14th Congress of the International Society of Photogrammetry

Hamburg 1980

Commission IV - 3

Invited Paper

Dr. J. G. Linders

University of Guelph
Canada

ABSTRACT

Georeferenced Data Base Design

The development of a comprehensive Georeferenced data base has been necessitated by the need to integrate and relate various land mass data to basic map survey data. The essence of a Georeferenced data base is to provide an information base from which a variety of information as well as graphic products may be derived upon demand. This paper discusses the design, implementation and operation for the integration of all Georeferencing data needs. Specific attention is given to the need to model complex spatial relations and their use for the production of map graphics as well as for specialized information processing functions.

Georeference Data Base

Professor J. G. Linders

Department of Computing & Information Science
University of Guelph
Guelph, Ontario
Canada

INTRODUCTION

Georeferencing is a common term which is coming into usage within the context of land mass information systems. It is concerned with the organization, storage, and management of all information related to the land mass. There are two major aspects to Georeferencing namely:

1. Geocoding which is the term being applied to the identification and coding of land mass features and,

2. A Georeference data base to provide for an information systems environment for storage, handling, and manipulation of all geocoded information.

The interest in Georeferencing has risen naturally from the variety of developments in spatial data management systems. These range from computer assisted cartography to the handling of computer based files for specific topographically related items, for example, resource inventories, control survey files, drainage systems etcetera. The experience and limitations of these developments has been the primary motivation for developing a more systematic and unified approach for handling georeference data.

The primary objectives of a Georeference data base are:

1. The definition, description and organization of topographic features within a single information system concept.

2. The integration of thematic information with basic survey data.

3. The storage of generic information from which various representations can be derived on demand.

4. The development of a

stem environment in which the information may be used for a variety of purposes ranging from the production of graphic products, reports, as well as analysis.

The data base approach emphasizes the collection of data in an application independent manner. The basic methodology of data base design (1) is predicated on the notion of various levels of schema within the CODASYL data base task groups report. The data base is viewed as a set of data items organized according to specific rules. This represents the data base schema while access to the schema is defined in terms of sub-schema, one for each individual application area. An alternative and more detailed proposal by the ANSI SPARC Group identifies at least three levels of schema which are:

1. Conceptual schema which provides the organizational framework for the description of the data base.

2. Internal schema for describing how the data base functions in terms of information processing.

3. External schema for providing access to the data base for all users.

The data base methodology represents a culmination of information processing technology in terms of a generalized system approach to handling diverse data needs. Within a practical framework a data base is normally realized in terms of a number of generalized products including:

1. A data dictionary for defining the semantics and elements of the total data base including accesss via subschema.

2. A data base management system for the orderly storage and management of data items.

3. A variety of process orientated tasks for effecting transactions against the data base.

The role of the data base designer is to ensure that the proper data organization is maintained within the information structures identified for realizing the data base, as well as to design the various functions and transactions which will be effected on the data base. This must be done within the framework of limited knowledge as it is seldom possible to anticipate all the needs of the data base at data design time. Consequently, the designer is faced with the need to recognize a generalized capability to permit ready enhancement of information, information structures as well as more sophisticated requirements such as data consistency checking, etcetera.

The role of the data base designer is to select information structures which will meet the needs of the community of users on the data base. Of primary consideration in loading of the data base is to ensure data base integrity. One aspect of this is the need to validate consistency of the data elements themselves. In the most naive example it would be necessary to ensure that an area defined to be a lake does not have superimposed upon it any other land related features, such as roads and contours etcetera. The data base designer is also concerned with maintaining a modular and hence extensible system which can be adapted to the needs of the user community without the requirement to re-organize the data base as new applications arise. The whole area of data base design is very much an art which is quickly evolving to a science (2,3).

MODELLING GEOREFERENCED INFORMATION

A Georeferenced data base exemplifies a wide variety of information needs. Traditionally most data base management systems have been modeled to reflect hierarchical, network or relational data. Within a Georeference system all three information types are readily present and hence must be modeled in order to realize the requirements of a single Georeference data base.

There are two basic approaches to the development of land mass information system. The first approach is to recognize and maintain an inventory of features within the land mass. The second approach, commonly referred to as digital terrain modeling involves the specification of terrain data according to a regular grid. Normally each has its own areas of application however, the former approach is the most preferred within a generalized Georeferenced data base framework.

Within a feature oriented data base the following information structures must be accommodated.

A. Hierarchies

Features are organized into hierarchical structures, for example a feature may consist of other features which are emselves defined in terms of components of features. Within a land registry framework, a particular area may be defined as a subdivision within which are lots, each of which is itself divided into parcels which in turn are made up of individual features such as boundaries.

B. Networks

To model to model links between features, for example road and drainage networks which define connectivety between sequences of features.

C. Associations

Within a spatial data management system the relationships the features bear to each other is often important. By way of example, within a planning environment the spatial dispersion of parcels in terms of neighbouring relationships are clearly important in making judgements and decisions about the land mass.

