Raster and Vector data model and structures

Geographic Information System is a computer system for capturing, storing, quiering, analyzing and displaying the geographical data. The GIS can be grouped to computer system, GIS software, Brain ware and infracture. Since late 1960's computer have been used to store and process geographic data. The Geographically referenced data distinguished GIS from other information system. Geographically referenced data describe both the location and characteristics of spatial features on the earth surface. GIS therefore involves two geographic data component.

The GIS model can be split into two parts

1. A Model of spatial form and
2. A model of Spatial Process.

The model of spatial form represents the structure and distribution of features in geographical space.

In order to model spatial process, the interaction between these features must be abstraction. By applying this abstraction process the GIS designer moves from the position of observing the geographical complexities of the real world to one of the stimulating them in the computer. These processes involved into

a. Identifying the spatial features from the real world that are of interest in the context of an application and choosing how to represent them in the conceptual model.
b. Representing the conceptual model by an appropriate spatial model.
c. Selecting an appropriate spatial data model within the computer.

Spatial Definition

Spatial data describes the location and shape of geographic features, and their spatial relationship to the features. The information contained in the spatial database is held in the form of digital co-ordinates which describe the spatial features it mainly depends on the latitude and longitude of the feature.

Spatial Entity

Spatial data describe the spatial situation of objects concerning their form and their relative situation in space. Usually, the spatial relation of individual points, lines or areas is made via the integration in a coordinate system resulting in the relation to the real world and the metrics. Geometry data can be available as raster data (pixel) or vector data (polygons / areas, lines, sites)

Spatial Data can be encoded using following spatial entities. They are

The Spatial data can be represented through using

1. The Point
2. The Line
3. The Area
4. Network
5. The Surface

Spatial Data Model

The GIS model can be split into two parts

A Model of spatial form and

A model of Spatial Process.

The model of spatial form represents the structure and distribution of features in geographical space.

In order to model spatial process, the interaction betweenthese features must be considered.

Computers require unambiguous instructions on how to turn data about spatial entities into graphical representations. This process is the second stage in designing and implanting a data model. At present there are trwo main ways in which computers can be display spatial entities. There are two method of spatial data model. They are

1. Raster Data Model
2. Vector Data Model

The Raster Data Model:

The Raster data model is one of the important spatial data model described as tessellation. In the raster world individual cells are used as the building block for creating images of point, line, area, and network and surface entities.

- In the raster data model the basic building block is the individual grid cell, and the shape and character of an entity is created by the grouping of cells. The size of the grid cell is very important as it Raster is a method for the storage, Processing and display of spatial data.

- Each area is divided into rows and columns, which form a regular grid structure. Each cell must be rectangular in shape, but not necessarily square.
- Each cell within this matrix contains location co-ordinates as well as an attribute value. The origin of rows and column is at the upper left corner of the grid.
- Rows function as the "y"coordinate and column as"x"coordinate in a two dimensional system. A cell is defined by its location in terms of rows and columns.

Vector Data Model

A vector spatial data model uses two dimensional Cartesian [x, y] co-ordinates to store the shape of spatial entity. In the vector spatial data can be represented by using point. It is the basic building blocks from which all spatial entities are constructed. The simplest spatial entity, the point is represented by a single co-ordinate pair. Line and area entities are constructed by connecting a series of points into chains and polygons. The more complex the shape of a line or area feature the greater the number of points required representing it.

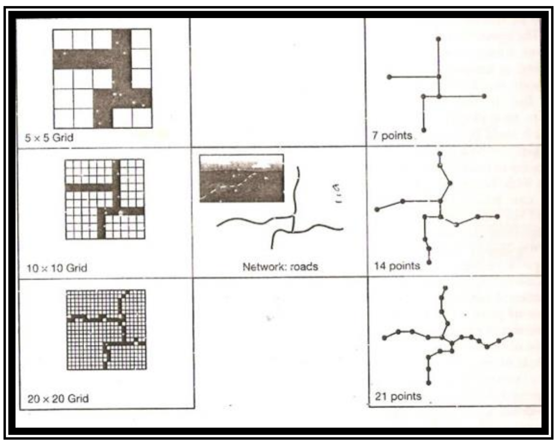
Spatial Data Structure

Data structures provide the information that the computer requires to reconstruct the spatial data model in digital form. There are many different data structure in use in GIS. Based on that the spatial data structures can be classified according to whether they are used to structure raster or vector data structure.

Raster Data Structures

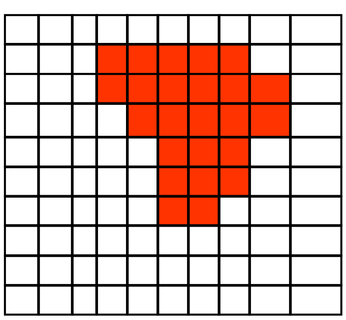
In the raster data model, each location is represented as a cell. The matrix of cells, organized into rows and columns, is called a grid.

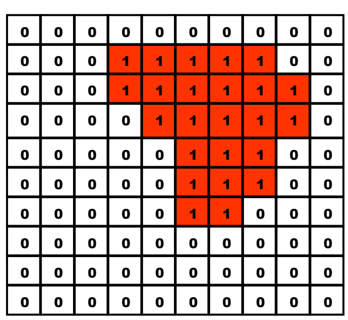
Each row contains a group of cells with values representing a geographic phenomenon. Cell values are numbers, which represent nominal data such as land- use classes, measures of light intensity or relative measures. The cells in each line of the image are mirrored by an equivalent row of numbers in the file structure.

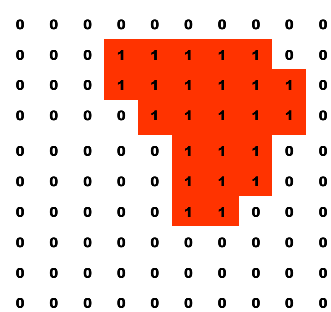


Effects of chaining resolution in the vector and raster data model

In a simple raster data structure, such as different spatial features must be stored as a separate data layers. However, if the entities do not occupy the same geographical location then it is possible to store them all in a single layer, with an entity code given to each cell.

