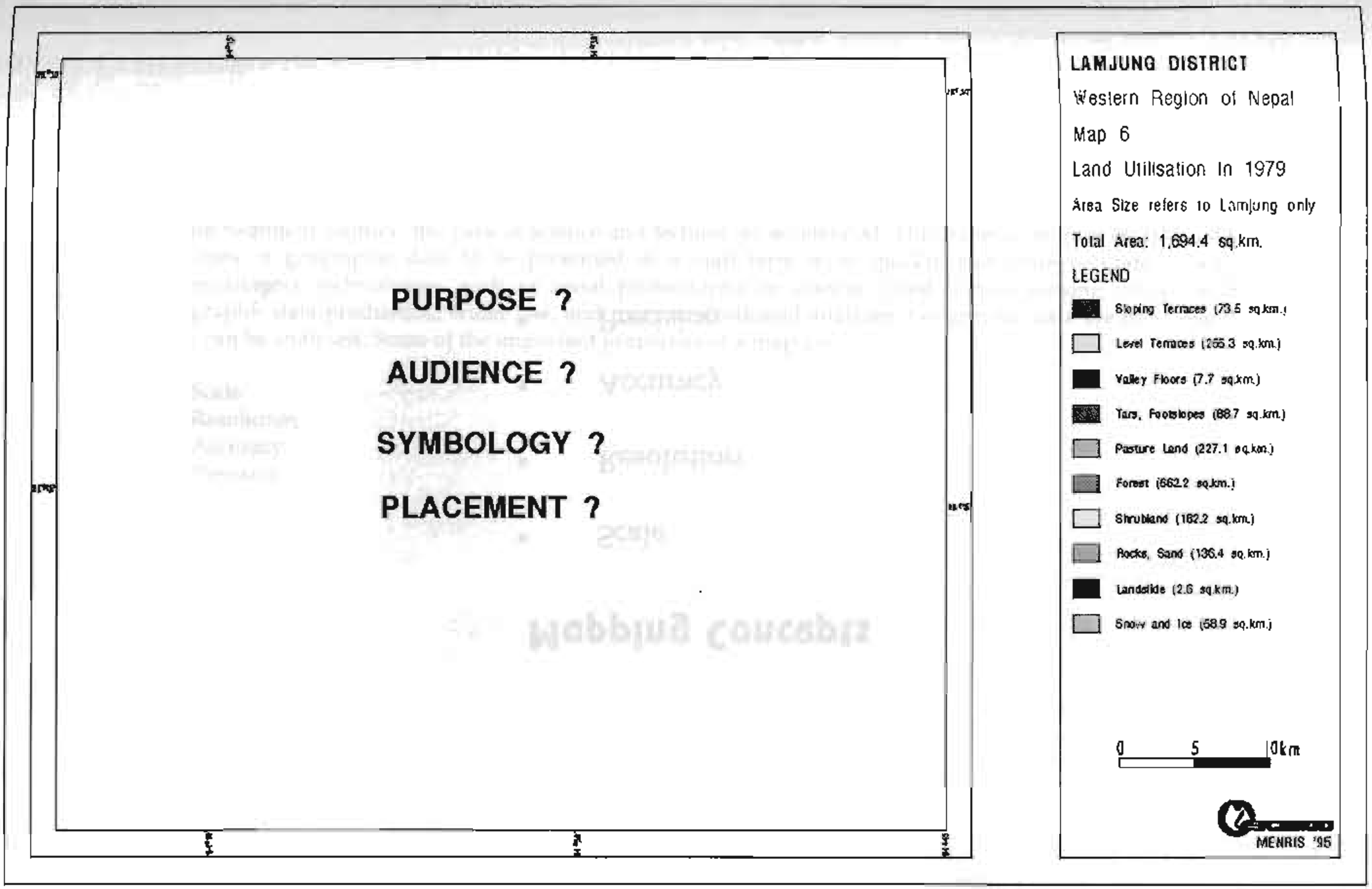
Chapter 3
GIS Concepts


PURPOSE ?
AUDIENCE ?
SYMBOLOGY ?
PLACEMENT ?
LAMJUNG DISTRICT
Western Region of Nepal
Map 6
Land Utilisation In 1979
Area Size refers to Lamjung only
Total Area: 1,694.4 sq.km.
LEGEND
Sloping Terraces (73.5 sq.km.)
Level Terraces (255.3 sq.km.)
Valley Floors (7.7 sq.km.)
Tars, Footslopes (88.7 sq.km.)
Pasture Land (227.1 sq.km.)
Forest (662.2 sq.km.)
Shrubland (182.2 sq.km.)
Rocks, Sand (136.4 sq.km.)
Landslide (2.6 sq.km.)
Snow and Ice (58.9 sq.km.)
0 5 10km
MENRIS '95

Mapping Concepts

- Scale
- Resolution
- Accuracy
- Precision

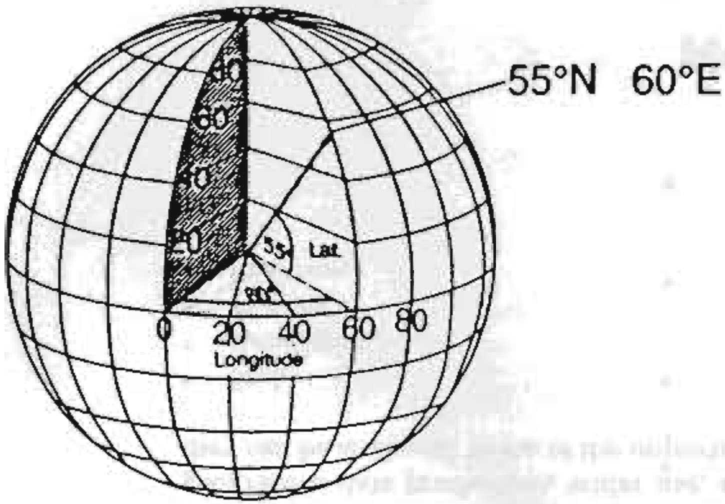
Mapping Concepts

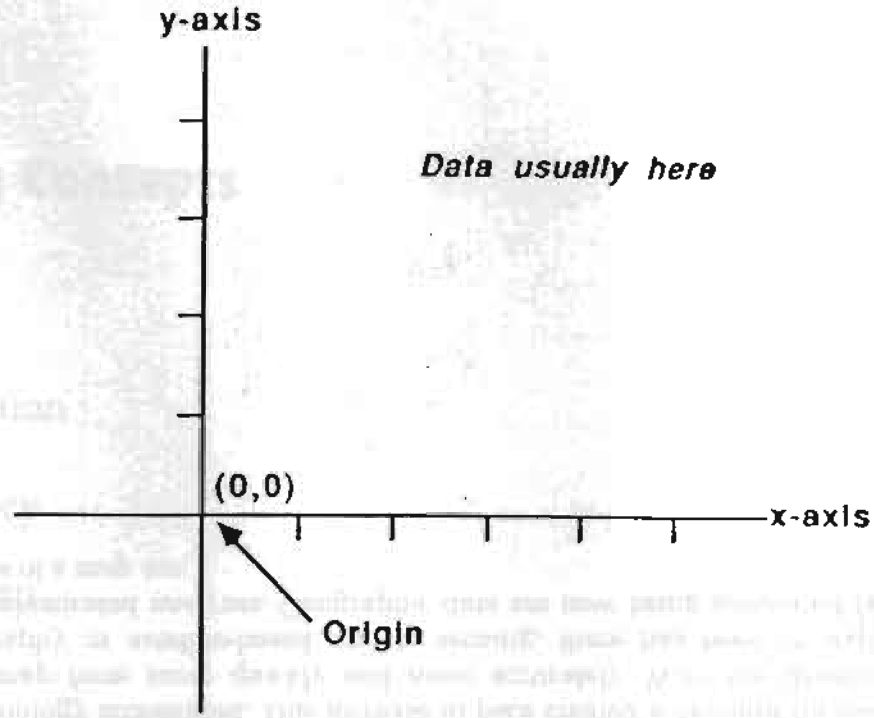
Maps have been used since the earliest times to portray information about the earth's surface. A map is also an information system, normally representing on a certain scale and on a flat medium a selection of material or abstract features on, or in relation to, the surface of the earth.

