****

**QGIS Lab Series**

**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:** **2014-08-15 (Final)**

**Author:**

Kurt Menke, GISP

****

**Contents**

[1 Introduction 2](#_Toc394677115)

[2 Objective: Learn How to Normalize Data and Import It Into a SpatiaLite Database 2](#_Toc394677116)

[3 How Best to Use Video Walk Through with this Lab 2](#_Toc394677117)

[Task 1 Investigate and Normalize Project Data 2](#_Toc394677118)

[Task 2 Create a New Database 8](#_Toc394677119)

[Task 3 Populate the New Database 10](#_Toc394677120)

[5 Conclusion 13](#_Toc394677121)

[6 Discussion Questions 14](#_Toc394677122)

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

1. 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.

1. How Best to Use Video Walk Through with this Lab

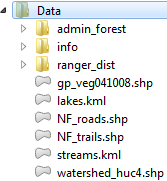
To aid in your completion of this lab, each lab task has an associated video that demonstrates how to complete the task. The intent of these videos is to help you move forward if you become stuck on a step in a task, or you wish to visually see every step required to complete the tasks.

We recommend that you do not watch the videos before you attempt the tasks. The reasoning for this is that while you are learning the software and searching for buttons, menus, etc…, you will better remember where these items are and, perhaps, discover other features along the way. With that being said, please use the videos in the way that will best facilitate your learning and successful completion of this lab.

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. The data for this lab is located in: ***GST103\Lab\_1\Data****.*
2. **Open QGIS Browser, navigate** to the **Lab\_2\Data** folder and select the **Data** folder so that you see the contents in the **Param** tab.
3. There are eight vector layers here. There are four shapefiles, two kml files and two coverages (admin\_forest and ranger\_dist) (**Figure 1**). 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.)



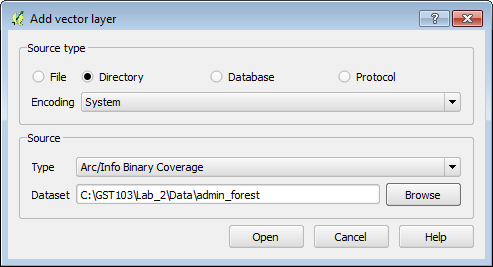
**Figure 1: Data Layers in QGIS Browser**

1. All the data for this project will need to be in UTM, zone 10, NAD83. Begin by identifying the coordinate reference system of each layer. Investigate the data layers and complete Table 1 below. This will tell you which layers will need to be saved to a new coordinate reference system.

|  |  |
| --- | --- |
| **LAYER** | **COORDINATE REFERENCE SYSTEM** |
| admin\_forest |  |
| ranger\_dist |  |
| gp\_veg041008 |  |
| lakes |  |
| NF\_roads |  |
| NF\_trails |  |
| streams |  |
| watershed\_huc4 |  |

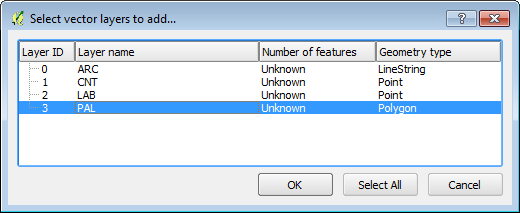
**Table 1: Coordinate Reference Systems of Lab Data**

1. **Open QGIS Desktop**. **Click** the **Add vector layer** button and change the **Source type** to **Directory.** Then **change** the **Type** to **Arc/Info Binary Coverage**. **Browse** to the **Data** folder and select the **admin\_forest** folder. **Click** **Select folder** (**Figure 2**). **Click Open.**



