Geographical Information Systems: Concepts

INTRODUCTION

Geography is the study of Earth's features and patterns of their variations in spatial location and time. Many questions of agricultural production are geographic in nature as the production depends on the environment and prevailing socio economic conditions, both of which vary spatially and in time.. Examples are questions related to natural resources management, precision agriculture, agroecological classification for land use planning, regional trends and patterns in technology adaptation, agricultural productivity and income, non-pont source pollution from agricultural lands, etc. Answering these questions requires access to large volumes of multidimensional geographical (spatial) information of weather, soils, topography, water resources, socio economic status, etc. Further, answers to even apparently simple questions require that the data from several sources be integrated in a consistent form. Geographical Information Systems or GIS enable representation and integration of such spatial information.

The traditional method of presenting geographical information in two dimensions is in the form of maps. Maps are graphic representations of the earth's surface on a plane paper They shape the way we visualize, assess and analyze spatial information. A map consists of points, lines and area elements that are positioned with reference to a common coordinate system (usually latitude and longitude).. They are drawn to specified *scales* and *projection*. Map scales can vary and depend on the purpose for which the maps are created. Projection is a mathematical transformation used to represent the real 3-dimensional spherical surface of the earth in 2-dimensions on a plane sheet of paper. The map legend links the non-spatial attributes (name, symbols, colours, thematic data) to the spatial data. The map itself serves to store and present data to the user. Such, analogue maps (on paper) are cumbersome to produce and use, particularly when there are a large number of them to be used for analysis. Computer based GIS facilitates both creation of maps and using them for various complex analyses. It allows working with geographic data in a digital format to aid decision making in resources management

GIS is a generic term implying the use of computers to create and display digital maps. The attribute data which describe the various features presented in maps may relate to physical, chemical, biological, environmental, social, economic or other earth surface properties. GIS allows mapping, modelling, querying, analyzing and displaying large quantities of such diverse data, all held together within a single database. Its power and appeal stem from its ability to integrate quantities of information about the environment and the wide repertoire of tools it provides to explore the diverse data. The history of development of GIS parallels the history of developments in digital computers and database management systems on one hand and those in cartography and automation of map production on the other. The development of GIS has also relied upon innovations made in several other disciplines – geography, photogrammetry, remote sensing, civil engineering, statistics, etc.

A GIS produces maps and reads maps. Its major advantage is that it permits identifying spatial relationships between specific different map features. It can create maps in different scales, projections and colours. But it is not just a map making tool. It is primarily an analytical tool that provides new ways of looking at, linking and analyzing data by projecting tabular data into maps and integrating data from different, diverse sources. This it does by allowing creation of a set of maps, each with a different theme (soils, rainfall, temperature, relief, water sources, etc.).

From its early beginnings, GIS has been an integrating technology both from the point of view of its development as well as its use. This is because, once geographic information of any kind is translated into the digital form in a GIS, it becomes easy to copy, edit, analyze, manipulate and transmit it. *This allows vital linkages to be made between apparently unrelated activities based on a common geographic location*. This has led to fundamental changes in the way resource management decisions are made in a variety of situations - forest management, marketing management, utility management, transportation, as well as in agricultural, environmental and regional planning and management. Some potential agricultural applications where GIS can lead to better management decisions are: precision farming, land use planning, watershed management, pest and disease management, irrigation management, resources inventory and mapping, crop area assessment and yield forecasting, biodiversity assessment, genetic resources management, etc.

DEFINITION OF GIS

A GIS is basically a computerized information system like any other database, but with an important difference: *all information in GIS must be linked to a geographic (spatial) reference* (latitude/longitude, or other spatial coordinates).

GIS = G + IS

Geographic reference + Information system

Data of spatial coordinates on the surface of the earth (Map) – location data

Database of attribute data corresponding to spatial location and procedures to provide information for decision making

GIS = IS with geographically referenced data

There are many different definitions of GIS, as different users stress different aspects of its use. For example:

(i) ESRI defined GIS as an organized collection of computer hardware, software, geographic data and personnel designed to efficiently capture, store, update, manipulate, analyze and display geographically referenced information.

(ii) ESRI also provided a simpler definition of GIS as a computer system capable of holding and using data describing places on the earth's surface).

(iii) Duecker defined GIS as a special case of information systems where the database consists of observations on spatially distributed features, activities or events, which are definable in space as points, lines or areas. A GIS manipulates data about these points, lines or areas to retrieve data for ad hoc queries and analyses.

-

The United States Geological Survey (USGS) defined provided *A GIS as a computer hardware and software system designed to collect, manage, analyze and display geographically (spatially) referenced data*. This definition is a fairly comprehensive and is suitable for agricultural applications of GIS

> Note that a GIS does not store a map or image. What it stores is a *relational database* from which maps can be created as and when needed. Relational database concepts are particularly crucial to the development of GIS. Each map (say a soil map) can be considered to hold a layer or a level of information. A GIS works with several layers of such thematic data. It can answer questions by comparing the different layers and also by overlaying them if all the layers are referenced to the same locations, that is *location is the common key* for all the thematic data sets. This ensures that every location (spatial reference point) is exactly matched to its location on other maps. Once this is done, the different layers can be compared and analyzed singly and in combination to identify spatial patterns and processes. Thus, GIS are related to other database applications, but with an important difference; all information in a GIS must be linked to a spatial reference. Other data bases may contain locational information (addresses, pin codes etc), but the GIS uses *georeferences* as the primary means of storing and accessing information.

WHAT A GIS CAN DO

There are five basic questions which a complete GIS must answer. These are:

What exists at a particular location? Given a geographic reference (eg lat,long) for a location, the GIS must describe the features of that location

Where can specific features be found? This is the converse of the first question. For example, where are the districts with rainfall greater than 500 mm and less than less than 750 mm?

Trends or What has changed over time? This involves answering both questions above. For example, at what locations are the crop yields showing declining trends?

What spatial patterns exist? if occurrence of a pest is associated with a hypothesized set of conditions of temperature, precipitation, humidity, where do those conditions exist?

Modelling or What if ...? This is a higher level application of GIS and answers questions like what would be the nitrate distribution in groundwater over the area if fertilizer use is doubled?

The abilities to separate great quantities of information about the environment into layers, explore each layer with a powerful suite of analytical tools, and then combine the layered information to use it in an integrated fashion is what makes the GIS a powerful and effective decision-support tool for agricultural and environmental management.