In the twentieth century, the pace of science and technology accelerated. This increase in pace created a demand for ever greater volumes of geographic data to be presented in a map form more quickly and more accurately. With the development of reconnaissance technologies, such as aerial photography or satellite-based remote sensing, there has been an explosion of geographic data production, wider use, and more sophisticated analyses. Geographic data are now being generated faster than they can be analysed. Some of the important properties of a map are:

- Scale
- Resolution
- Accuracy
- Precision

Coordinate System





- Spherical coordinate system
- Cartesian coordinate system

Projection System

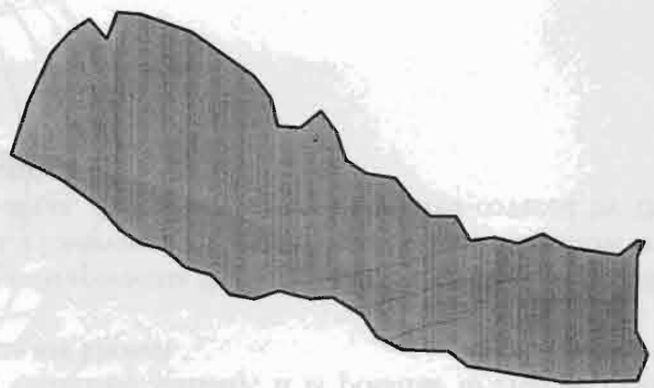
The important globally used coordinate systems are:

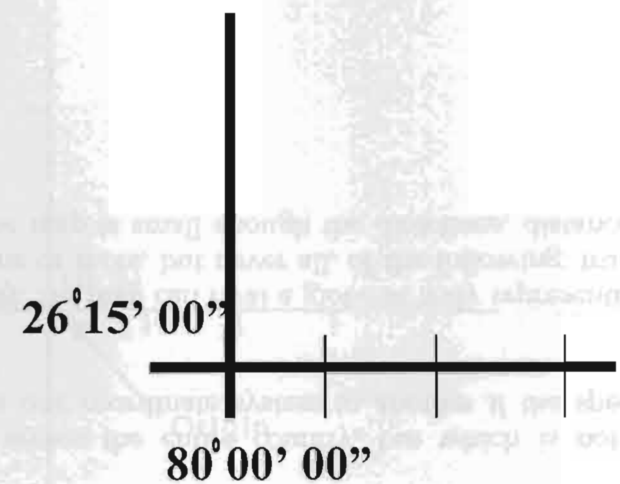
- geographic coordinates (latitude, longitude), (degree, minute, second: DMS),
- UTM (Universal Transverse Mercator) co-ordinates (X,Y) (metres), and
- Lambert Conformal Conic (metres)

Very often a country has its own coordinate system which serves the entire country, but which is not valid for other (neighbouring) countries. Usually, it is possible to change from one coordinate system to another if the specifications of the different systems are known.

Every flat map misrepresents the surface of the earth in some way. No map can rival a globe in truly representing the surface of the entire earth. However, a map or parts of a map can show one or more, but never all, of the following: true directions, true distances, true areas, and true shapes. If the area covered by the map is small enough the directions, distances, and areas are reasonably accurate.

Locating Nepal in Real-World Coordinates





- Global locations
- State Plane Coordinate System
- Latitude & Longitude

The Geographic Coordinate system is a spherical coordinate system composed of parallels of latitude and meridians of longitude. Both divide the circumference of the earth into 360 degrees which are further subdivided into minutes and seconds. Unlike the equator in the latitude system, there is no natural zero meridian. In 1884, it was finally agreed that the meridian of the Royal Observatory in Greenwich, U.K., would be the Prime Meridian.

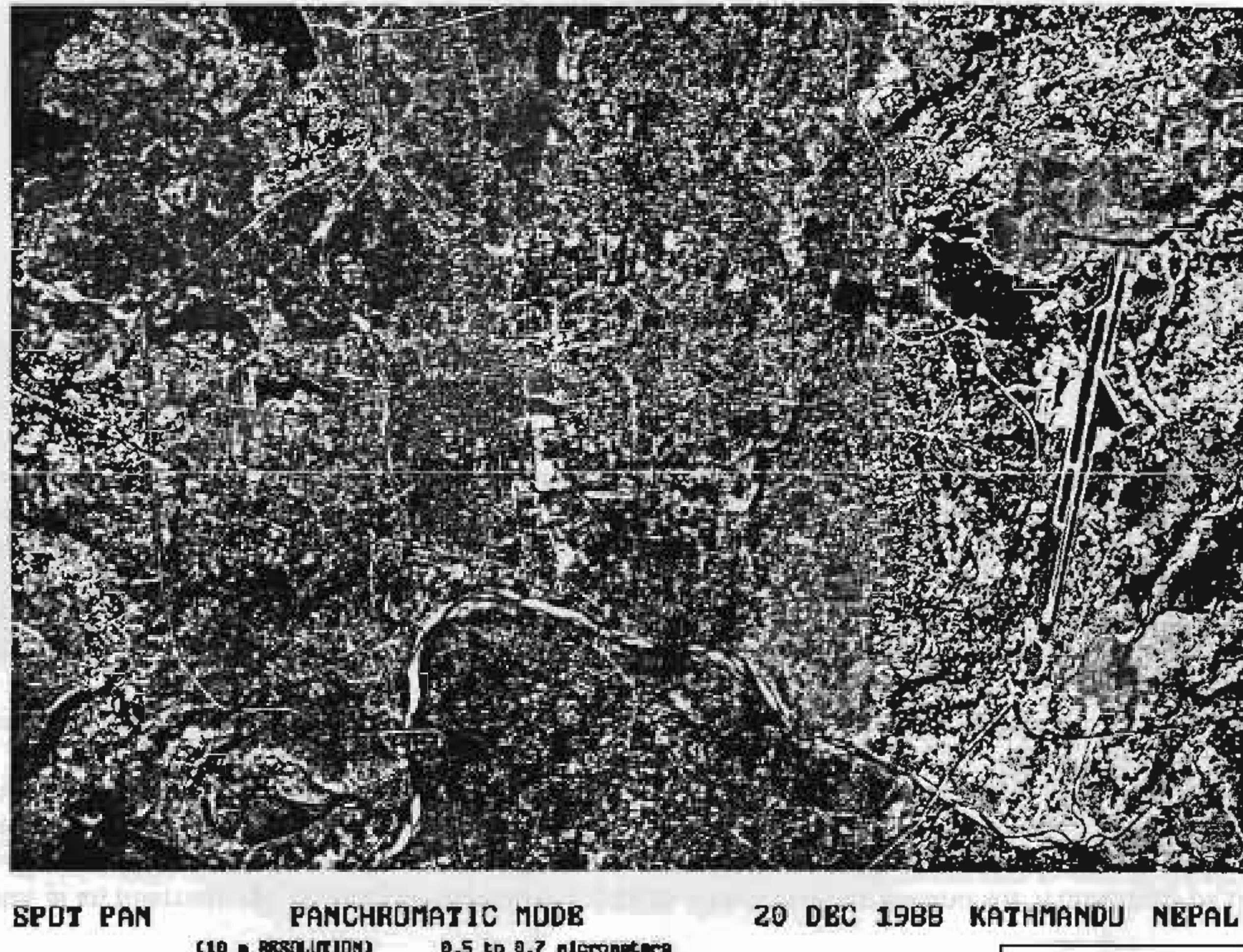
Mercator is an international rectangular coordinate system which extends around the world from 84 degrees North to 80 degrees South. The world is divided into 60 zones, each covering six (6) degrees longitude. Each zone extends three (3) degrees eastwards and three degrees westwards from its central meridian. Zones are numbered west to east from the 180 degree meridian. Because of the small area covered by each zone, a high degree of accuracy is possible. For example, Nepal falls under zones 44 and 45.

Problems arise when a map extends into two UTM zones. The entire map will then have to be projected into one of the two zones.

The advantage of using UTM coordinates is its metric nature. Normal calculations can be performed on UTM coordinates, while for geographic coordinates the minutes and seconds have to be first transformed into the decimal system.

