GIS DATA AND FILE TYPES IN DEP

Table of Contents

- Introduction
- File Formats in ArcGIS
  - Coverages
  - Shapefiles
  - Raster or Image Data
  - Personal Geodatabases
    - Personal Geodatabase Example
  - File Geodatabases
  - ArcSDE Geodatabases
    - Access to ArcSDE files from the R: Drive
    - Access to ArcSDE files from a Database Connection in ArcCatalog
- What is a Layer File?
- How Layer Files Work
  - Example 1
  - Example 2
- List of SDE Layer Files and R: Drive Shapefiles

Introduction

This document is one in a series of GIS reference documents compiled by the Connecticut Department of Environmental Protection to help staff better understand the ArcGIS line of GIS software products. These documents are meant to be a quick reference to an ArcGIS program or topic. If you need more extensive information, you may refer to the Help section of the software program you are using.

GIS Data in DEP has traditionally been created, accessed and edited using ArcGIS, a product of ESRI (Environmental Systems Research Institute), Inc. ESRI GIS products were first introduced in the 1980's and subsequent editions of the software resulted in more functionality. Along with increased functionality came new and different components of the software, including changes in how spatial data files were named, organized and accessed.

GIS Data in DEP has traditionally been stored on what is now commonly known as the R: drive. The data was in the form of coverages (older versions of ArcInfo) and more recently, shapefiles (ArcView). In recent years ArcGIS has evolved and more GIS file types were created. These include geodatabase feature classes and layer (.lyr) files. With all of these file types, it can be confusing to know which data files to add to your ArcMap documents.

The R: drive continues to provide GIS data to the agency. However, rather than searching through five or six folder levels to find a single shapefile, you can now add layer (.lyr) files stored on an ArcSDE server. Because of their ease of use and their pre-set symbology and scale thresholds, layer files are the recommended GIS file type to use with your ArcGIS documents.

SDE stands for Spatial Data Engine. It is a means by which spatial data is stored and accessed more efficiently. Thus, data stored in SDE draws faster when it is added to your ArcMap documents. DEP is now using SDE in more agency applications such as SIMS.

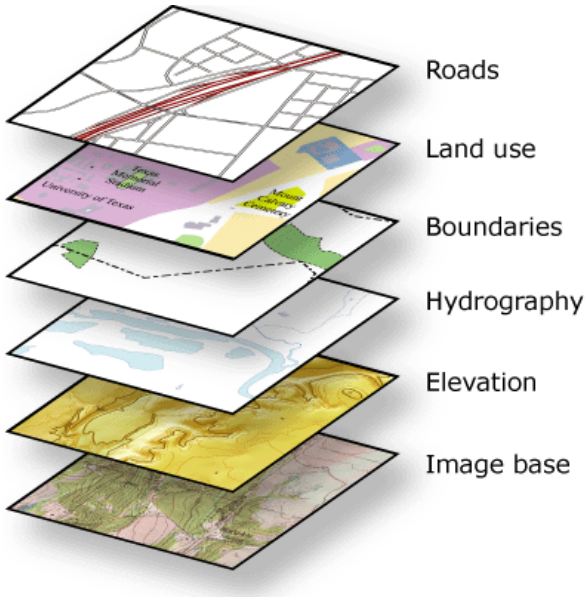
This document will discuss the different types of file formats used in ArcGIS, what they are, how they work, and when to use them. Included are explanations of Coverages, Shapefiles, Images or Raster data, Personal Geodatabases, File Geodatabases, and ArcSDE Databases.

This document will also discuss how to access ArcSDE data from the R: drive. It also discusses how layer files work and why they can help you to work more efficiently in ArcGIS.

GIS Data and File Types in DEP

File Formats in ArcGIS

In the current version of ArcGIS, the name for a specific set of spatial data is called a **"feature class"**. This term is used to refer to any spatial data that can be viewed in ArcGIS. Feature data classes can best be explained as "layers" on a map.



A feature class is characterized by the following:

- each class contains the same spatial data type (e.g. points, lines or polygons)
- each class contains the same feature type (e.g. roads, lakes, rivers)
- each class contains the same attributes (e.g. river name, river length, river depth in the Rivers feature class)
- each class contains the same spatial reference (e.g. Conn. State Plane Coordinate System).

Some examples of feature classes in DEP are "Basins", "Roads", "Waterbodies", "Dams", "Soils" and "Airports".

Feature classes can be stored in a variety of file types or **GIS file formats**:

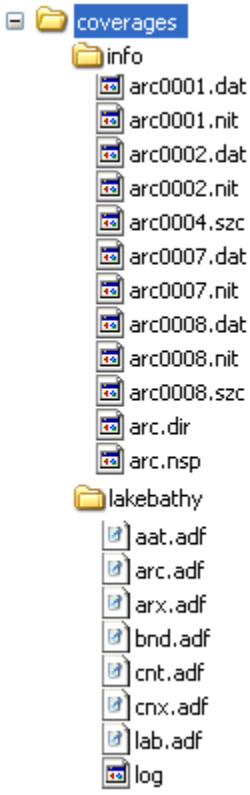
- Coverages
- Shapefiles
- Images or raster files and catalogs
- Personal Geodatabases
- File Geodatabases
- ArcSDE Databases

Each of these file formats is explained on the following pages.

