PDHonline Course L153 (5 PDH)

Geographic Information Systems (GIS) - Hardware and software in GIS

Instructor: Steve Ramroop, Ph.D.

2020

PDH Online | PDH Center

5272 Meadow Estates Drive
Fairfax, VA 22030-6658
Phone: 703-988-0088
www.PDHonline.com

An Approved Continuing Education Provider

Slide 1

Lecture 5 Content

- **Geographic Information Systems (GIS)**

Data in GIS – Acquisition and input

This lecture presents a very important aspect of GIS implementation which is, “Data in GIS-Acquisition and input”. Having understood the basics of hardware and software discussed in the four lectures of this course, this last lecture concludes this course in the series, but also as an incentive to take the second course in the series.

Slide 2

- **Introduction - GIS Origin**
- **What is GIS?**
- **Data Acquisition**
  - **Representing reality**
- **Data Categories**
  - **digital data**
  - **remote sensing data**
  - **automated surveying data**
  - **global positioning system data**

This slide shows the contents of this lecture. Each one of these categories will be discussed in more detail in the following slides. However, prior to presenting the data categories, one must understand that there are various definitions for a GIS. It is the intent for you to come up with your own definition which will be in your own words but contains the essential component of what is GIS. Also presented is the understanding of how difficult it is to select appropriate data sets for the intended GIS application.

Slide 3

Introduction

GIS origin :

- **according to COWEN (1989) first recorded use of a GIS is by a French Cartographer called: LOUIS ALEXANDER**

where:

- **Overlays of troop position to demonstrate military movement during the American Revolution (1775-1783)**
- **In 1854, Dr. John Snow demonstrated the use of map overlays by applying it to :**
  - **death from cholera in central London**
  - **well location**

- **In 1950's modern GIS began**

This slide gives the some basic information regarding the origin of GIS. GIS was initially used in a very crude manual manner to overlay troop positions in the military. In London, Dr. Snow in 1845 used GIS to find out the probable cause of death due to cholera.

Slide 4

In 1854, Dr. John Snow showed that the death from cholera in central London was due to a well location which spread cholera

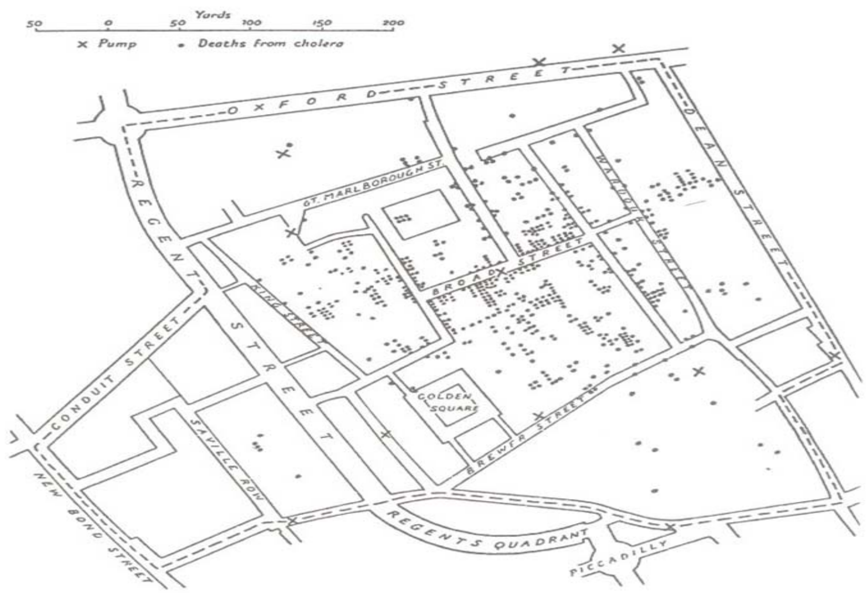


Figure 2. The Snow map of cholera incidence in the area of Broad Street, London, in 1854. The contaminated water pump is located at the center of the map, just to the right of the D in BROAD STREET. Source: Figure 1, page 174 in Gilbert (1958), reproduced by permission.