Calculation of a map projection requires definition of the spheroid. As the earth is not a perfect globe (i.e., the earth is 'flattened' at the poles), a spheroid is defined in terms of axes lengths and eccentricity of the earth. Several principal spheroids are in use by one or more countries. Differences are due primarily to calculation of the spheroid for a particular region of the earth's surface. Some important spheroids are the Clarke 1866, the Bessel and the New International 1967. In the HKH Region, in general the Everest spheroid is applied.

Geographic Data



City Name
Address range to the airport
Address range to the park
Street No., Direction of Travel

Geographic Data

Although the two terms, data and information, are often used indiscriminately, they both have a specific meaning. Data can be described as different observations which are collected and stored. Information is data which is useful in answering queries or solving a problem. Digitising a large number of maps provides a large amount of data after hours of painstaking work, but the data can only render useful information if used in analysis.

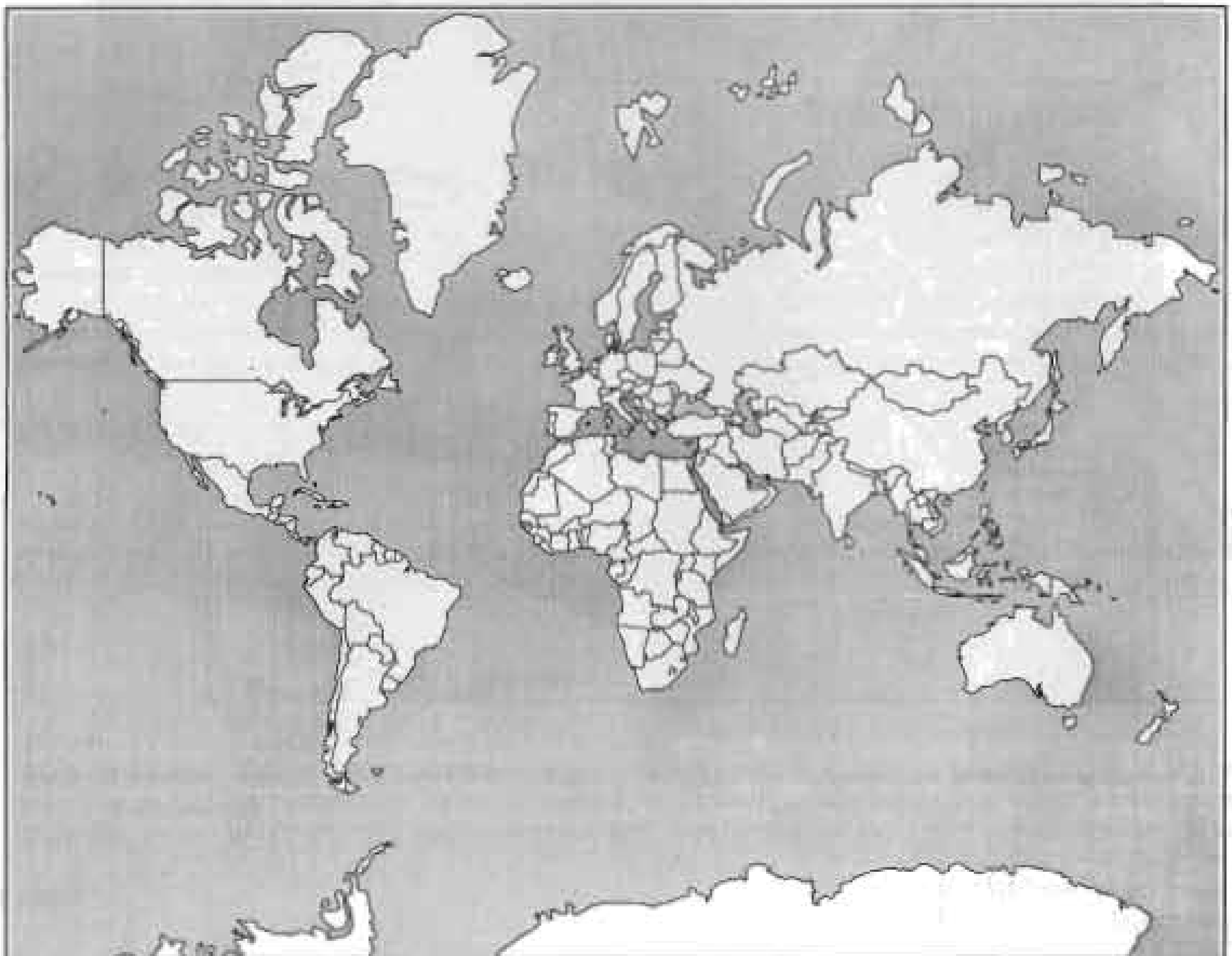
Spatial and Non-Spatial data

Geographic data are organised in a geographic database. This database can be considered to be a collection of spatially referenced data that acts as a model of reality. There are two important components of this geographic database: its **geographic position** and its **attributes or properties**. In other words, spatial data (where is it?) and attribute data (what is it?)

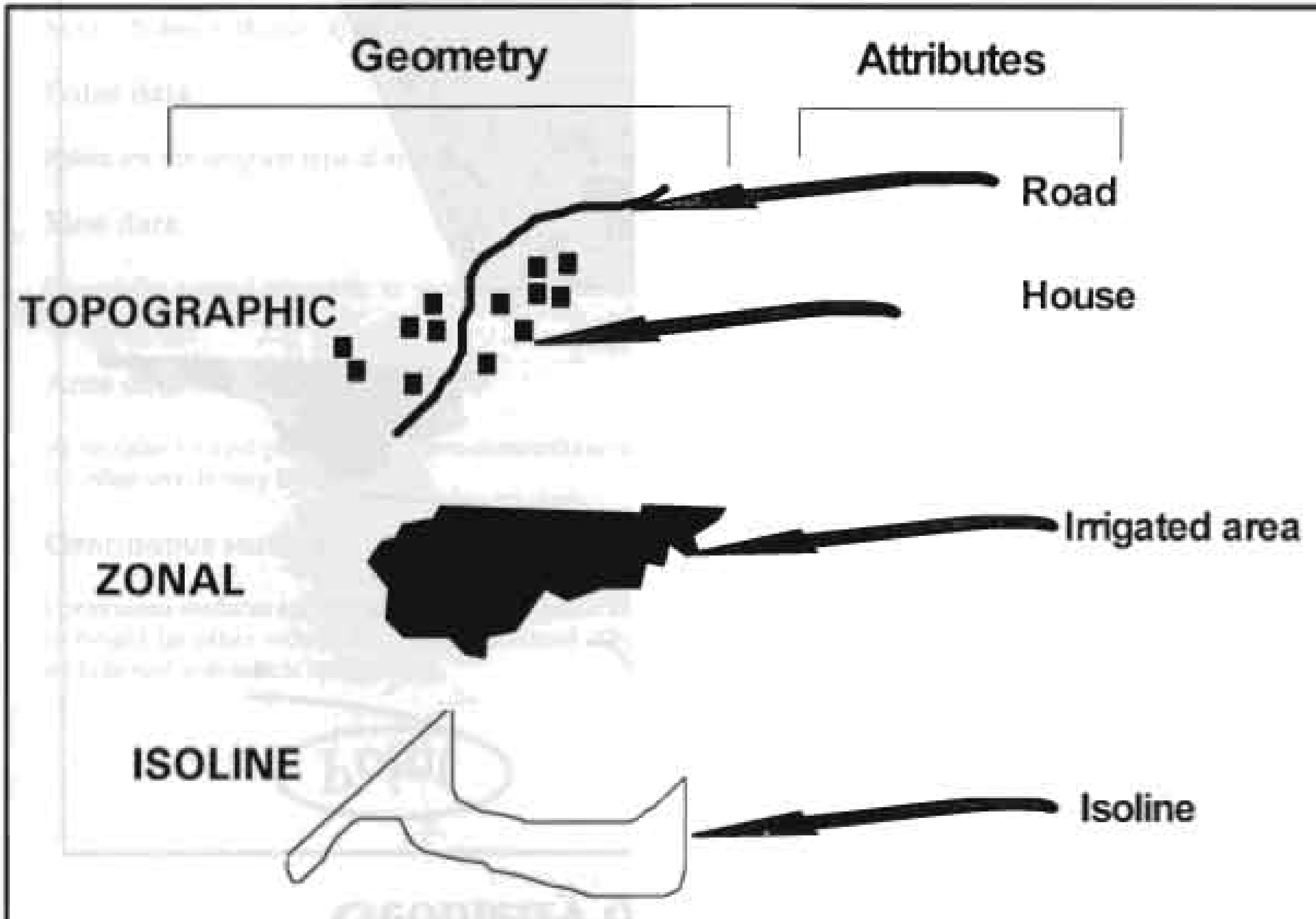
Spatial data

Geographic position refers to the fact that each feature has a location that must be specified in a unique way. To specify the position in an **absolute** way a **coordinate** system is used. For small areas, the simplest coordinate system is the regular square grid. For larger areas, certain approved cartographic projections are commonly used. Internationally, there are many different coordinate systems in use.