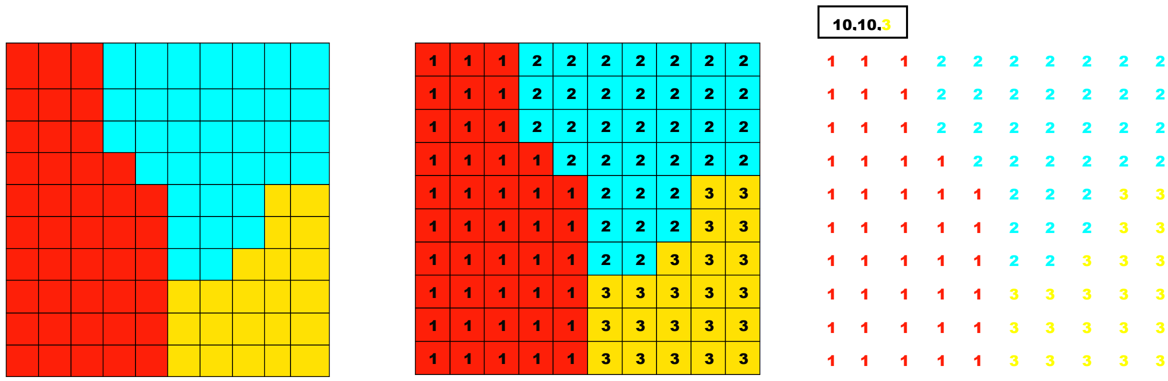






Entitv Model　　　Cell Value　　　File Structure

This figure shows the different land use can be coded in a raster layer. The values 1, 2 and 3 have been used to classify the raster cells according to the land use present at a given location. The values 1 represents residential area; 2 forest; and 3; farm land.



Entitv Model Cell Value File Structure

One of the major problems with raster data sets their size, because a value must be recorded and stored for each cell in an image. Thus a complex image made up of a mosaic of different features requires the same amount of storage space as a similar raster map showing the location of a single forest. To solve the problem in raster data model the compression or compaction methods are used for the real world representation.

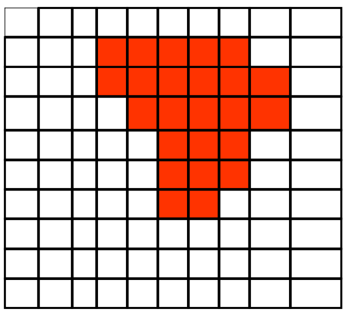
RASTER DATA COMPERSSIONAL METHOD

One of the major problems with raster data sets is their size, because a value must be recorded and stored for each cell in an image. A complex image made up of mosaic of different features requires the same amount of storage space as similar raster map. To overcome the problem the following raster representation method can be adopted. They are

1. Run Length Encoding Method
2. Block Coding Method
3. Chain Coding Method
4. Quadtree Method

Run Length Encoding Method

It is the raster image compression or compaction method. This technique reduces data volume on a row by row basis. It stores a single value where there are a number of cells of a given type in a group, rather than storing for each individual cell. The following diagram shows the RLE method of raster representation.

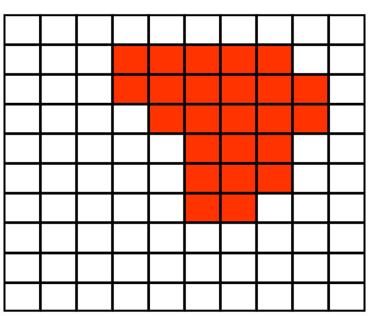


0	0	0	0	0	0	0	0	0	0
0	0	0	1	1	1	1	1	0	0
0	0	0	1	1	1	1	1	1	0
0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	1	1	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0

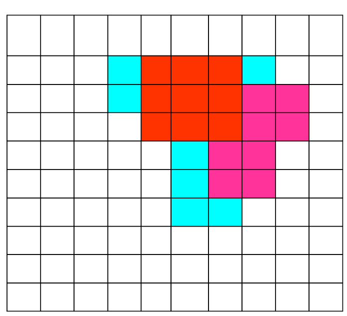
10,10,1		
0,10	**1,5**	**0,2**
0,3	**1,6**	**0,1**
0,3	**1,5**	**0,1**
0,4	**1,3**	**0,2**
0,5	**1,3**	**0,2**
0,5	**1,2**	**0,3**
0,5		
0,10		
0,10		

Block Coding Method

This approach extends the run length encoding idea to two dimensions by using a series of square blocks to store data. These are seven unit cells, two four- cell squares and one nine –cell square. Coordinate are required to locate the blocks in the raster matrix. The following diagram shows the block coding method of raster representation.



Entity Model



Cell Value

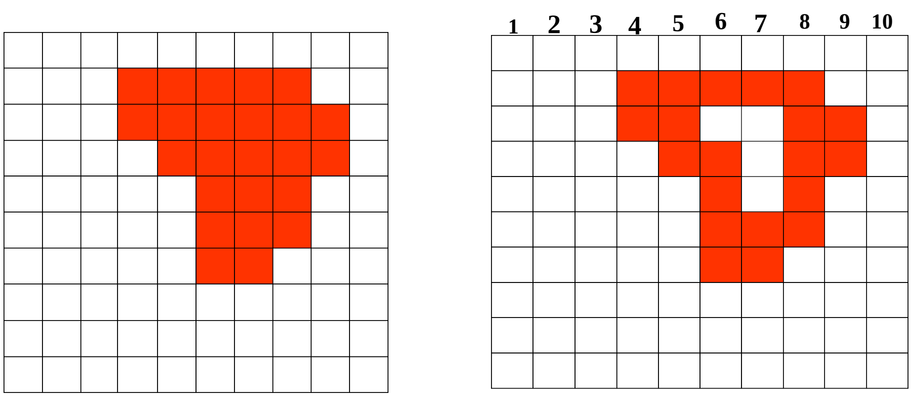
1		**4,2 8,2 4,3 6,5 6,6 6,7 7,7**
4		**8,3 7,5**
9		**5,2**

File Structure

Chain Coding Method

The chain coding method is an important technique to coding the raster data structure. This method of data can reduce the work by defining the boundary of the entity. The boundary is defined as a sequence of unit cells starting from returning to a given origin. The direction of travel around the boundary is usually given using a

numbering system. The following diagram shows the chain coding method.

