URPIS 15 SEMINAR

INTRODUCTION TO TOPOLOGICAL DATA STRUCTURES

November 30 - December 1, 1987

Hobart

John F O'Callaghan
Centre for Spatial Information Systems
CSIRO Division of Information Technology
Canberra, ACT

and

Andrew V Frank
Department of Civil Engineering
University of Maine
Orono, Maine

ABSTRACT

The Seminar introduces the concepts of geographic data structures which underlie geographic information systems. The concepts cover the geometric and topological properties of zone, network and cellular representations.

To illustrate these concepts, some of the techniques for manipulating and analysing geographic data are discussed. Indexing techniques for the efficient retrieval of spatially-referenced data are described.

The Seminar is designed for system analysts, programmers and potential users of geographic information systems.

URPIS 15 SEMINAR

INTRODUCTION TO TOPOLOGICAL DATA STRUCTURES

November 30 - December 1, 1987

Hobart

CONTENTS

SEMINAR PAPERS

1. A Classification of Software Components Commonly used in Geographic Information Systems, Jack Dangermond, In Proc. US/Australia Workshop on the Design and Implementation of Computer-based Geographic Information Systems, D Peuquet and J O'Callaghan (eds.), Amherst NY: IGU Commission on Geographical Data Sensing and Processing, 1983.

2. Cell Graphs: A Provable Correct Method for the Storage of Geometry, Andrew Frank and Werner Kuhn, Proc. Second Int'l Symposium on Spatial Data Handling, Seattle, USA, July 1986, pp. 411-436.

3. Standards Working Party, FICCDC, U.S. Federal Geographic Exchange Format Version 01.08, Dec. 1986, pp 1-10 (only).

basic characteristics: (a) the actual phenomenon or characteristic, such as the variable, its classification, value, name. etc.; and (b) its spatial location (i.e., that location within geographic space where it resides). However, a third characteristic which is particularly relevant to GIS systems is that of time.

Figure 2 is a conceptual diagram of the relationship of these three elements (i.e., locational data, nonlocational data (attributes), and time). From this diagram, one can observe that management of spatial data can become quite complex because locational data and attribute data often change independent of one another with respect to time.

Therefore, effective spatial data management requires that location data and nonlocation data be variable independent of one another. That is to say that attributes can change character but retain the same spatial location, or vice versa.

In conceiving of the overall structure of a data base for geographic information processing, it is useful to realize that data management must occur on both location data, as well as on nonlocation data. Different software and system designs have called for handling these two types of phenomena differently. In some cases, location data is considered as an additional attribute associated with the geographic characteristics. In other systems, the geographic location of a characteristic is kept separately from the actual attribute phenomena associated with that characteristic. The latter approach allows for more flexibility, with respect to handling data change -- specifically, change associated with time.

2. Types of geographic data and its spatial representation within GIS systems -- In general, there are three basic notations used for representing the spatial location of geographic phenomena: points, lines and polygons.

Figure 3 presents a breakdown of seven types of geographic data techniques for spatial representation, as mentioned above. They are (a) feature data; (b) aerial unit information; (c) network topological data; (d) sampling data; (e) surface information; (f) label text information; and (g) graphic symbol data. Graphics in the chart indicate how these seven data types are most commonly portrayed using point, line and polygon locational identifiers.

Points, lines, and polygons are most commonly defined on maps using x.y cartesian coordinates such as longitude/latitude based on principles of Euclidean geometry. This cartesian coordinate system is the most commonly used tool for measuring spatial location and analyzing various properties of spatial location, including measurement, etc. Figure 4 shows conceptually how a typical map with point, line, and polygon elements is expressed

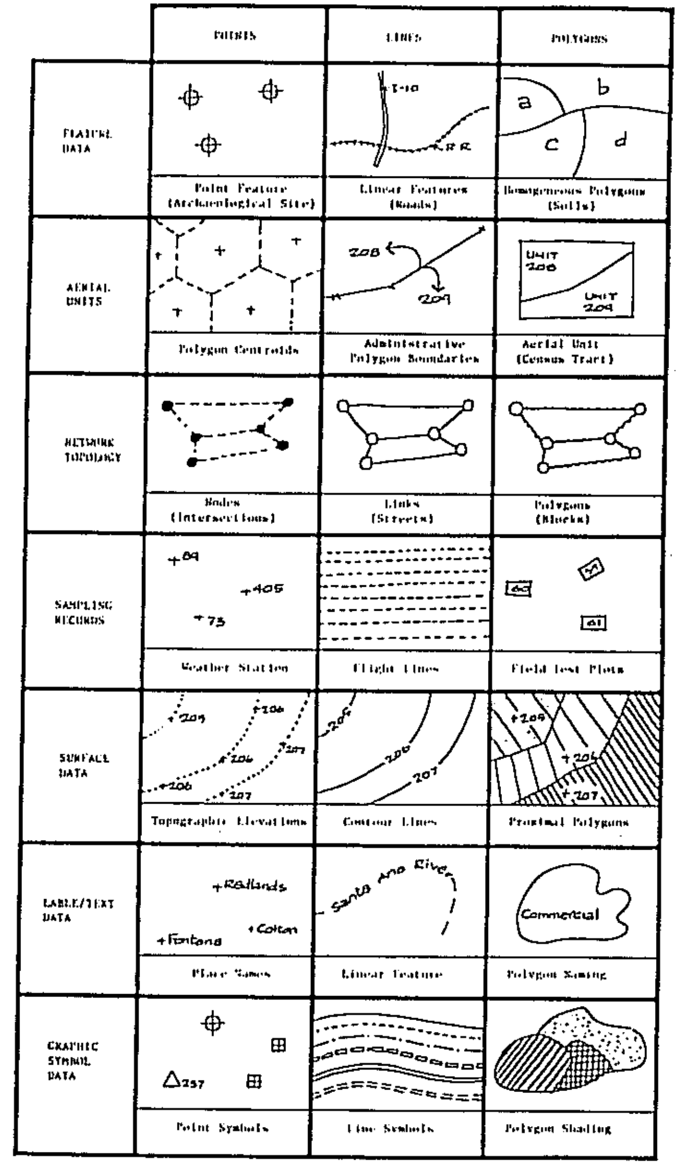
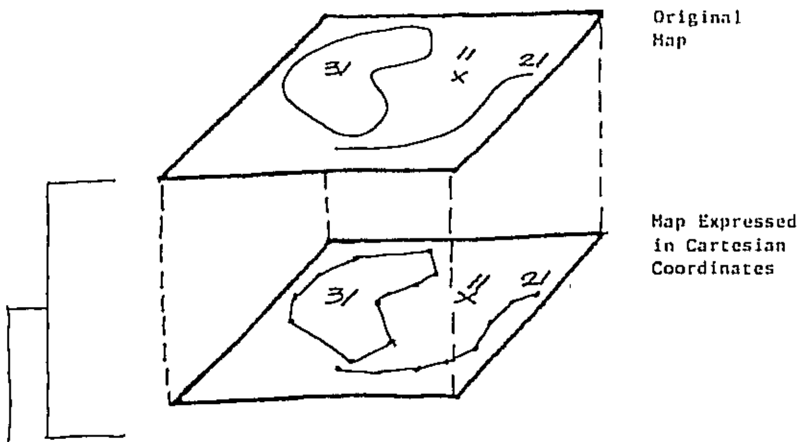


FIG. 3 BREAKDOWN OF GEOGRAPHIC DATA TYPES & METHODS OF REPRESENTATION



	Feature Number	Location
Point	11	X,Y (Single Point)
Line	21	$X_1Y_1, X_2Y_2, \ldots\ldots\ldots\ldots X_nY_n$ (string)
Polygon	31	$X_1Y_1, X_2Y_2, \ldots\ldots\ldots\ldots X_1Y_1$ (closed loop)

FIG. 4 X, Y COORDINATE FILE

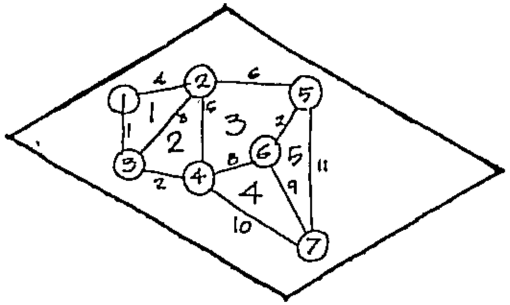
in cartesian coordinates and subsequently transferred to a x,y coordinate GIS file.

