8-16-1996

Modeling Spatial and Temporal Semantics in a Large Heterogeneous GIS Database Environment †

Sudga Ram
Department of Management Information Systems, College of Business and Public Administration, University of Arizona

Jinsoo Park
Department of Management Information Systems, College of Business and Public Administration, University of Arizona

Modeling Spatial and Temporal Semantics in a Large Heterogeneous GIS Database Environment†

Sudha Ram
Jinsoo Park
Department of Management Information Systems
College of Business and Public Administration
430 McClelland Hall
University of Arizona
Tucson, AZ 85721

Problems with Current GIS Data Management

One of the major challenges in GIS and large scientific databases is heterogeneous data management in a distributed environment. Data tends to be collected and archived locally before being shared with the scientific community at large. Data items are complex because of different types (e.g., integers, real numbers, maps and images), different resolutions, and different temporal and spatial properties under different formats. For example, the School of Renewable Natural Resources at the University of Arizona and the State Lands Department have large datasets dealing with the topography of the Tucson basin. However, the data may have been gathered from different sources. Often the same data (or supposedly the same data) is not only called by two different names and stored in different systems, but also has different levels of precision, different temporal and spatial resolution, and different levels of detail.

Another problem is a fairly large semantic gap between current GIS data and users. For example, consider an analyst who wants to run a simulation which requires remotely sensed information related to Hurricane Andrew. To support such a query, the analyst would first have to find the date, latitude, and longitude of the eye of Hurricane Andrew, and then formulate a query to retrieve data whose dates and geographic references would include the path of the hurricane. This could be a fairly extensive task because the user would first have to locate the appropriate datasets, and then create a query to retrieve and compare data across these datasets. Such an inefficient and time-consuming process presents a major bottleneck to the analyst contemplating such a study. In an ideal situation, the researchers would not have to worry about where these databases reside, or how they are organized in order to use them.

A Semantic Modeling Approach

Two broad approaches to resolving GIS data heterogeneity can be identified in the literature. One approach extends relational database concepts to support geographic object types and queries (Goodchild, et al., 1992). However, such an approach cannot effectively handle heterogeneous GIS databases. The other approach is to use object-oriented (OO) models to represent geographic data (Berrill and Moon, 1992; Herring, 1992; Milne, et al., 1993; Worboys, et al., 1990). However, proposed OO database models for GIS are typically tailored to specific systems and do not make a clear distinction between physical and conceptual representations. Thus, they are difficult to understand and use from a GIS analyst's point of view. Since they do not distinguish between data and models, they cannot provide a clear classification of the semantics of the data and models.

Semantic models offer solutions to these limitations (Hammer and McLeod, 1981; Hull and King, 1987; Peckham and Maryanski, 1988; Ram, 1995). A semantic model (SM) is a collection of the concepts such as *vegetation*, *weather*, or *sites*, used to describe observations, along with the logical relationships that hold these concepts together. SM can explicitly represent information that is often hidden in an object-oriented schema. Examples of such information are: the cardinality of a relationship between object classes, or whether an entity class is strong or weak (Ram, 1991; 1995). Migration of objects between classes, and support for views (Abiteboul, et al., 1995) and integrity constraints (Morehouse, 1990) are better managed with a semantic model. Further, in our GIS environment, the spatial and temporal semantics of data need to be explicitly modeled and made available to the users so that model definition, configuration and analysis of data is facilitated.

Modeling GIS using USM*

In another paper (Ram, et al., 1996) we presented the definition of a semantic model called USM* to deal with the heterogeneity in spatial and temporal GIS databases. In this paper we summarize this model and describe the tools that can be developed using this semantic model. The USM* is an extension of the Unifying Semantic Model (Ram, 1995). The USM* incorporates constructs to model spatial, temporal, and dynamic classes of objects. The uniqueness of this model is in that it can capture the behavior of object classes rather than just the static properties. Graphical representations for each construct are shown in Figure 1. The formal definition of the extended constructs in USM* can be found in (Ram, et al., 1996).

Figure 2 shows an example of a USM* model for a natural resource management application. Such a model can be used as a front end to multiple heterogeneous GIS databases. To enable this we are developing a comprehensive data management system. The architecture of this system is shown in Figure 3 (Ram, et al., 1996).