This slide shows the map used by Dr. Snow. From the slide Dr. Snow was able to infer that the reason for the high deaths is likely to be due to the water pump on Broad Street. Around this water pump there was a large record of deaths.

Slide 5

❑What is a GIS?

– **Many definitions for GIS.**

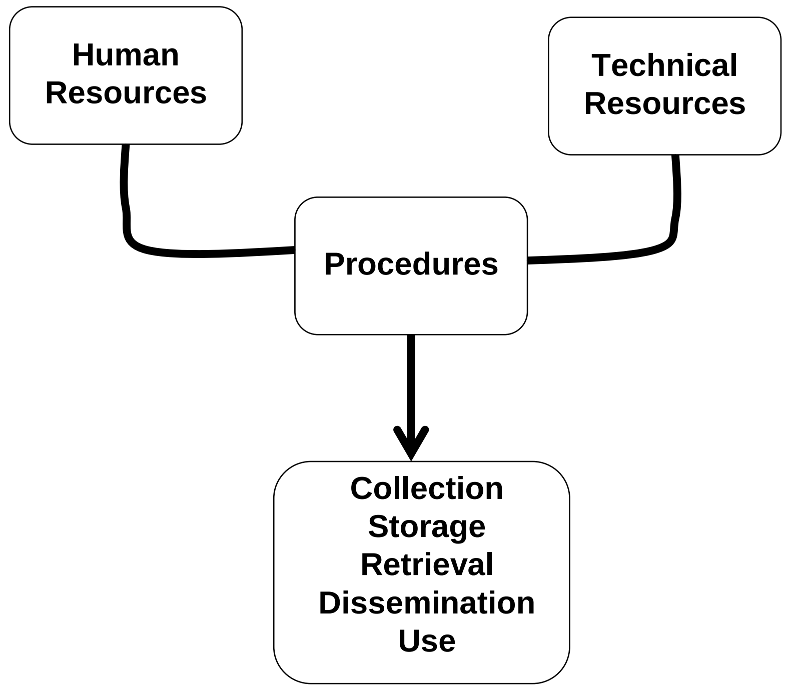
– **Sometimes referred to as Land Information Systems (LIS) but this is with particular reference to land ownership.**

– **Definitions:**

  - ***A GIS/LIS may be defined as a combination of human and technical resources, together with a set of procedures, which produces information to support decision makers.***

There are many GIS definitions. Traditional GIS definition was focused on the definition towards Land Information Systems (LIS). The focus now has been towards the linkage of the human, technical resources and a set of procedures which is defined by the GIS software. The end result is to support decision makers.

Slide 6



This slide shows the diagram summarizing the definition of GIS/LIS. Of all the definitions which are presented in this lecture you will be required come up with your own definition of GIS.

Slide 7

- a simple and concise definition given by Fisher and Lindenberg is :
  - " Geographical Information Systems are defined as the management, analysis, and manipulation of spatially referenced information in a problem solving environment."

- a modern description of GIS could be:
  - " a computer based system that is used to store, manipulate, analyze and output geographic information."

This slide gives two definitions of GIS. There are some obvious common characteristics between the two definitions. For example: management, store, manipulate, output geographic data, to assist in making better decisions.

Slide 8

- **the standards committee of the UK Association of Geographic Information (AGI) defined GIS as:**
  - **" A system for handling data which are directly and indirectly, spatially referenced to earth. It may be used for capturing , storing, validating, maintaining, manipulating, analyzing, displaying or managing such data. It is normally considered to involve a spatially referenced computer database and appropriate software. A primary function of a GIS is its ability to INTEGRATE data from a variety of sources."**

This is a lengthy definition of GIS offered by AGI. But this definition is lacking the notion of providing output to assist in making better decisions.

