GEOG 204

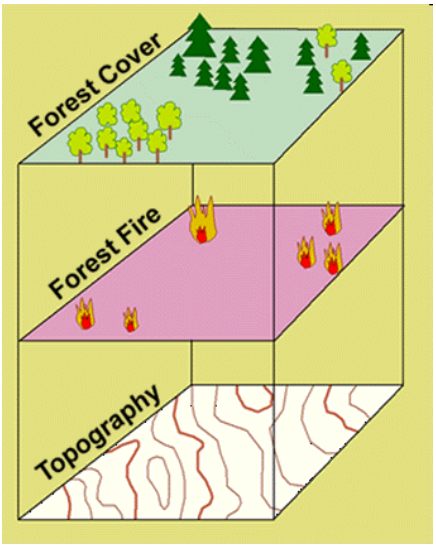
LECTURE 2

Data Collection

1

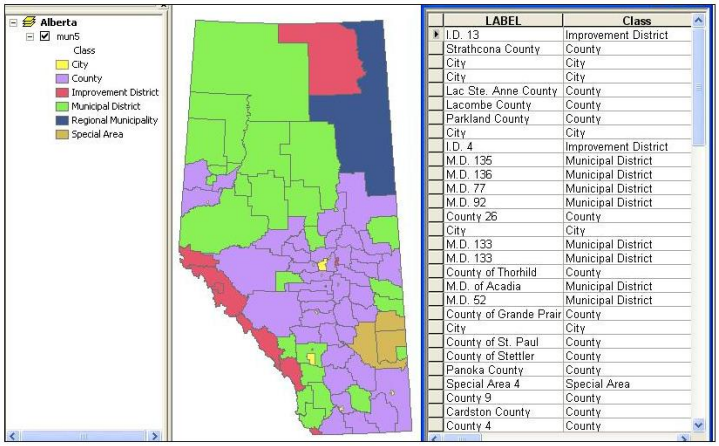
Properties of GIS data

- *Spatial (x,y location)*
  - *= 'where is it ?'*
- *Attributes (multiple)*
  - *= 'what is it' ?*
- *[Pattern]*
  - *= 'how are they related' ?*
- *Other questions?*



2

Location and Attributes



LABEL	Class
I.D. 13	Improvement District
Strathcona County	County
City	City
City	City
Lac Ste. Anne County	County
Lacombe County	County
Parkland County	County
City	City
I.D. 4	Improvement District
M.D. 135	Municipal District
M.D. 136	Municipal District
M.D. 77	Municipal District
M.D. 92	Municipal District
County 26	County
City	City
M.D. 133	Municipal District
M.D. 133	Municipal District
County of Thorhild	County
M.D. of Acadia	Municipal District
M.D. 52	Municipal District
County of Grande Prair	County
City	City
County of St. Paul	County
County of Stettler	County
Panoka County	County
Special Area 4	Special Area
County 9	County
Cardston County	County
County 4	County

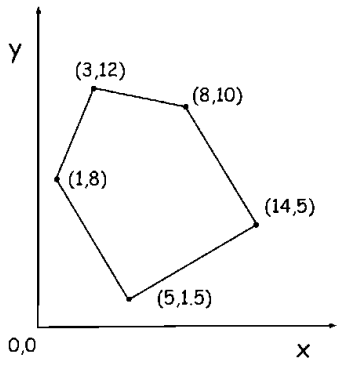
3

Location and Attributes

- **x (easting), y (northing) [z-elevation] coordinates**
- e.g. latitude / longitude (degrees, minutes, seconds)

Attributes of latlong_data2

OID	X	Y
0	-79.235444	43.207055
1	-79.261247	43.191196
2	-79.205194	43.149254
3	-79.207431	43.147622
4	-79.215648	43.15465
5	-79.250881	43.164771
6	-79.253802	43.170664
7	-79.267453	43.155617
8	-79.2638	43.16129
9	-79.21237	43.139409
10	-79.244187	43.12796
11	-79.196958	43.134143
12	-79.212195	43.132831
13	-79.279129	43.170963



Coordinates:
1,8
3,12
8,10
14,5
5,1.5
1,8

Attributes:
Lot # :1347
Street: Willow Lane
Town: Hopkins

4

Item (Description)