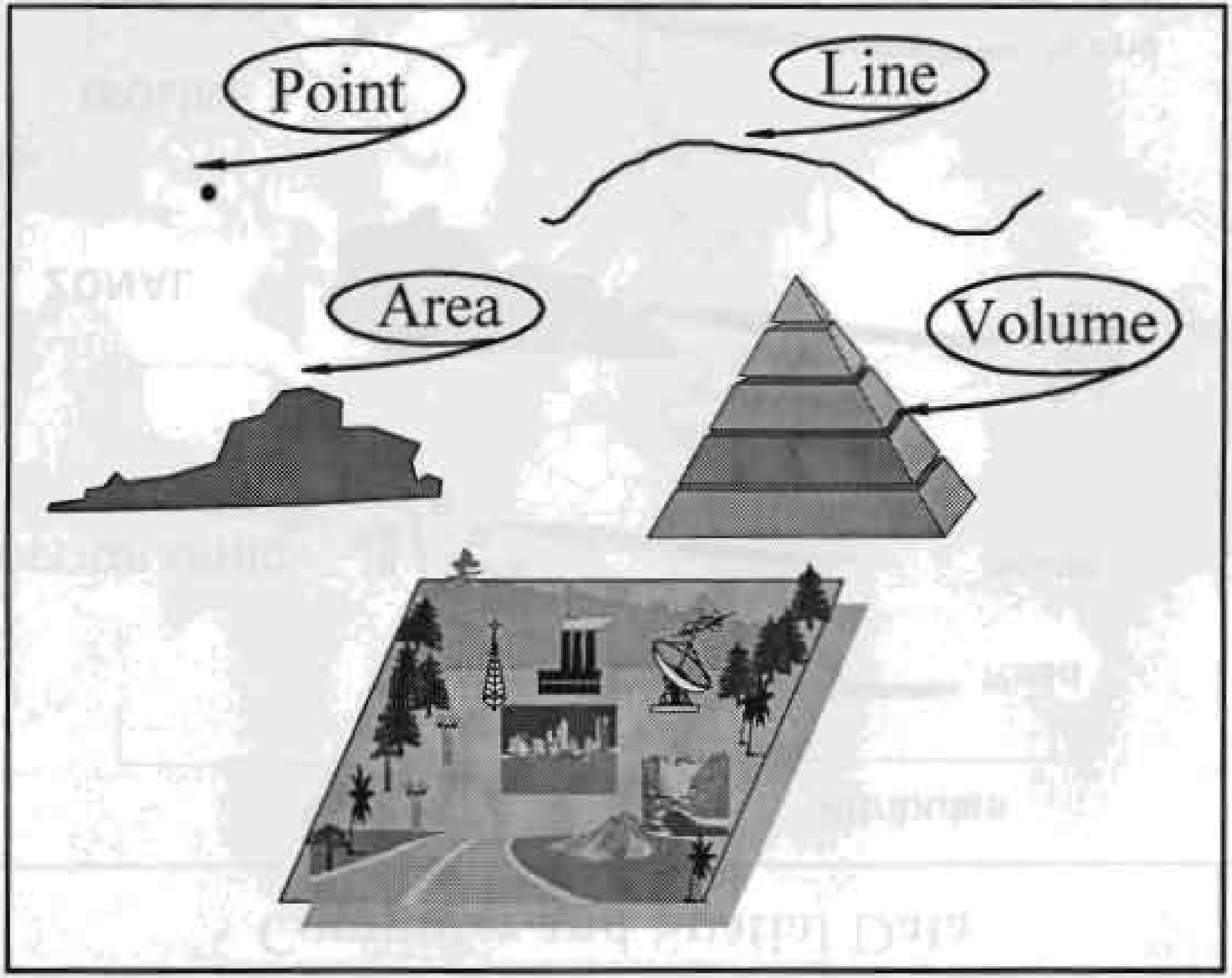
Spatial Data for GIS



Computer and Spatial Data



Geometry of Spatial Data



Basic Types of Spatial data

Most GIS deal with four (4) types of geographic data: points, lines, areas, and continuous surfaces.

Point data

Points are the simplest type of spatial data. They are zero-dimensional objects with only a position in space but no length.

Line data

Lines (also termed segments or arcs) are one-dimensional spatial objects. Besides having a position in space, they also have a length.

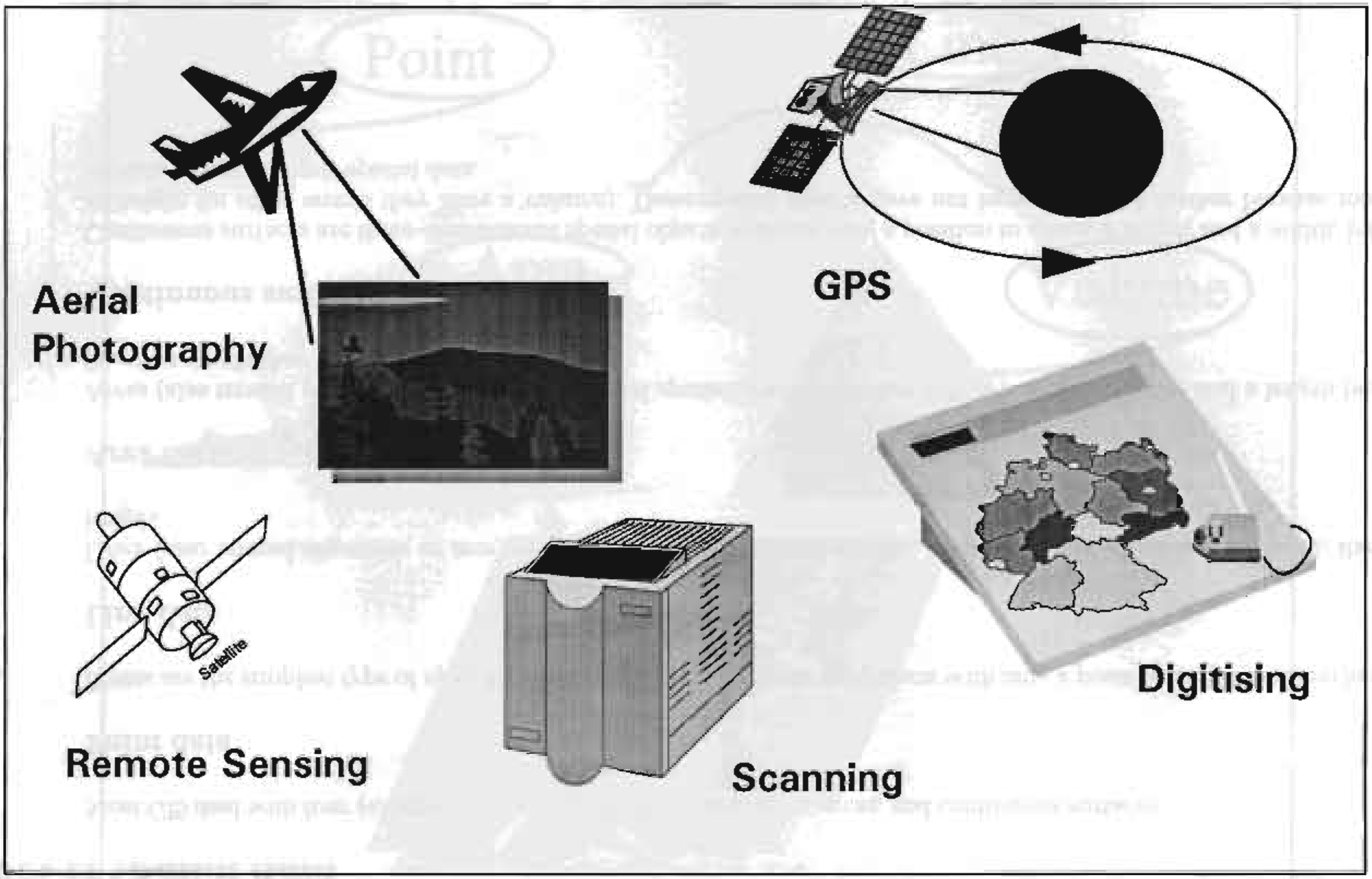
Area data

Areas (also termed polygons) are two-dimensional spatial objects with not only a position in space and a length but also a width (in other words they have an area).

