GEOGRAPHIC INFORMATION SYSTEMS

Getting the Map into the Computer

Getting Started with
Geographic Information Systems
Chapter 4

D. Z. Seker

ITU Photogrammetry Division

Getting the Map into the Computer

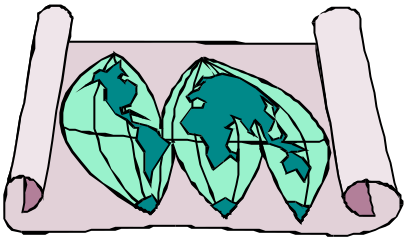
Analog-to-Digital Maps
Finding Existing Map Data
Digitizing and Scanning
Field and Image Data
Data Entry
Editing and Validation

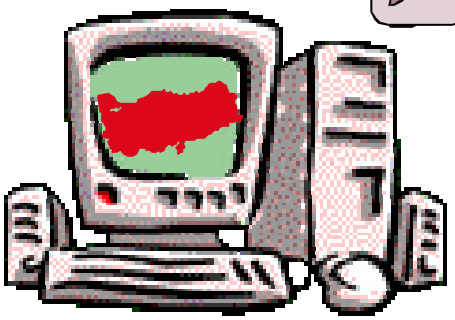
ITU Photogrammetry Division

GIS maps are digital not analog

- Maps have a communications function
- A map has a storage function for spatial data
- Somehow, the visually “stored” data must get digital
- Real and Virtual maps

real maps





virtual maps

a paper or analog form -to- a digital or number form

GIS Data Conversion

- Traditionally most of the cost of a GIS project, finding the right maps, and converting these maps from real to virtual form by geocoding, takes up anywhere between 60 and 90% of both the time and money spent on a typical GIS project.
- One time cost
- Depends on reuse, the map in digital form, we can use it in a GIS over and over again for different uses and projects unless it needs an update.
- Requires maintenance

Geocoding: the conversion of spatial information into computer-readable form

Finding Existing Map Data

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

The *trick* knows: where to look, what to do, when you find, what you want, and how to get the data into your GIS.

Finding Existing Map Data

- map library
- computer networks
- own map division — a state highway authority, park service, or industrial development organisation
- commercial companies

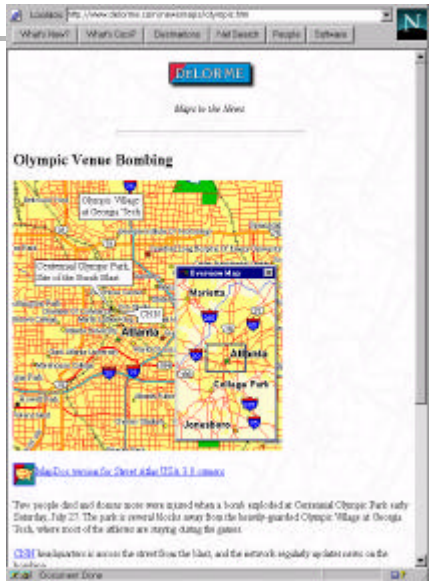
Finding Existing Map Data

- 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.

Finding Existing Map Data

In the United States, digital map data created by the federal government for its own use are the property of the people, with the obvious exception of sensitive data of use in national security although recently even spy satellite data have been made available.

