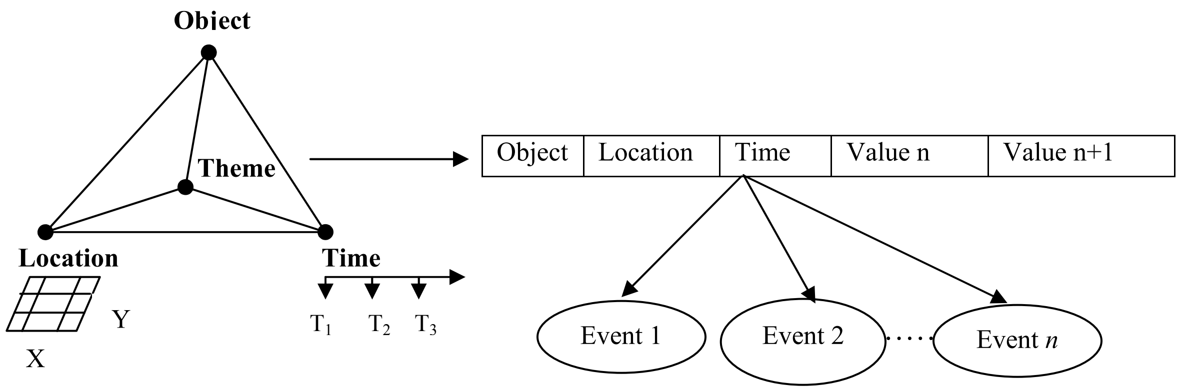

Figure 5: Modification based on event approach

Figure 5 illustrated pyramid framework that will modify, where component of time is classified by events. Events represent the changes that occurred in the real world. Our main reason of applying of this approach is to support time. As a conclusion, data retrieval can improved if we apply this framework table to our model also taking help from event based model.

6. 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.

7. 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]. 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 6. 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.

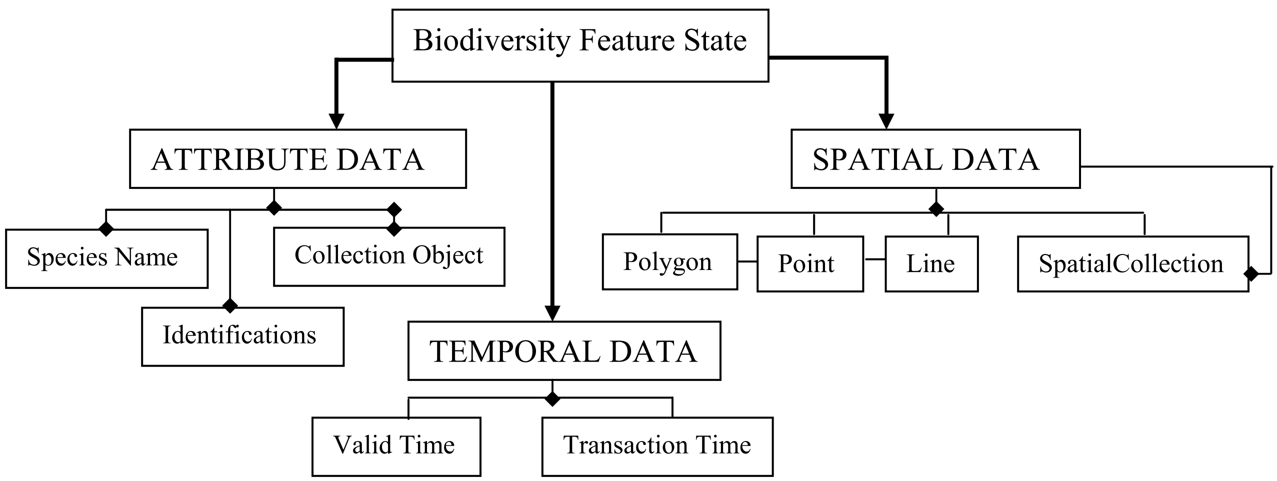

Figure 6: Overall conceptual of plant biodiversity data model

8. Conclusion

This paper provides an integrated conceptual model that partially overcomes some problems of spatio-temporal data model. 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