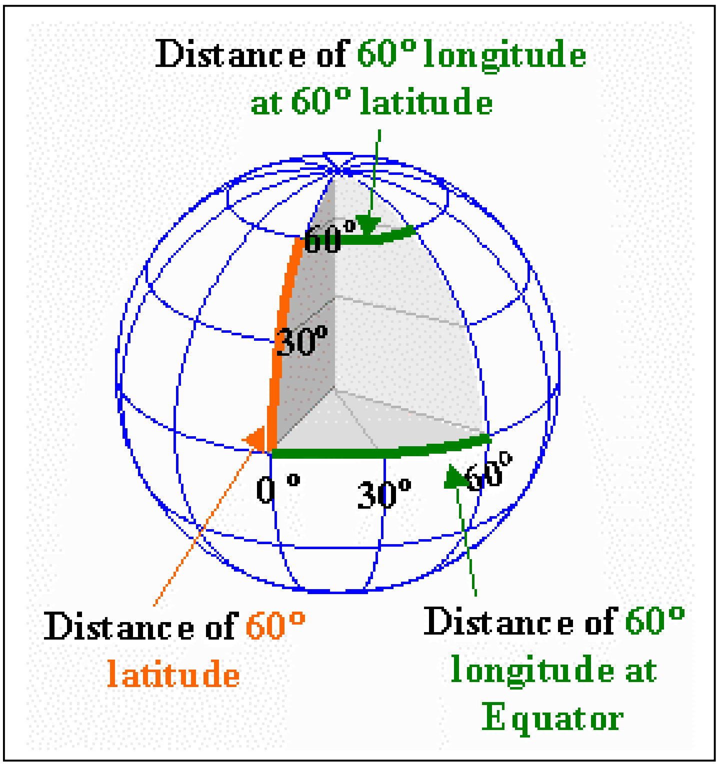
GEOGRAPHIC REFERENCING CONCEPTS

A GIS is to be created from available maps of different thematic layers (soils, land use, temperature, etc). The maps are in two-dimensions whereas the earth's surface is a 3-dimensional ellipsoid. Every map has a projection and scale.

To understand how maps are created by projecting the 3-d earth's surface into a 2-d plane of an analogue map, we need to understand the georeferencing concepts. Georefencing involves 2 stages: specifying the 3-dimensional coordinate system that is used for locating points on the earth's surface that is, the Geographic Coordinate System (GCS) and the Projected Coordinate System that is used for projecting into two dimensions for creating analogue maps.

Geographic Coordinate System

The traditional way of representing locations on the surface of the earth is in the 3-dimensional coordinate system is by its latitude and longitude.



Source: ESRI

Note that the distance between two points on the 3-d earth's surface varies with latitude. The 3-d system therefore does not provide a consistent measure of distances and areas at all latitudes.

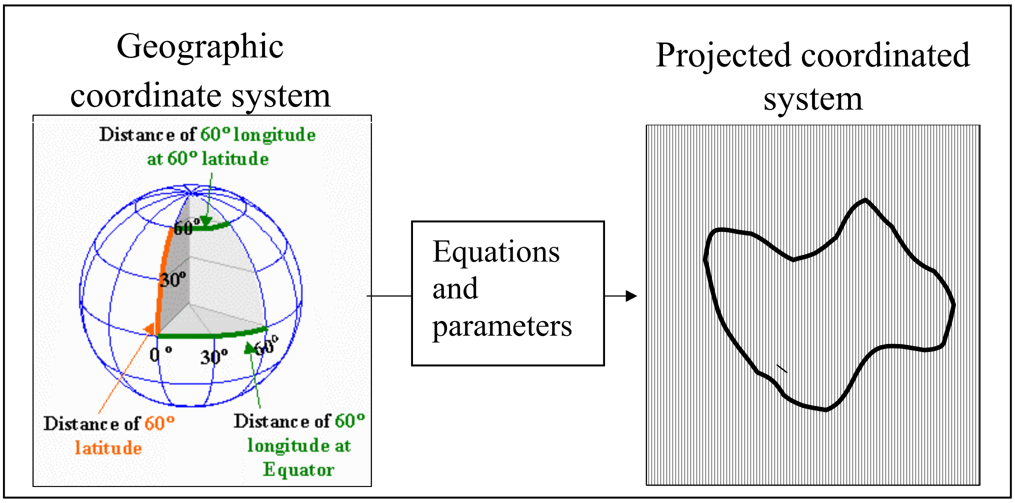
The true surface of the Earth is not the smooth ellipsoid shown in the figure but is quiet uneven and rugged. The GCS which is the surface used for specifying the latitude and longitude of a point on the earth's surface is also an approximation and a 3-d model of the earth. Several standard models of the ellipsoid are available to define the GCS (WGS 84, Everest ellipsoid) etc. The different models vary in their critical parameters (semi major or equatorial axis and semi minor or polar axis of the ellipsoid and the point of origin). The ellipsoid model that is used to calculate latitude and longitude is called the datum. Changing the datum, therefore, changes the values of the latitude and longitude.

Specifying the Geographic Coordinate System therefore requires specifying the Datum. The datum is a fixed 3-d ellipsoid that is approximately the size and shape of the surface of the earth, based on which the geographic coordinates (latitude and longitude) of a point on the Earth's surface are calculated. In fact describing a place by its lat/long is not complete without specifying its datum. In India the Everest Ellipsoid is used as the Datum for the Survey of India maps.

The ideal solution to would be a spheroidal model of the Earth that has both the correct equatorial and polar radii, and is centered at the actual center of the Earth. One would then have a spheroid, that when used as a datum, would accurately map the entire Earth. All lat/longs on all maps would agree. That spheroid, derived from satellite measurements of the Earth, is GRS80, and the WGS84. datum matches this spheroid.

