

GIS versus CAD versus DBMS: What Are the Differences?

Rashmi Sharma

Department of Computer Sc. & Engineering, M.J.R.P. University, Jaipur, Rajasthan, India
rashmi.sharma@hotmail.co.in

Abstract— Geographical Information System (GIS) is a hot topic. There is now a textbook and a journal devoted to the subject (Burrough, 1986; Coppock and Anderson, 1986) and it seems one can attend a conference on the subject practically any week. Although computer processing has been applied to geographical problems for more than a quarter of a century, it appears that there must be something different and more important about GIS that has caused so many diverse groups to take notice and invest so many resources. Yet there is much uncertainty about what the term GIS means. The purposes of this paper are to examine the alternative definitions of the dynamic field of GIS, and to attempt to distinguish GIS from other forms of automated digital analysis and mapping.

Keywords: *GIS, Database, OLAP, Datmining, CAD, GIS Management Tool.*

I. INTRODUCTION

While the origins of Geographic Information Systems usually have been traced to early work in computer mapping, there is a clear notion that the field is broader in scope today than simply automated map production (Dueker, 1979). The original work of the International Geographical Union Commission on Geographical Data Processing and Sensing resulted in a major two volume document that outlined the field and provided the basis for most of the subsequent efforts. In that compendium Tomlinson stated that GIS "is not a field by itself but rather the common ground between information processing and the many fields utilizing spatial analysis techniques" (Tomlinson, 1972). Based on Clarke's 1986 definition of GIS as "computer-assisted systems for the capture, storage, retrieval, analysis, and display of spatial data", it would appear that in the minds of many GIS is simply a catch-all for almost any type of automated geographic data processing (Clark, 1986). This paper argues that such vague definitions are doing a great disservice to the field by allowing the label of GIS to be applied to almost any software system that can display a map or map-like image on a computer output device. Four general approaches to defining GIS are found in the literature. All have some weaknesses.

II. THE PROCESS ORIENTED APPROACH

Process-oriented definitions, based on the idea that an information system consists of several integrated subsystems that help convert geographic data into

useful information were formulated originally in the early 1970s by Tomlinson and others (Catkins and Tomlinson, 1977). Logically, the entire system must include procedures for the input, storage, retrieval, analysis, and output of geographic information. The value of such systems is determined by their ability to deliver timely and useful information. Although the intentions of this process-oriented definition are quite clear, the application of the definition is far too inclusive to help distinguish GIS from computer cartography, location-allocation exercises, or even statistical analysis. By applying such a broad definition one could argue that almost any successful master's thesis in geography involves the creation of an operational GIS. Similarly, the production of an atlas also would seem to include all the necessary subsystems of a GIS. A process-oriented definition is, however, extremely valuable from an organizational perspective, as well as for establishing the notion that a system is something that is dynamic and should be viewed as a commitment to long term operation. Finally, any form of process-oriented definition of GIS emphasizes the end use of the information and, in fact, need not imply that automation is involved at all in the processing (Poiker, 1985).

III. THE APPLICATION APPROACH

A slight modification of the process-oriented approach yields a definition which categorizes GIS according to the type of information being handled. For example, Pavlidis' classification scheme includes natural resource inventory systems, urban systems, planning and evaluation systems, management command and control systems, and citizen scientific systems (Pavlidis, 1982). Applications in forestry may cut across several of these categories, but are primarily concerned with inventory, planning, and management. An area of greatly increased attention is the field of land records, or multi-purpose cadastre, systems that use the individual parcels as basic building blocks (McLaughlin, 1984). While defining GIS on the basis of applications may help to illustrate the scope of the field, it does not enable one to distinguish GIS from other forms of automated geographic data processing. Geographic information systems are independent of both scale and subject matter.

IV. THE TOOLBOX APPROACH

The toolbox definition of GIS derives from the idea that such a system incorporates a sophisticated set of computer-based procedures and algorithms for handling spatial data. Published works by Tomlinson and Boyle

(1981) and Dangermond (1983), for example, provide very complete delineations of the operational software functions that one should find in a full- featured GIS. Typically, these tools are organized according to the needs of each process-oriented subsystem (e.g., input, analysis, or output). The toolbox definition implies that all of these functions must be present and should work together efficiently to enhance the transfer of a variety of different types of geographical data through the system and ultimately into the hands of the end user. Therefore, even though they are important components of automated geography, neither digitizing image processing, nor automated mapping systems qualify as GIS because they do not possess all the necessary tools and do not provide the overall integration of functions. While check lists are very useful for evaluating different systems, they fail to provide a viable definition of the field.