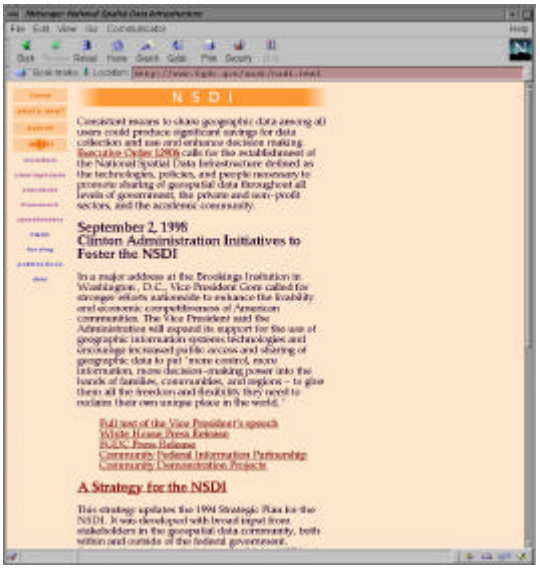
Commercial vendors



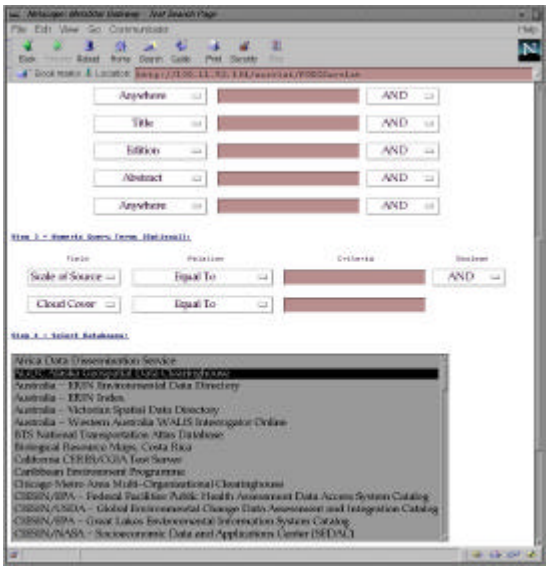
Federal Data Agencies

- ✓ USGS
- ✓ NOAA
- ✓ Census Bureau
- ✓ NIMA
- ✓ EPA
- ✓ many more…

National Spatial Data Infrastructure



National Spatial Data Clearinghouse



Finding Data on the Networks

- **The Internet is a network of computer networks and is accessible to all users through a computer that is attached to the system**

  http://www.usgs.gov

  http://www.noaa.gov

  http://www.census.gov/ftp/pub/geo/www/tiger

- **U. S. Geological Survey (USGS),**
  Department of the Interior, (partly)
  U. S. Bureau of the Census
  National Oceanic and Atmospheric Administration (NOAA)
  both part of the Department of Commerce

U.S. Geological Survey

digital data fall into six categories:

1-*digital line graphs* (DLGs),

2-*digital elevation models* (DEMs),

3-*land-use and land-cover digital data* (GIRAS),

4-*digital cartographic text* (Geographic Names Information System, GN-IS),

5-*digital orthophotoquads* (DOQ),

6-*digital raster graphics* (DRG)

USGS: National Mapping

U.S. Geological Survey

land cover derived from classifications of NOAA's AVHRR (advanced very high-resolution radiometer) measurements

the digital chart of the world (DCW): major cities, rivers, lakes, coastlines, contours, vegetation, and transportation routes.

DCW is a digital version of the Defense Mapping Agency's operational navigation chart (ONC) and jet navigation chart (JNC) at a scale of 1:1,000,000.

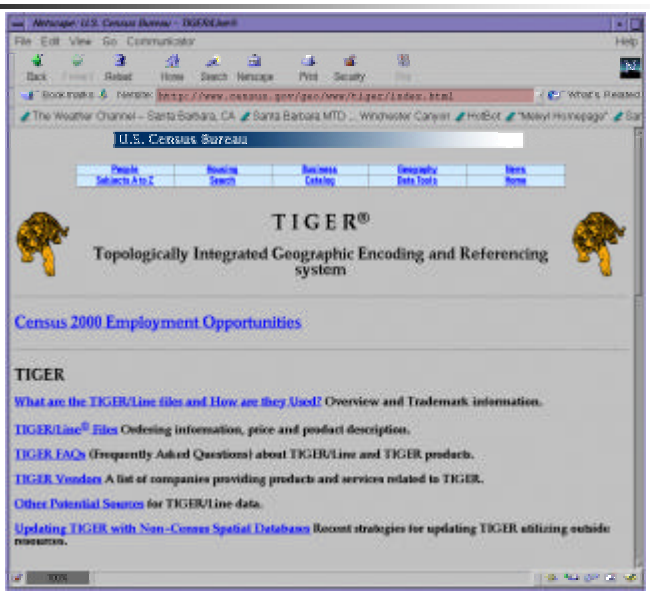
U.S. Bureau of the Census

- street-level address maps for use by the thousands of census enumerators (*address matching)*
- TIGER (topologically integrated geographic encoding and referencing)

