

Geographic Information System (GIS)



Dr. Eman Omar



Geographic Data Input & Output Methods & Data Quality.

Introduction



- For GIS to be useful it must be able to receive and produce information effectively.
- Data input and output are the ways that a GIS use to communicate with the world.
- Input and output technology has seen rapid advances.

Data Input

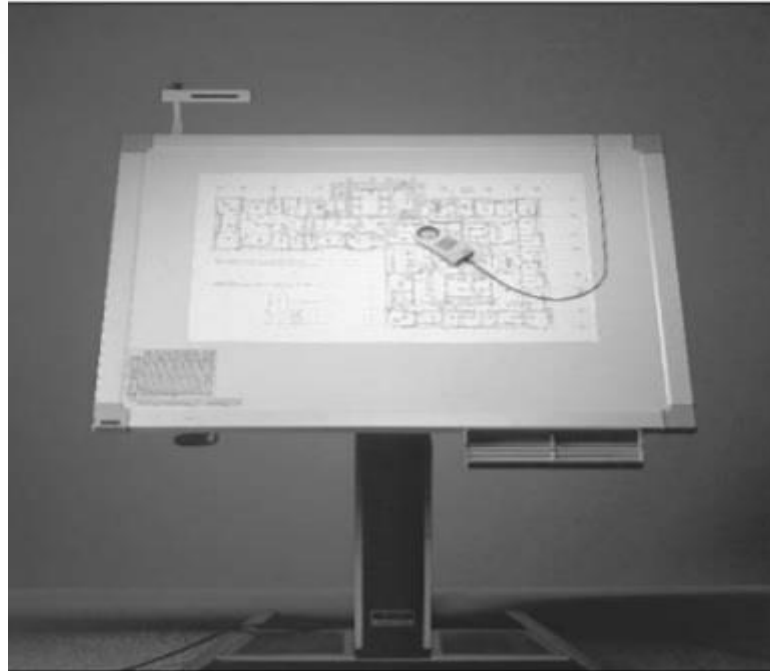


- Is the procedure of encoding data into a computer-readable form and writing the data to a GIS database.
- The initial cost of building the GIS database is 5 to 10 times the cost of GIS HW and SW.
- The creation of accurate well-documented geo-data is critical to a GIS.
- Documentation describes the quality of the data, which is important to assess its suitability for a particular application.
- Data quality information includes date of collection, accuracy, completeness and the collection method.

Data Input methods.



- **Manual-digitizing** using digitizing Tablet: Is the most widely used method for map features. The map is mounted to a digitizing table and a pointing device is used to trace the map features. It is time consuming.



Data Input methods cont.



- **Keyboard Data Entry:** The keyboard is used to input the **non-spatial data** such as field measurements that can be recorded using hand-held computer devices.
- It is not suitable for entering map features.
- **Coordinate Geometry Procedures:** Used with spatial features. Collect the geometry (coordinates) of features to create GIS compatible data files.
- It is more accurate than manual digitizing.

Data Input methods cont.

