# GST 101: Introduction to Geospatial Technology Lab Series

# Lab 2: Exploring Geospatial Data Models

**Document Version: 2013-07-30**

**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).

In this lab, students will explore and manage geospatial data using a program called ArcCatalog. ArcCatalog is part of the ArcGIS Desktop 10.1 program suite created by Esri. ArcCatalog is similar to Windows Explorer for your geospatial data. ArcCatalog can view data, set properties of that data, manage the metadata for the data, and much more.

This lab contains quite a bit of "handholding" to help students get started with the ArcGIS Desktop software. It is important that students learn the concepts in this lab since the skills introduced will be required in order to perform the subsequent labs in this lab series.

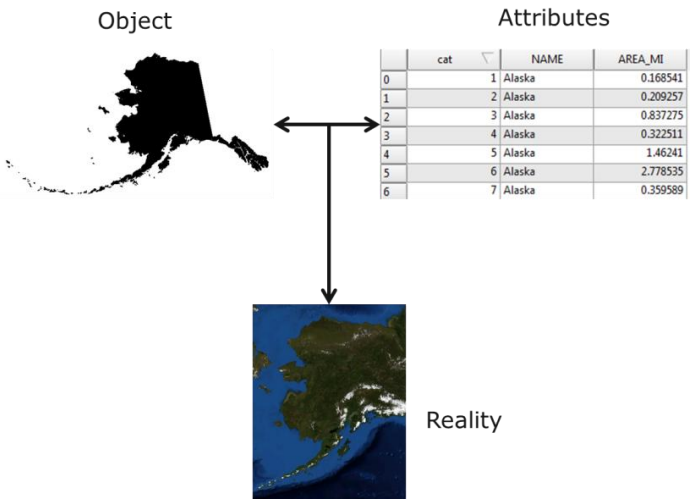
Your instructor may require that you provide screen captures and/or exported files. Please check with your instructor for requirements specific to your class.

This lab includes the following tasks:

1. Open ArcCatalog and Organize the Workspace
2. Become Familiar with Geospatial Datasets and Data Models

## Objective: Explore and Identify Geospatial Data Models

A geospatial data model is a formal means of representing spatially referenced information. A geospatial data model is composed of two parts: objects and attributes (shown in Figure 1) that, when combined, represent our model of reality.




| | cat | NAME | AREA_MI |
|---|---|---|---|
| 0 | 1 | Alaska | 0.168541 |
| 1 | 2 | Alaska | 0.209257 |
| 2 | 3 | Alaska | 0.837275 |
| 3 | 4 | Alaska | 0.322511 |
| 4 | 5 | Alaska | 1.46241 |
| 5 | 6 | Alaska | 2.778535 |
| 6 | 7 | Alaska | 0.359589 |

**Figure 1: Two parts of the geospatial data model**

There are three commonly used geospatial data models: vector, raster, and triangulated irregular network (TIN):

- **Vector Data Model** – best for modeling discrete objects. The three types of vector data are: point, line, and polygon.
- **Raster Data Model** – best for modeling continuous objects. A raster is composed of a matrix of contiguous cells, with each cell holding a single value.
- **Triangulated Irregular Network (TIN)** – best for modeling continuous objects with different sampling densities or when each measurement point needs to be preserved. A TIN is composed of a network of non-overlapping triangles that create a 3-dimensional surface.

Using one of the three geospatial data models above, reality can be modeled and stored in a computer. It is extremely important that the optimal data model be selected to model a facet of reality to make visualization, analysis, and queries clear, quick, and easy.

## Lab Settings

### Required Virtual Machines and Applications

| | |
|---|---|
| Windows Machine User Account | Train |
| Windows Machine User Password | Train1ng$ |

# 1 Open ArcCatalog and Organize the Workspace

The first step in working with geospatial datasets is to organize your workspace. It is important that the datasets are logically organized on the computer and are easy to find. In this task, you will obtain a copy of the lab data and explore how the data is organized in the folder.