GIS Data and File Types in DEP

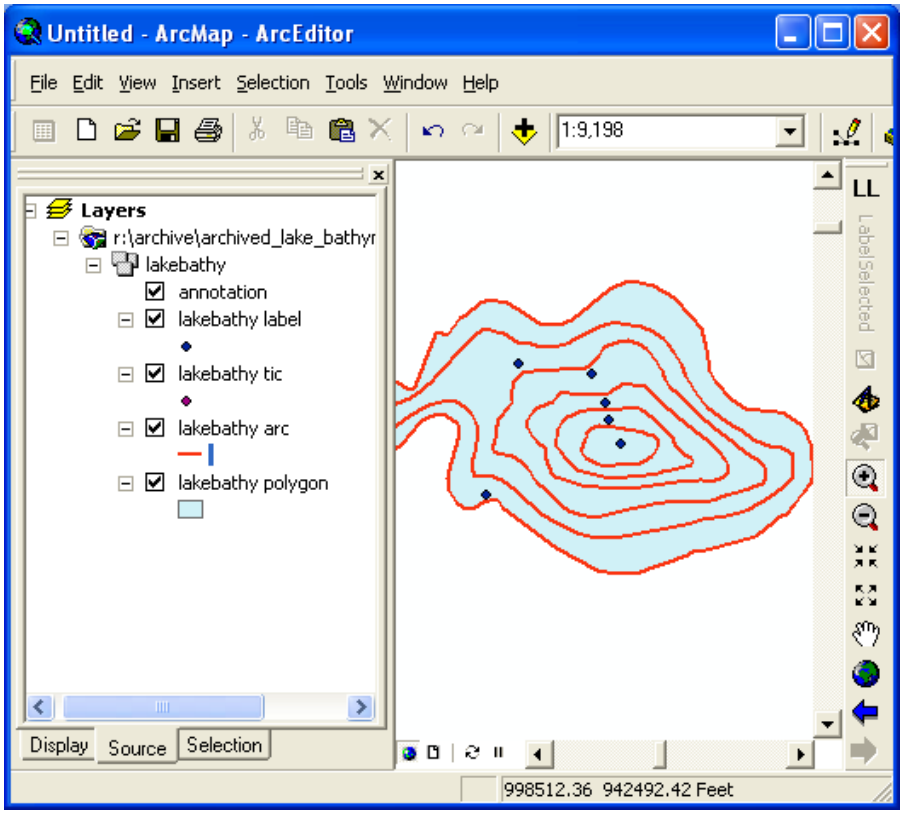
Coverages

When ESRI's GIS software was first introduced in the 1980's, the product was called ArcInfo and the GIS files were known as **Coverages**. Coverages are effective and functional within the ArcInfo software framework, but are not easily exported and contain a large number of individual files. Coverages are stored in two folders, an "info" folder and a folder with the coverage name, for example, "lakebathy" as shown in the illustration, for lake bathymetry.



All of the files in each folder have a specific purpose but all must be present in order for the user to view and edit the specific coverage. Some of the file types store attribute data, some store annotation, some store geometry, and some store topological information.

Coverages are not generally used today since more recent file formats have been introduced. In addition, coverages can be viewed in ArcGIS 9 but can no longer be edited in ArcGIS versions beyond 8.3.



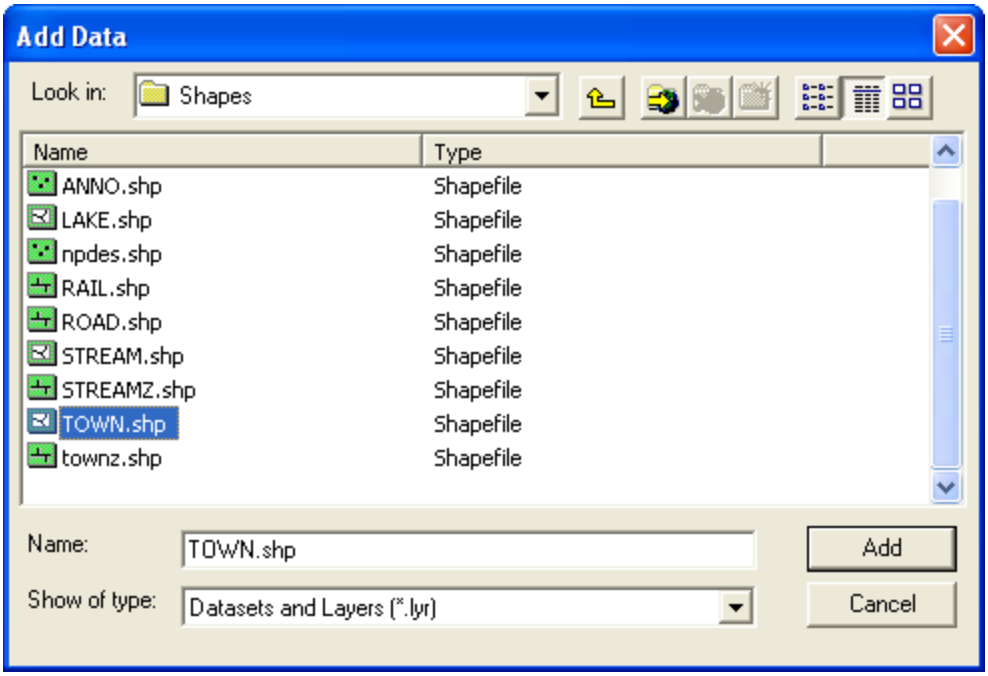
This illustration shows DEP's lake bathymetry coverage. Note that there are points, lines and polygons within this coverage.

More Information About Coverages

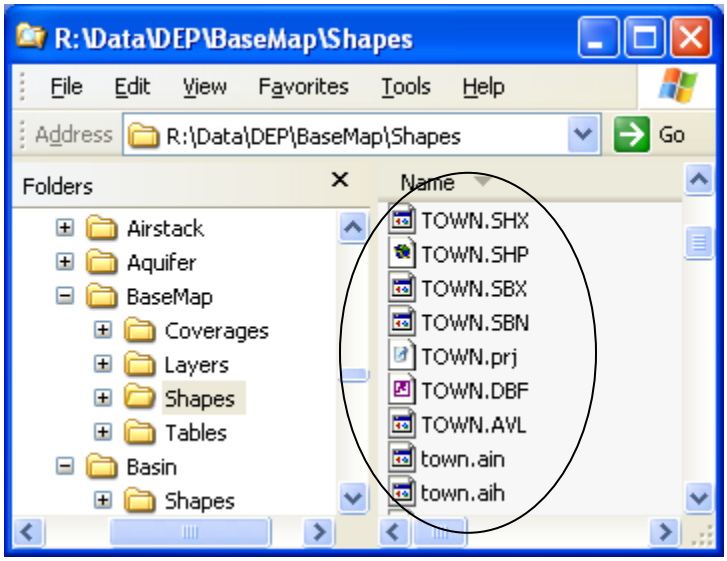
Shapefiles

The most common file format known to GIS users is the **Shapefile**. The shapefile was introduced by ESRI when the first version of ArcView became available. A feature class that is in shapefile format is also represented by multiple files, similar to coverages, but its format is slightly different.

When you click on the Add Data button in ArcView or ArcMap 9, you will see the names of feature classes that can be displayed in the map view. If you are looking at a folder containing shapefiles, you will see only the name of the shapefile. For example, if you are looking for towns in Connecticut, you might find TOWN.shp in the list of available feature classes.



However, when you navigate to the actual folder in Windows Explorer, you will find that a shapefile actually has a number of other files associated with it, all of which are necessary when viewing that specific feature class. In our TOWN.shp example, the accessory files in the illustration to the right are associated with our shapefile.



