GST 103: Data Acquisition & Management Lab Series

Lab 1: Reviewing the Basics of Geospatial Data

Document Version: 2013-07-31 (Beta)

Organization: Del Mar College
Author: Richard Smith

The development of this document is funded by the Department of Labor (DOL) Trade Adjustment Assistance Community College and Career Training (TAACCCT) Grant No. TC-22525-11-60-A-48; The National Information Security, Geospatial Technologies Consortium (NISGTC) is an entity of Collin College of Texas, Bellevue College of Washington, Bunker Hill Community College of Massachusetts, Del Mar College of Texas, Moraine Valley Community College of Illinois, Rio Salado College of Arizona, and Salt Lake Community College of Utah.

Contents

Introduction

This lab is part of a series of lab exercises designed through a grant initiative by the National Information, Security & Geospatial Technologies Consortium (NISGTC), funded by the United States Department of Labor in partnership with the Department of Education under the Trade Adjustment Assistance Community College and Career Training Grant Program (TAACCCT).

GIS data comes in many formats and many different structures. Data can be stored in structures such as an array in the case of raster data, a comma delimited file, or in a shapefile. These data structures can be organized and stored in a database or in a feature dataset within a geodatabase.

Your instructor may require that you provide screen captures and/or exported files. Please check with your instructor for the requirements specific to your class.

This lab includes the following tasks:

1. Review GIS Data Structures
2. GIS Data Organization

Objective: Explore and Understand Geospatial Data and Database Models

In this lab, we will be reviewing the various data formats of GIS data as well as looking at how to store the data efficiently. This lab will require the user to organize a large set of data.

Lab Settings

Required Virtual Machines and Applications

Windows Machine User Account	Train
Windows Machine User Password	Train1ng$

1 GIS Data Structures

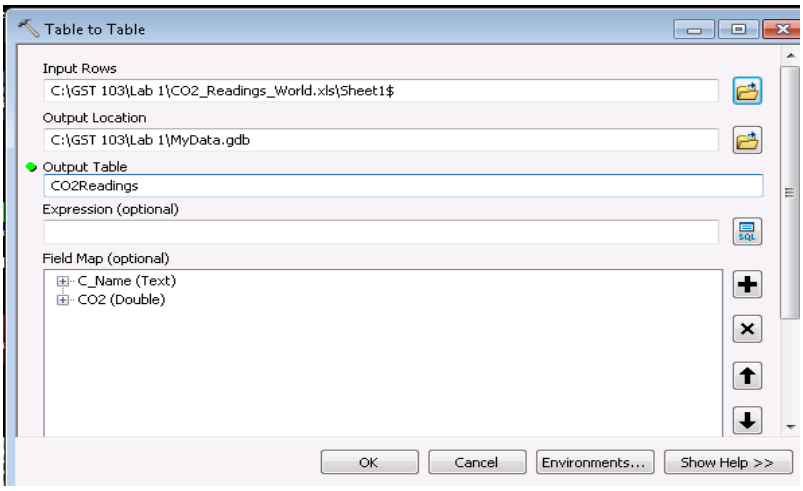
This task will be a review of some of the most commonly used GIS data types. Some of the data types cannot be "seen" in ArcCatalog but can be "seen" within ArcMap.

1. Log into the computer, using the information provided in the Lab Settings section.
2. Click **Start->All Programs->ArcGIS->ArcCatalog 10.1**. This will open ArcCatalog.
3. Expand your Catalog Tree for the D:\ drive by clicking on the plus sign next to the D:\ folder in the *Folder Connections*.
4. Connect to the *Lab 1* folder. Within the *Lab 1* folder are 7 data files that can all be read by the GIS.
5. Preview each data file using the Preview window. Do not close down ArcCatalog.
6. Click **Start->All Programs->ArcGIS->ArcMap 10.1**. This will open ArcMap. Open to a blank map.
7. Go to the **ArcToolbox** window in ArcMap and go to **Conversion Tools -> To Coverage**. One of the "invisible" files in your folder is an .e00 file that needs to be converted to a coverage before we can use it. Click on **Import from E00** and browse for the file called **clline.e00** and follow the steps to convert it.

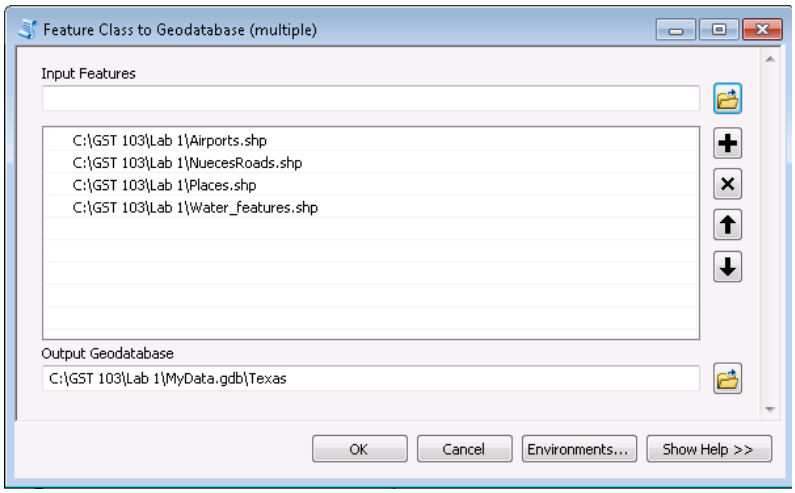
2 GIS Data Organization

In this task, we will look at ways to keep our data organized so we can avoid having our data in several folders for one project. The best way to organize data is to use a geodatabase. This allows for better access times to the shapefiles and keeps all the data in a container in the same place. When using a geodatabase, we store the data in a relational data structure. There are two types of geodatabases: personal and file. The personal database stores the data in an .mdb file also known as a Microsoft access database. This database can store a maximum of 2GB of data and can only be used by one person at a time. A file geodatabase is a database that can store large amounts of data with a large size limit on the data it holds. It runs faster than a personal geodatabase.