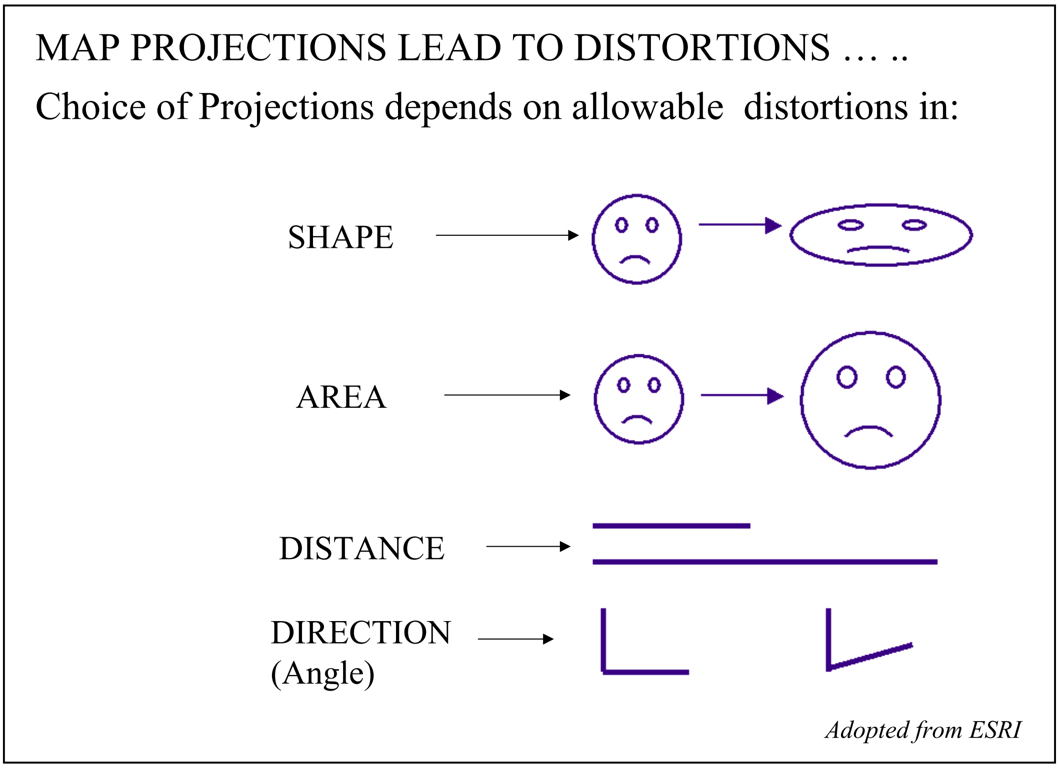
Projected Coordinate System

The development of GIS starts with an available map on paper (an analogue map). This map therefore represents a projection of a 3-d GCS in 2-dimensional form.

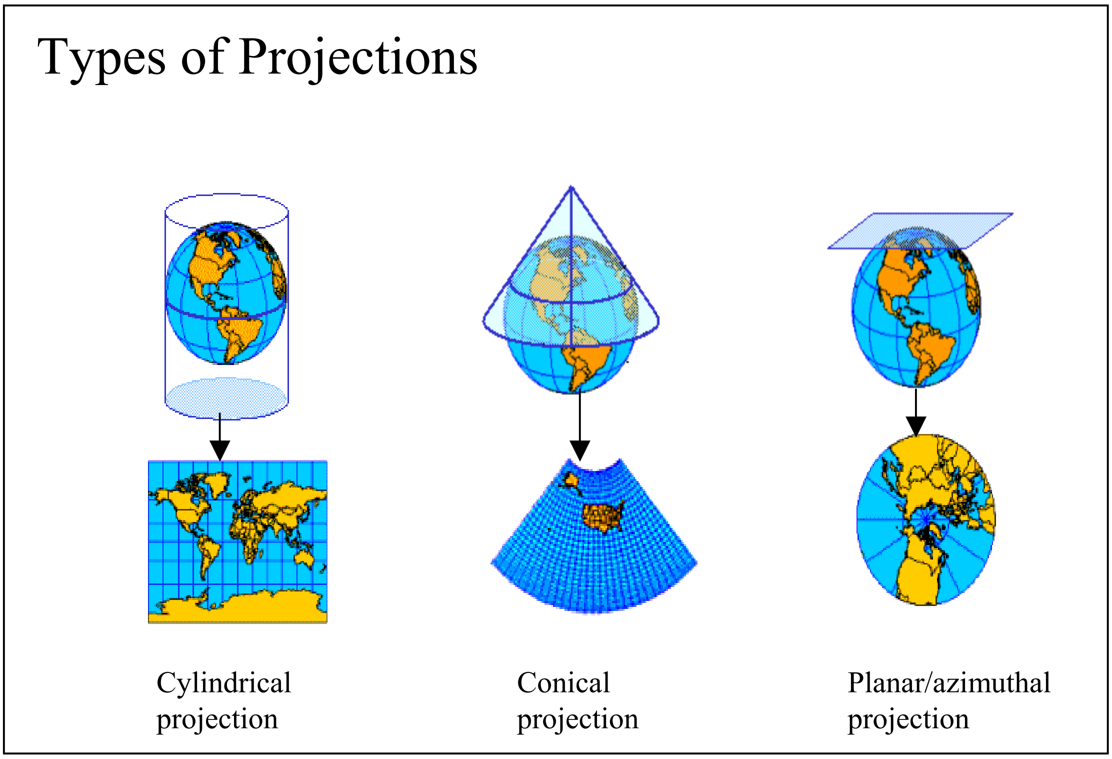


Adopted from ESRI

Projection is a mathematical transformation used to project the real 3-dimensional spherical surface of the earth in 2-dimensions on a plane sheet of paper. The projection causes distortions in one or more spatial properties (area, shape, distance, or direction).



There are many methods of map projections, since there are an infinite number of ways to project the 3-dimensional earth's surface on to a 2-dimensional planar surface. The 3-d to 2-d projections can be done to a plane or to the surface of a cone or cylinder leading to azimuthal, , conic or cylindrical projections respectively with many variations.

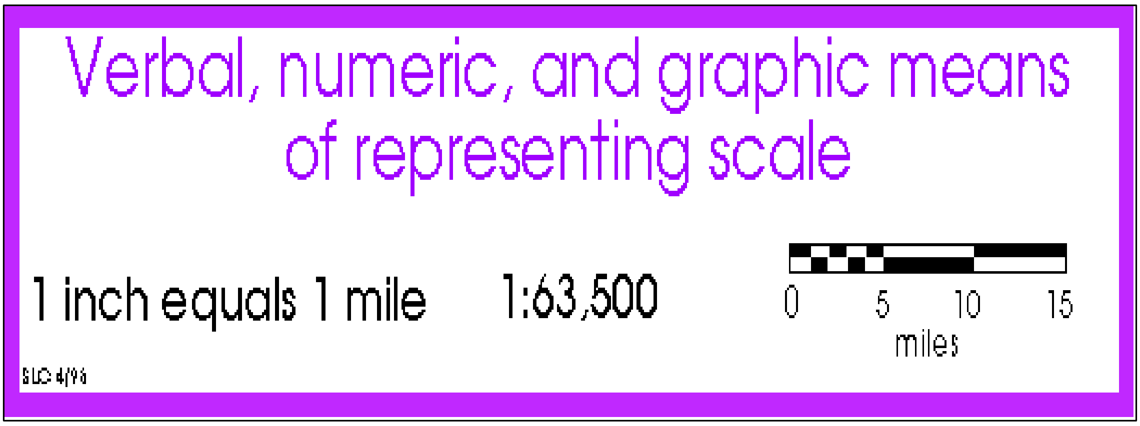


Adopted from ESRI

Depending on the scale and the agreeable tradeoffs with respect to distortions, a specific projection form is chosen. Different countries have adopted different standard projections at different map scales. In India, the polyconic projection is commonly used by Survey of India (SOI). All SOI toposheets are in the polyconic projection.

