Data Entry

Getting coordinates and attributes into our GIS
How we used to collect spatial data
How we collect spatial data now
DATA SOURCES, INPUT, AND OUTPUT

Manually digitizing from image or map sources
  • manually drawn maps
  • legal records
  • coordinate lists with associated tabular data
  • Aerial photographs

Field coordinate measurement
  • Coordinate Surveying
  • GPS

Image data
  • Manual or automated classification
  • direct raster data entry
Field Measurement

Coordinate Surveying

GPS

(courtesy NGS)
Satellite and Aerial Imagery

Image Data

Spatial Data in a GIS
Manual Digitizing

On screen or on a digitizing tablet
Connect the dots
Digitize each point once
Digitize each line segment once
Combine line segments to create area (polygon) features
Minimum Mapping Unit (MMU)
Hardcopy Maps – Fixed on “Permanent” Media
Map Type

Feature map
- Cities
- Highways

Choropleth map
- persons/sq.mi
- 0 - 25
- 26 - 50
- 51 - 100
- 101 - 1000
- > 1000

Dot density map
- Population
- 1 dot = 50,000

Contour map
- Height contours
Map Scale

The ratio of:
a distance on a map
to a distance on the ground

Commonly reported as a:

Unitless ratio: 1:100,000

Unit ratio: e.g., four inches to one mile

Scale bar: 0 50 100 200 Miles
Common Map Scale Confusion

People often say "large scale" map when they mean "large area".

Map scale is a number, a ratio of sizes, that is a fraction, e.g., 1:100,000 scale. As a number this is 0.000001.

A large scale map is one where the fraction is large. This happens when the denominator (bottom number) is small.

Example:

1 to 1 million map scale (1:1,000,000) expressed as a fraction is 0.000001;

a 1:200 map scale, expressed as a fraction, is 0.005.

Which is the larger scale, 0.000001, or 0.005?
If you have two map sheets which are 10 inches across, the 1:1,000,000 map (which is small scale) covers a distance of 10,000,000 inches.

The 1:200 map (large scale) covers about 200*10=2000 inches.

Remember, larger scale maps cover less area, but show more detail.
Which image has the larger scale?
Map Scale

"Scale" of GIS data

Digital spatial data have no scale. An input map had a scale, but many data don't come from maps, e.g., GPS data.

Spatial data in a GIS may be displayed on-screen at a broad ranges of scales – there is no one scale.

We must think of a source scale, if it exists
And
A display scale, the ratio of ground to on screen size
Map Scale
"Scale" of GIS data (there is none)
Determining Scale on Source Materials

If map scale is not available, the best method is to measure paired distances,

e.g., the distance between two road intersections on the map is 4.3 cm,

field measurements between the same two road intersections shows the distance to be 1220 meters. The scale of the map is then

\[
\text{map distance} / \text{ground distance}, \text{ or } \frac{4.3\text{cm} \times 1\text{m}/100\text{cm}}{1220 \text{ m}} = 0.043 / 1,220 = 1 / 28,372
\]
Concept of “grain size” or resolution of spatial data and the extent of spatial data

Grain size is the small thing you can discern and geographically place in the data

Extent is the area covered
Spatial Data Input from Hardcopy Sources

Common Input Methods:

- manual digitizing,
- automatic map scanning,
- text scanning
- format conversion
Manual Digitizing

Tracing the location of "important" coordinates

Done from an image or map source

"There must be an easier way to digitise the site, Cyril!"
Manual Digitization

Connect the dots - lines or points with an electrically sensitized puck.