U.S. Bureau of the Census

- TIGER uses the block face or street segment as a geographic building block and recognises cartographic objects of different dimensions, points (nodes), lines (segments), and areas (blocks, census tracts, or enumeration districts).
- In TIGER terminology, points are zero cells, lines are one cells, and areas are two cells

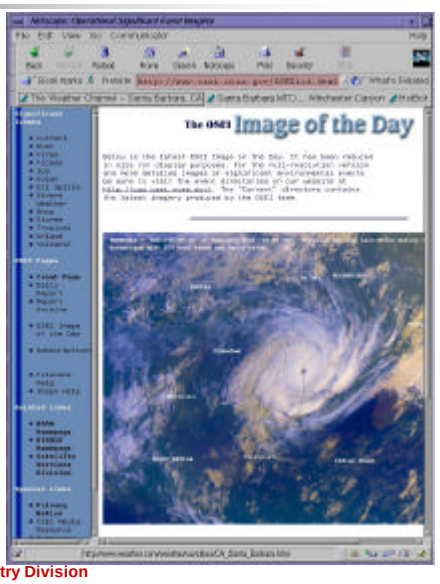
U.S. Bureau of the Census



National Oceanic and Atmospheric Administration - NOAA

- NOAA concentrates on marine and aeronautical navigation systems that electronically integrate digital charts, global positioning system-based locations, and real-time environmental information
- most recently including detailed bathymetry of the ocean and land-surface topography as well as geodetic and magnetic data for the earth's surface.

NOAA
Weather and other data

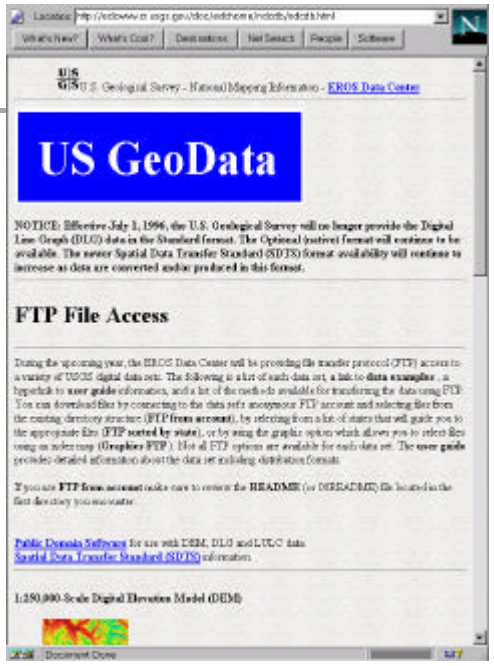


Eros Data Center

- Distributed active archive center
- Sioux Falls, SD
- Operated by USGS

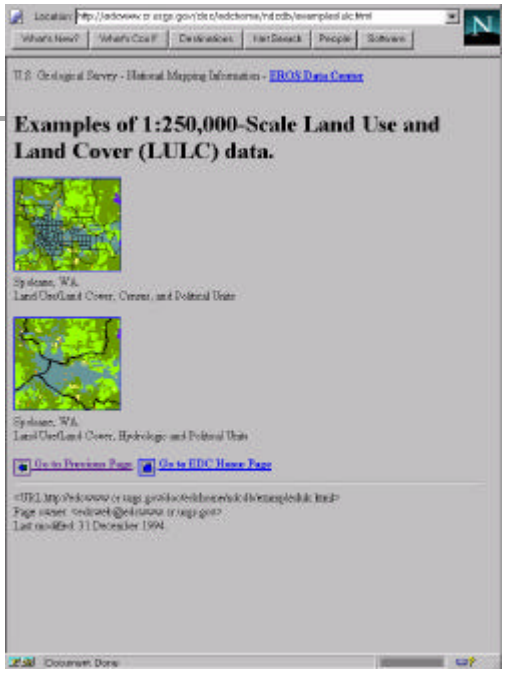
US GeoData

ftp access to
DEM
DLG
GNIS
GIRAS
etc.