In addition to expression of geographic phenomena using cartesian coordinates, one can also use principles of graph theory involving topological relationships to express the relative location of various map elements. Perhaps the first and most commonly used GIS system applying graph theory to maps is called Dual Independent Map Encoding (DIME) and was developed by the U.S. Bureau of the Census. DIME is actually a combined x,y coordinate and topological encoding system. The DIME system involves topologically structuring the graph elements of a map (i.e., nodes, line segments, and polygons). DIME also allows for inclusion of x,y coordinates associated with the node points or intersections connecting the line segments bounding polygon areas. Figure 5 shows how a typical polygon/network map can be abstracted into 7 nodes, 11 links or line segments, bounding 5 basic polygons. By numbering these links and associating them with nodes as well as polygons on the right and left, a basic map notation system can be derived. By adding x,y coordinate encoding at each of the nodes, one has a dual system for spatially identifying all elements of a map. This facilitates not only analysis using x,y coordinates but also use of mathematics associated with graph theory involving network, spatial aggregation, dropline, etc.

Topological quoting defines a location of geography phenomena relative to other phenomena, but does not require the use of the concept of distance in defining these relationships. Therefore, it is totally possible to have a spatially defined map without coordinates entirely.

A second technique which also applies certain of the relationship principles involves the use of a grid mesh to define a regular but arbitrary polygon framework for "holding" geographic data. The grid technique inherently involves association with a coordinate system, but it does not necessarily require precise association. Grid uses an i,j matrix for representing variations of geography to the computer. Figure 6 illustrates how an original polygon map overlaid on a grid can be abstracted into row/column values expressing geographic variation.

In summary, there are basically two methods for spatially identifying this information. The first is to use actual measurements in the form of x.y coordinates. The second involves defining geographic phenomena by way of point and line relationships, networks, polygon adjacencies, and cell contiguities. These are graphically portrayed in figure 2.



Coded Network Map

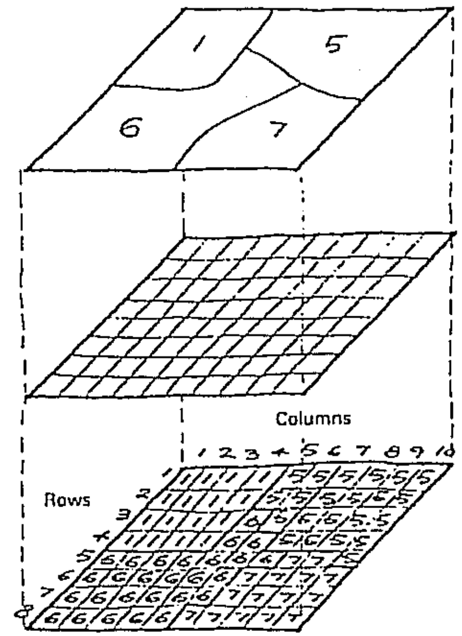
Link #	Right Polygon	Left Polygon	Node 1	Node 2
1	1	0	3	1
2	2	0	4	3
3	2	1	3	2
4	1	0	1	2
5	3	2	4	2
6	3	0	2	5
7	3	3	5	6
8	4	3	6	4
9	5	4	7	6
10	4	0	7	4
11	5	0	5	7

Topologically Coded Network & Polygon File

Node #	X Coordinate	Y Coordinate
1	23	8
2	17	17
3	29	15
4	26	21
5	8	26
6	22	30
7	24	38

X,Y Coordinate Node File

FIG. 5 X , Y COORDINATE NODE FILE



Original Map

Grid Overlay

Grid Cell Map

Row	Column	Value
1	1	1
1	2	1
1	3	1
1	4	1
1	5	5
1	6	5
.	.	.
.	.	.
.	.	.
.	.	.
.	.	.

Grid Cell File

FIG.6 GRID FILE

Automation Techniques for Locational Data

There are seven basic techniques used for spatial identification of geographic data. They include the following:

1. Encoding of geographic data to cells;
2. Topological coding of nodes, line segments, and polygons (DIME system);
3. Input of survey documents;
4. Manual entry of points, lines, and polygons using manual digitizer;
5. Automatic entry at time of data capture;
6. Automatic line following techniques; and
7. Optical scanning.

These techniques are described below:

1. Cell encoding - Manual encoding of cellular data involves a process wherein each point within an overall grid is encoded by observing the dominant feature within a given cell, writing it onto a coding form, and subsequently key entering the coding form into an automated file using a keypunching technique (see figure 7).

2. Topological coding of map features - This effort again involves manual coding from the map of the relationship of each point, line and polygon to each of its adjacent neighbors. For example, each line (link) is assigned 2 nodes which are its end points. Also each link is assigned the polygon which is at its right and left sides. In this manner, all graphic entities are fully documented, key entered, and edited for final corrections.

3. Input of survey documents - Input of survey documents is typically done through the manual key entry of key x,y reference points, bearings, and distances relative to specific features desired for encoding. This technique is often used for entry of ownership boundaries, roads, buildings, etc. The process involves the use of coordinate geometry techniques which have been developed especially for entry of the survey documents and generation automatically of associated cartesian coordinates.

4. Manual digitizing using a digitizer - Manual digitizing technique involves the use of an electromagnetic, electrostatic device called a digitizer. The digitizer converts movements of a cursor or "point locator" into electrical identified locations which are read directly into the computer. Digitizers are usually designed

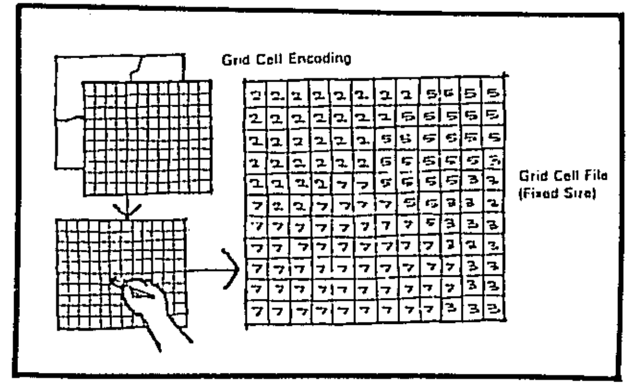


FIG.7 MANUAL CELL ENCODING

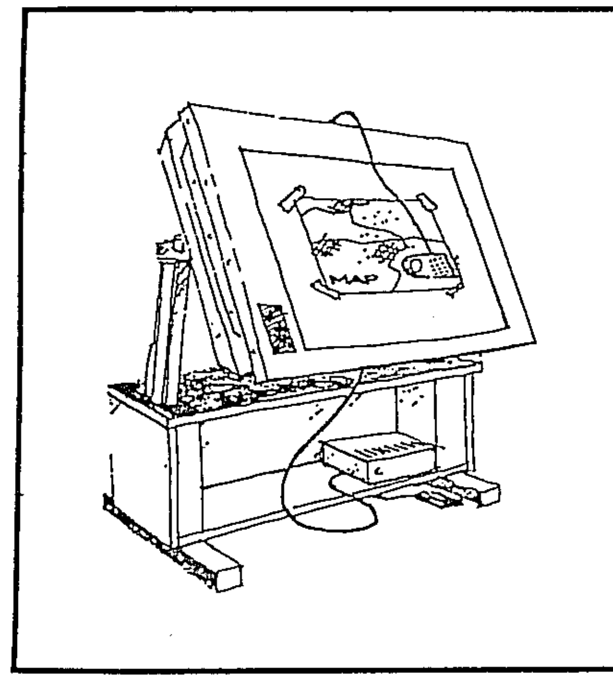


FIG.8 MANUAL DIGITIZER

to be very accurate and can be programmed with a minicomputer to capture data in varying formats of points, lines, and polygons. The digitizer can record the position of each point, line, or polygon in either a point by point mode or by a continuous mode incrementing by time or by distance moved of the cursor on the digitizer table. Figure 8 illustrates a common manual digitizer.

