1 Introduction

A geographic information system (GIS) is a computer-based tool that is used to store and manipulate the geographic information. It also involves the mapping and analyzing things that exist and events that happen on earth. GIS technology integrates common database operations such as query and statistical analysis with the unique visualization and geographic analysis benefits offered by maps. These abilities distinguish GIS from other information systems and make it valuable to a wide range of public and private enterprises for explaining events, predicting outcomes, and planning strategies.

Primary concern of geographic information system is the management, analysis, and visualization / mapping of spatial data. An oftenly quoted definition of GIS in the literature has came from Burrough (1986). He stated that **"GIS is a powerful set of tools for collecting, storing, retrieving at will, transform and displaying spatial data from the real world", (Burrough , 1986)**. Data types used in GIS are; geographical and attribute data in association with each other. Generally, the geographical data are represented as points, lines or polygons and attributes are linked to it. The data types are being handled easily by using convential database management systems (DBMS). According to Aronoff **"Geographic information systems are computer-based systems that are used to store and manipulate geographic information"(Aronoff , 1995).**

GIS technology has provided an exciting potential for geographic information to be used more systematically and by a greater diversity of disciplines than ever before. The major challenges we face in the world today; overpopulation, pollution, deforestation, natural disasters, has a critical geographic dimension. Whether selecting a new site for business, finding the best soil for growing a particular crop, or figuring out the best route for an emergency vehicle, local problems also have a geographical component. GIS will give you the power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before. GIS is a tool used by individuals and organizations, schools, governments, and businesses seeking innovative ways to solve their problems.

Mapmaking and geographic analysis are not new, but a GIS performs these tasks better and faster than do the old manual methods. Before GIS technology, only a few people had the skills necessary to use geographic information to help with decision making and problem solving.

GIS has emerged as a huge discipline employing hundreds of thousands of people worldwide. GIS is taught in schools, colleges, and universities throughout the world. Professionals in every field are increasingly aware of the advantages of thinking and working geographically.

2 Functional Subsystems

A GIS has four main functional subsystems. These are:

Data Input Subsystem
Data Storage and Retrieval Subsystem
Data Manipulation and Analysis Subsystem
Output/ Display Subsystem

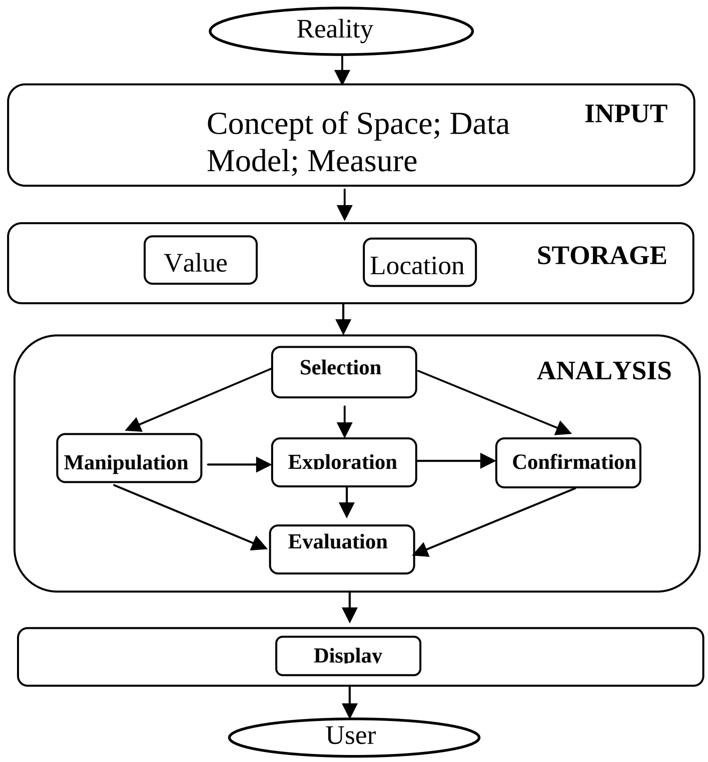

Figure 1 - Functions of GIS (after Anselin and Getis , 1992).

A **data input** subsystem allows the user to capture, collect, and transform spatial and thematic data into digital form. The data inputs are usually derived from a combination of hard copy maps, aerial photographs, remotely sensed images, reports, survey documents, etc.

Input subsystem functions in three phases: conceptualization, modeling and measurement of reality. Conceptualization is the phase where humans organize and structure their concept of reality. Space can, for instance, be perceived as a set of infinite number of dimensionless points which form a continuum, where each point is characterized by a coordinate value and a vector of attribute (Goodchild , 1990). He recognized this as geographic reality. There are many other ways for perceiving the space as Frank (1990) has indicated that " other ways of perceiving the space are as thematic layers , as Euclidean geometry, as a specific set of areas, or as graphs like those often used in transportation problems".

The concepts used are often based on notions which cannot be directly implemented in a computer system. In order to analyze the real world's features its necessary to reduce its infinite complexity and continuous variation to discrete, observable, and measurable units. This process of discretization , cxonverting the complex geographic reality into a finite number of database records is referred as *spatial data modeling* which constitutes the second input phase. Third input phase includes the measurement of spatial objects according to the data model developed in the second input phase so called spatial data modeling phase.