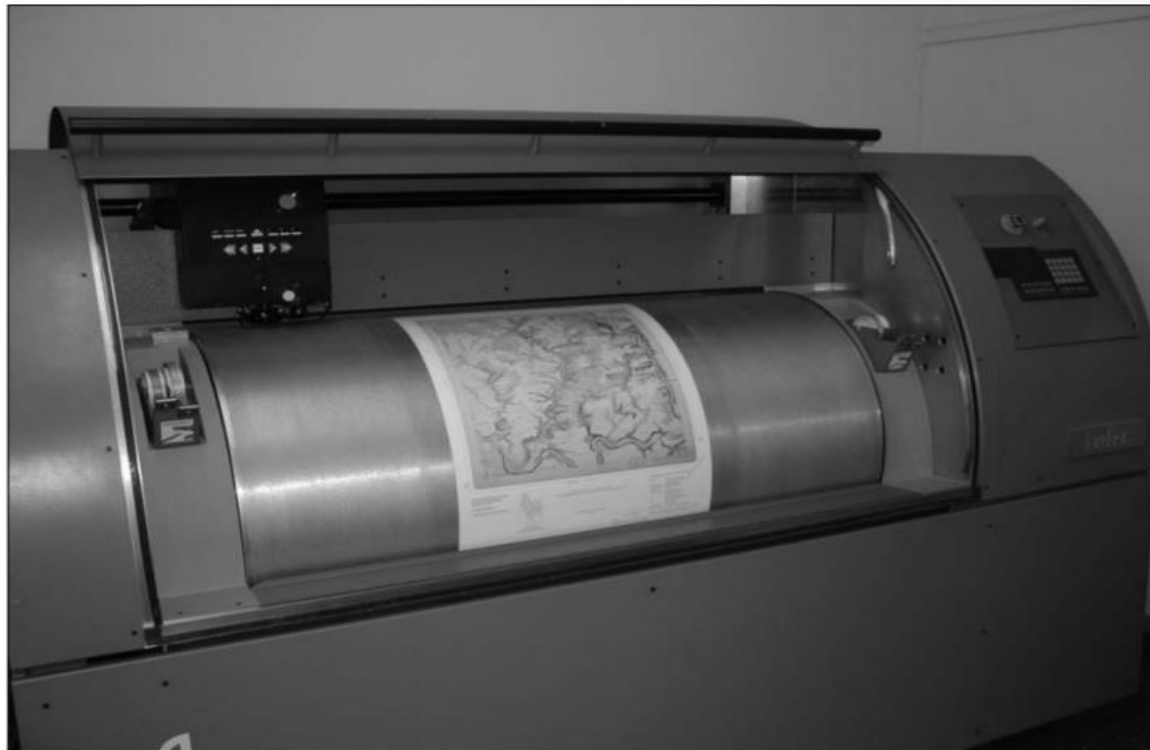


- **Photogrammetry** (the art and science of obtaining reliable spatial measurements from aerial photography and other remote sensing images).
- **Scanning**: . It is faster than manual digitizing.
- There are two types of scanners:
 1. Flat-bed scanners.
 2. Drum-scanners.

Data Input methods cont.



- **Scanning:** You place the map or image edge-first onto the drum inside, and the drum spins while a laser reads (scans) the document. Normally, the scanning software converts the document into a raster image that your GIS software can easily convert to vector format



Data Input methods cont.



- **Translation of existing digital files:** The data might be existing in a digital form, that can be easily converted to another format to be used with the GIS software.
- The web could be a good source of geo-spatial data.

Data Output methods.



- Output is the procedure by which information from the GIS is presented in a form suitable to the user.
- Data are output in one of three formats:
 1. Hardcopy: Permanent means of display.
 2. Softcopy: Can be viewed on a monitor. It can be interactive.
 3. Electronic: Computer-compatible files.

Hardcopy Devices



- Pen plotter.
- Line printer black and white.
- Dot matrix printer.
- Colour dot matrix printer.
- Ink jet plotters.
- Thermal plotters.
- Optical film writer.
- Laser printers and plotters.

Data Quality



- Why Data Quality is important?
 1. Recovery of an error in the data could be more expensive than implementing the whole system.
 2. Data error could make a significant error in the output.
 3. It is important for gaining the user trust in the system.
 4. Knowing the data quality could determine the suitable applications for its use.

Components of Data Quality



- **Micro Level components:**
related to individual data elements.
- **Macro Level components:**
Pertains to the dataset as a whole.
- **Usage components:**
Specific to the resources of the organization.

Components of Data Quality

cont.



- **Micro Level components:** include
 1. **Positional Accuracy:** Is the expected deviance in the geographic location of an object in the dataset (e.g. On a map) from its true ground position.
- It is tested by taking a sample of point coordinates and comparing them to data from an independent data source.
- Positional accuracy has two components:
 - **Bias:** measured by the mean or average error.
 - **Precision:** measured by the STD of the error.
- RMS is an alternative. $RMS = \sqrt{(X_{act} - X_{obs})^2 / n}$

Components of Data Quality

cont.



2. **Attribute Accuracy**: Variables can be discrete or continuous. Discrete variables can take only finite number of values. Continuous variable can take any value. Similarly, equation can be used to measure its accuracy.
3. **Logical Consistency**: How well logical relations between data is maintained. For example, data could change over time. These differences are called **Sliver**.
4. **Resolution**: The smallest unit represented also called minimum mapping unit.

Components of Data Quality

cont.



- **Macro Level components:** includes
 1. **Completeness:** It has three categories Coverage, Classification and Verification.
 2. **Time:** Time is critical in many Geo data. Most of geographic data can change over time.
 3. **Lineage:** History of data
 - Source of data
 - Processing Steps

Components of Data Quality

cont.



- **Usage components:** includes
 1. **Accessibility:** Refers to the ease of obtaining and using the data. The accessibility of the data may be restricted because it is privately-held.
 2. **Direct and Indirect Cost:** The direct cost is the cost of buying or collecting the data. The indirect cost includes all the time and materials used to make use of the data.

Sources of Errors



Error is introduced in **every step** in geo data generation and usage.

1. **Errors in data collection**

- Inaccuracies in field data collection.
- Errors in existing sources of data (maps).
- Errors in /or misunderstanding of remotely sensed data.

2. **Data Input**

- Incorrect digitizing of data.
- Errors due to uncertain edges of boundaries.

Sources of Errors cont.



3. Data Storage

- ✓ Insufficient precision for vector data.
- ✓ Insufficient resolution for raster data.

4. Data Manipulation

- ☐ Boundary Errors.
- ☐ Errors in data classification.
- ☐ Overlay Operations.

5. Data Output

- ❖ Scaling Inaccuracies.
- ❖ Instability of storage media.
- ❖ Errors due to output device.

Sources of Errors cont.



6. Use of results

- Errors due to misunderstanding.
- Inappropriate use.

Review



1. Data Input Methods.
2. Data Output Methods.
3. Data Quality.
4. Sources of Errors.



Thank You
Questions ??