1. Open **ArcCatalog**
2. Find the *Lab 1* folder in the Catalog Tree and **right-click** on the folder to select **New->File Geodatabase**. Name the geodatabase **`MyData`**.
3. **Right-click** on your newly created geodatabase and click **New->Feature Dataset**. This will be a container for a set of data that is in the same coordinate system. It is a nice method to group the data we have. Name the feature dataset **`Texas`** and set the projection to **GCS_North_American_1983 (NAD 1983).** Click **Next**. We will not be using any 3D data so no vertical coordinate system is necessary. Click **Next**. The default XY tolerances are sufficient. Click **Finish**. We will be storing all the data in Texas in this feature dataset.
4. Create another feature dataset in you geodatabase and name it **USA** and set the coordinate system to **GCS_WGS_1984 (WGS 1984)**. No 3D data and default XY tolerances will suffice.
5. We are going to import the data into the feature datasets and into the geodatabase. We will start with the non-spatial data such as the .csv and .xls file. Right-click on the geodatabase and select **Import->Table (Single)** and import the .xls file. Set the **Input Rows** to the **Sheet1$** within the CO2_Readings_World.xls. Name the output table to **`CO2Readings`** (see screenshot below). Click **OK**.

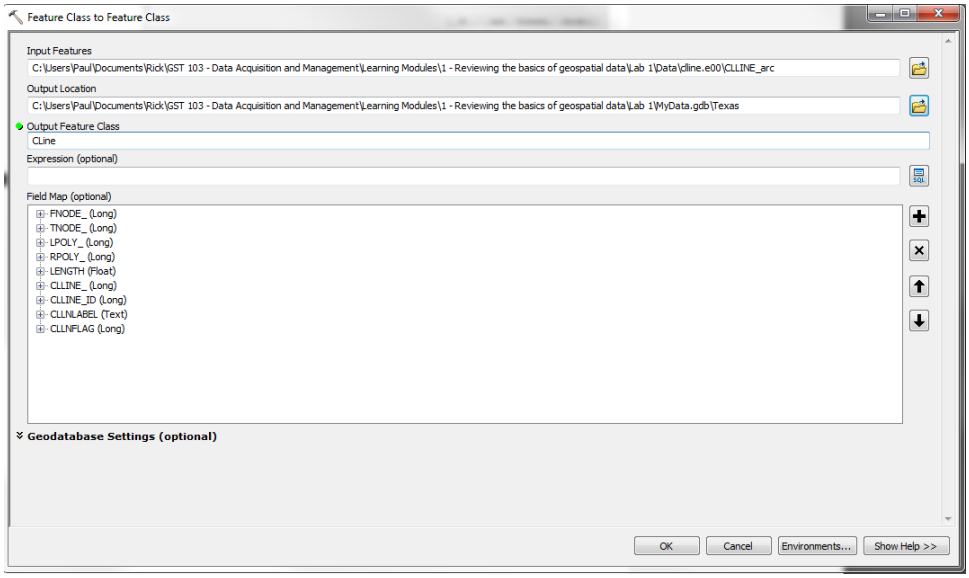


6. Do the same with the **Renewable_Energy_Percentages** and call the output **`Renewable_Energy`**.
7. Next, right-click on the **Texas** feature dataset and select **Import->Feature Class (Multiple)**. This is a batch process that will allow you to import more than one feature class into feature dataset. The data that will be selected for this will be **Airports.shp, Places.shp, NuecesRoads.shp, and Water_features.shp** (see screenshot below). (Note: if some of the features classes don't import initially, try them separately and the import may be successful the second time)



8. Next, import the **Counties.shp**, **Cities.shp** and the **Demographics.lyr** files into the **USA** feature dataset. (Notice the Demographics.lyr cannot be imported into a feature dataset because it is a layer file; however, the Counties and Cities are imported.)
9. Import the **World_Countries.shp** into the MyData geodatabase and name it **`Countries`**.
10. Import the raster by right-clicking on the geodatabase and **Import->Raster Datasets**. Set the **TX_terrain_hillshade_tsms.tif** file as the input and click **OK**.

11. Open **ArcMap** and go to the **Catalog** tab. We will now export the last dataset to our geodatabase. It is part of the converted .e00 file that could not be "seen" in ArcCatalog. Locate the *Lab 1* folder and find the clline coverage file you created. This file contains lines in Texas and hence it will be exported to the Texas feature dataset. Expand the clline file and and notice two feature classes. Right-click on the feature class called **arc** and select **Export->Geodatabase (single)**. Set the output location to the **Texas** feature dataset in the **MyData** geodatabase. Name the output feature class `Cline` (see the screenshot below). Click **OK**.



Conclusion

In this lab, we have looked at the datasets that we have been given and organized our data within our file geodatabase. Data organization is a valuable technique, especially when working with large amounts of data for different purposes. A geodatabase houses the data all centrally in one container.

Discussion Questions

1. Why do we want to use databases?

2. How does the use of feature datasets help to organize data? Provide an example.

3. What are the differences between a personal and file geodatabase?

