

Table of Contents.

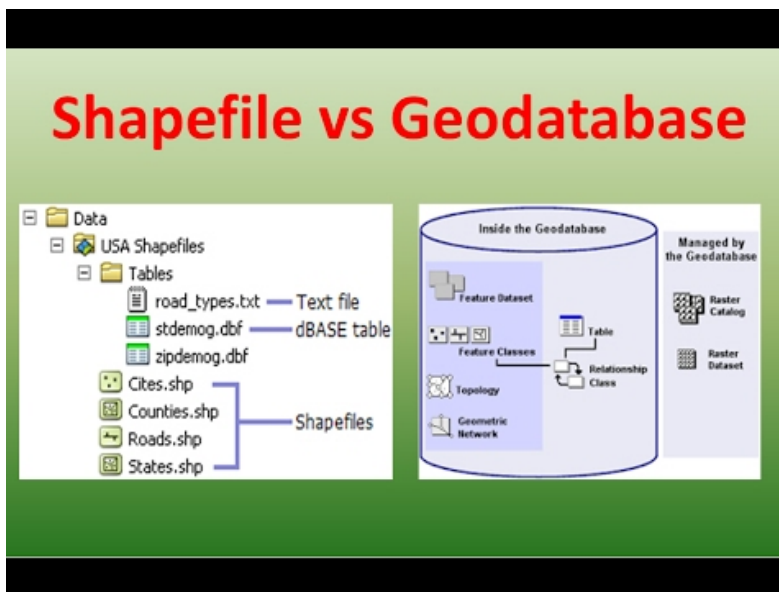
1. GIS Data Models
2. File Geodatabases (FGDBs)
3. Shapefiles (SHPs)

1) GIS Data Models.

There are currently two different file formats (or data models) for GIS data: shapefiles (SHPs) and geodatabases (GDBs). Shapefiles are the older file format, but still very popular in the GIS world.

- Most GIS data portals provide GIS data in either file format.
- You need to be able to handle either file format.

Both file formats work fine. GDBs are better.



2) File Geodatabases (FGDBs).

To make things more complicated: there are at least three types of geodatabases.

- File geodatabases. Stored as folders in a file system. Each dataset is held as a file that can scale up to 1TB in size. The file geodatabase is recommended over personal geodatabases.
- Personal geodatabases. All datasets are stored within a Microsoft Access data file, which is limited in size to 2 GB.
- ArcSDE geodatabases. Stored in a relational database using Oracle, Microsoft SQL Server, IBM DB2, IBM Informix, or PostgreSQL. These multiuser geodatabases require the use of ArcSDE and can be unlimited in size and numbers of users.

⇒ We will primarily use file geodatabases in this class.

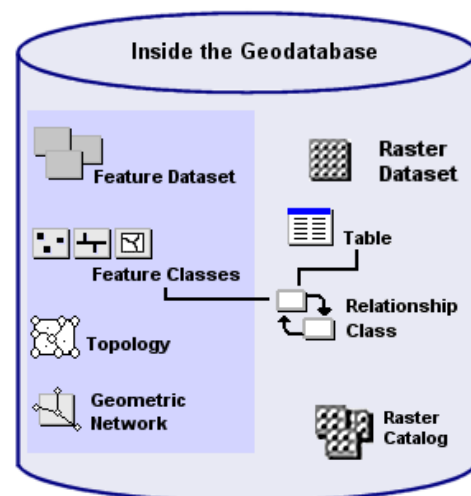
The File Geodatabase = The Duffle Bag Approach.

Everything (spatial data, attribute data, pictures, raster data, tabless etc.) is stored in a single, very complicated database table = all your GIS data are stored inside one single file and all you need is a GIS software to read that file.

You can easily import shapefiles as feature classes into a geodatabase and vice-versa. The same is true for raster data, networks, etc.

You can think of a file geodatabase (FGDB) as a container holding all kinds of GIS data sets. For example, vector data are stored inside geodatabases as so-called feature classes.

- Vector data (points, lines, polygons) are stored inside an FGDB as so-called feature classes: point feature class, polyline feature class, and polygon feature class.
- Raster data (e.g. grids).
- Database tables.
- Images.
- Topologies
- Relationships.
- Networks.
- Etc.

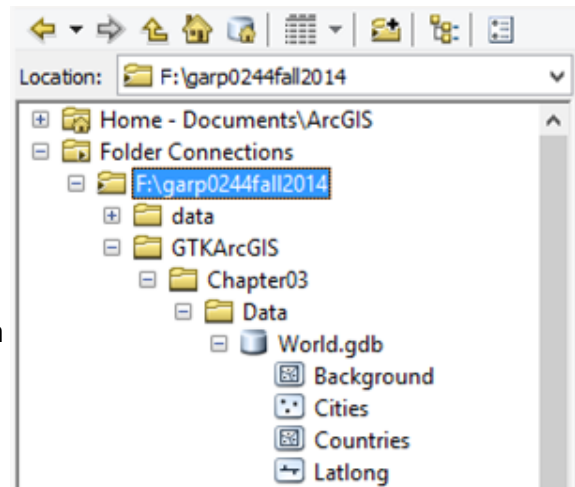


The internal structure of geodatabases is so complex that you need ArcGIS to look inside it.

