

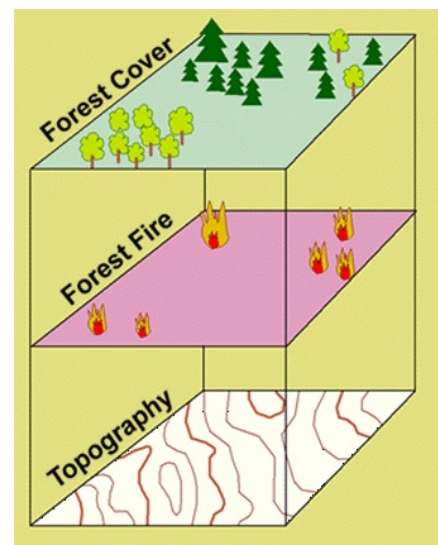
# GEOG 204

LECTURE 2  
Data Collection

1

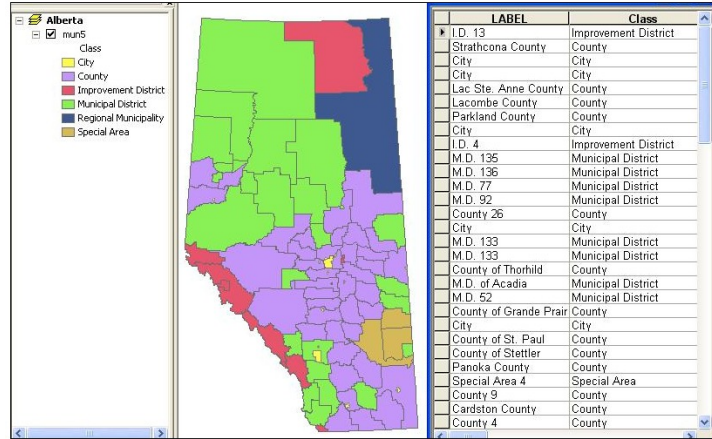
## Properties of GIS data

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



2

# Location and Attributes

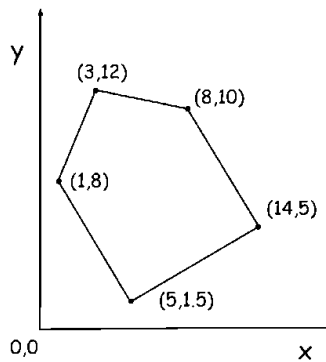


3

## 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)

	POLYGON	ESA_1	SPC1	PCT1	SPC2	PCT2	AGE_CL	HT_CL_IN	SITE_IDX	CRNCL_CL	SitePrep	Dist	YearDist	Regen	STTEND
<b>Record</b>	67		HW	40	S	40	2	1	16.6	8	B	R	1985	1999	F
<b>(place) -&gt;</b>	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

- Location: WHAT exists here**  
"What is at this location ?" e.g. Dig safe?
- Condition: WHERE are specific conditions**  
Where are all the pine dominated stands ?
- Trends: WHAT HAS CHANGED (over time)**  
How far has the riverbank receded in the past 10 years ?
- Patterns: HOW are features related**  
"How does proximity to salmon streams affect the number of bear attacks";
- 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 et 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
<b>Primary</b>	Digital satellite remote-sensing images Digital aerial photographs	GPS measurements Field survey measurements
<b>Secondary</b>	Scanned maps or photographs Digital elevation models from topographic map contours	Topographic maps 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

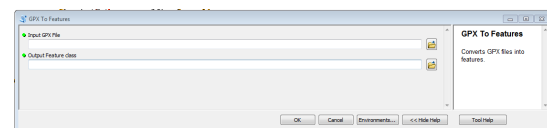
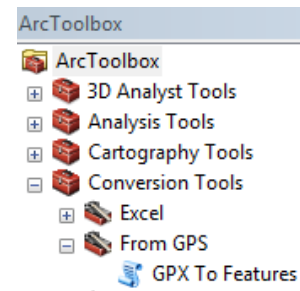
- Vector data
  - Data are captured by ground surveying, GPS and LiDAR
  - Ground surveying is based on the principle that the location of any point can be determined by measuring angles and distances from other known points.
    - It is highly accurate but time consuming and expensive
- The GPS consists of a system of 24 satellites each orbiting the Earth every 12 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
  - Conversion tools>From GPS>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