Continuous surface

Continuous surfaces are three-dimensional spatial objects with not only a position in space, a length and a width, but also a depth or height (in other words they have a volume). These spatial objects have not been discussed further because most GIS do not include real volumetric spatial data.

Data Capture



Data Capture

The functionality of GIS, however, relies on the quality of data available, which, in most developing countries, is either redundant or inaccurate. Although GIS are being used widely, effective and effiient means of data collection have yet to be systematically established. The true value of GIS can only be realised if the proper tools to collect spatial data and integrate them with attribute data are available.

The vector data model uses three methods of representing data: points, lines, and polygons, of which the x and y coordinates are stored, together with the relationships between the coordinates, lines, and polygons.

In practically all GIS systems, the conversion from analog maps into digital maps, called digitising, is carried out using the vector data structure. Digitising map data can be carried out in two ways, apart from manually entering the coordinates via the keyboard, which, for obvious reasons, is not very popular: manual digitising and raster to vector conversion.

Attribute Handling

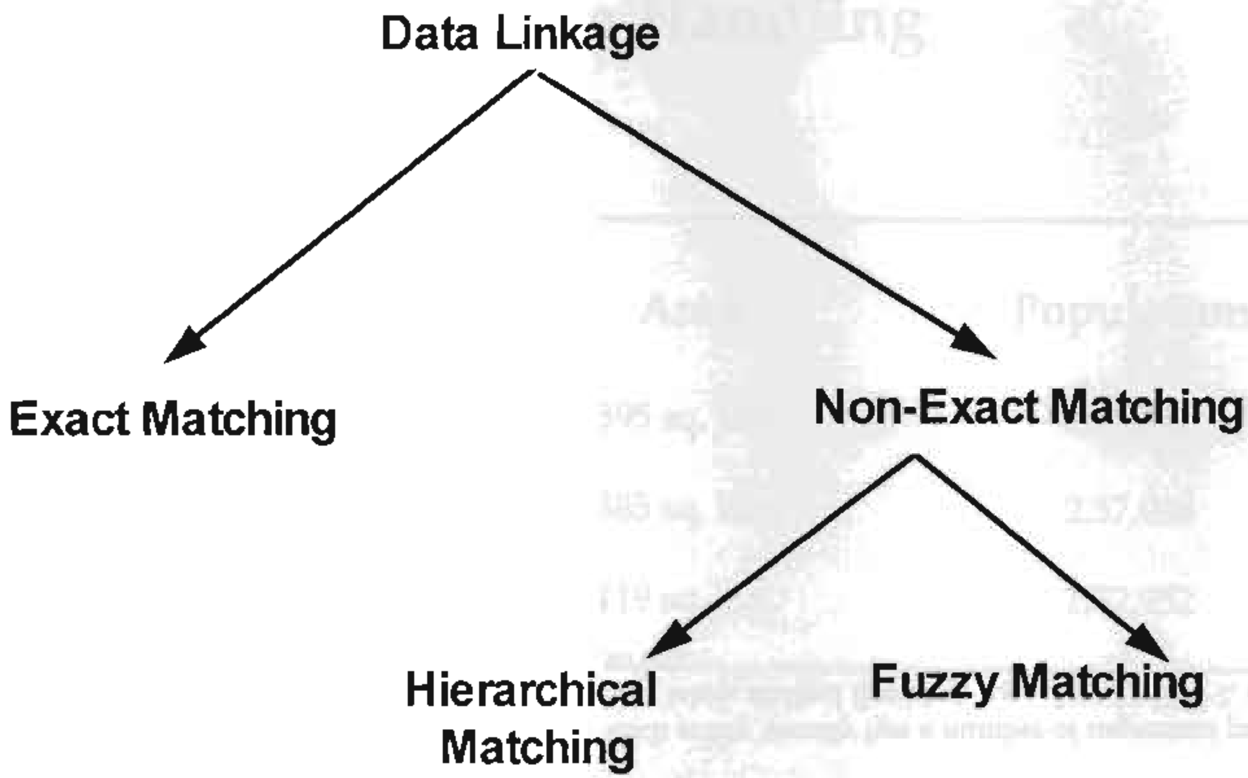
District Name	Area	Population
Kathmandu	395 sq. km.	6,75,341
Lalitpur	385 sq. km.	2,57,086
Bhaktapur	119 sq. km.	1,72,952

Attribute data

The attributes refer to the properties of spatial entities. They are often referred to as non-spatial data since they do not in themselves represent location information.

In addition to the spatial representation of a feature, each entity usually has a number of important properties or attributes. These attributes may be nominal (identity, e.g., maize, granite, lake), ordinal (ranking, e.g., class 1, class 2, class 3, and so on), or scalar (value, e.g., water depth, elevation, erosion rate, and so on).

Data Linkage



Data Linkage

A GIS typically links different sets. Suppose you want to know the mortality rate due to cancer among children under 10 years of age in each country. If you have one file that contains the number of children in this age group, and another that contains the mortality rate from cancer, you must first combine or link the two data files. Once this is done, you can divide one figure by the other to obtain the desired answer.

Exact Matching

Name	Population
Aenchok	4038
Baireni	7030
Naubesi	10777
Salyantar	5798
Gajuri	5606

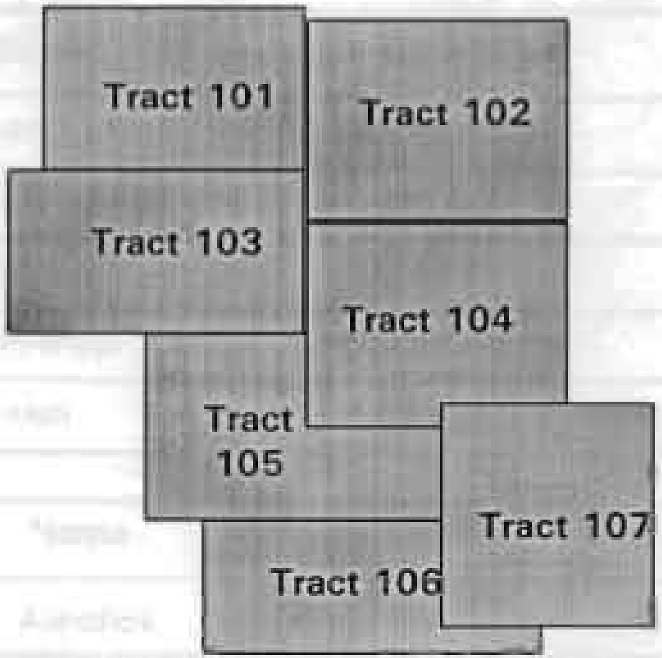
Name	Avg. Housing Cost
Aenchok	30,500
Baireni	22,000
Naubesi	100,000
Salyantar	24,000
Gajuri	24,000

Name	Population	Avg. Housing Cost
Aenchok	4038	30,500
Baireni	7030	22,000
Naubesi	10777	100,000
Salyantar	5798	24,000
Gajuri	5606	24,000

Exact Matching

Exact matching occurs when you have information in one computer file about many geographic features (e.g., towns) and additional information in another file about the same set of features. The operation to bring them together is easily achieved by using a key common to both files – in this case, the town name. Thus, the record in each file with the same town name is extracted, and the two are joined and stored in another file.