1. Log in to the computer using the settings provided in the Lab Settings section of this lab.
2. Using Windows Explorer on the virtual machine, create a new folder on your C: drive named: `GST 101`.
3. Copy the *Lab 2* folder from the *Shared_Drive**GST101* to the *GST 101* folder you just created on the C: drive
4. Click **Start->All Programs->ArcGIS->ArcCatalog 10.1**. ArcCatalog will open. ArcCatalog is composed of 4 main parts (shown in Figure 2 ).
   a. **Menu Bar** - The menu bar runs across the top of the program (File, Edit, View...) and provides access to options, extensions, file operations, etc.
   b. **Toolbars** - The toolbars located below the menu bar allow for access to more specific functions. On the toolbars are buttons, dropdown boxes, and so on. These are grouped together logically/according to function.
   c. **Catalog Tree** - The left pane of the program is the Catalog Tree. This shows the listing of contents on your hard drive, network drive, flash drive, and so on.
   d. **Display Window** - The right, large pane of the program is the Display Window. This is where you can view the contents of a directory, geospatial data, and metadata.

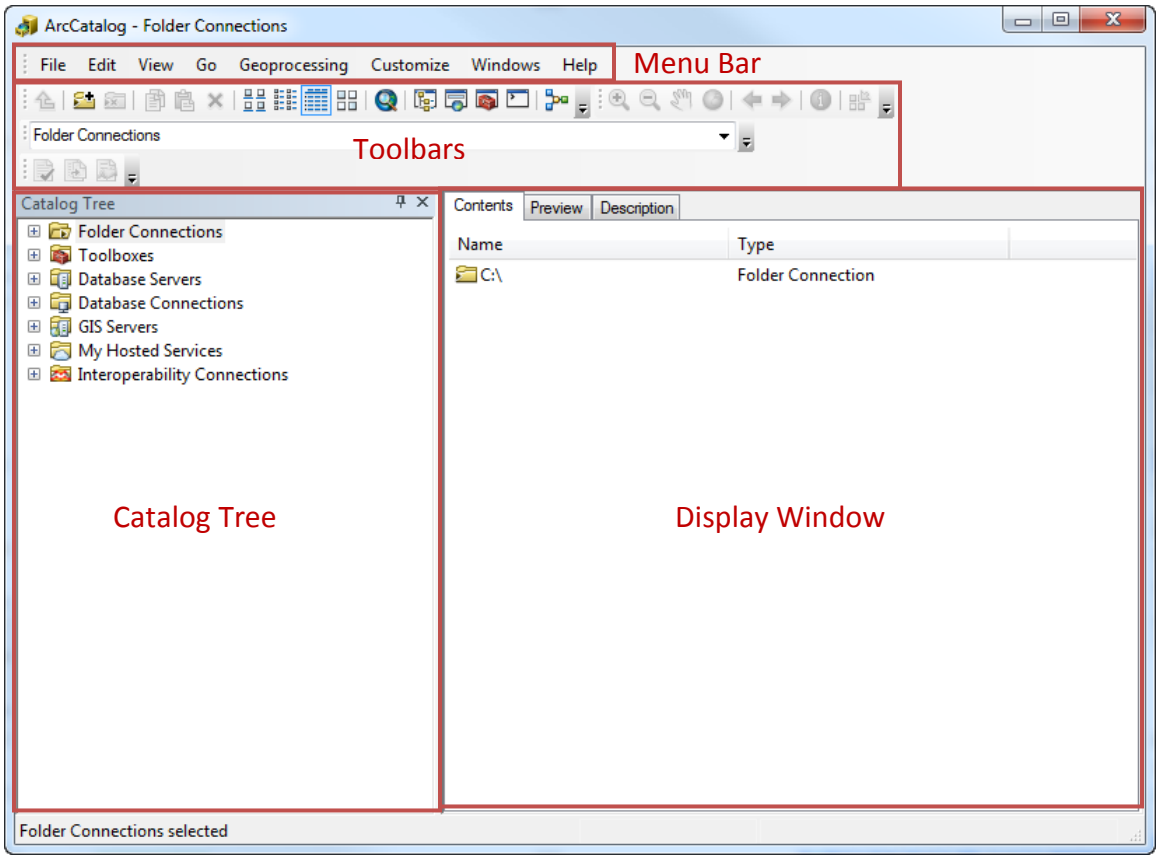



**Figure 2: Four main parts of ArcCatalog**

5. Let's make sure we have the appropriate toolbars turned on for this lab. In the menu bar, click **Customize->Toolbars** and make sure that there are checkmarks next to: **Main menu, Geography, Location, and Standard.** If they are not checked (turned on), click on them to turn them on. If you have to turn on any toolbars, you may see them floating on your screen. You can grab the toolbar and dock it to the ArcCatalog window by dragging the toolbar next to other toolbars and letting go. You can rearrange other toolbars as you see fit in the same manner.
6. Look in the Catalog Tree. You should see a tree containing multiple entries, one of which should be your computer's **Folder Connections**. Click the plus sign next to the Folder Connections to expand. If you do not see a connection to your C:\ drive, create one by clicking on the **Connect To Folder** button , then browsing to your C:\ drive and clicking **OK**.
7. Click on the plus sign next to the C:\ folder to expand. You will now see all of the subfolders directly under the C:\ folder.