GIS Data and File Types in DEP

Each of these accessory files has a function in the world of spatial data. ESRI defines these files as follows:

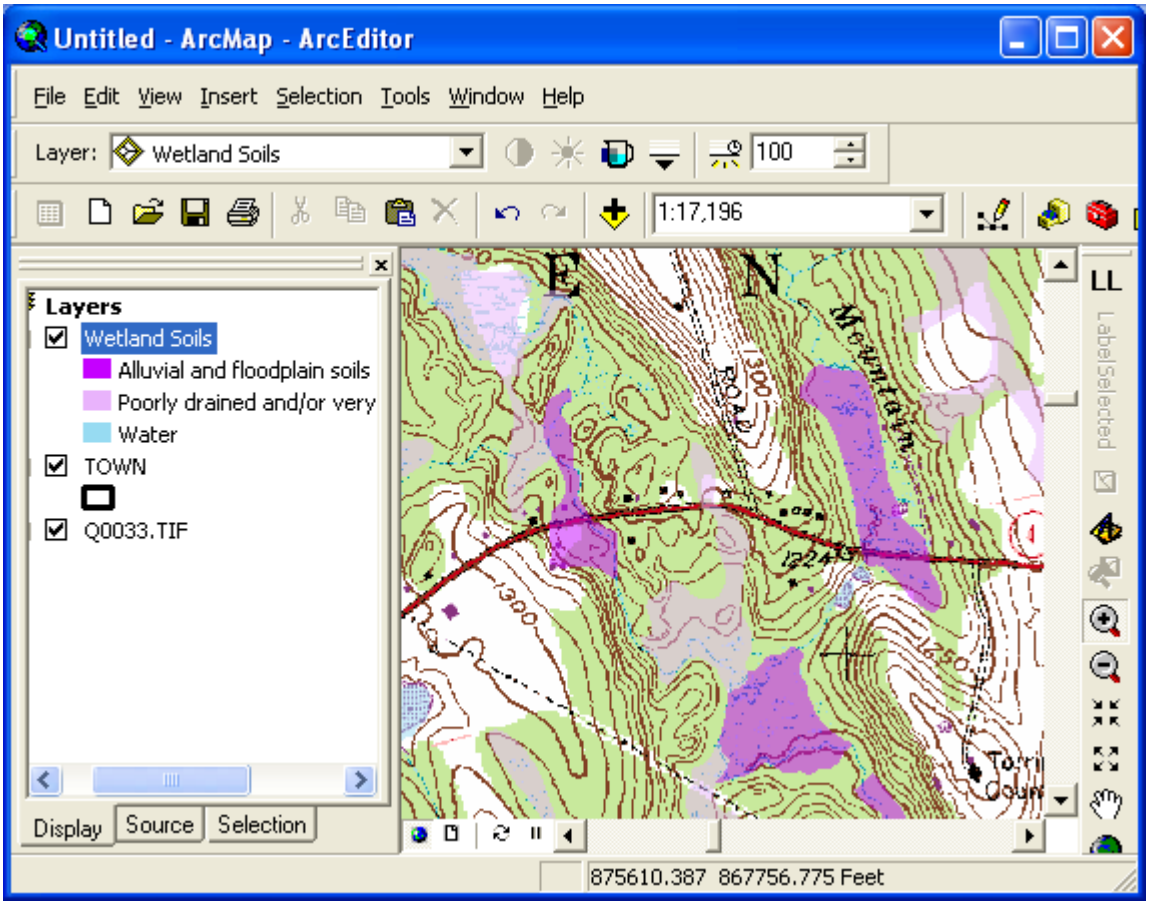
- **.avl**—The ArcView 3 file that stores the legend or symbolization for the shapefile.
- **.shp**—The main file that stores the feature geometry. Required.
- **.shx**—The index file that stores the index of the feature geometry. Required.
- **.dbf**—The dBASE table that stores the attribute information of features. Required. There is a one-to-one relationship between geometry and attributes, which is based on record number. Attribute records in the dBASE file must be in the same order as records in the main file.
- **.sbn** and **.sbx**—The files that store the spatial index of the features.
- **.ain** and **.aih**—The files that store the attribute index of the active fields in a table or a theme's attribute table.
- **.prj**—The file that stores the coordinate system information. Used by ArcGIS.
- **.xml**—Metadata for ArcGIS—stores information about the shapefile.
- **.lyr**—The file that stores legend or symbolization information for ArcGIS 9 feature classes.

It is not necessary to commit these file extensions to memory. Just be aware that when you add a shapefile to your ArcGIS project, there are associated files with it that must be stored within the same folder. When you create a new shapefile, all of the necessary files are created along with it. Some of the files seen above may not be necessary for your shapefile; thus, you may not have all of the above files in your folder.

Here in DEP, much of the data that we have been using in the past 15-20 years is in shapefile format and is stored on the R: Drive, a shared drive within DEP that contains GIS data.

Raster or Image Data Files

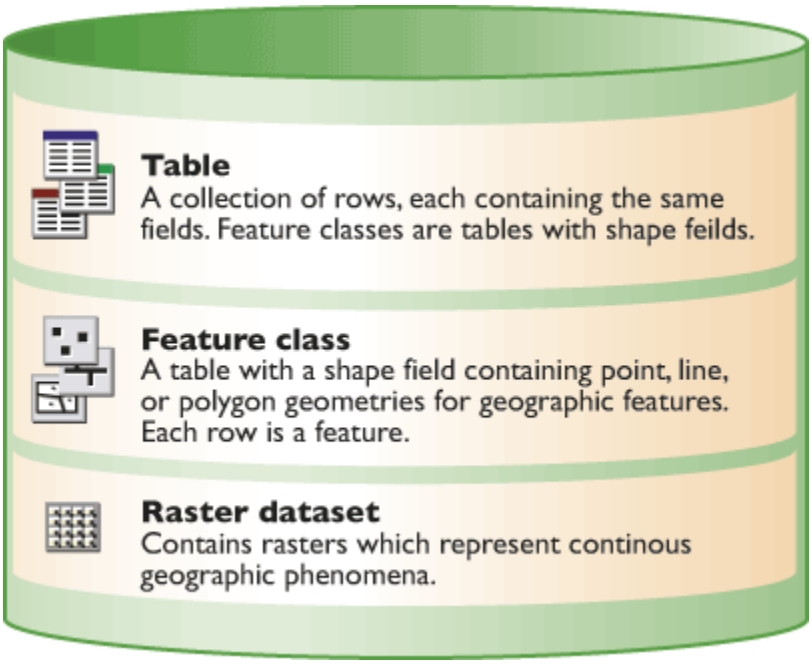
Thus far, the spatial data files we have addressed have contained vector data. Vector data is feature based spatial data. It must contain points, lines or polygons and it has attribute data attached to each feature. A raster dataset, on the other hand, is a spatial image that is represented only by equally sized pixels of differing colors or shades. It does not contain identifiable points, lines, or polygons on the map that are supported by attribute data. In the example below, the file known as q0033.TIF is an image, or raster file. It is topographic quadrangle #33 and it provides the background for the vector data (soils and town lines) that are drawn on top of it. Much of the raster data that we use in DEP is used as background or "base map" imagery. The topographic quadrangles, the 2004 orthophotos, and the coastal infrared imagery are all examples of raster data.



Any image file can be displayed in ArcGIS as raster data. This includes, but is not limited to, the following file types: .jpg, .tif, .png, .bmp, .gif, and MrSid files, which are a certain type of compressed file that minimized file storage space. In addition to topographic quadrangle images in .tif format, DEP also has a folder of these same quadrangles in MrSid format to maximize drawing time and minimize storage space.

