Data Conversion and data sources

Topics

1. Introduction

Geospatial data is created, shared, and stored in different formats. Creating your digital datasets from scratch can be slow and tedious. Quite a lot of data source are already in some kind of digital format. Taking this advantage can be much easier and time saving. However, existing data usually come from different sources/organizations and this resulted in having a wide variety of file formats, not all of which are compatible. In order to use these exist data sources, conversion need to be done before working in GIS project. In the GIS field it is said that at least 3/4 of your time is spent on data conversion, assembling and manipulation.

Vector Data File Formats

Layer File (.lyr)
A layer file is directly readable only by ArcGIS software and other newer software applications. This file does not contain actual geographic data, but rather contains specifications for the presentation of other datasets. Such specifications include color shading, naming, label properties (font, color, placements, etc.). A layer file allows these settings to be saved and shared.

Shapefile (.shp)
The ESRI Shapefile has become an industry standard geospatial data format, and is compatible to some extent with practically all recently released GIS software. A spatial data in shape file format should have several files associated with the spatial data. To have a complete shapefile, you must have at least 3 files with the same prefix name and with the following extensions: .shp = shapefile, .shx = header and .dbf = associated database file. Additionally, you may have a .prj = Projection file, a .lyr = layer file, and other index files. All these files must be saved in the same workspace.

ArcInfo Coverage
An ArcInfo **coverage** is a collection of feature classes used to represent geographic phenomena. A coverage does not have an individual file extension. Instead it is composed of two folders within a "workspace" which each contain multiple files. One of the two folders carries the name of the coverage, and contains a number of various .adf files. The other folder is an "info" folder, which typically contains .dat and .nit files for all the coverages and grids in the workspace.

Arc Export or Interchange Format (.e00)
.e00 (pronounced e-o-o or e-zero-zero) files are ArcInfo Interchange or export files, used to conveniently copy and move ArcInfo GIS coverages (see above) and grids (see below). An .e00 file must be "imported" in order to use the data in ArcView or other GIS software. It is a type of compressed version of coverage. The E00 file can be converted /imported to ArcGIS easily.

AutoCAD DXF files (.dxf)
DXF format is the standard interchange format used by most CAD systems. These CAD files can be viewed directly in ArcGIS

IGDS & Miscrostation files (.gdn)

IGDS and MicroStation design files both have the same file structure and are both referred to as DGN files. Both are binary files organized into levels containing one or more drawing elements, or entities. Graphic attributes are stored in the design file. Database attributes are not. Instead each element can store pointers to attributes in external database files. These pointers, or linkages, are preserved during the conversion. IGDSARC can translate design files created on either PC, Vax, or UNIX platforms.

ASCII text files (.txt)

ASCII text file can be used to store spatial and non-spatial information. These text file contains the locations of points, lines and polygons. The spatial information can be converted to other format for view and analysis purpose in GIS.

Geodatabase (.gdb)

The geodatabase is a collection of geographic datasets of various types, with the most basic types being vector, raster, and tabular data. There are three types of geodatabases: file, personal, and ArcSDE. Geodatabases are the native data format for ESRI's ArcGIS now.

Raster Data File Formats

Raster Grid

An ArcInfo Grid does not have an individual file extension. Instead it is composed of two folders within a "workspace" which each contain multiple files. One of the two folders carries the name of the grid, and contains a number of various .adf files. The other folder is an "info" folder, which typically contains .dat and .nit files for all the coverages and grids in the workspace. The best way to manage (copy, move, delete, rename) ArcInfo Grids is with ArcCatalog or ArcInfo Workstation (command line).

MrSID (pronounced "mister sid") is a proprietary format of LizardTech's GeoExpress software for imagery compression, and is commonly used on orthoimages. The MrSID file extension is .sid. A companion file with an .sdw extension and the same prefix name as the .sid is used as a world file for georeferencing an MrSID image.

ERMapper (.ecw)

ECW is a proprietary format of ERMapper for imagery compression. It is a more recent format than MrSID, but is gaining popularity because of free compression utilities available from ER Mapper's website.