Data Storage and Retrieval Subsystem

The **data storage** and retrieval subsystem organizes the data, spatial and attribute, in a form which permits it to be quickly retrieved by the user for analysis, and permits rapid and accurate updates to be made to the database. This component usually involves the use of a database management system (DBMS) for maintaining attribute data. Spatial data is usually encoded and maintained in a proprietary file format.

Data Manipulation and Analysis Subsystem

The **data manipulation and analysis** subsystem allows the user to define and execute spatial and attribute procedures to generate derived information. This subsystem is commonly thought of as the *heart of a GIS*, and usually distinguishes it from other database information systems and computer-aided drafting (CAD) systems.

Output/ Display Subsystem

The **data output** subsystem allows the user to generate graphic displays, normally maps, and tabular reports representing derived information products.

3 Components of a GIS

A working GIS integrates five key components: hardware, software, data, people, and methods.

3.1 Hardware

Hardware is the computer on which a GIS operates. Today, GIS software runs on a wide range of hardware types, from centralized computer servers to desktop computers used in stand-alone or networked configurations.

3.2 Software

GIS software provides the functions and tools needed to store, analyze, and display geographic information. Key software components are

Tools for the input and manipulation of geographic information
A database management system (DBMS)
Tools that support geographic query, analysis, and visualization
A graphical user interface (GUI) for easy access to tools

3.3 Data

Possibly the most important component of a GIS is the data. Geographic data and related tabular data can be collected in-house or purchased from a commercial data provider. A GIS will integrate spatial data with other data resources and can even use a DBMS, used by most organizations to organize and maintain their data, to manage spatial data.

3.4 People

GIS technology is of limited value without the people who manage the system and develop plans for applying it to real-world problems. GIS users range from technical specialists who design and maintain the system to those who use it to help them perform their everyday work.

3.5 Methods

A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organization.

4 GIS Data and Data Models

A GIS stores information about the world as a collection of thematic layers that can be linked together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to recording details of planning applications, to modeling global atmospheric circulation. The thematic layer approach allows us to organize the complexity of the real world into a simple representation to help and facilitate our understanding of natural relationships.

4.1 GIS Data Types

The basic data types in a GIS reflect traditional data found on a map. Accordingly, GIS technology utilizes two basic types of data. These are:

Spatial data

Describes the absolute and relative location of geographic features.

Attribute data

Describes characteristics of the spatial features. These characteristics can be quantitative and/or qualitative in nature. Attribute data is often referred to as tabular data.

The coordinate location of a forestry stand would be spatial data, while the characteristics of that forestry stand, e.g. cover group, dominant species, crown closure, height, etc., would be attribute data. Other data types, in particular image and multimedia data, are becoming more prevalent with changing technology. Depending on the specific content of the data, *image data* may be considered either spatial, e.g. photographs, animation, movies, etc., or attribute, e.g. sound, descriptions, narration's, etc.

4.2 Spatial Data Models

Traditionally spatial data has been stored and presented in the form of a map. Three basic types of spatial data models have evolved for storing geographic data digitally. These are referred to as

- Vector;
- Raster; and
- Image.

The following diagram reflects the two primary spatial data encoding techniques. These are vector and raster. Image data utilizes techniques very similar to raster data, however typically lacks the internal formats required for analysis and modeling of the data. Images reflects *pictures* or *photographs* of the landscape.

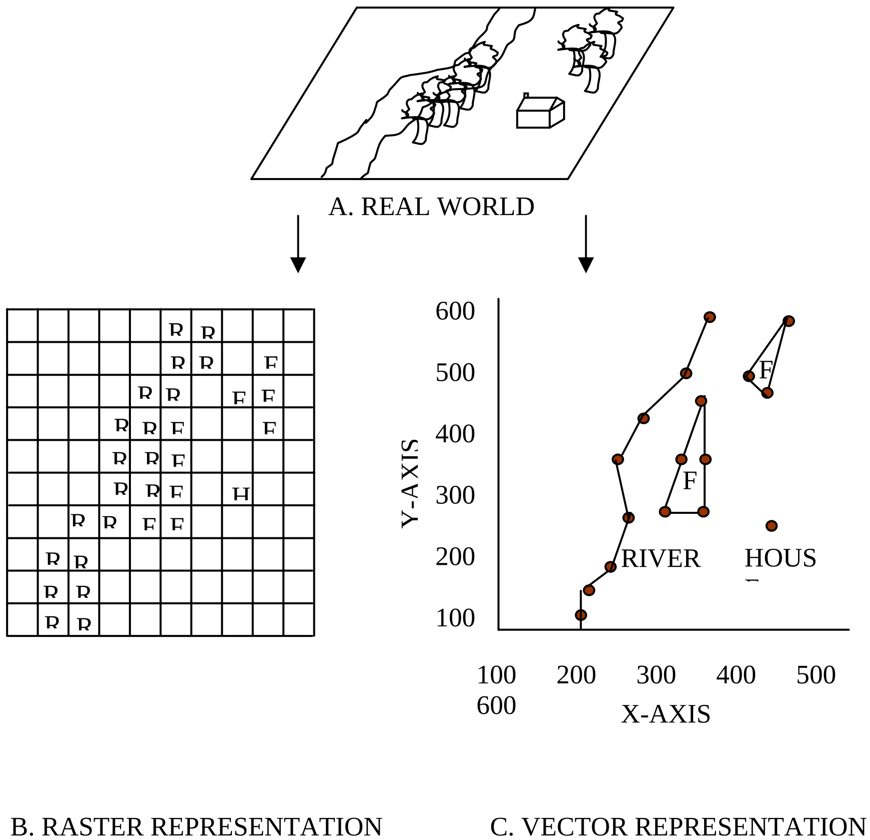

Figure 2: The Vector and Raster Data Model (Aronoff, 1995)

4.2.1 Vector Data Formats

All spatial data models are approaches for storing the spatial location of geographic features in a database. Vector storage implies the use of vectors (directional lines) to represent a geographic feature. Vector data is characterized by the use of sequential points or *vertices* to define a linear segment. Each vertex consists of an X coordinate and a Y coordinate.