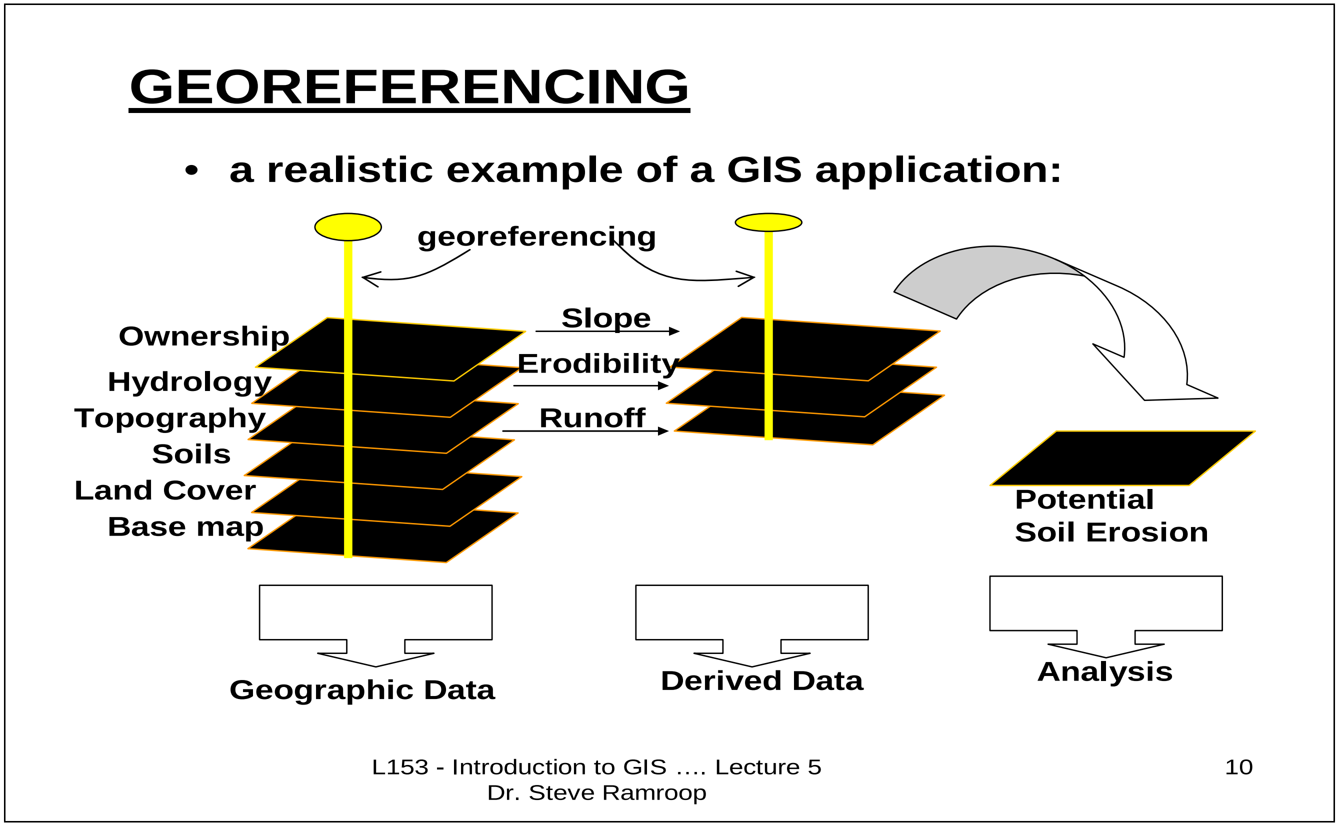
Slide 9

- **Other definitions of GIS:**
  - **Rhind (1989, p:28) “a computer system that can hold and use data describing places on the Earth’s surface”**
  - **Department of Environment (1987, p:132) “a system for capturing, storing, checking, integrating, manipulating, analysing and displaying data which is spatially referenced to the earth”**

Some other definitions are shown on this slide. Rhind’s definition is too simplistic while the second definition seems to have all that is needed for a good definition, except that it is lacking the notion of using the output to assist in making better decisions.