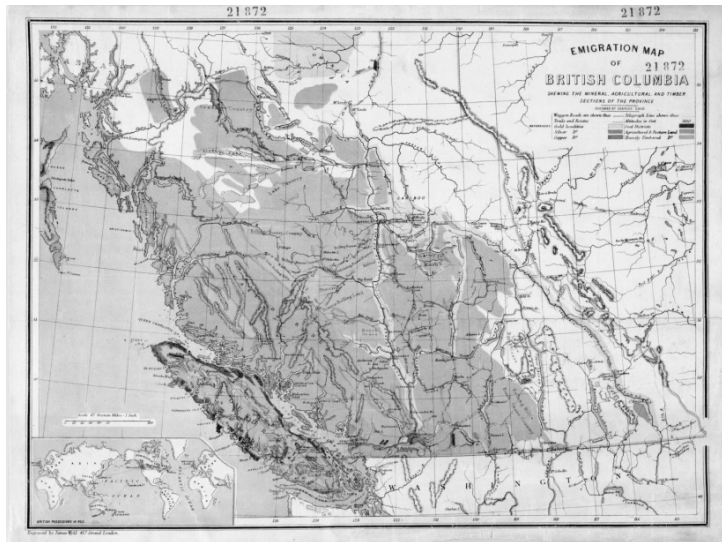
- Vector Data Capture
  - The digitization of vector objects from maps and other geographic data sources by heads-up digitizing and vectorization, photogrammetry, and COGO data entry
  - Heads-up digitizing and vectorization
    - creates vectors selectively from raster data
    - digitize vector objects 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



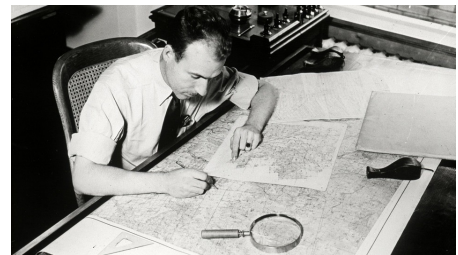
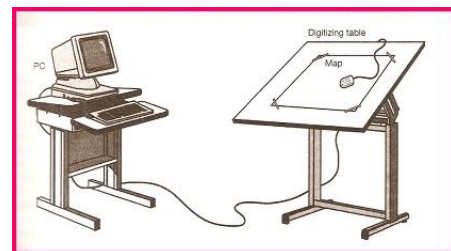
# Vector Data from Historic datasets and Maps



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

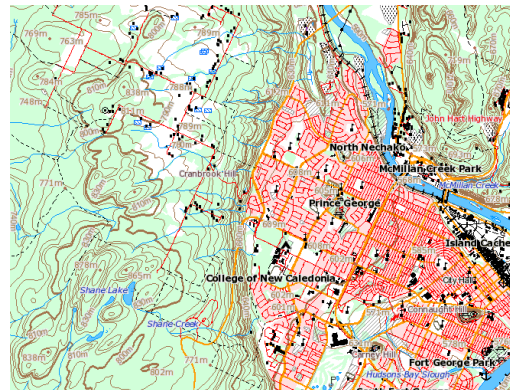
Digitizing is done in two ways:

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

or

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

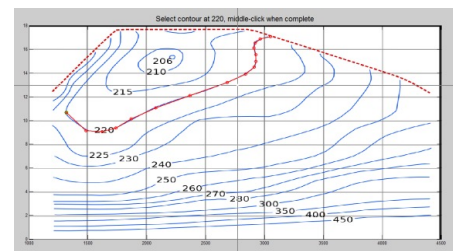
CONNECT  
THE DOTS



19

## Digitizing Procedure

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

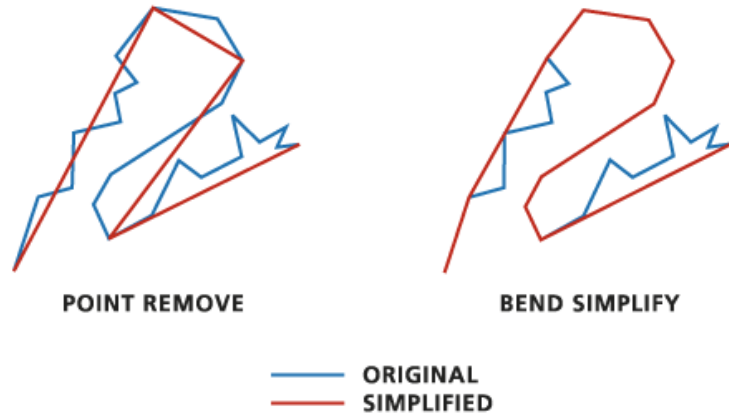


<http://forum.imagej.net/t/digitizing-contour-map/118>

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

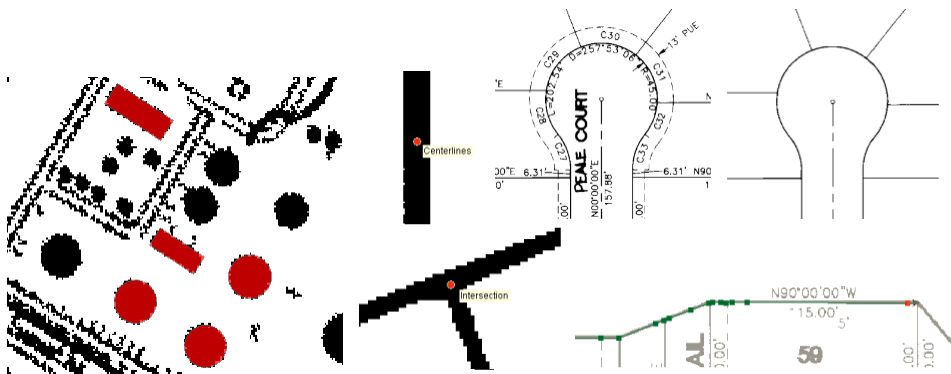


<http://pro.arcgis.com/en/pro-app/tool-reference/cartography/how-simplify-line-works.htm>

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

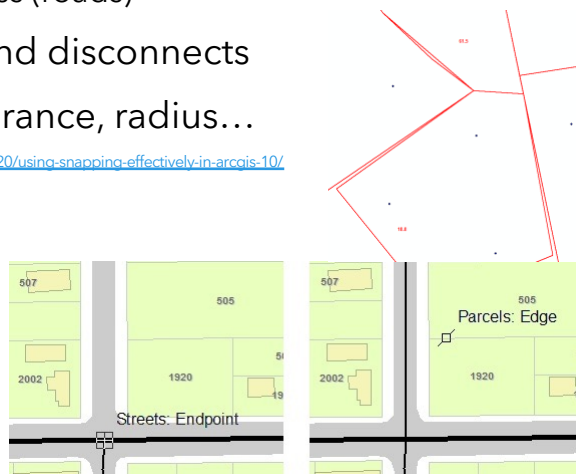
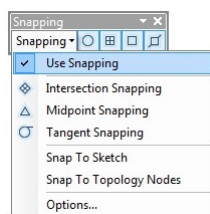
Edits: e.g. adding new features, modifying existing features, creating a new layer

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

23

## Snapping

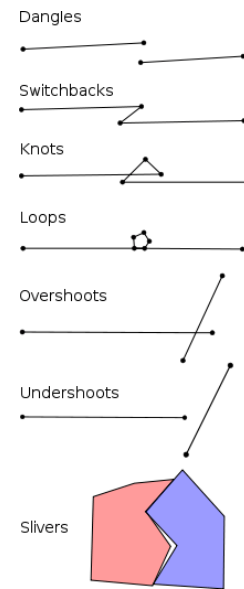
- Automatic connection to other features
  - Any features, selected features, feature class
  - Same feature class (roads)
- Prevents slivers and disconnects
- User-defined tolerance, radius...
- <https://blogs.esri.com/esri/arcgis/2010/09/20/using-snapping-effectively-in-arcgis-10/>