Vector lines are often referred to as *arcs* and consist of a string of vertices terminated by a *node*. A node is defined as a vertex that starts or ends an arc segment. Point features are defined by one coordinate pair, a vertex. Polygonal features are defined by a set of closed coordinate pairs. In vector representation, the storage of the vertices for each feature is important, as well as the connectivity between features, e.g. the sharing of common vertices where features connect.

The **topologic data structure** is often referred to as an *intelligent data structure* because spatial relationships between geographic features are easily derived when using them. Primarily for this reason the topologic model is the dominant vector data structure currently used in GIS technology. Many of the complex data analysis functions cannot effectively be undertaken without a topologic vector data structure.

The secondary vector data structure that is common among GIS software is the **computer-aided drafting (CAD) data structure**. This structure consists of listing elements, not features, defined by strings of vertices, to define geographic features, e.g. points, lines, or areas. There is considerable redundancy with this data model since the boundary segment between two polygons can be stored twice, once for each feature. The CAD structure emerged from the development of computer graphics systems without specific considerations of processing geographic features. Accordingly, since features, e.g. polygons, are self-contained and independent, questions about the adjacency of features can be difficult to answer. The CAD vector model lacks the definition of spatial relationships between features that is defined by the topologic data model.

4.2.2 Raster Data Formats

Raster data models incorporate the use of a *grid-cell* data structure where the geographic area is divided into cells identified by row and column. This data structure is commonly called *raster*.

The size of cells in a raster data structure is selected on the basis of the data accuracy and the resolution needed by the user. There is no explicit coding of geographic coordinates required since that is implicit in the layout of the cells. A raster data structure is in fact a matrix where any coordinate can be quickly calculated if the origin point is known, and the size of the grid cells is known. Since grid-cells can be handled as two-dimensional arrays in computer encoding many analytical operations are easy to program. This makes tessellated data structures a popular choice for many GIS software. Topology is not a relevant concept with tessellated structures since adjacency and connectivity are implicit in the location of a particular cell in the data matrix.

Several tessellated data structures exist, however only two are commonly used in GIS's. The most popular cell structure is the regularly spaced matrix or *raster* structure. This data structure involves a division of spatial data into regularly spaced cells. Each cell is of the same shape and size. Squares are most commonly utilized.
Since geographic data is rarely distinguished by regularly spaced shapes, cells must be classified as to the most common attribute for the cell. The problem of determining the proper resolution for a particular data layer can be a concern. If one selects too coarse a cell size then data may be overly generalized. If one selects too fine a cell size then too many cells may be created resulting in a large data volume, slower processing times, and

a more cumbersome data set. As well, one can imply accuracy greater than that of the original data captures process and this may result in some erroneous results during analysis.

As well, since most data is captured in a vector format, e.g. digitizing, data must be converted to the raster data structure. This is called *vector-raster conversion.* Most GIS software allows the user to define the raster grid (cell) size for vector-raster conversion. It is imperative that the original scale, e.g. accuracy, of the data be known prior to conversion. The accuracy of the data, often referred to as the resolution, should determine the cell size of the output raster map during conversion.

Most raster based GIS software requires that the raster cell contain only a single discrete value. Accordingly, a data layer, e.g. forest inventory stands, may be broken down into a series of raster maps, each representing an attribute type, e.g. a species map, a height map, a density map, etc. These are often referred to as *one-attribute maps*. This is in contrast to most conventional vector data models that maintain data as *multiple attribute maps*, e.g. forest inventory polygons *linked* to a database table containing all attributes as columns. This basic distinction of raster data storage provides the foundation for quantitative analysis techniques. This is often referred to as *raster or map algebra*. The use of raster data structures allow for sophisticated mathematical modelling processes while vector based systems are often constrained by the capabilities and language of a relational DBMS.

This difference is the major distinguishing factor between vector and raster based GIS software. It is also important to understand that the selection of a particular data structure can provide advantages during the analysis stage. For example, the vector data model does not handle continuous data, e.g. elevation, the raster data model is more ideally suited for this type of analysis. Accordingly, the raster structure does not handle linear data analysis, e.g. shortest path, very well while vector systems do. It is important for the user to understand that there are certain advantages and disadvantages to each data model.

4.2.3 Image Data

Image data is most often used to represent graphic or pictorial data. The term *image* inherently reflects a graphic representation, and in the *GIS world*, differs significantly from raster data.

Most often, image data is used to store remotely sensed imagery, e.g. satellite scenes or orthophotos, or ancillary graphics such as photographs, scanned plan documents, etc. Image data is typically used in GIS systems as background display data (if the image has been rectified and georeferenced); or as a graphic attribute. Remote sensing software makes use of image data for image classification and processing. Typically, this data must be converted into a raster format (and perhaps vector) to be used analytically with the GIS.