Record (place) ->

POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
133			0		0	0	0	0	0			0	0	
199		HM	40	HW	30	9	3	7.2	5		L	1980	0	
353		HW	90	BA	10	9	4	11.6	1	B	L	1980	1999	F
229		HW	70	HM	20	9	3	9.5	5	B	L	1980	1999	F
264		HM	50	HW	30	9	3	7.5	5	H	L	1980	1999	F
162			0		0	0	0	0	0			0	0	
393		HW	60	HM	20	9	3	8.5	5	H	L	1980	1999	R
165		HM	80	BL	20	9	3	7	4	H	L	1980	1999	R

Location and Attributes

- **Attribute data: allows us to ask the question … "what is it ?"**
  - > Every layer has an associated table
  - > These are linked to spatial location by a code number
  - Attributes are stored in columns as *items*
  - Rows display the attributes for each feature = *records*
  - Entries may be text strings, integers, float (decimal) or dates

5

Types of questions a GIS can answer – resulting from spatial location, attributes and patterns

a. **Location: WHAT exists here**
"What is at this location ?" e.g. Dig safe?

b. **Condition: WHERE are specific conditions**
Where are all the pine dominated stands ?

c. **Trends: WHAT HAS CHANGED (over time)**
How far has the riverbank receded in the past 10 years ?

d. **Patterns: HOW are features related**
"How does proximity to salmon streams affect the number of bear attacks";

e. **Modelling: WHAT IF** ..?
What if the climate warmed by 2 degrees? (e.g. effect on habitats)

6

Data Collection

- A GIS can contain a wide variety of geographic data types originating from many diverse sources
  - It is an important requirement for a GIS to integrate data from many forms of data from a diversity of sources
- Data collection is time consuming and expensive
  - In some cases costs are estimated to be 85% of the cost of a GIS (Longley at al)

7

7

Data Collection Classification

- Data collection can be classified by source
  - Primary Sources
    - captured by direct measurement specifically for use in GI systems
    - both raster and vector data can come from primary sources
  - Secondary Sources
    - reused from earlier studies or obtained from other systems
    - raster and vector data are created from maps, photographs, and other hardcopy documents

8

8

Data Collection Classification

	Raster	Vector
Primary	Digital satellite remote-sensing images	GPS measurements
	Digital aerial photographs	Field survey measurements
Secondary	Scanned maps or photographs	Topographic maps
	Digital elevation models from topographic map contours	Toponymy (place-name) databases

Longley, Goodchild, et al (2015) Geographic Information Science and Systems. John Wiley and Sons

9

9

Primary Data Collection

- Raster data
  - Data are collected by remote sensing
    - Remote sensing is the measurement of physical, chemical, and biological properties of objects without direct contact
  - Information is derived from measurements of the amount of electromagnetic radiation reflected, emitted, or scattered from objects.
    - Passive sensors rely on reflected solar radiation or emitted terrestrial radiation
    - active sensors (such as synthetic aperture radar) generate their own source of electromagnetic radiation
    - Sensors are mounted on earth-orbiting satellites or other airborne platforms

10

10

Primary Data Collection

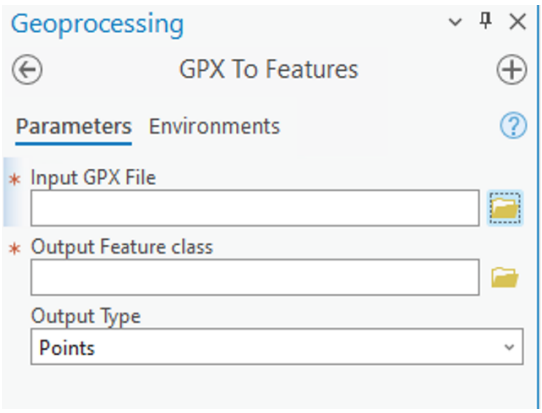
- Vector data
  - Data are captured by ground surveying, GPS and LiDAR
  - Ground surveying is based on the principle that the location of any point can be determined by measuring angles and distances from other known points.
    - It is highly accurate but time consuming and expensive
  - The GPS consists of a system of 24 satellites each orbiting the Earth every12 hours and transmitting radio pulses at precisely timed intervals
    - A receiver on the ground must make exact calculations from the signals, the known positions of the satellites, and the velocity of light in order to determine its position
    - GPS was developed by the US. Russia has GLONASS; China has BEIDU; Europe has GALILEO