Personal Geodatabases

Personal geodatabases were introduced with ArcGIS Version 8. Whereas a shapefile contains only one feature class and a coverage contains points, lines or polygons for a related feature class, a geodatabase is a spatial data storage "container" that can house many types of feature classes and their corresponding tables. A personal geodatabase works with Microsoft Access and has the file extension ".mdb". The spatial data and attribute data are contained in tables within the Access database.



The components of a geodatabase include tables, feature classes, raster datasets, and other items such as relationship classes and topology. The advantages of a personal geodatabase in ArcGIS lie in the interrelationships among the components.

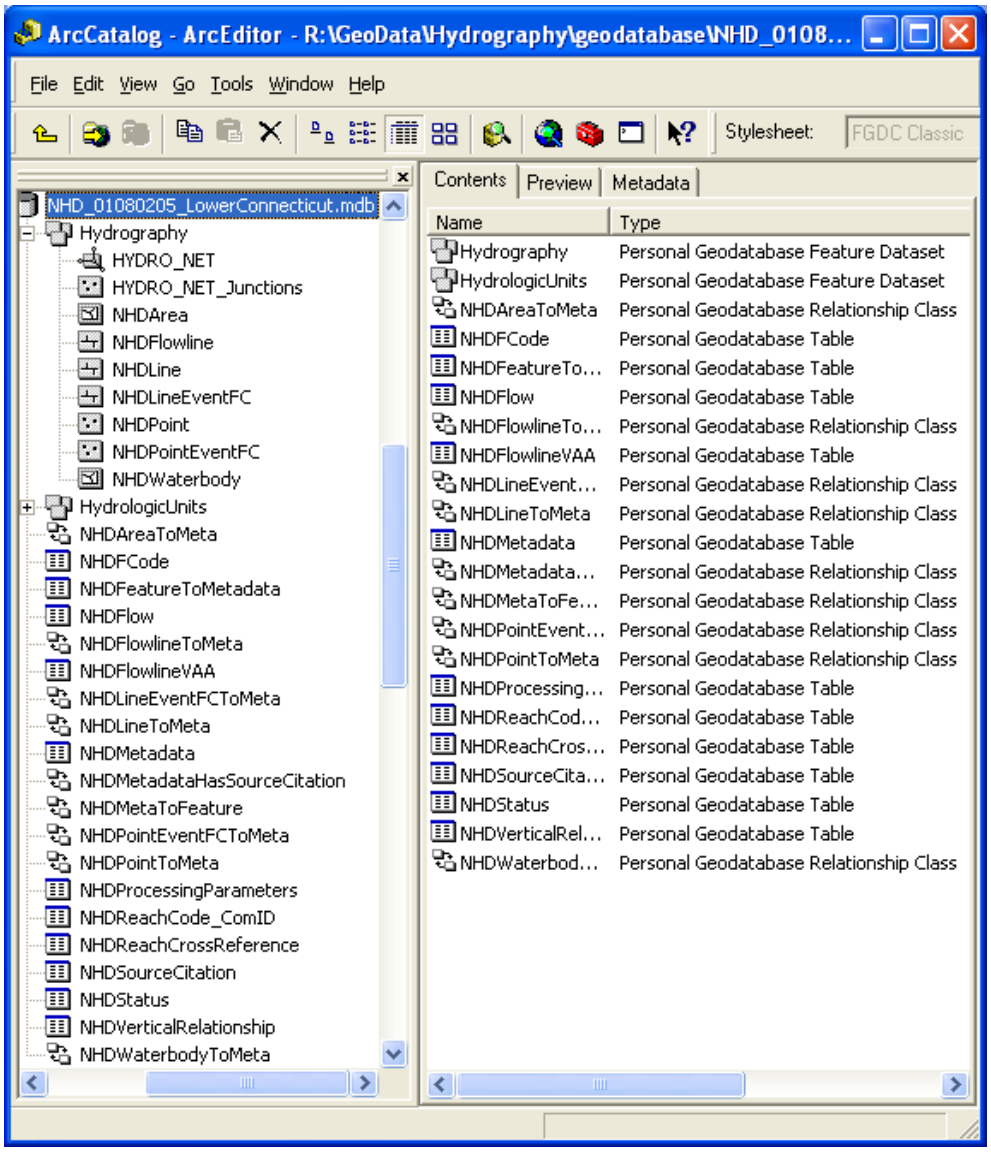
Some of the geodatabase capabilities include verifying data integrity by setting up domains for data entry, and topology rules for the spatial features.

For example, if the field entitled "Terrain_Code" in the Bedrock_Geology polygon feature class of the Geology Geodatabase can contain only 11 specific codes, then a rule can be created that allows only these codes to be entered and no other. This is called a **domain**. Likewise, within the polygon feature class itself, a **topology** rule can be set up to ensure that no two polygons ever overlap. If this occurs, an error will appear in the topology feature class. These types of controls can help with maintaining accurate and valid GIS data.

Note: Although personal geodatabases may be easy to use as a result of their Microsoft Access structure, keep in mind that the maximum size of a personal geodatabase is 2 gb. However, when your personal geodatabase reaches a size of between 200 and 500 mb, you might see a significant slowdown of processing. For this reason, ESRI recommends that the File Geodatabase format be used, especially if you anticipate your geodatabase growing.

Personal Geodatabase Example

In this example, the personal geodatabase is named “NHD_01080205_LowerConnecticut.mdb. It represents data from the National Hydrography Dataset for the Lower Connecticut River area.



Because so many feature classes, tables and other data can be contained in a geodatabase, there must be a logical schema. In the ArcCatalog illustration (left), you’ll see that the contents of the geodatabase are listed on the right. These include feature datasets, relationship classes, and tables. The two **feature datasets**, Hydrography and HydrologicUnits act as “containers” for the actual spatial data, i.e., the point, line, and polygon feature classes. The purpose of creating feature datasets is to apply topological rules for a related group of feature classes.

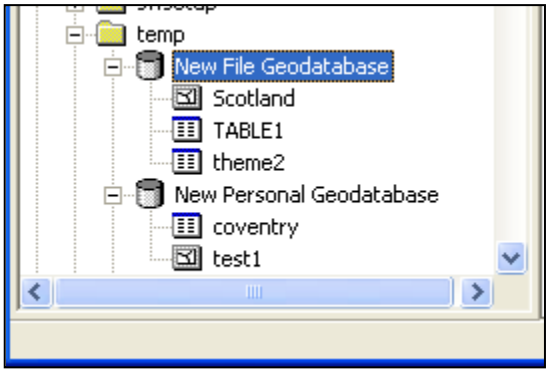
The Hydrography feature dataset has been expanded on the left. You’ll see the feature classes listed, beginning with “Hydro_Net”.