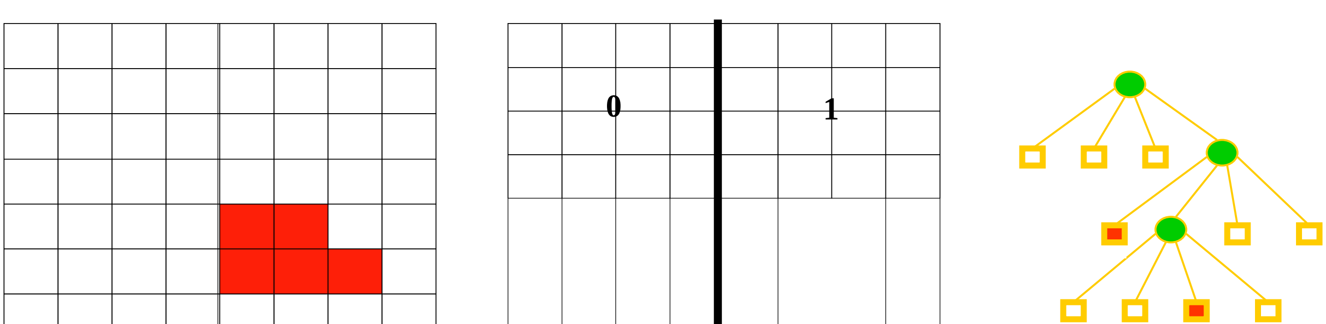


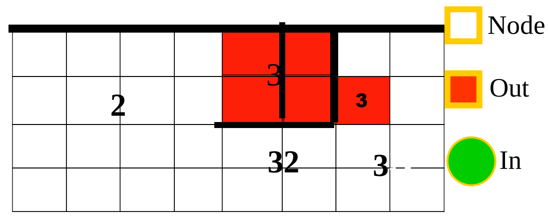
4,3

N,2 E,4 S,1 E,1 S,1 W,1 S,2 W,1 S,1 W,1 N,3 N,1 W,1

Quadtrees Method

One of the advantage of the raster data model is that each cell can be subdivided into smaller cells of the same shape and orientation._ Pequet in 1990.This data model has produced a range of innovative data storage and data reduction methods that are based on regularly subdividing space. The quadtree works on the principle of subdividing the cells in raster image into quads. This subdivision process continuous until each cell in the image can be classed as having the spatial entity either present or absent.



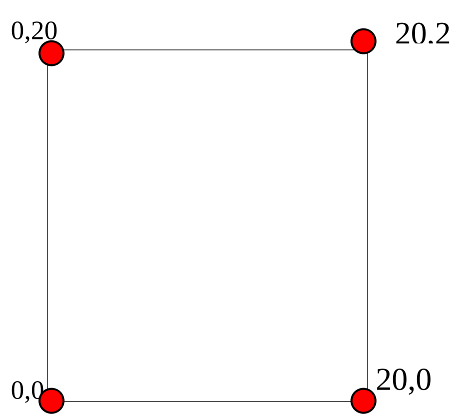


Vector Data Structure

The simplest vector data structure that can be used to produce a graphical image in the computer is a file containing (x,y) coordinate pairs that represent the location of individual point features.

The limitation of simple vector data structures start to emerge when more complex spatial entities are considered.

In the vector data all points in the data structure must be numbered sequentially and contain an explicit reference which record which points are associated with which polygon. This is known as Point Dictionary- Borrough 1986.



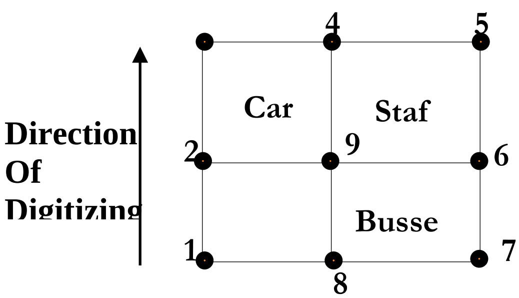
Spaghetti File	
X	**Y**
0	**0**
0	**20**
20	**0**
0	**0**
20	**20**

The vector data structure mainly ensure the following points. They are

- No node or line segment is duplicated;
- Line segment and nodes can be referenced to more than one polygon;
- All polygon have the unique identifiers
- Polygon can be adequately represented.

