Two sources
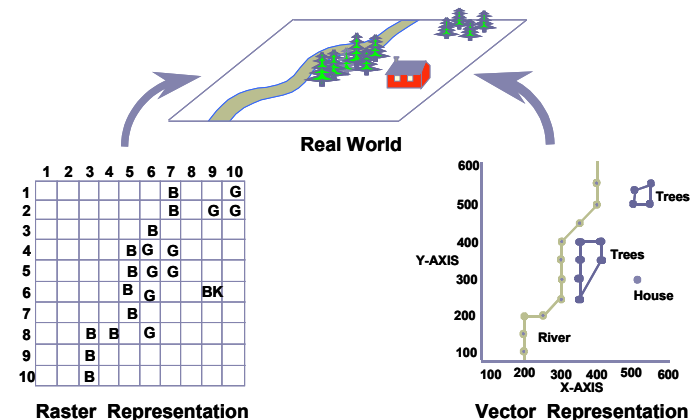
- ❑ Primary Data Capture (first-hand collection)
  - ◼ Digitizing
    - ❑ Tablet digitizing
    - ❑ Heads up digitizing
    - ❑ Automatic digitizing
  - ◼ Scanning
  - ◼ Other point measurements (in text files)
  - ◼ Census data
  - ◼ GPS collections (will be taught later)
  - ◼ Aerial photographs (will be taught later)
  - ◼ Remote sensing data (will be taught later)

- ❑ Secondary Data Capture (from others)
  - ◼ Published or released data (originally primary data)
  - ◼ All primary data from others are secondary data for you and me

Digitizing

- Often, digital data for a GIS project is not available, so it must be created from other existing sources like paper maps

- **Digitizing** is the process where features on a map or image are converted into digital format for use by a GIS.

- Digitizing converts the features on the map into three basic data types:
  - Points – zero dimensional objects
  - Lines – one dimensional objects
  - Polygons – two dimensional objects

Raster Representation   Vector Representation

©Arthur J. Lembo
Cornell University

Source: Defense Mapping School
National Imagery and Mapping Agency

Digitizing

□ There are three primary methods for digitizing spatial information:
- Manual Methods include:
  - Tablet Digitizing
  - Heads-up Digitizing
- An Automated Method includes:
  - Scanning and Vectorization

Tablet Digitizing

- Tablet digitizing requires a person to enter coordinate information through the use of a *digitizing tablet* and *digitizing puck*

    - A **digitizing tablet** is a hardened surface with a fine electrical wire grid under the surface.

    - Digitizing tablets are either hardened, more stationary tables or rollup boards designed for portability

    - A **digitizing puck** is an electrical device with cross hairs and multiple buttons to perform data entry operations

    - An operator then enters the information using the puck.

©Arthur J. Lembo
Cornell University

Tablet Digitizing – how it works

- When the user places the digitizing puck over a location on the tablet, and presses one of the buttons, the wire mesh beneath the tablet records the location of the puck

- Digitizing tablets are very accurate, with more expensive tablets able to measure objects to within 0.006 mm. This means that if you were to press the entry button on the puck continuously at one spot, the coordinate value received from the tablet would only vary by 0.006 mm.

- The coordinate, as referenced by the tablet is then stored in the computer.

x,y

©Arthur J. Lembo
Cornell University
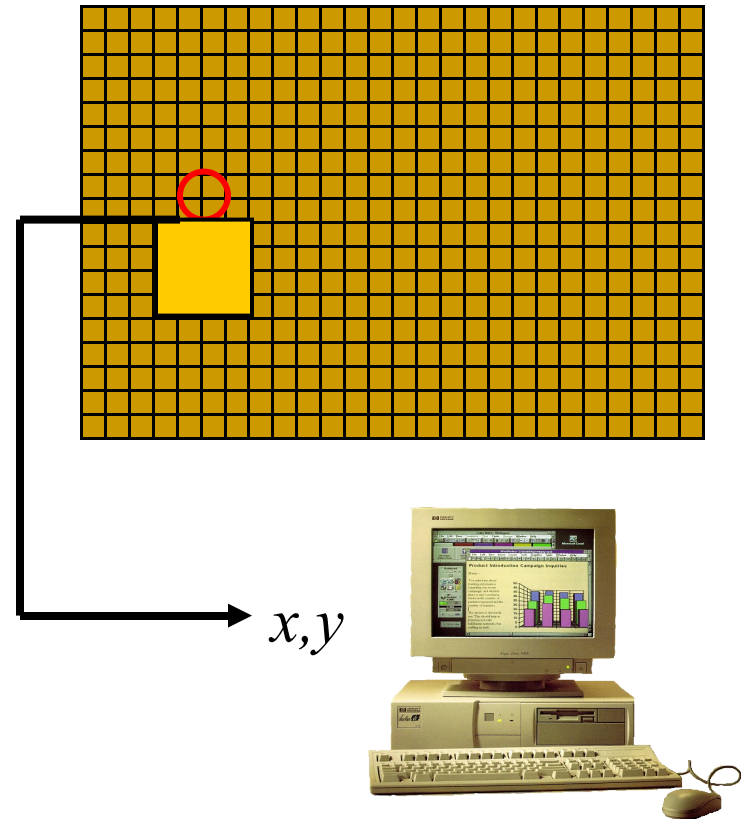
Tablet Digitizing – the problems

