

Key Components of GIS:

The GIS constitutes of five key components, namely, hardware, software, procedure, data and users. These five components need to be in balance to function any information system satisfactorily.

Hardware:

It consists of the computer hardware system on which the GIS software runs. The GIS run on the whole spectrum of computer systems ranging from portable personal computers to multi-user supercomputer. The hardware for GIS consists of input devices such as digitizers, scanners and GPS receivers, the storage devices such as magnetic tapes and disks, CD ROMs and other optical disks, central processing units and the output devices such as display devices, printer and plotters.

Software:

Software refers to the programmes that run on computers, these include programmes to manage the computer and to perform specific functions. The GIS software provides the functions and tools that are necessary to store, analyse and display

geographic information. Some common GIS software are ArcGIS, ArcView, ArcSDE, ArcIMS, MapInfo, Geomatica, etc. Web based GIS is another concept, which is becoming very popular nowadays, and it uses the web application software. A common practice in GIS is using MS-Access, Oracle, SQL Server, etc as a DBMS software along with the chosen GIS software.

Procedure:

A computer system for GIS consists of hardware, software, and procedures designed to support the data capture, storage, processing, analysis, modelling, and display of geospatial data. Besides the technical components like hardware, software and databases, institutional framework and policies are also important for a functional GIS. The interest and willingness of decision makers in exploiting GIS technology, and the organisational set up for collecting spatial data, analysis procedures, and using the results for planning and implementation form a very important component in a GIS. A successful GIS operates according to a well-designed plan and business rules, which are the models and operating practices unique to each organisation.

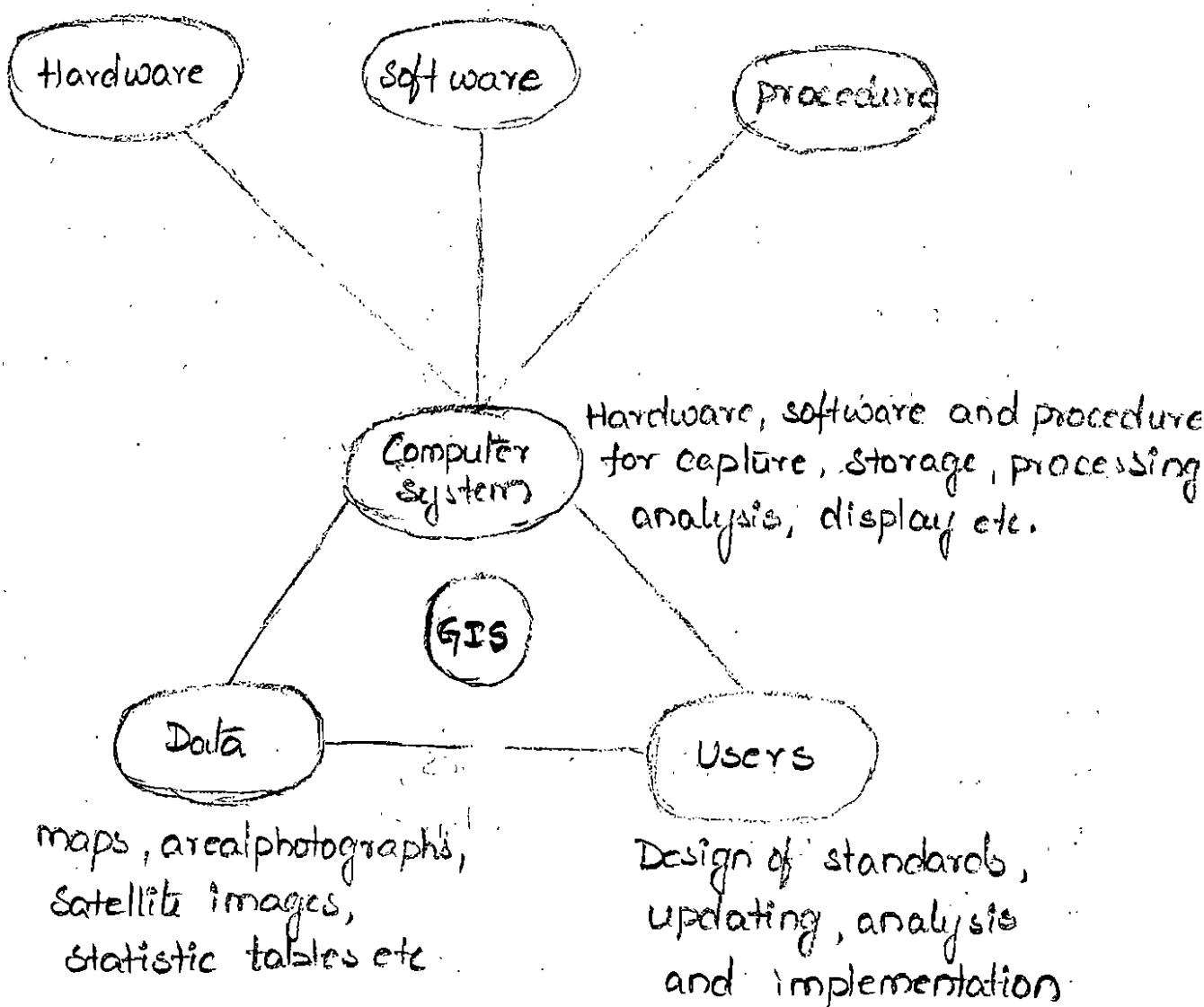
Data:

Data are named as geospatial and attribute data in GIS. The GIS facilitates integration of spatial and attribute data and this makes them unique in contrast to other database systems. The GIS data is handled in databases with special functional requirements as well as the general characteristics of any standard database. The sources of spatial data are digitized maps, aerial photographs, and satellite images. Attribute data sources are statistical tables and other related documents. The digital map forms are the basic data input for GIS. Tabular data related to the map objects are also necessary to be attached to the digital map data.

Users:

The roles of the users are to select pertinent information to set necessary standards, to design cost-efficient updating schemes, to analyse GIS outputs for relevant purposes, and plan the implementation. Most definition of GIS focuses on the hardware, software, data and analysis components. However, no GIS exist in isolation of the user. There must always be people to plan, implement, and operate the system as well as to make decisions based on the output. The GIS projects range from small research applications, where one user is responsible for design and

implementation and output, to international corporate distributed systems, where different type of users interact with the GIS in many different levels and ways.