The following diagram shows the vector data structure

Point Dictionary Of Vector Data Structure

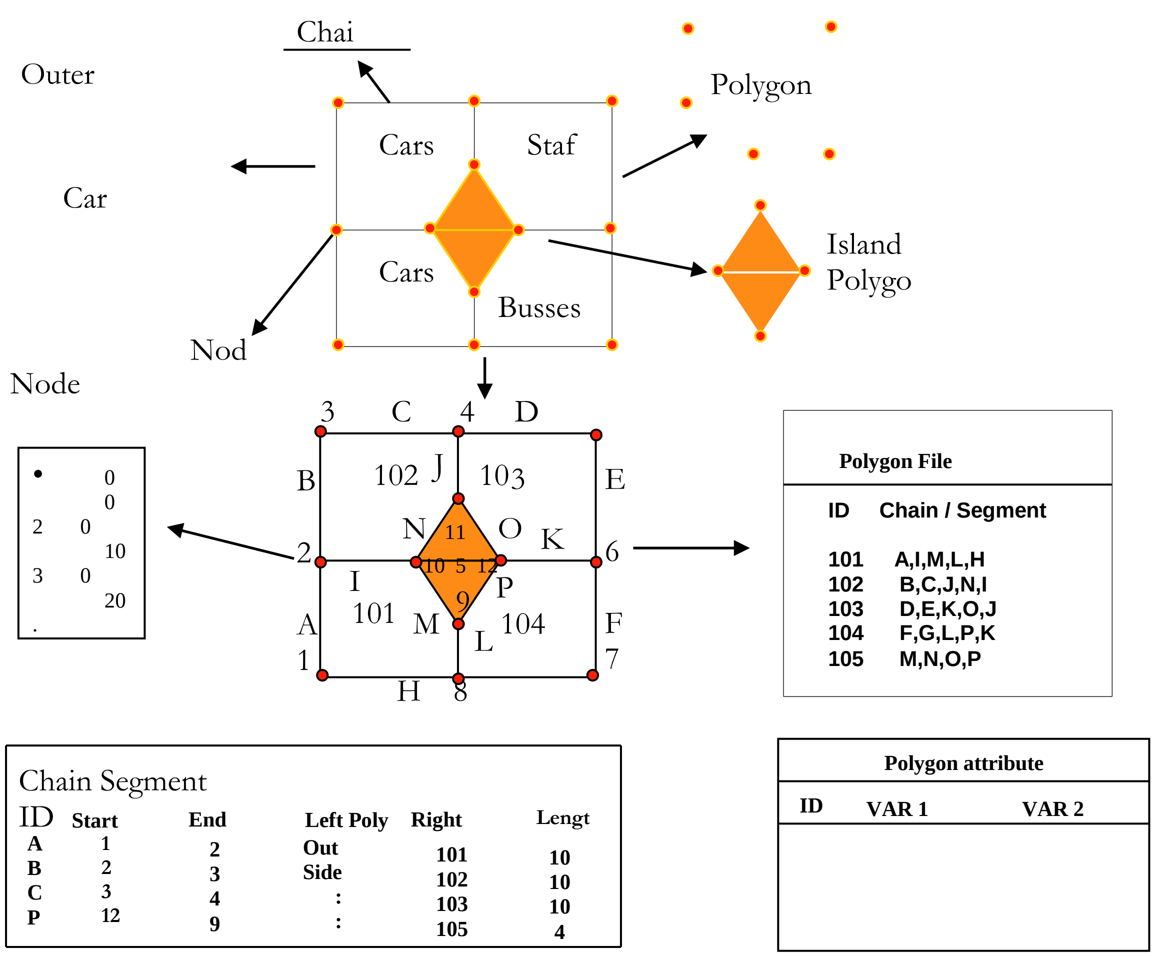


Polygon	
I	Point
Cars	1,2,3,4,8
Staff	,9
Busse	4.5.6.9

Point		
ID	X	
1	0	0
2	0	1
3	0	0
4	1	1
.	0	0
.	.	2
9	.	0
	1	.
	0	.

- In vector structure topology is concerned with connectivity between entities and not their physical shape.

- Boundaries are identified through network of arcs, checking polygons for closure, and linking arcs into polygons.
- The area of polygon can calculated and unique identification numbers are attached.
- This identifier would allow non spatial information to be linked to a specific polygon.



Comparison of raster and vector data models

The vector data model defines boundaries. There are no boundaries defined in the raster data model.

- The vector model represents location as x,y coordinates in a Cartesian coordinate system. The raster model represents location as cells, also in a Cartesian coordinate system. Rater data store rows and columns of cell values.
- The vector model represents features with well-defined boundaries; the raster model represents a more generalized view. The raster model can also represent gradual transition between features and surfaces, such as soil classification and elevation.
- The primary focus of the vector data model is the geographic feature; the primary focus of the raster data model is location.

- The vector model represents feature shape accurately; the raster model represents rectangular areas and thus is more generalized and less accurate.
- The vector model is used for high-quality cartography and where accuracy and precision are important, such as for cadastral (property) applications. The raster data model is useful for image/picture storage and is well suited to many spatial modeling operations such as modeling surface storm runoff and forest fire spread.
- The overlay operation examines two datasets to determine what geographic features exist at the same location. Overlaying vectors is a complex operation; the nature of the raster data model allows simple and fast overlays.

METHODS OF DATA INPUT

Data in analogue or digital form need to be encoded to be compatible with the GIS being used. This would be a relative straight forward task in all GIS packages uses the same spatial and attribute data models. How ever, there are many different packages and many different approaches to the handling of spatial and attribute data.

All data is analogue form need to be converted to digital form before they can be input into GIS. For methods widely used for the purpose are:

a) Key board entry

b) Automatic digitizing
c) Manual digitizing
d) Scanning

Key board entry may be appropriate for tabular data, or for small number of co ordinate pairs read from a paper may source or pocket GPS. Digitizing is widely used for the encoding of paper maps and data from interpreted air photographs. Scanning represents a faster encoding method for these data sources, although the resulting data source may require a considerable processing before analysis is possible.

Digital data must be down loaded from their source media (diskette, CD-ROM, or the internet) and may require reformatting to convert them to as appropriate format for the GIS being used. Reformatting or conversion may also be required after analogue data have been converted to digital format. For example, after scanning a paper maps, the file produced by the scanning equipment may not be compatible with the GIS, so reformatting may be necessary.

a) Key board Entry

Key board entry is after referred to as key coding. Key coding is the entry of data into a file at a computer terminal. This technique is used for attribute data that are only available on paper. E.g.: entering both the attribute and spatial data (location, is terms of postal codes) about the total number of hospitals with in the Trivandrum city limit etc. Key board entry operation is been simplified by the introduction of OCR. Text scanner and optical character recognition software can be used to read is characters automatically.