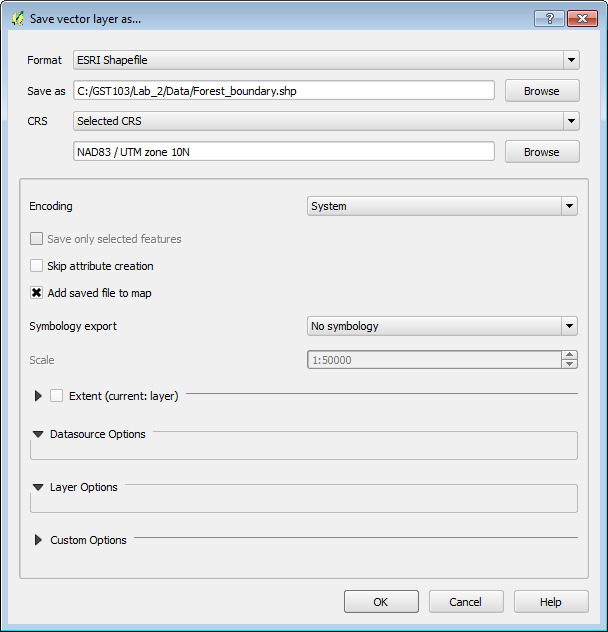
**Figure 2: Adding a Coverage to QGIS**

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



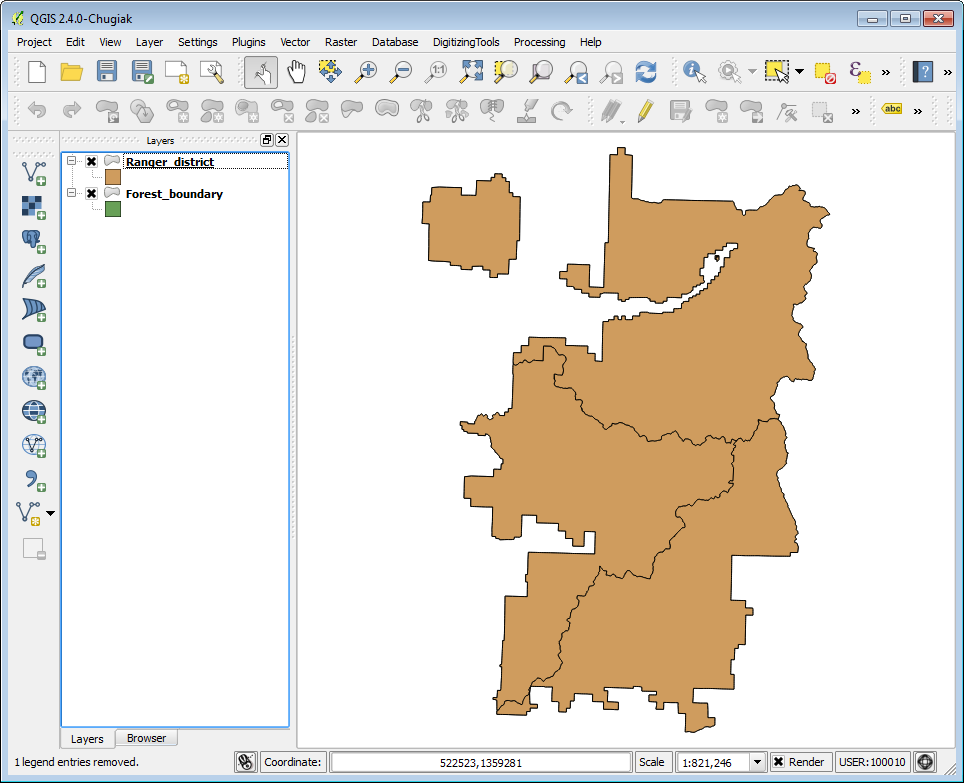
**Figure 3: Adding a the Coverage Polygon to QGIS**

1. **Save** your project as **Lab 2.qgs.**
2. This layer is in a custom Albers Equal Area coordinate system. Since they are in a custom CRS there is no EPSG code to use during import into a database. First, you will save this out to a shapefile in the desired CRS. **Right click** on the layer and choose **Save as**… Save the resulting dataset to the **Data** folder and set the CRS to UTM zone 10 NAD 83 (EPSG 26910). (**Figure 4**). Once this is done you can **Remove** **the PAL** layer from **QGIS Desktop**.



