
Geographic Information Systems and Problem Solving Environment

by [*Jinsoo Park*](#)

Introduction

Human activity over the last century has changed our natural landscape through construction and alteration of natural processes (e.g. fire suppression and dams). Incomplete understanding of ecosystem process interactions over large geographic regions impairs our ability to predict the consequences of natural resource management decisions. A lack of understanding how multi-level, large scale systems interact, and the inability to predict the consequences of management policies can be linked, in part, to the lack of tools for the management of heterogeneous data in large ecological datasets. The research reported in this article describes a semantic model and data management tools to help analyze ecological systems. Our research is part of a large multidisciplinary project aimed at developing the "Problem Solving Environment for Ecological System Analysis." The goal of this project is to develop a computer environment that enables ecological researchers to analyze and solve problems concerning real ecosystems at the level of complexity required of credible solutions for real world management concerns. The **Problem Solving Environment for Ecology System Analysis (PSE/ESA)** will enable researchers to:

- construct a representation of the physical environment,
- populate it with biological and physical process models,
- simulate ecosystem processes at high levels of resolution,
- observe the results of the simulation, and
- relate these results to real world observations.

The construction of this problem solving environment requires a state of the art system. It is being built on a high performance computing platform that integrates simulation with **Geographic Information System (GIS)** databases. The extended system is designed to:

1. support data/model management to enable access to a wealth of GIS data and models through an interface familiar to the users (ecosystem researchers or natural resource managers);
2. provide services to enable automated high performance optimization of model structures or policy parameters; and
3. provide advanced visualization services to enable automatic detection of events specified as significant or interesting to the user.

Such an environment is aimed at providing a platform for user-oriented development of spatially referenced, GIS-based, multi-process dynamic simulation models and the essential tools to employ such models to solve the problems of interest. The environment is expected to greatly reduce the effort required to build and test models by organizing data and models behind a high level interface that enables semantics-based access and reuse. The environment is currently being benchmarked on a challenging problem: wildfire management, which is a critical land use disaster prevention issue.

[Figure 1](#) illustrates the architecture of PSE/ESA. The outer layer consists of GIS databases, a visualization facility and a simulation engine. The middle layer provides components for model construction and archiving, model behavior interpretation, and optimization and search. These two layers provide the resources needed for problem solving. The users are at the center layer of the computational environment. Central to the PSE/ESA is an effective data management system that will facilitate interactions between the various layers shown in Figure 1. The middle layer, the core layer of PSE/ESA, consists of three components. The first component is the Data and Model Manager that provides support for model construction and archiving, with emphasis on semantics-based access to data and models, and multi-resolution model composition and configuration. The second component is the optimizer that facilitates effective and efficient search and optimization for model calibration and design/management alternative selection. It is based on a parallel, distributed multi-resolution strategy enabled by the Data and Model Manager. The third component is the Model Behavior Interpreter that detects user-defined significant events in simulation data, and controls the visualization display and simulation engine. This ensures that interesting behavior will not be overlooked in the massive stream of data.

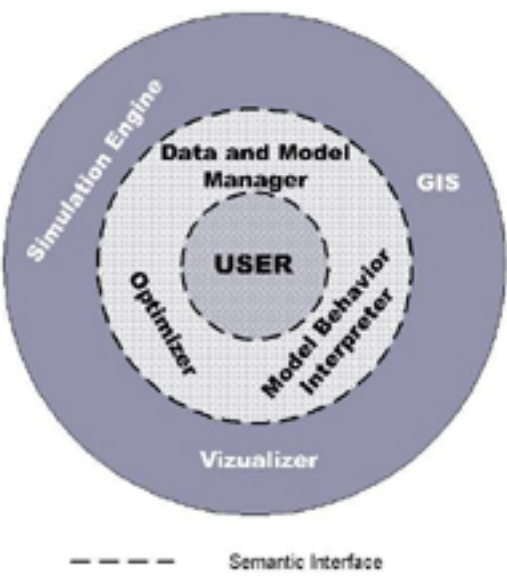


Figure 1. PSE/ESA Interaction of Users and the Ecological Simulation System.