TIFF File Format

TIFF is a container format including images and data within a single file with header files describing the internal properties.

JPEG 2000 (.jp2)

JPEG 2000 is a non-proprietary image compression format based on ISO standards, and typically uses .jp2 as the file extension. Its advantages are that it offers lossy and lossless compression, and world files (.j2w) can be used to georeference an image in GIS software. Compression ratios are similar to MrSID and ECW formats.

Other Data Sources

SAIF (pronounced safe) (.saf)

SAIF (Spatial Archive and Interchange Format) is a Canadian geomatics standard for the exchange of geographic data. SAIF uses an object oriented data model, and consists of definitions of the underlying building blocks, including tuples, sets, lists, enumerations, and primitives. It also includes the rules for defining higher level constructs, and the SAIF standard defines a base set of approximately 300 classes. These include such things as Points, Spatial Reference Systems, Map Projections, Arcs, and Polygons.

NTS, the **National Topographic System (NTS)** provides general-purpose topographic map coverage of Canada. These maps depict in detail ground relief (landforms and terrain), drainage (lakes and rivers), forest cover, administrative areas, populated areas, transportation routes and facilities (including roads and railways), and other man-made. The spatial data are available in 1:50000 and 1:250000 scale. The spatial data can be downloaded free of charge at *http://geogratis.cgdi.gc.ca/geogratis/en/index.html*

Satellite images

Satellite images are images taken from space of an area at any given time. The satellite images provides much more information such infra-red image, near infra-red, mid infra-red and thermal etc. than regular air photos

BC TRIM data

TRIM (Terrain Resource Information Mapping) is a series of digital maps at 20,000 scale produced and sold by Base Mapping and Geomatics Services Branch of the Ministry of Sustainable Resource Management BC. They are released in a binary, compressed form

BC Forest Coverage File / VRI (Vegetation Resources Inventory)

Forest coverage file (now is called VRI data) contains the information on vegetation such as species, harvest...

Canadian Soil Information Service (*http://res.agr.ca/cansis/nsdb/intro.html*)

National Soil DataBase (NSDB) is the national archive for soil and land resource information in Canada. Spatial datasets provide information on ecological groupings, soil properties and distribution.

File formats used in ArcGIS

Map document file (.mxd)	Raster data; jpeg, tif, etc.	
Geodatabase	Coverage - Polygon	Shapefile - Polygon
Geodatabase feature dataset	Coverage - Arc	Shapefile - Arc
Geodatabase feature class - Polygon	Coverage - Point	Shapefile - Point
Geodatabase feature class - Arc	Coverage feature class - Polygon	
Geodatabase feature class - Point	Coverage feature class - Arc	
Geodatabase feature class - Annotation	Coverage feature class - Tic	
	Coverage feature class - Label	

2. Text/CSV file to Shapefile

Make sure you have network drive K and L. K drive is your home directory and L drive is the home for all users and lab data

- Start Terminal Server Osmotar
- Create a course folder **geog413** in your K drive
- Create a subfolder **conversion** under your **geog413** for today's lab.
- Make sure you have extensions enabled, click Customize->Extension and check all extensions.

Spatial locations can be stored in ASCII text file containing x, y, z coordinates. This type of ASCII file can be converted directly to spatial data.

Take a look the **plots.txt** located at L:\labs\geog413\data\conversion\. You can use any text editor to view the file. We will use Notepad to open the file **plots.txt**.

```
1,562349,5988857
2,565429,5992530
3,560019,5994115
4,563737,5989856
5,557001,5994007
6,562323,5988959
7,562314,5994490
.....
```

Run Notepad

- Start -> All Programs -> Accessories -> Notepad
- Open the **plot.txt** from L:\labs\geog413\data\conversion

Each line in the file contains a feature's spatial location. The first column indicates the ID, the second column is X coordinate, and the third column is Y coordinate of the location.