b) Digitizing

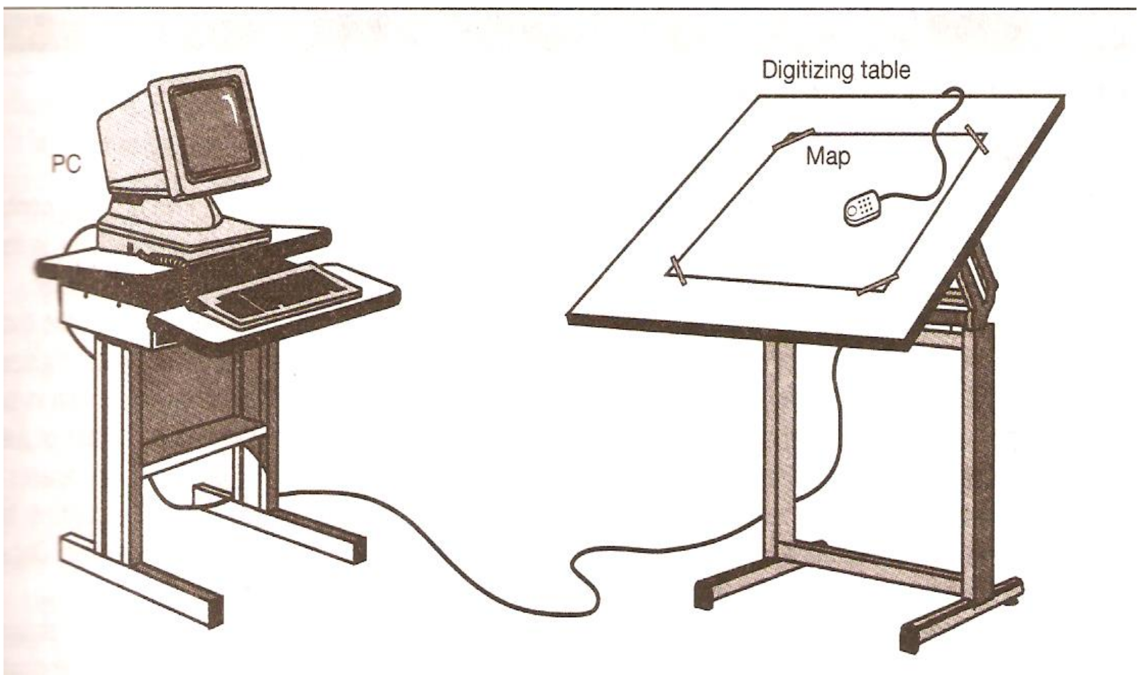
The most common method of encoding spatial features from paper maps is through digitizing. There are mainly two main methods is digitizing operation: they are.

i) manual digitizing
ii) automatic digitizing

i) <u>Manual digitizing</u>

A method employed is encoding spatial features from paper maps; is called as manual digitizing. It is an appropriate technique when a selection of features is required from a paper maps. Manual digitizing is also used for maps encoding where it is important to reflect the topology of features, since the information about the line feature can be included.

Manual digitizing requires a table digitizes that is linked to a computer work station. The table digitizer is a large flat tablet; the surface of which is underlain by a very fine mash of wires. Attached to the digitizer via a cable is a cursor that can be moved freely over the table. Buttons on the cursor allow the user to send instruction to the computer. The position of the cursor on the table is registered by reference to its position above the wire mesh.



Source: Ian Heywood & etal, An Introduction to Geographical Information System, Second edition 2003, Page No.93.

Most manual digitizers may be used in one of two modes: point mode or stream mode. In point mode the user begins digitizing each line segment with a start mode, records each change is direction of the line with a digitized point and finishes the segment with an end node. Thus a straight line can be digitized with just two points, the start and the end nodes. For more complex lines, a greater number of points are required between the start and end nodes. Smoot curves are problematic, since they require as infinite number of points to record their true shape. In practice, the user must choose a sensible number of points to represent a curve.

The manual digitizing of paper may is one of the main source of positional error in GIS. The accuracy of data generated by this method of encoding is dependent on many factors; including the scale and resolution of the source map, and the quality of the equipment and software being used. Errors can be introduced during the digitizing process by incorrect registration of map document on the digitize table or hand wobble. A shaky hand will produce differences between the line on the map and its digitized facsimile.

Five procedures to be followed while using a manual digitizer.

1. Registration

The map to be digitized is firmly fixed to the table top with sticky tape. Five or more control points are identified. The geographic co ordinates of the control points are noted and their locations digitized by positioning the cross hairs on the cursor exactly over there and pressing the digitize button on the cursor. This sends the co ordinates of a point on the table to the computer and save them is a file as digitizer co ordinates.

2. Digitizing point features

Point features for example, spot heights, location of buildings etc are digitized as a single digitized point. A unique code number is added so that attribute information may be attached later.

3. Digitizing line features

Line features are digitized as a series of points that the software will join with straight line segments.