5. Automatic entry devices - This technology actually refers to a variety of emerging technologies and instruments which directly capture automated spatial location at the time when other geographic data is recorded. These range in magnitude from satellites such as Landsat and related image processing technology to digital plotters, geometric equipment, and "auto surveying" devices which capture coordinates as they pass over terrain.

6. Automatic line following - Several hardware manufacturers have produced laser driven automated line following devices which are used to scan off a continuous string of coordinates associated with lines on a map. Although these scanners have some value, they have not proven to be overly economical at this time.

7. Scanning - Automated scanner technology typically operates in the form of raster scan measuring binary (on/off) positions representing the points, lines, polygons, and textual data appearing on a map.

   Subsequent to this raster scan is a raster-to-vector computer processing program which converts the binary information into topologically correct vector type information. At this time, the technology in this area is new and is only being used in a few locations.

Processing/Editing Techniques for Locational Data

All of the automation techniques described in the previous section involve a substantial amount of processing and/or editing subsequent to the initial data capture. Because these processes and steps are extremely system and technique dependent and vary considerably, only an outline of the basic steps of the processes is presented.

1. Plotting or printing out of digitized or coded data for visual editing.

2. Topological checking of the digitized data to insure correctness within the data set;

3. Splinter removal when more than one line has been captured to represent the same vector on the map;

4. Handling of "donut type polygons";

5. Polygonization of arc information into polygons;

6. Editing of x.y coordinate data (both interactively and batch modes) and

7. Edgematch analysis.

In the case where arc data is entered, the editing and processing activity may include things such as creation of topologically correct data structures including automatic left and right labeling of arcs based on codes assigned to centroid data associated with a polygon file.

Manipulation Techniques

This section describes the various analytic functions and data processing functions that can be performed on spatially automated data. Within this section, the actual function is described in a narrative outline and referenced to a pictorial representation of conceptually how each function is performed. No attempt is made to describe these functions in detail as different software systems approach the solution using alternative types of algorithms. Alternatively, emphasis is placed on defining the actual function which is being performed. It should be mentioned that these descriptions are meant to be representative and there will be examples left out.

1. Data Retrieval -- This series of techniques involves the basic extraction, query, and Boolean manipulation of information contained in an organized geographic information system. Figure 9 describes the following data retrieval functions:

   a) Browsing - This task would consist of setting up a CRT and browsing through the graphic and nongraphic data files associated with various maps and groupings of maps.

   b) Windowing - This function calls for the ability to allow the user to specify windows (either by x,y coordinate or textual information). Spatial data base management typically calls for the organization of map sheets into some form of modular relational structure allowing thereby the illusion to the user of a continuous map being contained within the computer.

   c) Query window generation - This function involves the ability to generate points, irregular shaped polygons, squares, circles, and corridors for interactively overlaying with data planes contained within the geographic file. These windows are typically used to perform spatial (coordinate) retrieval of points, lines, and polygons from various map layers which coincide in

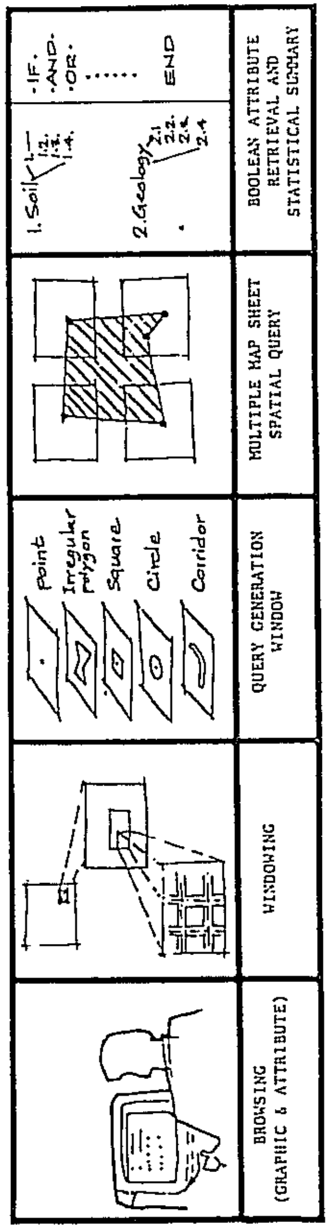


FIG.9 DATA RETRIEVAL

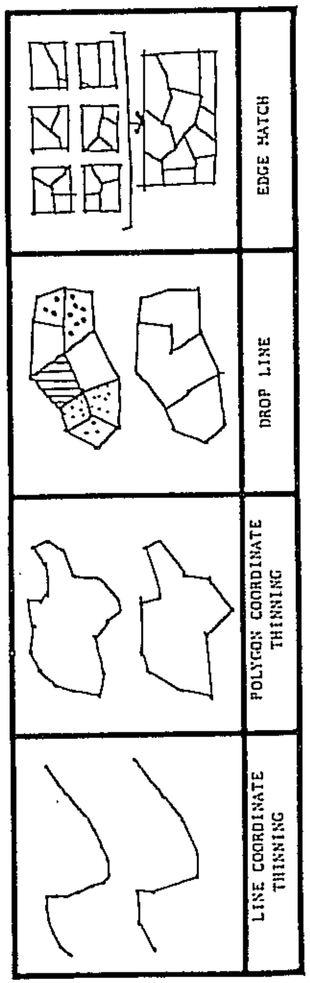


FIG.10 MAP GENERALIZATION

space with these generated query windows. Three techniques are used in extracting information from layers associated with these generated query windows. They are:

(1) Adjacency analysis (i.e., selection of specific points relative to a given point);

(2) Point and polygon retrieval (i.e., the selection of all points, lines, or polygons falling entirely within or partially within the query window); and

(3) Polygon overlay (i.e., selection of only those portions of geographic entities which fall within the boundaries of the query window polygon). In this case, all lines as well as parts of polygons which fall outside of the query window are snipped off using the polygon overlay routine.

These query windows are typically used in asking questions such as "Give me all polygons of the harvestable timber falling within the query window 'A' (a user specified polygon)."

d) Multiple map sheet spatial query - Frequently, the spatial query which is being conducted (i.e., described above) is being done over multiple map sheets. Therefore, the data retrieval software must have within it the ability to create a polygon which stands across map sheet boundaries, extracts those portions of the map sheet which are relevant to the query window, and organizes the map element data into a continuous window automatically. This is typically done in one or more steps on the current software modules which are available.

e) Boolean attribute retrieval and statistical summary - This function involves the ability to specify various Boolean type criteria for extraction of information based on nongraphic attribute data. An example of this would be "Give me all of the polygons of a certain type of soil which have an area of greater than 50 acres, and also produce a statistical summary of the polygon acreages, perimeters, and their total."

It should be noted that most users who have this retrieval type capability become most excited when they are able to conduct the query type retrievals using both the locational window as well as the attribute query. An example of this would be "Give me all of the polygons

1. Engineering mapping systems (typically a minicomputer CAD/CAM system for applications such as photogrammetry, topographic basemaps, road engineering, utilities, facility management, tax mapping and land cadastral information, geodetic data, etc.);

2. Property or parcel information systems (typically a mainframe based DBMS system to handle attributes associated with the land parcel);

3. Generalized thematic and statistical mapping systems (typically on both mini- and mainframe computers and being used for natural resource management, forest inventories, vegetation, geology, soils, census mapping, environmental planning and assessment, etc.);

4. Bibliographic systems which catalog a variety of bibliographic data sets about geographic documents;

5. Geographic base file systems associated with street networks and the aerial units which they define (most common of these is the GBF/DIME-File system developed by the U.S. Bureau of the Census; and

6. Image processing systems (typically associated with processing of Landsat and related satellite image data).

All of these technologies tend to support very specific types of information flows and information systems. It is not the purpose of this paper to go into detail about each of these information systems, but rather to extract from all of these systems the generic type of technologies which are commonly used for storage, manipulation, and display of the data contained within them.

