Spatio-Temporal Biodiversity Data Model Using Object Relational Modeling Approach

Mohd Taib Wahid[1], Abdur Rashid Mohammed Sharif[2], Suhaimi Napis[2], Noordin Ahmad[2]

[1]Faculty of Computer Science & Information Systems
Universiti Tekhnologi Malaysia, 81310 Skudai, Johor, Malaysia
Tel: (607)-5532422, Fax: (607) 5565044
taib@fsksm.utm.my

[2]Faculty of Engineering
Universiti Putra Malaysia, Serdang, Selangor, Malaysia
rashid@eng.upm.edu.my, suhaimi@putra.upm.edu.my, ndin@pc.jaring.my

Abstract

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 through database system. Nowadays, many applications need data modeling facilities for the description of complex objects with spatial and/or temporal facilities. Response to such requirements maybe found in Geographic Information Systems (GIS), in some other DBMS, or in other research area. A Geographical Information System (GIS) and its associated data sets are very powerful tools for mapping biodiversity and the intention of this paper is to demonstrate this using three important GIS functions: analysis, data integration, and data visualisation. Some others traditional database systems which are widely used for commercial applications but those models fail to meet biological application (biodiversity data). However, most of existing models cover only partly requirements either spatial or temporal modeling. Beside the commercial applications, there are few models have been designed for biodiversity data management (e.g., OSHADI, BODHI, APASD) but those models are relational model only handle information about biodiversity data does not support analysis, monitoring, manipulation and prediction. Integration between biodiversity data model and GIS data model can make enhancement for the biodiversity data model to make analysis, manipulation. This paper proposes object relational data modeling approach for spatio-temporal biodiversity data model.

Keywords: Biodiversity, Object relational, GIS, Spatio-temporal, data model

1. Introduction

Over the last few couple of years, there has been phenomenal growth in the area of plant bio-diversity studies, largely motivated by the economic and humanitarian potential that underlines the understanding of plant dynamics. New area has arisen which is called Bio-prospecting whose focus is solely on shifting through biodiversity data to locate potentially profitable biological sources [1].

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

Since the computer plays an important role in the solution of real world problems, it is almost impossible to find computer applications that do not have the time component as an important part of them (Zaniolo et al., 1997). Therefore, the lack of efficient management of time is a problem of real importance. In order to represent and operate upon real world phenomena, researchers in different areas, such as medicine, geography, or computer science, have been trying to find, during the last few years, the most suitable means to represent, store, manipulate, query, analyze and visualize data that change over time

Biological diversity plays a very 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.

A 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.

Geographical Information System (GIS) are becoming an increasingly popular tool for handling digital biodiversity data. GIS provides spatial data management, analysis, and display functions and facilitates the integration of various sources of data. GIS data models and analytical techniques are most appropriate for the representation and analysis of simple, spatially discrete, and temporally static entities, such as state, biodiversity area, road, and parcel boundaries. As biodiversity scientists have turned to GIS for the management and analysis of biodiversity data. GIS and its associated data sets are very powerful tools for mapping biodiversity data [2]. GIS deal with very complex phenomena which are often modeled as an interoperation of multiple spatial objects. In geology and other geo-sciences these spatial objects are intrinsically three dimensional and are characterized by extremely complex irregular shapes which require huge arrays of (point) data for their adequate representation. With in the context of GIS, geographic feature (also called entity or object) describes a part of geographic phenomenon. It has two data elements: attribute and geographic. Attributes are alphanumeric values that allow described characteristics of geographic features. The purpose of this research is to design a spatio-temporal biodiversity data model using object relational modeling approach.

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: There are various definitions on biodiversity has given by researchers. 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 or more specifically the availability and sustainability of high levels of biodiversity –is accorded great value under numerous perspectives. This definition captures the important concept that biodiversity is hierarchical in nature. Biodiversity manifests itself on the genetic species, ecosystem, and landscape levels. Interactions within and among levels all contribute to biodiversity. B*iodiversity* is the variety of different types of species that growth in period and various landscapes. 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 such as taxonomy data, geo-spatial data and temporal data.

2.1 Taxonomy Data

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 classify in 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 [11]. 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 as non identifying characteristics. The various characteristics on which this depends on these relationships may vary in time due to discover of new class characteristics, corrections to previously recorded characteristics etc.