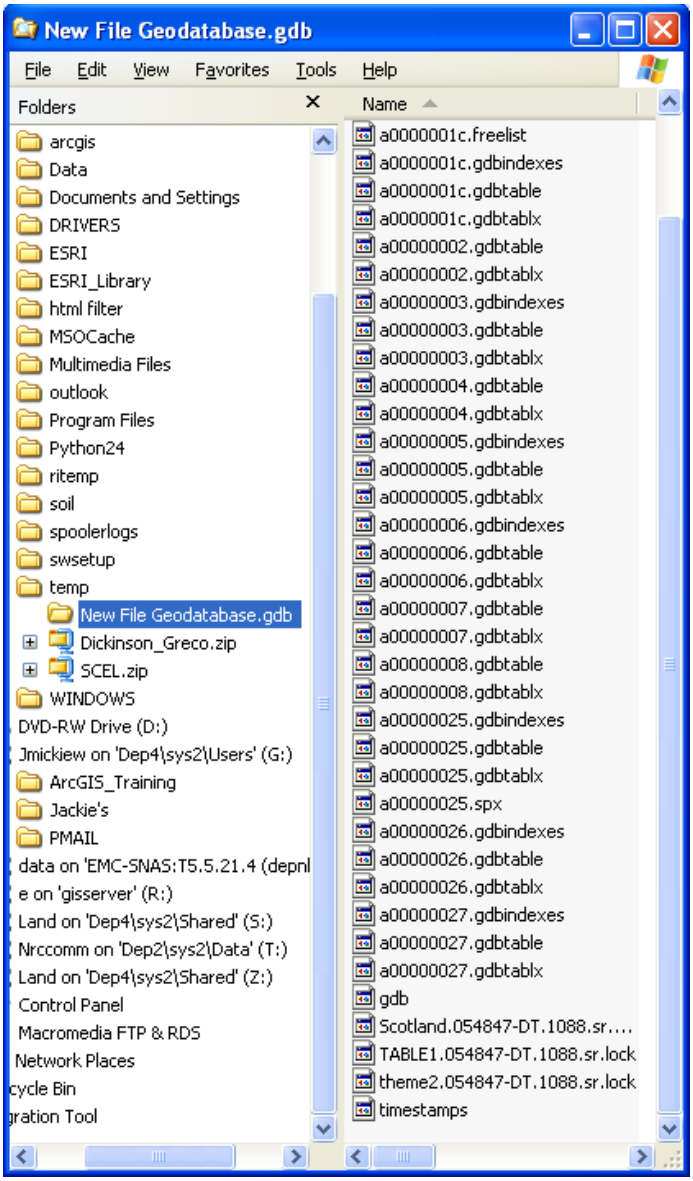
More Information on Types of Geodatabases from ESRI’s website

File Geodatabases

When you anticipate creating a geodatabase with large amounts of spatial and attribute data, it might be more efficient to create a file geodatabase. This type of spatial database can handle much more data than a personal geodatabase and is much faster in processing data.

File geodatabases store data in folders. When viewing this type of geodatabase in ArcCatalog, the contents are displayed in exactly the same way as a personal geodatabase.





However, when viewing it in Windows Explorer, you'll see that the structure is very different. There are numerous files associated with the feature classes in a file geodatabase.

The file extension for a file geodatabase is ".gdb", vs. the .mdb of a personal geodatabase.

Here in DEP, there are few, if any file geodatabases stored on our shared R: Drive. The reason for this has to do with the next section on ArcSDE Databases.

More Information on Types of Geodatabases from ESRI's website

ArcSDE Databases

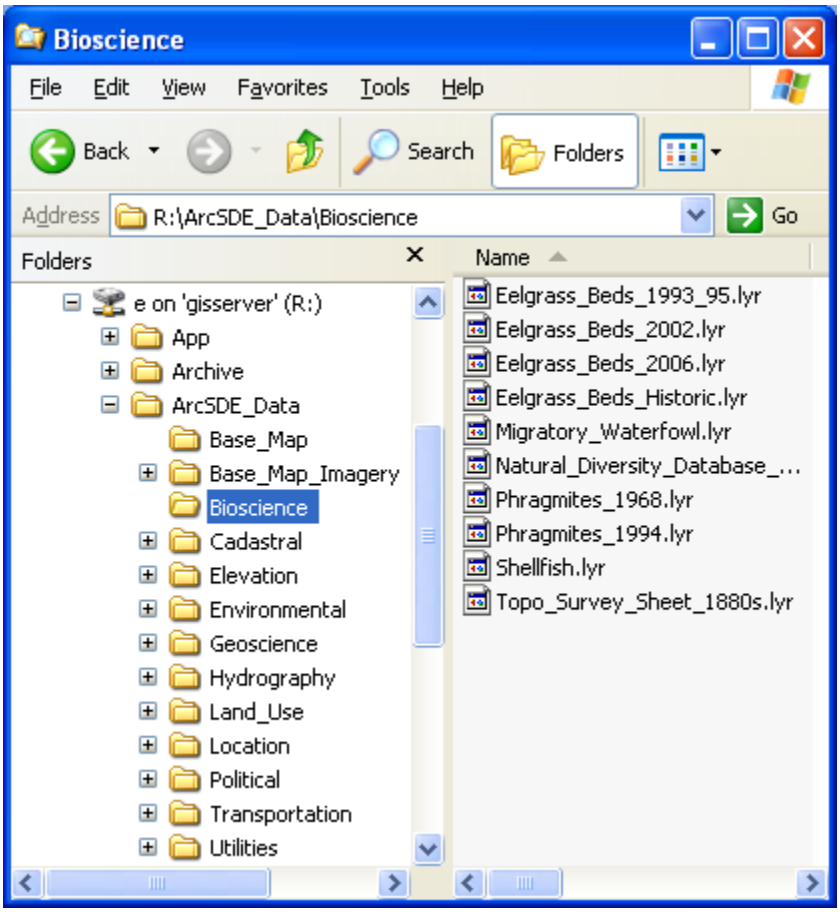
DEP is now actively moving towards storing all of its GIS data in ArcSDE. ArcSDE is ESRI's technology for accessing and managing geospatial data within relational databases. What this means for the GIS user is a much faster response time for displaying large amounts of spatial data and the ability to edit data in a multi-user environment. Without becoming too technical, the bottom line is that ArcSDE is the most efficient way to store and access spatial and attribute data within large organizations such as DEP. It is for this reason that, eventually, all of DEP's GIS data will be stored in this format.

Access to ArcSDE Files From the R: Drive

DEP's GIS data that resides within the ArcSDE framework is located on a network and is accessed two ways. The first way to access SDE stored feature classes is directly from the R: Drive via either Windows Explorer, ArcCatalog, or ArcMap. Although accessed from the R: Drive, the data itself resides within ArcSDE. Only the layer files, or "pointers" to the data, reside on R. Layer files are discussed in more detail in the following pages. For now, all you need to know is how to access the SDE data from this location.