It should be noted and strongly stressed that there are orders of magnitude in the difference in the accuracy of data among these various information systems. For example, engineering basemap systems typically use higher levels of precision and consume greater overhead for describing map features. Therefore, they are less practical for certain types of applications. Also, the vendors of systems use varying types of data management, spatial identification techniques, and approaches to the display of information. No attempt is made to distinguish among these technical approaches rather, focus is placed on generic subjects so that the reader may better understand the relationships and differences between and among these various techniques and systems.

Spatial Information Concepts

This section presents a series of concepts useful to generic definition of GIS components.

1. Basic characteristics of spatial data and its management -- Geographic information is most commonly thought of as having two

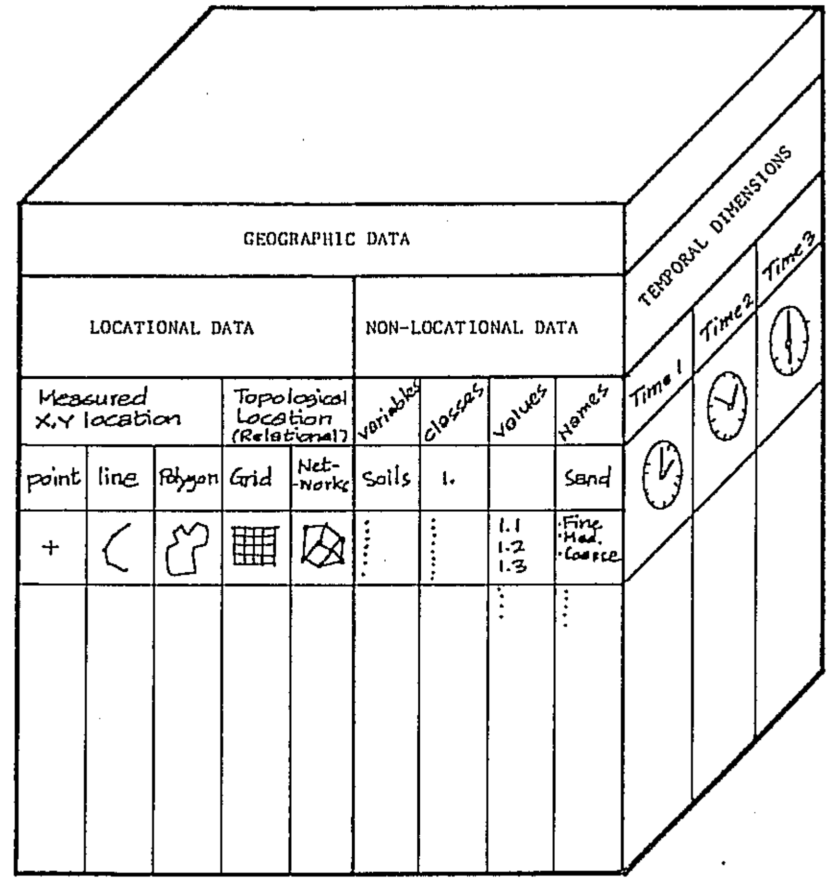


FIG. 2 THREE CONCEPTUAL COMPONENTS OF A GEOGRAPHIC BASED INFORMATION SYSTEM

loosely to the overlapping stages of activity involved in development of a region, state, or nation.

Graphics, and more specifically maps, have historically been a structure for most disciplines concerned with spatial analysis. For the past 4,000 years various cultures have used graphic symbologies to represent spatially distributed phenomena. Graphics in the form of maps have historically provided us with the useful medium for record keeping, conceiving of ideas, analyzing concepts, predicting the future, developing decisions about geography, and finally, communicating spatial concepts to others.

Today, geographic information systems and computer graphics are providing a similar framework for conducting spatial analysis. However, there exists a tremendous gap between the levels of technology that are available for geographic and graphic analysis and the various individuals and organizations who need to use it. It is felt that in order to advance this technology into the application environment, it is necessary to develop clear and concise information for GIS systems, the spatial data entities which they contain, and the various manipulation techniques which are possible from such systems.

In developing this information, it is useful to refer and keep in mind the historical methods of manual mapping as a benchmark and context for developing a common spatial information language. Therefore, this paper will from time to time describe functions performed by automated techniques in terms of this manual counterpart.

Why Use an Automated System?

Proponents of the automated geographic system technology point out a whole series of advantages which are enumerated in the literature. They include the following:

1. Data are maintained in a physically compact format (i.e., the magnetic file).

2. Data can be maintained and extracted at a lower cost per unit of data handled.

3. Data can be retrieved with much greater speed.

4. Various computerized tools allow for a variety of types of manipulation including map measurement, map overlay, transformation, graphic design, and data base manipulations.

5. Graphic and nongraphic (i.e., attribute information) can be merged and manipulated simultaneously in a "related" manner.

6. Rapid and repeated analytic testing of conceptual models about geography can be performed (i.e., land suitability/capability). This facilitates the evaluation of both scientific and policy criteria over large areas in short periods of time.

7. Change analysis can be efficiently performed for two or more different time periods.

8. Interactive graphic design and automated drafting tools can be applied to cartographic design and production.

9. Certain forms of analysis can be performed cost effectively that simply could not be done efficiently if performed manually (i.e., digital terrain analysis, calculations such as slope aspect, sun intensity, watershed, overlay analysis of multiple sets of complex polygon map sheets, etc.)

10. There is a resultant tendency to integrate data collections, spatial analysis, and decision-making processes into a common information flow context. This has great advantages in terms of efficiency and accountability.

Along with all of these advantages there are also a series of disadvantages that must be evaluated carefully before the user invests in the development or acquisition of such a system. They include the following:

1. The cost and associated technical problems of converting existing geographic records into an automated file (i.e., digitizing, scanning, or related data conversion);

2. The large amount of technical as well as financial overhead which is required to maintain automated files (i.e., computer, skilled technicians, software maintenance, etc.);

3. High cost of initial acquisition of systems; and

4. Marginal cost benefits in certain application areas.

Current Types of Systems

Within the past five years, GIS systems have begun to surface which have real and practical benefit to a variety of geographic problems. These systems are just beginning to be adapted to actual information flows and processes within governmental and private organizations. At this point, it is a very delicate technology in the sense that there has been considerable investment of various resources, and there is only beginning to emerge evidence of successful applications which make a difference.

At present, there are several categories of information technology which tend to be associated with vendor or government developed hardware/software systems. They generally fall into the following categories:

A CLASSIFICATION OF SOFTWARE COMPONENTS
COMMONLY USED IN GEOGRAPHIC INFORMATION SYSTEMS[1]

Jack Dangermond
Environmental Systems Research Institute
380 New York Street
Redlands, California 92373

Introduction

The purpose of this paper is to provide a general context for understanding the various software components which are commonly used in Geographic Information Systems (GIS). This is done by describing graphically and with narrative the actual analytic and data processing functions which are commonly performed in GIS systems. These functions are generic in nature and relate to no one system.

This paper initially presents a background of why these systems have evolved, the basic data types which are contained within them, and a series of concepts associated with spatial information language. This is followed with a description of GIS techniques including: map automation and data base creation· analytic manipulation techniques; data base manipulation techniques· and graphic manipulation techniques. Finally, a summary outline of the actual application of this technology is provided.

Background

In the past 20 years, a host of professions have been in the process of developing automated tools for the efficient storage, analysis, and presentation of geographic data. These efforts have apparently been the result of increasing demands by users for data and information of a spatial nature. This rapidly evolving technology has come to be known as "geographic information systems." The uses of this technology are vast and cross cutting across virtually all professions. This is well illustrated in figure 1 which is a matrix diagram indicating the relationship between a series of selected responsibilities which normally involve the use of geographic data and general tasks functions supported by geographic information systems (i.e., maintenance of geographic records, graphic display statistical reporting, planning, management, and policy making). The responsibilities listed from top to bottom of the chart correspond

---

1 Reprinted from Peuquet, Donna, and John O'Callaghan, eds. 1983. Design and Implementation of Computer-Based Geographic Information Systems. Amherst, NY: IGU Commission on Geographical Data Sensing and Processing.

