SPECULATIONS ON SEAMLESS, SCALELESS CARTOGRAPHIC DATA BASES

Stephen C. Guptill
U.S. Geological Survey
521 National Center
Reston, Virginia 22092

ABSTRACT

The idea of a seamless, scaleless data base of digital spatial data for use in automated cartography or geographic information systems has intrigued researchers for a number of years. Yet reviewing the plans of mapping agencies in the United States and Europe for their digital data bases in the 1990's shows partitioning by scale, space, and time. Why is this so? Is a seamless, scaleless data base still a quixotic quest, or have the conceptual models of spatial data and computer technology advanced to the point that such a goal is achievable? This paper contends that the impediments are now pragmatic concerns and not technological ones.

INTRODUCTION

The first uses of digital cartographic data were primarily geared toward the automation of the traditional map drafting process (Tomlinson, 1988; Rhind, 1988). This concept quickly expanded to the notion of interactive roaming through large cartographic data bases unhindered by map sheet boundaries or scale of presentation (Radlinski, 1974; Fields, 1978). Trailing these ideas was the actual creation of digital cartographic data bases and computer software to manipulate the data.

The first systematic collection of digital cartographic data sets began in the 1960's with the Canada Geographic Information System and was followed by efforts in the United Kingdom at the Experimental Cartography Unit, the Ordnance Survey, and the Military Survey, and in the United States by the Central Intelligence Agency, the Defense Mapping Agency, the Bureau of the Census, and the Geological Survey. Data manipulation software was largely written by the data-producing agency or academia, although some commercial software was available in the late 1960's and early 1970's.

Over the subsequent two decades, digital cartographic data has become much more prevalent, and capabilities of geographic information system (GIS) software to manipulate these data have increased dramatically. However, the concept of browsing through a seamless, scaleless data base has yet to be realized.

Publication authorized by the Director, U.S. Geological Survey

A seamless data base implies an ability to query, display, retrieve, or otherwise traverse the contents of a large spatial data base without limitations imposed by the spatial extent of the data. For example, a command to display the Mississippi River would yield the entire river, not just a portion of it. A scaleless data base implies an ability to transition from one level of detail to another appropriate to the scale of the display or precision of the data analysis. For example, select a feature, say Dulles Airport, and display its location with a point symbol at a scale of 1:2,000,000, zoom in on the airport, with runways appearing at a display scale of about 1:100,000, then, as the display scale increases, more detail, such as buildings, fuel tanks, and parking lots will appear.

Achieving these capabilities requires advances on two fronts: (1) availability of spatial data amenable to this environment, and (2) capacity of spatial data base management systems (DBMS) or GIS's to manipulate the data. Data producers must create data that can be easily placed into a seamless, scaleless data base environment. Data base system developers must then provide the tools necessary for handling the data.

DATA PRODUCERS

As noted above, national mapping agencies, including the U.S. Geological Survey have been producing digital cartographic data for over a decade. The data structures and underlying data models (some utilizing topology) used in collecting this information, for the most part, have not changed since they were developed in the late 1960's or early 1970's. However, during this time, the tasks for which the data were used became increasingly sophisticated, placing information demands on the data that were not planned for in their initial design.

In response, over the last several years, various mapping agencies have been designing data models for use in their current or future information systems. Much of this work is still ongoing and few references to the published literature are available. In those cases without published references, either agency representatives or internal reports are cited. The trend among these agencies is a data model built on a basic set of topological elements and superimposed with a set of cartographic features. This type of design may ease the construction of scaleless data bases. What is less clear is that national mapping agencies are committed to building seamless, scaleless data bases.

U.S. Geological Survey

In the case of the U.S. Geological Survey, one of the sources of the demands for change arises from the Survey's National Mapping Division, which is undertaking a major system development activity called Mark II. Mark II will be a digital cartographic production system with the National Digital Cartographic Data Base (NDCDB) at its hub. Information in the data base will reflect the data content of the National Mapping Program's standard map series. This information will be periodically revised and new graphic products generated using computer-assisted cartographic methods. Maintaining the information required to support these processes is a driving force behind the design of a more comprehensive data model (Morrison and others, 1987). Additionally, the growing sophistication of GIS's and the increasing diversity of applications involving GIS technology and spatial data are beginning to demand a more flexible and comprehensive model for spatial information. This linkage between advancing GIS capabilities and the need for more advanced data structures has been explored by Goodchild (1987).