GNIS Feature locations

GIRAS Land Use and Land Cover Data



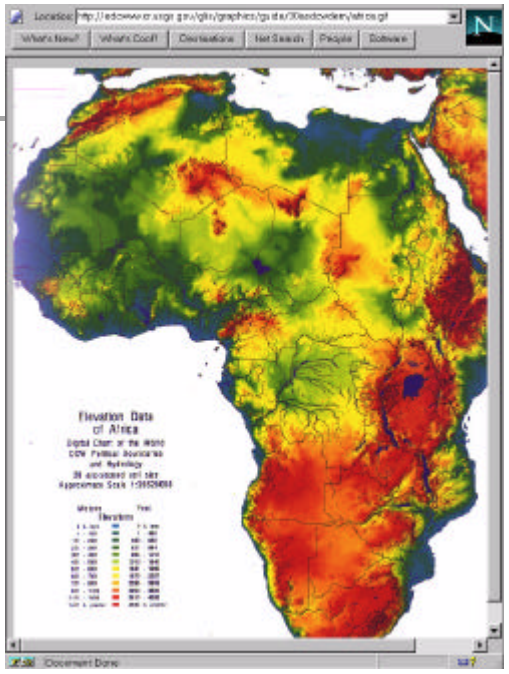
GIRAS into Arc/Info (GIRASARC)

Terrain data

DEM

DLG Contours

DCW Contours



CORONA (KH Satellites) Goleta, CA, 1967 Image

GIS data can be:

- Purchased
- Found from existing sources in digital form
- Captured from analog maps by GEOCODING

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

- The method of geocoding can influence the structure and error associated with the spatial information which results
- Example: scanning (raster), digitizing (vector)

Creating New Data

- different data formats is usually GIS analyst's problem:
  - digitised with different levels of precision
  - source maps that were out of date
  - become out of date since they were captured by computer,
  - errors or problems with their accuracy
  - …

Geocoding methods for maps

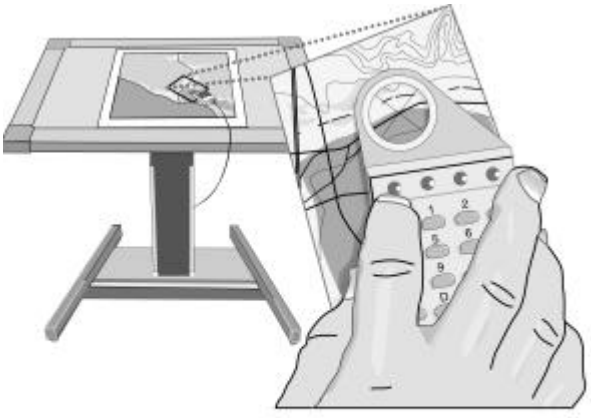
- Digitizing
- Scanning
- Field data collection

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

semiautomated digitising: means the use of a digitiser or digitising tablet.

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

process of digitising

- paper map is tailored or reprocessed

  the separate sheets should be digitised independently and digitally merged (zipped) later
- to derive a coordinate system for the map

  UTM, long-lat, hardware coordinates or map inches or millimetres
- transformed into geographic coordinates when the editing and proofing is complete

digitising

- selecting the control points and interactively entering their world coordinates
- After the map is taped to the tablet, it should not be moved without entering the control points again, and it is preferable to perform this step only once per map

Digitizing

- Cursor data entry
- Secondary tablet (menu/template)
- Voice command entry
- Point select
- Stream mode
- Distance mode

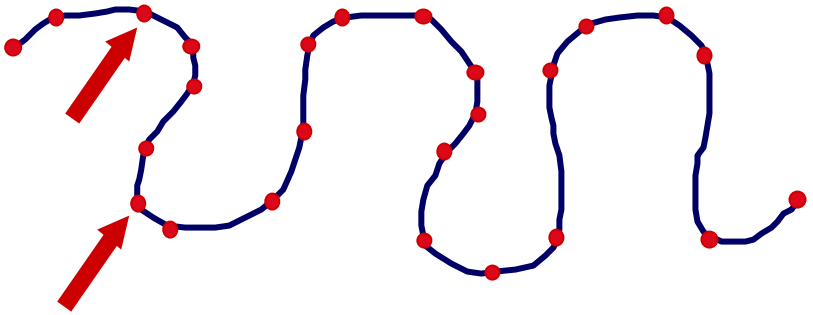
multiple collection modes for digitising :

- *Point mode* simply digitises one point each time the button on the cursor is pressed.
- *Stream mode* generates points automatically as the cursor is moved, either one point per unit of time or distance.
- *Point select mode* allows switching between point and stream mode. This mode is sometimes used when lines are both geometric and natural, such as when following a straight road and then a river.

digitising: editing and errors

- The digitising software or GIS may contain editing features, such as delete and add a line or move and snap a point.
- Errors in digitising can usually be eliminated using some very simple procedures and rules.