FIGURE 1 - SELECTED RELATIONSHIP BETWEEN RESPONSIBILITIES & GIS APPLICATIONS

		GENERAL TASK FUNCTIONS SUPPORTED BY GEO. INFO.					
Responsibilities	Examples of Data	Maintenance of Geographic Records (i.e., Operational & Inventory Data)	Graphic Display (Drafting)	Statistical Reporting	Planning	Management	Policy Making
Territorial Control	Boundary Survey	0	0				
Natural Resource Exploitation	Geophysical, Topo, Geology, Vegetation, Soils	0	+		0		
Taxation & Ownership Monitoring	Cadastral Surveys & Tax Records	0	0	+			
Land Use/ Infrastructure Planning	Opportunities & Constraint Data				0		
Land Use Zoning	Zoning Records	0	+		0		
Land Use/ Infrastructure Design/ Construction	Engineering Surveys/ Data		0		+		
Construction Facility Record Keeping & Management	As Built Drawings of Utilities, Buildings, etc.	0	0		+		
Development Measurement	Land Use Survey	+	+	0	+	0	0
Census of Statistics	Population, Housing, Health & Economic Data		+	0	0		0
Event Monitoring	Police/Fire Statistics		+	+	+	0	0
Natural Resource Management	Ongoing Forest Record Keeping	0	0		0	0	
Monitoring of Environment	Wildlife, Vegetation, Air, Soil, Water, etc.	0	+	0	0		0

0=Major
+=Minor

which fall in a certain window which are of a certain vegetation type and over a certain size." Typically, the output of such a search is displayed graphically on a screen, printed out in alphanumeric form on a hard copy unit. and if desired, plotted out on a plotter.

2. Map Generalization -- Figure 10 refers to four basic types of map generalization. This series of generalization tools is most frequently used when map scales are changed. They are described below.

a) Line coordinate thinning - This is a technique for reducing the number of coordinates defining a given line.

b) Dropline - This is a technique wherein a polygon boundary separating two polygons of similar characteristics can be dropped and the remaining line segments making up the two previously separate polygons can be rejoined together to form a new polygon unit comprising the area of the two previously separate polygons. This is often a technique used in Integrated Terrain Unit Mapping (ITUM) where multiple attributes are assigned each polygon and when single maps are desired for extraction from the data base. To accomplish this function, a check is made of the characteristics on the right and left side of each polygon line segment to determine which line segments should be dropped.

c) Edgematching - Edgematching consists of a series of procedures for bringing together a large number of map sheets and compositing them into one continuous map. Problems which must be resolved are: joining lines and polygons from adjacent maps, matching of the boundaries between the maps, and dropping the lines which separate polygons having the same characteristics. It should be kept in mind that although automated techniques are useful in this exercise, basic errors made on the cartographic manuscripts input into the systems are often the source of great problems in making clean edgematching occur.

d) Polygon thinning - This is a similar exercise to item "2a" above, except that the problem is somewhat more complex because when doing polygon thinning, vertices thinned for one polygon must be consistent with vertice thinning on adjacent polygons; otherwise, lines which are shared in common will have sliver errors, gaps, and overlaps due to indiscriminate selection of the vertices which can be dropped. Thinning of polygon arc's instead of complete polygon entities can resolve this problem.

3. Map Abstraction -- Map abstraction is closely associated with map generalization but involves five different forms of technology. These are graphically presented in figure 11 and described below:

a) Calculation of centroids;

b) Automatic contouring from randomly spaced data;

c) Proximal mapping (Thiessen polygons);

d) Reclassification of polygons; and

e) Conversion of x,y coordinate location data to a regular and uniform grid.

4. Map Sheet Manipulation -- Figure 12 presents a series of techniques which manipulate the x,y coordinates for a given map sheet. They include the following:

a) Scale change - This often involves usage of map generalization tools presented in figure 10 in conjunction with actual scale modification.

b) Distortion removal - This is done both using rubber-sheeting techniques as well as linear transformation.

c) Projection changes - Recent advances by the USGS and others have resulted in software which can conduct virtually any type of geographic coordinate and projection changes.

d) Coordinate rotation and translation - This function involves the alteration of coordinate sets either through their rotation or shifting so as to match correctly overlapping or sets of adjacent coordinates (including key reference points).

5. Buffer Generation -- Buffer generation involves the creation of new polygons from points, lines, and polygon features within the data bank (see figure 13). Specifically from a given point or series of points, circular as well as square buffers can be calculated. Likewise, from a string of points (i.e., a line) both narrow and broad buffers can be calculated and, in cases where these line buffers overlap, new polygons can be formed.

Polygons can be generated both around the exterior perimeter of an existing polygon as well as within the interior of a polygon, resulting in a dual set of polygons measuring horizontal distance relationships relative to a given geographic element.