Image data is typically stored in a variety of de facto industry standard proprietary formats. These often reflect the most popular image processing systems. Other graphic image formats, such as TIFF, GIF, PCX, etc., are used to store ancillary image data. Most GIS software will read such formats and allow you to display this data.

Source: http://www.interknowledge.com

Figure 3 - A typical image

4.2.4 Vector and Raster - Advantages and Disadvantages

There are several advantages and disadvantages for using either the vector or raster data model to store spatial data. These are summarized below.

Vector Data

Advantages :

Data can be represented at its original resolution and form without generalization.

Graphic output is usually more aesthetically pleasing (traditional cartographic representation);

Since most data, e.g. hard copy maps, is in vector form no data conversion is required.

Accurate geographic location of data is maintained.

Allows for efficient encoding of topology, and as a result more efficient operations that require topological information, e.g. proximity, network analysis.

Disadvantages:

The location of each vertex needs to be stored explicitly.

For effective analysis, vector data must be converted into a topological structure. This is often processing intensive and usually requires extensive data cleaning. As well, topology is static, and any updating or editing of the vector data requires re-building of the topology.

Algorithms for manipulative and analysis functions are complex and may be processing intensive. Often, this inherently limits the functionality for large data sets, e.g. a large number of features.

Continuous data, such as elevation data, is not effectively represented in vector form. Usually substantial data generalization or interpolation is required for these data layers.

Spatial analysis and filtering within polygons is impossible.

Raster Data

Advantages :
The geographic location of each cell is implied by its position in the cell matrix. Accordingly, other than an origin point, e.g. bottom left corner, no geographic coordinates are stored.
Due to the nature of the data storage technique data analysis is usually easy to program and quick to perform.
The inherent nature of raster maps, e.g. one attribute maps, is ideally suited for mathematical modeling and quantitative analysis.
Discrete data, e.g. forestry stands, is accommodated equally well as continuous data, e.g. elevation data, and facilitates the integrating of the two data types.
Grid-cell systems are very compatible with raster-based output devices, e.g. electrostatic plotters, graphic terminals.
The cell size determines the resolution at which the data is represented.;

Disadvantages:
It is especially difficult to adequately represent linear features depending on the cell resolution. Accordingly, network linkages are difficult to establish.
Processing of associated attribute data may be cumbersome if large amounts of data exists. Raster maps inherently reflect only one attribute or characteristic for an area.
Since most input data is in vector form, data must undergo vector-to-raster conversion. Besides increased processing requirements this may introduce data integrity concerns due to generalization and choice of inappropriate cell size.
Most output maps from grid-cell systems do not conform to high-quality cartographic needs.

It is often difficult to compare or rate GIS software that use different data models. Some personal computer (PC) packages utilize vector structures for data input, editing, and display but convert to raster structures for any analysis. Other more comprehensive GIS offerings provide both integrated raster and vector analysis techniques. They allow users to select the data structure appropriate for the analysis requirements. Integrated raster and vector processing capabilities are most desirable and provide the greatest flexibility for data manipulation and analysis.

4.3 Attribute Data Models

A separate data model is used to store and maintain attribute data for GIS software. These data models may exist internally within the GIS software, or may be reflected in external commercial Database Management Software (DBMS). A variety of different data models exist for the storage and management of attribute data. The most common are :

- Tabular
- Hierarchical
- Network
- Relational
- Object Oriented

The tabular model is the manner in which most early GIS software packages stored their attribute data. The next three models are those most commonly implemented in database management systems (DBMS). The object oriented is newer but rapidly gaining in popularity for some applications. A brief review of each model is provided.

4.3.1 Tabular Model

The simple tabular model stores attribute data as sequential data files with fixed formats (or comma delimited for ASCII data), for the location of attribute values in a predefined record structure. This type of data model is outdated in the GIS arena. It lacks any method of checking data integrity, as well as being inefficient with respect to data storage, e.g. limited indexing capability for attributes or records, etc.

4.3.2 Hierarchical Model

The hierarchial database organizes data in a *tree* structure. Data is structured downward in a *hierarchy* of tables. Any level in the hierarchy can have unlimited *children*, but any *child* can have only one *parent*. Hierarchical DBMS have not gained any noticeable acceptance for use within GIS. They are oriented for data sets that are very stable, where primary relationships among the data change infrequently or never at all. Also, the limitation on the number of parents that an element may have is not always conducive to actual geographic phenomenon.

4.3.3 Network Model

The network database organizes data in a network or *plex* structure. Any column in a plex structure can be linked to any other. Like a tree structure, a plex structure can be described in terms of *parents* and *children*. This model allows for children to have more than one parent.

Network DBMS has not found much more acceptance in GIS than the hierarchical DBMS. They have the same flexibility limitations as hierarchical databases; however, the more powerful structure for representing data relationships allows a more realistic modelling of geographic phenomenon. However, network databases tend to become overly complex too easily. In this regard it is easy to lose control and understanding of the relationships between elements.

4.3.4 Relational Model

The relational database organizes data in *tables*. Each table, is identified by a unique table name, and is organized by *rows* and *columns*. Each column within a table also has a unique name. Columns store the values for a specific attribute, e.g. cover group, tree height. Rows represent one record in the table.