The focus of this article is on the Data and Model Manager. One of the major challenges in this project (as well as in GIS and large scientific databases in general) 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 they are 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 (found online at <http://www.arizona.edu/academic/catalog/wsm.html>) at the University of Arizona and the State Lands Department (found at <http://www.land.state.az.us/asld/asldhome.html>) 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) are called not only by two different names and stored in different systems, but also have different levels of precision, different temporal and spatial resolution, and different levels of detail. The data from the School of Renewable Natural Resources may have higher resolution than data from the State Lands Department. In an ideal situation, the researchers would not have to worry how these databases are organized in order to use them.

Need For Semantic Models

Reusability of models is increasingly being recognized as an effective means for reducing the time and cost of model development. In ecological system analysis, a researcher starting from scratch would have to develop all the pieces before a useful model could be initiated. For example, the researcher first has to examine the contents of existing databases to determine whether there is sufficient relevant data for his/her research interest. This process is generally done before building a model. Such an inefficient and time-consuming process presents a major bottleneck to a researcher contemplating such a study. Ideally, the scientist should be able to describe a high-level need and have the PSE/ESA provide guidance on data and models that might satisfy this need. To reduce this preparatory burden, the PSE/ESA supports repository-based model development. Models that are developed for specific ecosystem processes can be archived in a model database for reuse later as the need arises.

Another challenge facing the ecological researcher interested in employing spatial data and reusing legacy models from a variety of sources is the lack of standards. The lack of standards to store and refer to data and models makes it difficult to integrate them into a current project. The difficulties discussed above for the shared reuse of physical data can be easily generalized to the reuse of the models that use that data. For example, the contextual semantics of "storm" include not only the information about how much rainfall the storm is carrying, or the storm's location, but also the various models that are used to create that storm in the simulated world.

As previously indicated, it is evident that the semantic gap between current data storage and retrieval technology and the needs of scientists is fairly large. For example, consider a researcher who wants to run a simulation which requires remotely sensed information related to Hurricane Andrew (for some fantastic images, see <http://www.nhc.noaa.gov/andrew.html>). To support such a query, the scientist 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.

Two broad approaches to this issue can be identified in the literature. The first approach extends

relational database concepts to support geographic object types and queries [4]. However, such an approach cannot effectively handle heterogeneous GIS databases. The second approach is to use object-oriented models to represent geographic data [7, 18, 19]. However, proposed **object-oriented (OO)** database models for GIS are typically tailored to specific systems and do not make a clear distinction between physical and representations. Thus, they are difficult to understand and use from an ecologist'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 [5, 8, 11]. 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. For example, hurricanes would be defined as a class of objects with properties such as area and time. This definition is then mapped to the physical representation of the actual data in one or more scientific datasets.

Even if **Object-Oriented Database Management Systems (OODBMSs)** are used to store data, it is advantageous to have a semantic model interface to the database(s). A semantic model 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. In addition, the semantics of complex entity classes such as aggregates can be identified explicitly in a semantic model and used for querying. This information is hidden in the "methods" or in the definition of the object in an object-oriented database (OODB). Migration of objects between classes and support for views and integrity constraints are better managed with a semantic model. Further, in our ecological environment, the spatial and temporal semantics of data need to be explicitly modeled and made available to users and to the different layers so that model definition, configuration and analysis of data is facilitated.