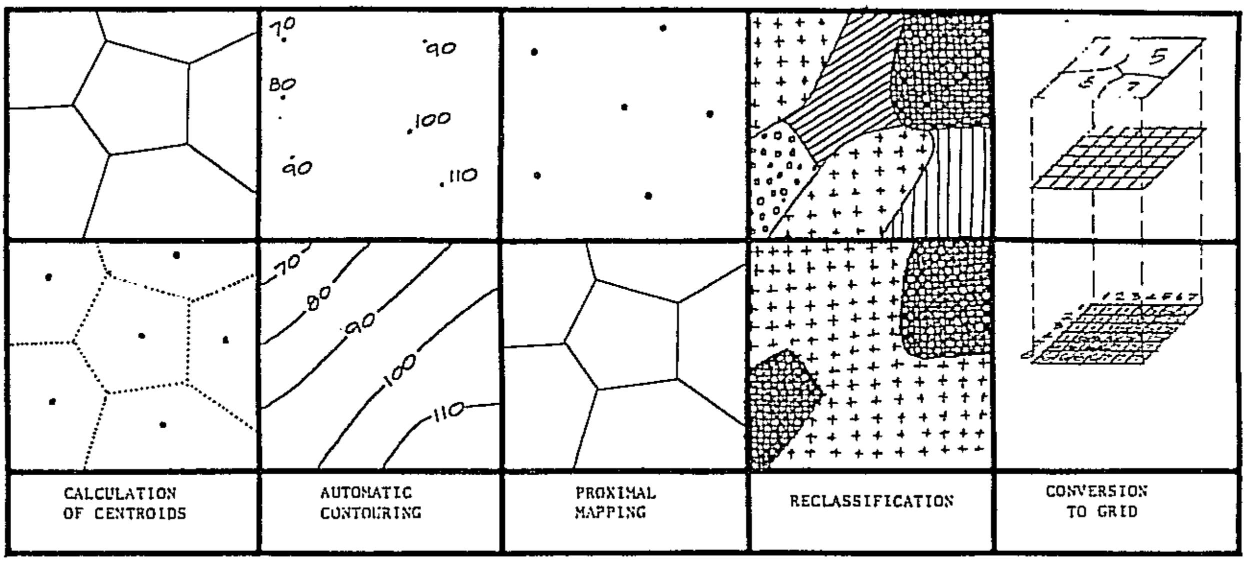


FIG.11 MAP ABSTRACTION

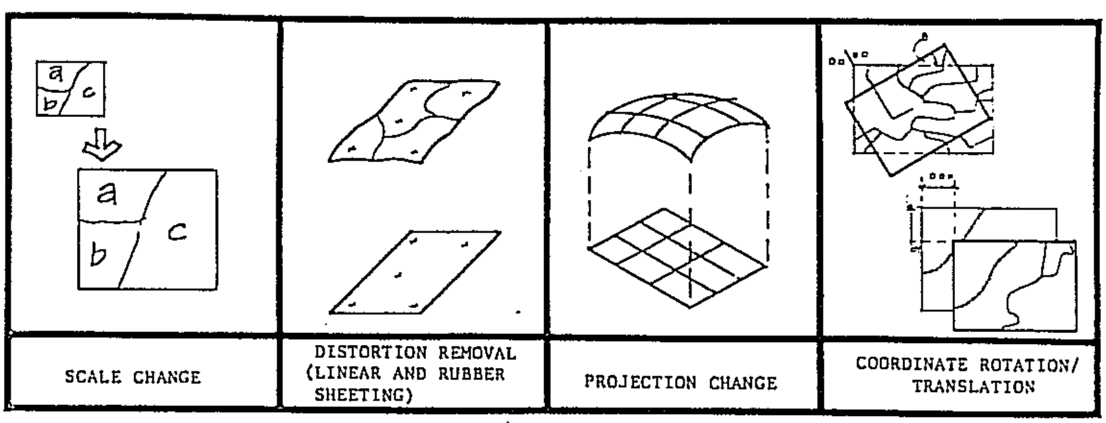


FIG.12 MAP SHEET MANIPULATION

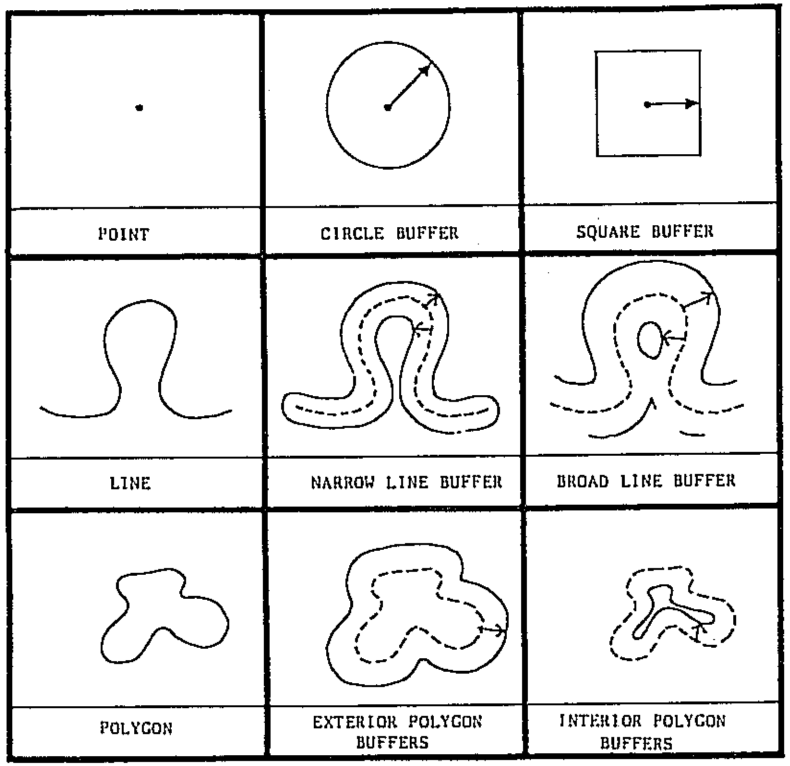


FIG. 13 BUFFER GENERATION

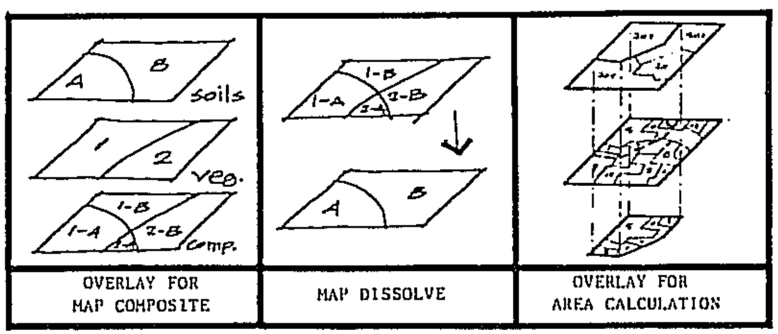


FIG.14 POLYGON OVERLAY/DISSOLVE

6. Polygon Overlay and Dissolve -- Polygon overlay and dissolve techniques involve the compositing (integrating) or extracting (dis-integration) of multiple maps (two or more) in order to create a new data set. These are shown in figure 14 and described below:

   a) Polygon overlay for creating new files - In the case of polygon overlay, a new data set is created, containing new polygons created from the intersection of the boundaries of the two or more sets of separate polygon layers. In addition to creating new polygons based on the overlay of the multiple layers, these polygons are assigned multiple attributes (i.e., the attributes which were attributed to each separate overlay before the composition occurred). The mathematical overlay of these maps is performed for the purpose of area and measurement, as well as multiple attribute (multiple parameter) modeling exercises. This involves the weighting of various parameters and their subclassifications in order to create "model interpretations"; an example of such an interpretation is the generation of land suitability/capability assessments.

   b) Map dissolve - Map dissolve function involves the inverse of polygon overlay. That is, the ability to extract from a multiple attribute polygon file single attributes, both by attribute description as well as locational definition. This function is the same as shown on figure 10 -- "Dropline." Dropline or map dissolve can be done both "graphically" (in which case the line is deleted on the output map only) or by actually altering the file by dropping the coordinates of the line segment separating the polygons.

   c) Polygon overlays for area calculation - A second type of polygon overlay is typically performed when the areas for a given data layer (for example, land use) need to be calculated and summarized within a second layer of polygons such as census tracts. The resulting output is the summary of statistics (i.e., land use areas by census tracts).

7. Measurement -- The four most common types of measurement tasks involve points, lines, and polygons and volumes. They are presented on figure 15 and outlined below:

   a) Points - The two typical measurement activities associated with points are enumeration of total number of points and also enumeration of total number of points falling within polygons. The latter technique involves the use of a "point-in-polygon" routine which

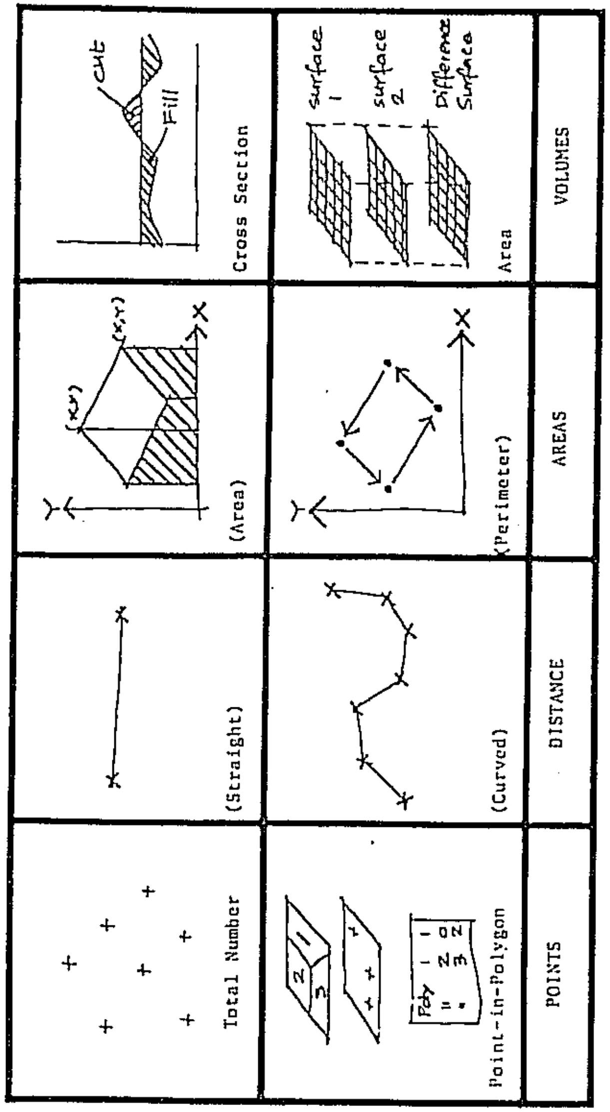


FIG. 15 MEASUREMENT

effectively counts up the number of various types of points falling within selected polygon types (i.e., crime incidents within police patrol districts).

b) Line measurement - Two basic forms of line measurement are point to point and measurement along a curvilinear line.

c) Measurement of areas - The two basic types of area measurement are: the area of a polygon and the perimeter of a polygon area.

d) The fourth category of measurement involves volumetric measurement which is performed either through a cross-section technique or through overlays of multiple surfaces (i.e., before grading, after grading, and difference computation).