☐ **Tablet digitizing is not without its problems**:

- ■ Tablet digitizing is tedious, and "operator fatigue" is very common. When digitizing for 8 hours a day, one should take plenty of breaks! As an operator gets tired, he may also tend to get careless.

- ■ Due to environmental conditions (such as humidity), the source materials may actually change (due to shrinking and expansion). This may not seem like a big deal, but consider that on a 1:100,000 scale map, a slight shift in the map on the tablet (if you accidentally bump or move the map), or a slight shrinking of the map of only 0.1 inches would equal an 800 foot shift in the real world. The high accuracy is only achieved with more expensive digitizing tablets.

☐ If you notice the person on the right, the actual map is usually a foot or two away from the operator. So, small subtitles in the geographic objects may be missed when you are so far away ©Arthur J. Lembo Cornell University

Tablet Digitizing – why bother?

❑ Digitizing objects on a map (like a river) requires that every bend and turn be captured using the puck. Therefore, digitizing is very labor intensive. So, you would be well within your rights to ask "why bother"? Here are just two reasons:

  ▪ Manual digitizing is quite accurate when performed with care.

  ▪ It is also the most common method for entering spatial information into a GIS.

❑ We are probably seeing a trend away from tablet digitizing toward heads-up digitizing or raster to vector conversion, especially with the reduction in price for scanners. But for now, tablet digitizing is one of the most effective methods for entering spatial information.

Digitizing – heads up digitizing

- ❑ Heads up digitizing is a combination of scanning and manual digitizing.
- ❑ The main steps in heads up digitizing typically include:
  - ▪ **Scanning the map** – a user can scan the map at a high resolution.
  - ▪ **Registering the map** – using the same transformation methods previously discussed, the user can enter control points on screen and transform the scanned image to real world coordinates.
  - ▪ **Digitizing the map** – the user can zoom to specific areas on screen and trace points, lines, or polygons on the map. Because the maps are already in the correct geographic coordinate system anything digitized on top of the map will also be in the correct coordinate system.

Digitizing – how its done

- Although methods vary, the typical steps for digitizing a paper map require the following:
  - Physical Preparation
    - Digitizer Set up
    - Map preparation
  - Digital Preparation
    - Map registration
  - Digital Collection
    - Feature collection
    - Feature correction
    - Save, save, save…

Digitizing - registration

□ The first step in registering a map requires selecting control points (locations on the map that have real world coordinates assigned to them. You may hear people call these "tic" marks. Good tic marks may be the grid tics on a USGS quad sheet, or you can create your own. But they should have a "bullseye" shape to it, and associated real world coordinates.

□ The user will place the puck over the tic marks, and press the button to record the tablet coordinates. The user then enters the real world coordinates by hand. After control points are selected, a match between the tablet coordinates and the real world coordinates are established and a mathematical relationship is formed.

