GIS Data Collection

Introduction

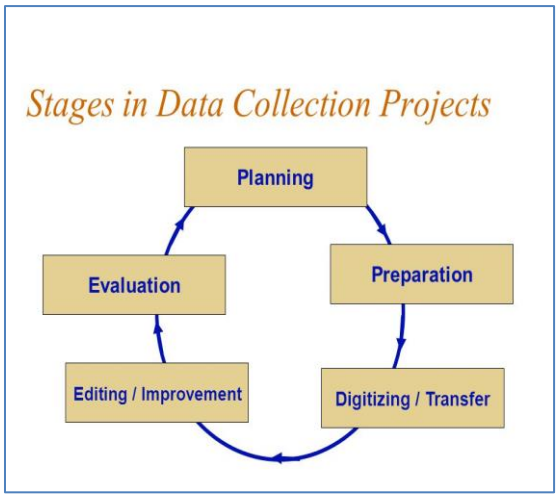
- **Data collection** is split into ***data capture*** (direct data input) and ***data transfer*** (input of data from other systems).
- Two main types of data capture are:

1. ***Primary data sources*** are those collected in *digital* format specifically for use in a GIS project.
2. ***Secondary sources*** are *digital* and *analog* datasets that were originally captured for another purpose and need to be converted into a suitable *digital* format for use in a GIS project.

- This lecture describes the data sources, techniques, and workflows involved in GIS data collection.
- The processes of ***data collection*** are also variously referred to as data capture, data automation, data conversion, data transfer, data translation, and digitizing.
- Data collection is a time consuming, tedious, and expensive process.
- Typically it accounts for 15–50% of the total cost of a GIS project
- If staff costs are excluded from a GIS budget, then in cash expenditure terms data collection can be as much as 60–85% of costs.

Data Collection Workflow

- Figure 1 shows the stages in data collection projects



- ***Planning*** includes establishing user requirements, garnering resources, and developing a project plan.
- ***Preparation*** involves obtaining data, redrafting poor-quality map sources, editing scanned map images, removing

noise, setting up appropriate GIS hardware and software systems to accept data.

- ***Digitizing*** and ***transfer*** are the stages where the majority of the effort will be expended.
- ***Editing*** and ***improvement*** covers many techniques designed to validate data, as well as correct errors and improve quality.
- ***Evaluation*** is the process of identifying project successes and failures.

Primary Geographic Data Capture

Raster Data Capture

- ***Remote sensing*** is a technique used to derive information about the physical, chemical, and biological properties of objects without direct physical contact
- ***Information*** is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects.

The Spatial and Temporal Characteristics of Commonly Used Remote Sensing Systems and Their Sensors

- ***Resolution*** is a key physical characteristic of remote sensing systems.
- ***Spatial resolution*** refers to the size of object that can be resolved and the most usual measure is the pixel size.
- ***Spectral resolution*** refers to the parts of the electromagnetic spectrum that are measured.
- ***Temporal resolution***, or repeat cycle, describes the frequency with which images are collected for the same area.
- ***Aerial photography*** is equally important in medium- to large-scale projects
- ***Photographs*** are normally collected by analog optical cameras and later scanned
- ***Aerial Photographs*** are usually collected on an ad hoc basis.
- Can provide stereo imagery for the extraction of digital elevation models.

Space Born Remote Sensing

- **<u>Advantages are</u>**

- Consistency of the data.
- Availability of systematic global coverage.
- Regular repeat cycles.
- **Disadvantages are**
- Resolution is often too coarse.
- Many sensors are restricted by cloud cover.

Vector Data Capture

- Two main branches are ***ground surveying*** and ***GPS***

Surveying

- ***Ground surveying*** is based on the principle that the 3-D location of any point can be determined by measuring *angles* and *distances* from other known points.
- Traditional equipment like transits and theodolites have been replaced by total stations that can measure both angles and distances to an accuracy of 1 mm
- Ground survey is a very time-consuming and expensive activity, but it is still the best way to obtain highly accurate point locations.
- Typically used for capturing buildings, land and property boundaries, manholes, and other objects that need to be located accurately.
- Also employed to obtain reference marks for use in other data capture projects.