11

11

GPX File Conversion

- GPX is a standard GPS file format
- QGIS reads it directly
- ArcGIS has one intermediate step
  - Analysis > Tools > Geoprossesing >GPX to Features
  - Input from GPS, output to shapefile



12

Primary Data Collection

- Vector data
  - Data are captured by ground surveying, GPS and LiDAR
  - LiDAR (light detection and ranging) employs a scanning laser range finder to collected accurate data
    - A LiDAR scanner is an active remote sensing instrument
    - It transmits electromagnetic radiation and measures the radiation that is scattered back to a receiver after interacting with the objects on the surface
    - The data collected from a LiDAR scanner is often referred to as a point cloud - a massive collection of independent points with (x, y, z)

13

13

Secondary Data Collection

- Raster Data Capture
  - Scanners
    - A scanner is a device that converts hardcopy media into digital images
    - Documents, such as building plans, CAD drawings, property deeds, and equipment photographs are scanned to reduce wear and tear, to improve access, to provide integrated database storage, and to index them geographically (e.g., building plans can be attached to building objects in geographic space).
    - Film and paper maps, aerial photographs, and images are scanned and georeferenced so that they provide geographic context for vector data layers
    - Maps, aerial photographs, and images are scanned prior to vectorization and sometimes as a prelude to spatial analysis

14

14

Scanners and Cameras

- High resolution raster

Contex

http://www.library.unt.edu/digital-projects-unit/scanners-and-scanning-systems

15

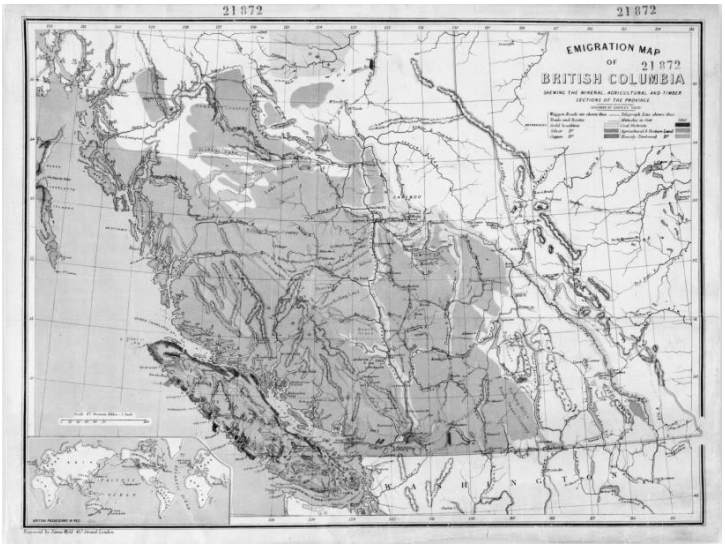
Secondary Data Collection

- Vector Data Capture
  - The digitization of vector objects from maps and other geographic data sources by heads-up digitizing and vectorization, photogrammetry, and COGO data entry
  - Heads-up digitizing and vectorization
    - creates vectors selectively from raster data
    - digitize vector objects manually straight off a computer screen using a mouse or digitizing cursor.
      - **heads-up digitizing** because the map is vertical and can be viewed without bending the head down.
    - Used to collect data for land parcels, buildings, and utility assets....

16

16

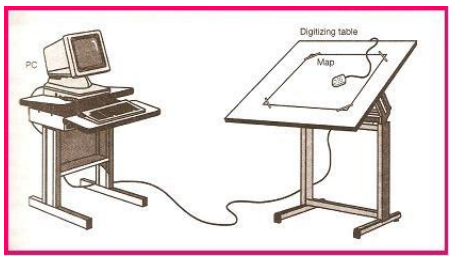
Vector Data from Historic datasets and Maps

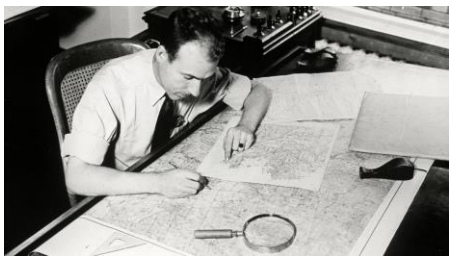


17

Vector Data from Historic datasets and Maps

- Digitizing centuries of hand-drawn maps…
- Guess who got to do this job??
  - Prisoners
  - GIS Techs
  - Students!
- Tedious and Painstaking