□ The tablet coordinates are then transformed to real world coordinates through the mathematical relationship
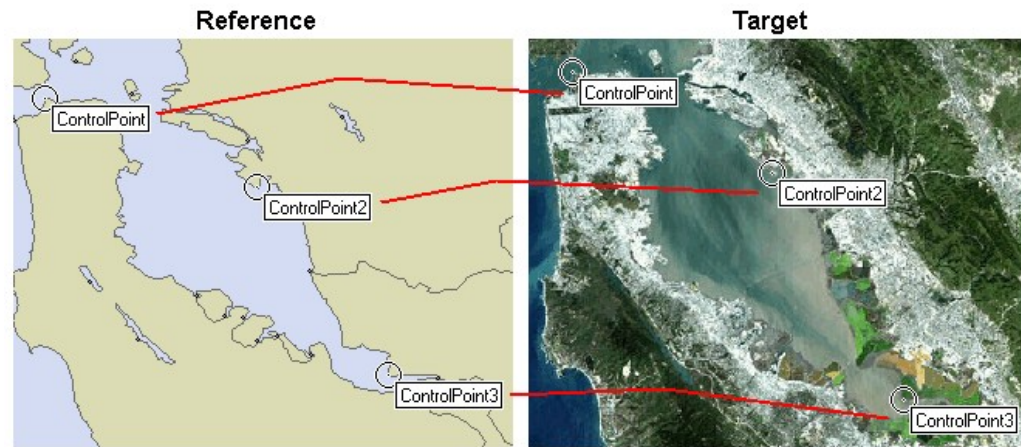
© Manifold

Digitizing - transformation

- Transforming paper coordinates to real world coordinates typically uses a polynomial transformation that models a rotation, translation, and scale change.

- That is, the paper coordinates have to be shifted (**translation**) both in the x direction and the y direction. Also, the map may not be oriented North/South. So, the map will have to be **rotated**. Finally, the tablet coordinates are in mm, and the real world coordinates may be in meters. So, the entire coordinate system has to be enlarged (**scaling**).

- A common transformation is the conformal coordinate transformation that uses the 4 transformation parameters we just spoke of (rotation, scale, translation of the x coordinate, translation of the y coordinate) to "map" the paper coordinates to the real world coordinates.

- After the coordinates are transformed, the transformed coordinates are compared against the original coordinates. The differences between the transformed coordinates and real coordinates represent the error. Lets look at a mathematical example

- Shift   - Rotation   - Scaling
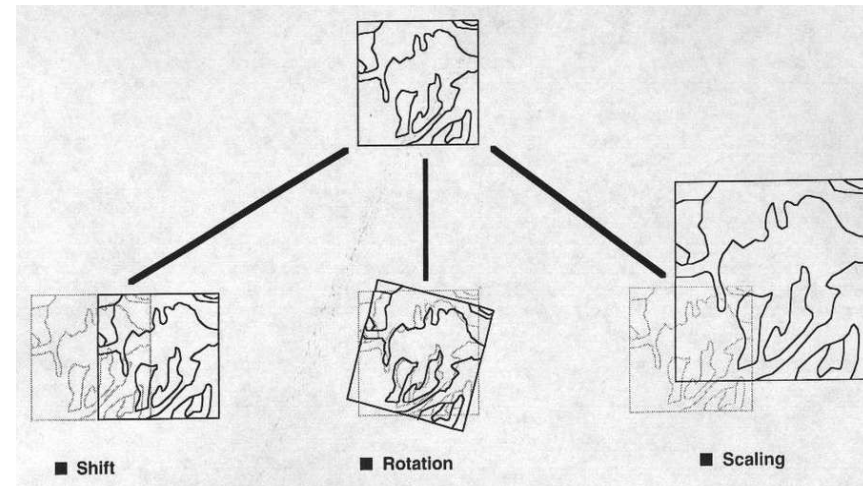
© ESRI

Automated Digitizing

- Tools to automatically convert a raster scan to vector lines
- Requires a very clean scan
- Scans can be cleaned using raster cleanup tools
- The vector files usually require cleanup after conversion
- If you start with a clean image it can save a lot of time
- If you image is not clean manual digitizing may be faster

Automated Tools

- Gridline
  - Command line ArcInfo command
  - Need a near prefect image
  - Can have a large amount of cleanup after conversion
- ArcScan Extension in ArcGIS
  - New extension since 8.3
  - Has clean up tools
  - First tools for ArcGIS that work well

The ArcScan toolbar

Raster Snapping Options button: Opens the Raster Snapping Options dialog box, which supports raster snapping settings.

Cell Selection menu and tools: Support tools and commands for selecting connected raster cells.

Raster target layer list: Sets the raster layer for selection, cleanup, and vectorization operations.

Vectorization Trace tool: This tool allows you to perform manual raster tracing.

Raster Line Width tool: This tool allows you to view the raster line width for connected cells.

ArcScan

Raster: 535p0704.tif ▼   Vectorization ▼   Raster Cleanup ▼   Cell Selection ▼

535p0704.tif
535p0703.tif
535p0702.tif
535p0701.tif
535p0604.tif
535p0603.tif
535p0602.tif
535p0601.tif

Vectorization Settings…
Show Preview
Generate Features…
Options…

Select Connected Cells…
Interactive Selection Target ▶   ✓ Foreground Cells
Clear Selected Cells   Background Cells
Save Selection As…

Vectorization menu: Supports batch vectorization tools and commands, such as vectorization settings, preview, and generate features.

