Getting the Map into the Computer: GIS Data Development

- 4.1 Analog-to-Digital Maps
- 4.2 Finding Existing Map Data
- 4.3 Digitizing and Scanning
- 4.4 Field and Image Data
- 4.5 Data Entry
- 4.6 Editing and Validation
GIS maps are digital not analog

• Maps have a communications function but...

• A map has a storage function for spatial data

• Somehow, the visually “stored” data must get digital

• Real and Virtual maps
GIS Data Conversion

• Traditionally most of the cost of a GIS project
• This is a one-time cost
• Depends on reuse
• Requires maintenance
GIS Data Development Approaches

- Research for **existing** data
- If analog maps exist, **convert analog maps to virtual maps** (GIS data)
- If no analog maps, **aerial photography, remote sensing, GPS** and **ground survey** usually are obtained.
Finding Existing Map Data

- Map libraries
- Reference books
- State and local agencies
- Federal agencies
- Commercial data suppliers e.g. GDT, Thompson, ETAK
Existing Map Data

- Existing map data can be found through a map library, via network searches, or on media such as CD-ROM and disk.
- Many major data providers make their data available via the World Wide Web, a network of file servers available over the Internet.
- GIS vendors package data with products.
Commercial vendors
Commercial vendors

- www.navteq.com
- www.teleatalas.com
- www.geographynetwork.com
- www.google.com
- www.mapquest.com
Federal Data Agencies

- USGS
- NOAA
- Census Bureau
- NIMA
- EPA
- many more...
National Spatial Data Infrastructure

Consistent means to share geographic data among all users could produce significant savings for data collection and use and enhance decision making. Executive Order 12296 calls for the establishment of the National Spatial Data Infrastructure defined as the technologies, policies, and people necessary to promote sharing of geospatial data throughout all levels of government, the private and non-profit sectors, and the academic community.

September 2, 1998
Clinton Administration Initiatives to Foster the NSDI

In a major address at the Brookings Institution in Washington, D.C., Vice President Gore called for stronger efforts nationwide to enhance the livability and economic competitiveness of American communities. The Vice President said the Administration will expand its support for the use of geographic information systems technologies and encourage increased public access and sharing of geographic data to put “more control, more information, more decision-making power into the hands of families, communities, and regions – to give them all the freedom and flexibility they need to reclaim their own unique place in the world.”

Full text of the Vice President’s speech
White House Press Release
FGDC Press Release
Community Federal Information Partnership
Community Demonstration Projects

A Strategy for the NSDI

This strategy updates the 1994 Strategic Plan for the NSDI. It was developed with broad input from stakeholders in the geospatial data community, both within and outside of the federal government.
National Spatial Data Clearinghouse
USGS: National Mapping
U.S. Bureau of the Census
The Decennial National Census

• This is a survey based, national-scale collection of data, conducted by the U.S. Census Bureau (a Federal agency) every 10 years

• This provides a snapshot of a wide range of socio-economic data for the entire nation, that can be used in two ways:

  • Comparisons can be made to conditions as captured by a previous census so that change in time can be studied AND
  • The data can be analyzed spatially, using levels of geography from local to national scales
Census 2000
Short Form and Long Form

Short form

Long form

The "Informational Copy" shows the content of the United States Census 2000 "long" form questionnaire. Each household will receive either a short form (100-percent questions) or a long form (100-percent and sample questions). The long form questionnaire includes the same 5 population questions and...

Please do not fill out this form. This is not an official census form. It is for informational purposes only.

United States Census 2000

This is the official form for all the people at this address. It is quick and easy, and your answers are protected by law. Complete the Census and help your community get what it needs — today and in the future!
Census Data as Spatial Data

• The information collected on the surveys form the attribute part of the data

• Individual surveys are aggregated spatially to geographic units, at the various levels of census geography, that uses a nested scheme that builds higher-level units out of lower ones
Small-Area Geography Overview

### Census Small-Area Geography

| Understanding the Relationships Among U.S. Census Bureau Geographic Entities |
|---|---|---|
| County | Minor Civil Division (MCD) or Census County Division (CCD) | Place |
| Smith County | Jones Township | Green City |
| Block 3014 | Block Group 3 | Census Tract 5.02 |
| Pinney Hollow Road | 3014 | 1001 |
| 3011 | 3012 | 1002 |
| 3013 | 1004 | 2002 |
| 2003 | 2004 | 2005 |
| 2006 | 2007 | 2008 |

David Tenenbaum – EEOS 265 – UMB Fall 2008
Geographic Products: The TIGER Data Base