Open Windows Explorer, then navigate to R:\ArcSDE_Data\. You will see categories of data beginning with Base_Map and ending with Utilities. Within each category are .lyr or layer files that can be dragged into your ArcMap document. These layer files contain not only symbology but also other pre-determined "rules" such as how much data to view at certain scales (scale thresholds).

In this example, when the Bioscience folder is opened, all of those layer files related to this category are shown.

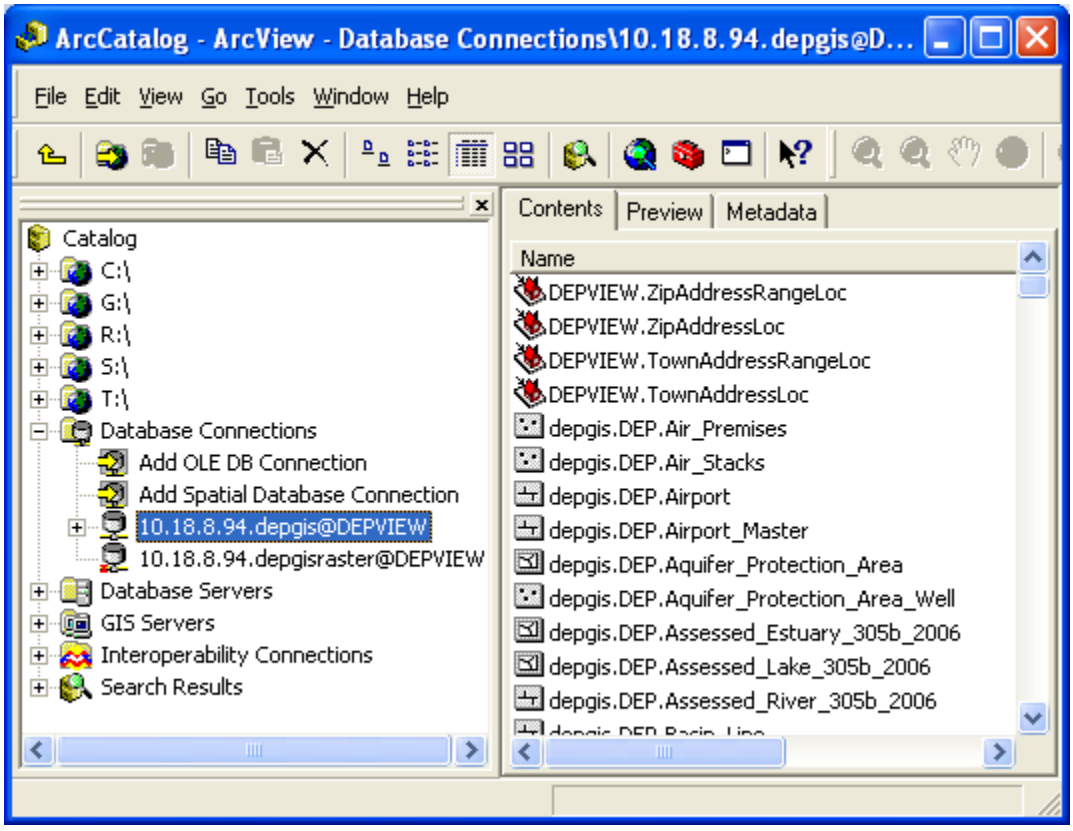


Access to ArcSDE Files From a Database Connection in ArcCatalog

The second way to access ArcSDE data is through a "Database Connection" in ArcCatalog. To establish this connection, you must place two files into a folder on your C: drive. These files will allow you to view the ArcSDE data from ArcCatalog and also from the "Add Data" dialog box in ArcMap.

The files are as follows:
10.18.8.94.depgis@DEPVIEW.sde (this file is for all feature data); and
10.18.8.94.depgisraster@DEPVIEW.sde (this file is for all raster data [imagery]).
Talk to your instructor to have a copy of each file emailed to you.

Copy both files to your profile folder as follows:
C:\Documents and Settings\[your profile name]\Application Data\ESRI\ArcCatalog\



Once you have copied these files to your C: Drive, you can now view them in ArcCatalog by clicking on "Database Connections", then double clicking on one of the above files. For all feature data, such as roads, hydrography, surficial materials, etc., click on the depgis@DEPVIEW file. For raster (image) data, double click on the depgisrater@DEPVIEW file. You will see a list of feature classes.

[More Information About ArcSDE from ESRI's Website]

What is a Layer File?

A layer file is different from a shapefile or a geodatabase feature class because it does not actually contain the data itself. It contains information about where to find the data and how to display it. Layer files tell ArcMap where the data is located, how to symbolize the data, at what scale to display the data (scale threshold), and how to display the attribute table. Virtually any property that can be set in the “Properties” dialog box can be assigned to a layer file.