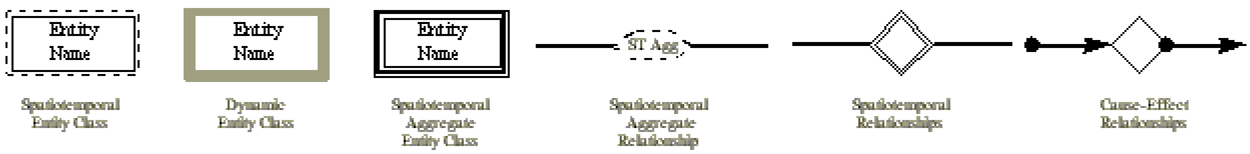

Figure 1. New Constructs in USM*.

Our system architecture is designed to help resolve some data heterogenieties using the mapping dictionary and the object-oriented (OO) data handler. The major components of the system are Data Modeler, Repository, and Data Accessor.

The Data Modeler: This component allows users/analysts to define a view of the underlying databases using the constructs in the USM*. It provides a graphical drawing palette to define entity classes, relationships, subclasses, and other information. It also allows users to explore the metadata associated with each entity class. For example, the user can browse through the properties of the entity class WEATHER and see its component classes (since WEATHER is a spatiotemporal aggregate). The data modeler also assists users in creating a correct USM* models, i.e. it points out errors and prompts for the correct sequence of modeling actions.

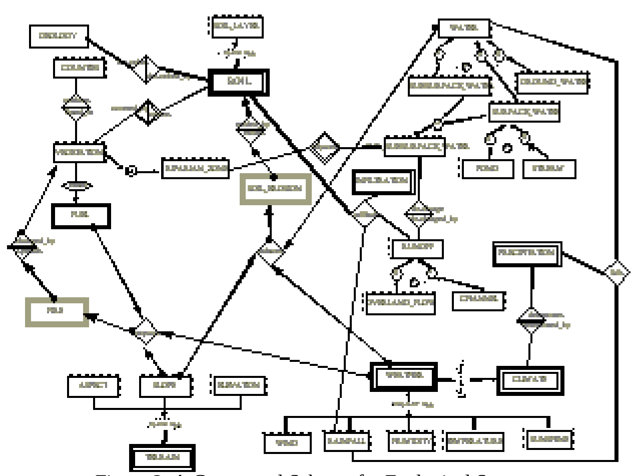
Figure 2. A Conceptual Schema for Ecological System.

The Repository: This component plays a very important role in resolving heterogenieties. It has three subcomponents (1) the USM* model created by users is stored in the metadata dictionary, (2) the metadata is mapped to each underlying database using the mapping dictionary, and, (3) a description of various simulation models is stored in the model descriptor. These three components interact with each other to effectively provide access to data from the underlying GIS databases.

The Data Accessor: This component uses the information provided in the metadata repository and extracts data from each database. It also extracts models from the simulation model base. It interacts with the underlying spatiotemporal databases through an OO data handler. The OO data handler provides data format independent interface for underlying heterogeneous data.

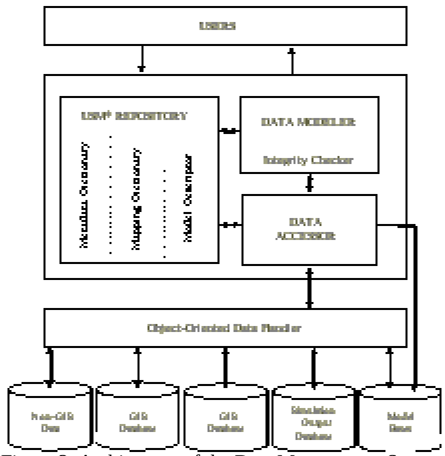

Figure 3. Architecture of the Data Management System

Conclusions and Future Research Direction

In this paper, we have presented a comprehensive data management system to access data in a heterogeneous GIS environment. These tools are being implemented on a network of IBM RISC System/6000 workstations running under the AIX Operating System. The programming language used to develop these tools is C; the graphical user interface is being developed using DB/UIMX. The repository has been implemented using Sybase system 10. The language to support the USM* is under development. It will not only support the definition and manipulation of spatial and temporal entities, but also provide event operators to support the behavior of dynamic entities.