V. THE DATABASE APPROACH

The database approach refines the toolbox definition of GIS by stressing the ease of the interaction of the other tools with the database. For example, Goodchild states, "A GIS is best defined as a system which uses a spatial database to provide answers to queries of a geographical nature. ...The generic GIS thus can be viewed as a number of specialized spatial routines laid over a standard relational data base management system" (Goodchild, 1985). Peuquet would agree that a GIS must start with an appropriate data model. Furthermore, she states that the success of the GIS will be determined by the efficiency that the data model provides for the retrieval, analysis, and display of the information (Peuquet, 1984). Some of the most important research in GIS is now concentrating on the design of optimal database management systems to link the geographic coordinate information with the attributes or variables associated with the geographic entities being represented in the system. From a conceptual viewpoint, questions relating to database design are more concerned with the performance of the system than with its essential functions. Even the recurring arguments over the merits of grid cell (raster) versus polygon (vector) based systems actually involve questions of representation rather than substance. There are comparable generic GIS operations in both types of software systems, and several sophisticated systems provide the user with the ability to transform data into either format to meet the needs of a specific task. While the technical issues surrounding database design are probably the most critical ones facing the field today, the database approach does not provide any better basis for defining the field than does the toolbox approach.

VI. INPUT AND OUTPUT

The confusion regarding the distinctions among different types of computer based geographic processing systems may be abated by examining the functions that such systems provide. The most appropriate way to accomplish this objective is to scrutinize the flow of data through the system and review the types of questions

the system is able to respond to at each stage of the process.

A. Data Capture

The first step in any form of automated geographic data processing consists of the transformation of analog models of objects on the Earth's surface into machine readable formats. Because any data that can be displayed on a map are the raw material for geographic research, maps represent the most common building blocks for all spatial data processing. Maps are two-dimensional representations of the Earth's surface; therefore, there exists a direct translation of all geographical entities into the equivalent elements of planar geometry, i.e., "points, lines, or polygons. Conceptually, the process of map digitization is an exercise in transferring these objects into a machine readable format. Another source of geographic data is the direct capture of images of the Earth. Once the digital representations of air photos or multispectral scanner data from satellites are mathematically converted into map-like products, they become suitable inputs into a GIS. Notwithstanding the technical problems involved with these transformations, all image-based data are simply spatially registered matrices of numbers and are, therefore, no different from any other grid cell layer of geographic data (e.g., that portrayed in a digital terrain model). In summary, geographical entities can be captured from maps or images and subsequently represented as points, lines, polygons, or a matrix of numbers. The most important questions in data capture relate to scale, resolution, and the efficient storage and retrieval of the spatial entities with respect to the ultimate use of the data. Bad decisions at the data capture stage are often difficult to correct at later stages in the process.

B. CAD – The Graphic Approach

In many cases maps and images are converted into a digital format simply for selective retrieval and display. A surprising amount of digital cartography is merely electronic drafting. For cartographic applications, graphical entities are often traced electronically from existing maps only to be selectively redrawn with additional annotation and other embellishments. These operations are analogous to those involved in electronic drafting and are being handled increasingly by computer aided design (CAD) systems. In essence, CAD systems handle geographic data in the same manner as photographic separations are used for the production of topographic maps. Different types of geographic features are placed on individual layers that are then combined and printed with different colors and line styles to generate the final product. Although the concept is the same, CAD systems provide much more versatility in terms of display functions than do their photographic counterparts, and are particularly beneficial for editing and updating.

While offering major improvements over photo-mechanical methods of map production, CAD systems have severe limitations when it comes to analytical tasks. In particular, it is difficult to link attributes in a database to

specific geographical entities and then automatically assign symbology on the basis of user-defined criteria (Cowen et al., 1986). For example, a CAD system could be used to create a graphical representation of a residential subdevelopment consisting of all the property lines separating individual land parcels. In fact, the CAD system would generate smooth curves for cul-de-sacs and would force all the lines to join perfectly. The system would also enable the cartographer to point to a particular land parcel and shade it with a pattern. The CAD system by itself, however, could not automatically shade each parcel based on values stored in an assessor's database containing information regarding ownership, usage, or value. In other words, a CAD system is merely a graphic system. This is not to suggest that such systems are not useful. In fact, a PC-based CAD system linked to World Data Bank II has provided the State Department's Office of the Geographer with the valuable capability to quickly generate base maps for any part of the world (Heivly and White, 1987).

C. Computer Mapping-Graphics Linked to a Database

Automated mapping systems provide, at a minimum, a rudimentary linkage between a database and a graphical display system. Even the earliest computer mapping systems, such as SYMAP, allowed one to automatically assign symbology to geographical entities on the basis of attributes or variables in the database. For example, in the land parcel illustration presented above, SYMAP could be used to assign differential gray shades to the parcels according to their assessed value or any other variable in the assessor's files. Theoretically, by changing class intervals and symbolism, it would be possible to generate an infinite number of maps with the same data.

