(b)(3)-P.L. 86-36

Geographic Information Systems

[redacted]

Geographic Information System (GIS) technology is being introduced to the Department of Defense community. A Geographic Information System is a computer system that facilitates the storage and intelligent use of geographic data. This technology has typically been used in land-use planning and in city, state, and county governments. Several organizations within the National Security Agency are currently looking at how Geographic Information System technology can aid intelligence analysts in their jobs. This article will explain the historical development of these systems along with four basic components necessary to a Geographic Information System: data capture and input processing, data storage and management, data manipulation, and data display and output. It will also discuss and contrast raster and vector digital data storage formats; give guidelines for the necessary skills an employee must have for successful GIS implementations; discuss software packages leading in today's market; and give examples of how GIS technology could be applied to assist the National Security Agency in its SIGINT mission.

INTRODUCTION

Geographic Information System (GIS) technology is relatively new to the intelligence community. Several organizations within the National Security Agency are looking at how Geographic Information System technology can assist intelligence analysts in their jobs. This paper will (1) present a historical perspective on GIS technology and will explain the different components that make up a Geographic Information System; (2) show how GIS technology could be used to aid the National Security Agency in its SIGINT mission; (3) inform the reader about the leading GIS software packages in today's marketplace; and (4) list organizations within the National Security Agency that are looking at the different GIS software packages.

In order to begin a discussion on Geographic Information Systems, we must define what this new technology is all about. A Geographic Information System can be defined as a system that facilitates the storage and intelligent use of geographic data.[1] It can further be defined as "the use of sophisticated computer hardware and software to collate, store, manipulate, and process this geographic data."[2] Basically, a GIS consists of two main components. The first is an extensive database of geographic information that involves

1. Ross Babbage and Desmond Ball, eds., *Geographic Information Systems for Defense Applications*, Strategic and Defense Studies Centre, Research School of Pacific Studies, Australian National University, MacArthur Press, Australia, 1989, 3.

2. Ibid., 5.

positional data about the earth's surface as well as descriptive nonlocational data about these features. The second component is the set of programs or applications that enable the data held in the database to be input, accessed, manipulated, analyzed, and reported.[3]

HISTORY

The concept of a Geographic Information System was developed in the mid- to late 1960s. GIS development has its roots in two areas of study, one being the computer-aided design/facilities discipline and the other those areas developed with geographic analysis in mind. The latter include land-use planning and natural resources management. These systems, which evolved with the geographic analysis emphasis, should be what the National Security Agency is interested in. These systems go beyond providing computer "mirrors" of manually produced maps. They are designed with a spatial analysis perspective, with an emphasis on the overlay and study of relationships among land-use patterns. With this type of system, map creation is not an end in itself, but a by-product of more important database and analytical applications.[4] The ability to perform "analysis" on the combination of geographic data and intelligence-derived data should be of great interest to our community.

GIS FUNCTIONAL COMPONENTS

Because of the different origins these systems have, it's important to understand each of the functional components of a GIS. Each of the various GIS software packages emphasizes certain aspects of spatial data handling and deemphasizes or omits others.[5] The degree of emphasis placed on development of certain features depends on the marketplace a vendor is targeting. None of the current GIS software packages place their emphasis on the defense community; therefore, an understanding of the basic components of a GIS is needed. Understanding these underlying concepts will help users in their research of the technology and also in the development of applications for the defense community.

DATA CAPTURE AND INPUT PROCESSING

The first GIS component is Data Capture and Input Processing. In order to successfully implement a GIS, one must have data available for the study area. This includes the acquisition of the data, its digitization, and the appropriate tagging of attributes. Sources may include hard copy maps, existing digital data, imagery, and

3. Babbage and Ball, eds., *Geographic Information Systems for Defense Applications*, 5.

4. Glenn E. Montgomery, "GIS Software Selection," *GIS World*, July/August 1989, 44.

5. Barry J. Glick, "A Functional Definition and Categorization of Geographic Information Systems," *Spatial Data Sciences*, Virginia, 1988, 2.

tabular data. [redacted] The format, coordinate system, and geographic projection of the data must be known before input into a GIS. One must also be able to generalize the data and select only the amount of data that are necessary for a particular project. Digital data take up a huge amount of storage space on computer systems; therefore, the thinning and proper selection of this data is important. Most of the GIS software packages have algorithms and methods to aid in this process. One must also be aware of the amount of error that exists in the data being used and fully understand its limits and implications in project accuracy. Inherent data errors have been a traditional bottleneck in the development of GIS technology. Research is currently being done to help quantify this error.