Map Scale

Map Scale is the ratio of distances on map to distances to on the surface of the earth. It is specified in verbal, numeric or graphical form on all standard maps.



A graphical scale should be present on all maps that are used in GIS as it ensures that any changes in scale in photocopying, etc. are accounted for.

The standard map scales are:

1:1000,000	Country level or State level
1: 250, 000	State or District level
1: 50,000	District level
1: 12,500	Micro level

Survey of India maps are available at all the above levels except the micro-level.

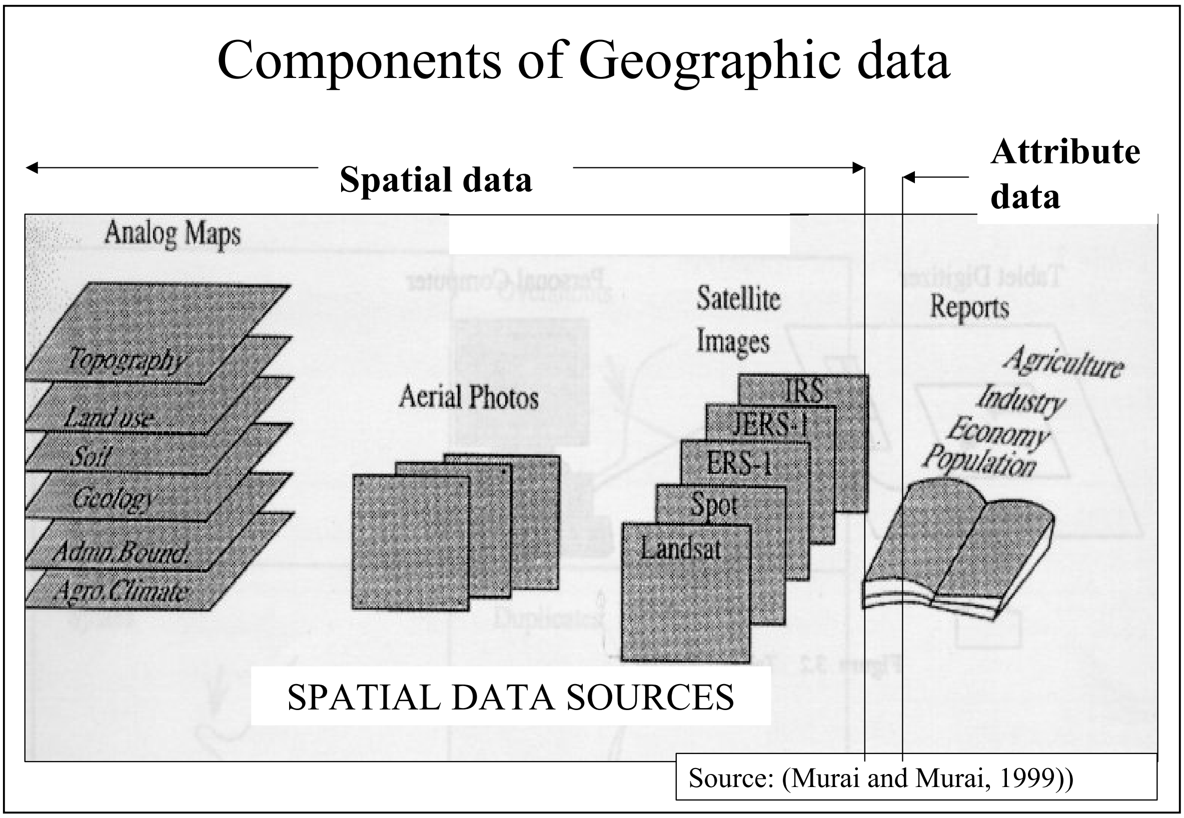
CREATING A GIS

Like for any other Information System , creating a GIS involves 4 stages:

(i) Data input
(ii) Data Storage
(iii) Data Analysis and modelling, and
(iv) Data Output and presentation

The distinction from other Information Systems is that for a GIS the data inputs are of two types:

(i) Spatial data (latitude/longitude for georeferencing,the features on a map, eg soil units, administrative districts), and
(ii) Attribute data (descriptive data about the features, eg soil properties, population of districts, etc.)



Source: (Murai and Murai, 1999))