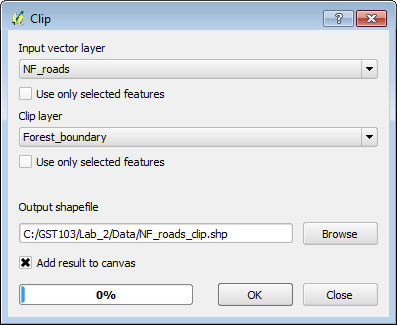
**Figure 4: Saving the Coverage out to a Shapefile in UTM**

1. **Repeat steps 5-7** for the **Ranger District** coverage. Once this has been completed, QGIS Desktop should resemble **Figure 5**.



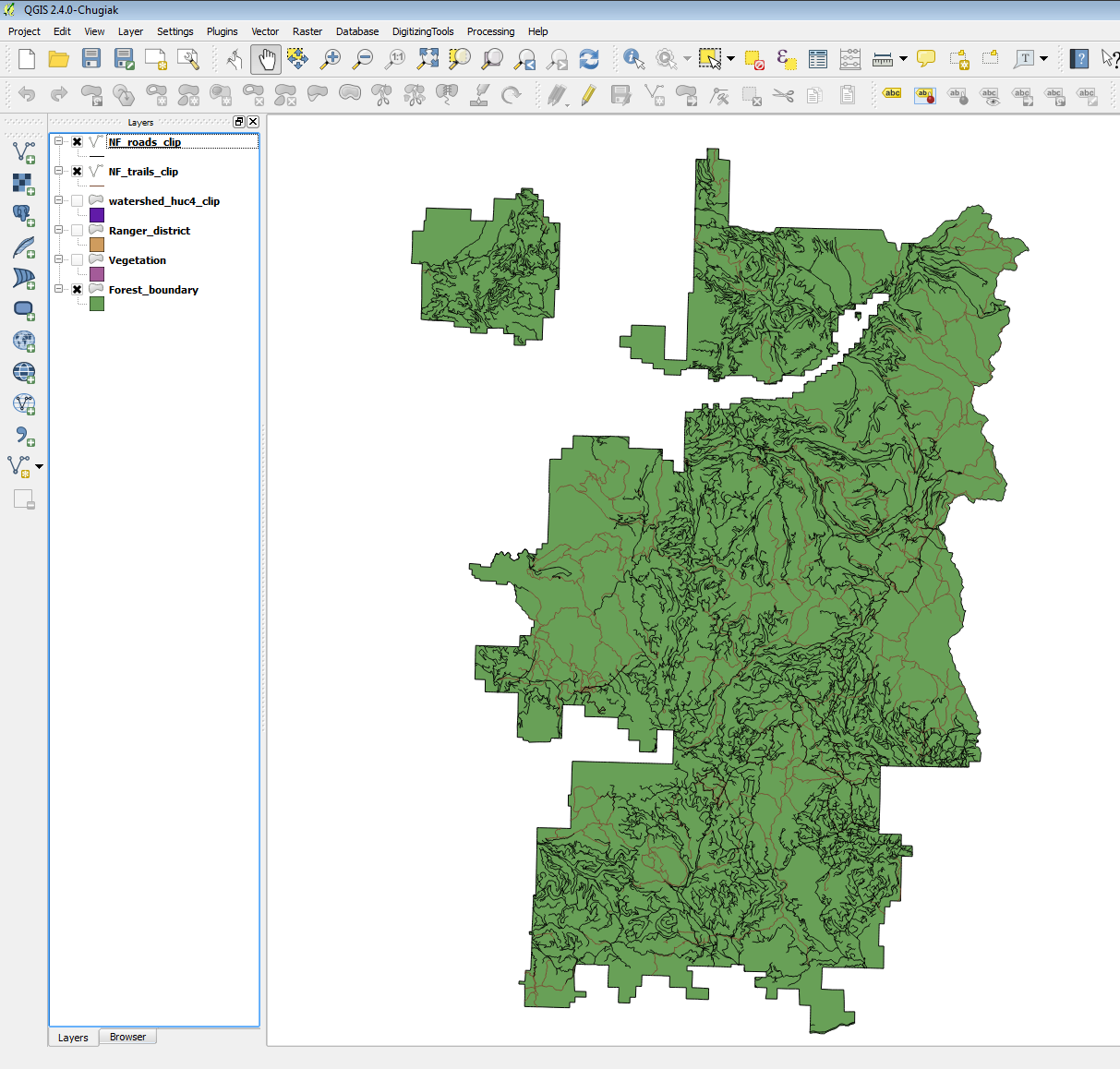
**Figure 5: Shapefile Versions of Both Coverages in UTM**