DIGITAL DATA FORMATS

"All data that can be mapped have both a locational (x,y) and nonlocational (i.e., attribute) characteristics. These attributes can be both qualitative (e.g., the land use at a location) and quantitative (e.g., the elevation at the same location). In addition, the attributes at a location can be monitored through time. These three components – location, attribute, and time – represent the content of most GIS."[6] This information has to be somehow represented inside the GIS. Map data in this particular format are called digital data, and the process of getting the data into this format is called digitization. Digital data are represented in the computer as large sets of numbers, not as analog images. Two different data structures are commonly used to represent map data inside the GIS: raster and vector representations. Four fundamental types of geographic data have to be stored within a GIS: points, lines, polygons, and surfaces.[7] Raster and vector data representations use different techniques to store points, lines, polygons, and surfaces. These techniques will be discussed because they provide a basis for comparison of the two different types of digital data representations.

In raster representation, a grid data structure is used. A grid matrix is actually superimposed over the terrain such that the attribute information is collected within an array of grid cells.[8] Each grid cell is referenced by a row and column number and contains a value for the type of attribute that is being referenced. Points are represented by a single grid cell in this data structure. A line is a number of neighboring grid cells strung out in a given direction. An area is an agglomeration of neighboring cells (see fig. 1).[9] Grid cell representations define the interior of an area and imply the boundaries that define that

---

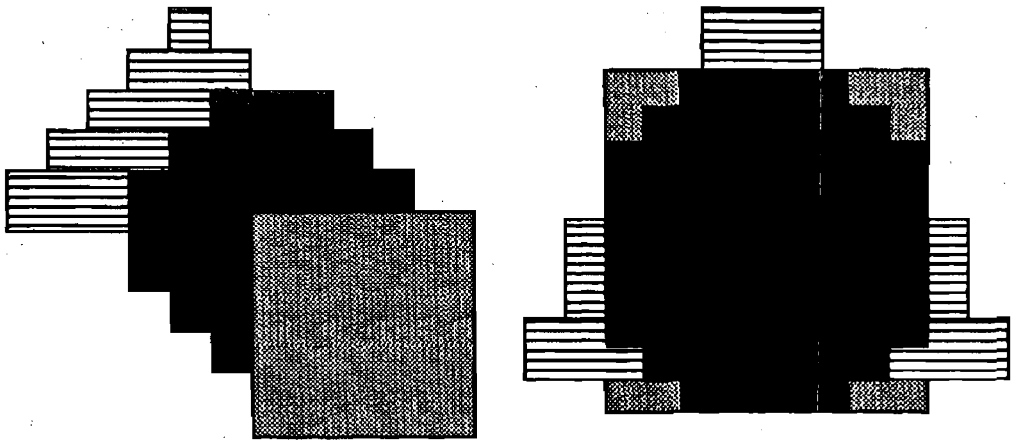
6. John R. Jensen, *Introductory Digital Image Processing – A Remote Sensing Perspective*, Englewood Cliffs, New Jersey: Prentice-Hall, 1986, 255.

7. K.J. Deuker, "Land Resource Information Systems: A Review of Fifteen Years Experience," *Geo-Processing*, 255.

8. Jensen, *Introductory Digital Image Processing – A Remote Sensing Perspective*, 263.

9. Joseph K. Berry, "Beyond Mapping," *GIS World*, July/August 1989, 34.

area. The selection of the grid cell size plays an important part in determining the precision in which the raster image reproduces the original. Smaller grid cell sizes depict the original more accurately than the larger cell sizes. A data storage increase is seen when small grid cell sizes are implemented. On the other hand, large grid cell sizes generalize the original data; that causes detail to be lost but does have the advantage of decreasing storage requirements of the data. One must clearly understand his data set, the storage capacity limits in his hardware, and the problem that he is solving in order to make a choice on grid cell size when using the raster representation for storing geographic data.



a. Triangle, circle, and square behind each other   b. Graphic overlay of the geometric shapes