How Layer Files Work

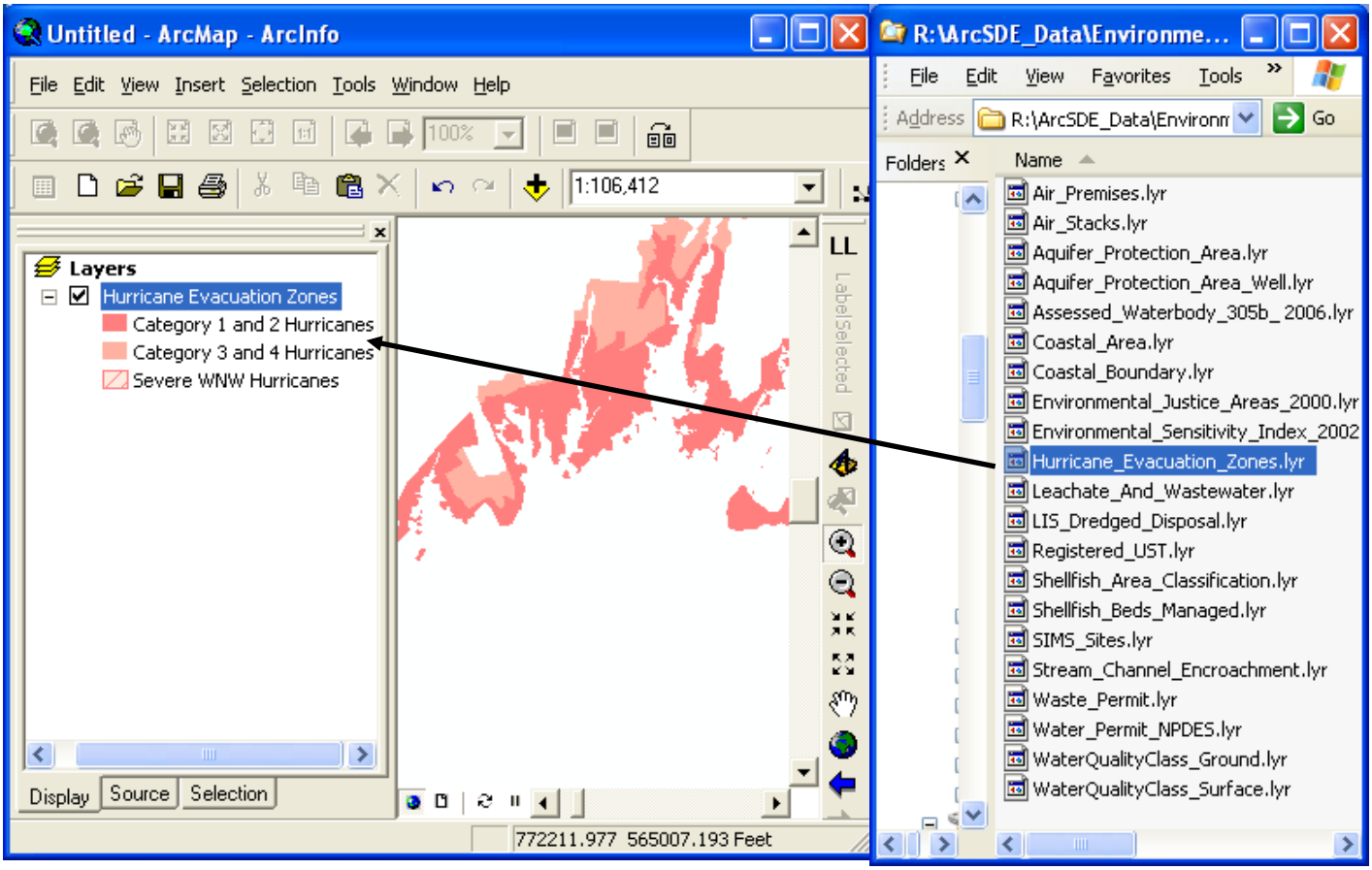
Once you have dragged and dropped a layer file into ArcMap, it displays with pre-determined symbology and scale levels, as well as any other property that you have set and that was saved with the layer file. This is the advantage of using a layer file vs. a shapefile or geodatabase feature data class.

Layer Files in DEP

Layer files in DEP are used especially in ArcSDE. The list of .lyr files that are displayed in the R:\ArcSDE_Data\ folder are layer files that were created specifically for use by DEP staff and were created with settings that would be most helpful to staff.

On the following pages are examples of layer files and how they help with the display of GIS data.

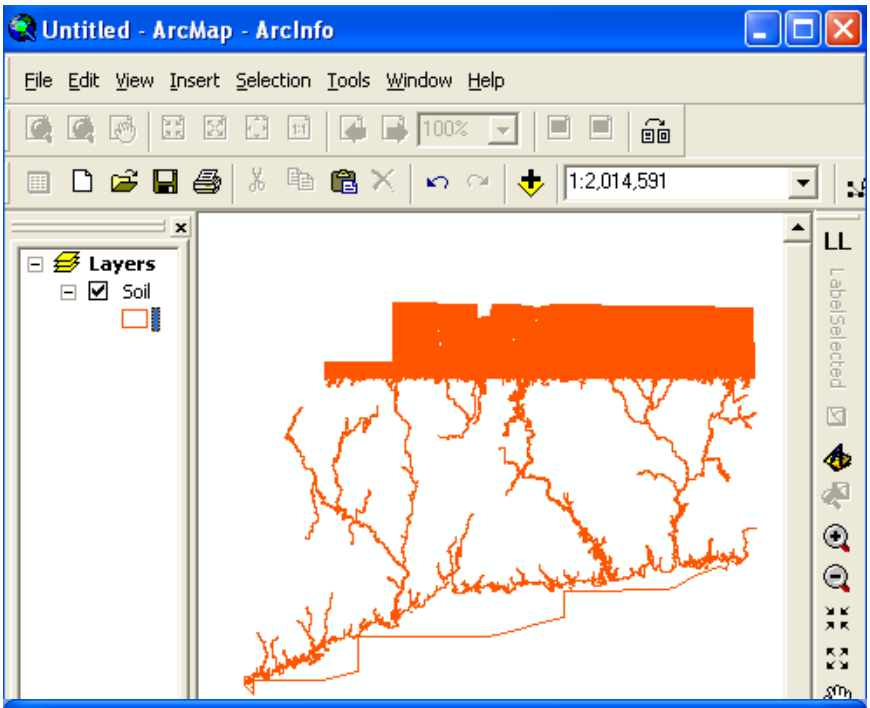
EXAMPLE 1



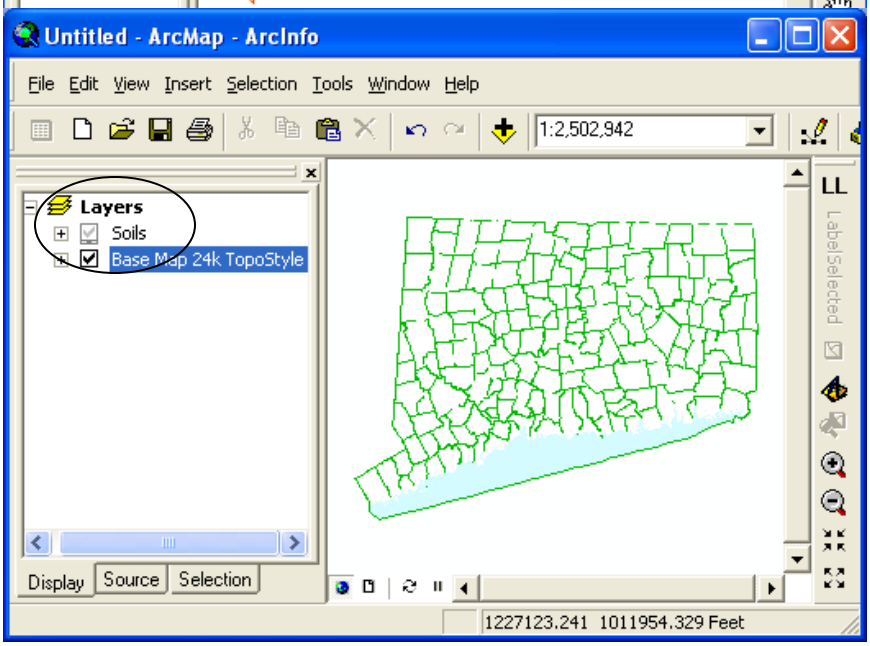
In the above Example 1, Hurricane Evacuation Zones are automatically symbolized by Category. The colors have been preset. In addition, if you zoom out beyond 1:200,100 the layer will not longer display since a scale threshhold has been set.