Start Cleanup
Stop Cleanup
Raster Painting Toolbar
Erase Selected Cells
Fill Selected Cells
Save
Save As…

Raster Painting

Raster Cleanup menu: Supports tools and commands for editing and exporting raster data.

Raster Painting toolbar: Supports tools that allow you to perform raster edits, such as brush, fill, draw, and erase.

Secondary Data

- Large amount of data is now available
- Always check for existing data before creating it
- Several groups of data exist
  - Free data from the government
  - Government data available for a fee
  - Internet map servers
  - Commercial data
  - Data from other GIS users

Free Government Data

□ Large amounts of data are available from
   ■ USGS
   ■ Census
   ■ NOAA

USGS Data

- DEM = Digital Elevation Model
- NED = National Elevation Dataset
- NHD = National Hydrography Dataset
- DRG = Digital Raster Graphic
- DLG = Digital Line Graph
- DOQQ = Digital Ortho Quarter Quad
- GNIS = Geographic Names Information System
- LULC = Land Use Land Cover
- NLCD = National Land Cover Data

DEM

- Available in Several Resolutions
  - 3-arc-second
    - 1:250,000 Intended scale
    - Approximately 90m Cell size
  - 2-arc-second
    - 35% of US coverage complete
    - 100% for New Mexico
    - 1:100,000 intended scale
    - Approximately 60m cell size
- 30m DEM
  - 1:24,000 intended scale
  - Production stopped in favor of 10m DEMs
- 10m DEM
  - 1:24,000 intended scale

NED

- Created from the 30m and 10m DEMs
- No Gaps, Units are standardized
- 100% coverage for the US

NHD http://nhd.usgs.gov

- ◻ National Hydrography Data Set
- ◻ Contains information about naturally occurring and constructed bodies of water
- ◻ Has links to EPA RC3 river reach files
- ◻ Has links to GNIS
- ◻ Is setup for network modeling
- ◻ Two resolutions are available
  - ◼ 24k In progress
  - ◼ 100k Complete for New Mexico
- ◻ Files are being reformatted from Coverages to Geodatabases

DRG

- ❑ Scans of the USGS Topographic Maps
- ❑ Scanned at 250 dpi or greater and resampled back to 250 dpi
- ❑ Earlier maps may be at a lower resolution
- ❑ Useful as base maps

DLG

□ http://edc.usgs.gov/products/map/dlg.html

Layer	Feature Type
Public Land Survey System (PLSS)	Township, range, and section lines
Boundaries	State, county, city, and other national and State lands such as forests and parks
Transportation	Roads and trails, railroads, pipelines and transmission lines
Hydrography	Flowing water, standing water, and wetlands
Hypsography	Contours and supplementary spot elevations
Non-vegetative features	Glacial moraine, lava, sand, and gravel
Survey control and markers	Horizontal and vertical monuments (third order or better)
Manmade features	Cultural features, such as building, not collected in other data categories
Vegetative surface cover	Woods, scrub, orchards, and vineyards

GNIS http://gnis.usgs.gov

❑ Geographic Names Information System

About GNIS

The Geographic Names Information System (GNIS), developed by the USGS in cooperation with the U.S. Board on Geographic Names (BGN), contains information about almost 2 million physical and cultural geographic features in the United States and its territories. The Federally recognized name of each feature described in the data base is identified, and references are made to a feature's location by State, county, and geographic coordinates. The GNIS is our Nation's official repository of **domestic** geographic names information.

The GNIS also contains information on geographic names in Antarctica which are approved by the BGN for use by the Federal Government.

Information about **foreign** geographic feature names can be obtained from the GEOnet Names Server (GNS), developed and maintained by the National Imagery and Mapping Agency.

LULC

□ http://edc.usgs.gov/products/landcover/lulc.html

Product Description

Land Use and Land Cover (LULC) data consists of historical land use and land cover classification data that was based primarily on the manual interpretation of 1970's and 1980's aerial photography. Secondary sources included land use maps and surveys. There are 21 possible categories of cover type. Along with the LULC files, associated maps are included which provide additional information on political units, hydrologic units, census county subdivisions, and Federal and State land ownership.