Fig. 1. Raster Data Representation

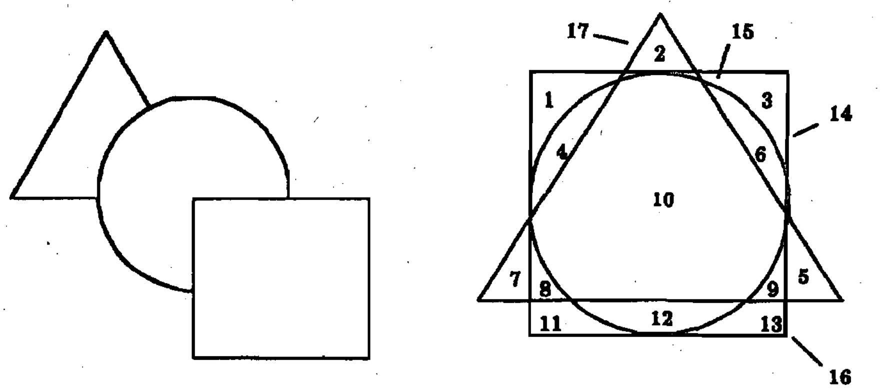
The second and more traditional method of data encoding is the vector representation format. This is based on the use of Cartesian coordinates such as latitude/longitude founded on the principles of Euclidean geometry. This is often called polygon or vector coding.[10] A singe point, or node, is represented by an (x,y) coordinate value. A line, or link, is represented by a string of (x,y) pairs. An area, or polygon, is represented by a closed loop of (x,y) points. Vector representations lend themselves nicely to applying graph theory involving topological relationships to express the relative location of various map elements. Figure 2 illustrates how topology is derived using the basic point, line, and polygon data structures.[11] By numbering the nodes and lines and also determining where they create closed loops that define polygons, one can use this to depict a geographic area. These points, lines, and polygons are associated among themselves to depict the spatial relationships. By adding (x,y) coordinate encoding at each of the points, one has a dual system for spatially identifying all elements of the map.[12] This concept of spatially

10. Jensen, *Introductory Digital Image Processing – A Remote Sensing Perspective*, 255.

11. Berry, "Beyond Mapping," 34.

12. Jensen, *Introductory Digital Image Processing – A Remote Sensing Perspective*, 258.

relating the (x,y) coordinates creates the basis for a topological data structure. Vector representations do just the opposite of grid cell representations. They imply the interior of an area while precisely defining its boundaries. Vector representations are good for spatial precision of map data and for the efficient storage of base maps. The coordinate space is also assumed to be continuous, not quantized as it is with raster storage. Vectors can represent an object exactly while raster representations generally define the object.



a. Triangle, circle, and square behind each other   b. Graphic overlay of the geometric shapes

Fig. 2. Vector Data Representation

When purchasing a GIS software package, one must keep in mind whether the raster or vector data representation format is supported. Educating oneself on the advantages and disadvantages of each data format, along with having an idea of the type of data that will be needed to solve a particular problem, will help narrow down the number of GIS packages that one needs to look at.

DATA STORAGE AND DATA MANAGEMENT

The second functional component of a GIS is the role of data storage and the management of these data. Once the data are encoded in their proper digital format, they must be stored in the GIS. Most of the GISs use a database mode to store these data. Today's Geographic Information Systems can use either a hierarchical, network, or a relational database model to achieve this component. The geographic information is arranged in files of related information, each file being called a layer. Each of these layers can be combined or overlaid upon each other to form new layers. These newly created layers form the basis for geographical analysis and can be queried to answer questions of interest to the user. Each of the attributes associated with this geographic data is stored in the database alongside its data structure and is queriable.

Since the amount of data needed is usually large, it's recommended that someone interested in setting up a GIS obtain the proper amount of disk storage. With the price of disk space decreasing and the storage capability on each hard disk drive increasing, the space requirements needed for these databases are becoming obtainable. One must always pay attention to the storage capacity demands of both the GIS package itself and the digital data being used for the project.

DATA MANIPULATION

The third functional component of a GIS is the role of data manipulation. To extract meaningful information from a GIS database, one must be able to query it and ask logical questions.[13] The leading database model used in GIS technology is the relational database. Relational databases have the ability to join different attribute tables to create new relationships among the data. This concept is important to the GIS in that the geographic data are stored in the database along with its attribute tables, which enhance the geographic data. This relationship helps make all features within a GIS queriable. When layers of data are combined, the attribute information for this material is carried along and also becomes queriable. As new information is added to the database and geographic layers are combined among themselves, the newly created geographic and spatial queries aid in performing actual analysis on the data. For example, a typical query could be to find all features of a defined type within a certain area. Another is to find all features that are adjacent to a specified feature. A third is to find all features that are a certain distance from another defined feature. A point and click type query could also be implemented (e.g., point to a road and tell me the attributes that are associated with that road).