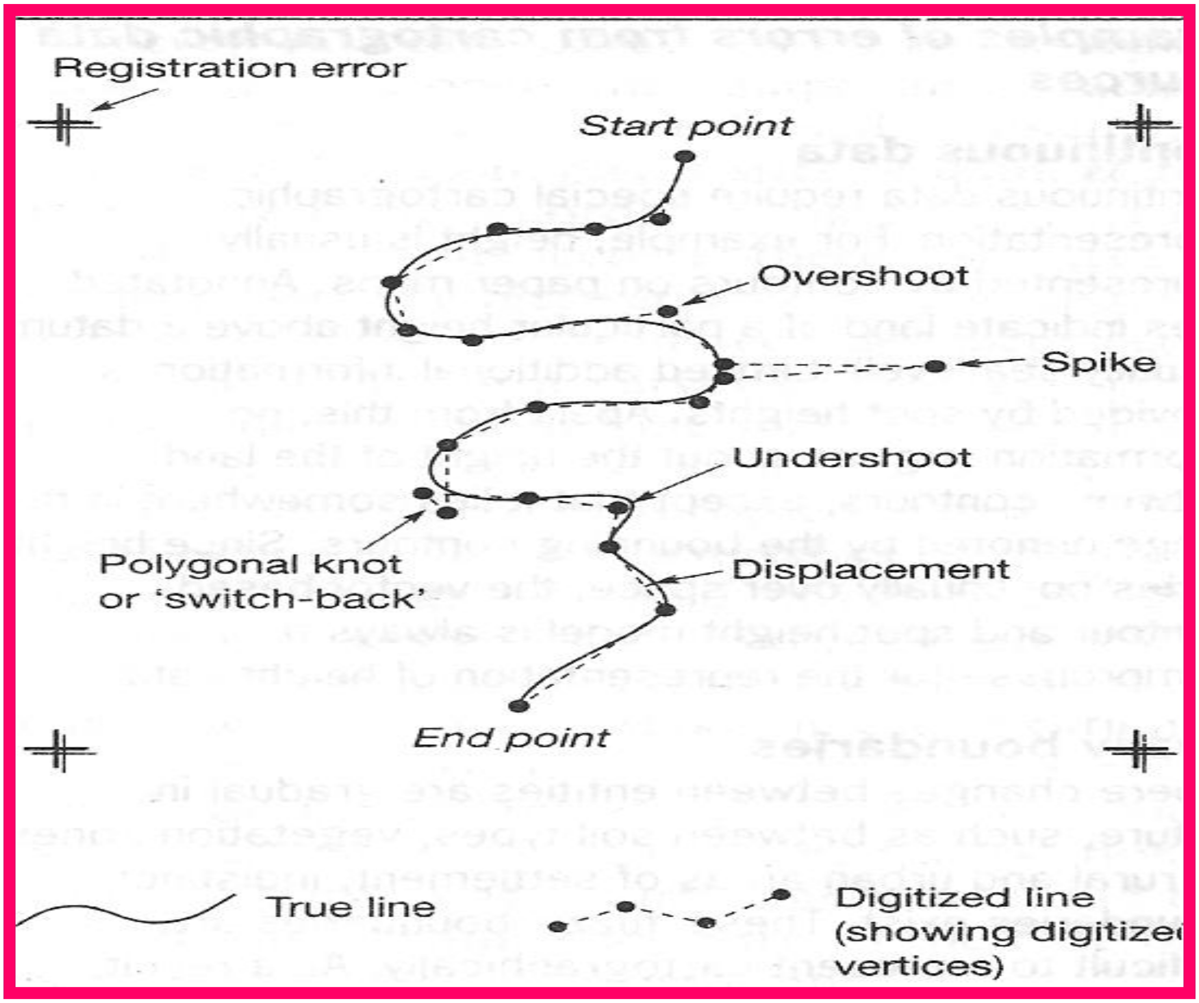
4. Digitizing area features

Area features for example, forested areas or administrative boundaries are digitized as a series of points linked together by line segments in the same way as line features. Here it is important that the start and end points join to form a complete area. Polygons can be digitized as a series of individual lines, which are later jointed to form areas.

5. Adding attribute information

Attribute data may be added to digitized polygon features by linking then to a centroid in each polygon. These are digitized manually or created automatically once the polygon has been encoded. Using a unique identifier or code number, attribute data can then be linked to the polygon controids of appropriate polygon.

SOME SOURCES OF ERRORS IN DIGITIZING



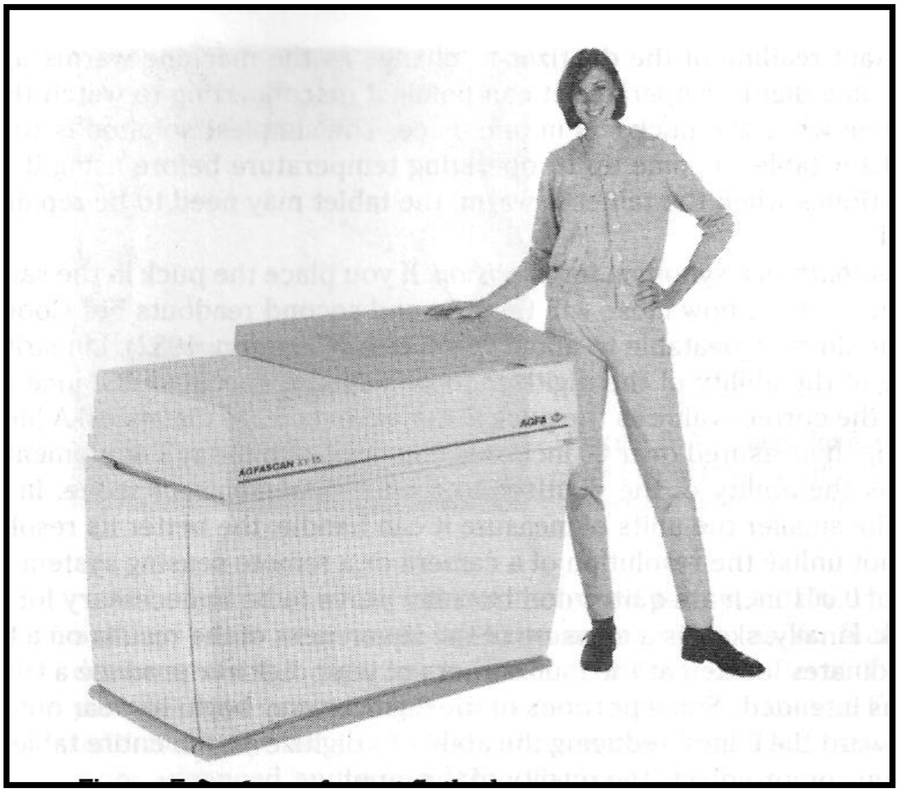
Source: Ian Heywood & etal, An Introduction to Geographical Information System, Second edition 2003, Page No.198.

b) Automatic digitizing

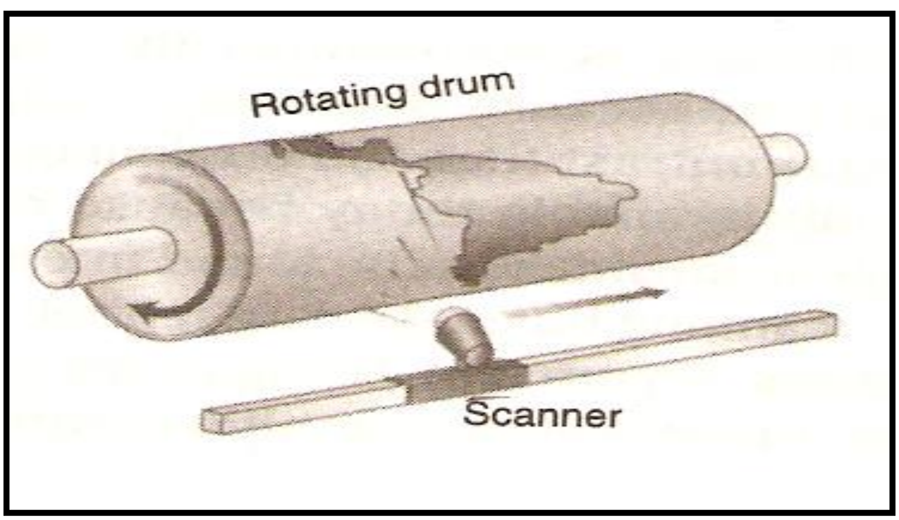
Manual digitizing is a time consuming and tedious process. If large number of complex maps need to be digitized it is worth. Considering the alternative, although often more expensive, methods. Two automatic digitizing methods are considered here; scanning and automatic line follower or following.