18

Digitizing

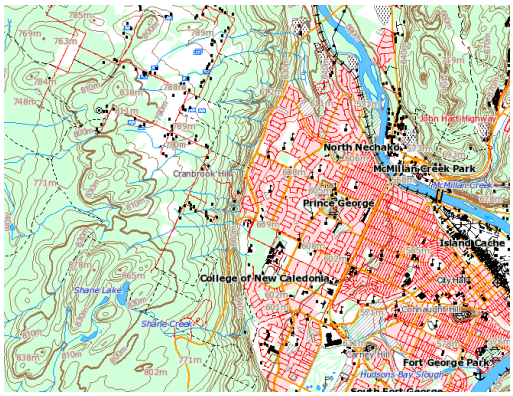
Digitizing is done in two ways:

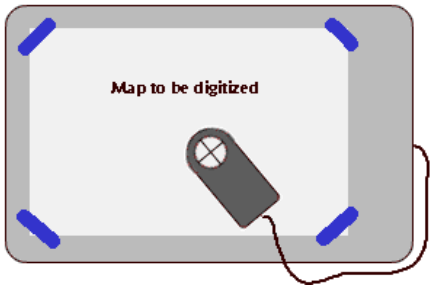
Tracing lines on maps initially using a tablet with map taped down,

or

onscreen / 'heads-up' (= copying a map) after 1995

CONNECT
THE DOTS

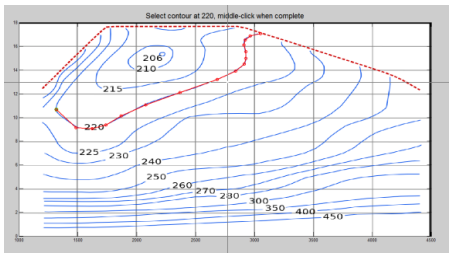




19

Digitizing Procedure

- Lines = connected points
  - Manual point selection
  - Timed point selection
  - Interval point selection

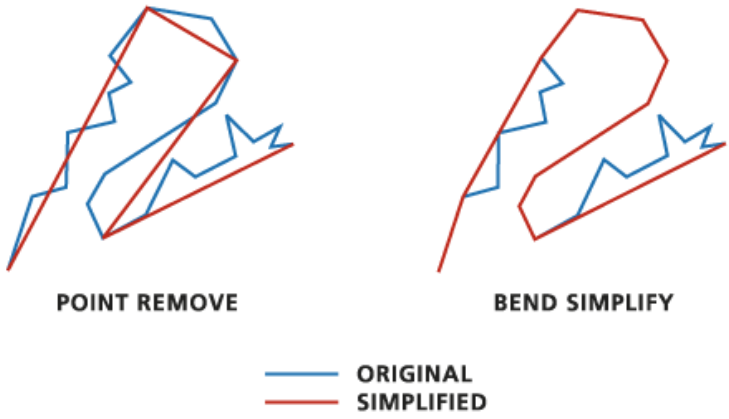


http://forum.imagej.net/t/digitizing-contour-map/118

20

Simplifying Lines

- Each vertex has a storage cost
- How much is enough? Too many?
- If too many, simplify in post process
  - Point remove: maintain essential shape
  - Bend simplify: maintain "important" bends

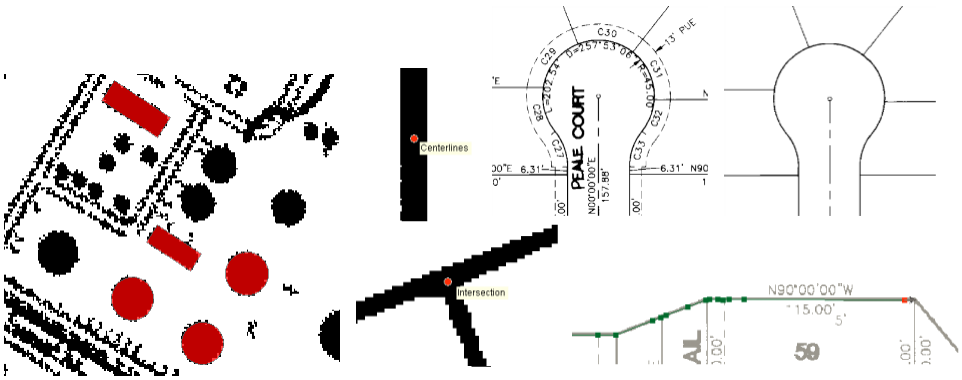


http://pro.arcgis.com/en/pro-app/tool-reference/cartography/how-simplify-line-works.htm