24

## Digitizing errors

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



Source: Caitlin Dempsey, GIS Lounge

25

25

## Sources of Error

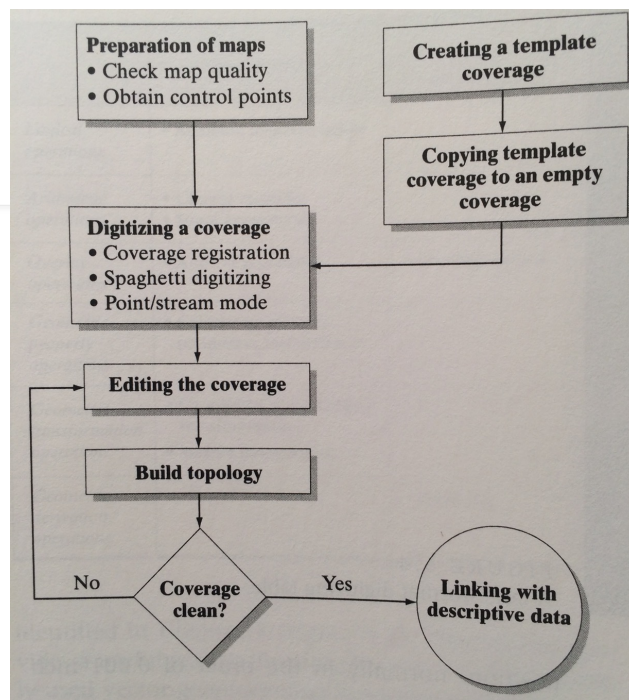
- Precision:
  - If points +/- 25m on creation
    - Similarly +/- 25m error introduced on digitization
  - Conceivably 50m total error
- Accuracy:
  - Paper may have shrunk, stretched or torn
  - Symbols rearranged to prevent overlap
  - Map sheet boundaries
  - Human boredom, fatigue, humor or malice

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

28

## Secondary Data Collection

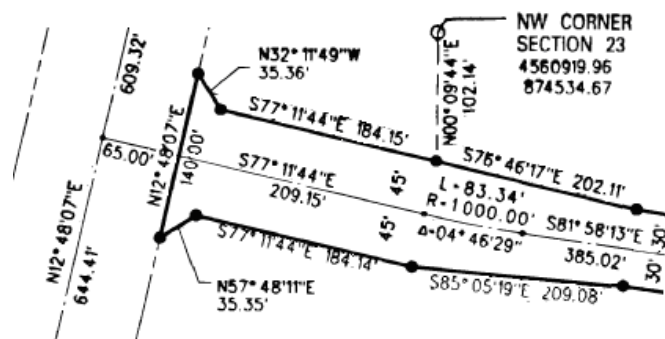
- Vector Data Capture
  - Photogrammetry
    - Measurements are taken from pictures, aerial photographs, and images
    - Measurements are captured from overlapping pairs of images using stereoplotters.
- COGO and Other Data Entry
  - COGO is short for coordinate geometry and a method for data entry
  - 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](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	STEND	
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

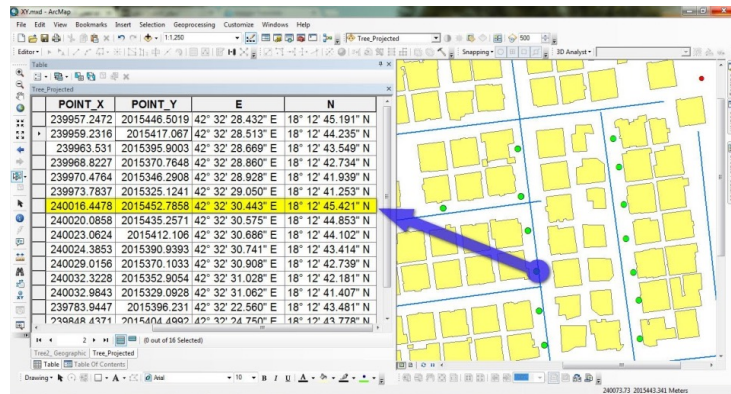
Field data usually entered in Excel or similar  
GPS data straight to Arc