Key Components of GIS

Application Areas of GIS:

To understand the advantages and importance of GIS in a better way, it is very important to know the applications and uses of GIS. A GIS enables us to better understand and evaluate our data by creating graphic displays using information stored in our database. With a GIS, we can change the display of our geographic data by changing the symbols, colours, or values in the database tables. The GIS can be applied in various distinct application areas. Major areas of application of GIS can be grouped into five categories.

Facilities Management:

Large scale and precise maps and network analysis are used mainly for utility management.

Environment and Natural Resources Management:

Medium- or Small-scale maps and overlay techniques in combination with aerial photographs and satellite images are used for management of natural resources and environmental impact analysis.

Street Network:

Large- or medium- scale maps and spatial analysis are used for vehicle routing, locating house and streets, etc.

Planning and Engineering:

Large- or medium- scale maps and engineering models are used mainly in civil engineering.

Land Information System:

Large- scale cadastral maps or land parcel maps and spatial analysis are used for cadastre administration, taxation, etc.

Agriculture:

The GIS is used in a variety of agricultural applications such as managing crop yields, monitoring crop rotation techniques, and projecting soil loss for individual farms or entire agricultural regions, etc.

Business:

A GIS is a tool for managing any kind of business information according to its location. We can keep a track of where the customers are located, target marketing campaigns, optimize

Sales territories, and model retail spending patterns such an added advantage is provided by the GIS to enhance in marketing the company more competitive and successful.

Electric/gas utilities:

The GIS is used on a daily basis by the cities and utilities to help them in mapping, in invent systems, track maintenance, monitor regulatory compliance or model distribution analysis, transformer analysis, and load analysis.

Environment:

The GIS is used everyday to help protect the environment. An environmental professional uses GIS to produce maps, inventory species, measure environmental impact, or trace pollutants. The environmental applications for GIS are almost endless.

Forestry:

Nowadays, managing forests is becoming a more complex and demanding challenge. With GIS, foresters can easily see the forest as an ecosystem and manage it efficiently.

Geology:

Geologists use GIS every day in a wide variety of applications. The GIS can also be used to study

geological features, analyse soils and strata, assess seismic information, or create three-dimensional (3D) displays of geographic features.

Hydrology:

The GIS can also be used to study drainage systems, assess groundwater, and visualize watersheds, and in many other hydrologic applications.

Land-use planning:

people use GIS to help visualize and plan the land-use requirements of cities, regions, or even national governments.

Local government:

The local government uses GIS every day to solve several problems, for taxation, and so on. Often the data collected and used by one agency or department can be used by another.

Mapping:

Mapping is an essential function of a GIS. People in a variety of professions use GIS to help others understand geographic data.

Military:

Military analysts and cartographers use GIS in a variety of applications such as creating base-maps, assessing terrain, and aiding in tactical decisions.

Risk Management:

A GIS can help with risk management and analysis by showing us which areas are likely to be prone to natural or manmade disasters. When such forthcoming disasters are identified, preventive measures can be developed that deal with the different scenarios.

Site Planning:

People around the world use GIS to help them locate sites for new facilities or locate alternative sites for existing facilities.

Transportation:

A GIS can be used to in managing transportation infrastructure or in managing logistical problems. They can help us in monitoring rail systems and road conditions or finding the best way to deliver goods and services.

Water/Wastewater Industry:

People in the water/wastewater industry use GIS with the planning, engineering, operations, maintenance, finance and administration functions of their water/waste water networks.

1. 3. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

1. 1. 1. 1. 1. 1. 1.

Raster Data Representation:

In raster representation, the terrain is divided into a number of parcels or quantised the space into units. A parcel or a unit is called a grid cell. Although a wide variety of raster shapes like triangles or hexagons are possible, it is generally simpler to use a series of rectangles, or more often squares, called grid cells. Grid cells or other raster forms generally are uniform in size, but this is not absolutely necessary. For the sake of simplicity, we will assume that all grid cells are of the same size and that, therefore, each occupies the same amount of geographic space as any other.

Raster data structures do not provide precise locational information because geographic space is now divided into discrete grids, as much as we divide a checkerboard into uniform squares. Instead of representing points with their absolute locations, they are represented as a single grid cell. This stepped appearance is also obvious when we represent areas with grid cells. All points inside the area that is bounded by a close set of lines must occur within one of the grid cells to be represented as part of the same area. The more irregular the area, the more stepped the appearance.

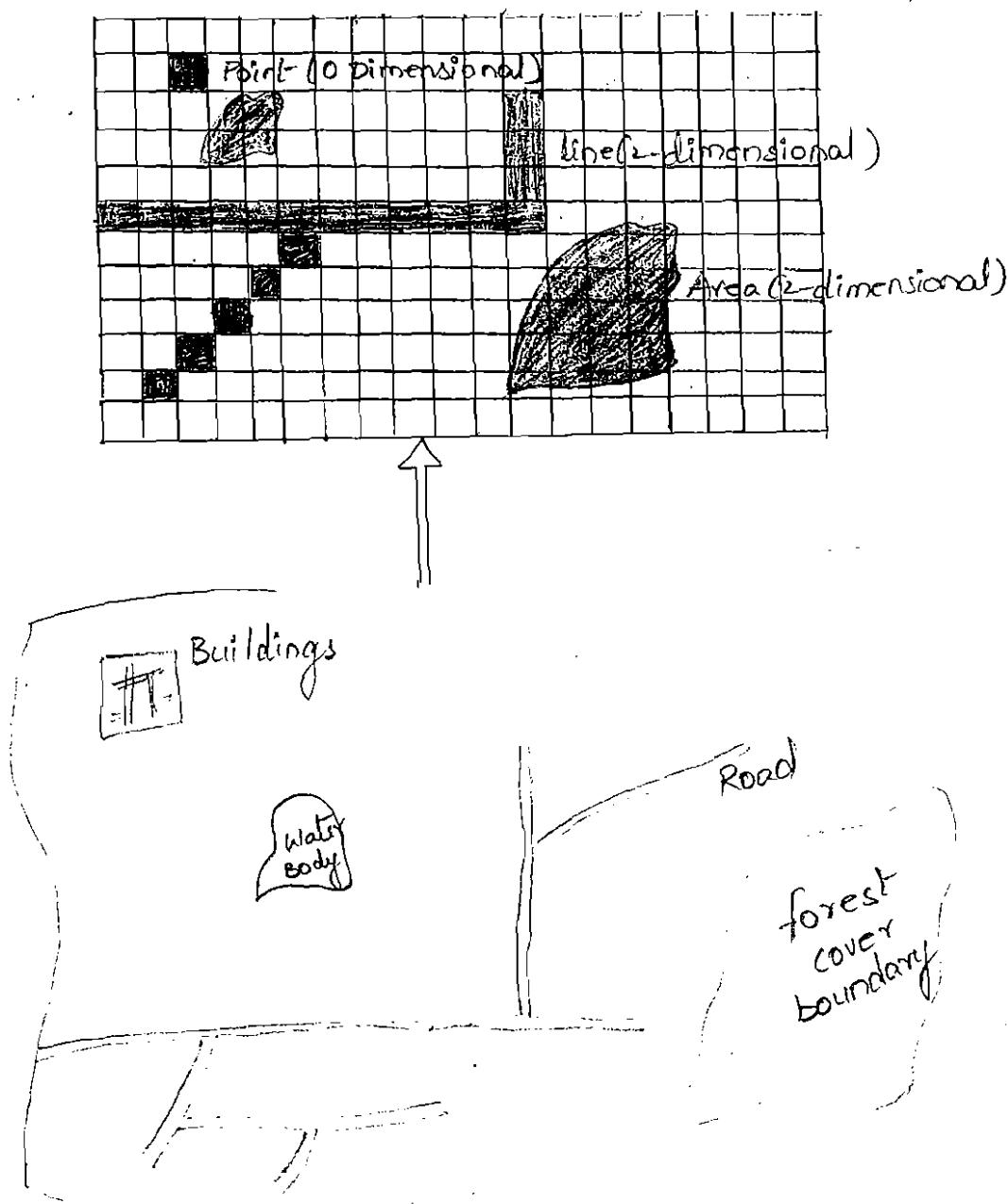
Raster structures, especially square grid cells, are pieced together to represent an entire area.