8. Grid Cell Analysis -- Figure 16 illustrates five examples of basic analytic activities that are performed using grid cell data. These techniques are similar to the types of map analysis that are performed in x,y coordinate data structure but have more generalized spatial resolution. It should be pointed out that the grid cell technique for map manipulation is typically much more efficient both in data storage as well as in the operation of the analytic tasks. The five are described in the following pages:

a) Grid cell overlay - This involves the development of Boolean type map overlays in order to create composite maps. Typically, this involves the weighting of selected classifications for multiple map layers, resulting in a numeric surface of composite data, expressing the users interpretation. This surface is then typically displayed using computer graphics.

b) Area calculation overlay - This is a similar type of polygon overlay as was shown in figure 14, except that it uses a gridded technique, rather than a polygon technique.

c) Search radius aggregation - This is a technique for moving a circle with a specified radius across a grid cell by cell searching out from each cell. This routine counts up the number of events, activities or related geographic phenomena that fall within the circle relative to each grid cell and stores that enumerated value in a new grid array. The output from such a computation is a frequency/accessibility array surface.

d) Distance calculation - This is a routine similar to the previous one, in that, each cell is examined relative to its adjacency to other phenomena in surrounding cells.

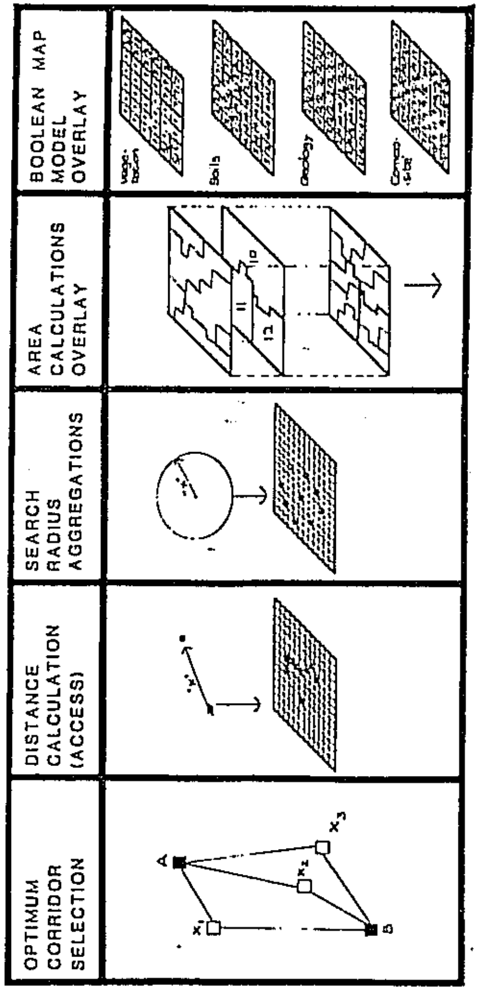


FIG.16 GRID CELL TECHNIQUE

In this case, however, the computation applied is a distance calculation to the nearest event or geographic characteristic which is defined by the user. A typical application is calculation of the distance from each cell to a stream network, road network, etc. The output from the analysis is the generation of a file of accessibility.

e) Optimal corridor selection - This analysis involves the computation of a least cost between two points. Specifically, a friction surface is computed away from the origin and also away from the destination. These friction surfaces are numerical summaries of cost based on an original input grid of costs computed using the grid weighted overlay technique described above. This original map (actually surface of grid data) expresses for each cell the cost to move across it. These costs can be economic costs, environmental costs, social costs, etc., and are summarized based on multiple parameter overlay. After computing the two friction maps (origin and destination), they are overlaid and the resulting surface is an optimum route selection with gradient computations of increased costs away from the optimum.

9. Digital Terrain Analysis — Digital terrain analysis involves the computation of a variety of outputs from a digital elevation model (see figure 17). There are various forms of digital elevation models and therefore different forms of actual analysis that can be performed. The following is a description of the most common of these analyses:

a) Visual display - These consist primarily of cross-sections and three-dimensional views.

b) Interpolation/contouring - Typically, this involves the ability to take random or regularly spaced data, generate a grid or other structured model framework on which automated contouring can occur. Based on this interpolated model, a contouring program actually computes and subsequently draws out the contour lines.

c) Slope/aspect/sun intensity - This involves primarily the calculation of slope aspect and in some cases, the relationship of this slope aspect to solar radiation in the form of sun intensity.

d) Watershed computation - This involves the computation of the watershed boundaries, drainages, topographic pits and peaks and, in some cases, this data is used to compute surface flow.

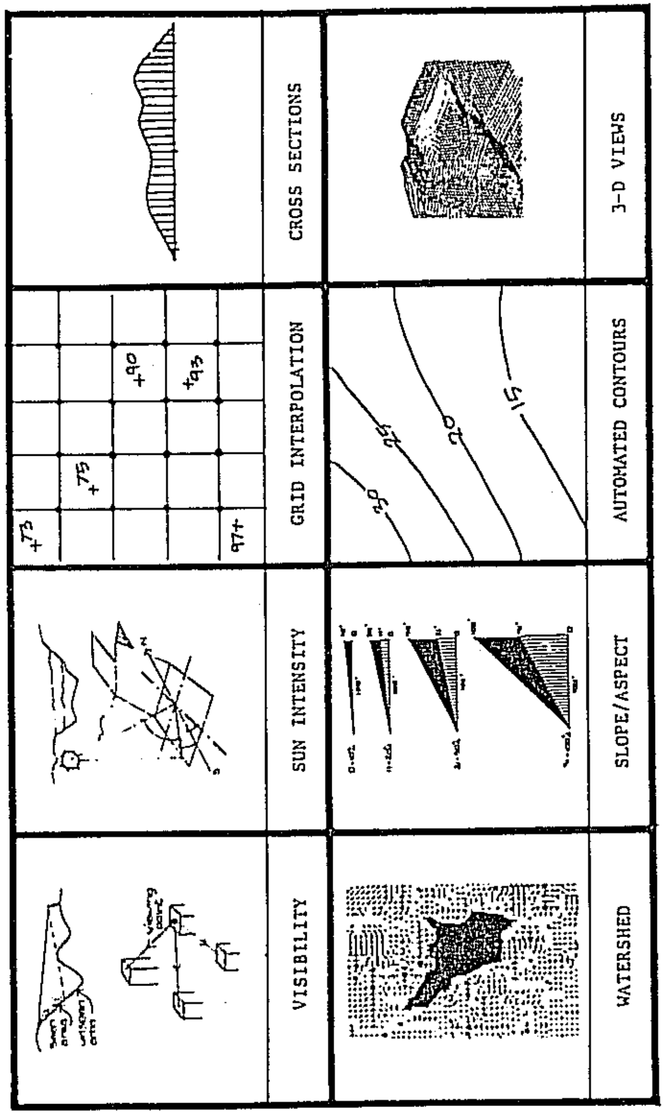


FIG. 17 DIGITAL TERRAIN ANALYSIS

e) Visibility - This analysis is primarily used in conjunction with determining what can be seen and not seen from a given x,y and z coordinate. Output is in the form of a grid of values indicating seen and nonseen area values.

10. Output Techniques -- Figure 18 illustrates the basic four output formats from a GIS system. They are as follows:

a) Hard copy maps;

b) Statistical tabulations;

c) CRT interactive displays (both graphics and attributes); and

d) Computed data files resulting from various manipulations generated on the geographic data.