As flower is composed of several parts namely petals, sepals, inflorescence carpals, ovaries, anthers, style and stigma. Each of these characters could be used for identification of a species and hence the yare also identifying characters. Similarly a leaf consist sparts such as blades sheath 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. A species may have synonyms the geographical location details and the textual in formation associated with it.

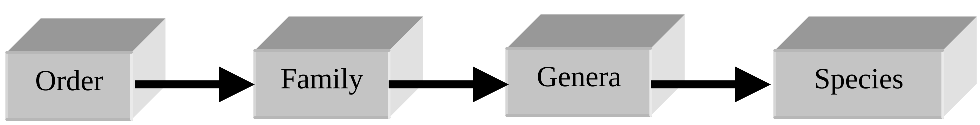


Figure 1: taxonomy of species

2.2 Geo-spatial Data

Geographical organizations are classified into four types – the *bio-geographic provinces*, *the biomes* (vegetative regions, e.g., evergreen, deciduous), *the administrative organizations* (e.g. district, state), and *the forest reserves*. All classification stated above alone with botanical gardens and herbariums, represents the geographical data and also regional map associated with it. The environmental conditions (e.g. soil, temperature, pressure, humidity) in these regions need to be recorded in two forms. One is spatial (i.e. regional) variation of the condition and the other is the temporal (i.e. over time) variation of the condition. The combination of two forms becomes spatio-temporal data.

Geographic data and taxonomic data had a rich inter-relationship. The distribution of species in a geographical organization is an example of this type relationship. Another instant is that, for all species found in a geographic region, the spatial and temporal variation in population have to be recorded.

2.3 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 plant species.

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 [12].

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%) [13]. 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), 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, 5]. Recently the academy of natural science of America developed a relational database and implemented for biodiversity [6]. 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].

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, and temporal data. Early 2000, biodiversity object database arcHItecture (BODHI) developed to handling plant taxonomies. To support spatial and temporal biodiversity data, one robust data model is required to develop.

The main purpose of this research is to develop a data model that better facilitates the exploration and analysis of biodiversity data to support temporal data. The goals of this research is to design a biodiversity data model which is combination of GIS 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.

5. Geographical Information System (GIS)

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, biodiversity data also geographical data should be in multi-dimensional mode and be interacted preferably from various sites.

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 (Figure 2); 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, forest area). Geographic data component can be classified into two main categories: Knowledge component and data component.

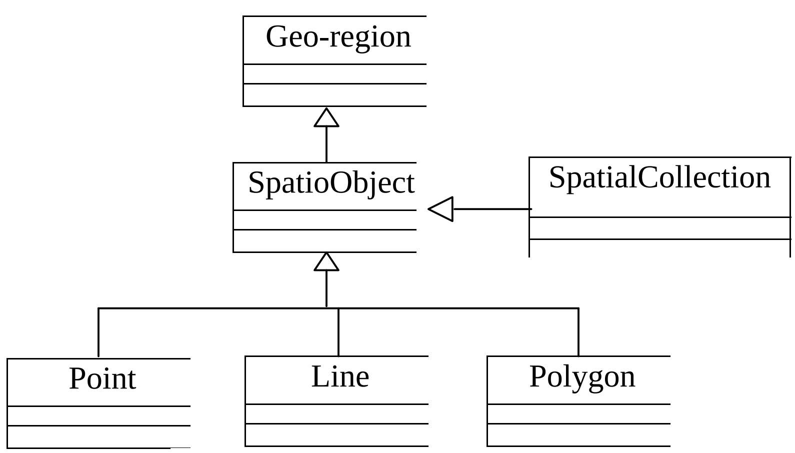


Figure 2: spatial data as object

5.1 GIS Components for Biodiversity Data

GIS components first designed and implemented for semantic GIS database model for the exploration spatio-temporal environmental data [14]. J Mennis, Donna J. Peuquet have designed conceptual framework which is also called pyramid framework. This frame work is interrelated with two separate parts one is data component and another is knowledge component (Figure 3). 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.