1. The only other dataset in Albers Equal Area is the vegetation shapefile. **Add** the **gp\_veg041008.shp** shapefile to **QGIS Desktop** and **save** this as a **new shapefile in UTM**.
2. **Add** the **NF\_roads, NF\_trails** and **watershed\_huc4** shapefiles to **QGIS Desktop**. These layers are all shapefiles in the correct CRS. However, they extend beyond the forest boundary. Use the **Vector 🡪 Geoprocessing Tools 🡪 Clip** tool to clip each to the forest boundary (**Figure 6**). You can give them the same output name, but end it with clip. For example, NF\_roads will become NF\_roads\_clip.



**Figure 6: 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 **Figure 7.**

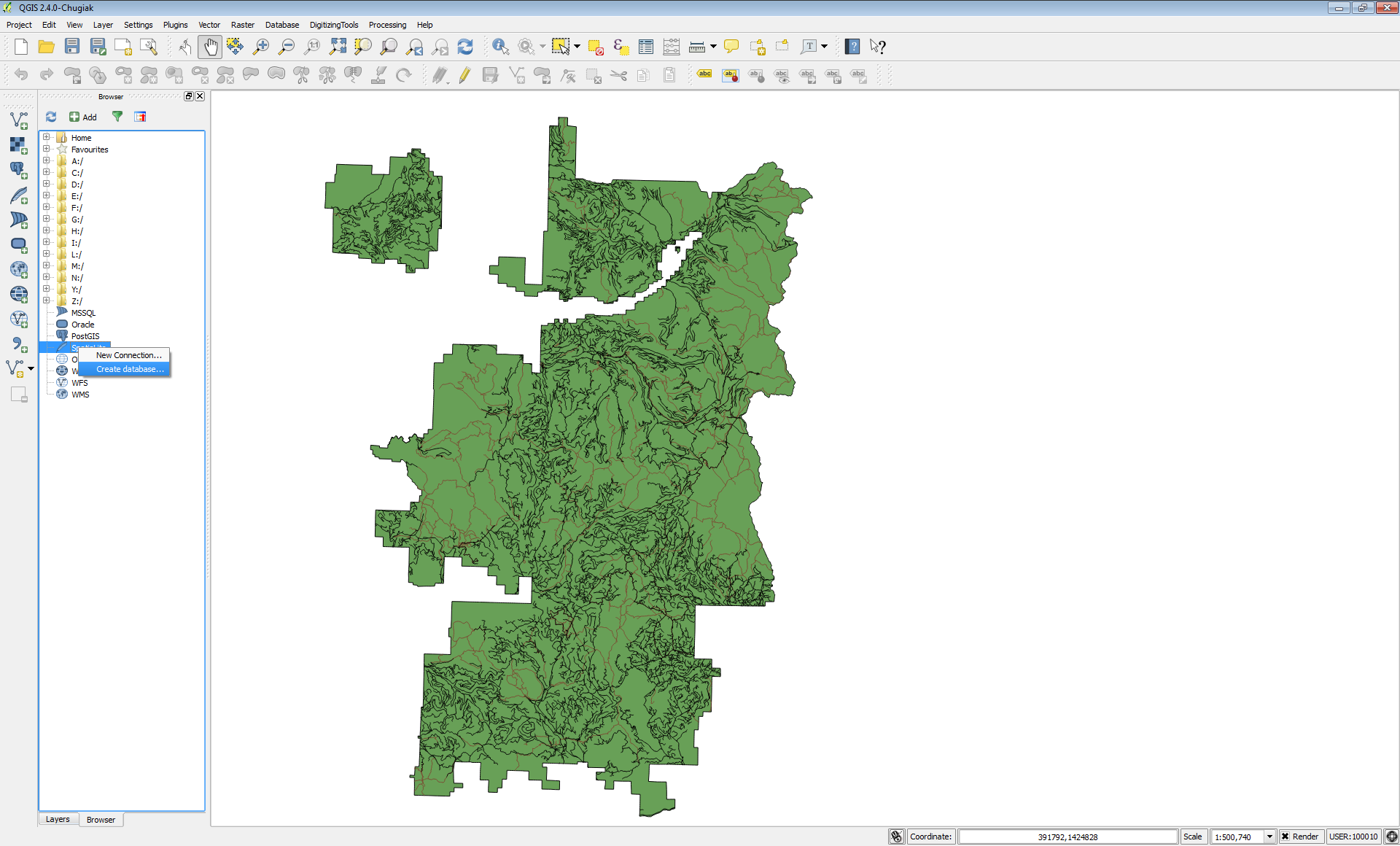


**Figure 7: Normalized Layers**

1. You have taken the initial steps to normalize the data. The rivers and lake layers are KML. KML is always in a geographic CRS with an EPSG code of 4326. These can be repojected when importing into the SpatiaLite database.