UNIQUE STAND NUMBER	DOMINANT COVER GROUP	AVG. TREE HEIGHT	STAND SITE INDEX	STAND AGE
001	DEC	3	G	100
002	DEC-CON	4	M	80
003	DEC-CON	4	M	60
004	CON	4	G	120

In a GIS each row is usually linked to a separate spatial feature, e.g. a forestry stand. Accordingly, each row would be comprised of several columns, each column containing a specific value for that geographic feature. The following figure presents a sample table for forest inventory features. This table has 4 rows and 5 columns. The forest stand number would be the *label* for the spatial feature as well as the *primary key* for the database table. This serves as the linkage between the spatial definition of the feature and the attribute data for the feature.
Data is often stored in several tables. Tables can be joined or referenced to each other by common columns (relational fields). Usually the common column is an identification number for a selected geographic feature, e.g. forestry stand polygon number. This identification number acts as the *primary key* for the table. The ability to join tables through use of a common column is the essence of the relational model. Such relational joins are usually ad hoc in nature and form the basis of for querying in a relational GIS product. Unlike the other previously discussed database types, relationships are implicit in the character of the data as opposed to explicit characteristics of the database set up.

There are many different designs of DBMSs, but in GIS the relational design has been the most useful. In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link them together. This surprisingly simple design has been so widely used primarily because of its flexibility and very wide deployment in applications both within and without GIS, see example below.

In fact, most GIS software provides an *internal* relational data model, as well as support for *commercial off-the-shelf* (COTS) relational DBMS'. COTS DBMS' are referred to as *external* DBMS'. This approach supports both users with small data sets, where an internal data model is sufficient, and customers with larger data sets who utilize a DBMS for other corporate data storage requirements. With an external DBMS the GIS software can simply *connect* to the database, and the user can make use of the inherent capabilities of the DBMS. External DBMS' tend to have much more extensive querying and data integrity capabilities than the GIS' internal relational model. The emergence and use of the external DBMS is a trend that has resulted in the proliferation of GIS technology into

more traditional data processing environments.

The relational DBMS is attractive because of its:

- Simplicity in organization and data modelling.
- Flexibility - data can be manipulated in an ad hoc manner by joining tables.
- Efficiency of storage - by the proper design of data tables redundant data can be minimized; and
- The non-procedural nature - queries on a relational database do not need to take into account the internal organization of the data.

4.3.5 Object-Oriented Model

The object-oriented database model manages data through *objects*. An object is a collection of data elements and operations that together are considered a single entity. The object-oriented database is a relatively new model. This approach has the attraction that querying is very natural, as features can be bundled together with attributes at the database administrator's discretion. To date, only a few GIS packages are promoting the use of this attribute data model. However, initial impressions indicate that this approach may hold many operational benefits with respect to geographic data processing. Fulfillment of this promise with a commercial GIS product remains to be seen.

5 Spatial Data Relationships

The nature of spatial data relationships is important to understand within the context of GIS. In particular, the relationship between geographic features is a complex problem in which we are far from understanding in its entirety. This is of concern since the primary role of GIS is the manipulation and analysis of large quantities of spatial data. To date, the accepted theoretical solution is to *topologically structure* spatial data.

Most GIS software segregate spatial and attribute data into separate data management systems. Most frequently, the topological or raster structure is used to store the spatial data, while the relational database structure is used to store the attribute data. Data from both structures are linked together for use through unique identification numbers, e.g. feature labels and DBMS primary keys. This coupling of spatial features with an attribute record is usually maintained by an internal number assigned by the GIS software. A *label* is required so the user can load the appropriate attribute record for a given geographic feature. Most often a single attribute record is automatically created by the GIS software once a clean topological structure is properly generated. This attribute record normally contains the internal number for the feature, the user's label identifier, the area of the feature, and the perimeter of the feature. Linear features have the length of the feature defined instead of the area.

The topologic model is often confusing to initial users of GIS. Topology is a mathematical approach that allows us to structure data based on the principles of feature adjacency and feature connectivity. It is in fact the mathematical method used to define spatial relationships. Without a topologic data structure in a vector based GIS most data manipulation and analysis functions would not be practical or feasible.
The most common topological data structure is the *arc/node* data model. This model contains two basic entities, the *arc* and the *node*. The arc is a series of points, joined by straight line segments, that start and end at a node. The node is an intersection point where two or more arcs meet. Nodes also occur at the end of a *dangling* arc, e.g. an arc that does not connect to another arc such as a dead end street. Isolated nodes, not connected to arcs represent point features. A polygon feature is comprised of a closed chain of arcs.

In GIS software the topological definition is commonly stored in a proprietary format. However, most software offerings record the topological definition in three tables. These tables are analogous to relational tables. The three tables represent the different types of features, e.g. point, line, area. A fourth table containing the coordinates is also utilized. The *node table* stores information about the node and the arcs that are connected to it. The *arc table* contains topological information about the arcs. This includes the start and end node, and the polygon to the left and right that the arc is an element of. The *polygon table* defines the arcs that make up each polygon. While arc, node, and polygon terminology is used by most GIS vendors, some also introduce terms such as *edges* and *faces* to define arcs and polygons. This is merely the use of different words to define

Since most input data does not exist in a topological data structure, topology must be *built* with the GIS software**.** Depending on the data set this can be an CPU intensive and time consuming procedure. This building process involves the creation of the topological tables and the definition of the arc, node, and polygon entities. To properly define the topology there are specific requirements with respect to graphic elements, e.g. no duplicate lines, no gaps in arcs that define polygon features, etc.