DATA DISPLAY AND OUTPUT

The fourth and final functional component of a GIS is the role of data display and output. All GISs should include software for this capability, and they should provide means for both soft- and hardcopy output. The ability to interface with output peripherals such as wax thermal printers and plotters to be able to produce a map depicting the results of analysis is important, to say the least. Report generation and business graphic generation are necessary for some applications. Geographic Information Systems have been found to be lacking in this arena. Tabular data could be imported into desktop publishing packages or spreadsheet packages with little trouble to compensate for this deficiency. Spreadsheet packages could be used to produce graphical output displays such as histograms and time/frequency plots. One should also understand the types of maps he wants to produce when evaluating GIS software packages. Consider this list of maps when deciding what type of output capability a package has: polygon/cloropleth maps, contour/isarithmic maps, three-dimensional/perspective maps, and grid-cell maps.

---

13. John R. Jensen, *Introductory Digital Image Processing – A Remote Sensing Perspective*, 263.

SKILL REQUIREMENTS FOR GIS IMPLEMENTATION

Since the development of a GIS involves many different skills and skill levels, organizations must understand what they are getting themselves into when investigating GIS technology. GIS development requires the skills of many different people. According to Thomas A. Wikle's study on GIS educational needs, skills are needed in the following areas for successful GIS development efforts: computer cartography, database management, map reading, statistics, aerial photography and interpretation, computer programming, remote sensing, and knowledge of physical geography.[14] This list should warn the potential user of a GIS that the development of a GIS will not be an easy task. GIS vendors can help by offering extensive training courses in their software packages. But most require that the user be knowledgeable in the fundamental concepts behind GIS. Most GIS systems require the users to customize the software package for their specific needs. These are called toolkit systems versus the standard turnkey system. Even though these systems offer flexibility, their use does require the user to be knowledgeable, specifically in computer science. Having people work on the project who understand the digital map data that will be used, who have basic skills in geography, and who have computer programming skills will aid in the successful development of a GIS.

NSA APPLICATIONS OF GIS TECHNOLOGY

[redacted]

14. Thomas A. Wikle, "Survey Targets GIS Educational Needs," *GIS World*, June/July 1990, 90–91.

15 [redacted]

16. Ibid., 9.

[redacted]

LEADING GIS SOFTWARE PACKAGES

An enormous number of software packages in today's marketplace claim to be GIS. Vendors seem to think that if they say their product is a GIS, it will sell. A lot of these systems, however, do not have full GIS capabilities. Some products incorporate only one of the functional components of a GIS described previously. A few of the packages incorporate all four of the functional components that define a GIS. A few of the leading packages will be mentioned here so that the serious reader looking at GIS technology will be given a baseline of products to evaluate.

ARC/INFO by Environmental Systems Research Incorporated (ESRI) is probably the most traditional GIS software package around and is the current market leader. It implements a vector data structure. The package has excellent traditional GIS functionality that is fairly complete. It uses the INFO database by HENCO, Inc. It currently has a weak spatial-to-attribute database link, but new releases of the product will include interfaces to the following commercial databases: ORACLE, INGRES, and SYBASE. This weak database link will be eliminated with the release of these interfaces. It has incorporated the open systems concept of development. A tool kit is provided for users depending on their specific needs. A macro language is included in the package that enables customization to be done. A down side is that the package is very expensive to procure, averaging about $10,500 per copy for a complete copy of the software.

GEOVISION is another vector-based GIS that performs traditional GIS functions fairly well. It has a stronger spatial-attribute database link than the ARC/INFO package. It uses the ORACLE database package as its database interface. It also has a better user interface with it than ARC/INFO. Its main disadvantage is that it has a weak analytical tool kit. This is a major consideration when looking for a package to perform analytical functions.

(b)(1)
(b)(3)-50 USC 403
(b)(3)-18 USC 798
(b)(3)-P.L. 86-36

GRASS (Geographic Resources Analysis Support System) is a raster-based GIS package, developed by the U.S. Army Construction Engineering Research Laboratory (CERL). It allows for the storage, analysis, and display of maps, images, and other spatially referenced data. This product is distributed by DBA Systems, Inc., and its main advantage is that it is pubic domain software. DBA Systems does charge a fee – $1,500 for the basic product – for the service of GRASS distribution and support. Its big disadvantage is that it has an incomplete spatial database package. The product also has poor feature support. It is good for data overlaying and for displaying imagery data and is recommended as a first product purchase for GIS research and development work. It will give the new GIS user experience with a GIS without the investment in an expensive package.