Hierarchical Matching



Kathmandu District

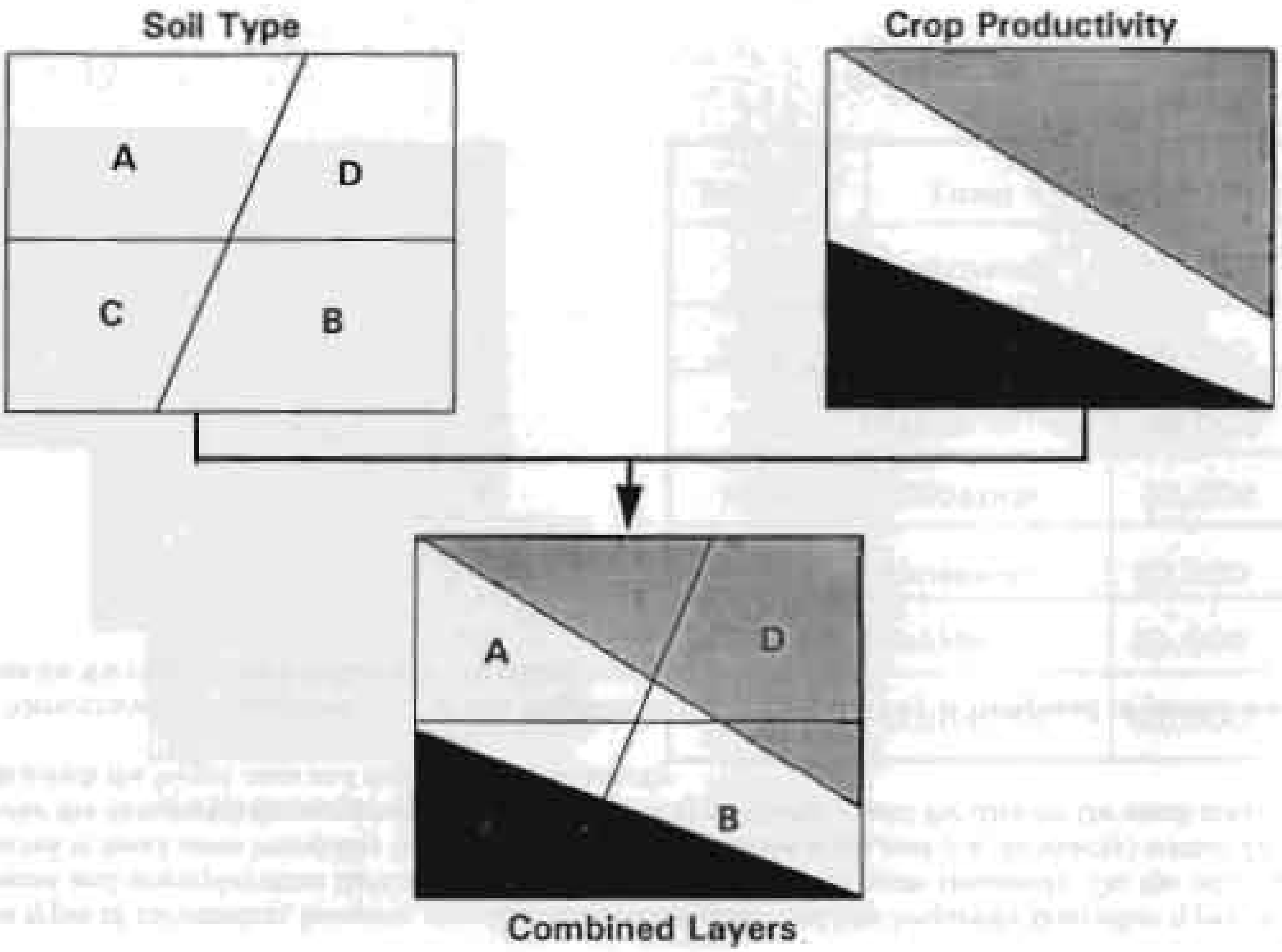
Tract	Town	Population
101	Kathmandu	60,000
102	Maharangung	45,000
103	Patan	35,000
104	Dillibazaar	36,000
105	Baneswor	57,000
106	Nakkhu	25,000
107	Kupondole	58,000

Hierarchical Matching

Some types of information, however, are collected in more detail and less frequently than other types of information. For example, financial and unemployment data covering a large area are collected quite frequently. On the other hand, population data are collected in small areas but at less frequent intervals. If the smaller areas nest (i.e., fit exactly) within the larger ones, then the way to make the data match the same area is to use hierarchical matching — add the data for the small areas together until the grouped areas match the bigger ones and then match them exactly.

The hierarchical structure illustrated in this diagramme shows that this city is composed of several tracts. To obtain meaningful values for the city, the tract values must be added together.

Fuzzy Matching



When data boundaries between layer do not match, the layers can be jointed, creating a new layer containing the characteristics of both layers.

Fuzzy Matching

On many occasions, the boundaries of the smaller areas do not match those of the larger ones. This occurs often while dealing with environmental data. For example, crop boundaries, usually defined by field edges, rarely match the boundary between soil types. If you want to determine the most productive soil for a particular crop, you need to overlay the two sets and compute crop productivity for each and every soil type. In principle, this is like laying one map over another and noting the combinations of soil and crop productivity.

A GIS can carry out all of these operations because it uses geography, or space, as a common key between the data sets. Information is linked only if it relates to the same geographical area.

Why is data linkage so important? Consider a situation where you have two data sets for a given area, such as yearly income by county and average cost of housing for the same area. Each data set might be analysed and/or mapped individually. Alternatively, they may be combined. With two data sets, only one valid combination exists. Even if your data set may be meaningful for a single query you will still be able to answer many more questions than if the data sets were kept separate. By bringing them together, you add value to the database. To do this, you need GIS.

Four Major GIS Functions

- data capture
  - graphic data: digitised, converted from existing data
  - attribute data: keyed-in, loaded from existing data files
- data storage and manipulation
  - file management
  - editing
- data analysis
  - database query
  - spatial analysis
  - modelling
- data display
  - maps
  - reports

Major Functions of GIS

Data Capture. Data used in GIS often come from many different sources, are of many types, and are stored in different ways. A GIS provides tools and a method for the integration of different data into a format to be compared and analysed. Data sources are mainly manual digitisation/scanning of aerial photographs, paper maps, and existing digital datasets. Remote-sensing satellite imagery and GPS are promising data input sources for GIS.

Database Management and Update. After data are collected and integrated, the GIS must provide facilities which can contain and maintain data. Effective data management has many definitions but should include all of the following aspects: *data security, data integrity, data storage* and *retrieval*, and *data maintenance abilities*.

Geographic Analysis. Data integration and conversion are only a part of the input phase of GIS. What is required next is the ability to interpret and to analyse the collected information quantitatively and qualitatively. For example, a satellite image can assist an agricultural scientist to project crop yield per hectare for a particular region. For the same region, the scientist also has the rainfall data for the past six months collected through weather station observations. The scientist also has a map of the soils for the region which shows fertility and suitability for agriculture. These point data can be interpolated and what you get is a thematic map showing isohyets or contour lines of rainfall.

Presenting Results. One of the most exciting aspects of GIS technology is the variety of different ways in which the information can be presented once it has been processed by GIS. Traditional methods of tabulating and graphing data can be supplemented by maps and three dimensional images. Visual communication is one of the most fascinating aspects of GIS technology and is available in a diverse range of output options.

Types of GIS

- Vector GIS
- Raster GIS
- Hybrid GIS

Technology is moving towards hybrid GIS

Raster Vector Integration

Representation of Geographic Data

The representation of geographic data is primarily driven by the types of data structure. This is often termed a data model. A data model is a formal system in which a set of precisely defined objects can be manipulated in accordance with a set of precisely predefined rules, without any regard for the 'meaning' or real-world interpretation of those objects or rules. Reality is an informal system, a system of immense complexity and a system with an infinite amount of information. The difficulty in defining a comprehensive and useful data model is to find a formal system whose behaviour mimics the informal behaviour of the real world as closely as possible

A GIS data model, often known as a geo-relational model, is a formal collection of spatial operators that act on a spatial database in order to relate the user to the real world. GIS uses primarily two spatial data models: **vector and raster**.