Selecting points to digitize



Some common digitizing errors

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

Scanning (*automated digitising)*

- *desktop (flat bed) scanner*
- *drum scanner*

A major difference with this type of digitising is that lines, features, text, and so on, are scanned at their actual width and must be reprocessed for the computer to recognise specific cartographic objects.

Scanning

- *Flat bed (Desktop)*
- *Drum*
- DPI
- File size

scanning

- An alternative scanner is the automatic line follower, a scanner that is manually moved to a line and then left to follow the line automatically which are used primarily for continuous lines, such as contours.
- CADD(computer-aided drafting and design) systems

scanning

- *of scale* and *resolution* must be a clear concept
- *Dropout:* the loss of data due to scanning at coarser resolution than the map future to be captured. Features smaller than half the size of a pixel can disappear entirely.
- noise

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 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

Field Data Collection

Using two GPS receivers in what is called *differential mode* it is possible to locate control points to sub-meter accuracy

Field data collection

GPS Data Collection

Surveying Methods

GPS Data Collection

24 orbiting satellites (about 20,000km)

each transmitting a time signal.

at least four of the satellites

when a GPS receiver is activated, the nearest satellites are located and the signals are received from each visible satellite.

GPS Data Collection

Decoding the time differences between the signals from each satellite, combined with data from the satellite itself about its orbit (called *ephemeris data)* are used to solve the three unknowns of latitude, longitude, and elevation.

Many receivers can do direct conversion into any of several coordinate systems and datums, and most can download the data directly to a computer.

Some GPS equipment can download directly in common GIS formats.

GPS Data Collection

The GPS signal is accurate in its coarse acquisition (C/A) code mode to about 75 to 100 meters, but this is because the signal is deliberately degraded under a system called *selective availability.* !!

When selective availability is turned off, accuracy of 15 to 25 meters is possible.

GPS Data Collection

differential-mode use of the GPS: one receiver measure the timing errors and then provide correction information to the other receivers that are roving around. That way virtually all errors can be eliminated from the system, even the trouble Selective Availability error.

by a radio receiver or a cellular telephone link to receive real-time differential corrections

private services http://www.trimble.com/

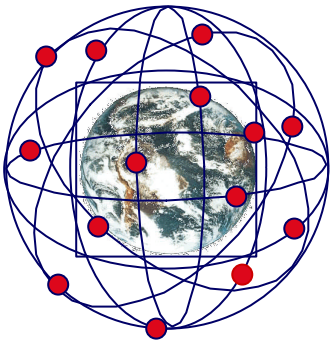
Global Positioning System: GPS

Point Positioning by Distances only:

$$(Ri+\mathbf{D}r)^2 = (Ei-En)^2 + (Ni-Nn)^2 + (Hi-Hn)^2$$

GLOBAL POSITIONING SYSTEM (GPS) Utilises a combination of Carrier and Modulation Frequency Phase Shift measurements.

24 NAVSTAR satellites in near circular orbits at 20 200 km altitude provide for usually up to six satellites simultaneously in sight of an earth bound survey station. In it's most accurate mode, accessible to US military only, positional accuracy to a few centimetre can be achieved.

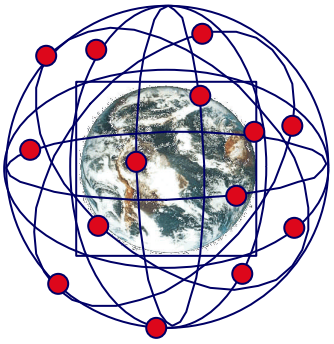


(schematic illustration only)

Global Positioning System: GPS

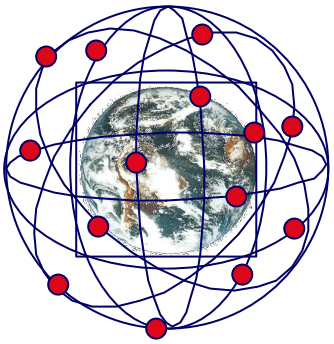
Navigation and positioning are the two practical uses of GPS:

NAVIGATION: Obtaining and monitoring the movement of a transportation vehicle or ship. Involves constantly moving platform, allowing little time for measuring any one unique position along path of movement (Reduced absolute position accuracy). Also requires determination of current and past direction of movement (in form of north referencing angle, azimuth), and velocity of movement along path. Position accuracy to within 30 to 100 m direct, 3 to 30 m in differential mode.