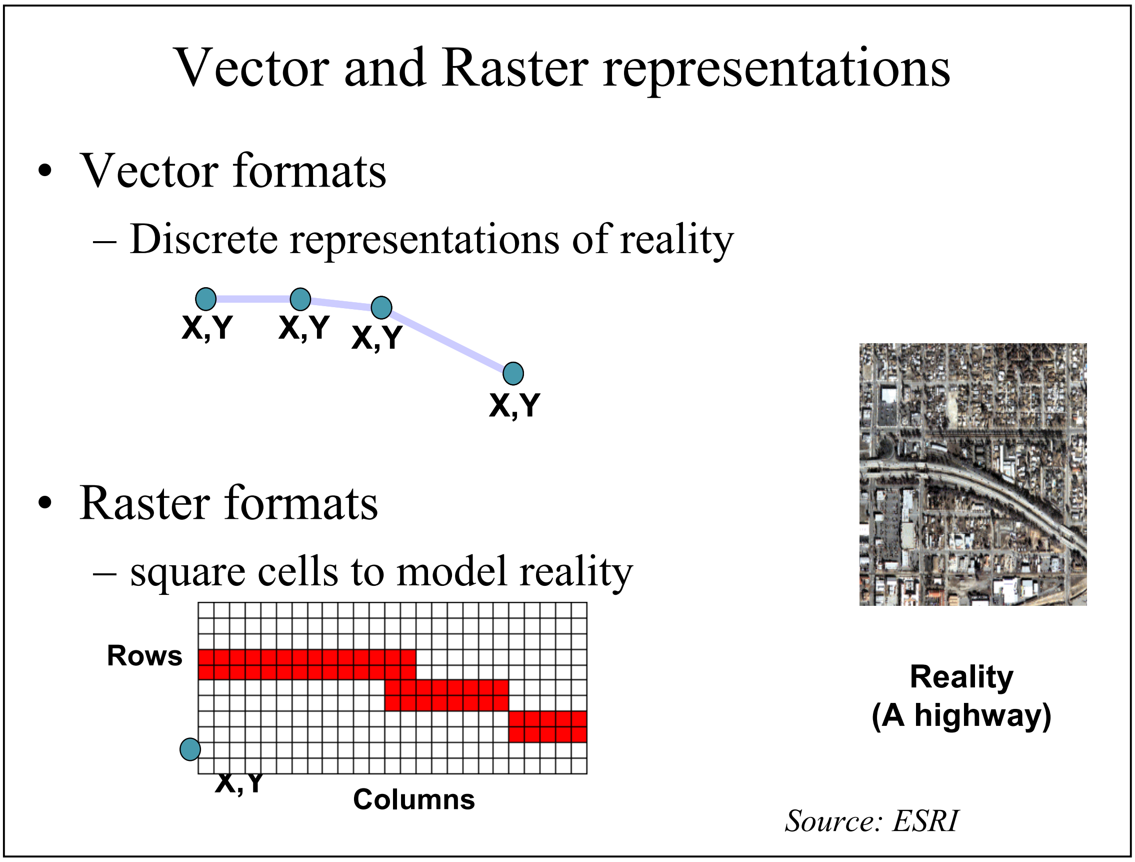
Spatial data sources for creating a GIS are analogue maps (soil map, land use map, administrative districts, map, agroecological zone map, etc.) or aerial photographs and satellite imageries. Data input is the process of encoding analogue data in the form of maps, imageries or photographs into computer readable digitized form and writing data into the GIS database.

GIS Data Input

Spatial Data capture (representing locations in a database) can be in two basic formats:

(i) Vector format
(ii) Raster format

In the Vector format reality is represented as points lines and areas and in the raster format reality is represented as grid of cells/pixels. The Vector format is based on dicrete objects view of reality (analogue maps) and the raster format is based on continuous fields view of reality (photographs, imageries, etc. In principle, any real world situation can be represented in digital form in both raster and vector formats. The choice is up to the user. Each format has its advantages and disadvantages.



Vector data capture

This is generally used for capturing data fro analogue maps. It is based on th observation that any map consists of 3 basic kinds of features –

(i) point features,
(ii) line features and
(iii) polygon or area features.

Points do not have length, width or area. They are described completely by their coordinates and are used to represent discrete locational information on the map to identify locations of features such as, cities, towns, well locations, rain gauge stations, soil sampling points, etc.

A line consists of a set of ordered points. It has length, but no width or area. Therefore it is used to represent features such as roads, streams or canals which have too narrow a width to be displayed on the map at its specified scale.

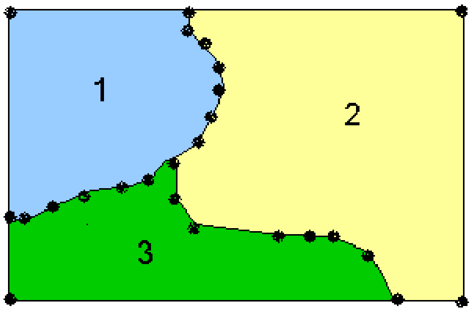
A polygon or area is formed when a set of ordered lines form closed figure whose boundary is represented by the lines. Polygons are used to represent area features such as land parcels, lakes, districts, agroecological zones, etc. A polygon usually encloses an area that may be considered homogeneous with respect to some attribute. For example, in a soil map, each polygon will represent an area with a homogeneous soil type.

A vector based system displays graphical data as points, lines or curves, or areas with attributes. Cartesian coordinates (x, y) or geographical coordinates (latitude, longitude) define points in a vector system.

Data is captured from a map in the form of known x-y coordinates or latitude-longitude by first discretizing the features on the map into a series of nodes (dots) and digitizing the points one by one directly after placing the map on a digitizer. The digitizer can be considered to be an electronic graph paper with a very fine grid. The map is placed on the digitizer and the lines and areas are discretized into a series of points. The digitizer's cursor is used to systematically trace over the points. The points on the map are captured directly as point coordinates. Line features are captured as a series of ordered points. Area features are also captured as an ordered list of points. By making the beginning and end points/nodes of the digitization the same for the area, the shape or area is closed and defined. The process of digitizing from a digitizer is both time consuming and painstaking. Alternately, the map can be scanned and the scanned image digitized on-screen with appropriate software tools. The latter process is relatively simpler, more accurate and is often preferred.

Digitization is usually done feature by feature. For example, all point features are on a map (say cities, towns, etc.) are digitized in one layer. Similarly all line features (eg. Roads, rivers, drainage network, canal network, etc.) are digitized as a separate layer. So are the polygon features (soils, districts, agroecological zones, etc.) For the points feature, the digitization process builds up a database of the points' identification number (ID) and their coordinates. For the lines it builds up a database of their ID, the starting and end nodes for

Spatial data Generation in Vector Format



• discretize lines into points (nodes) and digitize as straight-line segments called vectors or arcs.

• data of X,Y coordinates of points and vectors and their connections (topology) are generated and stored in a database