1) Scanning.

Scanning is the most commonly used method of automatic digitizing. Scanning is an appropriate method of data encoding when raster data are required; since this is the automatic output format from most of scanning software's.



LARGE FORMAT FLAT BED SCANNERS



ROTATTING DRUM SCANNER

A scanner is a piece of hardware for converting as analogue source document into digital raster format. All scanners work by sampling the source document using transmitted or reflected light.

The cheapest scanners are the small flat bed scanners a common pc peripheral. High quality and oral large format scanners require the source document to be placed on the rotating drum; and sensors move along the axis of rotation.

There are some practical problems faced when scanning source document; they are

- The possibility of optical distortion when using the flat bed scanners.
- The automatic scanning of unwanted information.
- The selections of appropriate scanning tolerance to ensure important data are encoded and background data are ignored.
- The format of files produced and the input of data to GIS software.
- The amount of editing required producing data suitable for analysis.

Any way there are three different types of scanners in wide spread use; they are the flat bed scanners, rotating drum scanners and large format feed scanners.

Large format feed scanners are most suitable for capturing data for input to GIS as they are relatively cheap, quick and accurate. Flat bed scanners, while being a common Pc peripheral are too small and in accurate and drum scanners despite being very accurate tend to be too slow and expensive.

All scanners works on the same principle. A scanner has a light source, a background and lenses. During scanning absence or presence of light is detected as one of the three components move part the other two.

It the output is to be used as a visual backdrop for other data is GIS then quality may not be a major issue. (Quality can be influenced by the setting of a threshold above which all values are translated as white and below which the values are black. Setting brightness and contrast level can also affect the quality of images obtained.

Resolution of scanners usually measured is dots per inch (dpi) as a linear measurement along the scan line. Commonly used resolutions are 200 dpi for text, 300 dpi for line maps and 400 dpi for high quality orthophotos.

2. Automatic line follower

Another type of automatic digitizer is then automatic line follower. This method mimics manual digitizing and uses a laser and light sensitive device to follow the lines on the map. Where scanners are raster devices, the automatic line follower is a vector device and produce output as (x, y) co ordinate strings. Difficulties may be faced when digitizing features such as dashed or contour lines. Hence arises the necessity of considerable editing and checking of data.

ELECTRONIC DATA TRANSFER

The difficulties and the time associated with keyboard encoding, manual digitizing, and automatic digitizing, the prospect of using data already in digital form is appealing.

Spatial data may be collected in digital form and transferred from devices such as GPS receivers, total stations and data toggers attached to all manner of scientific monitoring equipments. Electronic data transfer is also necessary if the data have been purchased form a data supplier, or another agency than originally encoded the data.

Thus electronic data transfer is an appropriate method of data encoding where the data are already available in digital form. Some times reformatting is also needed for the data before it can be used for further works.

MODELLING THE THIRD DIMENSION AND FOURTH DIMESION IN GIS

Over the recent years various datasets have been collected. This can be serve different purposes in general, spatial data represents of reality. When used in a GIS environment, can create various landscape scenarios are able to visualize them in two dimensional, three dimensional or even four dimensional.

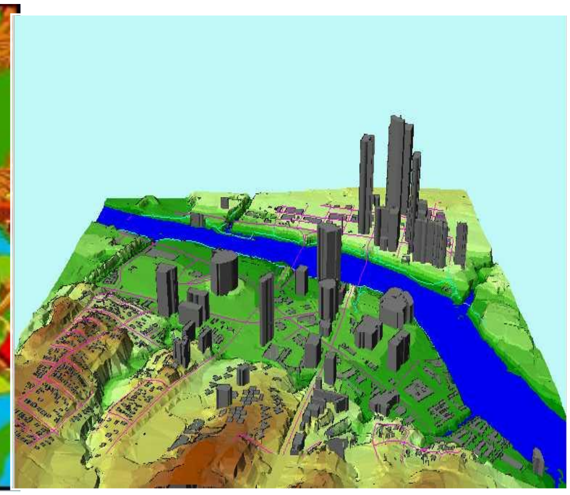
MODELLING THE THIRD DIMENSION

- In the real world all the features are seen exactly on 3D
- In terms of its display capabilities the computer screen is a two-dimensional display device even though the use of clever graphics it is possible to stimulate the appearance of the third dimension.
- In GIS the only medium for 3D is the display of computer screen. To produce system capable of representing the complexities of the real world, we need to portray the third dimension in more than a visual way.
- To produce system capable of representing the complexities of the real world. we need to portray the third dimension in more than a visual way.
- The representation of the third dimension of an entity can also help us model the form of entity and associated spatial process. the 3D is an integrated part of GIS tool box.
- Worboys (1995) describes how the simple raster grid cell an e extended into 3D representation; known as a voxel is that it is unable to record topological information .worboys suggests that the solution to this problem lies in extending the vector approach by using constructive. Solid geometry to create 3D objects such as cubes, spheres and cylinders.

TWO DIMENSIONAL MAP

THREE DIMENSIONAL MAP

THREE DIMENSIONAL MAP WITH Z COORDINATES

MODELLING THE FOURTH DIMENSION

- Gis is often a long –term process it is more than likely that his collage of data will include entities at different period in time.
- Modelling time is made more complex since there are several different sorts of time that Gis developers might to consider work practice time, data base time, and future time.
- Work practice times the temporal state of a gis data base used by many people.
- The 4D is having a temporal character.
- The problem with developing a data structure that is flexible enough to handle the time elements of entities include what to include how frequently to update the information.
- Most users of as represent the different temporal states of a spatial entity either by creating a separate data layer of each time period or y recording the state of the entity at a given time.

RELATIONAL DATABASE MODEL

Data Base Data models

There are a number of different database data models. Amongst those that have been used for attribute data in GIS are the hierarchical, network, relational and object-oriented data models of these the relational data model has become the most widely used.

