A NEW DESIGN FOR THE U.S. GEOLOGICAL SURVEY'S NATIONAL DIGITAL CARTOGRAPHIC DATA BASE

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

ABSTRACT

In analyzing the requirements for the U.S. Geological Survey's next generation of digital cartographic systems, several enhancements to the National Digital Cartographic Data Base regarding spatial data structures and data management systems were identified as key items. These included the development and implementation of a new data structure having enhanced attribute and feature identification characteristics to support product generation and analytical requirements and the development and implementation of an improved data management capability. A data base design incorporating these enhancements has been completed. Research and development activities pursuant to these design concepts are underway. A prototype of the spatial data base environment will be developed on a standalone workstation (32-bit supermicro) using relational data base software and anticipating a network of workstations connected to a central data archive.

INTRODUCTION

The U.S. Geological Survey (USGS) has been collecting information for the National Digital Cartographic Data Base (NDCDB) for almost a decade. The initial plans were for the NDCDB to consist of boundary, public land net, stream and water-body, and transportation data collected from 1:24,000-scale maps; and elevation data usually obtained concurrently with the orthophotoquad program. Over the years, other components have been integrated into the NDCDB including: planimetric data from the 1:2,000,000-scale sectional maps of the National Atlas of the United States of America; elevation data from the 1:250,000-scale map series; land use and land cover and associated map data; and geographic names data.

Recently the scope of the data base has again been increased to include digital cartographic data collected from 1:100,000-scale base maps. These data will provide nationwide coverage of transportation features (such as roads, railroads, powerlines, and pipelines) and hydrographic features (such as streams, rivers, and water bodies) by the end of the decade.

The digital planimetric data are produced and distributed in the form of digital line graphs (DLG). The DLG concept uses the principles of graph theory and topology to represent a graph as a set of nodes, lines, and areas that explicitly records the spatial relationships inherent in the graph. The use of the DLG structure in the NDCDB has helped to hasten the adoption of topologically structured data by the mapping and charting community.

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

REQUIREMENTS FOR CHANGE

The USGS has embarked on a major development program to implement advanced technology and production procedures to satisfy National Mapping Program product and process requirements by the year 2000 with increased efficiency and acceptable operational costs. This program will exploit state-of-the-art technology available over the next 6 years, followed by an intensive production effort during the 1990's. The NDCDB will evolve to become the central focus of most National Mapping Division activities, including maintenance and revision of our basic map series and provision of data to geographic information systems.

The current NDCDB system, which manages an archive of cartographic data files, will have to be enhanced to meet the design goals of the future system. The major requirements for improvements and changes fall into four areas:

(1) Development and implementation of a data management facility within each Mapping Center to better support increased data production and product generation requirements.

(2) Acquisition and implementation of mass storage systems to support improved data storage and retrieval.

(3) Development and implementation of a new data structure and spatial data base management system having enhanced attribute and feature handling characteristics that will support expanded data production, product generation, and analysis requirements of the digital data.

(4) Development and implementation of an NDCDB data distribution subsystem to support the public sale of standard and derivative digital and graphic products.

CONCEPT OF OPERATIONS

To meet these requirements a concept of operations has been developed to define the future data base operations and data structure. The NDCDB is envisioned as functioning in two environments: archival and operational. The archival data base environment will consist of a central data repository, including information and indices about the data (metadata), and will support retrieval by cartographic feature and geographic partitions. The operational environment will provide data staging and tracking facilities for the movement of data between the archival and operational data bases. It also allow for the manipulation, analysis, and display of the spatial information in a data base context. The queries, retrievals, and uses are different within the archival and the operational environments.

Archival Data Base

The archival data base will consist of spatial data and metadata. The spatial data would be partitioned by series, quadrangles, and categories. These partitions are defined as:

Series: a partition by data content (imagery, elevation matrices, cartographic data) or scale (1:24,000, 1:100,000); 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.

