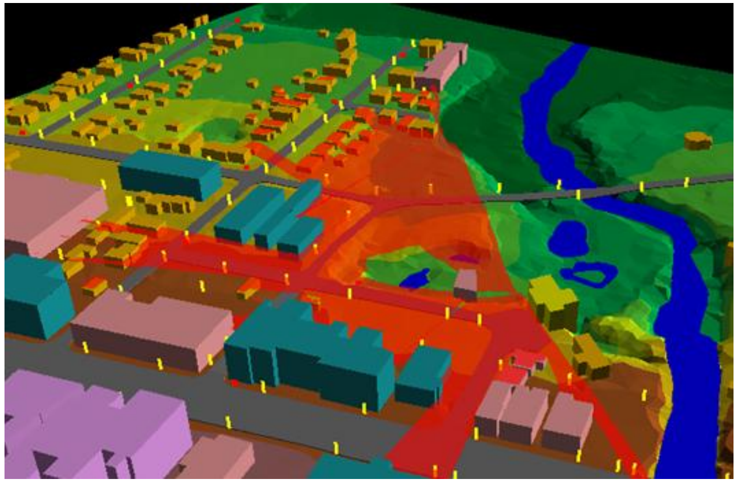
DATA SOURCES AND INPUT IN GIS



By

Prof. A. Balasubramanian

Centre for Advanced Studies in Earth Science,

University of Mysore, Mysore

1. GIS stands for 'Geographic Information System'.

It is a computer-based system that accepts and stores geographically referenced data, links it with their attributes (data in tables) and allows the user to carry out a wide range of information processing including manipulation, analysis and modeling. The GIS is an organized collection of Data ,Software, Hardware, Network , People, Procedures and Methods of Applications.

GIS represents geographic data and map information as data layers.

GIS is helpful to perform a lot of statistical and numerical analysis, on the data. It is also useful in proper visualization of data for making cartographic output, tables and charts.

2. GIS includes 5 basic components.
They are: 1) Data, 2) Software, Hardware, 3) People, 4) Procedures and   Methods and 5) Applications**.**

People are the most important part of a GIS.
Without these components, GIS will not function.

3. GIS-DATA & Sources:

The most important and expensive component of the Geographic Information System is Data which is generally known as fuel for GIS.
GIS data is combination of graphic and tabular data.

Perhaps the most important component of a GIS is in the part of data used in GIS. The data for GIS can be derived from various sources.
A wide variety of data sources exist for both spatial and attribute data.

People can use topo maps, aerial photographs, satellite images, data of ground surveys, readily available reports and government or research publications.

The most common general sources for spatial data are: hard copy maps;
aerial photographs;
remotely-sensed imagery;
point data, samples from surveys; and
existing digital data files.

Existing hard copy maps, e.g. sometimes referred to as *analogue maps*, provide the most popular source for any GIS project.

Attribute data has an even wider variety of data sources.

They are: GIS Data from Libraries,

Data from National and International Mapping Agencies,

Elevation Data , Bathymetry Data**,** Georeferenced Images,

Time Series Multispectral Satellite Images,

State and national agencies and Detailed District/ Municipal Data.

4. The data input process in GIS:

Any textual or tabular data than can be referenced to a geographic feature, e.g. a point, line, or area, can be input into a GIS.

Attribute data is usually input by manual keying or via DBMS software.

ASCII format is the default standard for the transfer and conversion of attribute information.

5. Automated Surveying:

Directly determines the actual horizontal and vertical positions of objects.

Two kinds of measurements are made: distance and direction.

Distance is measured using pedometer, chains and tapes. The direction measurements were made with transits and theodolites.

The modern surveyors have a number of automated tools to make distance and direction measurements easier.

More electronic systems measure distance using the time of travel of beams of light or radio waves.
The total station captures distance and direction data in digital form.

6. Global Positioning System (GPS):
It is the new tool used for determining accurate positions on the surface of the earth computes positions from signals received from a series of satellites.

7. Four basic procedures of data Input: Data entry is the operation of encoding both types of data into the GIS database formats. There are at least four basic procedures for inputting spatial data into a GIS. These are:

Manual digitizing;
Automatic scanning;
Entry of coordinates using coordinate geometry;
and the Conversion of existing digital data.
GIS Data are obtained from data capture, map digitization, scanning and digitization, semi-automatic digitization and automatic digitization.
The creation of a clean digital database is the most important and time consuming task upon which the usefulness of the GIS depends.

The process of creating the GIS data from the analog data or paper format is called digitization. **Digitization process** involves registering of raster image using few GCP (ground control point) or known coordinates.

This process is widely known as rubber sheeting or georefrencing. Polygon, lines and points are created by digitizing raster image.

8. Digitizing:

GIS spatial data entry is done by manual digitizing.
A digitizer is an electronic device consisting of a table upon which the map or drawing is placed.
The user traces the spatial features with a hand-held magnetic pen, often called a *mouse* or cursor.
All points that are recorded are registered against positional control points.
The coordinates are recorded in a user defined coordinate system or map projection.

Latitude and longitude and UTM is most often used.

Digitizing can be done in a *point mode*, where single points are recorded one at a time, or in a *stream mode*, where a point is collected on regular intervals of time or distance, measured by an X and Y movement, e.g. every 3 metres.
The most common features to digitize are: 1) points, 2) lines, and 3) polygons.
The Point features are simply created by identify the x,y coordinate.

The Lines can be snapped together by setting the snapping environment so that two lines can share the same node feature

The Polygon can be created using an Auto-complete function that automatically creates closed polygonal areas and facilitates the creation of polygon features that share a common line with an adjacent polygon.