In grid-based or raster GIS, there are two general ways of including attribute data for each entity.

The simplest is to assign a single number representing an attribute like a class of land cover, for each grid cell location. By positioning these numbers, we, ultimately, are allowing the position of the attribute value to act as the default location for the entity. For example, if we assign a code number of 10 to represent water, then list this as the first number in the X or column direction, and the first in the Y or row direction, by default the upper left grid cell is the location of a portion of the earth representing water. The larger the grid cell, the more land area is contained within. It - a concept called resolution. The coarser the resolution of the grid, the less we know about the absolute position of points, lines, and areas represented by this structure.

Raster structures, especially square grid cells, are pieced together to represent an entire area. Raster data structure may seem to be rather undesirable because of the lack of absolute locational information. Raster data structures have numerous advantages over other structures. Notably, they are relatively easy to conceptualise as a method of representing space. Remotely sensed data acquired by a sensor is one of the well known example of raster data representation. In fact, the relationship between the pixel used in remote sensing and the grid cell used in GIS allows data from satellites to be readily

incorporated into raster based GIS without any changes. A characteristic feature of grid based systems is that many functions, especially those involving the analysis and modelling of surfaces and overlay operations, are simple to perform with this type of data structure. The major disadvantages of the raster data structure are a reduced spatial accuracy, decrease of the reliability of area and distance measures, and the need for large storage capacity associated with having to record every grid cell as a numerical value.



Types of Raster GIS Models :

The grid based GIS spatial data can be stored manipulated ,analysed and referenced basically in any one of the three methods/models . These three models are GRID/LUNAR/MAGI model, IMGRID model and MAP model. All of these models use the grid cell values, their attributes , coverages and corresponding legends. These models are developed depending upon the requirements from time to time. Based on the applications of interest , availability of softwares and other related information, any one of the above models can be selected for the execution of a particular GIS project. There are a number of ways of forcing a computer to store and reference the individual grid cell values, their attributes , coverage names and legends. These models can be understood by considering a checkerboard. Red indicates water and black, land. Perhaps you could think of a checkerboard ,with its red and black squares If each of these squares is taken to represent a simple map of land cover we have produced a simple cover But the problem is , how are the attributes of our landcover physically connected to these grid squares We can pick up the entire checkerboard because it is a physically connected structure. Likewise, when we pick up a thematic map , it also represents all the

different changes in the theme as a single, connected object. The similarity between the checker board as a single unit of play for a game and the map as a single unit of storage for spatial information is natural. All these problems can be resolved if we use anyone of the above mentioned models as discussed in the following paragraphs.

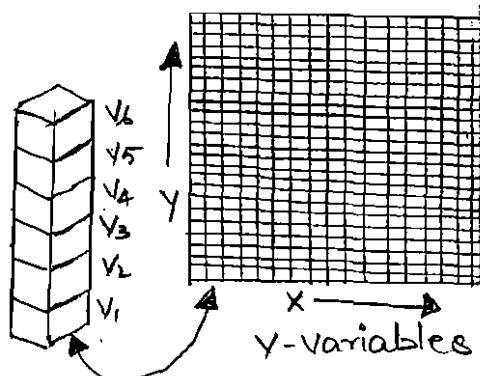
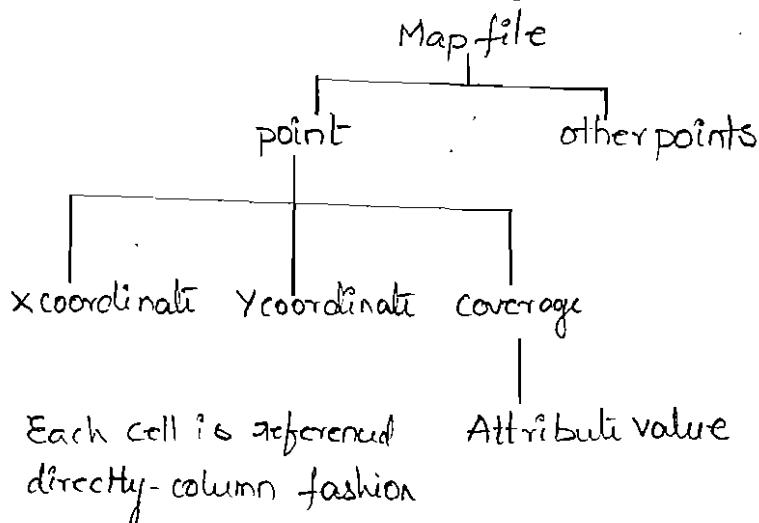
GRID Model:

The first and foremost model for the representation of raster data is the GRID model. The model method of storing, manipulating, and analysing the grid based data was first conceptualised by an attempt to develop GRID model. In this method, each grid cell is referenced and addressed individually and is associated with identically positioned grid cells in all other coverages, rather like a vertical column of grid cells, each dealing with a separate theme. Comparisons between coverages are therefore performed on a single column at a time.