Example.

Here are the feature classes contained in the file geodatabase World.gdb.

- Each feature class is one vector data set (point, line, or polygon).
- ArcGIS allows you to preview the geography and attribute table of each feature class. That's quite useful so you can get an idea of what the feature class contains before adding it to ArcGIS.



A New Language.

GIS data = geospatial data, stored as feature classes inside geodatabases or as shapefiles. The map you create in ArcMap does NOT contain the feature classes (or shapefiles) underlying each map layer, but a link back to the location where the actual feature classes (or shapefiles).

GIS maps and GIS data are independent, but linked.

The File Geodatabase and ArcGIS Pro.

There are a few Big-Picture concepts here.

1. Everything is organized around Projects.
2. All GIS data used in a Project is stored inside the Project's FGDB.
3. You need to use ArcGIS to handle your GIS data.

That's a different approach from using shapefiles (see below).

More Information.

- Feature class basics:
<http://desktop.arcgis.com/en/desktop/latest/manage-data/geodatabases/feature-class-basics.htm>
- Geospatial Data Formats: <https://www.lib.ncsu.edu/gis/formats.html>
- Commonly Encountered File Types/Extensions in ArcGIS:
https://www.wou.edu/las/physci/taylor/es341/arc_gis_file_extensions.pdf

3) Shapefiles (SHPs).

Most MassGIS data layers are still distributed as so-called shapefiles. The same is true for many other online data sources. Shapefiles are still very popular with GIS professionals for the simple reason that the data and file structure of shapefiles is simple and manageable without relying on ArcGIS Pro.

You will encounter and use shapefiles for years to come.

A shapefile has to include (at minimum):

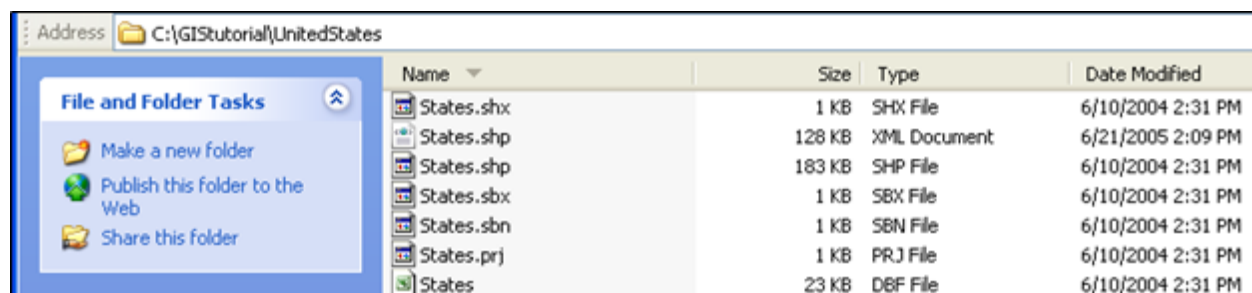
- The geographic data of the spatial feature.
- The attribute data of the spatial feature.
- An index, linking the geographic data to the attribute data.

Therefore.

1. You can look at a shapefile as a map layer (= map the geographic data).
2. You can look at a shapefile as a table (= view the attribute data).
3. Use the attribute data to change how the associated geographic data are displayed on the map.

Example.

The file format for MS Word is *.docx = a MS Word document is stored on your computer as a *.docx file. The file format for GIS data is a shapefile, meaning that GIS data are saved on your computer as shapefiles. As mentioned above, shapefiles are being gradually replaced with a new file format called geodatabases.



Name	Size	Type	Date Modified
States.shx	1 KB	SHX File	6/10/2004 2:31 PM
States.shp	128 KB	XML Document	6/21/2005 2:09 PM
States.shp	183 KB	SHP File	6/10/2004 2:31 PM
States.sbx	1 KB	SBX File	6/10/2004 2:31 PM
States.sbn	1 KB	SBN File	6/10/2004 2:31 PM
States.prj	1 KB	PRJ File	6/10/2004 2:31 PM
States	23 KB	DBF File	6/10/2004 2:31 PM

Shapefile Structure.

Now it gets a little weird: 1 shapefile is actually between 3 and 8 individual files. For example, the screenshot above shows you 1 shapefile (States.shp) but that shapefile actually consists of 7 individual files.