Data Base Manipulation Techniques for Location Data

This section outlines the technologies which are commonly used in association with the maintenance and manipulation of the data base. Many of the software techniques used for spatial data base management are the same techniques used for initial data automation and in analytic manipulation of the data base. They are simply used differently for different purposes (i.e.. for making routine data base adjustments). These functions are outlined as follows:

1. Basic file creation and update

2. File management

3. Basic search, retrieval and listing/display

4. Query of selected attributes

5. Query of selected attributes by geographic area

6. Windowing manipulations

7. Edgematching of the files

8. Grouping of multi-map modules for larger area contiguous map measurement and display, etc.

9. Conversion from x,y coordinates into grid and also from grid into x.y coordinates.

In addition to these spatial or locational data management functions, GIS systems must also address the host of typical attribute related DBMS

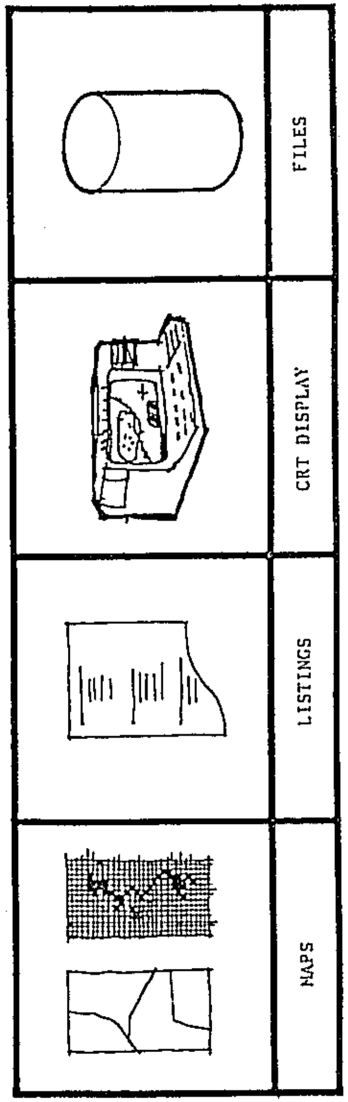


FIG.18 OUTPUT FORMAT

functions which are beyond the scope of this paper.

Graphical Manipulation Techniques

Within the past five years, substantial efforts have been made to adapt the CAD/CAM interactive graphic capabilities to automated cartography. These types of technologies, although working with many of the same data elements, are typically not classified in the same context as the more analytic geographic information systems. They are primarily focused on the creation, storage, manipulation and display of a variety of graphic map symbolisms and. in fact, treat the geographic elements contained in a GIS system as graphic files, rather than topologically correct analytic files. These systems have great use, particularly in the utility/engineering basemap and related tax/cadastral map environment. They are also finding considerably high acceptance in the photogrammetric field where stereo plotters are being interfaced in support of basic mapping activities.

Figure 19 indicates the typical labels and text, as well as graphic symbolism vocabulary, which is operable on virtually all of these interactive graphic systems. Functions include:

1. Text label data - These systems have multiple fonts with various sizes, scales. and a variety of capabilities including multiple font selection sizing, slanting, scaling, rotating, and in some cases, skewing along a curvilinear line.

2. Creation of a symbol library from a digitizer and association of symbols to an interactive menu.

3. Storage of symbol menus for interactive recall and display on maps and graphs. These symbols are typically associated with points. but they also involve symbols along lines and shading within polygon areas.

4. Some of the basic interactive functions are:

    a) Windowing, enlarging, and graphic symbol manipulation;

    b) Editing of cartographic elements (i.e., rotation, enlargement, transformation, and deletion);

    c) Graphical compositing of multiple graphic overlay files; and

    d) Overlay of a graphic reference grid.

5. Entry interactively (typically from CRT or digitizer) of points, lines, arcs, curvilinear lines, and polygons for various forms of graphical display.

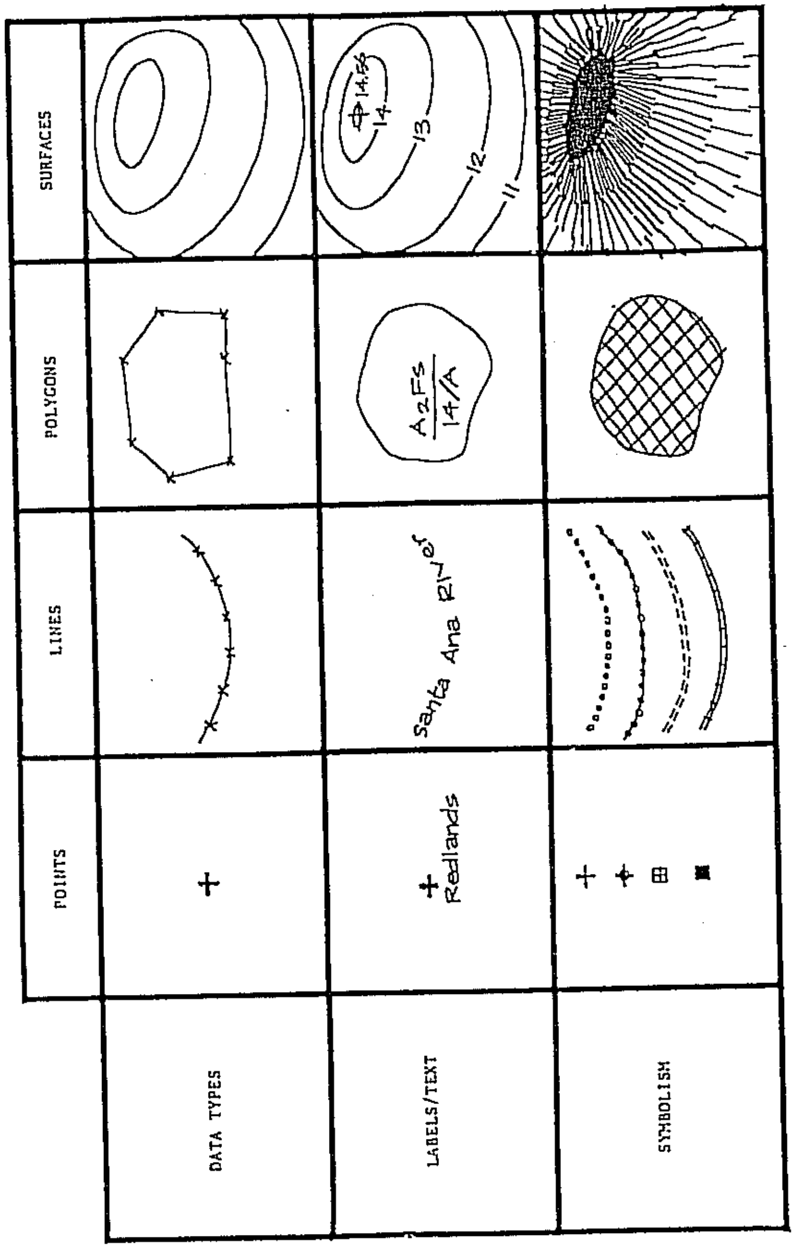


FIG.19 MAP SYMBOLISM

Common Applications

Although automated geographic information systems are really just beginning to have practical benefits, they have been adapted to many application areas. The following is a summarized list of some of these:

1. Engineering mapping
2. Automated photogrammetry
3. Subdivision design (cut/fill, street layout, parcel layout)
4. Tax mapping (cadastral mapping)
5. Highway mapping
6. Utility facility mapping and management
7. Geodetic mapping
8. Event mapping (accidents, crime, fire, facility breakages, etc.)
9. Census and related statistical mapping
10. Management of well log data
11. Land use planning and management
12. Environmental impact studies
13. Natural resource mapping and management (also included would be forest management, agricultural management. ecological and biological studies)
14. Routing of buses and similar service vehicles
15. Urban and regional planning
16. Route selection of highways/pipelines.

Summary

It is apparent that there is increasing awareness and interest in geographic phenomena (i.e., natural resources, population distribution, land uses, etc.). This interest is increasing the pressure to develop better ways to record, store, analyze, manage, retrieve, and display geographic information. In the future two decades, we will see a revolution in the integration of hardware and software analytic techniques.

As this trend continues, there will be a continued need to better understand and define these analytic functions that are possible and, even more importantly, a need to develop human skills for adapting this technology to very specific applications which will be so demanding of the knowledge GIS systems will inspire.