Over the past two decades an extensive number of statistical mapping systems have been developed. Some of these have been incorporated into large scale statistical analysis programs that provide excellent interaction between mapping procedures and data manipulation operations. In many cases, these statistical systems treat choropleth or thematic mapping in much the same manner as they handle bar graphs and pie charts. The combined database components serve as an electronic filing cabinet that supports the query, sorting, and selection procedures, while the maps and graphs are just specialized output functions. Other current computer mapping systems provide very versatile formats, a wide range of symbolism which includes graduated symbols and dot maps, and publication quality fonts. Even though these modern computer mapping systems produce a much higher quality of output than the line printer maps of 20 years ago, they still are restricted to the functions of data retrieval, classification, and automatic symbolization.

While linking a database to the pictorial representation of geographical entities enables the researcher to address an extensive array of geographical questions, a computer mapping system still is not a GIS. Attempts to oversell mapping systems as GIS usually have

led to failure. For example, the Domestic Information Display System (DIDS) probably represents the most elaborate attempt to build a GIS around a choropleth mapping system. Even though it utilized the most advanced technology available at the time and had the support of numerous Federal agencies, DIDS was a failure as an information system. Although there were major organizational obstacles that contributed to its demise, the failure of DIDS was the result of its inability to provide appropriate answers to relevant spatial problems (Cowen, 1983). For example, an investigation of the influence of the Interstate Highway System on population growth in the 1970s was limited within the DIDS environment to county level estimates of population change and number of miles of interstate highway. This sophisticated choropleth mapping system could not integrate linear features for analysis, nor could it even properly incorporate such features for reasonable display.

By combining standard database management operations with automated symbol assignment, computer mapping systems provide a much better linkage between geographic information and display than do simple drafting or CAD systems. However, such systems fall far short of the type of capabilities that are now available with full featured geographic information systems.

VII. THE UNIQUE SCOPE OF GIS

Carstensen's recent investigation of the needs of a local government provides a basis for pinpointing the unique capabilities of a GIS (Carstensen, 1986). He approached the selection of an automated system on the basis of each candidate system's ability to determine which parcels of land met a set of six criteria for industrial site selection. The site had to be at least five acres in size, zone commercial, vacant or for sale, not subject to flooding, not more than one mile from a heavy duty road, and have no slope over ten percent. All of the information needed to select such a site could have been gathered from maps and searches at the appropriate local offices. The important question from the standpoint of geographic data processing and the field of GIS is the determination of whether the information could be generated automatically from digital representations of the relevant maps.

If the information for each parcel of land already existed in a database, then a standard database management system (DBMS) would have been able to deliver a list of addresses of the parcels that met all six criteria. A computer mapping system could have retrieved the same parcels and generated a resultant map. It is interesting to note that the street addresses might very well have been more useful in the decision process than a map. However, for even a moderate size area, either of these solutions would have required more manual effort to build the database than would have been justified for this single problem.

The dependency on manual creation of a database provides the basis for distinguishing a GIS from a computer mapping system. One could expect a full featured GIS to support the entire creation of the database, as well as the storage, retrieval, analysis, and report generation required to select the appropriate subset of geographic entities. For example, by utilizing a GIS, the size of each parcel would have been calculated automatically from the boundary coordinates, the type of zoning for each parcel would have been determined from the overlay of a zoning map, and the ownership status would have been updated automatically from transactions at the assessor's office. Inclusion in a flood-prone area would have been determined by another overlay created from maps of water bodies and topography. The same sources would have been used to determine the slope. Finally, the distance to different types of geographical entities could have been calculated from existing map inputs. In every case, variables or attributes relating to each parcel would have been created from other layers of geographical information. Most significantly, the GIS actually would have created new information rather than just have retrieved previously encoded information. This ability to both automatically synthesize existing layers of geographic data and to update a database of spatial entities is the key to a functional definition of a GIS.

A. Spatial Search and Overlay

It is important to note that of all of the operations that commonly are included in GIS toolboxes, spatial search and overlay are the only ones unique to GIS. Furthermore, it can be illustrated that most spatial searches are merely special forms of the overlay process. For example, in order to identify all of the parcels located within a mile of heavy duty roads, one would generate a buffer zone, or polygon, one mile wide around each such road. A polygon overlay algorithm would then be used to identify which parcels fell within these polygons. The emphasis of the GIS operations must be on the integration of different layers; not their creation. Concentration on the integration process results in the classification of the digitizing step as one that simply preprocesses maps into machine readable formats. Cartographic systems reverse the digitizing process by converting digital information into an analog format. Whereas digitizing is an essential part of the GIS process, the cartographic output subsystem of a GIS is often a convenient by-product.