Responding to these demands, the Geological Survey (as both a data supplier and data user) has begun the design of an enhanced version of the digital line graph, termed Digital Line Graph - Enhanced (DLG-E). In simple terms, the DLG-E begins with the topological model used in the Survey's present DLG format (U.S. Geological Survey, 1986) and builds a cartographic feature layer upon the topology. The feature definition is open-ended, allowing users to define additional features of interest. Cartographic entities will be described using objects, attributes, and relationships. Additionally, recommendations contained in the "Proposed Standard for Digital Cartographic Data" (<u>The American Cartographer</u>, 1988) regarding data quality information and formatting will be followed. Details on the DLG-E design are given "Designing an Enhanced Digital Line Graph" (Guptill, and others, 1988).

However, the contents of the NDCDB are envisioned as neither seamless or scaleless. The archival portion of the NDCDB will be partitioned by series, quadrangles, and categories (Guptill, 1986). These partitions are defined as:

> Series: a partition by data content (imagery, elevation matricies, cartographic data) or scale (1:24,000, 1:100,000, 1:2,000,000); interseries topological consistency is not required.
>
> Quadrangles: partitions along latitude, longitude boundaries; matching across boundaries is required.
>
> Categories: a logical subdivision of a series into classes of related data (transportation, hydrography); intercategory topological consistency is required.

The main unit of data collection and revision will remain the map quadrangle. The data are scale dependent, although some thought has been given to creating 2-3 times reduction products (such as 1:50,000-scale map graphics from 1:24,000-scale source or 1:250,000-scale map graphics from 1:100,000-scale source).

U.S. Bureau of the Census

The U.S. Bureau of the Census has developed the Topologically Integrated Geographic Encoding and Referencing (TIGER) system for its use in automating the geographic support system required for the 1990 Decennial Census. The geographic data contained in TIGER consists of a set topological elements with a set of feature directories and lists. The topological elements are represented by 0-, 1-, and 2- cells. Feature lists (containing items such as landmarks, road names, and county names) reference the appropriate set of topological elements that make up the features (Marx, 1986; Kinnear, 1987). Spatial partitioning for the TIGER data base corresponds to county boundaries. Within a county, the data are seamless with all categories of data "vertically integrated" into one planar graph. In addition, more detailed data covering the urban areas has been spliced into the TIGER data base replacing (along 7.5-minute quadrangle boundaries) the data from the 1:100,000-scale maps (Marx, 1987).

U.S. Defense Mapping Agency

As part of its Mark 90 modernization effort, the U.S. Defense Mapping Agency has developed a new data structure called MINITOPO (also referred to as the Advanced Mapping, Charting and Geodesy format). The data model underlying the MINITOPO structure consists of the following elements: nodes, edges, faces, point feature components, line feature components, area feature components, and features. The nodes, edges, and faces correspond to 0-, 1-, and 2-dimensional topological objects. The point, line, and area feature components are groupings of topological objects of the same dimension. Features are collections of feature components or of other features. Attributes are associated with either the features or feature components. The data base must support a wide variety of products covering various geographic extents across the globe. To support this diversity, the cartographer is presented with a seamless, scaleless window into the data base, although some physical partitioning based on geographic extent and scale is incorporated into the design of the data base archive (C. Kottman, oral commun., 1988).

Institut Géographique National

The Institut Géographique National, France, is in the process of creating two digital cartographic data bases: one with data commensurate with 1:25,000-scale mapping, and one with data commensurate with 1:100,000-scale mapping. The model used in these data bases consists of a set of "elementary objects" corresponding to topological elements and a set of "complex objects" made up of the elementary objects. Descriptive information is associated with the complex objects. The data bases are scale dependent, although, like the Geological Survey case, conceived to support the production of generalized output products with scale reduction factors of two to five. In lieu of an external spatial indexing scheme, the data base is physically partitioned into sets corresponding to a 1-2 km grid (Benard and Piquet-Pellorce, 1986; Salgé and Piquet-Pellorce, 1986; and Salgé, oral commun., 1988).

Landesvermessungamt Nordrhein-Westfallen

The Landesvermessungamt Nordrhein-Westfallen (Surveying and Mapping Agency, North Rhine-Westfalia) is participating in the design of both a digital cadastral map data base (Automatisiertes Liegenschaftskataster - ALK) and a Digital Land Model (Automatisiertes Topographisch-Kartographisches Informationssystem - ATKIS). These activities in the North Rhine-Westfalia region are part of a nationwide project to create the Official Topographic-Cartographic Information System. ATKIS will contain stratified data sets with information appropriate for mapping at scales from 1:5,000 to 1:1,000,000. The ATKIS data model consists of objects that were classified into "point-shaped, line-shaped and area-shaped objects" and are further characterized with attributes. An object has pointers to other objects and to object parts. For each object several attributes of different attribute types and references of different types to other objects are defined (Barwinski and Brüggemann, 1986; Brüggemann, written commun., 1987).