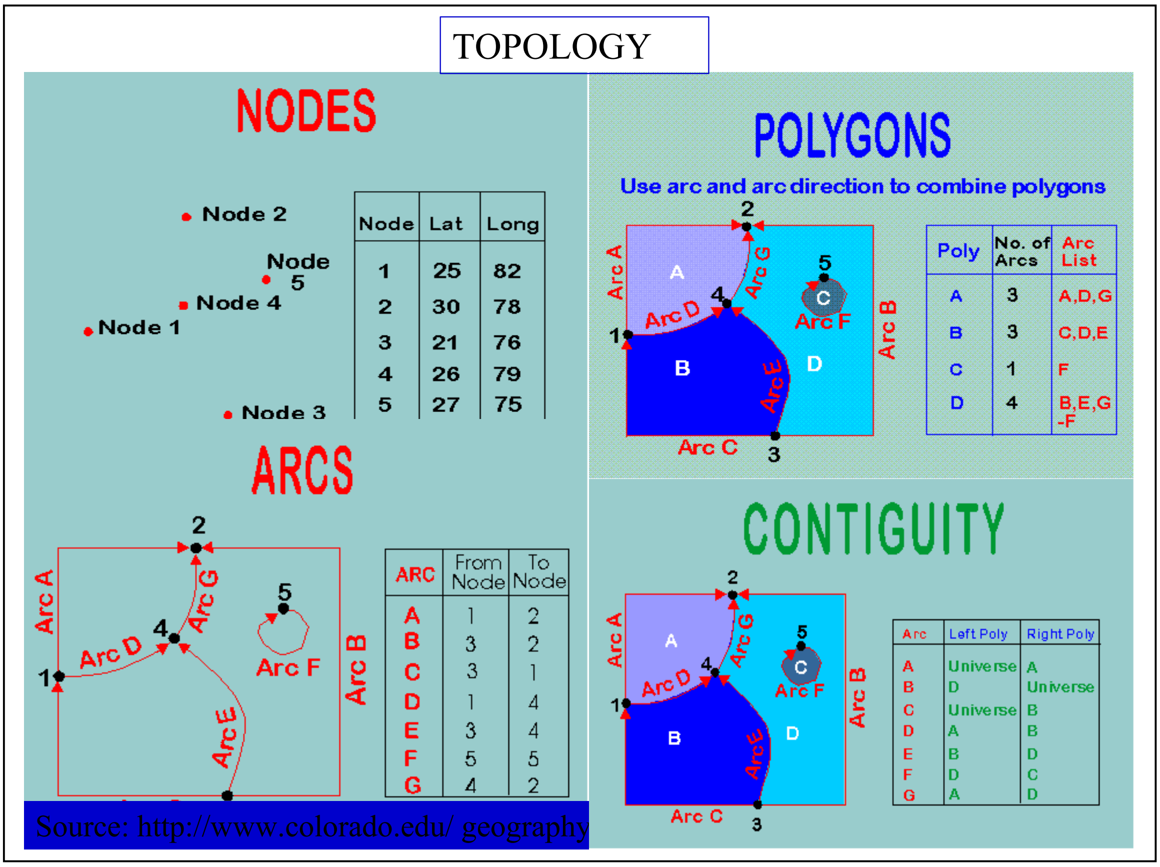
• for areas, geometry (area, perimeter) data are generated

• points, lines and areas have independent database tables

• Add attribute data to database

Adopted from FAO

the line and its length. In addition the GIS also creates a database of the topology, that is, the spatial relationships between the lines. For the polygons also it develops the database of their ID, lines or arcs which comprise it, its toplogy and its area and perimeter.



The identification number (ID) is the key field in each data base (points, lines, areas databases) as it can be used to relate the spatial data with the attribute data. The data resolution depends on the discretization of the digitized points on the initial map. Vector systems are capable of very high resolution (<.001 in) and graphical output is similar to hand drawn maps. But it is less compatible with remote sensing data for which the raster system is preferred.

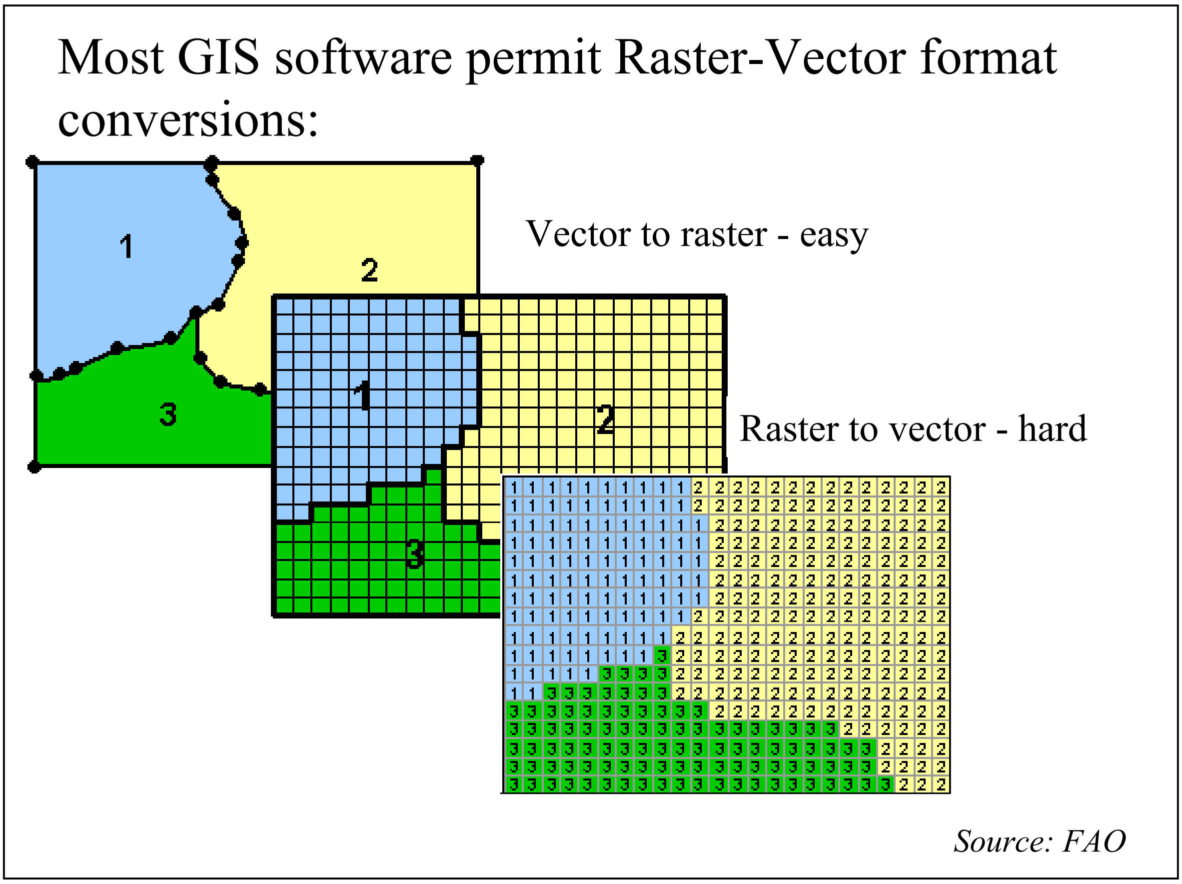
<u>Map projections and Scale</u>:Note that all standard maps which are to be digitized are drawn to specific *projection* and *scale*. But, the digitizer which facilitated the computerized map has its own scale and units and the digitized maps are in these units and scale. Translating information from the digitized map into the real world information of locations, lengths and areas requires information about the mathematical equations used for the projection as well as the scale in which the original analogue map is prepared. In case several map layers are to be digitized (topography, soils, districts, etc), it is necessary to ensure that they are all assembled in the same projection and scale before any spatial analysis is done using them. Most standard GIS have the facility to convert from one map projection to the other and to transform scales from the digitizer scale to map scale to ensure that all map layers have the same locational reference.

