# GST 103: Data Acquisition and Management

## Lab 2 - Setting Up a Project Database

### Objective – Learn How to Normalize Data and Import It into a SpatiaLite Database

Document Version: 4/26/2015

**FOSS4G Lab Author:** Kurt Menke, GISP Bird's Eye View GIS

**Original Lab Content Author:** Richard Smith, Ph.D., GISP Texas A&M University - Corpus Christi

The development of the original document was 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. This work is licensed under the Creative Commons Attribution 3.0 Unported License. To view a copy of this license, visit http://creativecommons.org/licenses/by/3.0/ or send a letter to Creative Commons, 444 Castro Street, Suite 900, Mountain View, California, 94041, USA.

This document continues to be modified and improved by generous public contributions.

### 1. Introduction

There are two main data models for GIS data: vector and raster. Additionally, GIS data comes in many file formats. When gathering data for a project, it is common to acquire data from several sources. Therefore, it is also common for the data to be in several different file formats. In this lab you will create a project geodatabase for the Gifford-Pinchot National Forest in Washington State. First, you will normalize the data. This means that you will put all datasets in the same coordinate reference system (CRS) and clip them to the study area boundary. Lastly, you will put them all into the same file format: a SpatiaLite geodatabase.

This lab includes the following tasks:

* Task 1 Investigate and Normalize Project Data
* Task 2 Create a New Database
* Task 2 Populate the New Database

### 2 Objective: Learn How to Normalize Data and Import It Into a SpatiaLite Database

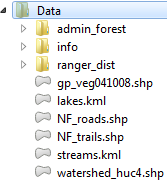
The objective of this lab is to explore and understand geospatial data models and file formats.

### Task 1 Investigate and Normalize Project Data

In this task, you will familiarize yourself with the lab data and will begin to normalize the data.

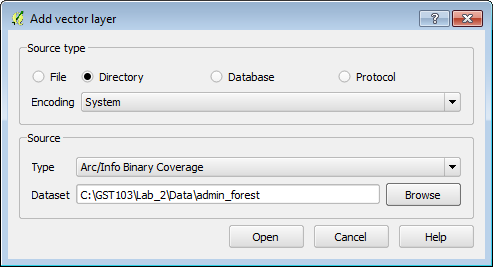
1. Open QGIS Browser, navigate to and expand the Lab 2 Data folder.

There are eight vector layers here. There are four shapefiles, two KML files, and two coverages (in the figure below). Each of these file formats will be treated in different ways. (*Note*: when there are multiple coverages in the same workspace they share the same info folder.)



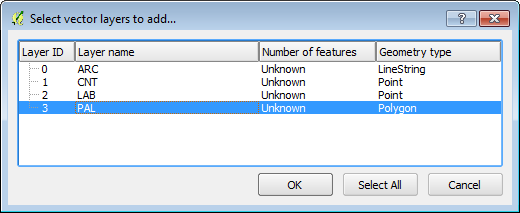
Data Layers in QGIS Browser

1. All the data for this project will need to be in UTM, Zone 10, NAD83. Unfortunately, not all of the data are in that coordinate system, so we will convert all data to UTM Zone 10, NAD83.
2. Open QGIS Desktop.
3. Click the Add Vector Layer button and change the Source type to Directory. Then change the Type to Arc/Info Binary Coverage.
4. Browse to the Lab Data folder and select the admin\_forest folder. Click Select folder (in the figure below). Click Open.