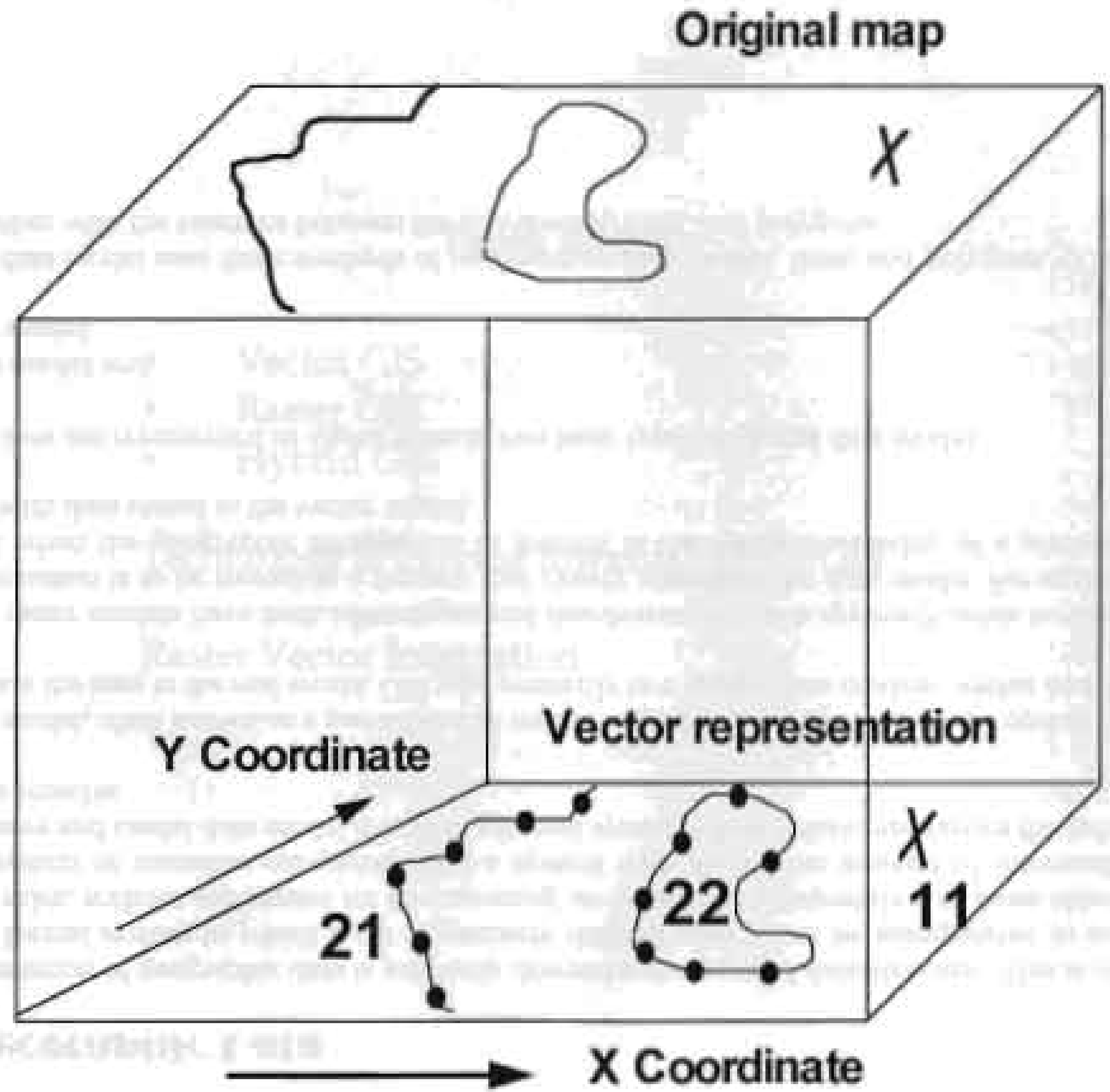
Vector and raster models have both advantages and disadvantages. Each approach tends to work best in situations in which the spatial information is to be treated in a manner that closely resembles the data model. For example, the raster model is generally well suited when the geographic information of interest is the spatial variability of a phenomenon; network analysis is best performed with data stored in the vector model.

The spatial data are represented in digital form in two basic types of spatial data model.

- **a vector model** and
- **a raster model.**

The vector data model uses three methods of representing data: points, lines, and polygons, of which the x and y coordinates are stored, together with the relations between the coordinates, lines, and polygons.

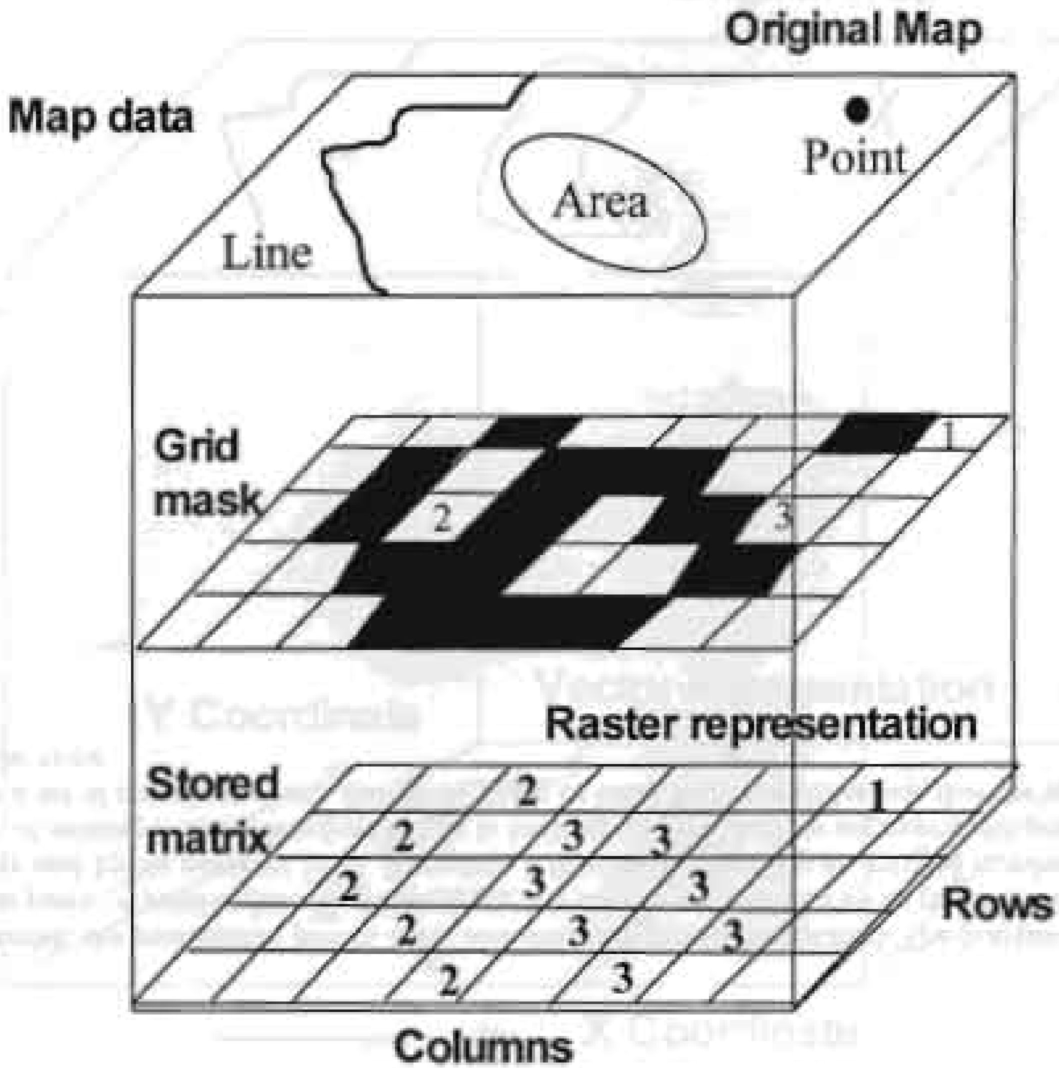
Vector Representation



The vector Model

In a vector model, the positions of points, lines, and areas are precisely specified. The position of each object is defined by a (series of) coordinate pairs. A *point* is described by a single X-Y coordinate pair and by its name or label. A **line** is described by a set of coordinate pairs and by its name or label. In reality, a line is described by an infinite number of points. In practice, this is not a feasible way of storing a line. Therefore, a line is built up of straight line segments. An *area*, also called a *polygon*, as a line is described by a set of coordinate pairs and by its name or label, with the difference that the coordinate pairs at the beginning and the end are the same.