- Add the a line at the beginning of the file

      ID,X_coord,Y_coord

- Save the file as **my_plots.txt** to your local folder
- In ArcCatalog, navigate to your local folder, right-click **my_plot.txt** and choose Create Feature Class->From XY table.
- Make sure that the X Field is set to X_coord and Y Field is set to Y_coord.
- Set the output dataset to K:\geog413\conversion***plots.shp***.

- You can set the Coordinate system by clicking on Coordinate System of Input Coordinate. Click Select button, this will pop up the coordinate system selection. Click Projected Coordinate Systems-> UTM -> NAD 1983 and select NAD 1983 UTM Zone 10N.prj and click Add. Click OK to apply the changes and close the window.
- Click OK. You will have a shape file ***plots.shp*** generated under your directory.

3. E00 to ArcInfo coverage

An Arc/Info interchange file is a type of file format (old version of ArcGIS file format) that enables a coverage to be portable by parceling all its files into one file, and compressing file size. These files are recognized by the extension **.e00** and are formally known as 'ArcInfo export interchange files'. E00 file format is commonly used for data transfer between different machines or platform. When transferring ArcInfo coverages, they are converted to the interchange file format. To use the files they must be converted back to coverages.

We have Forest Coverage files in E00 format located at L:\labs\geog413\data\conversion\. The spatial data are stored in four separate files.

fc93i01.e00, fc93i11.e00, fc93j10.e00, fc93j20.e00

As the study area cross overs several 1:20,000 mapsheets, so we need all these mapsheets to cover study area.

E00 file is not in a spatial data format, you can't see them in ArcCatalog by default.

- In ArcCatalog. Click the Customize->ArcCatalog Options and click File Types tab. Here you can add file types to allow ArcCatalog to list them in the catalog tree.
- Click New Type. In the popped up window, give E00 for the file extension and 'ArcInfo Inter-exchange file' for the Description and click OK button.
- Back to Catalog tree and navigate to L:\labs\geog413\data\conversion\, you should be able to find all **.e00** files. If not, right-click on the folder **conversions** and chooses Refresh
- In ArcToolbox, click   **Conversion->To Coverage->Import form E00**

A small window pops up asking for input file and output dataset.

Input File:	L:\labs\geog413\data\conversion**fc93i01.e00**
Output Folder:	**K:\geog413\conversion**
Output Dataset:	*fc93i01*

- Convert the rest of E00 files to coverage and name them ***fc93i11***, ***fc93j10, fc93j20*** respectively under your **conversion** folder.
- Examine the datasets in ArcCatalog.

4. DXF and IGDS to coverages and shapefile

Some spatial data are stored in DXF and DGN (AutoCAD) format.

These are ***streams.dfx*** (DXF) and ***roads.dgn*** (IGDS) located at L:\labs\geog413\data\conversion\

Started from ArcGIS 9.3, DXF and IGDS files can be viewed and used directly in ArcGIS. But it would be better to have all datasets in the same format. Another issue is that DXF and IGDS data usually do not have coordinate system defined that may cause problem for later analysis.

Here we will convert two datasets: ***streams.dfx*** (DXF) and ***roads.dgn*** (IGDS) to shapefile format. This can be done easily in ArcCatalog.

- In ArcCatalog, navigate to L:\labs\geog413\data\conersion folder

- Expend the dataset by clicking the plus sign beside *roads.dgn*. You will see that there are more than one feature class in this dataset. The useful feature class is **Polyline** which indicates road lines.
- Right-click **Polyline** and choose Export-> To Shapefile(single)...
- In the popped up window, specify the Output Location to K:\geog413\conversion
- Set Output feature Class as *roads* and in your local folder.
- In the same way export Polyline of *streams.dxf* to a shapefile and name it to *streams* in your local folder.

Check the datasets *roads* and *streams* in your **conversion** folder.