There are three mandatory files in each shapefile:

1. States.shp = the shapefile 'proper' = the spatial data: shape + location = geometry of the features.
2. States.dbf = database table with the associated attribute data of the features.
3. States.shx = a spatial index file to link the geometry of features to their attributes.

There are several optional files in each shapefile:

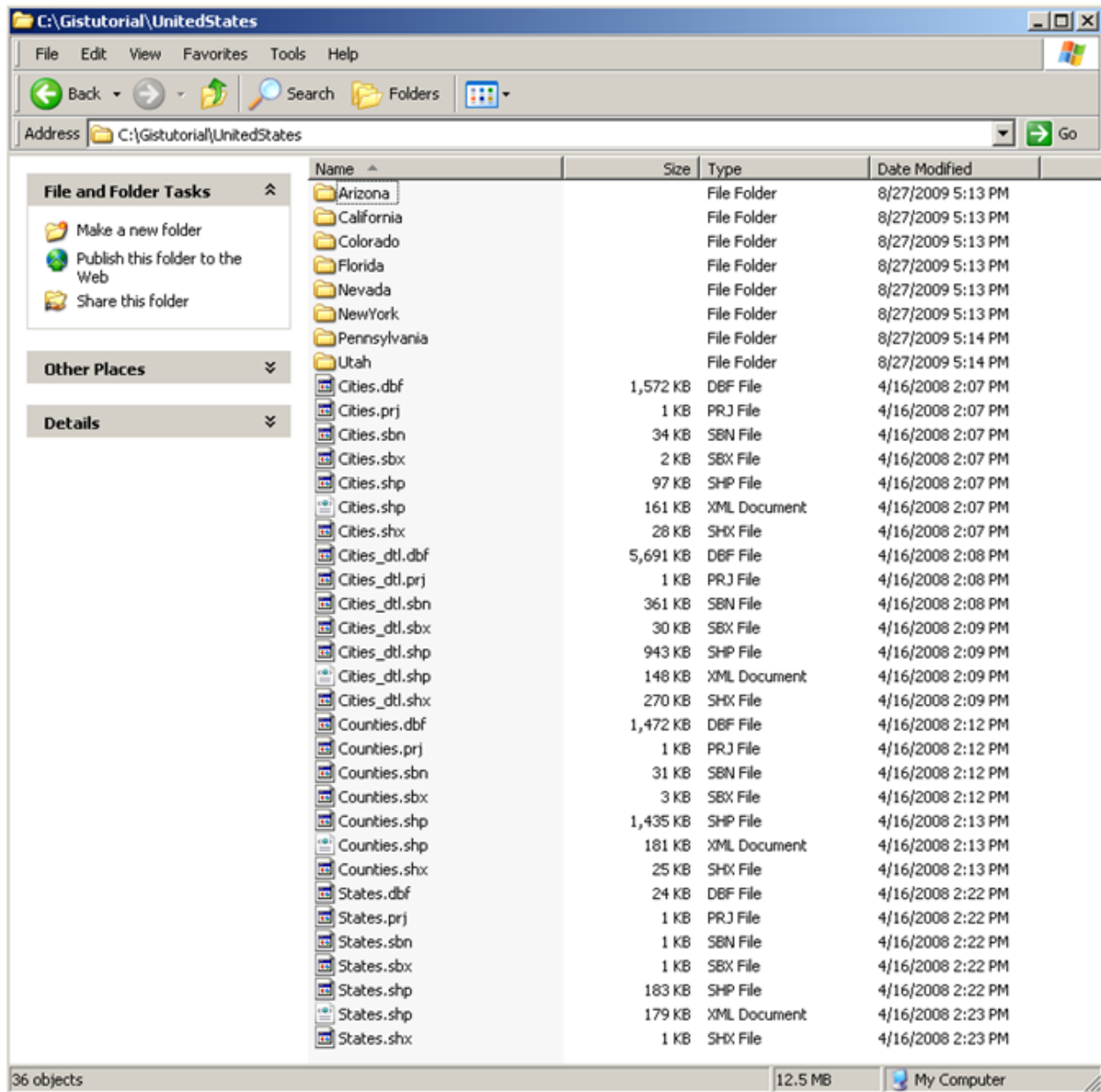
.sbn/.sbx	Additional index files as needed.
.prj	The projection file,
.shp.xml	meta data for the shapefile (= information about the data),

We are used to a different situation. For example, you type a paper in MS Word and save it as a single *.docx file. This one single file contains all your text, all your figures, all your tables – everything. That single file you can attach to an Email and, for example, send your professor.

Well, a shapefile is not a single file, but a family of up to eight files – even though we refer to it simply as a shapefile. Therefore you need to combine or zip the individual files together as a single archive for sharing or downloading.