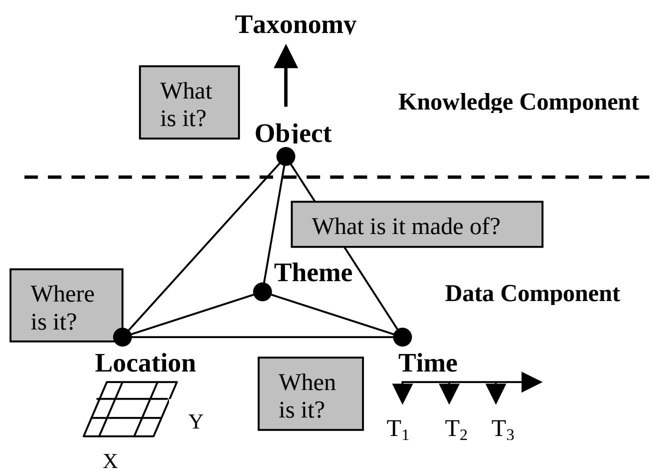


Figure 3: Pyramid Frame work: Data Component and Knowledge Component

Knowledge component stores information about higher-level semantic 'objects' the geographic entities or process that are describe by the data. Information concerning an 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 *time* 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.

6 Temporal Extension by 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 [15]. 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. More about pyramid framework describes in literature review. Pyramid framework consists of main three components; objects, location and time [16]. These features will be referred to accessed data from the database. This framework can be converted to table in stated figure 5.

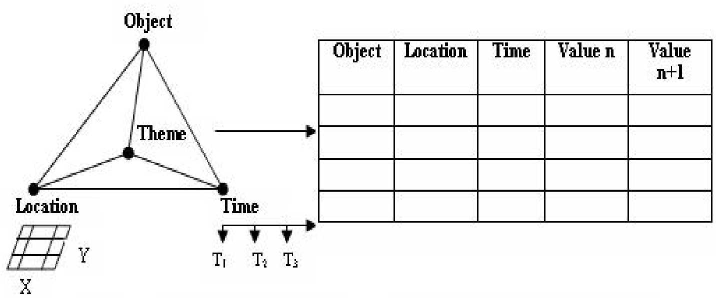


Figure 5: Transaction Pyramid Frame work to Table

Figure 6 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.