2. **Save your project**.
3. Create a New Database

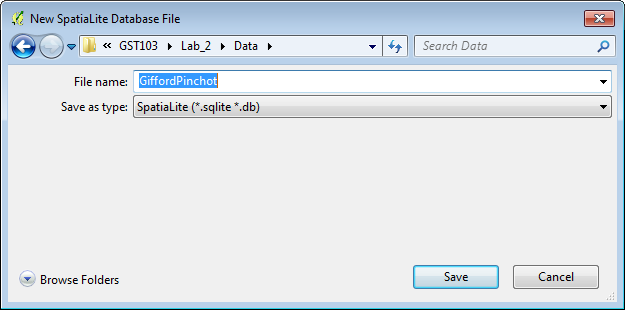
Now that you have taken the initial steps to prepare your data, you will create a new empty SpatiaLite database in which you will import your datasets.

1. **Open** your **Lab2.qgs** project in **QGIS Desktop** if it not already.
2. **Click** on the **Browser** tab at the bottom of the Table of Contents. If your Browser tab is not there go to **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 8**).



**Figure 8: Browser Tab 🡪 SpatiaLite 🡪 Create New Database**

1. **Navigate** to the **Data** folderand name the new database **GiffordPinchot and click Save (Figure 9).**

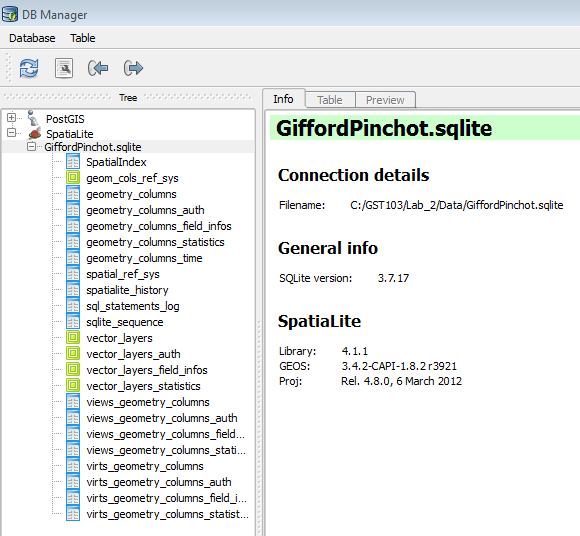
****

**Figure 9: Naming Database**

1. The database will now appear under the SpatiaLite database connection. **Switch** back to the **Layers** tab.
2. **Save your project.**
3. Populate the New Database