32



## Keyboard Data Entry 2

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

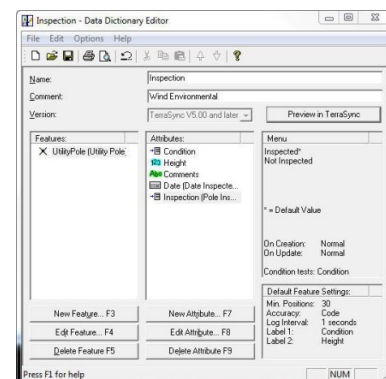


33

## 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:

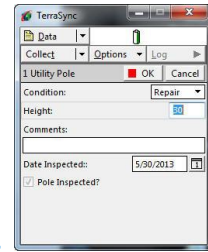


<http://support.windenviro.com/knowledgebase.php?article=10>

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

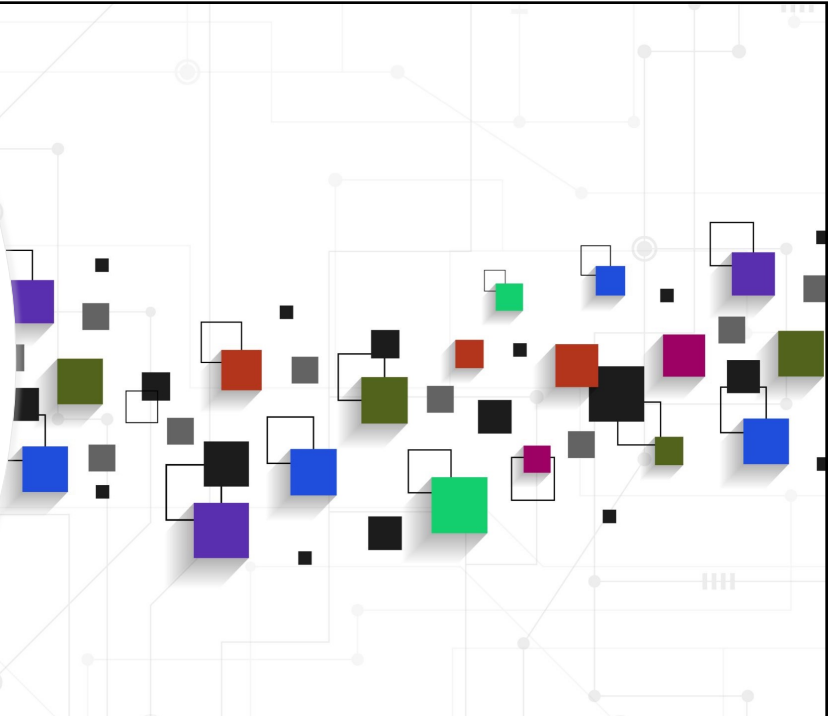


<http://support.windenvironmental.com/knowledgebase.php?article=106>

35

## Data Acquisition

- [Data BC](#)
- [ESRI Open Data](#)
- [GeoDiscover Alberta](#)
- [BC MoFLNRO](#)
- [Google](#)
- [Openstreet Map Data](#)
- Municipal Open Data Portals
- Spatially referenced 99% of the time
- Most data publicly available (no \$\$)



36



## Canadian Data Sources

- Canada: <https://www.mcgill.ca/library/find/maps/geospatial-online>
- BC: <https://catalogue.data.gov.bc.ca/dataset>
- Alberta: <https://geodiscover.alberta.ca/geoport/catalog/main/home.page>
- Saskatchewan: <https://www.isc.ca/Pages/Content%20Gallery/GeoSask.aspx>
- Manitoba: <http://www.manitoba.ca/gem/gis/index.html>
- Ontario: <https://www.ontario.ca/page/land-information-ontario>
- New Brunswick: <http://www.snb.ca/geonh1/e/DC/catalogue-F.asp>
- Nova Scotia: <https://geonova.novascotia.ca/>
- Quebec: <https://www.mcgill.ca/library/find/maps> (Gov't data in a non-ESRI format)
- Yukon: <http://www.geomaticsyukon.ca/>
- Northwest Territories: <http://www.geomatics.gov.nt.ca/dldoptions.aspx> (must register)
- Nunavut: <http://ntilands.tunnqavik.com/maps/>

37

## U.S. and Other Data Sources



NASA: <http://data.giss.nasa.gov/>



[Census Data](#)



[US Geological Society](#)



[Wikipedia's List of GIS Data Sources](#)

38

## Searching the Web: Keywords

- [Google tips and tricks](#)
- [Shapefile British Columbia Wildfire](#)

**How To Google It**

**Don't ask Google questions.** Think about how an answer would be phrased, and search for that (ie, never search for 'What is the air speed velocity...').