The topological model is utilized because it effectively models the relationship of spatial entities. Accordingly, it is well suited for operations such as contiguity and connectivity analyses. Contiguity involves the evaluation of feature adjacency, e.g. features that touch one another, and proximity, e.g. features that are near one another. The primary advantage of the topological model is that spatial analysis can be done without using the coordinate data. Many operations can be done largely, if not entirely, by using the topological definition alone. This is a significant advantage over the CAD or *spaghetti* vector data structure that requires the derivation of spatial relationships from the coordinate data before analysis can be undertaken.

The major disadvantage of the topological data model is its static nature. It can be a time consuming process to properly define the topology depending on the size and complexity of the data set. For example, 2,000 forest stand polygons will require considerably longer to *build* the topology that 2,000 municipal lot boundaries. This is due to the inherent complexity of the features, e.g. lots tend to be rectangular while forest stands are often long and sinuous. This can be a consideration when evaluating the topological building capabilities of GIS software. The static nature of the topological model also implies that every time some editing has occurred, e.g. forest stand boundaries are changed to reflect harvesting or burns, the topology must be rebuilt. The integrity of the topological structure and the DBMS tables containing the attribute data can be a concern here. This is often referred to as *referential integrity*. While topology is the mechanism to ensure integrity with spatial data, referential integrity is the concept of ensuring integrity for both linked topological data and attribute data.

A GIS stores information about the world as a collection of thematic layers that can be linked together by geography. This simple but extremely powerful and versatile concept has proven invaluable for solving many real-world problems from tracking delivery vehicles, to recording details of planning applications, to modeling global atmospheric circulation.

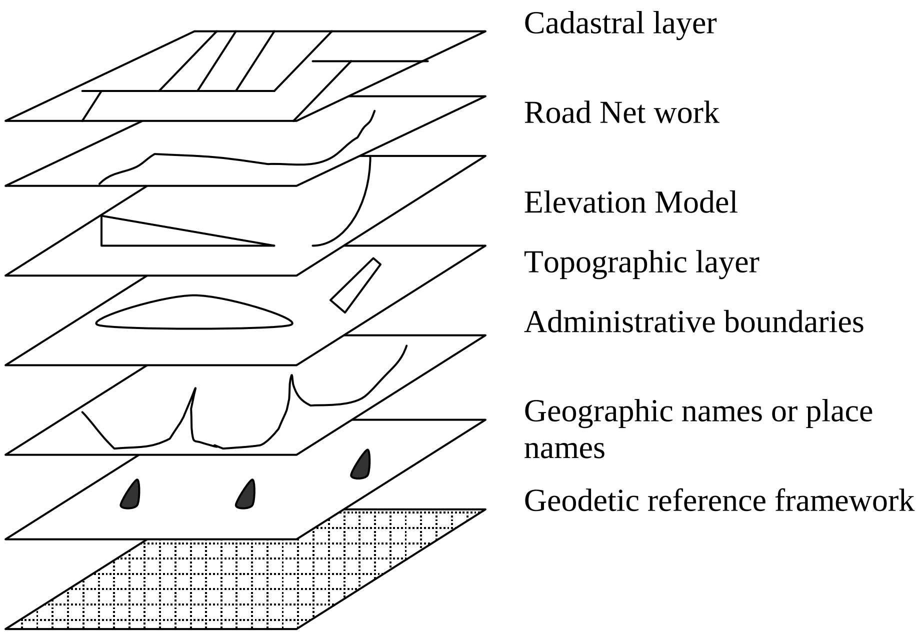

Figure 4: Reality in the different thematic layers (Farha, 1999)

Geographic information contains either an explicit geographic reference, such as a latitude and longitude or national grid coordinate, or an implicit reference such as an address, postal code, census tract name, forest stand identifier, or road name. An automated process called geocoding is used to create explicit geographic references (multiple locations) from implicit references (descriptions such as addresses). These geographic references allow you to locate features, such as a business or forest stand, and events, such as an earthquake, on the earth's surface for analysis.

7 GIS Tasks

General purpose geographic information systems essentially perform five processes or tasks:

- Input
- Manipulation
- Management
- Query and Analysis
- Visualization

7.1 Input

Before geographic data can be used in a GIS, the data must be converted into a suitable digital format. The process of converting data from paper maps into computer files is called *digitizing*.
Modern GIS technology can automate this process fully for large projects using scanning technology; smaller jobs may require some manual digitizing (using a digitizing table). Today many types of geographic data already exist in GIS-compatible formats. These data can be obtained from data suppliers and loaded directly into a GIS.

7.2 Manipulation

It is likely that data types required for a particular GIS project will need to be transformed or manipulated in some way to make them compatible with your system. For example, geographic information is available at different scales (detailed street centerline files; less detailed census boundaries; and postal codes at a regional level). Before this information can be integrated, it must be transformed to the same scale (degree of detail or accuracy). This could be a temporary transformation for display purposes or a permanent one required for analysis. GIS technology offers many tools for manipulating spatial data and for weeding out unnecessary data.