The three information structures described above are normally not found within a single data base system, only the relational data base model as proposed by Codd (4) is capable of modeling all three types of information structure. It should however, be pointed out that no acceptable practical implementation of the relational data base currently exists to be considered a suitable candidate for developing Georeference data base systems. With this in mind it has been necessary to develop special purpose data base management systems for recognizing and realizing the requirements of Georeference data base systems.

It should be realized that the information structures are in essence requirements over and above those of the choice of basic data items themselves. The items to be stored have been identified in other systems and relate very much to the individual applications (5).

GEOREFERENCE DATA BASE SYSTEM

The Georeference system under development at the University of Guelph is characterized by it's data base approach. The conceptual design is based on the notion of a feature as a basic entity within the data base. The feature represents any object or construct, real or conceptual, which can be related to the earth surface. A feature can be simple such as a river, road, contour, etc. or it may be a more plex entity defined in terms of other features - for example subdivision plan. Within the data base there is a taxonomy of features into classes which groups like features together for definition and description. Each feature has associated with it a variety of information types as follows:

1. Co-ordinate data to define all of the positional aspects associated with the feature.

2. Attribute information defining feature characteristics.

3. Descriptive information in the form of text to enhance the interpretation of the data base.

4. Representation of information which describes the depiction of the feature according to it's graphical characteristics.

5. Topological information for building relational and network structures and maintaining associations between the features.

6. Feature definition - information to describe user access to the data base.

The following basic components can be recognized within the Georeference system:

1. Data dictionary for defining all the data entities within the data base. The data dictionary also stores access sub-schema.

2. The file manager, which is the implementation vehicle for the data base management system. This file manager is unique to the Georeferencing requirement and is described elsewhere in this paper.

3. The data collection sub-system represents the sub-schema for entering and loading the data base. The approach has been to do as much consistency checking of the data at data collection time in order to ensure that the integrity of the data base is maintained.

4. Query sub-language representing a specific set of access schema to the data base for the casual user.

5. Display processor, a conceptual mechanism for deriving representations from the data base on demand.

ms 1 and 2 essentially constitute the model for the Georeference data base while items 3,4 and 5 represent the transaction processing functions which manipulate data within the data base environment.

The fundamental consideration of the design of the data base has been that data will require constant enhancement. It is often appropriate to encode primitive data items which are then used to construct larger spatial entities, for example subdivision plans would be defined in terms of parcels which are themselves defined in terms of boundaries and other land mass features. The editing function implied has severe consequences in that it does not represent graphic editings but represents modifications to the data base and hence data base integrity and consistency are essential ingredients to manage.

FILE MANAGER

The file manager represents the data management component for the data base and is responsible for the organization and handling of all topographically related data. The file manager is designed to emulate a high level data base machine. It does this by defining the data organization and a set of functions which can be used to reference, retrieve and manipulate features which are stored within the file manager.

The file manager is structured to recognize a feature as the basic entity within the data base. A feature is any real or conceptual object (e.g. river, lake, contour, etc.) which has meaning within the data base system. A feature may in turn be composed of constituent elements, each of which is an entity within its own right to any level of abstraction necessary.

Conceptually the data base can be construed as providing coverage for the total land mass. The data is partitioned logically on a map sheet basis; each map sheet constituting part of the total mosaic. Even though data is entered on a map sheet basis the data base user is not aware of the arbitrary map sheet boundaries unless he so desires, for example a user may window on any part of the data base which may cross a number of the map sheet partitions. Within a logical partition of the data base (map sheet) the features are organized according to a basic classification. All features of the same class

.. be referenced through a class directory which provides access to every feature within the class. Each feature in turn has a multiplicity of record types which can be potentially associated with it. Each record type corresponds to an information type required to maintain the definition and description of the feature. The following record types are currently stored within the sytem:

1. Feature code record - provides a set of system defined parameters for maintaining the information associated with the storage of the feature. The geocode is also stored within this feature record.

2. Co-ordinate data records - provide the positional information with respect to a frame of reference for all points defining the feature.

3. Attribute data records - defined according to a set of masks for inputing and outputing attributes associated with the feature.

4. Text data records - for maintaining descriptive information about the feature.

5. External data records - for building spatial structures and maintaining relationships between the various features.

Other featured types may be defined as required, and are used in some instances, for example, representation records, historical records, etc.

User access to the file manager is defined in terms of a set of high level parameters. These are structured to reference the data entities on a feature basis and a record type within the feature, for example a user wishing to pass through the class of features called contour could set internal pointers to the feature class directory for contours. Through a set of parameters he is able to reference the first contour and have the information provided to him on a block basis. In order to process a feature the normal procedure would be to ask for the future code record from which the feature characteristics can be ascertained. From the feature code record the user can gain access to either the attribute data records associated with the feature as well as the co-ordinate data records. If the feature is segmented, access to the individual features is obtained in a like manner.