8. Expand the *GST 101/Lab* 2 folder that you created in step 2. Click on the *Lab* 2 folder to display the contents in the display window. When you click on a folder, the display window will display the contents of that folder.
9. Collapse the *C:*\ folder by clicking the minus sign.

Data is often stored deep inside a hierarchy of folders. It is often tedious and time consuming to navigate deep inside the folders to gain access to the data. ArcCatalog provides a way to create a shortcut directly to any folder so that you have one-click

access. ArcCatalog refers to these shortcuts as *connections*. A connection is a link to a drive or folder. You will now create a connection to the *Lab 2* folder so you will have direct access to the folder without having to navigate through the parent directories.

10. Click **File->Connect To Folder…** navigate to, and select *C:\GST 101\Lab 2* from the folder listing and click **OK**. You will now see a connection to your folder for this lab. Your catalog tree should look like the figure below.

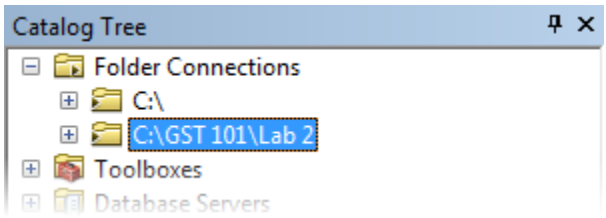



**Figure 3: Folder Connection to C:\GST 101\Lab 2**

## 2 Become Familiar with Geospatial Datasets and Data Models

Now that you have a connection to your *Lab 2* folder, let's take a look inside and see what we have.

1. Select the *Lab 2* connection in the catalog tree. The display window shows the data inside that folder. You should see 7 files.
   a. If you do not see 7 files, right-click on the *Lab 2* folder in the catalog tree, and choose **Refresh.**
2. Take a moment to read the names of the files and the file types. Below is an explanation of the icons and types you see.
   a.  **Shapefiles**
   Shapefiles are a file format that contains data in the vector geospatial data model. Each shapefile can only contain one feature type (point/line/polygon) at a time. From left to right, the icons represent point, line, and polygon. A shapefile is actually a collection of files on the disk with a common name, but different extensions.
   b.  **Raster**
   Rasters are raster data files. Many different formats of the raster geospatial data model exist. Rasters may be a single file, or a collection of files on disk.
   c.  **Triangulated Irregular Network (TIN)**
   TINs store surface information, such as elevations using a set of nodes and triangles. A TIN is a collection of files on the disk contained in a folder.