21

Automatic Feature Recognition

- ArcScan toolbar in ArcMap
  - Automated, semi-automated, or manual modes
  - Scanned image must be georeferenced
  - Toolbar:



22

Digitising – editing is still needed: updates and errors

Coordinate locations are based on underlying georeferencing e.g. NAD 1983

Edits: e.g. adding new features, modifying existing features, creating a new layer

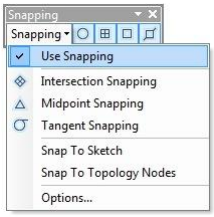
ArcEdit: https://www.youtube.com/watch?v=6dY3x-5qX6U

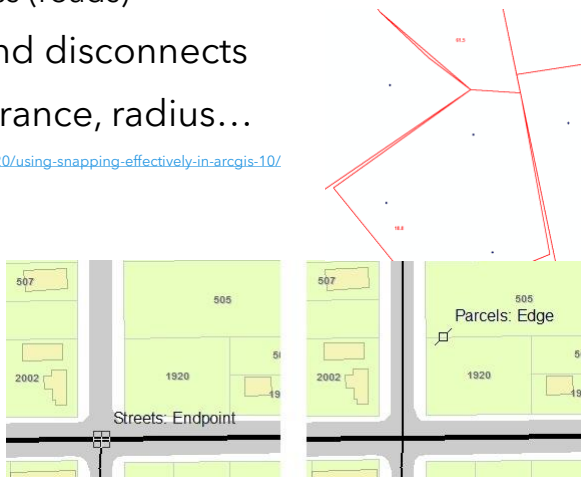
23

Snapping

- Automatic connection to other features
  - Any features, selected features, feature class
  - Same feature class (roads)
- Prevents slivers and disconnects
- User-defined tolerance, radius…

- https://blogs.esri.com/esri/arcgis/2010/09/20/using-snapping-effectively-in-arcgis-10/

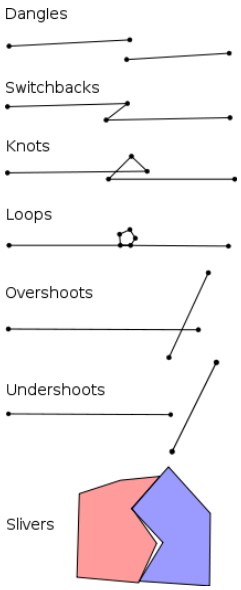




24

Digitizing errors

- Common errors
  - Dangles
  - Switchbacks
  - Knots
  - Loops
  - Overshoots
  - Undershoots
  - Slivers



Source: Caitlin Dempsey, GIS Lounge

25

25

Sources of Error

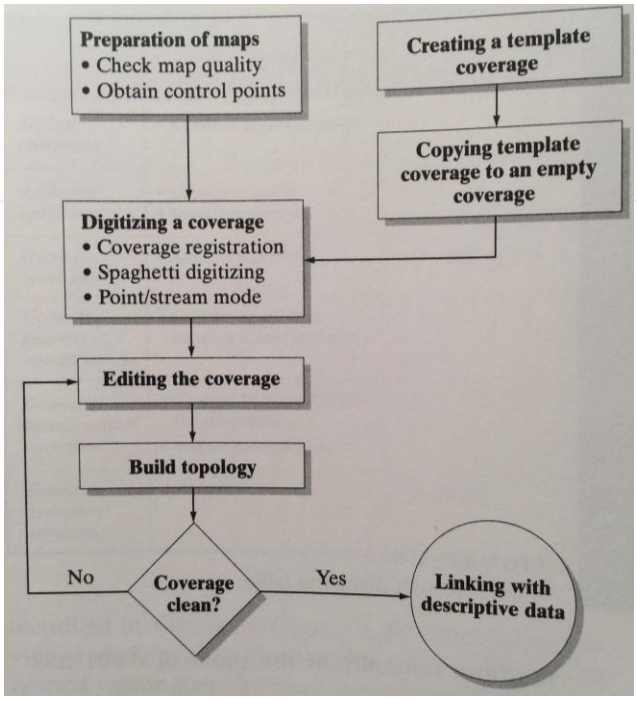
- Precision:
  - If points +/- 25m on creation
    - Similarly +/- 25m error introduced on digitization
    - Conceivably 50m total error
- Accuracy:
  - Paper may have shrunk, stretched or torn
  - Symbols rearranged to prevent overlap
  - Map sheet boundaries
  - Human boredom, fatigue, humor or malice