Modeling an Ecological System Using a Semantic Model USM*

In this section, I demonstrate how a semantic model, called **USM***, can be used to facilitate ecological system analysis. The **USM*** is an extended version of the **USM (Unifying Semantic Model)** [12]. The **USM*** serves as a formal specification mechanism for describing a Universe of Discourse (GIS data and model bases). It incorporates new constructs to support multidimensional objects and other types of spatial and temporal objects typically found in data and model bases which are relevant to GIS analysis. It also captures the behavior of dynamic and process-oriented spatial objects used for decision support. The formal definition of the extended constructs in **USM*** can be found in [13]. The schema developed by **USM*** provides a semantic interface between users (both experts and non-experts) and the various layers of [Figure 1](#). [Figure 2](#) shows a subset of the **USM*** for describing the semantics and interrelationships among spatial and temporal objects for a part of our ecological system. The spatial entity class **WATER** consists of three subclasses, **SURFACE_WATER**, **GROUND_WATER**, and **SUBSURFACE_WATER**. Each subclass inherits all attributes of **WATER**. The collection of these subclasses is totally exhaustive and every pair of these subclasses is mutually exclusive.

SURFACE_WATER is a superclass of POND and STREAM. Note that the USM* allows the definition of a generalization hierarchy that can be tailored to an individual user's point of view. Thus, different users may have different hierarchies. SUR_SUBSURFACE_WATER is the union of SUBSURFACE_WATER and SURFACE_WATER, which has an *adjacent* spatial relationship with RIPARIAN_ZONE. There is an Interaction entity class, INFILTRATION, among the entity classes SOIL, RAINFALL, and RUNOFF. Infiltration is the process by which water enters the soil surface. It consists of attributes, such as Inf_capacity and Inf_equation, to describe the interaction relationship among the three spatial entity classes.

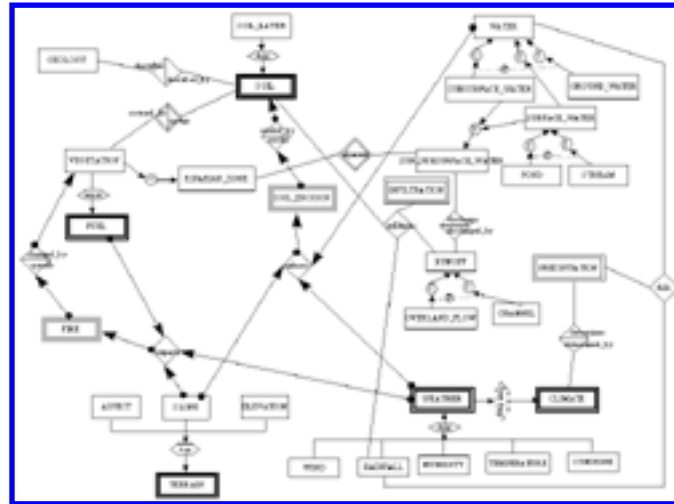


Figure 2. A Schema for Ecological System.

WEATHER is an aggregate of WIND, RAINFALL, HUMIDITY, TEMPERATURE, and SUNSHINE, each of which is a different entity class. Thus, WEATHER is heterogeneous. CLIMATE is a time-based aggregate of WEATHER over a period of time, which is affected by PRECIPITATION. WATER, SLOPE and WEATHER cause SOIL_EROSION which in turn affect SOIL. FUEL is a Composite entity class defined by listing its members from VEGETATION, that is, different types of vegetation provide different types of fuel. SLOPE, WEATHER and FUEL influence the behavior of FIRE. For example, a fire tends to burn uphill, and is affected by the direction of wind. In addition, the greater the slope and the faster the wind speed, the greater the fire intensity and the faster the spread of fire.

USM* is being implemented on a network of workstations running under the Microsoft Windows NT Operating System version 4.0. The programming languages used to develop the USM* and the graphical user interface are Java and C++. The data and model repository has been implemented using the Oracle relational database management system version 7.3 and the Microsoft SQL Server version 6.5. The language used to interface with 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.