**filetype:**  
Searches only results of the file type you designate. Can use for pdf, doc, jpg, etc.

**intitle:**  
Only shows results with that word in the title (in this case, 'velocity').

**filetype:pdf air speed intitle:velocity of swallow**

Replaces itself with common terms in your search (in this case, Red Rumped swallow and Lesser Striped swallow will both be searched, along with many others).

## How I google something

PC overheated solutions

## How my MOM googles something

My PC is overheated how do I get it to cool down because this pc is my life

We Know Moms

39

## Publicly Available KMLs

- Sometimes Google Earth is all you get
- It's enough...
  - It comes in like this:
  - Right click Polygons>File>Export data>Export as shapefile
  - Result: usable shapefile/attribute table (turn off extra items)

**KML To Layer**

Input KML File

Output Location

Output Data Name (Optional)

☐ Include Ground Overlay (Optional)

OK Cancel Environment... Help

**KML To Layer**

Converts a KML or KMZ file into feature classes and a layer file. This layer file maintains the symbology found within the original KML or KMZ file.

**ArcToolbox**

- ArcToolbox
  - 3D Analyst Tools
  - Analysis Tools
  - Cartography Tools
  - Conversion Tools
    - Excel
    - From GPS
    - From KML
    - KML To Layer
    - From Raster
    - From WFS
    - JSON
    - Metadata
    - To CAD
    - To Coverage
    - To dBASE
    - To Geodatabase
    - To KML
    - To Raster
    - To Shapefile

☒ DFB PG Fixed Wing All

☒ Polygons

SymbolID

0

ID	ShapeID	Name	Code/Path	Symbology	Attribute	Base	Clipped	Estimated	Support
1	Polygon 1	DFB_PG_Fixed_Wing_All	DFB_PG_Fixed_Wing_All	0	0	0	0	0	0
2	Polygon 2	DFB_PG_Fixed_Wing_All	DFB_PG_Fixed_Wing_All	0	0	0	0	0	0
3	Polygon 3	DFB_PG_Fixed_Wing_All	DFB_PG_Fixed_Wing_All	0	0	0	0	0	0

40

## Shapefiles

One "shapefile" = 3 or more files

**.shp**: shape format/geometry

**.shx**: shape index format (file navigation)

**.dbf**: attribute data (the 'spreadsheet')

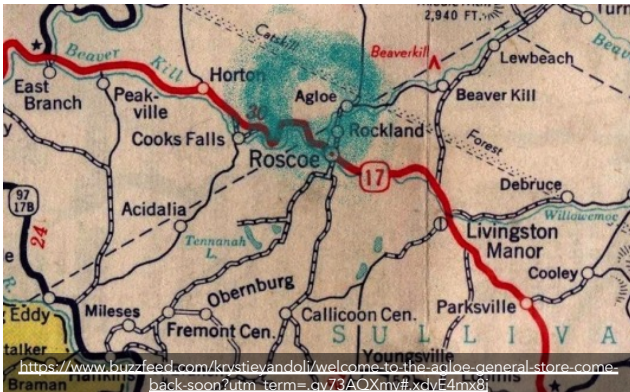
**.prj**: projection data

.sbn, .sbx, .fbn, .fbx, .ain, .aih, .ixs, .mxs, .atx, .shp.xml, .cpg, .qix: [other formatting files](#)

**ALL HAVE TO MOVE TOGETHER**

41

## Paper Towns: False Input



- Agloe, New York
  - Copyright "trap"
    - Agloe General Store later built at location
- Beatosu (Beat OSU) and Goblu (Go Blue)
  - Also copyright trap

42