3. Let's take a look at the data in Windows Explorer to understand how files are stored on the computer versus how they display in ArcCatalog.
   a. Minimize ArcCatalog and open **My Computer** and navigate to your *Lab 2* folder. You will see three folders and a number of files. Let's examine the Airports shapefile. Switch back to ArcCatalog and find the *Airports.shp* file in the display window. You will see that it is a point shapefile. Now, switch back to Windows Explorer and find the files named *Airports*. Note that there are 6 files named *Airports* but they have different file extensions. All 6 of these files compose a shapefile.
   b. Now find the *CCBoundary* shapefile. Notice that this shapefile is composed of 7 files! So, how can one shapefile only be 5 files and another be 7 files? Not all of the files are required. In fact, only 3 files are required to compose a shapefile. The other files add features.
   c. In ArcCatalog, access ArcGIS Desktop Help (**Help->ArcGIS Desktop Help**) and search the terms "shapefile file extensions" to learn more about shapefile file extensions.
4. In ArcCatalog, find the *ccelevation* raster file in the Lab 2 folder. Look back in Windows Explorer and find the two folders named *ccelevation* and *info*. These two folders contain the files that compose the raster dataset.
5. Go back to ArcCatalog again, and find the TIN dataset. Note the name. Go back to Windows Explorer and look at the files and folders. The rasttin folder contains the files that compose the TIN dataset.

Now that you have seen how complex shapefiles, rasters, and TINs are, it is important to note that:

**Geospatial datasets should be moved, copied, renamed and deleted through ArcCatalog only.**

If you attempt to move, copy, rename, or delete these files through Windows Explorer, you leave yourself open to error by missing some files or even corrupting other datasets.

## 2.1 Viewing Vector Geospatial Datasets

Now that you know how the geospatial datasets are stored on your computer, let's see what the data they contain looks like.

1. In ArcCatalog, once again, navigate to your *Lab 2* folder. Click on *Airports.shp* in the display window. Now, click the **Preview** tab above the display window. You will see a single dot on your screen. This tells us that there is only 1 feature in this shapefile…but what is it? That is where the attribute table comes in. Let's take a look at it.
2. At the bottom of the display window, you will see the **Preview dropdown box**. Since Geography is selected, we see the geometry stored in the shapefile. Now, **Click the Preview dropdown box and select Table**. The attribute table associated with the Airports shapefile is now displayed. Scroll to the right and read the column headers and feature's attributes. Can you determine which airport is in this shapefile?
3. So, from the Attribute table, you should have found that it is the Athens/Ben Epps airport in Athens, GA. So the attributes of the feature helped us determine what we were looking at and where it is located. But, what if we weren't familiar with the Athens area? How would we know which Athens this airport is in? Greece, Georgia, Ohio? This is where Metadata comes in to save the day.
4. **Click the Description tab** above the display window. Descriptive information about the airports dataset is now at your fingertips, all of which describe the data inside the shapefile.
5. Take a few moments and read the metadata for the *Airports* shapefile.

So, now you know how to find geospatial datasets, preview them, and read the metadata about them. Now, let's rename a dataset, navigate around the preview section, and create a thumbnail for the metadata.

1. Find the shapefile named ***Roads.shp***. Look at the metadata for this shapefile. From the metadata description, we see that this dataset contains roads for Athens-Clarke County in Georgia. Now that we know where these roads are, let's make the filename a little more descriptive so it is easier for us to identify what is contained in the shapefile at first glance in the Catalog Tree.
2. In the Catalog Tree, **right-click on the Roads.shp** shapefile and **select Rename**. Rename the file to something more descriptive. (i.e. **`Athens-Clarke County Roads.shp`**). Don't be afraid to make the filename long; the more descriptive the better.
3. Now, let's look at the roads and navigate around a little. **Click the Preview tab** to see the roads in Athens. **Locate and Select the *Zoom In* Tool**  on the Geography toolbar. Your mouse pointer will now be a magnifying glass when you move the pointer over the display window.
4. **Click to zoom in** on the roads in Athens. You can also click and drag a box around an area to zoom into that area.
5. Locate and Select the **Pan** Tool  on the **Geography** toolbar. Click and drag your mouse pointer to move around the roads.