Metadata are descriptions of the data content of the NDCDB. These descriptions apply to sets of records such as a quadrangle of a series and include such information as data source, processing history, and accuracy. The metadata also contain feature keys to quadrangle coverage. For example a key to all the quadrangles that contain the Potomac River would be maintained in the metadata. Queries are against only the metadata, not the spatial data. The spatial data would be retrieved by quadrangle multiples according to series, category, or feature designations.

Operational-Level Data Base

The operational environment will consist of two subsystems; a data staging and tracking system, and a spatial data management system. The data staging and tracking system provides facilities for the transfer of information between the archival and operational data bases and will manage the local storage of data in work at a production center. The operational-level data base will be initialized by quadrangle multiples of data from the NDCDB Archive. This downloading of data from the archive (where it will reside on a high-density mass storage device such as an optical disk or magnetic tape cassette) will probably occur over a local area network either to a local host computer or directly to a computer workstation.

The spatial data management system will provide the capabilities to manipulate spatial information with a set of data base tools. The schema used in this system will consist of the various feature, location, and topological components that make up a cartographic data structure. It is postulated that these data elements can be placed in a relational data base and used in the workstation environment. Data retrievals by feature, attribute, topology, or location will be required in the operational environment.

CARTOGRAPHIC DATA MODELS

To understand the conceptual design being proposed for the spatial data base management system, it is helpful to see the role of digital cartographic data as a model of reality. All maps can be considered as models of the real world. Future cartographic applications and activities will center around a spatial data base which is, like a map, a multifaceted model of geographic reality. Before such a spatial data base can be designed, there is, therefore, a need to define an efficient model of the geographer's abstraction of the real world.

The key here is to define and develop abstract global descriptions for digital cartographic and geographic data. This issue of data abstraction arises with respect to the identification of spatial entities. Definitions based on dimensionality of space (that is, points, lines, and areas) address the issue of feature description from only one perspective, that of object geometry. This is not sufficient.

To a user whose applications are to be supported by spatial analysis operations, individual points, lines, and polygons are but abstractions of the user's view of spatial reality. It would benefit the user if spatial entities and respective operators were defined at a user-logic level. Spatial reality from a user's view probably comprises visually or logically discernible geographic point, linear, and areal features (such as landmarks, roads, and counties), each with their respective descriptive attributes. We will term these spatial entities cartographic features.

More precisely, a cartographic feature is the set of points, lines, and/or areas that meet some specified attribute or spatial criteria. Examples include highways, by route number or by name; named complexes, such as Dulles Airport; unnamed complexes, such as a drive-in theater; and named natural features such as the Potomac River, including its shorelines, islands, falls, rocks, and mudflats. A spatial data base management system would allow the user not only to pose questions, but also to receive responses expressed in the user-logic level of abstraction, that is, in terms of cartographic features. The actual manipulations would occur at a lower level of abstraction, one that contains the geometric information about the spatial entity.

The representation of shape and location at the geometric and subordinate levels of abstraction are usually achieved by using either raster (typically using grid cells to enumerate spatial occupancy) or vector (using sequences of x,y coordinate pairs to represent boundaries) techniques. In this design, the geometric description of spatial entities will be presented in terms of a vector representation. However, it should be noted that the conceptual schema for the spatial data base management system separates the geometric descriptions of spatial entities from the feature description of the entity. In other words, alternate (or even multiple) geometric models can be used to describe the shape characteristics of the objects (point, line and area entities) of the cartographic model. For example, roads could be represented by vector line segments and county areas could be represented by quad trees (Samet, 1984).

In addition to modeling the geometry of cartographic features, the relationships of these features to each other need to be described in the model. One set of these relations describes the topology of the features (that is, relationships such as adjacent to, connected to, contained in, and bounded by). A second set of relations describes the class of objects that comprise a given cartographic feature.

In summary, the cartographic data model consists of spatial entities (objects), descriptive attributes for the objects, geometric descriptions of those objects, and the relationships among the objects. The concepts and terminology of the extended relational model RM/T (Codd, 1979; Date, 1983) have been used to develop the design for the cartographic data model. A formal description of this model is given in Appendix A. For data transfer this schema will be implemented in an enhanced digital line graph (DLG-E) format. The DLG-E will conform to the data standards guidelines being proposed by both the Federal Interagency Coordinating Committee on Digital Cartography (1985, p. 25-38) and the National Committee for Digital Cartographic Data Standards (Moellering, 1986).