Data Collection Strategies

The results of this brief survey are rather mixed. Those agencies with strong traditions of quad-based standardized mapping (the civilian mapping agencies in the United States, France, and the German Federal Republic) have continued this practice in the construction of their digital data bases. However, the military mapping agencies are supporting a wider range of products over a much greater territory and appear to have adopted a more flexible approach to accomplish their mission. The Census Bureau is in a unique position, acting more like a data user than a producer, modifying and adding value to the base category digital data supplied by the Geological Survey. The Census Bureau has a seamless, scaleless (or at least multiresolution) data base over the areas (counties) of major importance to its mission. The county partitions were the most suitable for Census operations. The conceptual design of TIGER would allow the aggregation of counties to form State or National data bases if desireable, although computer limitations at the time did not allow this aggregation to be implemented (F. Broome, oral commun., 1989).

ENABLING TECHNOLOGY

Seamless, scaleless spatial data bases reflect a nontraditional view of cartography: "cartography as an information transfer process that is centered about a spatial data base which can be considered, in itself, a multifaceted model of geographic reality" (Guptill and Starr, 1984). Spatial data models based on the concept of a set of feature objects that represent aspects of the real world, that is geographic reality, provide a logical framework for seamless, scaleless data bases. But how can these concepts be implemented within a data-base context? Several researchers and commercial firms have implemented spatial data bases within a commercial relational data base. These include the GEOVIEW project (Waugh and Healey, 1987), SYSTEM 9 (Schuch, 1988), and GeoVision (Madill, 1987). However, each of these implementations has had to work around the limited set of data types and operations supported by existing relational DBMS's.

In recent years, computer scientists have sought to extend the capabilities of data base systems, creating a class of "extensible" DBMS's. These systems include university research systems such as EXODUS (Cary and others, 1986) and POSTGRES (Stonebraker and Rowe, 1986), as well as several commercial systems. Several features of extensible DBMS's are of potential use in the implementation of seamless, scaleless spatial data bases. For example, of particular interest are the following design goals of POSTGRES: better support for complex objects; user extendibility for data types, operators, and access methods; and facilities for active data bases (such as alerters and triggers) and inferencing including forward- and backward-chaining.

Using the DLG-E data model and the POSTGRES facilities, speculation on some design characteristics of a seamless, scaleless data base is possible. The seamless requirement to traverse the entire data base and retrieve various elements implies that the DBMS must support abstract data types and user-defined indexes. The scaleless requirement to vary data base resolution implies that the DBMS support multiple representations, user-defined operators, and rules.

The DLG-E spatial objects (points, nodes, chains, and areas) are defined as new atomic abstract data types (ADT) using the POSTGRES abstract data type definition facility (the "define type" command). Creation of these data types allows a further definition of sets of spatial operators (Claire and Guptill, 1982) to work on those data types (using the "define operator" command). Feature objects are each defined as ADT's of type POSTQUEL. This definition allows the feature objects to be represented by a set of shared subobjects, that in DLG-E may either be spatial objects or feature objects. Using field of type POSTQUEL (containing a sequence of commands to retrieve data from other relations) for the DLG-E feature objects will allow for multiple representations of the object (for example, three representations of Dulles Airport). The proper representation from a set of representations (or even a modification of a given representation by a spatial operator) could be invoked using the POSTGRES rule management facilities (Stonebraker and others, 1987). Finally, the "define index" command of POSTGRES allows a user to define secondary indexes on various relations in the data base. The use of external R-Tree or quad-tree indexes, coupled with appropriate access method software, should allow reasonably quick retrievals from large seamless data bases.

CONCLUSIONS

Spatial data users are the driving force toward seamless, scaleless spatial data bases. Data producers would have difficulty justifying such capabilities unless they were needed to satisfy their own internal use (for example, the Census Bureau). GIS users, on the other hand, would probably prefer to view their study area in total, not arbitrarily partitioned by map sheet edges or effected by varying resolution data within the area. With extensible DBMS's, the technology to handle seamless, scaleless data bases is almost at hand. The burden is, therefore, placed on data producers to create future data-base designs that do not preclude users from creating seamless, scaleless versions. Toward this end, the adoption by many mapping agencies of a feature-based, object-oriented data model is a positive step. Assiduous edge-matching across map sheets would be another step forward.

REFERENCES

Barwinski, Klaus, and Brüggemann, Heinz, 1986, Development of Digital Cadastral and Topographic Maps - Requirements, Goals, and Basic Concept: Proceedings of Auto Carto London, Imperial College, South Kensington, London, Vol. 2, pp. 76-85.

Bernard, Antoine, and Piquet-Pellorce, Daniel, 1986, A Workstation for Handling Located Data : PISTIL: Proceedings of Auto Carto London, Imperial College, South Kensington, London, Vol. 1, pp. 166-174.