7.3 Management

For small GIS projects it may be sufficient to store geographic information as simple files. However, when data volumes become large and the number of data users becomes more than a few, it is often best to use a database management system (DBMS) to help store, organize, and manage data. A DBMS is nothing more than computer software for managing a database.

There are many different designs of DBMSs, but in GIS the relational design has been the most useful. In the relational design, data are stored conceptually as a collection of tables. Common fields in different tables are used to link them together. This surprisingly simple design has been so widely used primarily because of its flexibility and very wide deployment in applications both within and without GIS.

7.4 Query and Analysis

Once you have a functioning GIS containing your geographic information, you can begin to ask simple questions such as:

- Who owns the land parcel on the corner?
- How far is it between two places?
- Where is land zoned for industrial use?
- And analytical questions such as
- Where are all the sites suitable for building new houses?
- What is the dominant soil type for oak forest?
- If I build a new highway here, how will traffic be affected?

GIS provides both simple point-and-click query capabilities and sophisticated analysis tools to provide timely information to managers and analysts alike. GIS technology really comes into its own when used to analyze geographic data to look for patterns and trends and to undertake "what if" scenarios. Modern GISs have many powerful analytical tools, but two are especially important.

7.5 Proximity Analysis

- How many houses lie within 100 m of this water main?
- What is the total number of customers within 10 km of this store?
- What proportion of the alfalfa crop is within 500 m of the well?

To answer such questions, GIS technology uses a process called **buffering** to determine the proximity relationship between features.

7.6 Overlay Analysis

The integration of different data layers involves a process called overlay. At its simplest, this could be a visual operation, but analytical operations require one or more data layers to be joined physically. This overlay, or spatial join, can integrate data on soils, slope, and vegetation, or land ownership with tax assessment.

7.7 Visualization

For many types of geographic operation the end result is best visualized map or graph. Maps are very efficient at storing and communicating geographic information. While cartographers have created maps for millennia, GIS provides new and exciting tools to extend the art and science of cartography. Map displays can be integrated with reports, three-dimensional views, photographic images, and other output such as multimedia.

8 Related Technologies

GISs are closely related to several other types of information systems, but it is the ability to manipulate and analyze geographic data that sets GIS technology apart. Although there are no hard and fast rules about how to classify information systems, the following discussion should help differentiate GIS from desktop mapping, computer-aided design (CAD), remote sensing, DBMS, and global positioning systems (GPS) technologies.

8.1 Desktop Mapping

A desktop mapping system uses the map metaphor to organize data and user interaction. The focus of such systems is the creation of maps: the map is the database. Most desktop mapping systems have more limited data management, spatial analysis, and customization capabilities. Desktop mapping systems operate on desktop computers such as PCs, Macintoshes, and smaller UNIX workstations.

8.2 CAD

CAD systems evolved to create designs and plans of buildings and infrastructure. This activity required that components of fixed characteristics be assembled to create the whole structure. These systems require few rules to specify how components can be assembled and very limited analytical capabilities. CAD systems have been extended to support maps but typically have limited utility for managing and analyzing large geographic databases.

8.3 Remote Sensing and GPS

Remote sensing is the art and science of making measurements of the earth using sensors which are not in direct contact with the objects under observation. The sensors collect data in the form of images and provide specialized capabilities for manipulating, analyzing, and visualizing those images. Lacking strong geographic data management and analytical operations, they cannot be called true GISs.

8.4 DBMS

Database management systems specialize in the storage and management of all types of data including geographic data. DBMSs are optimized to store and retrieve data and many GISs rely on them for this purpose. They do not have the analytic and visualization tools common to GIS.

9 What Can GIS Do For Us?

- Perform Geographic Queries and Analysis
- The ability of GISs to search databases and perform geographic queries has saved many companies literally millions of dollars. GISs have helped reduce costs by
- Streamlining customer service.
- Reducing land acquisition costs through better analysis.
- Reducing fleet maintenance costs through better logistics.
- Analyzing data quickly, as in this example:
- A realtor could use a GIS to find all houses within a certain area that have tiled roofs and five bedrooms, then list their characteristics.
- The query could be further refined by adding criteria - the house must cost less than $100 per square foot. You could also list houses within a certain distance of a school.

9.1 Improve Organizational Integration

Many organizations that have implemented a GIS have found that one of its main benefits is improved management of their own organization and resources. Because GISs have the ability to link data sets together by geography, they facilitate interdepartmental information sharing and communication. By creating a shared database, one department can benefit from the work of another - data can be collected once and used many times.

As communication increases among individuals and departments, redundancy is reduced, productivity is enhanced, and overall organizational efficiency is improved. Thus, in a utility company the customer and infrastructure databases can be integrated so that when there is planned maintenance, affected customers can be sent a computer-generated letter.

9.2 Make Better Decisions

The old adage "better information leads to better decisions" is as true for GIS as it is for other information systems. A GIS, however, is not an automated decision making system but a tool to query, analyze, and map data in support of the decision making process. GIS technology has been used to assist in tasks such as presenting information at planning inquiries, helping resolve territorial disputes etc.

GIS can be used to help reach a decision about the location of a new housing development that has minimal environmental impact, is located in a low-risk area, and is close to a population center. The information can be presented succinctly and clearly in the form of a map and accompanying report, allowing decision makers to focus on the real issues rather than trying to understand the data. Because GIS products can be produced quickly, multiple scenarios can be evaluated efficiently and effectively.