DATA BASE OPERATORS

As described, the cartographic data model consists of the basic components (entities, designations, and properties) used in the extended relational data model. This then, allows relational data manipulations to be performed using the set of relational operators (select, project, product, union, intersection, difference, join, and divide (Date, 1986a)). Thus, the spatial data base management system has full manipulative power of the relational model available to process the descriptive attributes and the other non-spatial components of the cartographic data model. For the spatial (geometric) components of the model, it is proposed to use a set of spatial operators (Claire and Guptill, 1982; Claire, 1984). The commonly referenced spatial operators are listed in table 1.

Table 1.--Commonly referenced spatial operators (from Claire and Guptill, 1982).

OPERATOR	CLASS	OPERAND	RESULT	COMMENTS
LENGTH	Mon	1-cell	Nu	
AREA	Mon	2-cell	Nu	
BOUNDARY	Mon	2-cell	1-cell	Set of points that comprise limit of α
COMPLEMENT	Mon	n-cell	n cell	Set of points that are not members of α.
EXTEND	Mon	n-cell	2 cell	Set of points within a distance 'd' of α.
SEPARATION	Dya,sym	n-cell	Nu	Minimum of distances between points of α and points of β.
OVERLAP	Dya,sym	n-cell	B	True if α and β have at least one point in common
EQUALS	Dya,sym	n-cell	B	True if all points are members of both α and β.
CONTAINS	Dya,asy	n-cell	B	True if all points of β are members of α.
INTERSECTION	Dya,sym	n-cell	n cell	Set of points that are members of α AND members of β
UNION	Dya,sym	n-cell	n cell	Set of points that are members of α OR members of β
DIFFERENCE	Dya,asy	n-cell	n cell	Set of points that are members of α and not members of β

Mon : Monadic
Dya : Dyadic
Sym : Symmetric
Asy : Asymmetric
Nu : Numeric
B : Boolean
n-cell : 0-cell, 1-cell, 2-cell or combination
α, β : Operands

Spatial operators represent the basic and fundamental manipulations of spatial entities. The purpose of both the relational and spatial operators is to allow for the writing of expressions, which themselves serve many purposes, including data retrieval. The spatial operators are the analogs for spatial data to the set of relational operators used with non-spatial data. As such, spatial operators would be applied to various geometric components of the cartographic data model and relational operators to the remaining elements of the model.

Taken in concert, the relational and spatial operators provide full manipulative capabilities over all elements of the cartographic data model. They are the building blocks that may be aggregated in various ways to support the increasingly complex forms of geoprocessing. Thus we have both the data structure and the data manipulation components that are necessary for a spatial data base management system.

SUMMARY AND CONCLUSIONS

A preliminary design for the next generation of the NDCBD has been completed. The key features of this design are the creation of a comprehensive cartographic data model by adding a feature data element to the DLG schema and the separation of data base operations into archival and operational environments. It is thought that commercial relational data base management systems can be used to manipulate the non-spatial components of the data model. Spatial operators will perform fundamental geometric manipulations of spatial entities. These design concepts are being investigated and test implementations are underway.

REFERENCES

Claire, R.W., 1984, Algorithmic Development of Spatial Operators: The Ninth William T. Pecora Memorial Remote Sensing Symposium, Proceedings, Sioux Falls, South Dakota, October 2-4, 1984, p. 213-221.

Claire, R. W., and Guptill, S. C., 1982, Spatial Operators for Selected Data Structures: Auto-Carto V/Commission IV, International Society for Photogrammetry and Remote Sensing Symposium, Arlington, Virginia, Proceedings, August 22-28, 1982, p. 189-200.