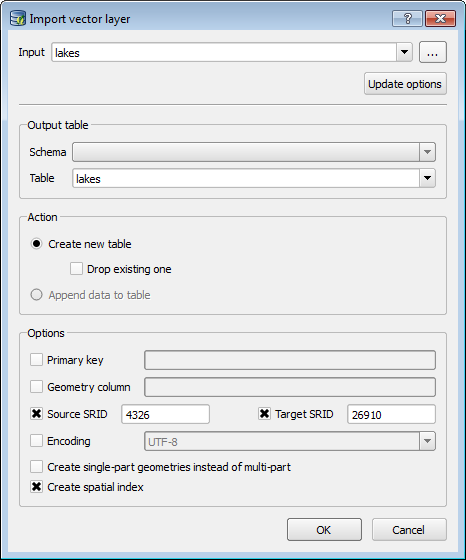
Now you will populate the database with the eight layers.

1. **Open** your **Lab2.qgs** project in **QGIS Desktop** if it not already.
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 (**Figure 10**). You will see the GiffordPinchot.sqlite database. If you expand the database, you will see many tables but no GIS layers yet.



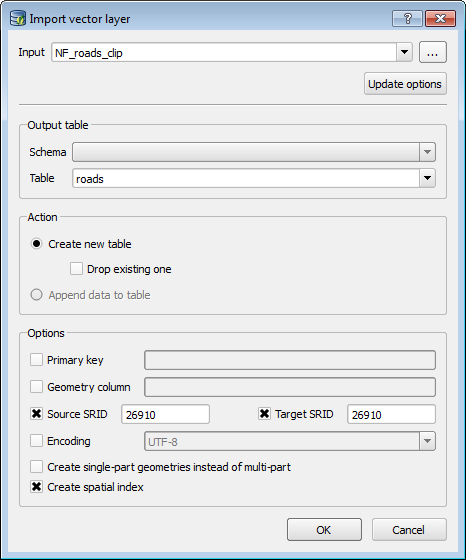
**Figure 10: DB Manager**

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 .
3. Set up the **Input vector layer** window as follows (**Figure 11**):
   1. Set the Input as **lakes**
   2. Name the table **lakes**
   3. Click on Source SRID and enter **4326**
   4. Click the Target SRID and enter **26910**
   5. Click **Create spatial index**
   6. Click **OK**



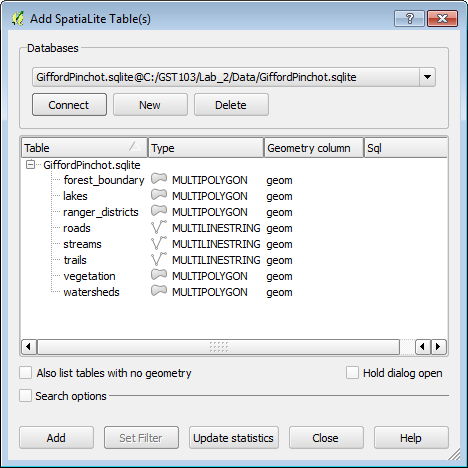
**Figure 11: Import KML Lakes into SpatiaLite DB**

1. Once the operation has completed successfully **click** the **Refresh** button  to see the lakes layer in the database.
2. **Repeat steps 5-7 for streams.**
3. 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).
4. **Click** the **Import layer/file** button .
5. Set up the **Input vector layer** window as follows (**Figure 12**):
   1. Set the Input as **NF\_roads\_clip**
   2. Name the table **roads**
   3. Click on Source SRID and enter **26910**
   4. Click the Target SRID and enter **26910**
   5. Click **Create spatial index**
6. Click **OK**

****

**Figure 12: Import Shapefiles into SpatiaLite DB**

1. **Repeat steps 10-12 for trails, watersheds, ranger districts, vegetation and forest boundary layers.**
2. Now that all eight layers have been imported you can remove the layers you have in QGIS Desktop.
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 , 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 13**).



**Figure 13: Add SpataiLite Table(s)**

5 Conclusion

In this lab, you took data in several different file formats and CRS’s 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 to fall back on. 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.

6 Discussion Questions

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