Cary, M., DeWitt, D., Frank, D., Graefe, G., Richardson, J., Shekita, E., and Muralikrishna, M., 1986, The Architecture of the EXODUS Extensible DBMS: Proceedings of the International Workshop on Object-Oriented Database Systems, Pacific Grove, California, September, 1986.

Claire, R.W., and Guptill, S.C., 1982, Spatial Operators for Selected Data Types: Proceedings of Auto Carto 5, Arlington, Virginia, pp. 189-200.

Fields, Craig, 1978, Beyond "Electronic Paper", Harvard Papers on Geographic Information Systems, Volume Seven: Laboratory for Computer Graphics and Spatial Analysis, Harvard University, Cambridge, Massachusetts, 7 p.

Goodchild, M.F., 1987, Towards an Enumeration and Classification of GIS Functions: Proceedings of International GIS Symposium, Arlington, Virginia, November, 1987, [in press].

Guptill, S.C., 1986, A New Design for the U.S. Geological Survey's National Digital Cartographic Data Base: Proceedings of Auto Carto London, Imperial College, South Kensington, London, Vol. 2, pp. 10-18.

Guptill, S.C., Fegeas, R.G., and Domaratz, M.A., 1988, Designing an Enhanced Digital Line Graph: American Congress on Surveying and Mapping, 1988 ACSM-ASPRS Annual Convention, St. Louis, Missouri, Vol. 2, pp. 252-261.

Guptill, S.C., and Starr, L.E., 1984, The Future of Cartography in the Information Age, in Computer-Assisted Cartography Research and Development Report: International Cartographic Association, p. 2.

Kinnear, C., 1987, The TIGER Structure: Proceedings of Auto Carto 8, Eighth International Symposium on Computer Assisted Cartography, Baltimore, Maryland, pp. 249-257.

Madill, R.J., 1987, Content Management - The Challenge for Geographic Information Systems: Proceedings of International Cartographic Association Conference, Morelia, Mexico, October 12-21, 1987, Vol. 1, pp. 141-147.

Marx, R.W., 1986, The TIGER System: Automating the Geographic Structure of the United States Census: Government Publications Review, Vol. 13, pp. 181-201.

Marx, R.W., 1987, The TIGER System: Six Years to Success: Proceedings of the 13th. International Cartographic Conference, Morelia, Mexico, October 12-21, 1987, Vol. IV, pp. 633-645.

Morrison, J.L., Callahan, G.M., and Olsen, R.W., 1987, Digital Systems Development at the U.S. Geological Survey: Proceeding of International Cartographic Association Conference, Morelia, Mexico, October 12-21, 1987, Proceedings, Vol. 4, pp. 201-214.

Radlinski, W.A., 1974, Untitled Keynote Address in Proceedings of the International Conference on Automation in Cartography, Reston, Virginia, December 9-12, 1974: American Congress on Surveying and Mapping, Falls Church, Virginia, pp. 3-7.

Rhind, D.R., 1988, Personality as a Factor in the Development of a Discipline: The Example of Computer-Assisted Cartography: The American Cartographer, Vol. 15, No. 3, pp. 277-89.

Salgé, Francois, and Piquet-Pellorce, Daniel, 1986, The I.G.N. Small Scale Geographical Data Base (1:100,000 to 1:500,000): Proceedings of Auto Carto London, Imperial College, South Kensington, London, Vol. 1, pp. 433-446.

Schuch, H., 1988, Wild SYSTEM 9: A Perspective for the User: Technical Papers, 1988 ACSM-ASPRS Annual Convention, St. Louis, MO, March 13-18, 1988, Vol. 2, pp. 149-158.

Stonebraker, M.R., and Rowe, L.A., 1986, The Design of POSTGRES: Proceedings of the ACM-SIGMOD International Conference on Management of Data, Washington D.C., May, 1986.

Stonebraker, M.R., Hanson, E., and Hong, C.H., 1987, The Design of the POSTGRES Rules System: Proceedings of IEEE Conference on Data Engineering, Los Angeles, California, February, 1987.

The American Cartographer, 1988, The Proposed Standard for Digital Cartographic Data: Vol. 15, No. 1, 144 p.

Tomlinson, R.F., 1988, The Impact of the Transition From Analogue to Digital Cartographic Representation: The American Cartographer, Vol. 15, No. 3, pp. 249-61.

U.S. Geological Survey, 1986, Digital Line Graphs from 1:24,000-Scale Maps: U.S. Geological Survey Data Users Guide 1, 109 pp.

Waugh, T.C., and Healey, R.G., 1987, The GEOVIEW design. A relational data base approach to geographical data handling: International Journal of Geographical Information Systems, Vol. 1, No. 2, pp. 101-118.