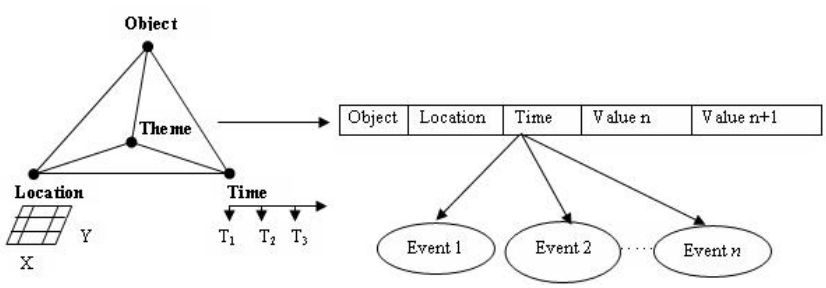
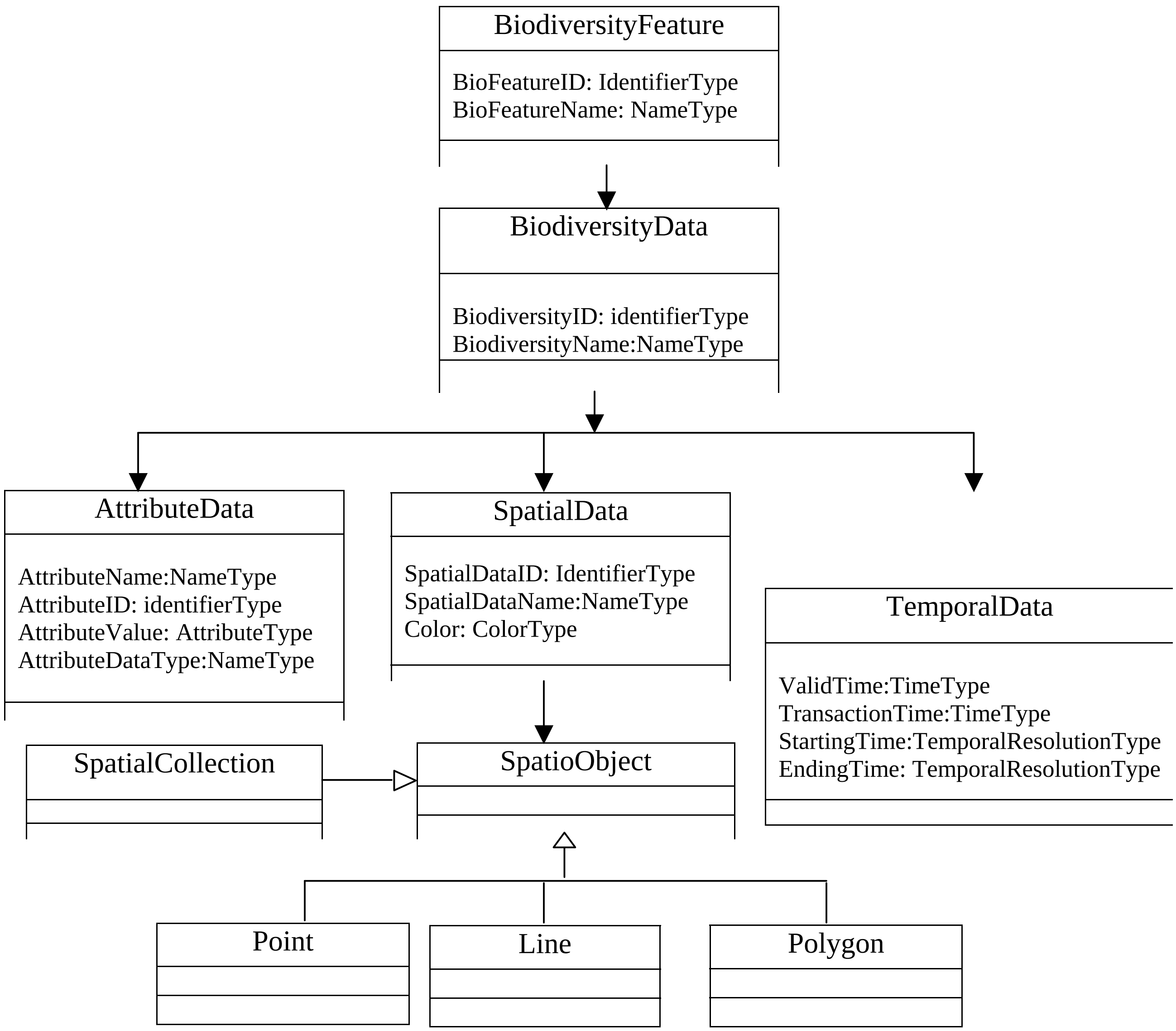


Figure 6: Modification based on event approach

6. 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 7. 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).



7. Conclusion

This chapter provides an integrated object-relational 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. This unification means the effectiveness both in function, consistency and computational complexity.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 biodiversity. Unlike overall design of spatio-temporal 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. This model is built up from the basic attributes and behaviors of spatio-temporal objects. Based on the object-oriented strategy, it's extensible to any applications by generalization, specialization, association and aggregation. This, this model can completely support the different semantics of space and time. Another Advantage of this model is no matter querying time or querying space, this model has the same efficiency. Not like temporal model or the spatial model.

References

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

[2] Lars Brabyn, (2004), "The use of GIS for Biodiversity Mappin in New Zealand", Department of Geograhy, University of Waikato, Hamilton, New Zealand.

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

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

[5] Srikant B.J, Jayant R. Haritsa, "Design and implementation of Biodiversity Information Management 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] P C Vasishta, (1974), "Taxonomy of Angiosperms", S Chandand Co. Delhi

[12] Thang, H.C. (2004). Forest Management and Development in Malaysia. Paper presented at the Malaysian Timber Council's Familiarisation Programme for European Trade Representatives, 12 July 2004, Kuala Lumpur, Malaysia. Unpublished. UNEP (2002) Malaysia-Study Area. http://www.rrcap. unep.org/lc/cd/ html/countryrep/Malaysia/studyarea.html

[13] 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.

[14] Jeremy L, Mernnis, Donna J. Peuquet, Diansheng Guo, (2000), "A Semantic GIS Database Model for the Exploration of Spatio-temporal Environmental Data", GIS/EM4, Alberta, Canada, September 2-8, 2000

[15] 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

[15] 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