Adding a Coverage to QGIS

1. In the Select vector layers to add window choose 3 PAL (Polygon layer). Click OK.

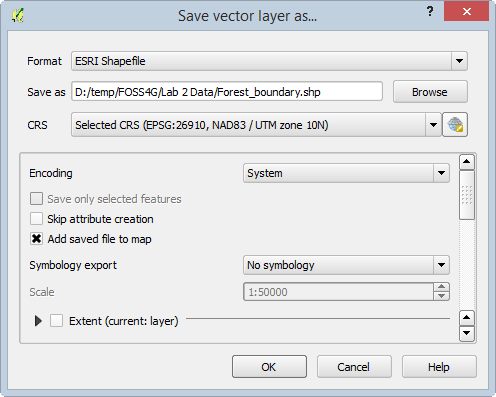


Adding\_the\_Coverage\_Polygon\_to\_QGIS

1. Save your project as Lab 2.qgs.

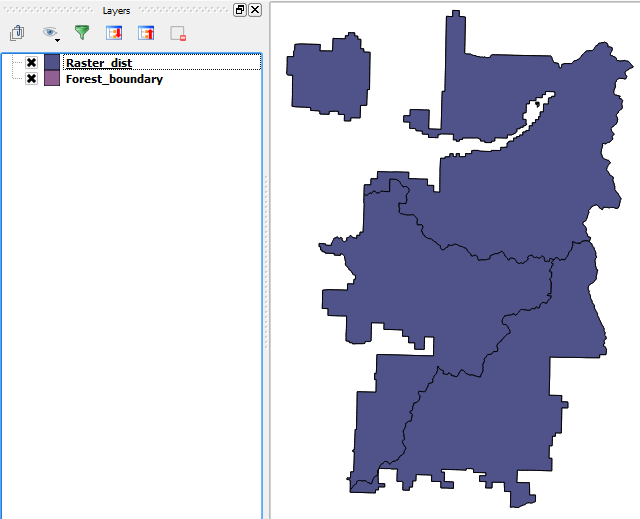
This layer is in a custom Albers Equal Area coordinate system. Since it is in a custom CRS, there is no EPSG code to use during import into a database. Therfore, we will first save this out to a shapefile in the desired CRS.

1. Right-click on the layer in the Layers panel and choose Save as…
2. Save the resulting dataset as a shapefile named Forest\_boundary.shp to the Lab Data folder.
3. Set the CRS to UTM Zone 10 NAD83 (EPSG 26910) (completed dialog shown in figure below).



Saving the Covererage out to a Shapefile in UTM

1. Click OK to save the new shapefile and add it to the map.
2. Remove the PAL layer from the map as we no longer need it.
3. Export the Ranger\_dist coverage PAL polygon to a shapefile with the same coordinate system as the Forest\_boundary layer. Once this has been completed, thet map should resemble the figure below.



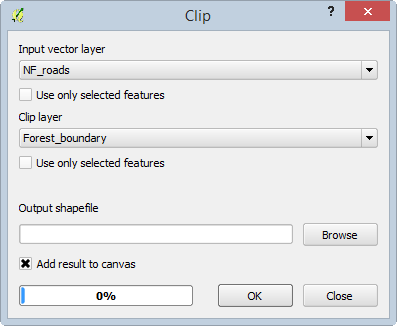
Shapefile Versions of Both Coverages in UTM

The only other dataset in Albers Equal Area is the vegetation shapefile.

1. Add the gp\_veg041008.shp shapefile to QGIS Desktop and save this as a new shapefile in UTM Zone 10 NAD83.
2. Add the NF\_roads, NF\_trails, and watershed\_huc4 shapefiles to QGIS Desktop.

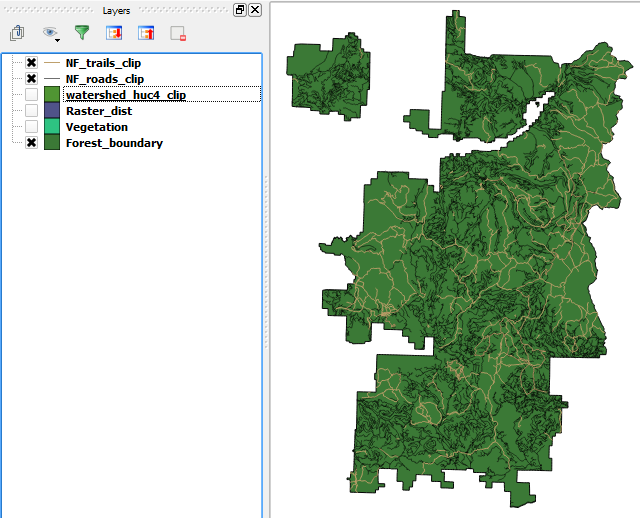
These last three layers are all shapefiles in the correct CRS. However, they extend beyond the forest boundary. Let's clip the layers to within the forest boundary.