A unique feature of our system is that these tools allow users to develop models in a collaborative research environment. Users can work on individual workstations to retrieve, view and modify the

schema and definitions of other designers simultaneously. Each authorized user logs onto the system with his/her user name. Every update of the definitions done by the user is recorded by the system. This group environment for building a schema using the USM* provides multiple views of the modeling activity [1, 6] at the different levels of abstractions and detail. For example, consider a researcher who is interested in building a fire simulation model. The researcher may choose to use one of thirteen pre-existing fuel models. Each fuel model has different parameter values and belongs to one of the following groups: grass, shrub, timber and logging slash. Alternatively, the user can change the default values provided by each fuel model to construct models to suit their own interests. The user can also construct (add, modify or delete) their own schema and simulation model from an existing model. The necessary information to build a fire simulation model may include fuel, fuel moisture, slope, wind speed and direction. Fire is a dynamic entity class because, as the fire moves across the landscape, the fuel composition and the terrain characteristics may change. Wind can also change in both speed and direction in relatively short time periods. Precipitation, time of day and aspect may affect the fuel moisture. [Figure 3](#) shows an example of fire simulation model derived from [Figure 2](#).

In [Figure 3](#), a new cause-effect relationship *affects* is added by the researcher. This reflects the fact that, in the fire simulation model, containment of FUEL, one of the affectors determining the characteristics of fire behavior, is affected by several entities, such as PRECIPITATION, ASPECT, and WEATHER at a specific time (specifically Time_of_day). In addition, temporal information (Time_of_day) determines the degree of moisture contained in the fuel, because humidity is highest during early morning and decreases during the day. Note that our software tools afford a lot of flexibility to the user. Also, as the system is used more and more, the USM* schema will evolve and become more comprehensive as new relationships and entity classes are added, updated or removed.

Semantics-based data access using the USM* offers several advantages. First, the user need not be familiar with the contents of the actual data in advance. The semantic interface allows researchers to interact with data in ways that are congruent to their normal ways of thinking about them. The semantic model helps locate pertinent data and models using an interactive dialog with the researcher. Having located the data and models that can serve as components for model development, the semantic model helps construct and configure larger models from these components. For example, the USM* toolkit can satisfy a request to invoke a particular fire simulation model and configure it for a specific geographic region without posing a lot of burden on the user.

Second, the schema helps those users who do not have a clear idea of the entire ecological system. For instance, a researcher may know what they want to investigate but has an incomplete idea of the data and models that are available and pertinent. Such a researcher can use the USM* semantic interface as a browsing facility [15]. Since most scientific databases are extremely large as well as heterogeneous and distributed, it is impossible for the user to know all existing data, data types and their format. Further, the user may not know how to access each dataset. Since the USM* tools provide a graphical user interface for the schema of the whole ecological system and allows automated support for multi-scale and multi-view modeling, users can navigate the schema and access the scaled-down datasets without specific knowledge about their locations and formats.

modeling. The predefined set of entity classes, relationships and built-in integrity constraints support a unified way to construct models by multiple users in a collaborative environment.

We plan to evaluate our tools by examining their actual use by a group of ecologists at the University of Arizona. Our experimental study will provide us with insights to expand and/or modify the constructs of the USM* as well as the toolkit.

While there are several topics for further work, our current focus is on resolving semantic heterogeneities in spatial databases. Semantic heterogeneity exists between two or more entities which represent the same world, but have different semantics and interpretation. For example, one schema may use the spatial relationship *near* to represent a distance relationship between two spatial entities, say TUCSON and PHOENIX (two cities in Arizona). Another schema may define a *far* relationship between the same spatial entities because the person who designed the schema had a different perception of distance. A third schema may describe the relationship by stating that TUCSON is located *south_of* PHOENIX. These three schemas describes the same world with different semantics. Many attempts have been made to identify and integrate semantic heterogeneity in heterogeneous and distributed database environments, such as multidatabases and federated databases [2, 3, 6, 9,10, 14]. However, few of these efforts have addressed semantic heterogeneity in geoscientific databases [16, 17]. We are currently investigating techniques to identify and resolve *spatial semantic heterogeneity*. Other topics of interest are schema evolution and techniques to assist in keeping track of changes in the schema without affecting existing applications, and semantic query processing in a spatial database environment.