Semantics-based data access using the USM* offers several advantages. First, users need not to be familiar with the contents of the actual data in advance. Users can use the USM* semantic interface as a browsing facility (Smith and Frank, 1990). The semantic interface allows analysts to interact with data in ways that are congruent with their normal ways of thinking about them. Since most GIS databases are extremely large as well as heterogeneous and distributed, it is impossible for the user to know all existing data, and their types and format. However, the semantic model can help locate pertinent data and models using an interactive dialog with the user. Finally, our model allows access to spatial and temporal data because it explicitly captures spatial and temporal entity classes and relationships. We believe our approach is applicable to other similar scientific databases as well, such as atmospheric sciences and space sciences.

While there are several topics for further work, our current focus is on resolving semantic heterogenieties in spatial databases. Few efforts have addressed semantic heterogeneity in geoscientific databases. Semantic heterogeneity exists between two or more entities which represent the same world, but have different semantics and interpretation (Ram and Ramesh, 1995). For example, one schema may use a spatial relationship *adjacent* to represent a spatial relationship between two counties (e.g. PIMA and YUMA in the State of Arizona). Another schema may define a *near* relationship between the same spatial entities because the person who designed the schema has a different perception of distance. These two schemas describes the same real world with different semantics. We are currently investigating techniques to identify and resolve *spatial semantic heterogeneity*. Another important area of work is in spatial schema evolution. Changes to underlying databases (i.e. content or schema) will trigger changes to the global or federated schema. These changes have to be incorporated in the global schema without affecting existing applications. Finally, semantic query processing in GIS databases is an unexplored area.

References

Abiteboul, S., Hull, R., and Vianu, V. *Foundations of Databases*. Addison-Wesley Publishing Company, Reading, MA, 1995.Berrill and Moon, 1992]

Berrill, A., and Moon, G. "An Object Oriented Approach to an Integrated GIS System," in *Proceedings of GIS/LIS '92*, November 10-12, 1992, pp. 59-63.Goodchild, et al., 1992]

Goodchild, M., Haining, R., and Wise, S. "Integrating GIS and Spatial Data Analysis: Problems and Possibilities," *International Journal of Geographical Information Systems* (6:5), 1992, pp. 407-423.Hammer and McLeod, 1981]

Hammer, M., and McLeod, D. "Database Description with SDM: A Semantic Database Model," *ACM Transactions on Database Systems* (6:3), September 1981, pp. 351-386.Herring, 1992]

Herring, J.R. "TIGRIS: A Data Model for an Object-Oriented Geographic Information System," *Computers & Geosciences* (18:4), 1992, pp. 443-452.Hull and King, 1987]

Hull, R., and King, R. "Semantic Database Modeling: Survey, Applications, and Research Issues," *ACM Computing Surveys* (19:3), September 1987, pp. 201-260.Milne, et al., 1993]

Milne, P., Milton, P., and Smith, J. "Geographical Object-Oriented Databases: A Case Study," *International Journal of Geographical Information Systems* (7:1), 1993, pp. 39-55.Morehouse, 1990]

Morehouse, S. "The Role of Semantics in Geographic Data Modelling," in *Proceedings of the 4th International Symposium on Spatial Data Handling*, Zurich, Switzerland, July 23-27 1990, pp. 689-698.Peckham and Maryanski, 1988]

Peckham, J., and Maryanski, F. "Semantic Data Models," *ACM Computer Surveys* (20:3), September 1988, pp. 153-189.Ram, 1991]

Ram, S. "Heterogeneous Distributed Database Systems," *IEEE Computer* (24:12), December 1991, pp. 7-12.Ram, 1995]

Ram, S. "Intelligent Database Design Using the Unifying Semantic Model," *Information and Management* (19), 1995, pp. 191-206.Ram

Ram, S, and Ramesh, V. "Schema Integration: Past, Current and Future," in *Heterogeneous Database Systems*, Sheth, A., et al. (eds.), Morgan Kaufman, 1995.Ram, et al., 1996]

Ram, S., Park, J., and Ball, G. "Semantic Model Support for Geographic Information Systems," Submitted for Journal Publication, 1996.Smith and Frank, 1990]

Smith, T.R., and Frank, A. "Report on Workshop on Very Large Spatial Databases," *Journal of Visual Languages and Computing* (1:3), 1990, pp. 291309.Worboys, et al., 1990]

Worboys, M., Hearnshaw, H., and Maguire, D. "Object-Oriented Data Modeling for Spatial Databases," *International Journal of Geographical Information Systems* (4:4), 1990, pp. 369-383.