- Right-click on *roads* and choose Properties. Click XY Coordinate System tab at top. Notice that "Unknown" is showing in the Current coordinate system box which indicates the coordinate system is not defined. This may cause problem for later analysis.
- Define the coordinate system for *roads* and *streams*.in NAD_1983_UTM_Zone_10N

5. BC TRIM data in SAIF format

Beside the information we have above, we also want the information on **rivers, lakes, elevation, contours, break lines**, and some other ground features in large scale. The TRIM data contain all the ground features and are stored in SAIF format. BC TRIM data are in 1:20,000 large scale which provide more details on ground features

The datasets we will work are located at L:\labs\geog413\data\conversion

93i001.saif, 93i011.saif, 93j010.sail, 93j020.saif

The data come with four mapsheets to cover the study area. We need to convert these datasets to shapefile format.

To convert SAIF data format, we will use [FME](#) software, the third party conversion software
- Click Start->FME Desktop->Utilities->FME Quick Translator
- Click Run button and go to C:\Program Files\FME\gallery\trim2000\ folder. Highlight **trim2_to_shp.fme** and click Open. This is a mapping file generated from FME allowing for TRIM data conversion.
- Set the parameters as the following

TRIM SAIF input file name:	Click the Browse button [...] and go to the conversion folder. Select *93i001.saif, 93i011.saif, 93j010.saif, 93j020.saif* (use the SHIFT key to add all datasets)
Shape Directory	K:\geog413\conversion**trim**
Version	Updated
Attribution Type	Full Attribution

You will have a lists of shapefiles generated in the output folder **trim**. FME automatically merges data if there are more than one source files.

- Examine the TRIM layers in your conversion/trim folder

Now we have all datasets ready for our study area. These datasets can be used for analysis in the later labs.

We have known that datasets *plots, roads, streams* are in UTM projection. How are the trim data?

- Check the coordinate system of TRIM data. You should be able to find the coordinate system information.

Q1. What is the coordinate system for forest cover data (fc data)?
Q2. What is the coordinate system for trim data (in trim folder)?
Q3. These two types of data are in different projection system. Why can they be displayed perfectly in ArcMap

6. KML & ArcGIS Layer

Google earth is a very handy tool for browsing earth, location, and landscape etc. KML and KMZ file formats used in Google earth are more widely used for carrying geographic locations collected from Google earth.

KML, stands for Keyhole Markup, is a file format used to display geographic data in an Earth browser such as Google Earth, Google Maps, and Google Maps for mobile. KML uses a tag-based structure with nested elements and attributes and is based on the XML standard. All tags are case-sensitive and must be appear exactly as they are listed in the KML Reference.

KMZ stands for KML-Zipped. It is the default format for KML because it is a compressed version of the file. One of the more powerful features of KMZ is that it allows any images you use - say custom icons, or images in your descriptions - to be zipped up within the KMZ file.

After version 10, ArcGIS provides a tool for easy converting KML/KMZ file to ArcGIS layer and vice versa. The resulting output will be saved into a geodatabase.

We have two files, one is in KML format and the other is in KMZ format that contains UNBC campus roads and campus features collected from Google earth. We will convert them to ArcGIS spatial layer

The two files are **campus_roads.kml** and **campus_features.kmz** located at L:\labs\geog413\data\conversion\

Now take a look the KML and KMZ files in Google Earth first.

- On the local Linux machine, click Google Earth icon on the Desktop to launch Google Earth
- Click Add->Network Link from the top menu bar
- In the popped up window, set the Name to UNBC Campus Roads
- Click Browse button for Link and navigate to /home/labs/geog413/data/conversions/ and double click **campus_roads.kml** to add it in. Click OK.
- In the Google Earth, double click UNBC Campus Roads in the left panel to zoom into the area
- In the same way add UNBC Campus Features **campus_features.kmz**

- Start ArcMap, in ArcToolbox, click Conversion Tools->From KML->KML to Layer

    Input KML file: L:\labs\geog413\data\conversion**campus_roads.kml**
    Output Location: K:\geog413\data\conversion
    Output Data Name: *campus_roads*
- Click OK. A new layer will be added to ArcMap
- In the same way convert **campus_features.kmz** and save it as *campus_features* in your local folder

These data can be used together with other existing data

~~~~~~~~~~~~~ The End ~~~~~~~~~~~~~