Raster Representation



The raster model

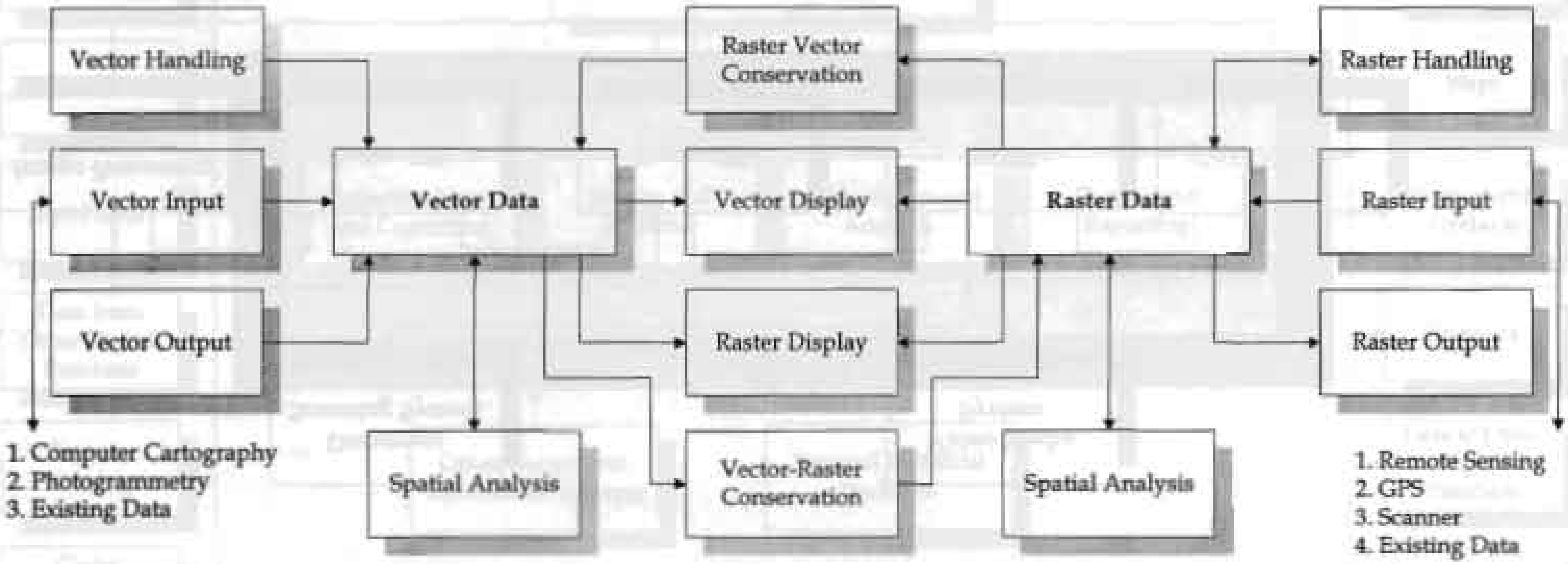
The simplest way of storing spatial data is through the raster model. In this model, the spatial data are organised in cells or pixels (hereafter referred to as pixels). These pixels are the basic units for which information is explicitly recorded. Each pixel is only assigned one value.

A *point* is described by the position of a single pixel. Usually, the position of the cell is defined by a row and column number. A pixel is assigned one numerical value. The name of the point is not explicitly recorded within the pixel itself. A legend has to be available to determine which name belongs to which digital number. A *line* and an *area* are described by a set of connected pixels with one numerical value. In the raster model there is no basic difference in how points, lines, and areas are stored.

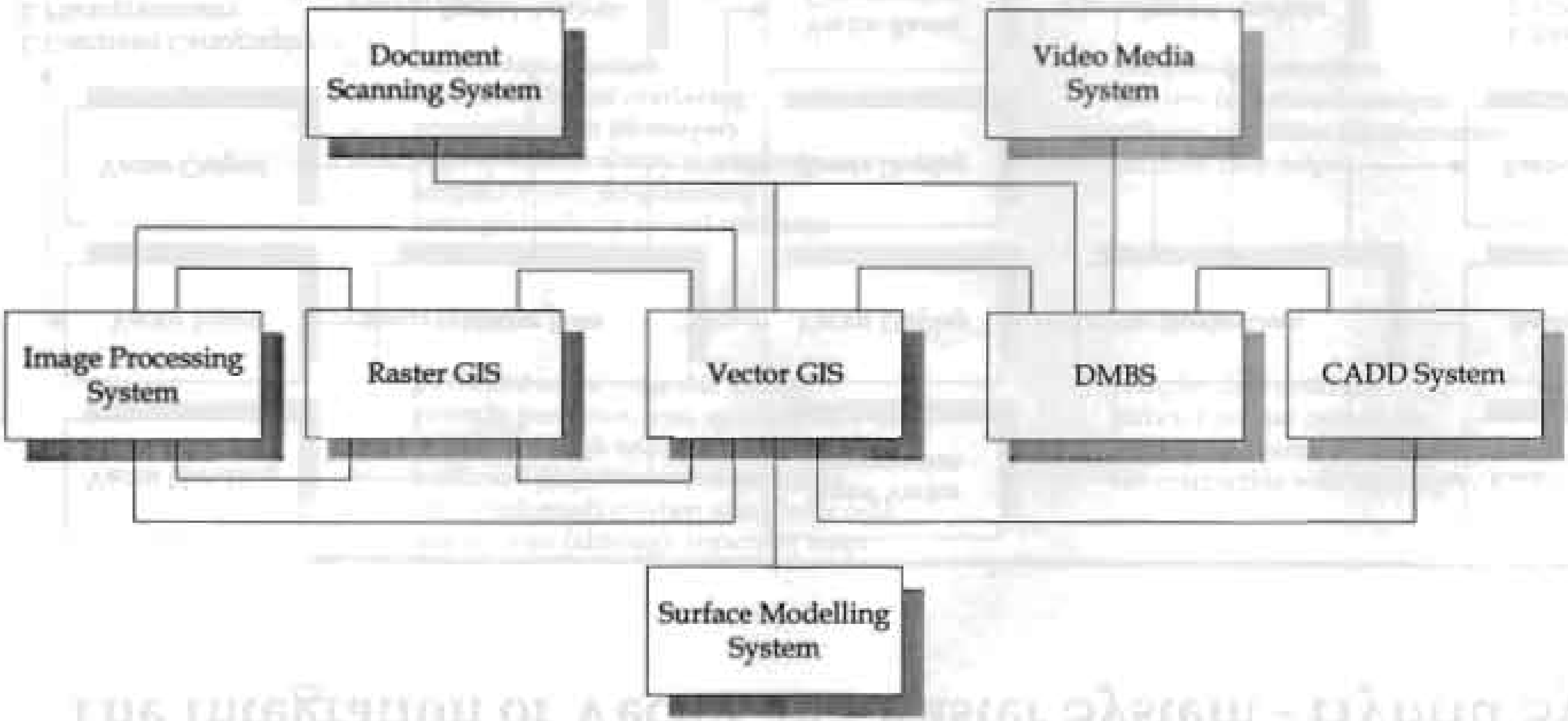
Advantages and Disadvantages of the Raster and Vector Data Models

RASTER MODEL	VECTOR MODEL
Advantages	**Advantages**
- simple data structure	- compact data structure
- easy and efficient overlaying	- efficient for network analysis
- compatible with RS imagery	- efficient projection transformation
- high spatial variability is efficiently represented	- accurate map output
- simple for own programming	
- same grid cells for several attributes	
Disadvantages	**Disadvantages**
- inefficient use of computer storage	- complex data structure
- errors in perimeter, area, and shape	- difficult overlay operations
- difficult network analysis	- high spatial variability is inefficiently represented
- inefficient projection transformations	- not compatible with RS imagery
- loss of information when using large cells	
- less accurate (although attractive) maps	

The Integration of Vector and Raster System - Hybrid System



An Example of Raster and Vector Integration



Principal Components and Functions of an Ideal GIS