Acknowledgements

I would like to give credit to Dr. Sudha Ram and Dr. George Ball. Both of them helped me write this article, and certainly deserve special thanks.

References

- 1 Alonso, G. and A.E. Abbsadi, Cooperative Modeling in Applied Geographic Research. *International Journal of Intelligent and Cooperative Information Systems* 3, 1 (1994), pp. 83-102.
- 2 Batini, C. and M. Lenzerini, A Methodology for Data Schema Integration in the Entity Relationship Model. *IEEE Transactions on Software Engineering SE-10*, 8 (1984), pp. 650-664.
- 3 Batini, C., M. Lenzerini, and S.B. Navathe, A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys* 18, 4 (1986), pp. 323-364.
- 4 Boursier, P. and M. Mainguenaud. Spatial Query Languages: Extended SQL vs. Visual Languages vs. Hypermaps. In *Proceedings of the 5th International Symposium on Spatial Data*

Handling(Aug. 3-7, Charleston, South Carolina). 1992, pp. 249-259.

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

6 Hayne, S. and S. Ram. Multi-User View Integration System (MUVIS): An Expert System for View Integration. In *Proceedings of the 6 th International Conference on Data Engineering* (Feb. 5-9, Los Angeles, CA). 1990, pp. 402-409.

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

8 Hull, R. and R. King, Semantic Database Modeling: Survey, Applications, and Research Issues. *ACM Computing Surveys* 19, 3 (1987), pp. 201-260.

9 Kim, W. and J. Seo, Classifying Schematic and Data Heterogeneity in Multidatabase Systems. *IEEE Computer* 24, 12 (1991), pp. 12-18.

10 Navathe, S.B. and S.G. Gadgil. A Methodology for View Integration in Logical Database Design. In *Proceedings of the 8th International Conference on Very Large Data Bases* (Mexico City, Mexico). 1982, pp. 406-416.

11 Peckham, J. and F. Maryanski, Semantic Data Models. *ACM Computing Surveys* 20, 3 (1988), pp. 153-189.

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

13 Ram, S., J. Park, and G. Ball, Semantic Model Support for Geographic Information Systems. *Submitted for Journal Publication*, 1996.

14 Ram, S. and V. Ramesh, Schema Integration: Past, Current and Future, In *Heterogeneous Distributed Databases*, A. Elmagarmid, M. Rusinkeiwicz, and A.P. Sheth, (Eds.). Morgan Kaufman, 1996.

15 Smith, T.R. and A.U. Frank, Report on Workshop on Very Large Spatial Databases. *Journal of Visual Languages and Computing* 1, 3 (1990), pp. 291-309.

16 Stonebraker, M., Sequoia 2000: A Reflection on the First Three Years. *IEEE Computational Science & Engineering*, (Winter 1994), pp. 63-72.

17 Stonebraker, M., et al. Tioga: Providing Data Management Support for Scientific Visualization Applications. In *Proceedings of the 19th International Conference on Very Large Data Bases*

(Aug., Dublin, Ireland). 1993, pp. 25-38.

18

Wiegand, N. and T.M. Adams, Using Object-Oriented Database Management for Feature-Based Geographic Information Systems. *Journal of the Urban and Regional Information Systems Association* 6, 1 (1994), pp. 21-36.

19

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

Jinsoo Park received his M.B.A. and M.S. in MIS from the University of Pittsburgh, and is currently a Ph.D. candidate in MIS at the University of Arizona. His research interests include Geographic Information Systems (GIS) and Advanced Database Management, such as spatial databases and heterogeneous database integration. His current focus is to develop technologies for integrating heterogeneous GISs with databases and a methodology for spatial schema integration.