LULC data is available for the conterminous U.S. and Hawaii, but coverage is not complete for all areas. The data is based on 1:100,000- and 1:250,000-scale USGS topographic quadrangles.

All LULC files are cast to the Universal Transverse Mercator (UTM) projection, and referenced to the North American Datum of 1983 (NAD83). The files are available in GIRAS (Geographic Information Retrieval and Analysis System) or CTG (Composite Theme Grid) format.

The spatial resolution for all LULC files will depend on the format and feature type. Files in GIRAS format will have a minimum polygon area of 10 acres (4 hectares) with a minimum width of 660 feet (200 meters) for manmade features. Non-urban or natural features have a minimum polygon area of 40 acres (16 hectares) with a minimum width of 1320 feet (400 meters). Files in CTG format will have a resolution of 30 meters.

All LULC data is available via FTP download only. LULC data in ArcInfo export format is also available from the EPA (1:250,000 scale only).

NLCD

□ http://landcover.usgs.gov/prodescription.asp

Product Description:
Derived from the early to mid-1990s Landsat Thematic Mapper satellite data, the National Land Cover Data (NLCD) is a 21-class land cover classification scheme applied consistently over the United States. The spatial resolution of the data is 30 meters and mapped in the Albers Conic Equal Area projection, NAD 83. The NLCD are provided on a state-by-state basis. The state data sets were cut out from larger "regional" data sets that are mosaics of Landsat TM scenes. At this time, all of the NLCD state files are available for free download as 8-bit binary files and some states are also available on CD-ROM as a Geo-TIFF.

The TM multi-band mosaics were processed using an unsupervised clustering algorithm. Both leaves-off and leaves-on data sets were analyzed. The resulting clusters were then labeled using aerial photography and ground observations. Clusters that represented more than one land cover category were also identified and, using various ancillary data sets, models developed to split the confused clusters into the correct land cover categories. Click here for a more detailed discussion of the mapping and classification procedures. A hierarchical land cover classification scheme of 21 classes was developed and applied in a consistent manner across the entire United States. The classification scheme follows:

21-Class National Land Cover Data Key:

NOTE - All Classes May NOT Be Represented in a specific state data set. The class number represents the digital value of the class in the data set (for a detailed description of each class, see any state "readme" file).

TIGER

- **Topologically Integrated Geographic Encoding and Referencing**
- http://www.census.gov/geo/www/tiger/overview.l
- is the name for the system and digital database developed at the U.S. Census Bureau to support its mapping needs for the Decennial Census and other Bureau programs.
- The design of the TIGER® database adapts the theories of topology, graph theory, and associated fields of mathematics to provide a disciplined, mathematical description for the geographic structure of the United States and its territories. The topological structure of the TIGER® data base defines the location and relationship of streets, rivers, railroads, and other features to each other and to the numerous geographic entities for which the Census Bureau tabulates data from its censuses and sample surveys.

  It is designed to assure no duplication of these features or areas.

NOAA

- ☐ Shoreline Data
  - ▪ http://www.csc.noaa.gov/shoreline/data.html
- ☐ Estuarine Bathymetry
  - ▪ http://sposerver.nos.noaa.gov/bathy/
- ☐ **NCDC = National Climatic Data Center**
  - ▪ **NCDC is the world's largest active archive of weather data**
  - ▪ **http://www.ncdc.noaa.gov/oa/ncdc.html**

NCDC

- NCDC archives 99 percent of all NOAA data, including over 320 million paper records; 2.5 million microfiche records; over 1.2 petabytes of digital data residing in a mass storage environment. NCDC has satellite weather images back to 1960. NCDC annually publishes over 1.2 million copies of climate publications that are sent to individual users and 33,000 subscribers. NCDC maintains over 500 digital data sets, receives almost 2,000,000 requests each year, and records over 100 million hits per year on the website.

Remote sensing data

- NASA (USGS) EOS datagate way
  - http://edcimswww.cr.usgs.gov/pub/imswelcome/
- NASA lunar and planetary science data
  - http://nssdc.gsfc.nasa.gov/planetary/planets/mar
- ESA (European Space Agency) Planetary Science Archive
  - http://www.rssd.esa.int/index.php?project=PSA

Other Important GIS Data Websites

- http://seamless.usgs.gov/viewer.htm
- http://gisdata.usgs.gov
- http://edcsns17.cr.usgs.gov/EarthExplorer/
- http://data.geocomm.com/
- http://www.texasview.org/
- http://www.tnris.state.tx.us/