Wire grid typically used to identify puck location on tablet

Puck location recorded relative to an arbitrary table coordinate system

Points locations are signaled by pressing buttons on the puck

Accuracies of between 0.01 and 0.001 inches
Manual Digitizing Process from digital image:

1. Scan map or image
2. If image not referenced, collect ground coordinates of control points
3. Digitize control points (tics, reference points, etc.) of known location
4. Transform (register) image to known coordinate system
5. Digitize feature boundaries in stream or point mode
6. Proof, edit linework
7. Re-edit, as necessary
Digitize Primarily from Cartometric Images

Taken with specialized cameras, from specialized planes

Control based on coordinate surveys
Digitize Primarily from Cartometric Maps

- Based on coordinate surveys
- Plotted and printed carefully
Manual Map Digitization, Pros and Cons

Advantages
- low cost
- poor quality maps (much editing, interpretation)
- short training intervals
- ease in frequent quality testing
- device ubiquity

Disadvantages
- upper limit on precision
- poor quality maps (much editing, interpretation)
- short training intervals
- ease in frequent quality testing
- device ubiquity
DATA SOURCES, INPUT, AND OUTPUT

Problems with source maps:

- Dimensional stability (shrink, swell, folds)
- Boundary or tiling problems
- Maps are abstractions of Reality
- Features are generalized:
  - classified (e.g., not all wetlands are alike)
  - simplified (lakes, streams, and towns in a scale example)
  - moved (offsets in plotting)
  - exaggerated (buildings, line roadwidths, etc).
Map Generalization: An Example

1:24,000 scale map  1:250,000 scale map
Types of Map Generalization

Truth

Grouped

Exaggerated
- True Scale Road Width
- Standard Symbol Road Width

Offset

Omitted

Categorized
- water
- marsh
- fen
Look for generalizations
Look for generalizations
Manual Digitizing

- nodes at line endpoints
- vertices define line shape
Manual Digitizing
common errors that require editing

- undershoot
- dangling node
- pseudonode
- polygon label point
- missing label point
- line crossing - no node
- overshoot
Editing

**Manual editing:**
Line and point locations are adjusted on a graphic display, pointing and clicking with a mouse or keyboard. Most controlled, most time-consuming.

**Interactive rubbersheeting:**
Anchor points are selected, again on the graphics screen, and other points selected and dragged around the screen. All lines and points except the anchor points are interactively adjusted.
Manual editing - removing a feature
Manual editing - modifying an edge
Editing

Attribute consistency analysis:
Identify contradictory theme types in different data layers, and resolve

Line snapping:
When a vertex or node is “close” to a line or end point, the lines are “snapped” together

Point snapping:
Points which fall within a specified distance of each other are snapped (typically, on point eliminated).
Snapping

before snapping:

undershoots

snap tolerance

overshoot

after snapping:
Manual Digitizing – Vertex Density

Too few vertices

Too many vertices
To Few Vertices – Spline Interpolation

Create smooth, curving lines by fitting piecewise polynomial functions.

One polynomial equation is fit to points one through five, \( y = f(x) \)

At the join point \( f(x) = g(x) \)
Slope of \( f(x) \) = slope of \( g(x) \)
Change in slope of \( f(x) \) equals the change in slope of \( g(x) \)

A second polynomial equation is fit to points five through eleven, \( y = g(x) \)
Common problem:

Features which occur on several different maps rarely have the same position on each map.

What to do?

1. **Re-drafting the data** from conflicting sources onto the same base map, or
2. **Establish a "master" boundary**, and redraft map or copy after digitizing.
Digitizing Maps - Automated Scanners

• Main alternative to manual digitizing for hardcopy maps

• Range of scanner qualities, geometric fidelity should be verified

• Most maps are now available digitally – however many began life as paper maps
Digitizing Maps - Automated Scanners

- Suitable thresholding allows determination of line or point features from the hardcopy map.

- Scanners work best when very clean map materials are available.

- Significant editing still required (thinning, removing unwanted features)
Cell Thinning and Vectorizing—After Scan-Digitizing

before thinning

after thinning

Raster

Cell center points

Smoothed line
Map Data Entry Summary

- Manual digitizing and scanning,
- Use cartometric maps at correct scale
- Ensure appropriate map type, generalization
- Line/node digitizing, with snapping, thinning, and editing