26

Input Error

- Very susceptible to errors
  - Does not cause error messages in digitization process
  - Outlier analysis sometimes catches mistakes
  - Easily goes unnoticed until publication

27



Coverage: A layer that can contain points, lines and polygons

28

Secondary Data Collection

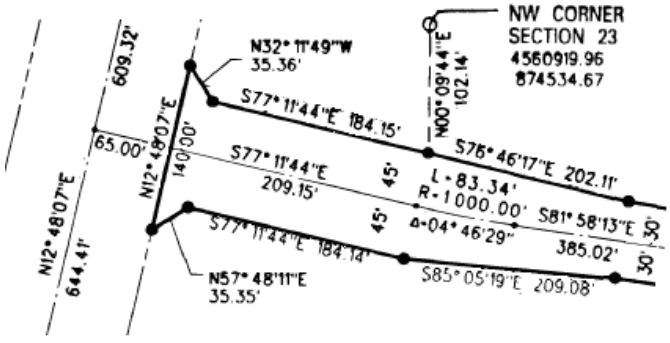
- Vector Data Capture
  - Photogrammetry
    - Measurements are taken from pictures, aerial photographs, and images
    - Measurements are captured from overlapping pairs of images using stereoplotters.
  - COGO and Other Data Entry
    - COGO is short for coordinate geometry and a method for data entry
    - Uses bearings and distances to define each part of an object
    - The COGO system is widely used in North America to represent land records and property parcels

29

29

Secondary Data Collection

- COGO descriptions for a road centerline and parcel boundaries adjoining the road



Source: ESRI

30

30

File Conversion

- FME Universal Translator
  - GIS Lab has a license
- ArcMap File formats
  - Read-only
  - Read + Write
- Raster: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/raster-and-images/supported-raster-dataset-file-formats.htm
- Vector: http://desktop.arcgis.com/en/arcmap/10.3/manage-data/datatypes/about-geographic-data-formats.htm
- QGIS: https://docs.qgis.org/2.2/en/docs/user_manual/working_with_vector/supported_data.html

31

Keyboard Data Entry

POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
133			0		0	0	0	0	0			0	0	
199		HM	40	HW	30	9	3	7.2	5		L	1980	0	
353		HW	90	BA	10	9	4	11.6	1	B	L	1980	1999	F
229		HW	70	HM	20	9	3	9.5	5	B	L	1980	1999	F
264		HM	50	HW	30	9	3	7.5	5	H	L	1980	1999	F
162			0		0	0	0	0	0			0	0	
393		HW	60	HM	20	9	3	8.5	5	H	L	1980	1999	R
165		HM	80	BL	20	9	3	7	4	H	L	1980	1999	R

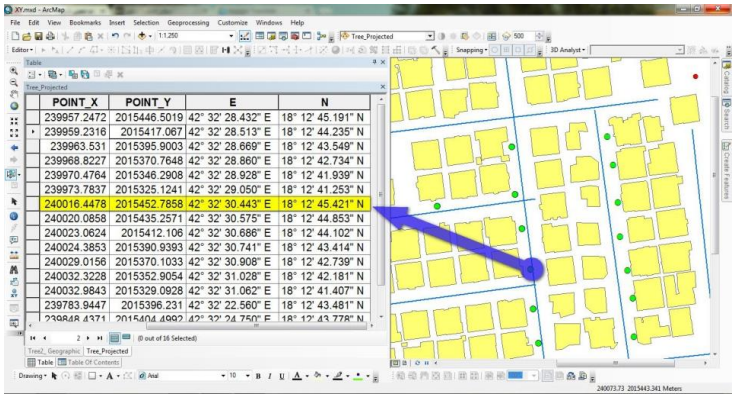
Spreadsheet --> Attribute Table

Spatial reference: GPS	GPS point # = row ID # ("Join" function)
Row ID links spatial and tabular data	Field data usually entered in Excel or similar GPS data straight to Arc

32

Keyboard Data Entry 2

- Coordinates + all data in spreadsheet
- Geographic data, no projection (unless…)