My recommendation of a definition of a GIS is as follows:
GIS is a system used to capture, store, manage, manipulate, and output spatially referenced data sets to assist in making better decisions.

Slide 10



This slide gives a realistic example of a GIS application and the notion of “georeferencing”. This term refers to the ability to attach a geographic coordinate system to data sets. From the example, on this slide all of the data sets are using the same coordinate system which allows them to be overlaid.

From the example, the raw data set are the original data sets collected and when they are processed (from the example: Slope, Erodibility, Runoff), the new data sets are called Derived Data. The final output is the Potential Soil Erosion Map is generated as output which will be used for further analysis.

Slide 11

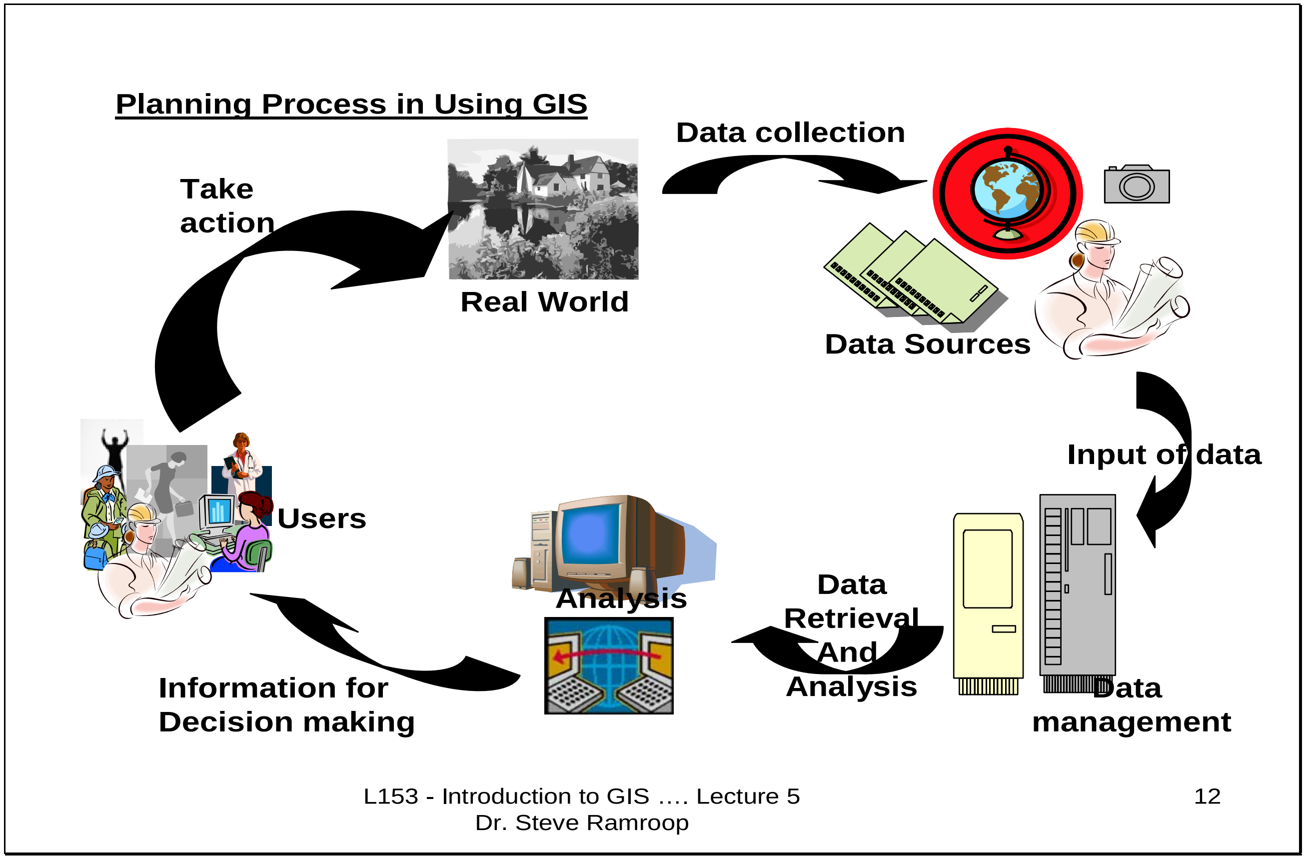
From the previous diagram of a realistic example of a GIS application, the geographic data sets are collected and referenced using the same coordinate system which is called *"georeferencing"*.