The calculation of slope, in Cartensen's example, emphasizes the need to combine different data structures in a GIS framework. Such functions are special cases of GIS operations. Slope is usually measured by analyzing the elevation of a particular cell with respect to its eight neighbors. Because the calculation of slope involves the manipulation of a matrix of numbers, the operation is analogous to the manipulation of a Landsat scene. In both cases it could be argued that information is being preprocessed in order to generate a layer of information that conforms to the needs of the GIS. The

incorporation of remotely sensed images into a GIS has led to an interesting debate regarding the interface between the fields. For example, Fussell et al. (1986) raised the following questions:

What will be the role of remote sensing vis-a-vis the current trend toward Geographic Information Systems (GIS) technology? Is our future role to be reduced to providing input to GIS activities?

By restricting the definition of GIS to those operations that integrate geographic information, a concise method of classifying systems and activities results. More importantly, the polygon overlay process provides a mechanism that places GIS into the broader context of geographical research. White (1984), a mathematician, lists five geographical questions that such a system should address. Each of the five is actually a variation on a single question, "What regions cover a given region?" Further, it should be noted that his fundamental question would also apply to points and lines because the only true geometrical lines on the mapped model of space are the boundaries of legally defined polygons. All other points and lines on maps are generalizations of polygons on the Earth's surface. Therefore, GIS operations are restricted to polygon or grid cell overlays.

B. The Geographical Matrix.

The functional definition of GIS that focuses on integration provides a link to the Geographic Matrix that Brian Berry (1964) proposed almost 25 years ago. Berry suggested that all geographical information could be arranged in a matrix of infinite dimensions in which the columns are places and the rows are characteristics of the places. In such an ordering, regional analysis would involve the detailed study of a particular column of the matrix. In other words, if all the world were partitioned into a discrete, non-overlapping set of places, then a regional study would consist of a series of polygon overlays that would include every possible layer of geographical information concerning the area being studied. Systematic, or thematic, studies involve the detailed evaluation of a particular row (layer), such as land use, of the matrix. In a data processing context, these are reduced a further to a simple descriptive database retrieval from a flat file. Similarly, spatial association is a form of polygon overlay that covers a number of places (columns) or two entire rows of the matrix. If a third dimension representing time were added to the matrix, the resultant geographical cube would provide an extension to handle historical geography, sequent occupation, and change detection. Berry probably envisioned some of the technical and philosophical problems associated with his matrix when he stated:

Now assume [that] a whole series of characteristics has been recorded for a whole series of places. Perhaps we can imagine that complete "geographical data files" are available (whether such a dream may really be a nightmare is another topic) (Berry, 1964).

This is perhaps a warning to be heeded by the researchers who currently are involved in building global databases. Geographers have always considered themselves to be the great synthesizers of human and physical processes. The great interest in GIS would appear to reside in its technical basis for implementing integration methodologies. As Muller (1985) states:

The ah- f e i a s

- Undergraduate Geography: A Contemporary Dilemma, *The Operational Geographer*, Vol. 8, pp. 34-38.
- [16] Heivly, C., and T. White, 1987. Accessing Large Spatial Data Bases Via a Microcomputer, *Proceedings, Auto Carto VIII*, pp. 623-629.
- [17] Marble, D. F., et al., 1983. Geographic Information Systems and Remote Sensing, *The Manual of Remote Sensing*, Vol. 1, American Society of Photogrammetry, Falls Church, Virginia, pp. 923-958.
- [18] McLaughlin, J. D. 1984. The Multipurpose Cadastre Concept: Current Status, Future Prospects, Seminar on the Multipurpose Cadastre: Moderning Land Information Systems in North America, University of Wisconsin Institute of Environmental Studies, Madison.
- [19] Muller, J. C., 1985. Geographic Information Systems: A Unifying Force for Geography, *The Operational Geographer*, Vol. 8, pp. 41-43.
- [20] Pavlidis, M. G., 1982. Database Management for Geographic Information Systems, *Proceedings, National Conference on Energy Resource Management*, Vol. 1, pp. 255-260.
- [21] Peuquet, D. I., 1984. A Conceptual Framework and Comparison of Spatial Data Models, *Cartographica*, Vol. 2, pp. 66-113.
- [22] Poiker, T. K., 1985. Geographic Information Systems in the Geographic Curriculum, *The Operational Geographer*, Vol. 8, pp. 38-41.
- [23] Tomlinson, R. F., and A. R. Boyle, 1981. The State of Development of Systems for Handling National Resources Inventory Data, *Cartographica*, Vol. 18, pp. 65-95.
- [24] Tomlinson, R. F. (ed), 1972. *Geographical Data Handling*. IGU Commission on Geographical Data Sensing and Processing, Ottawa.
- [25] White, M. S., 1984. Technical Requirements and Standards for a Multipurpose Geographic Data System, *The American Cartographer*, Vol. 11, pp. 15-26.