(schematic illustration only)

Global Positioning System: GPS



(schematic illustration only)

Navigation and positioning are the two practical uses of GPS:

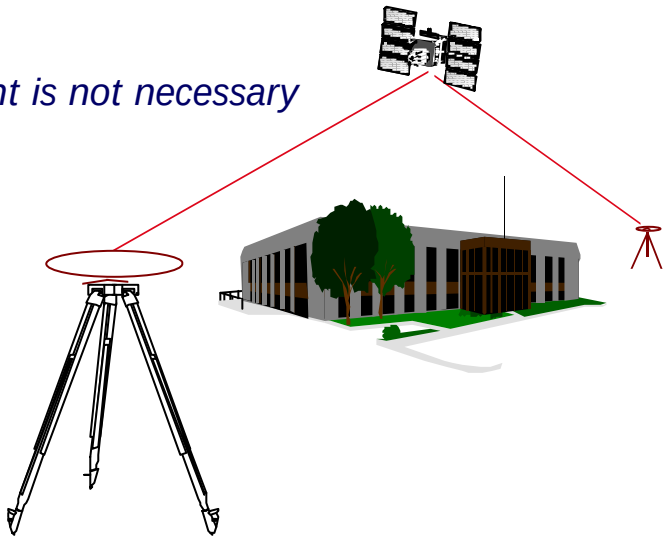
POSITIONING: Even though the full accuracy potential of GPS is limited to military use only, GPS, when applied correctly can provide for accuracy more than sufficient for Geodetic Network Determination.

For highest accuracy, the so called STATIC GPS SURVEYING method is used. Relative accuracy between stations anywhere in the net of ±(5 mm ± 1ppm) or better can be achieved, eg. ±1 m for 1000 km point separation (1ppm = 1 : 1 000 000).

(for more detail, see Textbook, chapter 20)

Why We Use Satellites for Mapping

Line of sight is not necessary



Need 4 Satellites for 3D Position

If we are looking at the previous slides in 3d then we need

- 4 satellites for 3d position (X, Y, Z, time)
- 3 satellites for 2d position (X, Y, time) - user must enter Z value
  - Problem - if user enters poor Z, then X and Y will be incorrect!
  - Solution - work only in 3D!

Knowing Where the Satellites Are

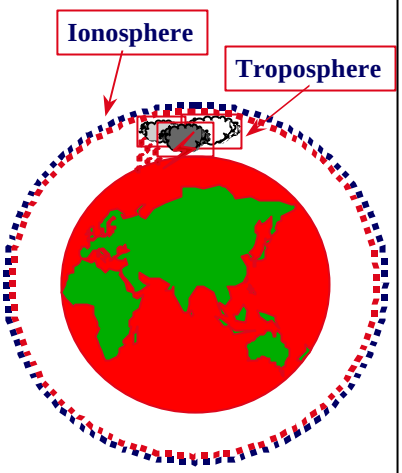
After all, they're 20,000 km up

- High orbit
  - Very stable orbits
  - No atmospheric drag
  - Survivability
  - Earth coverage
- Monitored by US Defense Department
  - DOD transmits corrections back to satellite
- Corrections transmitted from satellites to us
  - Status message

Atmospheric Corrections

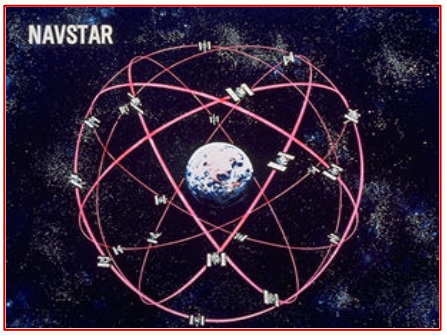
Apply estimated corrections

- The signals are delayed by the ionosphere and troposphere
- Receiver makes estimated corrections for these delays