The data sets are processed too generate a new sets of data which is called *"derived"* data.

Using the derived data, analysis is done to address areas where there is potential for soil erosion

This slide gives more information on the previous slide.

Slide 12



This slide shows the planning process in developing a GIS application. It is cyclic. The diagram is self explanatory and further notes are given on the following slide.

Note that the process starts and ends in the real world….but in actual fact, it does not really end because the end user may decide to update the data sets after the first iteration…..hence the reason why it can be a continuous cycle.

Slide 13

From the previous diagram of the cyclic process:

- **task begins and ends with the real world**
- **there are varied data sources**
- **data management have been increasing at the pace of computer development**
- **analysis concerns with GIS and its associated software**
- **users will use the analysis to take the relevant action**
- **entire process is cyclic and is seen as iterative such that an acceptable solution is obtained**

This slide gives some notes on the previous slide. The end used has the ability to say if the final result is satisfactory or do they need refinement or more data sets. If so, then the procedure will have to be repeated.

Slide 14

❑Data Acquisition

– availability of data is a critical input into the final usefulness of the GIS

– in reality the world has a multitude of data

– because of this multitude →a decision is needed in terms of which data set(s) is/are important for the intended GIS application

Data Acquisition is the third aspect of this lecture which will be discussed. Data for a GIS is critical and there are various methods available used to acquire data. Reality is complex and there is a countless amount of data that exists. A decision is needed to identify what data is important to achieve the objectives of the intended GIS application. The amount of data collected is related to the cost of the GIS application project. Collecting unnecessary will mean wastage of money. Therefore a balance is needed.

Slide 15

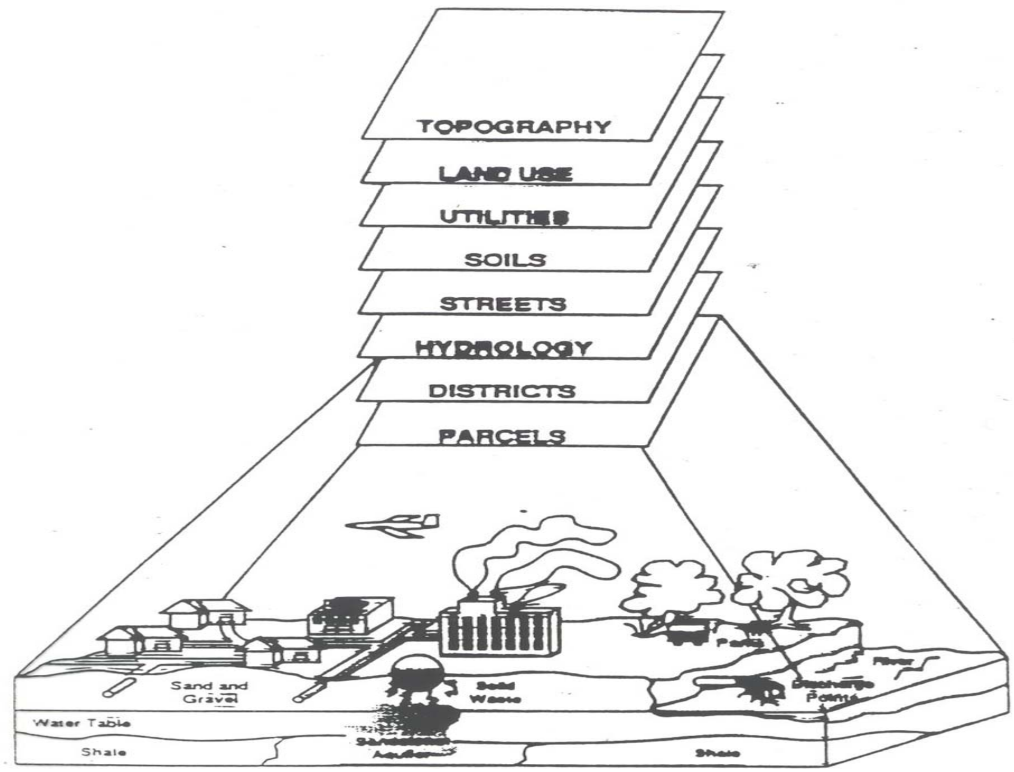
- **Representing reality**
  - **the real world consist of many geographies.**