1. Use the Vector | Geoprocessing Tools | Clip tool to clip the roads, trails, and watershed to the forest boundary (example clip parameters shown in the figure below). You can give them the same output name, but end it with clip. For example, NF\_roads will become NF\_roads\_clip.



Clipping Roads to the Forest Boundary

1. Remove the original unclipped roads, trails, and watershed layers once the three clip operations are complete. Your map should now resemble the figure below (note that not all layers are visible in the figure below).



Normalized Layers

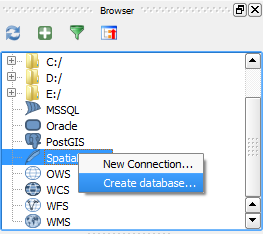
You have taken the initial steps to normalize the data. There are two more layers that we will put in the SpatiaLite database, rivers and lakes, however, KML files are always in a geographic CRS of WGS 1984 (EPSG:4326). These KML files can be repojected when importing into the SpatiaLite database without having to be converted to another format.

1. Save your project.

### Task 2 Create a New Database

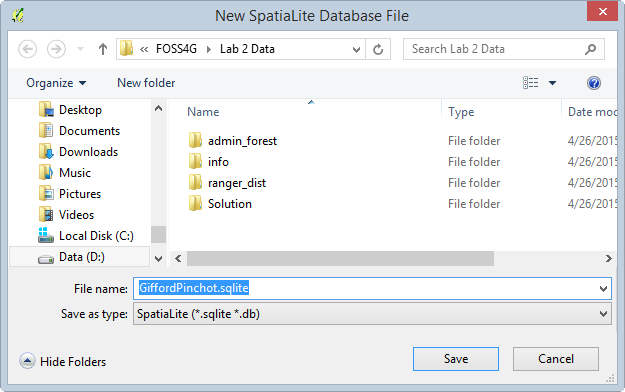
Now that you have taken the initial steps to prepare your data, you will create a new empty SpatiaLite database that you will import your datasets into.

1. Open your Lab2.qgs project in QGIS Desktop if it not already open.
2. Find the Browser panel. If your Browser panel is not visible, click View | Panels | Browser to turn it on.
3. Find the SpatiaLite database connection below your hard drives. Right-click on it and choose Create database (figure below).



Create New Database Context Menu

1. Navigate to the Lab Data folder and name the new database GiffordPinchot.sqlite and click Save (figure below).



Naming Database

1. A dialog will appear stating that the database has been created. Click OK to dismiss.

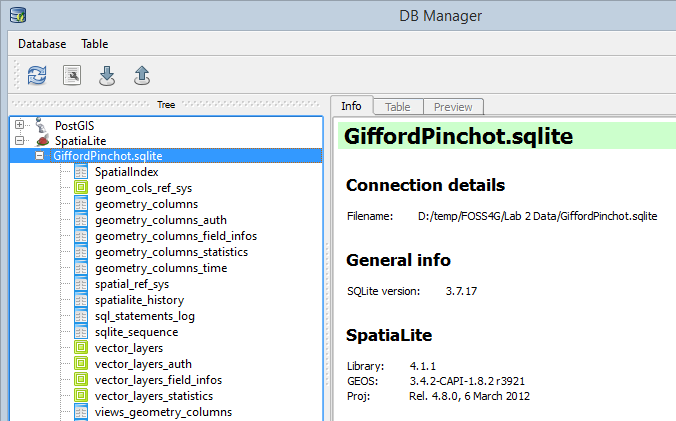
The database will now appear under the SpatiaLite database connection.

1. Save your project.

### Task 3 Populate the New Database

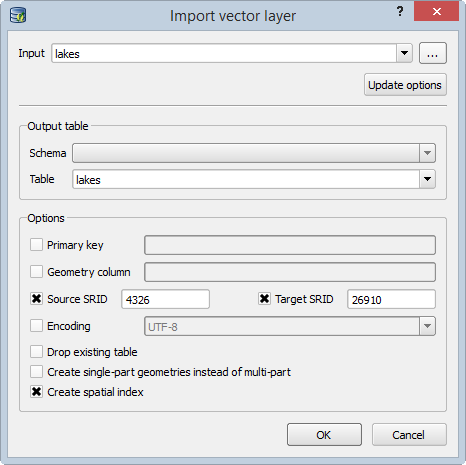
Now you will populate the SpatiaLite database with the eight layers.