For example : To compare soil attributes in one coverage with vegetation attributes in second coverage, land used/land covered attributes in a third coverage, each x and y location must be examined individually. So a soil grid cell at location must be examined individually. So a soil grid

cell at location $x_{10}-y_{10}$ will be compared to its vegetation counterpart and third layer (land used) land covered at location $x_{10}-y_{10}$. You might be able to envision this by imagining a geological core in which each rock type is lying directly on top of the next; and to get a picture of the entire study area, it will be necessary to put a large number of cores together.

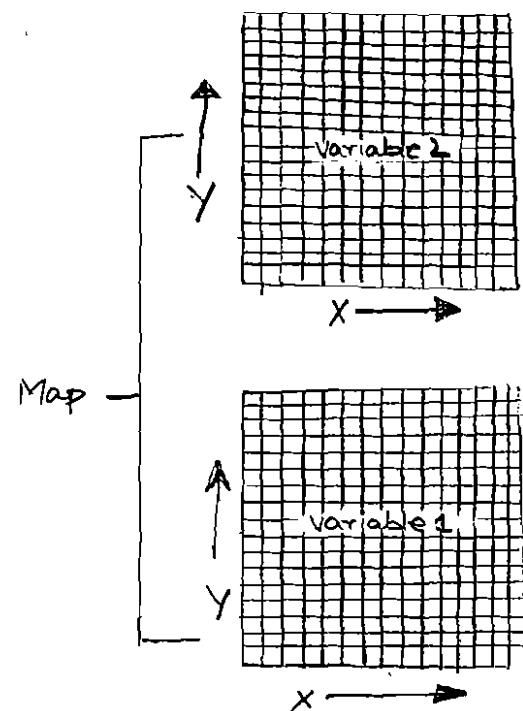
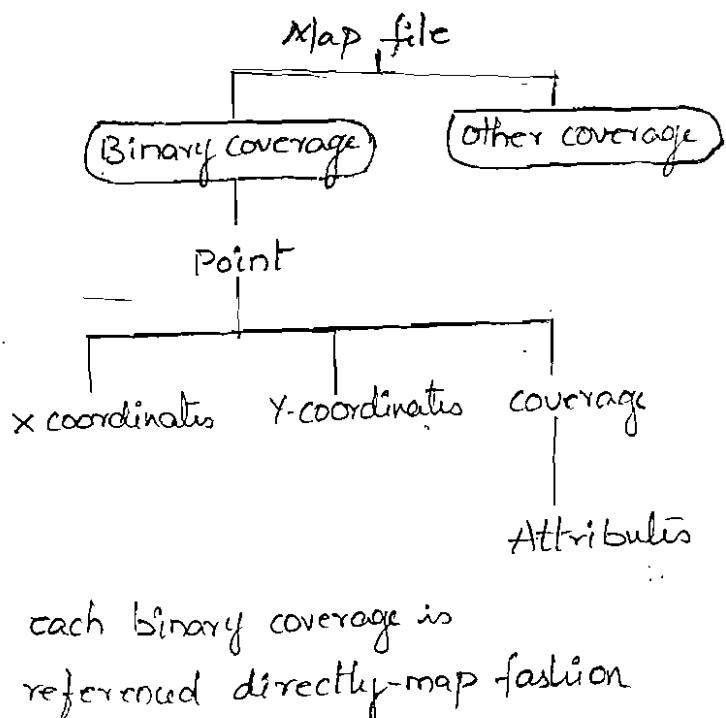
The advantage of this model is that computation/comparison of multiple themes or coverages for each grid cell location is relatively easy. This is a resonable approach and has proven successful. The main disadvantage is that it limits the efficient examination of relationships of themes to one-to-one relationships within the spatial framework. Second disadvantage is more storage space for the cell data and the representation is vertical rather than horizontal, which would more closely resemble our notion of maps.



IMGRID Model :

With a slight modification of the checkerboard analog, the second basic raster data model, that is the IMGRID data model, can be used. This model is also used in the early GIS system. Let us assume that the red squares on checkerboard map serve to contain a single attribute, rather than just a theme. Instead, we can use the number 1 'red squares' to represent water and 0 'black squares' to indicate the absence of water.

A thematic map of land use that contains four categories, namely, recreation, agriculture, industry, and residences is represented as, each of these four attributes would have to be separated out as an individual layer. One layer would stand for agriculture only, with 1's and 0's representing the presence or absence of this activity for each grid cell. Recreation, industry and residences would be represented in the same way, with each variable referencing directly, rather than referencing the grid cells as did in the GRID/LUNAR/MAGI data model.



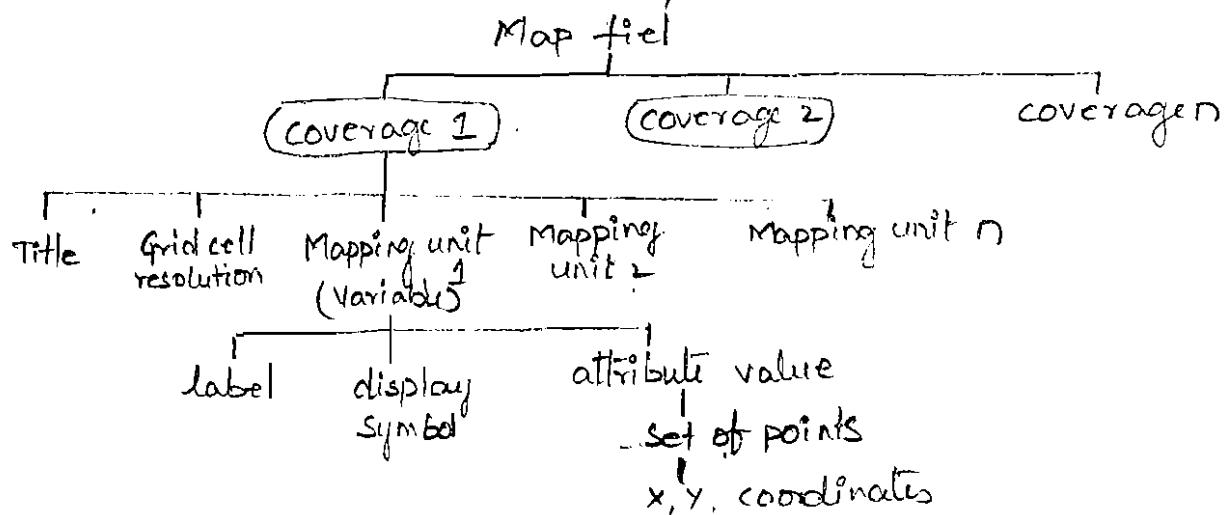
MAP Model:

The third raster GIS model MAP Analysis Package (MAP) model developed by C. Dana Tomlin, formally integrates the advantages of the above two raster data structure model methods. In this data model each thematic coverage is recorded and accessed separately by map name or title. This is accomplished by recording each variable, or mapping unit, of the coverage's theme as a separate number code or label, which can be accessed individually when the coverage is retrieved. The label corresponds to a portion of the legend and has its own symbol assigned to it. In this way, it is easy to perform operation on individual grid cells and groups of similar grid cells, and the resolution changes in value requires rewriting only a single number per mapping unit,

thus simplifying the computations. The overall major improvement is that the MAP method allows ready manipulation of the data in a many-to-one relationship of the attribute values and the sets of grid cells.

The MAP data model is compatible to almost all computer systems from its original mainframe version to Macintosh and PC versions and modern UNIX-based work station versions. It can be used as a teaching version of GIS as it is very flexible and also becomes a major module in commercial GIS packages like ARC/INFO.

Although raster GIS systems have traditionally been developed to allow single attributes to be stored individually for each grid cell, some have evolved to include direct links to existing database management systems. This approach extends to include the utility of the raster GIS by minimising the number of coverages and substituting multiple variables for each grid cell in each coverage. Such extensions to the raster data model have also allowed direct linkage to existing GIS systems that use a vector back and forth from raster to vector.



VECTOR DATA MODEL :

The vector model is close to the traditional mapping approach where the objects are represented as point lines, or areas. In a vector model, the positions of point lines, and areas are precisely specified. The position of each object is defined by a coordinate pairs.

Vectors are graphical objects that have geometric primitives such as points, lines, and polygons to represent geographical entities in computer graphics. Vectors have a precise direction, length, and shape, and can be defined by coordinate geometry.

Object-based Vector Model :

The vector model is ideal to represent discrete entities. According to this concept, discrete entities are represented as points, lines and areas. Point is simpler to input and analyse. Points are required to represent entities whose areas are negligible or not important; such as electric poles, postbox, tube well, etc.

Lines, defined by two points, are used to represent features that are linear in nature, for example, roads are pipelines. They can also be used to represent linear features that do not have any physical existence, such as a line showing an international border. It is often not possible to represent real-world linear entities by a single line. Multiple sequentially connected segments of line are used to create an object. These multiple lines are collectively called as polyline. Polyline have different

structure than lines. Multiple lines are multiple objects but each of the polylines having multiple linear segments is a single object. Arc is another term used in ARC/INFO synonymously with polyline.

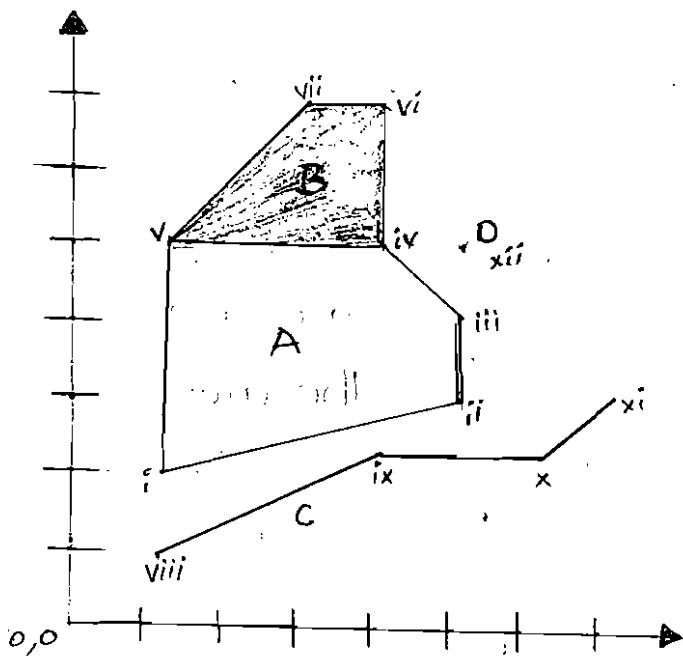
Areas are represented by a close set of lines and areas used to define features such as play-grounds, buildings, or administrative areas. These closed set of lines are referred to as polygons or regions. As with line features, some of these polygons exist on the ground, while others are imaginary. Polygons need only points to input but the area, perimeter, and other geometric attributes may be computed by GIS software rather than by manual input.

The object-based vector model, points / lines / area can be represented via different vector model. Four such models are important to discuss (1) spaghetti (2) vertex dictionary, (3) dual independent map encoding DIME and (4) topological model.

(1) Spaghetti Model:

The spaghetti model uses the simplest type of data structure. All objects are defined as single items and no reference is made to other objects. Spaghetti data are collection of points and lines segments with no real connection. There are no specific points that designate where the lines cross, nor there any logical relationships between the objects.

The common boundaries between adjacent polygons are stored twice. This model cannot handle holes within a polygon. The structure is inefficient in data storage and queries, and consistency checks are not possible.