Raster data capture

A raster based GIS locates and stores map data by using a matrix of grid cells or pixels. Each cell or pixel is represented either at its corner or centroid by a unique reference coordinate (cell address). Each cell also has discrete attribute data assigned to it.

The raster data resolution is dependent on the pixel or grid cell size. Data can be conveniently captured from remote sensing imageries, areal photographs, and other such imageries of the earth's surface in a raster data format. In this format, the various features are identified by superposing the imageries over a fine rectangular grid of the earth's surface which they represent. Raster data capture does not build topography, that is derive spatial relationships between the identified features. But it facilitates simple scalar operations on the spatial data which a vector format does not permit. Raster data requires to be converted to vector format before topology can be built and spatial operations can be carried out. The raster format also requires more storage space on the computer than the vector format.

Most standard GIS software have the facility to transform maps from raster formats and vice versa.



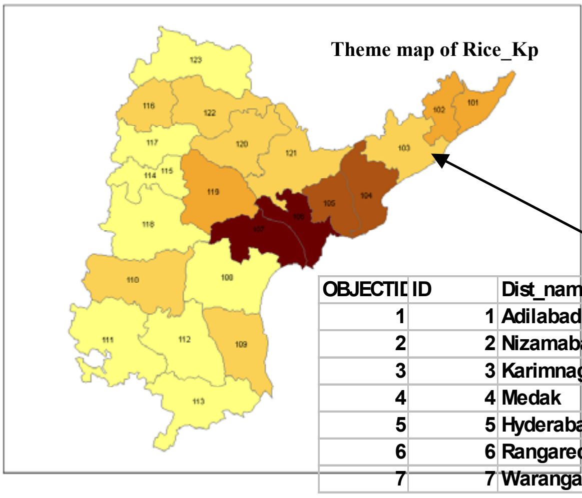
Attribute data

Attribute data are descriptive data of point, line and area features. For points, this may be the name of the location, its elevation, etc. For lines attribute data could be the name of a road, or canal and other descriptions associated with them. For polygons, the attribute data may relate to name of a district and its population, area, area under specific crops in the district, etc.

Attribute data about points/lines/areas features can be entered into different database files. The files can be linked to the default spatial database generated after digitizing by creating an identification key in each data file which is also common to the spatial database generated by the GIS after digitization.

Maps representing several layers of spatial and thematic or attribute information (soil map, rainfall map, agroecozone map, district map, States map, etc.) can be digitized in this fashion independently.

LINKING SPATIAL AND ATTRIBUTE DATA



Attribute table of Dist

OBJECTII	ID	Dist_name	Dist_ID	Geometry	Shape_Len	Shape_Area
1	1	Adilabad	123	1Bo	753337	16838673859.80
2	2	Nizamabad	116	1BoG(	430537	7655003826.81
3	3	Karimnagar	122	1BoG	662231	11771979304.40
4	4	Medak	117	1BoG(	573986	9510748717.53
5	5	Hyderabad	115	1BoG\	53851	203218046.01
6	6	Rangareddy	114	1BoG(	587854	7813591915.82
7	7	Warangal	120	1BoG	674389	12701135906.20

Rice_1995 table

OBJECTII	DISTID	DISTNAME	NO_	YEAR_1995	RICEKA	RICEKP	RICERA	RICERP
23	123	Adilabad	113	1995	63	64	7	19
16	116	Nizamabad	78	1995	102	206	34	84
22	122	Karimnagar	108	1995	104	260	88	252
17	117	Medak	83	1995	64	111	41	78
15	115	Hyderabad	73	1995	1	1	1	1
14	114	Rangareddi	68	1995	22	47	18	35

