4.4

DATA MODELING FOR EMERGENCIES

MICHAEL F. GOODCHILD

AS A PROCESS, data modeling is often so implicit that its significance is hardly recognized. Whenever observations are made about the world and assembled in some framework, such as a table or a collection of marks on a sheet of paper, the framework constitutes a data model. A table, for example, provides a framework in the form of a collection of cells whose values can be inserted from field observations. In turn, the table provides data for analysis, which ultimately conditions and constraints the types of models and theories that can be developed as contributions to science or to practical problems. Computer databases require data models to be explicit, especially if users are to take advantage of functions related to specific data models. For example, if data are to be assembled into the framework of a table by inserting numerical values into the table's cells, then in addition to the framework itself (the basic table), functions can be provided in advance for routine table functions, such as totaling columns or printing. Microsoft's Excel represents a software environment built on this simple model of data assembled in a table, whereas Word's implicit data model is a linear stream of text. This paper describes the significance of data modeling in the context of emergencies.

GEOSPATIAL DATA MODELING

Geospatial data modeling tends to be comparatively complex relative to such applications as Excel and Word. This is due to the wide range of models in use, such as those based on rasters or vectors, and those focused on georelational or object-oriented models (for a basic introduction to GIS see Longley et al. 2001; for a review of GIS data modeling see Worboys 1995).

The earliest GIS, such as the Canada Geographic Information System constructed in the mid-1960s (Foresman 1998), developed their own unique approaches to the handling of data that were specifically adapted to the needs of geospatial applications. Two types of data models emerged during this period: vector models, in which all features on the landscape are modeled as points, lines, or areas; and raster models, in which all variation on the landscape is expressed in terms of attributes of regularly shaped rectangular cells in fixed locations. A highly specialized adaptation of the vector approach, the topological data model, became popular during the 1970s because of its high level of internal consistency and hence advantages in quality control, and its straightforward application to the representation of a wide range of geographic themes, including land ownership, political boundaries, land cover, and land use. Rather than focusing on the areas on such maps, the topological model treats each common boundary between two areas as its basic unit, and records its location as a series of coordinate pairs, together with the identities of the areas on each side of the common boundary. The properties of areas are stored in a separate data structure as a table. Thus the model has two distinct types of data elements: the properties of areas in a tabular structure, and common boundaries as a set of digitized lines, with varying numbers of coordinate pairs. In contrast, the raster model records variation as an ordered sequence of values, each corresponding to the value of a single and consistent property in a cell. Order is sufficient to establish the geographic location of each cell, and the entire raster is registered to the Earth's coordinate system at its corner points.

Neither of these early data models has much in common with the needs of other computer applications, except perhaps for the similarity between GIS rasters and digitized images, so there was little to be gained by adopting common approaches. This situation changed dramatically in the 1980s, however, with the computing industry's widespread adoption of the relational data model (Date 1975). In this framework, all information is expressed as a series of tables. Each table provides the characteristics, in its columns, of a series of similar objects, each object occupying one row. Tables are linked by *keys*, which allow information in one table (for example, characteristics of patients) to be linked to information in another table (such as characteristics of doctors; one of the properties of a patient would be the identity of a doctor).

The popularity of the relational model caught the attention of GIS designers, who recognized the advantages of using this framework to store the tables of information about areas in the topological data model. By doing so, they could link areas to tables of cartographic symbols, for example, or to tables of information about other features on the landscape. Note, however, that the relational model was suitable only for the area attribute tables; the definitions of common boundaries contain variable numbers of coordinate pairs, and could not be fitted simply into the relational structure's rectangular tables. ARC/INFO, the GIS from Environmental Systems Research Institute that appeared in the early 1980s, adopted the relational model of the INFO software, but coupled it with specialized software for handling the common boundaries arcs) of the topological data model, in what became known as the hybrid approach (Burrough 1986). The relational model was not widely adopted as a framework for raster data. It also was not as useful for representing the hierarchical relationships that commonly exist between geographic features, such as the relationship between the runways, hangars, terminals, and other component parts of an airport and the airport itself; or between counties and their parent state. Thus, the adoption of the relational model was, at best, a partial solution to the needs of geospatial data; not all geospatial data fits easily into its tabular framework.

