

CONCEPTUAL DESIGN OF BIODIVERSITY DATA MODEL (BiDaM) USING OBJECT RELATIONAL AND EVENT BASED APPROACH

A.Z.M Kamruzzaman, Dr. Harihodin Selamat, Mohd Taib Wahid,
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
akzaman@siswa.utm.my, harihodin@itp.utm.my, taib@fsksm.utm.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 with database system. Plant biodiversity is the variety of different types of plants species that growth in various landscape. Various data models, query languages and techniques have been proposed by many researchers. Traditional database systems which are widely used for commercial applications but those models fail to meet biological application (biodiversity data). This paper discusses about the conceptual design, how spatio-temporal data can make enhancement plant biodiversity data model design for long term stewardship of biodiversity information. Plenty of data model has been developed (such as BODHI), which is only support spatial data. In addition to this, most of the data models are relational model. Geographical data have an object and location, attribute and time. Integration between plant Biodiversity data and geographical data with event-based approach can make an enhancement for the plant biodiversity data analysis and data manipulation. 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.

KEYWORDS

Biodiversity, Data Model, Object Relation, Event Based, GIS, BODHI

1. Introduction

Malaysia is undoubtedly on the richest mega biodiversity country 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 losing 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 [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, 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 has 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).

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: 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 such as taxonomy data, geo-spatial data and temporal 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 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), 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, and temporal data. 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.

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.

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.
- 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.

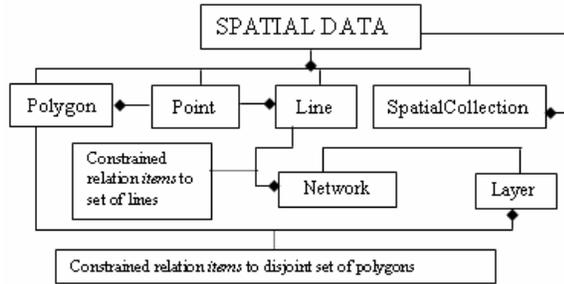


Figure 2: Class diagram of Spatial Data Model in BODHI

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:

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

5. 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:

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

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.

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

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:

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

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.

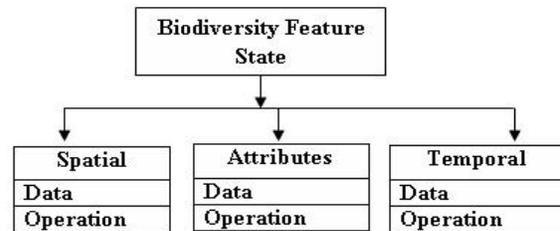
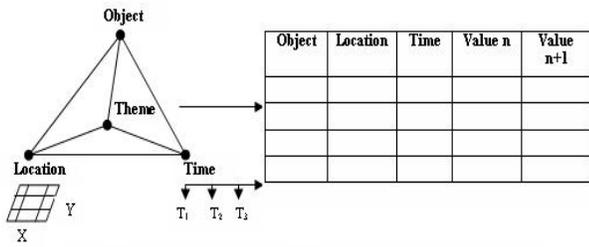


Figure 3: biodiversity feature state objects

6. 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. More about pyramid framework describes in literature review. 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.



ill

Figure 4: Transaction Pyramid Frame work to Table

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.

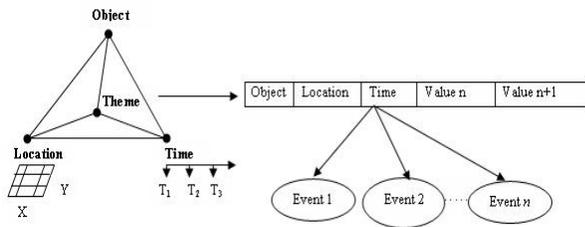


Figure 5: Modification based on event approach

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

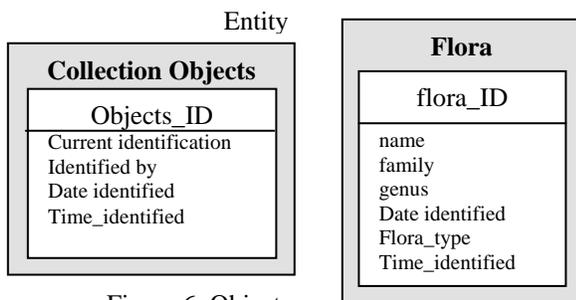


Figure 6: Objects

Relationship: 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.