1. Open your Lab2.qgs project in QGIS Desktop if it is not already open.
2. Add the stream and lake KML layers to QGIS Desktop.
3. From the menu bar choose Database | DB Manager | DB Manager. Expand the SpatiaLite database connection . You will see the GiffordPinchot.sqlite database. If you expand the database, you will see many tables but no GIS layers yet (shown in figure below).



DB Manager

1. First, you will load the streams and lakes layers. Since these are KML, they are in a geographic CRS with an EPSG code of 4326. This is the case for all KML datasets.
2. Click the Import layer/file button Import layer file button .
3. Set up the Input vector layer window as follows (shown in figure below):
   1. Set the Input as lakes
   2. Name the table lakes
   3. Check Source SRID and enter 4326
   4. Check Target SRID and enter 26910
   5. Check Create spatial index
   6. Click OK

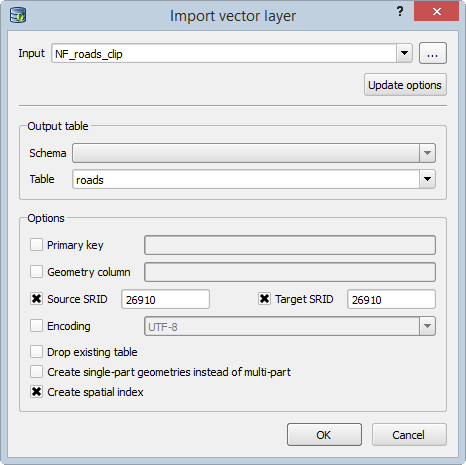


Import KML Lakes into SpatiaLite DB

1. Once the operation has completed successfully click the Refresh button Refresh button to see the lakes layer in the database.
2. Repeat the previous steps to import the streams KML.

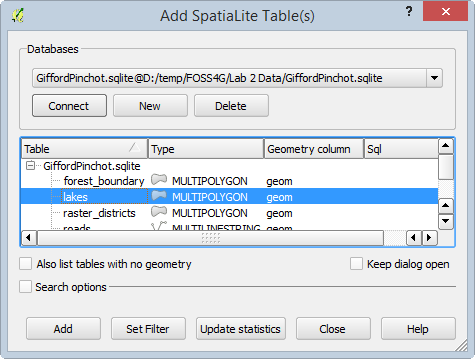
Streams and Lakes were the final two layers that required a CRS reprojection. The remaining six UTM layers can now be imported. The only change is that both the input and target SRID’s will be 26910 (UTM Zone 10 NAD83).

1. Click the Import layer/file button Import layer file button .
2. Set up the Input vector layer window as follows (shown in figure below):
   1. Set the Input as NF\_roads\_clip
   2. Name the table roads
   3. Check Source SRID and enter 26910
   4. Check Target SRID and enter 26910
   5. Check Create spatial index
   6. Click OK



Import Shapefiles into SpatiaLite DB

1. Repeat the above steps for trails, watersheds, ranger districts, vegetation, and forest boundary layers.
2. Now that all eight layers have been imported you can remove the original layers from the map.
3. The layers in the database can be added via the DB Manager or the Add SpatiaLite Layer button.
4. If using the DB Manager, right click on a layer and choose Add to canvas.
5. If using the Add SpatiaLite Layer button SpatiaLite layer button, select the database and click the Connect button. Once the layers appear, you can select them and click Add to add them to QGIS (figure below).



Add SpatiaLite Tables

### 3. Conclusion

In this lab, you took data in several different file formats and CRSs and normalized them. They are all now in the same CRS, clipped to the forest boundary, and in a geodatabase. This methodology has the benefit of creating a working copy of the data. The raw data still exist. Therefore, if you accidentally delete or corrupt a dataset, you still have the original files. Additionally, the data now all reside in a tidy database. Since they are all in the same CRS, you can run any geoprocessing or analysis tools against them knowing they are all in UTM Zone 10 NAD83.

### 4 Discussion Questions

1. What are the steps involved in setting up a SpatiaLite database?
2. What are the advantages of normalizing project data?