Codd, E.F., 1979, Extending the Database Relational Model to Capture More Meaning: Association for Computing Machinery, Transactions on Database Systems, vol. 4 no. 4, December 1979.

Date, C.J., 1983, An Introduction to Database Systems: Addison-Wesley, Reading, Mass., vol. II, p. 241-289.

Date, C.J., 1986a, An Introduction to Database Systems: Addison-Wesley, Reading, Mass., vol. 1, p. 257-280.

Date, C.J., 1986b, Relational Database: Selected Writings: Addison-Wesley, Reading, Mass.

Federal Interagency Coordinating Committee on Digital Cartography, 1985, Coordination of Digital Cartographic Activities in the Federal Government: 3rd Annual Report to the Director, (U.S.) Office of Management and Budget, 50 p.

Moellering, H., 1986, Testing the Interim Proposed Standards and Elevation of Digital Cartograhic Objects [abs.]: Technical Papers, American Congress on Surveying and Mapping and American Society of Photogrammetry and Remote Sensing Annual Convention, Cartography and Education, v. 1, p.4.

Samet, H., 1984, The Quadtree and Related Hierarchial Data Structures: Computing Surveys, vol. 16, no. 2, p. 187–260.

APPENDIX A
DATA BASE CONCEPTUAL SCHEMA

This description of the design for the spatial data base uses the general terminology of the RM/T data model (Codd, 1979), and more specifically, Date's "pseudo-DDL" formal expressions (Date, 1986b). The main elements of this semantic are entities (classified as being either kernel, associative, or characteristic), designations, and properties. Each entity (kernel, associative, characteristic) maps into a base table. Each designation maps into a field (foreign key) within the base table for the designating entity type. Each property maps into a field in a base table. Primary keys provide the only guaranteed record-level addressing mechanism in the relational model. Foreign keys provide the referencing mechanism in the relational model.