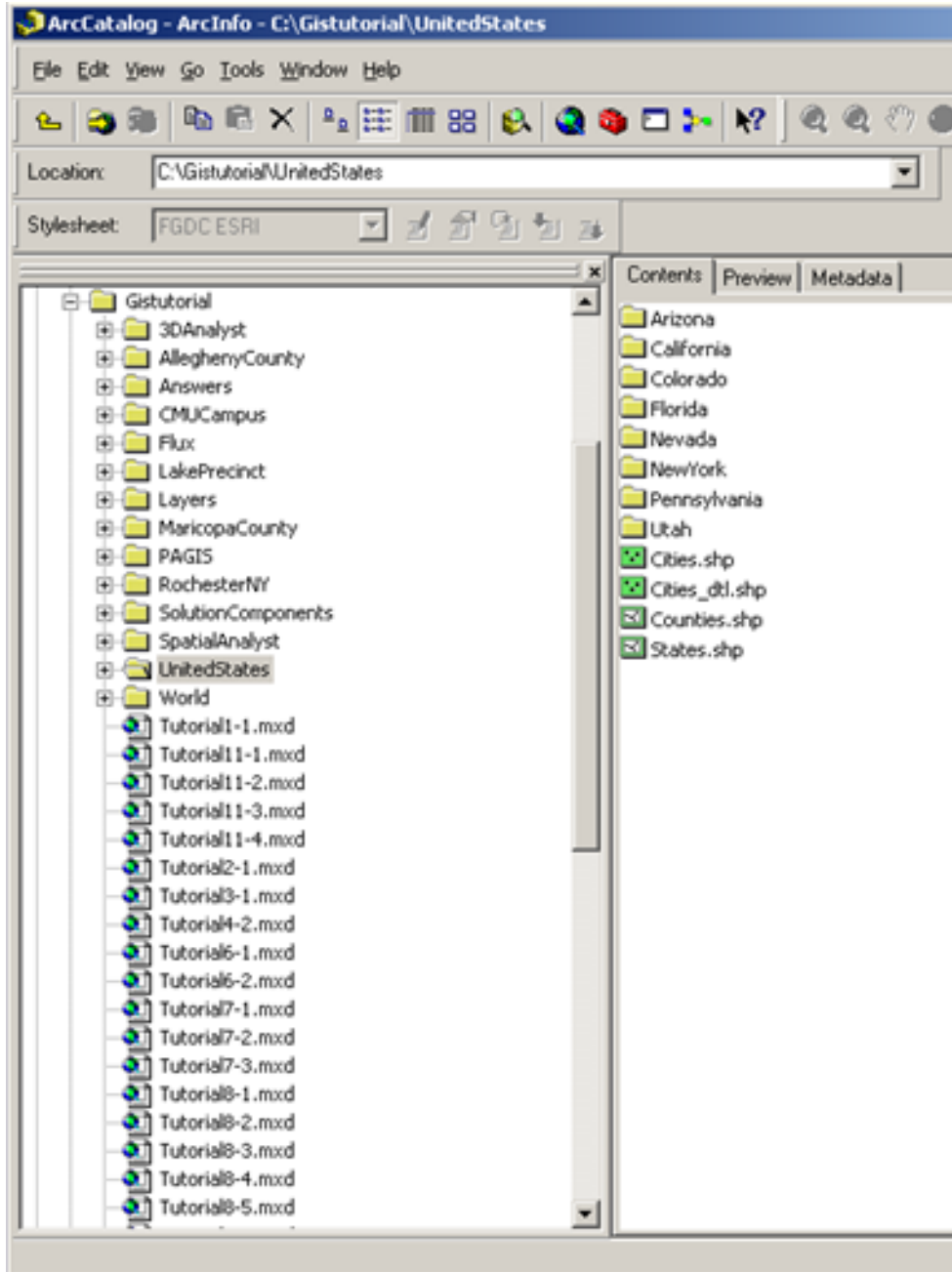
More Information.

- What is a shapefile?
<http://desktop.arcgis.com/en/desktop/latest/manage-data/shapefiles/what-is-a-shapefile.htm>
- Parts of a shapefile:
<http://gisgeography.com/arcgis-shapefile-files-types-extensions/>



Here is a screenshot of 4 shapefiles using Windows Explorer: Cities, Cities_dtl, Counties, and States. Each shapefile here consists of 7 individual files.

⇒ A shapefile is not one single file, but rather a family of 3 to 8 files.



There is the same folder when viewed in ArcGIS. ArcGIS manages and organizes the shapefiles for you and only shows you Cities.shp, Cities_dtl.shp, Counties.shp, and States.shp and hides the rest of the files of each shapefile family. This makes it easier to manage the files as you only deal with 1 file and ArcGIS handles the rest.

⇒ ArcGIS only shows you the 'head of the family'.