Topologically Integrated Geographic Encoding & Referencing

The source of ALL census geographic products
Basic TIGER/Line File Topology

One census block:
3 GT-polygons
1 point landmark
(school)
1 area landmark
(park)
NOAA Weather and other data
Eros Data Center

• Distributed active archive center
• Sioux Falls, SD
• Operated by USGS
US GeoData

ftp access to

DEM

DLG

GNIS

GIRAS

e tc.
GNIS Feature locations
GIRAS
Land Use and Land Cover Data
GEOCODING

• Geocoding is the **conversion of spatial information into digital form**

• Geocoding involves **capturing the map**, and sometimes also **capturing the attributes**

• Often involves **address matching**
Address Matching

• Most GISs contain capability for address matching
• Start with 100 Morrissey Boulevard, Boston, MA 02125
• End with coordinates (lat-long for UMB)
• May need to interpolate along blocks
• Street number range, left and right side e.g. 101-199
TIGER/Line Address Range Basics

The complete chain has two address ranges: left: odd-numbered right: even-numbered.

Potential address ranges along a complete chain have values that encompass the addresses of existing structures, as well as those not yet built.
GEOCODING LEAVES A “STAMP” ON DATA

• The **method** of geocoding can influence the **structure and error** associated with the spatial information which results

• **Examples**: scanning (raster), digitizing (vector)
Geocoding methods for maps

• Digitizing
• Scanning
• Field data collection
Digitizing

• Captures map data by tracing lines from a map by hand

• Uses a cursor and an electronically-sensitive tablet

• Result is a string of points with (x, y) values
The Digitizing Tablet

1. Digitizer cursor transmits a pulse from an electromagnetic coil under the view lens.

2. Pulse is picked up by nearest grid wires under tablet surface.

3. Result is sent to computer after conversion to x and y units.
Digitizing

- Stable base map
- Fix to tablet
- Digitize control
- Determine coordinate transformation
- Trace features
- Proof plot
- Edit
- Clean and build
Digitizing

- Cursor data entry
- Secondary tablet (menu/template)
- Voice command entry
- Point select
- Stream mode
- Distance mode
Selecting points to digitize
Some common digitizing errors

- Slivers
- Duplicate lines
- Duplicate nodes
- Unended lines
- Gaps
- Zingers
Scanning

• Places a map on a glass plate, and passes a light beam over it
• Measures the reflected light intensity
• Result is a grid of pixels
• Image size and resolution are important
• Features can “drop out”
Scanning

- **Types** of scanners:
  - Flat bed vs. Drum

- **Characteristics**
  - DPI & File size
Scanning example

- 15 x 15 cm (3.6 x 3.6 km)
- grid is 0.25 mm
- ground equivalent is 6 m
- 600 x 600 pixels
- one byte per color (0-255)
- 1.08 MB
Field data collection
Global Positioning System (GPS)

- A space-based 3-dimensional measurement and positioning system that operates using radio signals from satellites orbiting the earth.
- Created and maintained by the US Dept. of Defense and the US Air Force.
- The system as a whole consists of three segments:
  - satellites (space segment)
  - receivers (user segment)
  - ground stations (control segment)
- Note: Russia and a European consortium are implementing similar systems.
GPS – Space Segment (Satellites)

• 24 NAVSTAR satellites in the GPS constellation
  • orbit the Earth every 12 hours
  • ~11,000 miles altitude (a very high orbit)
  • positioned in 6 orbital planes (4 per plane)
  • orbital period & planes are designed to keep 4-6 satellites above the horizon at any time everywhere on the planet
  • controlled and monitored by five ground stations around the globe
GPS Nominal Constellation
24 Satellites in 6 Orbital Planes
4 Satellites in each Plane
20,200 km Altitudes, 55 Degree Inclination
GPS – User Segment (Receivers)

• Ground-based devices that can read and interpret the radio signals from several of the NAVSTAR satellites at once

• Use timing of radio signals to calculate the receiver’s position on the Earth's surface

• Calculations result in varying degrees of accuracy that depend on:
  • quality of the receiver
  • user operation of the receiver
  • local & atmospheric conditions
  • current status of system
GPS – Control Segment (Ground Stations)

- Five control stations
  - master station at Falcon (Schriever) AFB, Colorado
  - monitor satellite orbits & clocks
  - broadcast orbital data and clock corrections to satellites
GPS - How Does it Work?