For polygon A
A,5 [identifier of polygon and number of vertices]
1,2 [coordinates of vertex i]
2,3 [coordinates of vertex ii]
3,4 [coordinates of vertex iii]
4,5 [coordinate of vertex iv]
5,6 [coordinate of vertex v]
1,2 [coordinate of vertex i again]

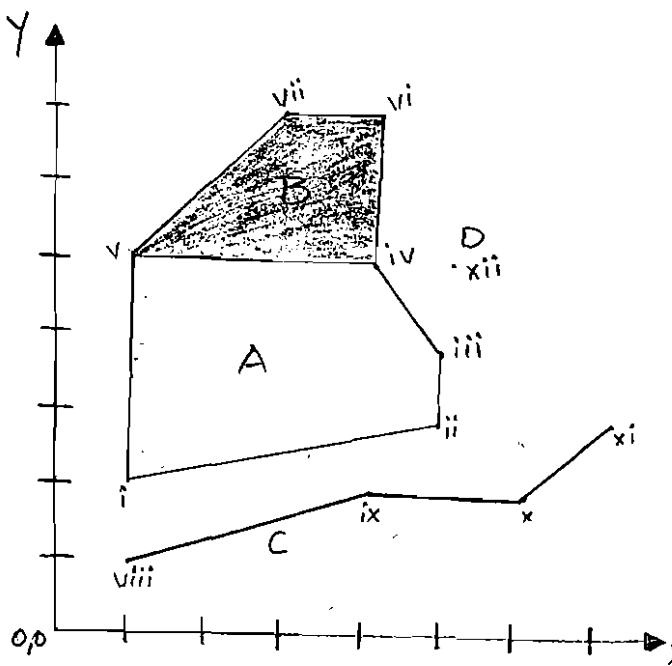
For polygon B
B,4 [identifier of polygon and number of vertices]
4,5 [coordinate of vertex v]
4,7 [coordinate of vertex vi]
3,7 [coordinate of vertex vii]
1,5 [coordinate of vertex v]
4,5 [coordinate of vertex iv again]

For line C
C,4 [identifier of line and number of vertices]
1,1 [coordinate of vertex viii]
4,2 [coordinate of vertex ix]
6,12 [coordinate of vertex x]
7,3 [coordinate of vertex xi]

For point D
D,1 [identifier of point and number of vertices]
5,5 [coordinate of the point, vertex xi]

(2) Vertex Dictionary Model:

Vertex dictionary uses a similar approach as spaghetti but a smarter structure. It uses two files to store the vector data. The first file stores the vertices and the second file stores the description of objects. In this data model, if some vertices are shared by two adjacent polygons, these vertices are not required to store twice. However, topological relationships are not well defined in this model. The problem of island polygons still exists. Therefore it is inferior for data analysis and query.



File 1
vertex x y

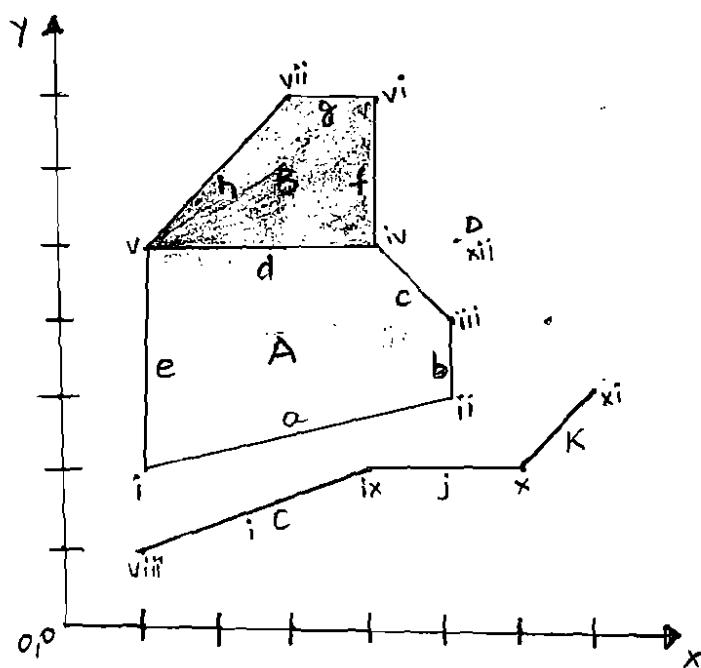
i	1	2
ii	5	3
iii	5	4
iv	4	5
v	4	5
vi	4	7
vii	3	7
viii	1	1
ix	4	2
x	6	2
xi	7	3
xii	5	5

File 2
polygonA : i, ii, iii, iv, v
polygonB : iv, vi, vii, viii
Line C : viii, ix, x, xi
Point D : xii

(3) DIME (Dual Independent Map Encoding) Model / GBF (Geographical Base File) :-

The DIME model was developed by the United States Bureau of the census. This is a complex model but more intelligent. It uses three files to represent the vector data. This model also avoids duplication of data for adjacent polygons and can establish several relationships between objects.

Directions of objects can also be determined by the starting vertex and ending vertex, however, this approach is not very efficient. The DIME was the first attempt to build explicit topological relationships. However it has several limitations compared to a fully topological model.



File 1

Vertex	x	y
i	1	2
ii	5	3
iii	5	4
iv	4	5
v	1	5
vi	4	7
vii	3	7
viii	1	1
ix	4	2
x	6	2
xi	7	3
xii	5	5

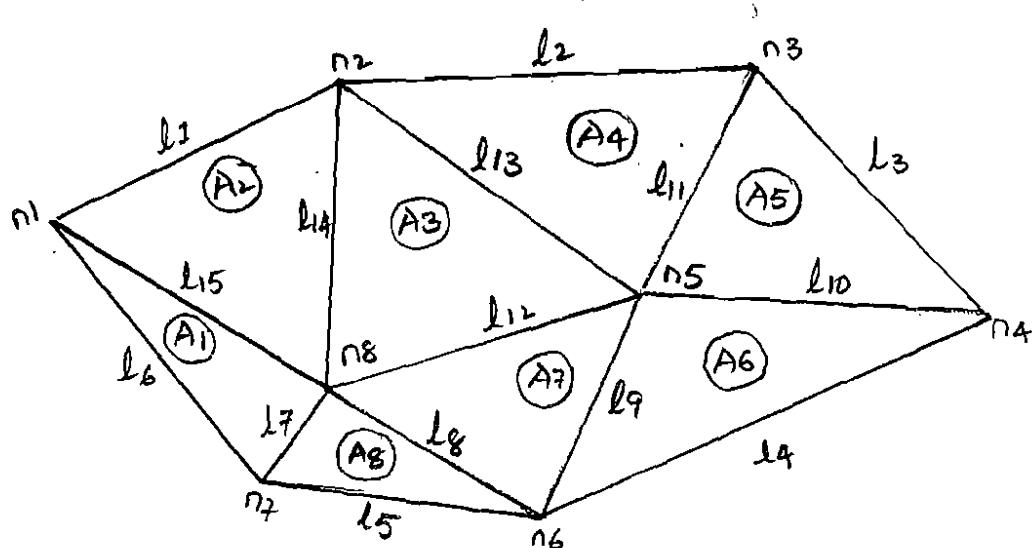
File 2

Segment	Right polygon	Left polygon	from vertex	To vertex	File 3
a	external	A	i	ii	Polygon segments A ab,c,d,e
b	external	A	ii	iii	B f,g,h,d
c	external	A	iii	iv	Line segments
d	B	A	vi	v	C i,j,k
e	external	A	v	vi	
f	external	B	iv	vi	
g	external	B	vi	vii	
h	external	B	vii	v	
i	-	-	viii	ix	
j	-	-	ix	x	
k	-	-	x	xi	