EXAMPLE 2

Example 2 below shows what happens when you add a soils shapefile to an ArcMap document without using the layer file. If you turn the shapefile on when you are at a statewide scale level, the soils will begin to draw. There is no scale threshold and the drawing may take several minutes. With a scale threshold within a layer file, the soils will only display when zoomed in to a more manageable scale. This is why it is easier and faster to use layer files vs. shapefiles.



The soil shapefile draws slowly when no scale threshold is set.

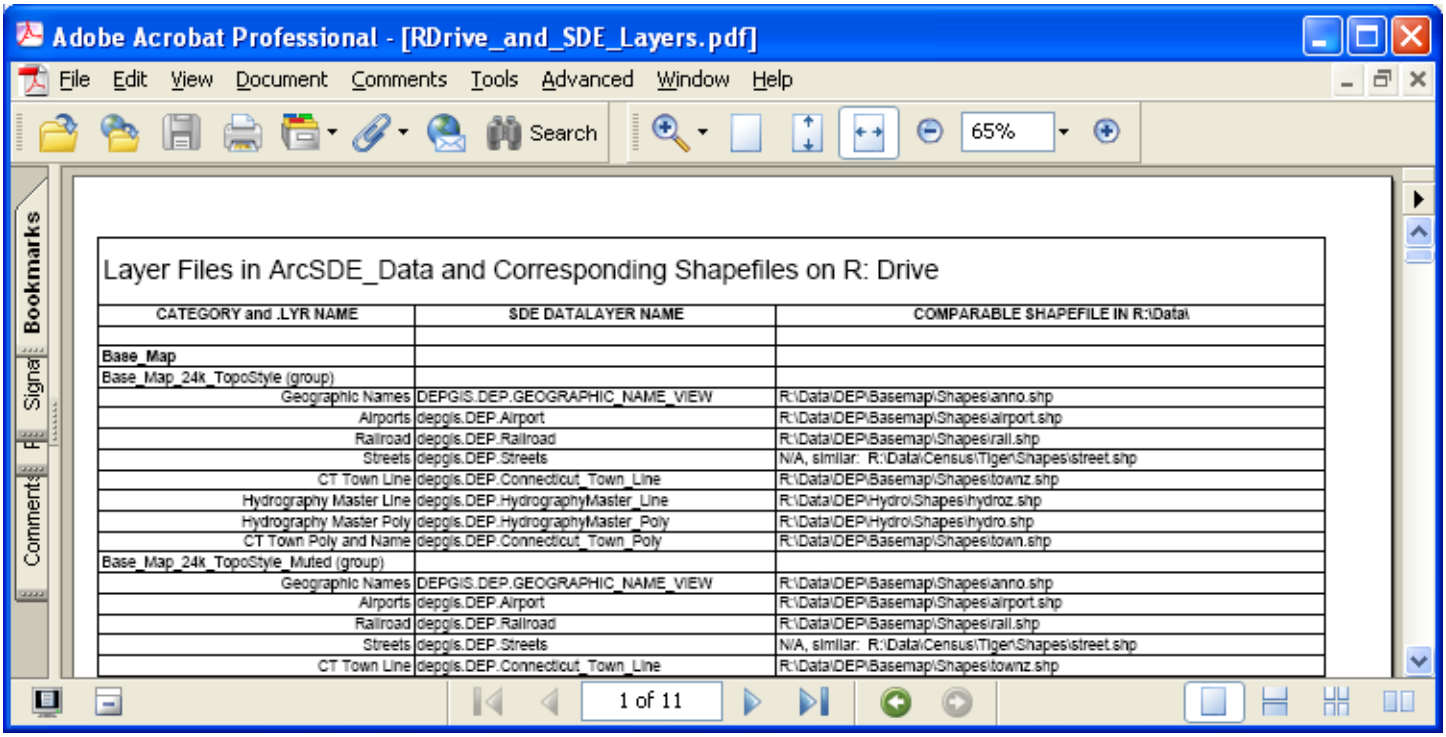


If you set a scale threshold on a layer, then it will not draw beyond the scale you specified and a grayed-out check mark will be shown in the layer box.

List of SDE Layer Files and R: Drive Shapefiles

To see the relationship between the newer SDE layer files and the older R: Drive shapefiles, open the following document. This document shows the SDE layer name and the corresponding "older" R: Drive shapefile name.

R: Drive and SDE Files



Layer Files in ArcSDE_Data and Corresponding Shapefiles on R: Drive

CATEGORY and .LYR NAME	SDE DATALAYER NAME	COMPARABLE SHAPEFILE IN R:\Data\
Base_Map		
Base_Map_24k_TopoStyle (group)		
Geographic Names	DEPGIS.DEP.GEOGRAPHIC_NAME_VIEW	R:\Data\DEP\Basemap\Shapes\anno.shp
Airports	depgis.DEP.Airport	R:\Data\DEP\Basemap\Shapes\airport.shp
Railroad	depgis.DEP.Railroad	R:\Data\DEP\Basemap\Shapes\rail.shp
Streets	depgis.DEP.Streets	N/A, similar: R:\Data\Census\Tiger\Shapes\street.shp
CT Town Line	depgis.DEP.Connecticut_Town_Line	R:\Data\DEP\Basemap\Shapes\townz.shp
Hydrography Master Line	depgis.DEP.HydrographyMaster_Line	R:\Data\DEP\Hydro\Shapes\hydroz.shp
Hydrography Master Poly	depgis.DEP.HydrographyMaster_Poly	R:\Data\DEP\Hydro\Shapes\hydro.shp
CT Town Poly and Name	depgis.DEP.Connecticut_Town_Poly	R:\Data\DEP\Basemap\Shapes\town.shp
Base_Map_24k_TopoStyle_Muted (group)		
Geographic Names	DEPGIS.DEP.GEOGRAPHIC_NAME_VIEW	R:\Data\DEP\Basemap\Shapes\anno.shp
Airports	depgis.DEP.Airport	R:\Data\DEP\Basemap\Shapes\airport.shp
Railroad	depgis.DEP.Railroad	R:\Data\DEP\Basemap\Shapes\rail.shp
Streets	depgis.DEP.Streets	N/A, similar: R:\Data\Census\Tiger\Shapes\street.shp
CT Town Line	depgis.DEP.Connecticut_Town_Line	R:\Data\DEP\Basemap\Shapes\townz.shp

The first column, CATEGORY and .LYR NAME, is organized by category. These categories are:

- Base Map
- Base Map Imagery
- Bioscience
- Cadastral
- Elevation
- Environmental
- Geoscience
- Hydrography
- Land Use
- Location
- Political
- Transportation
- Utilities

The second column, SDE DATALAYER NAME, displays the name of the Layer File stored in ArcSDE.

The third column, COMPARABLE SHAPEFILE IN R:\DATA, displays the corresponding Shapefile on the R: Drive, if available.