Geography is defined as:

" The science dealing with the aerial differentiation of the earth's surface, as shown in the character, arrangement, and interrelations over the world of such elements as climate, elevation, soil, vegetation, population, landuse, industries, or states, and the unit areas formed by the complex of these individual elements."

The next topic to be discussed in this lecture is representing reality. The various data sets that exist in reality are an indication of the many geographies that exists. A definition of Geography is given. You are not required to learn this definition but note that there are some common features between GIS and this definition. For example using earth related data sets and their interrelations between them. This is indicative to the capabilities of a GIS.

Slide 16



The real world consists of many geographies

L153 - Introduction to GIS …. Lecture 5
Dr. Steve Ramroop

16

This slide shows the complexity of reality. In this example only 8 data sets were extracted but the other data sets were omitted such as pollution, sub-surface, airspace, and such like.

Slide 17

– **From the previous figure, the real world consist of many geographies:**

- **real world is very complex**
- **not possible to record/store all the data**
- **balance is sought between the final acceptable GIS result against what is the required**
- **there is the cost for collecting the minimum data at the minimum cost**

This slide gives some brief notes on the previous slide. It is not possible to collect all data that exist in reality because of the cost required to collect them, and the need for enough computer memory to store the multiple data sets. Therefore a balance is sought when deciding on suitable data sets for the intended GIS application.

Slide 18

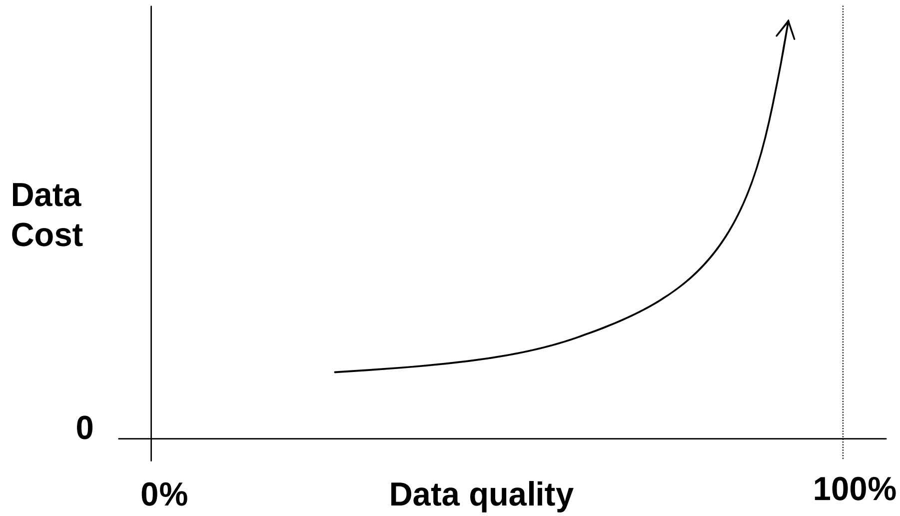
Note:-

- the data used in a GIS represent something about the real world at some point in time
- abstraction of reality
- not all data is required
- constraint is placed on the GIS in that you can’t use data you don’t have
- data are costly to collect
- the most cost-effective data collection is to collect only the data you need
- the optimum data quality is the minimum level of quality that will do the job
- as the need for improved data quality increases then the data cost also increases

This slide gives some consideration about selecting data sets from reality. They are self explanatory. You are expected to review the considerations and be able to write a paragraph about selecting data sets from reality.