• GPS allows us to determine a position by calculating the **distance** between a receiver and multiple satellites
  • Distance is determined by **timing** how long it takes the signal to travel from satellite to receiver
  • Radio signals travel at **speed of light**: 186,000 mi / sec
  • Satellites and receivers generate **exactly the same signal** at exactly the same time
  • Signal travel time = delay of satellite signal relative to the receiver signal

![Diagram of GPS signal timing](image)
GPS - Satellite Signals

• Satellites have accurate **atomic clocks** and all 24 satellites are transmitting the same time signal at the same time

• The satellite signals contains information that includes
  – Satellite number
  – Time of transmission

• Receivers use an **almanac** that includes
  – The **position** of all satellites every second
  – This is updated monthly from control stations

• The satellite signal is received, compared with the receiver’s internal clock, and used to calculate the **distance** from that satellite

• **Trilateration** (similar to triangulation) is used to determine location from multiple satellite signals
Start by determining the distance between a single GPS satellite and your position (a sphere).

Adding a second distance measurement to another satellite narrows down your possible positions to a circle where the spheres intersect.
Adding a third satellite narrows down the position to two points where the three sphere intersect, and usually only one point is a ‘reasonable’ answer.

The intersection of four spheres occurs at one point, but the 4th measurement is not needed, and is used for timing purposes instead.
GPS - Using the 4th Signal

• How do we know that satellites and receivers generate the same signal at the same time?
  • The satellites have atomic clocks, so we know they are accurate
  • But receivers do not -- so can we ensure they are exactly accurate? No! If the receiver's timing is off, the location in 3-D space will be off slightly...

• So: We use the 4th satellite to resolve any signal timing error instead by
  • determining a correction factor using the 4th satellite
  • (like solving multiple equations ... there will only be one solution that satisfies all equations)
GPS - Sources of Error

- Satellite errors
  - satellite position / orbit error
  - satellite clock error
- Atmospheric errors
  - Speed of electromagnetic waves in the atmosphere
  - Path taken by the signal
- Multi-path distortion errors
- Receiver errors
- (Selective availability)
GPS - Sources of Error

• Satellite Errors
  – Although the satellites are in high orbits to minimize their deviations, sometimes there is a slight ‘wobble’ due to local gravitational forces
  – While the atomic clocks used in the satellites are extremely accurate (and quadruple redundant), sometimes clock errors can occur

• These can contribute up to 1-5 meters of error
GPS - Sources of Error

• Atmospheric Delays/Bending
  – The speed of light is only precisely 186,000 miles per second in a vacuum, and is slightly \textit{slower in the atmosphere}, varying by composition
  – The signal can be \textit{bent} as it moves through the atmosphere (sphere size based on a straight path)

• \textbf{Up to 30m} of error
GPS - Sources of Error

- Multi Path Interference
  - The signal can **bounce off** of buildings, trees, etc. and this again distorts the time and distance between the receiver and the satellite

- **Up to 1m** of error
GPS - Sources of Error

• Receiver Errors (Timing/Rounding)
  – Satellites have quadruple redundant atomic clocks that are **accurate to nanoseconds** (about $800,000 of clock hardware on each satellite), e.g. “the time is 2:02:01.23456789012”
  – Receivers are powered by 4 AA batteries worth about $2.99, generate their clock signal with an oscillating crystal that is sensitive to battery current, e.g. “the time is 2:02:01.2345”

• **Up to 10 meters** of error
GPS - Sources of Error

- Satellite Coverage in Sky
  - Position Dilution of Precision (PDOP)
The Role of Error

• **Enforcement** for map data is usually by using topology

• Map and attribute data errors are the **data producer's responsibility**, but the GIS user must understand error

• **Accuracy and precision** of map and attribute data in a GIS affect all other operations, especially when maps are compared across scales
Precision and Accuracy

• When describing error we need to distinguish between **two characteristics**:  
  – **Accuracy** refers to the **amount of distortion** from the true value in a measurement  
  – Precision refers to the **variation among repeated measurements**, and also to the **amount of detail** in the reporting of a measurement
Precision and Accuracy

• These related concepts are often confused:
  • **Precision** refers to the exactness associated with a measurement (i.e. closely clustered)
  • **Accuracy** refers to the extent of systematic bias in the measurement process (i.e. centered on the middle)
Getting the Map into the Computer: GIS Data Development

- 4.1 Analog-to-Digital Maps
- 4.2 Finding Existing Map Data
- 4.3 Digitizing and Scanning
- 4.4 Field and Image Data
- 4.5 Data Entry
- 4.6 Editing and Validation
Next Topic:

What is where?