

GIS File Data Formats

In the GIS world, you will encounter many different GIS file formats. Some file formats are unique to specific GIS applications, others are universal. For this course, we will focus on a subset of spatial data file formats: **shapefiles** for vector data and **file geodatabases** for both vector and raster data.

Vector Data File Formats

Shapefile

A **shapefile** is a file-based data format native to ArcView 3.x software (a much older version of ArcMap). Conceptually, a shapefile is a feature class—it stores a collection of features that have the same geometry type (point, line, or polygon), the same attributes, and a common spatial extent.

Despite what its name may imply, a “single” shapefile is actually composed of at least three files, and as many as eight. Each file that makes up a “shapefile” has a common filename but different extension type.

The list of files that define a “shapefile” are shown in the following table. Note that each file has a specific role in defining a shapefile.

File extension	Content
.dbf	Attribute information
.shp	Feature geometry
.shx	Feature geometry index
.aih	Attribute index
.ain	Attribute index
.prj	Coordinate system information
.sbn	Spatial index file
.sbx	Spatial index file

File Geodatabase

A **file geodatabase** is a relational database storage format. It's a far more complex data structure than the shapefile and consists of a **.gdb** folder housing dozens of files. Its complexity renders it more versatile allowing it to store multiple feature classes and enabling topological definitions (i.e. allowing the user to define rules that govern the way different feature classes relate to one another). An example of the contents of a geodatabase is shown in the following figure.

Figure 3.1: Sample content of an ArcGIS file geodatabase.

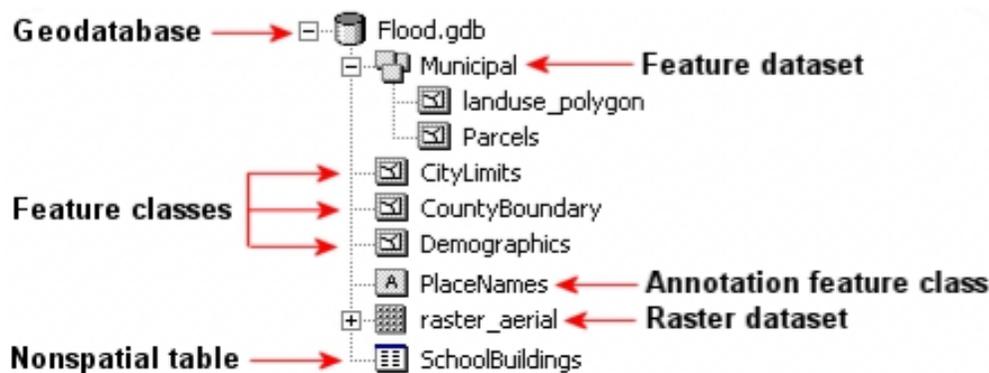


Figure 1. Sample content of an ArcGIS file geodatabase.

GeoPackage

This is a relatively new data format that follows [open format standards](#) (i.e. it is non-proprietary). It's built on top of SQLite (a self-contained relational database). Its one big advantage over many other vector formats is its compactness—coordinate value, metadata, attribute table, projection information, etc..., are all stored in a *single* file which facilitates portability. Its filename usually ends in **.gpkg**. Applications such as QGIS (2.12 and up), R and ArcGIS will recognize this format (ArcGIS version 10.2.2 and above will read the file from ArcCatalog but requires a script to create a GeoPackage).