A more satisfactory solution became available with the object-oriented data model (Zeiler 1999). In this model all objects are instances of classes, and classes can be specializations of more general classes. For example, State Street is an instance of the class *streets*, and streets are a specialized class of transportation link. The object-oriented model also allows the coordinates defining the outline of an area to be stored as merely one additional attribute of an area, avoiding the need for the somewhat awkward hybrid model of early ARC/INFO. Object-oriented data models, implemented as special interfaces to relational database management software, are now the standard for GIS. But they still leave some problems unresolved; perhaps most importantly, there are many examples of geographic phenomena that do not naturally fit the concept of an object. Many phenomena are conceived instead as continuous *fields*, or functions of location, as is the case for elevation, or atmospheric temperature, or soil moisture content. Fields do not naturally fit a model of geographic phenomena as discrete objects littering an otherwise empty space, and are much more closely aligned with the raster model.

There are obvious advantages to being able to model geographic phenomena, and to do so within the framework of a data model developed for a wide range of computer applications. While some types of geospatial information fit the popular relational and object-oriented models well, others do not. But for those that do, the advantages are clear. Computer-assisted software engineering (CASE) tools are available to facilitate the design of data models, and to automate their implementation.

Software, in the form of database management systems such as Oracle or Access, are available cheaply and off-the-shelf, relieving the GIS programmer of the need to handle many routine data input, output, and housekeeping operations. In principle, then, GIS software development is much simpler today than it was thirty years ago.

ESRI's ArcGIS is a good example of the contemporary implementation of object-oriented data modeling in GIS. A design is first developed in a convenient graphics design package such as Microsoft's Visio. Each class of objects is represented graphically as a box, with its name, attributes, and any methods or functions closely associated with the class, using the notational standards of the Unified Modeling Language (UML). Various types of relationships between classes are represented symbolically. The graphic nature of the process makes for easy participation by users, managers, and others associated with the application. When the design is complete, a software wizard is used to build an ArcGIS Geodatabase, with all of the specified tables and links. The database is then populated from a variety of sources using data loading software.

EMERGENCY RESPONSE DATA MODELS

A comprehensive data model for a large geospatial application, such as responding to emergencies or managing the distributed facilities of a utility company, must encompass a great variety of features, capturing their characteristics and locations on the landscape. A data model for the World Trade Center (WTC) response, for example, would have to represent the locations of buildings, streets, and underground utility lines, and at a much more detailed scale, the insides of buildings and the locations of workers. In fact, this data model would need to represent any features of relevance to the response operation. The full data model for this type of complex application might contain hundreds or thousands of distinct types of information.

In the past, such data models were constructed and populated with data ad hoc in the immediate aftermath of the emergency, often without a comprehensive design. Indeed, data models tended to evolve as data sets were acquired. There are great advantages, however, to planning and constructing data models in advance. First, a data model can be largely independent of the area to which it is applied, so a data model constructed for a response at the WTC in Lower Manhattan could be equally useful in a similar situation anywhere in the nation. Thus there are strong economies of scale in data modeling. Second, an existing data model provides a framework into which data can be inserted quickly, at minimal effort, and using standardized procedures.

Such data loading procedures can be organized in advance to provide consistency checks and other quality control mechanisms. Planning in advance for emergencies by developing data models thus conveys enormous benefits in a response. Third, by organizing a data model in advance, it is possible to develop the associated functions that are needed by a geographic information system (GIS) within an emergency operations center. A GIS can be up and running much more quickly if a comprehensive data model and associated functions have been designed in advance. Finally, there is the potential for learning from past emergencies, by analyzing the data models that evolved in a specific response, and refining them for future responses in cooperation with communities of users and decision makers.

CONCLUSION

In an emergency response situation, it is crucial that GIS capabilities be available as quickly as possible. These include both the database needed to support decisions and also the procedures needed to analyze data and to present information to decision makers. I have argued in this paper that contemporary approaches to GIS database design, including visual database layout, object-oriented modeling, and semi-automated database creation and data loading, are very significant improvements over earlier approaches. They allow data models for specific purposes to be developed well in advance of applications, and to be populated rapidly.

In the past few years, much effort has gone into developing essential data model designs using these techniques for specific GIS application domains. In the case of ESRI, many of these are available on the company's web site (http://www.esri.com). Researchers at the University of California, Santa Barbara, have led the development of one of these, a data model for transportation applications of GIS known as UNETRANS. The development of the model involved extensive discussions over a period of two years with users in the transportation field. A similar effort conducted within the community concerned with emergency response is warranted. Making use of the experience of events such as the World Trade Center disaster, and the construction and use of a GIS database in its immediate aftermath, would ensure that emergency response employs the best of contemporary GIS database design techniques, and would do much to speed the response to similar events in the future.