Space Segment Description

- 24 satellites
  - 6 planes with 55° rotation
  - Each plane has 4 or 5 satellites
- Very high orbit
  - 20,000 km
  - 1 revolution in approximately
  - 12 hours
  - For accuracy
  - Survivability
  - Coverage



Copied from 'GPS Navstar User's Overview' prepare by GPS Joint Program Office, 1984

Image and Remote Sensing Data

- Digital orthophotos are at an equivalent scale of 1:12,000 and have a 1-meter ground resolution.
- The Landsat program has been generating imagery of many locations in the world since 1972.

Image and Remote Sensing Data

- French SPOT satellite series
- the Canadian RADARSAT
- NOAA polar orbiting satellites carrying the AVHRR (advanced very high resolution radiometer) for small scale
- the geostationary GOES weather satellite
- …

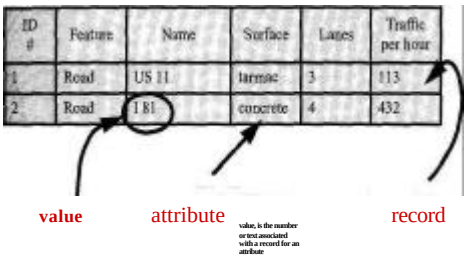
Attribute data

- Logically can be thought of as in a flat file
- Table with rows and columns
- Attributes by records
- Entries called values

Relevant road attributes:

- its state route number,
- the year it was built,
- what the surface is made up of,
- how many traffic lanes are on the road,
- if the road is one-way or two-way,
- how many bridges it goes over,
- how many cars travel along the road per hour,
- ….

An attribute table organised as flat file



ID #	Feature	Name	Surface	Lanes	Traffic per hour
1	Road	US 11	tarmac	3	113
2	Road	I 81	concrete	4	432

value attribute record

value, is the number or text associated with a record for an attribute

Value, is the number or text associated with a record for an attribute

Attribute has a name and value for each record

Record, all attributes for one future

Database elements

- Type of value
- Range
- Missing data
- Duplicate data
- Key

```
Attribute_labels = "ID #", "Feature",
"Name" , "Surface" , "Lanes", "Traffic" , "per hour"
"1",
"Road",
"US 11",
"tarmac",
"3",
"113"

"2",
"Road",

"I 81",
"concrete",
"4",
"432"
"3",
"Road",
"Lisk Bridge Road",
"tarmac",
"2",
"12",
"4"
```

attribute characteristics

- What is the *type* of the value? (text, number, decimal value or units such as meters, vehicles per hour…)
- What is the legitimate *range* of the values?
- What happens when there is a missing cell in the table? (-999 or NULL)
- Are duplicates allowed? (Name)
- Which attribute is the *key*? (ID)

Database Management Systems

- Data definition module sets constraints on the attribute values
- Data entry module to enter and correct values
- Data management system for storage and retrieval
- Legal data definitions can be listed as a data dictionary
- Database manager checks values with this dictionary, enforcing data validation.

data entry

data definition module: *many of questions must be answered in the database that allows attribute setup*

data dictionary: *a catalog of all the attributes includes range and type of values, category lists, legal and missing values, with of legal field*

data-entry module: *as part of database manager, allow the entry and modification of values*

EDITING AND VALIDATION

- limited editing capabilities
- re-entry
- …
- The connection between lines, known bordering of areas and inclusion of points in areas is called map *topology.*

the attribute data

is a data listing or *report*

errors: formatting, spelling, writing, language, misclassification, …

the roads and rivers could both be mislabeled as power lines

printing and checking

Logical consistency

- A data set that is correctly geocoded both positional and with attributes is not necessarily logically consistent
- checking by topological data

scale and precision

- generalisation of the features
- digital maps as absolutely correct maps instead of the digital alternative form of the analog maps

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

Conclusion

The intelligent GIS user should know and understand the amount and distribution of error in a GIS database.

Many of the sources of error are due to the method and process of geocoding. Some of the errors multiply as we move through the stages of data management, storage, retrieval, GIS use, and analysis.

An understanding of error is essential to working effectively with GIS.

Coming Next....

What

is

where?