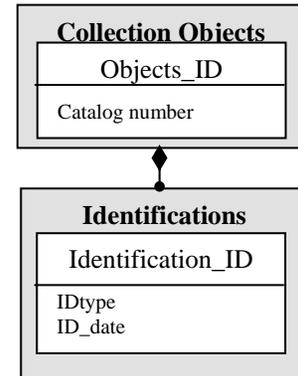


Figure 7: Identifies relationship

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”.

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 multivalued (with a multiset value);
- Mandatory (with a value in every instance) or optional (with a value in some instances and no value in others).

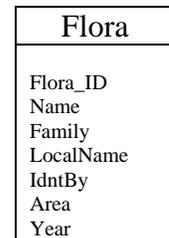


Figure 8: Entity and its Attributes

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

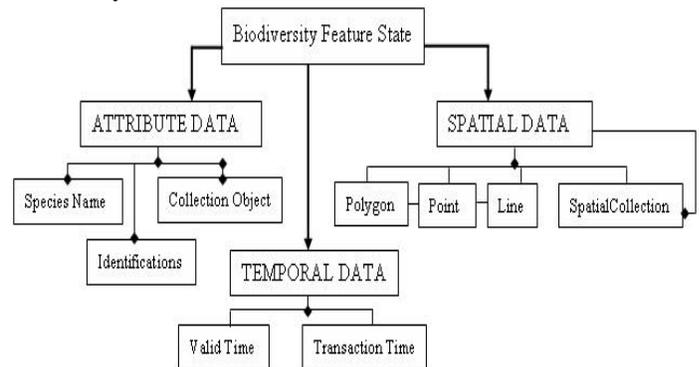


Figure 10: Overall conceptual data model of plant biodiversity

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 10. 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.

8. 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). After the conceptual model we have been designing logical and physical model to complete the enhancement on biodiversity data model. Once data model completed, implementation and testing will be done accordingly.

9. Acknowledgement

This document is a part of my master's project work about plant biodiversity data modeling. I would like to thank to Dr. Harihodin Selamar and Mohd Taib Wahid for their guidance and valuable suggestions. Beside that, I would like to thank to my examiner Dr. Nomie and PM Daut Daman.

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] Srikant B.J, Jayant R. Haritsa, Uday S. Sen "The Building of BODHI, a Biodiversity Database System", 2000
- [5] Paul J. Morris, "Relational Database Design and Implementation for Biodiversity Informatics", The Academy of Natural Sciences 1900 Ben Franklin Parkway, Philadelphia, 2005, PA 19103 USA
- [6] Suhani Napis, Kamarudin, Khairudin & A.Latif, "Biodiversity Database for Malaysian Flora and Fauna: Update" University Putra Malaysia, 2001
- [7] Mat Salleh, K. & A Latif., "Towards a comprised database of Malaysian medicinal and aromatic plants", 1997.
- [8] Joseph Maada Korsu Kendeh, "Geographical Information Systems for Rural Applications", international institute for geo-information science and earth observation, enschede, the Netherlands, March 2002,
- [9] Berendsohn, W. G., "The concept of "potential taxa" in databases", *Taxon* 44: 207-212. 1995. ISSN 0040-0262
- [10] 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
- [11] 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
- [12] Srikant B.J, Jayant R. Haritsa, "Design and implementation of Biodiversity Information Management System", 2000
- [13] Adnan Yunus, "Spatio-temporal object Relational Model for GIS", Dissertation, Universiti Teknologi Malaysia, Skudai Johor, 2005,