Slide 19

Relationship of data quality and data cost



This slide shows, as the quality of original data increases so to will the cost of the data. Therefore a balance is sought so that the cost of data sets does not make the overall GIS project very costly.

Slide 20

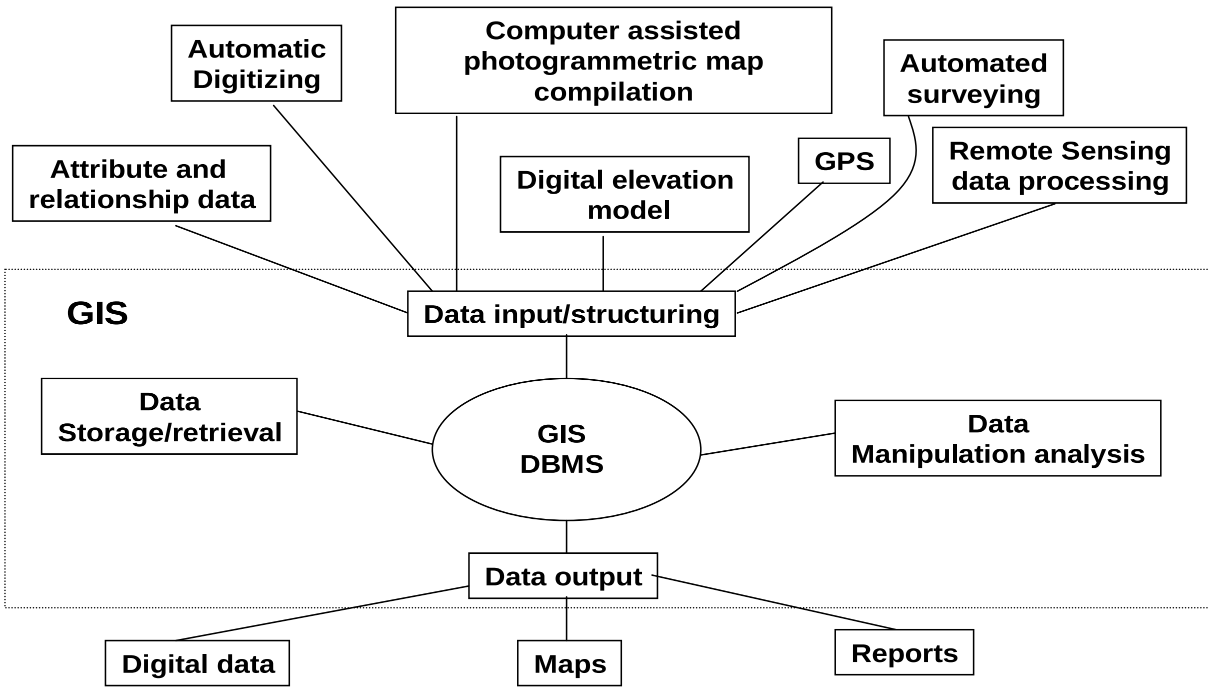
- **From the data quality and data cost table the cost of collecting geographic data increases with the improvement of the quality of geographic data**

- **Therefore, to reduce the cost of data the minimum quality of data is sought which is striking a balance with the intended GIS application and the available budget**

This slide gives some notes on the previous slide comparing data cost with data quality.

Slide 21

❑Data Categories



This slide summarizes the different data categories available to a GIS. These categories will be discussed further in the following slides. Other data categories are Internet, other organizations, and such like.