Relational data base model:

A relational database is a collection of tables also called relations, which can be connected to each other by keys.

i. *The relational data model is based on concepts proposed by Codd (1970)*
ii. *Data are organized, in a series of two-dimensional tables, each of which contains records for one entity.*
iii. *These tables are linked by common data known as keys. Queries are possible on individual tables or on groups of tables.*

Fig. 1 illustrates an example of one such table

Hotel ID	Name	Address	Number of rooms	Standard
001	**Maintain View**	**23 High Street**	**15**	**Budget**
002	**Palace Deluxe**	**Pine Avenue**	**12**	**Luxury**
003	**Ski lodge**	**Lo Ski School road**	**40**	**Standard**

Fig. 1 Relational Database table data for Happy Valley

Each table in a relational database contains data for one entity. In the example this entity is 'hotel'. The data are organized into rows and columns, with the columns containing the attributes of the entity. Each of the columns has a distinctive name, and each of the entries in a single column must be drawn from the same domain (where a domain may be all integer values, or data, or text). With in a table, the order of the columns has no special significance.

The terminology of relational database can be confusing, since different software vendors have adopted different terms for the same thing. Table 1 represents the relationship between relational database terminologies and the traditional table, or simple computer file. Fig.2 applies this terminology to the table suggested for the Happy Valley database. A useful shorthand way of describing a table is using its intension. For the table in Fig.1 this would be:
HOTEL (Hotel ID, Name, Address, No of rooms, Standard)

Paper Version	File Version	RDBMS
Table	File	Relation
Row	Record / Case	Tuple
Column	Field	Attribute
Number of Columns	Number of Fields	Degree
Number of Rows	Number of Cases	Cardinality
Unique ID	Primary Key Possible Values	Index Domain

Table 1: Relational Database Terminology

SQL (structural query language)

i. The data is a relational database is stored as a set of base tables with the characteristics described above.

ii. The table structure is extremely flexible and allows a wide variety of queries on the data.

iii. The advantages of SQL for database users are its

- **Completeness**
- **Simplicity**
- **Pseudo English – Language Style and**
- **Wide application**

RELATION: HOTEL

Attributes

Hotel ID	Name	Address	Number of rooms	Standard
001	**Mountain View**	**23 High Street**	**15**	**Budget**
002	**Palace Deluxe**	**Pine Avenue**	**12**	**Luxury**
003	**Ski Lodge**	**10, Ski School road,**	**40**	**Standard**

Primary Key **Tuples**

Fig.2 Database terminology applied to Happy Valley Table

i. The availability of SQL is one of the advantages of the relational database model.

ii. Additionally, the model has a sound theoretical base in mathematics, and a simple logical data model that is easy to understand.

iii. The relational database model is more flexible than either of the previously used hierarchical or network models.

<u>LIMITATIONS:</u>

Seaborne (1995) considers that many of the limitations of relational databases in GIS stem from the fact that they were developed to handle simple business data, not complex multi-dimensional spatial data. However they have been widely adopted and successfully used.

Relational databases are predominantly used for the handling of attribute data in GIS. For example, ARC/INFO maintains an attribute table in relational database software, using a unique ID to link this to spatial data.

LINKING SPATIAL AND ATTRIBUTE DATA

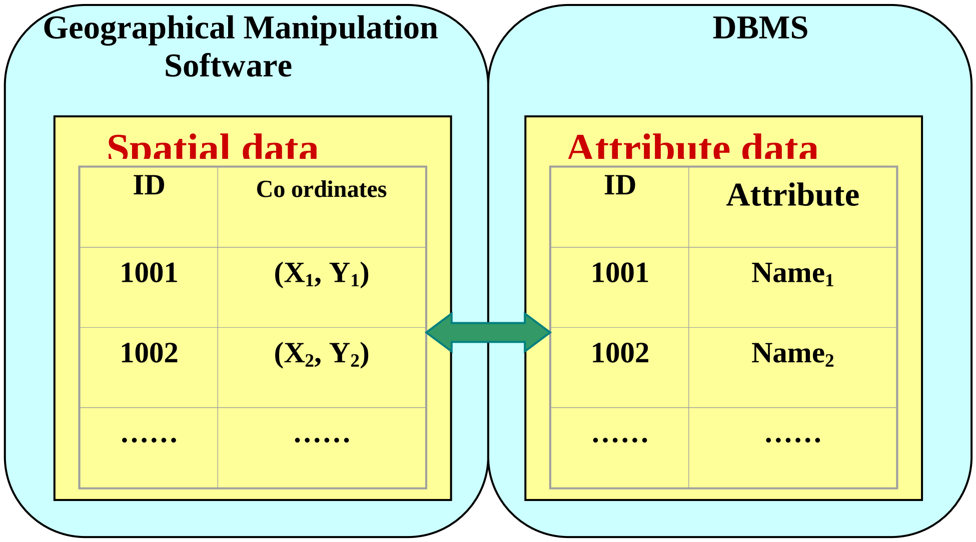
***SPATIAL DATA*:** Describes the location and shape of geographic features, and their spatial relationship to other features. The information contained in the spatial database is held in the form of digital co-ordinates, which describe the spatial features. These can be points (for example, hospitals), lines (for example, roads), or polygons (for example, administrative districts.) Normally, the different sets of data will be held as separate layers, which can be combined in a number of different ways for analysis or map production.

THE ATTRIBUTE DATABASE: The attribute database is a more conventional type; it contains data describing characteristics or qualities of the spatial features (i.e., descriptive information)

Linking spatial and attribute data: The relationship between GIS and databases varies. For a simple raster GIS, where one cell in a layer of data contains a single value that represents the attributes of that cell, a data base is not necessary. Here the attribute values are likely to be held in the same file as the data layer itself. However there are few 'real' GIS like this and those which exist are designed for analysis, rather than attribute data handling. An improvement on this approach is the ability to handle attribute values in a file separate from the raster image. Although this method also lacks the flexibility of a true relational DBMS, it is possible to link the GIS software with propriety relational DBMS to upgrade the capabilities. Most GIS, particularly vector-based systems, offer a hybrid approach. In this case spatial data are stored as part of the GIS data structure, and attribute data are stored in a relational DBMS. This approach allows integration of existing databases with graphics by the allocation of a unique identifier to each feature in the GIS.

Finally, an alternative approach is an extended GIS, where all aspects of the spatial and attribute data are in a single DBMS.

GIS TOOLS



Linking Spatial and attribute data in GIS