POINT_X	POINT_Y	E	N
239957.2472	2015446.5019	42° 32' 28.432" E	18° 12' 45.191" N
239959.2316	2015417.067	42° 32' 28.513" E	18° 12' 44.235" N
239963.531	2015395.9003	42° 32' 28.669" E	18° 12' 43.549" N
239968.8227	2015370.7648	42° 32' 28.860" E	18° 12' 42.734" N
239970.4764	2015346.2908	42° 32' 28.928" E	18° 12' 41.939" N
239973.7837	2015325.1241	42° 32' 29.050" E	18° 12' 41.253" N
240016.4478	2015452.7858	42° 32' 30.443" E	18° 12' 45.421" N
240020.0858	2015435.2571	42° 32' 30.575" E	18° 12' 44.853" N
240023.0624	2015412.106	42° 32' 30.686" E	18° 12' 44.102" N
240024.3853	2015390.9393	42° 32' 30.741" E	18° 12' 43.414" N
240029.0156	2015370.1033	42° 32' 30.908" E	18° 12' 42.739" N
240032.3228	2015352.9054	42° 32' 31.028" E	18° 12' 42.181" N
240032.9843	2015329.0928	42° 32' 31.062" E	18° 12' 41.407" N
239783.9447	2015396.231	42° 32' 22.560" E	18° 12' 43.481" N
239848.4371	2015404.4992	42° 32' 24.750" E	18° 12' 43.778" N

33

Data Dictionaries (1)

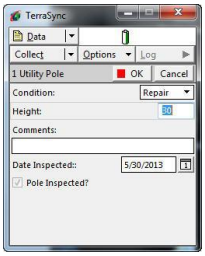
- Trimble and other survey/map grade GPS
- Establish database design first
  - Features may be culverts, bridges, signs, poles…
  - In this case, poles:

http://support.windenvironmental.com/knowledgebase.php?article=10

34

Data Dictionaries (2)

- Populate attributes while collecting points
  - Takes ~90 seconds to average enough points
  - Data entry taken care of at no additional time cost
- Attribute table ready to use
- Requires proper prior planning



http://support.windenvironmental.com/knowledgebase.php?article=106

35

Data Sources

Canadian Sources

UNBC GIS Server (L: Drive)

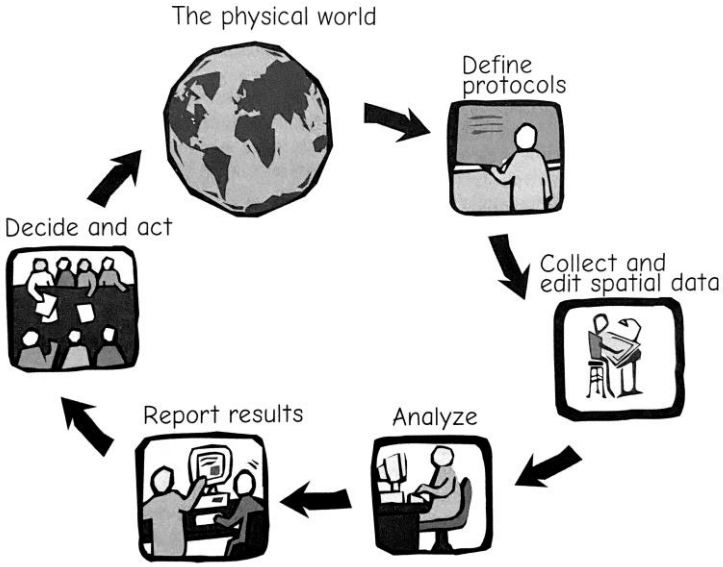
- DataBC
- GeoDiscover Alberta
- Other provincial data portals
- Municipal Data Sources
  - Prince George, BC
  - Vancouver Open Data

International Sources

- Wikipedia List of GIS Data Sources
- US Census Data
- OpenStreetMap Data
- ArcGIS Open Data

36

36



GIS functionality requires a set of protocols and an integration into the data collection, analysis, decision, and action loop of an organization.

Source: Bolstad, P. (2019). GIS Fundamentals: *A First Text on Geographic Information Systems*. XanEdu

Summary

- GIS data has location and attribute
- Classification of Sources: Primary and Secondary
- Methods of Capture for Primary and Secondary data
- Sources of Errors
- Data File Conversion