LiDAR

- Relatively new technology that employs a scanning laser rangefinder to produce accurate topographic surveys
- Typically carried on a low-altitude aircraft that also has an inertial navigation system and a differential GPS to provide location.

Secondary Geographic Data Capture

Raster Data Capture Using Scanners

Three main reasons to scan hardcopy media are:

1 Documents are scanned to reduce wear and tear, improve access, provide integrated database storage, and to index them geographically.

2 Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for other data.

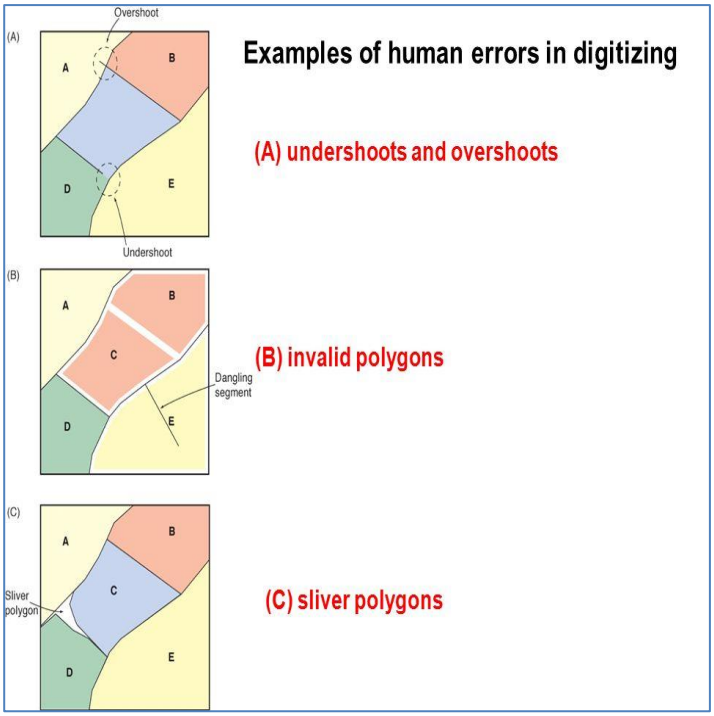
3 Maps, aerial photographs and images are scanned prior to vectorization.

Vector Data Capture

- Secondary vector data capture involves digitizing vector objects from maps and other geographic data sources.

Heads-Up Digitizing and Vectorization

- ***Vectorization*** is the process of converting raster data into vector data.
- The simplest way to create vectors from raster layers is to digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor.
- Describes how automated vectorization is performed.



Measurement Error

- Figure 2 presents some examples of human errors that are commonly introduced in the digitizing procedure including overshoots, undershoots, invalid polygons, and sliver polygons

Photogrammetry

- Is the science and technology of making measurements from pictures, aerial photographs, and images.
- Measurements are captured from overlapping pairs of photographs using stereo plotters.
- ***Orientation*** and ***triangulation*** are fundamental photogrammetry processing tasks.
- ***Orientation*** is the process of creating a stereo model suitable for viewing and

extracting 3-D vector coordinates that describe geographic objects.

- ***Triangulation*** (also called _block adjustment') is used to assemble a collection of images into a single model so that accurate and consistent information can be obtained from large areas. ***Orthoimages*** are images corrected for variations in terrain using a DEM.
- ***Photogrammetry*** is a very cost-effective data capture technique that is sometimes the only practical method of obtaining detailed topographic data.

COGO Data Entry

- ***COGO*** is a contraction of the term coordinate geometry, a methodology for capturing and representing geographic data.
- COGO uses survey-style bearings and distances to define each part of an object
- COGO data are very precise measurements and are often regarded as the only legally acceptable definition of land parcels.

Obtaining Data from External Sources (Data Transfer)

- The best way to find geographic data is to search the Internet.

Capturing attribute data •

- Attributes can be entered by direct data loggers, manual keyboard entry, optical character recognition (OCR) or, increasingly, voice recognition. • An essential requirement for separate data entry is a common identifier (also called a key) that can be used to relate object geometry and attributes together following data capture