Slide 22

The following slides shows the various data categories

1. Digitizing data

a) Manual digitizing

- manual tracing of all graphical elements
- makes use of existing map, (or field survey), digitizing table and a pointing devise
- efficiency of digitizing depends on the quality of the digitizing software and the skill of the operator
- time consuming and error prone

This slide is self explanatory and presents digitized data. Digitizing is the conversion features from a hardcopy map into digital data. One method described here is Manual Digitizing. This slide shows some characteristics of manual digitizing.

Slide 23

b) Scan digitizing

- enables automatic conversion of graphic documents into digital data
- however, difficult to extract and structure desired features (especially when the input document is not very simple)
- need to separate the similar features on separate maps and then scan
- Present GIS software has intelligent line following software which can digitize automatically

The second method of digitizing is Scan Digitizing. This method requires data preparation before any hardcopy maps are scanned. That is, maps to be scanned must be clean from unwanted details. The final scanned data set will require more post processing steps in order to get the features onto separate data layers. This can be very time consuming.

Slide 24

2. Remote sensing data

- **Data collected by measuring the GEOMETRIC and THEMATIC PROPERTIES of OBJECTS in the environment without touching them and using various devices in the AIR or SPACE**
- **based upon the following measurables:**
  - **spectral entities:**
    - **i.e. recognition of reflected or emitted parts of different electro-magnetic wavelengths**
  - **spatial entities:**
    - **i.e. the shapes of objects and their relations to other objects**
  - **temporal entities**
    - **i.e. changes in objects with time and the time at which the data set was taken**

This slide shows the second data category which is from satellites. Through the use of the reflected energy from objects on the earth surface, we are able to capture data from a distance which is called remote sensing data. It captures the spectral, spatial, and temporal entities. Further details on these are shown on the slide.

Slide 25

3. Computer assisted photogrammetric compilation

- use of aerial photographs to create a realistic (stereo) model of the terrain at the time of exposure
- model is used to make:
  - maps (hardcopy)
  - or stored as digital data (x,y,z)

4. Attribute and relationship data

- textual data having numerical and/or character value
- can be stored in a normal database
- e.g. :- census related data sets; Landuse, Land cover data; Natural resource data; etc.

Two other data categories are shown here. The first is through the use of aerial photos. The photos are used together with ground control information such as, coordinates are used to generate a three dimensional model of the area captured in the photos. This is all done using a softcopy workstation and the computer. Therefore, using such a model one can plot features directly from the 3-D model.

The next data categories are related attribute data. This data typically refers to textual data which can be collected through census fieldwork, or from other organizations that collect textual data.

Slide 26

5. Digital elevation model collection

- **set of elevation measurements for location distributed over the land surface**
- **used to analyze the topography of an area**
  - **e.g. calculate cut-and-fill, locate flooded areas, intervisibility analyses, etc.**
- **data capture methods are :**
  - **Regular grid – landform is created using a grid**
  - **Contours – landform is created using contour lines**
  - **Profiles – landform is created using a set of profiles**
  - **Triangular Irregular Network (TIN) – landform is created using a set of irregular triangles**

This next data category is the existence of a digital representation of the landform from which features can be extracted which has ground control coordinate systems. There are various data capture methods which are used to create such a digital elevation model (DEM).

Slide 27

6. Automated Surveying data

- process whereby the field survey data are directly stored in digital format
- data is down-loaded onto the computer and directly processed
- examples:- total station (with data logger) satellite imagery, etc.

Another data category is through the use of automated equipment that captures data directly from the field. The data sets collected are typically in digital format in the filed and is usually easily downloaded onto computers.

Slide 28

7. Global Positioning System (GPS) data

- use of satellites orbiting the earth to determine positions on the earth's surface
- x, y, z, coordinates are recorded
- digital field data can be down-loaded onto the computer for further processing

The last but not the least data category is through GPS receivers. This is the use of a receiver and a set of satellites which are used to position points on the earth surface. This is common and well known by many professions. The data collected is in digital format and can be easily downloaded in the computer.

Slide 29

… The End …