9.3 Making Maps

Maps have a special place in GIS. The process of making maps with GIS is much more flexible than are traditional manual or automated cartography approaches. It begins with database creation. Existing paper maps can be digitized and computer-compatible information can be translated into the GIS. The GIS-based cartographic database can be both continuous and scale free. Map products can then be created centered on any location, at any scale, and showing selected information symbolized effectively to highlight specific characteristics.

The characteristics of atlases and map series can be encoded in computer programs and compared with the database at final production time. Digital products for use in other GISs can also be derived by simply copying data from the database. In a large organization, topographic databases can be used as reference frameworks by other departments.

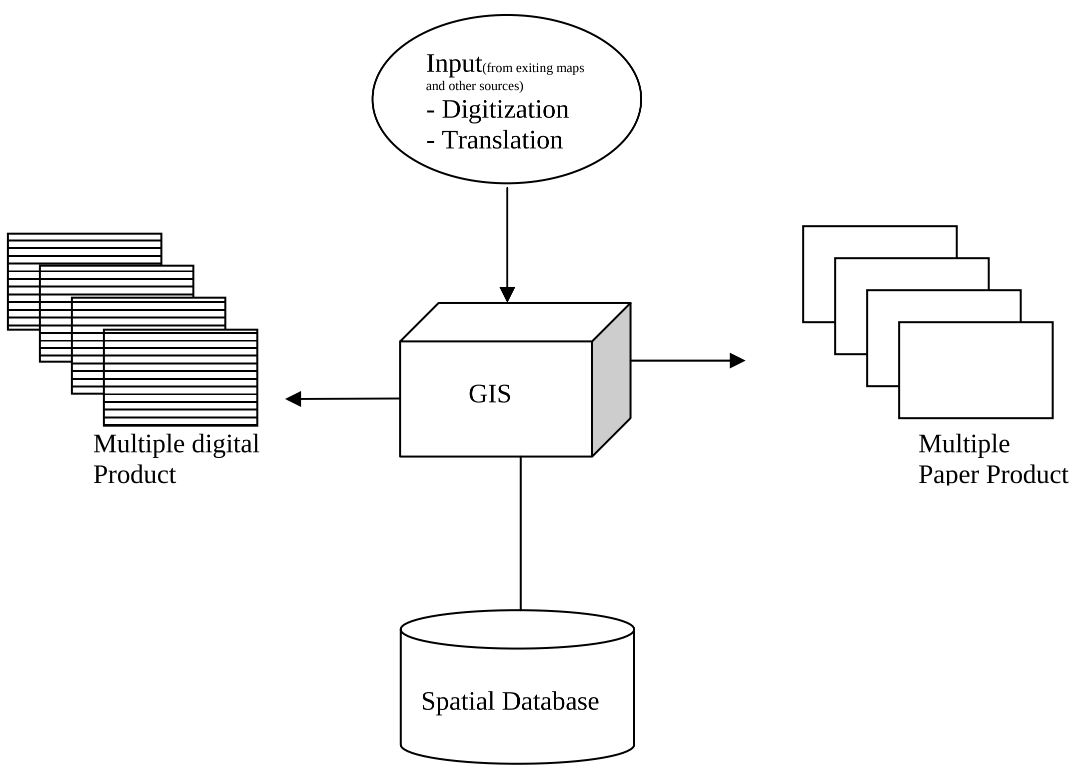

Figure 5 - GIS, spatial database and multiple products.

10 GIS in Everyday Life

In today's global community, the more information you have at your fingertips, the easier it is to make an informed decision. In today's high-tech world, information comes in many different ways, from company reports and statistics from down the hall to digital photos and multimedia from across the world.

Information can be overwhelming and the need for timely decisions calls not only for innovative ways to access accurate, up-to-the minute information, but also tools to help present the information in useful ways.

A geographic information system or GIS allows you to bring all types of data together based on the geographic and location component of the data.

But unlike a static paper map, GIS can display many layers of information that is useful to you. You will be able to integrate, visualize, manage, solve, and present the information in a new way. Relationships between the data will become more apparent and your data will become more valuable. GIS will give you the power to create maps, integrate information, visualize scenarios, solve complicated problems, present powerful ideas, and develop effective solutions like never before. GIS is a tool used by individuals and organizations, schools, governments, and businesses seeking innovative ways to solve their problems.

References

Aronoff , S. *"Geographic information Systems : A Management Perspective"*,wdl Publications , Ottawa, Canada, forth edition , 1995.

Burrough , P., *"Principles of Geographic Information Systems for Land Resource Management"*, Oxford University Press, Oxford 1986.

Farha, S., "Web Based Agent Oriented Spatial Data Infrastructure", M.Sc Thesis, International Institute of Aerospace Survey and Earth Sciences (ITC), The Netherlands, 1999.

Frank, A.U., *"Spatial concepts, geometric data model and data structure, in: Frank and Goodchild,Two perspectives on geographical data modelling"*,National Center for Geographical Information and Analysis. Technical Paper Series 90-11, Santa Barbra, USA.1990.

Goodchild, M.F., *"Geographical Data Modelling, in: Frank and Goodchild, Two perspectives on geographical data modelling"*, National Center for Geographical Information and Analysis. Technical Paper Series 90-11, Santa Barbra, USA.1990.