(4) Topological Models:

In order to use the data manipulation and analysis subsystem more efficiently and obtain the desired results, to allow advanced analytical techniques on GIS data and its systematic study in any project area, much explicit spatial information is to be created. The topological data model incorporates solution to some of the frequently used operations in advanced GIS analytical techniques. This is done by explicitly recording adjacency information into the basic logical entity in topological data structures, beginning and ending when it contacts or intersects another line, or when there is a change in the direction of the line. Each line then has two sets of numbers: a pair of coordinates and an associated node number. The node is the intersection of two or more lines, and its number is used to refer to any line to which it is connected. In addition, each line segment, called a link, has its own identification number that is used as a pointer to indicate the set of nodes that represent its beginning and ending polygon. These links also have identification codes that relate polygon numbers to see which two polygons are adjacent to each other along its length. In fact, the left and right polygon are also stored explicitly, so that even this tedious step is eliminated. This design feature allows the computer to know the

actual relationship among all its graphical parts, to identify the spatial relationships contained in an analog map. fundamentally, the topological models available in gis ensure (a) that no node or line segment is duplicated, (b) that line segments and nodes can be referenced to more than one polygon, and (c) that all polygons can be adequately represented. One possible topological data structure for the vector representation is shown below.



To understand the topological vector data structure, let us consider a network with 8 nodes encoded as n1 to n8. The links joining all these nodes encoded as l1 to l14 and the polygons created by all these line segments /links are coded as A1 to A8. The creation of this structure for complex area features is carried out in series of stages. Burrough identifies these stages as identifying a boundary network of arcs, checking polygons for closure, and linking arcs into polygon

The area of polygons can then be calculated and unique identification numbers attached. This identifier would allow nonspatial information to be linked to a specific polygon.

Table 4 provides the spatial data base along with the coordinate file, Table 1(a) of all the nodes and the corresponding attribute information can be given to each point, line and polygon by keeping the identification numbers.

Table 1

Link No.	Left node	Right node	Left polygon	Right polygon
l ₁	n ₁	n ₂	o	A ₂
l ₂	n ₂	n ₃	o	A ₄
l ₃	n ₃	n ₄	o	A ₅
l ₄	n ₆	n ₄	A ₆	o
l ₅	n ₇	n ₆	A ₈	o
l ₆	n ₁	n ₇	A ₁	o
l ₇	n ₇	n ₈	A ₁	A ₈
l ₈	n ₈	n ₆	A ₇	A ₈
l ₉	n ₅	n ₆	A ₇	A ₆
l ₁₀	n ₅	n ₄	A ₅	A ₆
l ₁₁	n ₃	n ₅	A ₄	A ₅
l ₁₂	n ₈	n ₅	A ₃	A ₇
l ₁₃	n ₂	n ₅	A ₃	A ₄
l ₁₄	n ₂	n ₈	A ₂	A ₃

Table 1(a)

node No.	x-coordinate	y-coordinate
n ₁	x ₁	y ₁
n ₂	x ₂	y ₂
n ₃	x ₃	y ₃
n ₄	x ₄	y ₄
n ₅	x ₅	y ₅
n ₆	x ₆	y ₆
n ₇	x ₇	y ₇

Field-based Vector Model:

Although vectors are ideal for representing discrete objects , it can also be used for field-based or continuous data such as elevation , temperature , etc.

Mass-points , contour lines / isolines are used to represent elevation or other continuously changing values. However it should be borne in mind that raster like continuity cannot be obtained by any of the aforementioned models.

Point Model:

Multiple points are used to represent the surface. For example , a pyramid can be represented by a minimum of five points , four for base and one for peak. Mass-point is a technique to represent surfaces using several points in a very dense manner.

But point model is not a good approach to represent surfaces, because it requires the viewer to imagine the surface joining the points.

Contour / Isoline Model:

Contour is an imaginary line of constant elevation on the ground surface. The corresponding line on a map is called a 'contour line', a line on a map that joins places of the same elevation (height) above sea level. Contour interval is the difference in elevation between two contour lines. Isoline is a line on a surface, connecting points of equal value such as temperature, rainfall, etc. Contour / Isolines can be used to represent surfaces. But even in this case, it requires the viewer to imagine the surface between two contour lines.

Raster vs Vector:

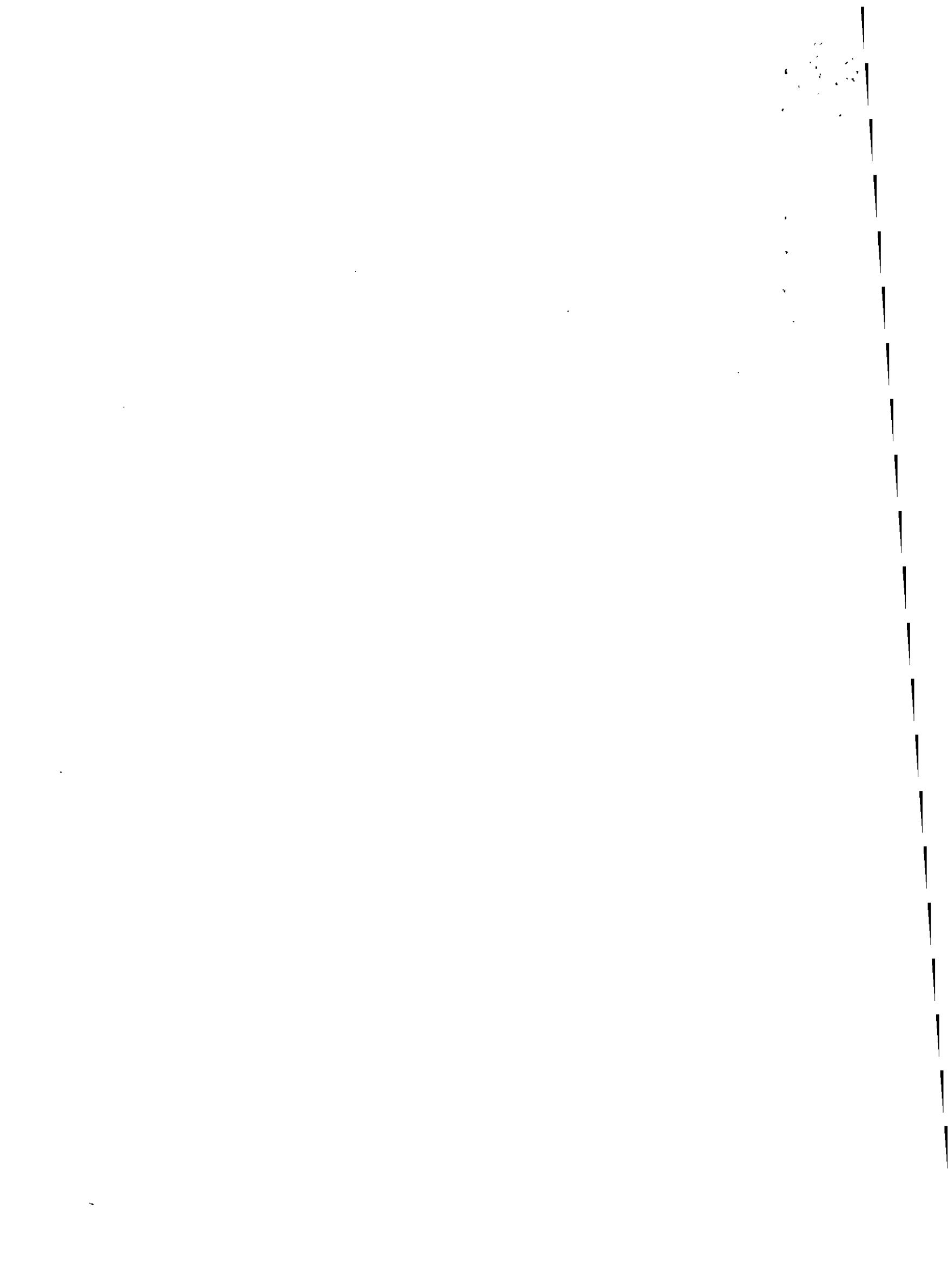
The traditional advantages and disadvantages of raster versus spatial data structures have been documented by Kennedy and Meyers. The basic issues include data volume, retrieval efficiency, data accuracy, data display, correctness to perturbation, and data manipulation, efficiency, and processing capabilities. Comparisons of data volume between raster and vector systems are entirely dependent upon the database elements, as well as consideration of accuracy and precision. A detailed comparison between raster model and vector model.

Raster Model Advantages	Vector Model Advantages
<p>1. It is a simple data structure</p> <p>2. Overlay operations are easily and efficiently implemented.</p> <p>3. High spatial variability is efficiently represented in a raster format.</p> <p>4. The raster format is more or less required for efficient manipulation and enhancement of digital images.</p>	<p>1. It provides a more compact data structure than the raster model</p> <p>2. It provides efficient encoding of topology, and, as a result, more efficient implementation of operations that require topological information, such as network analysis</p> <p>3. The vector model is better suited to supporting graphics that closely approximate hand-drawn maps</p>

Disadvantages	Disadvantages
<p>1. The raster data structure is less compact.</p> <p>2. Topological relationships are more difficult to represent</p> <p>3. The output of graphics is less aesthetically pleasing because</p>	<p>1. It is more complex data structure than a simple raster</p> <p>2. Overlay operations are more difficult to implement.</p> <p>3. The representation of high spatial variability is inefficient.</p>

boundaries tend to have a blocky appearance rather than the smooth lines of hand-drawn maps. This can be overcome by using a very large number of cells, but it may result in unacceptably large files.

4. Manipulation and enhancement of digital images cannot be effectively done in the vector domain.



Data Input Methods:

There are three methods of data input which are widely used: keyboard entry, automatic digitisation and scanning. Digital data must be downloaded from their source media and may require reformatting to convert them to an appropriate format for the GIS being used. Reformatting or conversion may also be required after analogue data have been converted to digital form. For example, after scanning a paper map, the file produced by the scanning equipment may not be compatible with the GIS, so it needs reformatting. For both the analogue and digital data, keyboard entry method, ~~manual digitising~~ and automatic digitising and scanning methods are very important.

Keyboard Entry:

Keyboard entry, often referred to as key coding, is the entry data into a file at a computer terminal. This technique is used for attribute data that are available only on paper. This technique can be mixed with digitising process for the creation of GIS database. The attribute data, once is digital format, are linked to the relevant map features in the spatial database using identification codes. There are unique codes that are allocated to each point, line and area feature in the dataset.

The coordinates of spatial entities like point, line and area features can be encoded by keyboard

entry. This method is used when the coordinates of these spatial entities are known and there are not too many of them. If the coordinates are more in number, this data can be encoded using digitising. The procedure of keyboard entry can be used to enter land record information. This method leads to obtain very high level of precision data by entering the actual surveying measurements. This method is used for the development of cadastral information system.

Scanning and Automatic Digitising

Scanning is the most commonly used method of automatic digitising. Scanning is an appropriate method of data encoding when raster data are required, since this is the automatic output format from most scanning software. Thus scanning may be used as a background raster dataset for the over-plotting of vector infrastructure data, such as pipelines and cables.

A scanner is a piece of hardware for converting an analogue source document to a digital raster format. There are two types of scanners, (i) flatbed scanner and (ii) rotating drum scanners. The cheapest scanners are small flatbed scanners, and high quality and large format scanners are rotating drum scanners in which the sensor moves along the axis of rotation.

A digital image of the map is produced by moving an electronic detector across the map surface. The size of the map area viewed by the detector

and scanner should be processed or edited to improve the quality and convert the raster to vector after online digitisation. The accuracy of the scanned output data depends on the quality of scanner, the quality of the software used to process the scanned data, and the quality of the source document. A very important feature that a GIS user should observe after scanning the paper map is the occurrence of splines, which is black appearance on the scanned output. This can be removed by a process called thinning.

The resolution of the scanner used affects the quality of output data. The cheaper flat-bed scanners have resolutions of 200-500 mm whereas the more expensive drum scanners use resolutions of 10-50 mm. The higher the resolution, the larger the volume of the data produced.