ERDAS is a package that has limited GIS capabilities but is an excellent image processing system. It is a raster-based system that has a raster modeling subpackage called GISMO. It also has a raster/vector integration capability called the ERDAS – TARC/INFO live link. This allows the ERDAS satellite imagery to interact with vector data stored in the ARC/INFO GIS package. This link helps close the gap in its limited capabilities. These two GIS packages are special in that they can interact with one another; other packages currently don't integrate with one another. ERDAS is an expensive package, costing about $15,300 for the basic package, including the ERDAS – ARC/INFO Live Link.

There are several other GIS packages that are worth mentioning briefly. TYDAC Technologies Spatial Analysis System (SPANS) package stores its data in quadtree data structure. This is another, less common, data structure used for storing digital data and is supposed to be the happy medium between raster and vector storage methods. "Quadtrees optimize data storage by reducing vector or raster files to a factor of between one third to one tenth of their original size."[17] This package is mainly a PC-based system, but the company is in the process of porting its software to the UNIX-based IBM RISC 6000 workstation.

ULTIMAP by the Ultimap Corporation was designed solely for city and county planning purposes, which makes it hard to adapt it in the defense community. The system does have some good features, such as an ability to link into other external databases, that one might want to use within a GIS.

The last package worth mentioning is a group of software tools produced by Dynamic Graphics, Inc. The company provides a set of interactive software packages that perform surface modeling, slope analysis, coordinate transformation, and digital data importation. It is not a GIS, but it is a package that can be used well within the defense community for surface modeling applications and in conjunction with GIS systems.

17. TYDAC Technologies Corporation, "Technical Specifications for TYDAC Technologies Spatial Analysis System SPANS," 17.

FUTURE TRENDS IN GIS TECHNOLOGY

Advancements and research are continuously taking place in the GIS arena. Work is being done to improve user interfaces for these software packages. In addition, systems are advancing and moving towards point-and-click-type interfaces that involve windowing. Typical raster-based systems are moving towards integrating vector data, and the same is happening to the vector-based systems. The amount of available digital data is rapidly increasing for the continental United States and for the world. New digital products are being produced by the Defense Mapping Agency. Research is being done to further define and quantify the error that exists in data. This research will eventually offer the ability to inform the user about margin of error when analysis is being performed. Improvements are also being done by the various database companies to improve the quality of their software packages, which will obviously increase a GIS's power.

NSA ORGANIZATIONS INVOLVED IN GIS

[redacted]

CONCLUSION

This paper has attempted to introduce Geographic Information System technology to its reader. It has described the basic components that make up a GIS and provided the basic background necessary for one to begin researching the technology. It has listed the potential uses for GIS and described how it can be applied in assisting in solving current SIGINT problems. It concluded with presenting a market survey on the various products that are currently available. One can watch the enormous growth and effect the GIS industry is having on city, state, and country governments; in the land-use planning industry; and in a large number of environmentally related industries. Many potential uses exist for applying GIS technology within the defense community. The implementation of this technology could prove to be an effective means of assisting the National Security Agency in accomplishing its mission.

(b)(3)-P.L. 86-36

~~(FOUO)~~

[redacted]

BIBLIOGRAPHY

Babbage, Ross, and Desmond Ball, eds. *Geographic Information Systems: Defense Applications*. Strategic and Defense Studies Centre, Research School of Pacific Studies, Australian National University, MacArthur Press, Australia, 1989.

Berry, Joseph K. "Beyond Mapping." *GIS World*, Vol. 2, No. 4, July/August 1989, 34–49.

Deuker, K.J. "Land Resource Information Systems: A Review of Fifteen Years Experience." *Geo-Processing*, Vol. 1, 1979.

[redacted]

——— "A Functional Definition and Categorization of Geographic Information Systems." *Spatial Data Sciences*, Inc., Virginia, October 18, 1988.

Jensen, John R. *Introductory Digital Image Processing: A Remote Sensing Perspective*. Englewood Cliffs, New Jersey: Prentice-Hall, 1986.

Montgomery, Glenn E. "GIS Software Selection." *GIS World*, Vol. 2, No. 4, July/August 1989, 31–47.

TYDAC Technologies Corporation. "Technical Specifications for TYDAC Technologies Spatial Analysis System SPANS." Promotion package, Arlington, Virginia, 1989.

(b)(1)
(b)(3)-50 USC 403
(b)(3)-18 USC 798
(b)(3)-P.L. 86-36

Vendor Literature for the following software packages: ARC/INFO, DYNAMIC GRAPHICS, ERDAS, GEOVISION, GRASS, SPANS, and ULTIMAP.

Wikle, Thomas A. "Survey Targets: GIS Educational Needs." *GIS World*, Vol. 3, No. 3, June/July 1990, 90–92.