It is intended that the file manager be used for both on-line and batch oriented transaction processing. There are however, a number of stipulations about the operation of the data base. In order to achieve updates an individual record type would be flagged as being amended and a pointer would be established to the new copy of the data records. In like manner the amendment record may itself be amended etcetera. During archiving it is possible to maintain either all of the amendment records or only the final record itself.

The primary requirement in the operation of the file manager is the ability to provide new information as the information content of the data base is enhanced. Essentially this means writing new records associated with the features, for example in building up topological structure as would occur in defining the polygon structure from a set of existing segments; it is necessary to create new external reference records to define the polygon. This is done dynamically by chaining together all the records which are of type external reference. It should be remembered that many different external reference records may exist for a feature, each of which would model a different type of association, for example connectivity, neighbouring relations, etcetera.

The essence of the data base is the file manager and the success of the data base is predicated on the efficiency of operation of the file manager.

THEMATIC DATA BASE APPLICATIONS

The primary consideration in developing the data base is to demonstrate the ability to integrate thematic information with the basic survey data and topography. Within the current pilot production project a number of thematic applications are being integrated into a single data base structure. The same file manager maintains the information for all thematic applications as well as the base map data.

1. OBM Maps

The Ontario base maps provide for coverage of Ontario at scales of 1:10,000 in the urban areas, and 1:20,000 in the northern rural areas, as well as 1:5,000 and 1:2,000 in the urban areas for planning purposes. To date the primary data collection activity has concentrated on collecting data at scales of

.0,000 and 1:20,000. This is being done by the development of a number of specific design procedures (6) which are carefully planned to facilitate and expedite the orderly and cosistent collection of data into the data base. A number of design procedures have been developed for the OBM maps including drainage, culture, contours, and vegetation. In each case a number of consistency checks are executed in real time, and the procedure for entry of data is described to the user on a visual display terminal. The data is collected from the stereo compilation manuscripts which are provided from photogrammetry. As each feature is digitised any attributes or information associated with the feature is entered. In some cases the information is mandatory, for example in the case of contours - heights, boundary descriptions, etcetera. Much of the information is collected by formatted screen procedures and the appropriate data record is created for transmission to the file manager.

Each feature is defined as a free standing entity and any attempt to resolve cartographic interference does not occur until the data display phase. This is consistent with the desire to maintain a definition instead of a representation as would be the case if the representation resolving cartographic interference were stored within the file system.

2. Forest Resource Inventory

A primary motivation for the development of the Georeferencing project was a need to create forest resource inventory maps and provide a management information system to operate in conjunction with resource inventory. It is intended that the information associated with resource inventory be treated as a thematic which is superimposed on the OBM map base. This is achieved by drawing out selected portions of the OBM maps, namely drainage and culture and manually superimposing the forest and boundaries on this base. The individual forest and boundaries are then digitised and the segments which constitute a closed sequence are defined as polygons. This polygon creation process can either be automatic or user assisted through another data collection procedure. As the data is collected in the data base a number of parameters are computed in real time, including perimeter, area, etcetera. Experiments were conducted in the automatic placement of the forest stand text, however, it has been found to be more convenient to have the user identify the location of the text so no automatic placement is necessary.

The forest resource inventory in essence constitutes a separate data base structure within the Georeference data base. However, the primary consideration has been integration within a single data base structure.

3. Lake Bottom Contours

A less significant area of the data base has been the extraction of lake boundaries from the basic topographic data base and the superimposition of transect data on the lake surfaces from which sequence of lake bottom contours is derived through contouring process. The contours can then be stored within the Georeference data base and a number of values computed from this information including lake volume as well as the lake volumes within specific depths.

4. Land Registry

The cadastral element of the data base defines the large scale and high resolution component of the data base. At this time a variety of information structures are available which can be used to collect data using various design procedures and the data is subsequently stored in the file manager. The structures are specific instances of the polygon structures stored for forest resource inventory. The specific application is the last area of major consequence to be developed and is currently in the implementation phase.

5. Summary

The primary rationale for the Georeferencing project has been to build a single integrated data base for the storage, management, retrieval of land mass information. Within the data base structure the land mass is stored as a sequence of definitions as opposed to representations. Whenever a representation is required the features are extracted from the data base and projected onto a cartesian space through a variety of graphic representation procedures. It is hoped to provide a generalized utility for the common sharing of Georeferencing information.

REFERENCES

1. CODASYL, Data Base Task Group Report, ACM, New York, 1971

2. ANSI/X3/SPARC Study Group on Data Base Management Systems: Interim Report, Bulletin of ACM SIGMOD 7, No 2, 1975

3. Ullman, T. D., Principles of Data Base Systems, Computer Science Press, 1980

4. Codd, E. F., A Relational Model of Data for Large Structured Data Banks, CACM 13, 6(1970)

5. Linders, J. G., Distributed Data Bases, Computers & Geosciences, Vol 2, pp 293-297, 1976

6. Linders, J. G., "Data Base Design for Cartographic Applications, presented at International Symposium Commission IV, New Technology in Mapping, Ottawa, Canada, October 1978

- - - - -