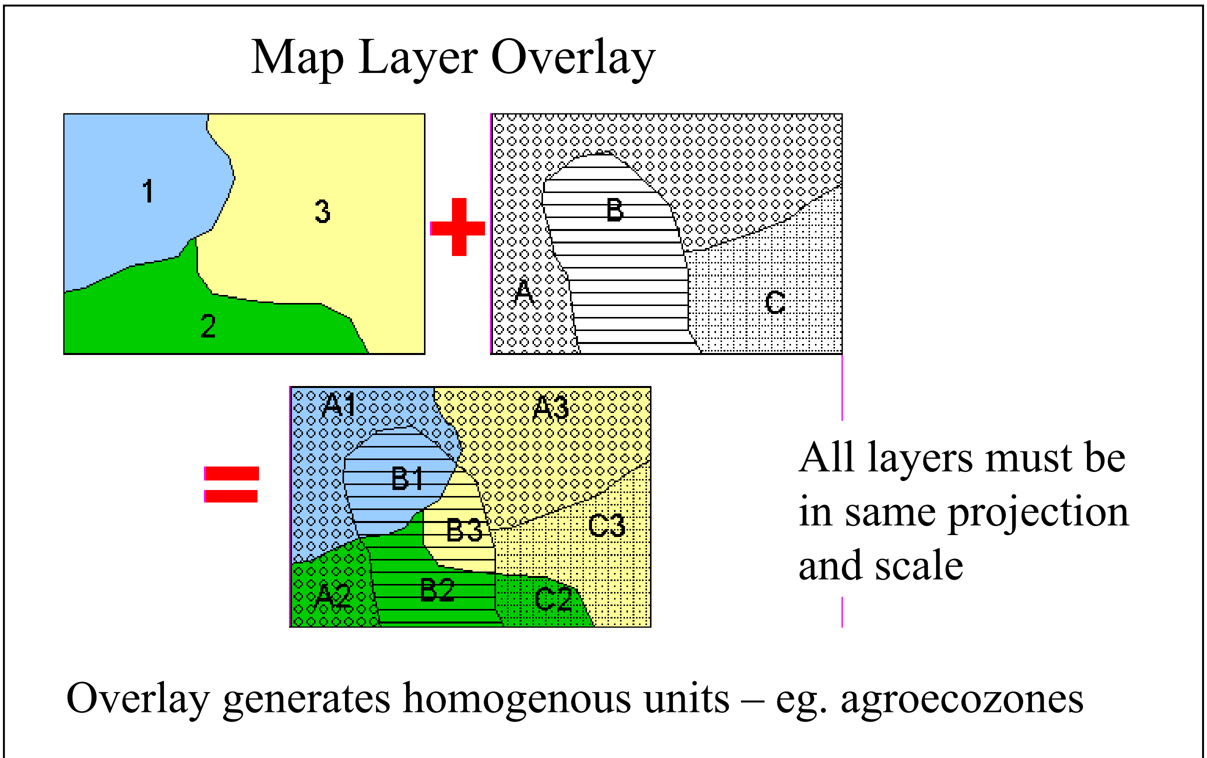
Data Storage and Retrieval

A GIS does not store maps. It stores data organized into a database. The locational data of different features (coordinates, topology) are generated during the digitization process. The attribute data of locations are created separately. the GIS must provide the link between the locational and attribute data. The relational database model is most suitable to ensure such linkage and the database query language can be used to retrieve data. Relational database concepts are therefore central to organizing and managing data in GIS. The specific format of data storage varies with the GIS software. For example, Geomedia GIS stores the spatial and attribute data in a Microsoft Access databases. The feature attribute database created during digitization is created in a specific folder called the Warehouse. The map connections are stored in file created in the Geoworksapces folder Retrieval of data is possible by employing the appropriate query language for the database model. Other attribute databases can be storesd as MS Access files anywhere in the system and connections to them can be established if the share a common ID with the feature attribute table.

Geographic Analysis

What distinguishes GIS from other databases or information systems are its spatial analysis functions. These functions use spatial and non-spatial data to answer questions about the real world. The answers could relate to a presentation of the current data (first level use), some patterns in the current data (second level use) and predictions of what the data could be at a different place or time (third level use). Geographic analysis is carried out using the layers of map information created in vector or raster data formats and associated attribute data to find solutions to specific problems. In each case the problem needs to be defined clearly before the relevant map layers and analysis procedures can be identified. For instance, if the problem is to find optimal locations for siting of wells for conjunctive use in an irrigation project area, information about the geographical features influencing the groundwater recharge will be required. These will include maps of existing well locations, rainfall, land use, soils and command area of the project, all of which influence recharge. Regions with recharge above a selected threshold value may be considered suitable for additional wells. Further, if the area happens to be near the coast, a buffer zone may be required within which no wells can be sited to prevent sea water intrusion. Similarly buffer zones may be required on either sides of canals to prevent drawal of canal water by the wells. What could happen to the ground water levels and quality in the area if the present use is persisted with or changed could be the subject of another study where the GIS can help to provide more realistic answers.

Most standard GIS software come with basic analytical tools that permit overlays of thematic maps, creation of buffers, etc., in addition to calculations of lengths and areas. Overlay operations permit overlaying one polygon over the other to generate a new map of their intersections which are new polygon combinations with desired homogenous properties with respect to specified polygon attributes.



Source: FAO

STEPS IN BUILDING A GIS

The way in which a GIS is built will depend on the way information will be used in the decision-making process. Building a GIS proceeds through at least 4 stages:

(i) Defining the objectives
(ii) Building the spatial and attribute data bases
(iii) Database management for geographic analysis
(iv) Presenting results in the form of maps, etc.

The definition of objectives or the problem to be solved using GIS is critical to the choice of spatial and attribute databases. Once the problem is defined and the relevant map layers and attribute data are identified, building databases involves:

(i) database design
(ii) entering spatial data
(iii) creating topology
(iv) entering attribute data

Designing the database requires identifying:

(i) study area boundaries
(ii) coordinate system
(iii) data layers
(iv) features in each layer
(v) attributes for each feature type
(vi) coding and organizing attributes

Depending on whether the map sources are two dimensional maps of the area or remote sensing imageries, data is entered in vector or raster format. In the vector format, entering spatial data and creating topology are components of overall digitization process. Raster data will need to be vectorized before topologies can be built. Attribute data is created in the form of database files with one field, the feature identification field, in common with the spatial data base created during the spatial data entry process.

Data base management refers to translating the digitized map into real world coordinates, identifying coverages for analysis and maintaining the data base.

Presenting the maps for decision-making is facilitated by creating customized maps using the various facilities available in GIS software.

SUMMARY

A GIS is a computer based tool for geographical analysis of information. It is not simply a digitized map, nor does it hold maps. It holds a database of spatial data and attribute or descriptive information about features on a map which can be used to create desired maps. *The crucial concept of GIS is the separation of spatial or geographic reference information and attribute or descriptive information of map features for data entry and database development, and their linkage during analysis*. Central to both spatial and attribute information is the database management concept. The separation of the two types of information facilitates entering the spatial information (map) into computers in a digitized form and establishing connectivity (topology) between different stored map features (points, lines and polygons). The feature attribute data is entered independently taking care to introduce an identification variable which is in common with the identification variable for each feature that is common with the spatial database. For geographic analysis, the spatial and attribute data are linked through this unique identifier variable common to the two types of data bases.

Initially, spatial data capture is in spatial units and coordinates of the data capture tool. To translate the map information into real world information of locations, distances and areas these need to be translated to real world units through appropriate transformations of scale and map projections.

The digitized maps and their associated feature attributes are the building blocks of the GIS. The maps can be created and stored in different layers, with each layer containing information about one feature. They can be overlaid over each other to obtain new maps (coverages) with new polygons that are homogeneous with respect to specified feature attributes of maps that were used in the overlays. The overlay operations must be between maps with exact boundary fits. Exact fits are obtained between maps only if they are created in the same projection and scale. To make exact fits, appropriate map projection and scale transformation operations will be needed before geographic analysis can be performed using overlay operations.

References

1. Aronoff, S. (1991) Geographic Information Systems: A management Perspective, WDL Publications, canada.
2. ESRI (1990) Understanding GIS
3. FAO : http://www.fao.org/sd/eidirect/gis/EIgis000.htm
4. Heywood I, Cornelius S.,and Carver,S. (1998) An Introduction to Geographical Information Systems, Longman pub., 279 pp.
5. Longley, P.A., Goodchild,M.F., Maguire,D.J. and Rhind,D.W (eds) (1999) Geographic Information Systems, Volumes 1 & 2, Wiley pub.
6. *http://www.colorado.edu/geography/gcraft/notes*