6. Locate and Select the **Zoom Out** Tool  on the Geography toolbar. Click to zoom out. If you get lost, use the Full Extent button on the Geography toolbar to zoom out to the original view.
7. **Zoom to the Full Extent of the dataset by clicking on the Full Extent**  **icon** on the Geography toolbar**.** Let's make this view the thumbnail that displays in the Metadata.
8. **Click the *Create Thumbnail***  **icon** on the Geography toolbar. Now switch to the Metadata view and see that the thumbnail has been created and displays the area you were zoomed in to. You can always overwrite a thumbnail by simply creating a new one.

## 2.2 Viewing a Raster Dataset

Now that you are familiar with vector geodatasets, let's take a look at a raster.

1. In ArcCatalog, once again, navigate to your *Lab 2* folder. **Click on ccelevation** in the Catalog Tree. If you still had the Preview tab selected, you will be presented with a dialog box asking to create pyramids. If you do not have the Preview tab selected, select it now. Pyramids allow for rapid display of raster data at different resolutions. Generally, it is preferred that you do create pyramids. **Click Yes** to create pyramids. After a few moments wait, the raster will display.
2. **Zoom in really close to a part of the raster**. Can you see the pixels that make up the surface? **Zoom back to full extent**.
3. This is a Digital Elevation Model (DEM) of the Athens-Clarke County area. A DEM is a continuous surface where each cell represents an elevation value. What does it all mean?
4. Let's find out what the dark areas represent and what the light areas represent. On the **Geography** toolbar, **find and select the Identify Tool**. **Select the darkest area of the raster** you can find. An *Identify Results* box will display. It tells you the X,Y coordinate that you clicked at, and the Stretched and Pixel value at that location.
    a. The *Stretched* value represents the color value between 0 and 255; 0 being black and 255 being white. This has nothing to do with the elevation; it just represents the color displayed on the screen to you.
    b. The *Pixel* value represents the elevation in feet at that point on Earth.
5. Now, **select the lightest area of the raster** you can find. Observe the *Pixel* value again. The pixel value of the lightest area should be much larger than the darkest pixel value. This is because the darker colors on *this* DEM represent lower elevation values, and lighter values represent higher elevation values.
6. Close the **Identify Result** box.

## 2.3 Viewing a TIN Dataset

1. In ArcCatalog, once again, navigate to your *Lab 2* folder. **Click on rasttin** in the Catalog Tree. **Preview the dataset.** You will see a colorful surface. This, too, is a DEM of Athens, but represented using the TIN geospatial data model instead of a raster geospatial data model.
2. **Zoom in really close to a part of the TIN**. Can you see the triangles that make up the surface? **Zoom back to full extent**.
3. **Use the Identify tool** and click on a blue area of the TIN (previously black on the raster). The identify box will display again but this time, there are 5 attributes associated with this location on the TIN. The *Elevation* attribute contains the elevation at that point.
4. **Note what the elevation is for the identified area. Preview the raster dataset again and click on the same (or close) spot. Compare the elevations**. The elevation values should be the same, or very similar. Both geospatial datasets are representing the same continuous spatial phenomenon (elevation) but using two different geospatial data models (raster and TIN).

## Conclusion

In this lab, you have explored datasets that use the three common geospatial data models: vector, raster, and TIN. ArcCatalog is used to manage, preview, and explore geospatial datasets. It is always important to manage the geospatial datasets using ArcCatalog, since each dataset is actually composed of multiple files on disk. If you do not move every file on disk together, you risk corrupting the geospatial datasets.

## Discussion Questions

1. What are the 14 possible file extensions for files that compose a shapefile? What do these files contain? Which files are required?

2. Why should folder connections be used in ArcCatalog?

3. In the *Lab 2* folder, explore Parcels.shp using ArcCatalog. After exploring, can you determine how many parcels are in Athens-Clarke County Georgia? (hint: preview the attribute table, the answer is below the table. Include everything except for Roads_Row).