4) The digitizing process is just one element in a series that contribute to the overall error present in a final database.

It is important that a final digitized product is only as accurate as the map which is being digitized (and is probably somewhat less accurate because of the error involved in the digitizing process).

9. Spaghetti mode of digitizing :

Most GIS's use a *spaghetti mode* of digitizing.
This allows the user to simply digitize lines by indicating a start point and an end point.
Data can be captured in point or stream mode.

10. Automatic Scanning:

A variety of scanning devices exist for the automatic capture of spatial data.

All have the advantage of being able to capture spatial features from a map at a rapid rate of speed.

Scanners are generally expensive to acquire and operate.

Large data capture work are done using scanning technology.

Today, a variety of *line following* and *text recognition* techniques are available.

11. COGO- Coordinate Geometry:

A third technique for the input of spatial data involves the calculation and entry of coordinates using coordinate geometry (COGO) procedures. This involves entering, from survey data, the explicit measurement of features from some known locations.

This method is useful for creating very precise cartographic definitions of property, and accordingly is more appropriate for land records management at the cadastral or municipal scale.

12. Conversion of Existing Digital Data:

A fourth technique that is becoming increasingly popular for data input is the conversion of existing digital data.

A variety of spatial data, including digital maps, are openly available from a wide range of government and private sources.

The most common digital data to be used in a GIS is data from CAD systems.
A number of data conversion programs exist, mostly from GIS software vendors, to transform data from CAD formats to a raster or topological GIS data format.
Most GIS software vendors also provide an ASCII data exchange format specific to their product.
Attribute data is usually handled as ASCII text files.

Some of them include:

1. IGDS - Interactive Graphics Design Software (Intergraph / Microstation)
2. DLG - Digital Line Graph (US Geological Survey)
3. DXF - Drawing Exchange Format (Autocad)
4. GENERATE - ARC/INFO Graphic Exchange Format
5. EXPORT - ARC/INFO Export Format .

Georeferencing

Georeferencing is the process of converting a image in file coordinates or page coordinates to a file in map coordinates in a specific map project, coordinate system, map projection and datum.

For example, a scanned map can have an origin point and a raster association where each point on the map is identifiable by it's file coordinates (e.g. 1244, 1515). The task then is to convert these file coordinates to map coordinates.

Satellite images can be bought 'unprocessed', in which case the user is responsible for georeferencing to known coordinates.

A satellite image is simply a raster dataset with file coordinates (rows/columns) that need to be converted to map coordinates.

Likewise, aerial photography is commonly used in GIS operations and sometimes comes as already geo-referenced data.

The other type of product that is often georeferenced using GIS software is a hardcopy map which will be used as a backdrop for georeferencing. An example is a topographic map.

13. Data editing and quality assurance:

Data editing and verification is in response to the errors that arise during the encoding of spatial and non-spatial data.

The editing of spatial data is a time consuming. Several kinds of errors can occur during data input. They can be classified as: Incompleteness of the spatial data.

Locational placement errors of spatial data.

Distortion of the spatial data, Incorrect linkages between spatial and attribute data.

Attribute data is wrong or incomplete.

This includes missing points, line segments, and/or polygons.

14. Spatial Data Errors and Attribute Data Errors

15. Integrating different data sources:

A GIS can integrate spatial data with other existing data resources, often stored in a corporate DBMS.

The integration of spatial data (often proprietary to the GIS software), and tabular data stored in a DBMS is a key functionality afforded by GIS.

<u>Formats:</u> many different format standards exist for geographical data.

A good GIS can accept and generate datasets in a wide range of standard formats.

Projections: map projections are very common, e.g. Mercator, Universal Transverse Mercator (UTM).

A good GIS can convert data from one projection to another, or to latitude/longitude.

Scale: data may be input at a variety of scales. maps of the same area at different scales will often show the same features.

Acquiring geographic data is an important factor in any geographic information system (GIS) effort.

It has been estimated that data acquisition typically consumes 60 to 80 percent of the time and money spent on any given project.

Primary Data Capture

Primary data capture is a direct data acquisition methodology that is usually associated with some type of in-the-field effort.

In the case of vector data, directly captured data commonly comes from a global positioning system (GPS) or other types of surveying equipment such as a total station.

Use of a total station allows field crews to quickly and accurately derive the topography for a particular landscape.

Crowdsourcing is a data collection method whereby users contribute freely to building spatial databases. Raster data obtained via direct capture comes more commonly from remotely sensed sources.

Secondary Data Capture

Secondary data capture is an indirect methodology that utilizes the vast amount of existing geospatial data available in both digital and hard-copy formats.

Topology

GIS contains many sophisticated tools for creating topological relationships between features within and between datasets.

For example, GIS also uses a rules-based methodology for maintaining the topological integrity of datasets.

Feature attributes/ID's

After digitizing features, the next step is to assign ID values to each feature that has been digitized.

This is only necessary if we need to uniquely identify each feature.

Discovering Geographic Data

The great thing about geographically referenced data is that datasets compiled by independent agencies can be combined together and will align with each other in a coherent way.

In addition to representing things graphically, GIS data includes attributes that reflect measurements, classifications or other observations that may be critical to understanding what things are and how they are related.

Exploring GIS Data

When we finally find some GIS data, it is necessary to understand whether the data is suitable for our purposes or not.

Critical aspects of a dataset should be addressed in the data documentation, or **Metadata** for example:

- The subject of the dataset?
- What time period is represented?
- Who collected the data and why?

Formats for Geographic Data

Geographic data formats are the vehicles that we use for **encoding** and **exchanging** observations and ideas about places and things that happen in space.