Pseudo-DDL description of the Spatial Data Base

```
CREATE TABLE N /* NODES(KERNEL) - DESIGNATE NG */
   FIELDS (N#, NG#, NAME, ATTRIBUTES,...)
   PRIMARY KEY (N#)
   FOREIGN KEY (NG# IDENTIFIES NG
           NULLS ALLOWED
           DELETE OF NG NULLIFIES
           UPDATE OF NG.NG# CASCADES)

CREATE TABLE L /* LINES(KERNEL) - DESIGNATE LG */
   FIELDS (L#, LG#, DIRECTION, NAME, ATTRIBUTES,...)
   PRIMARY KEY (L#)
   FOREIGN KEY (LG# IDENTIFIES LG
           NULLS ALLOWED
           DELETE OF LG NULLIFIES
           UPDATE OF LG.LG# CASCADES)

CREATE TABLE A /* AREAS(KERNEL) - DESIGNATE AG */
   FIELDS (A#, AG#, NAME, ATTRIBUTES,...)
   PRIMARY KEY (A#)
   FOREIGN KEY (AG# IDENTIFIES AG
           NULLS ALLOWED
           DELETE OF AG NULLIFIES
           UPDATE OF AG.AG# CASCADES)

CREATE TABLE F /* FEATURES(KERNEL)*/
   FIELDS (F#, NAME, ATTRIBUTES,...)
   PRIMARY KEY (F#)
```

```
CREATE TABLE NL /* NODES AND LINES - ASSOCIATE LINES AT A
            NODE */
   FIELDS (N#, L#, PLACE,...)
   PRIMARY KEY (N#, L#)
   FOREIGN KEY (N# IDENTIFIES N
           NULLS NOT ALLOWED
           DELETE OF N CASCADES
           UPDATE OF N.N# CASCADES)
   FOREIGN KEY (L# IDENTIFIES L
           NULLS NOT ALLOWED
           DELETE OF L CASCADES
           UPDATE OF L.L# CASCADES)

CREATE TABLE LBA /* LINES BOUNDING AREAS - ASSOCIATE LINES
            AND AREAS */
   FIELDS (L#, A#, ORIENT, DIRECTION, INOUT, PRIOR_L#,
           NEXT_L#,...)
   PRIMARY KEY (L#,A#)
   FOREIGN KEY (L# IDENTIFIES L
           NULLS NOT ALLOWED
           DELETE OF L CASCADES
           UPDATE OF L.L# CASCADES)
   FOREIGN KEY (A# IDENTIFIES A
           NULLS NOT ALLOWED
           DELETE OF A CASCADES
           UPDATE OF A.A# CASCADES)

CREATE TABLE FF /* FEATURES COMPOSED OF FEATURES -
            ASSOCIATE FEATURES */
   FIELDS (MAJOR_F#, MINOR_F#,...)
   PRIMARY KEY (MAJOR_F#, MINOR_F#)
   FOREIGN KEY (MAJOR_F# IDENTIFIES F
           NULLS NOT ALLOWED
           DELETE OF F CASCADES
           UPDATE OF F.MAJORF# CASCADES)
   FOREIGN KEY (MINOR_F# IDENTIFIES F
           NULLS NOT ALLOWED
           DELETE OF F CASCADES
           UPDATE OF F.MINOR_F# CASCADES)

CREATE TABLE FA /* FEATURES COMPOSED OF AREAS - ASSOCIATE
            FEATURES AND AREAS */
   FIELDS (F#, A#,...)
   PRIMARY KEY (F#,A#)
   FOREIGN KEY (F# IDENTIFIES F
           NULLS NOT ALLOWED
           DELETE OF F CASCADES
           UPDATE OF F.F# CASCADES)
   FOREIGN KEY (A# IDENTIFIES AREAS
           NULLS NOT ALLOWED
           DELETE OF A CASCADES
           UPDATE OF A.A# CASCADES)
```

```
CREATE TABLE FL /* FEATURES COMPOSED OF LINES - ASSOCIATE
            FEATURES AND LINES */
   FIELDS (F#, L#,...)
   PRIMARY KEY (F#,L#)
   FOREIGN KEY (F# IDENTIFIES F
          NULLS NOT ALLOWED
          DELETE OF F CASCADES
          UPDATE OF F.F# CASCADES)
   FOREIGN KEY (L# IDENTIFIES L
          NULLS NOT ALLOWED
          DELETE OF L CASCADES
          UPDATE OF L.L# CASCADES)

CREATE TABLE FN /* FEATURES COMPOSED OF NODES - ASSOCIATE
            FEATURES AND NODES */
   FIELDS (F#, N#,...)
   PRIMARY KEY (F#,N#)
   FOREIGN KEY (F# IDENTIFIES F
          NULLS NOT ALLOWED
          DELETE OF F CASCADES
          UPDATE OF F.F# CASCADES)
   FOREIGN KEY (N# IDENTIFIES N
          NULLS NOT ALLOWED
          DELETE OF N CASCADES
          UPDATE OF N.N# CASCADES)

CREATE TABLE NGEOM /* GEOMETRY OF NODES (KERNEL) -
            DESIGNATE NODES */
   FIELDS (NG#, N#, GEOM,...)
   PRIMARY KEY (NG#)
   FOREIGN KEY (N# IDENTIFIES N
          NULLS NOT ALLOWED
          DELETE OF N CASCADES
          UPDATE OF N.N# CASCADES)

CREATE TABLE LGEOM /* GEOMETRY OF LINES (KERNEL) -
            DESIGNATE LINES */
   FIELDS (LG#, L#, GEOM,...)
   PRIMARY KEY (LG#)
   FOREIGN KEY (L# IDENTIFIES L
          NULLS NOT ALLOWED
          DELETE OF L CASCADES
          UPDATE OF L.L# CASCADES)

CREATE TABLE AGEOM /* GEOMETRY OF AREAS (KERNEL) - DESIGNATE
AREAS */
   FIELDS (AG#, A#, GEOM,...)
   PRIMARY KEY (AG#)
   FOREIGN KEY (A# IDENTIFIES A
          NULLS NOT ALLOWED
          DELETE OF A CASCADES
          UPDATE OF A.A# CASCADES)
```