ES341 – Overview of key file formats and file extensions in ArcGIS

Commonly Encountered File Types/Extensions in ArcGIS

.mxd   A file containing a map, its layers, display information, and other elements used in ArcMap. Double-click to open the document within ArcMap.

.mxt   A template file containing the information that can be contained in a map document. Often used to re–create standardized map layouts or provide standardized interfaces. Double-click to open the template with ArcMap.

.mxt   Map templates make it easy to reuse or standardize a layout, or even the same data or customization of the ArcMap interface, on a series of maps. Using a template can save you time since you don't have to manually reproduce the common parts of the maps. Like maps and layers, templates can be shared within an organization to increase productivity and standardize the maps that the organization produces.

.lyr   A layer is a set of rules for displaying and working with datasets in ArcMap, ArcGlobe, and ArcScene. Layer definitions include symbol assignments, classifications, labeling rules, and other information describing the appearance and actions of the layer. A group layer helps organize related kinds of layers in a map and can be used to define advanced drawing order options. To add a layer or group layer, drag-and-drop it into ArcMap, ArcGlobe, or ArcScene.

.pmf   A read-only map file created using the ArcGIS Publisher extension. Read-only maps can be used in ArcReader and ArcMap. They can also be served on the Web. Double-click to open it with ArcReader.

.xml   Metadata documents for individual file-based datasets are stored in XML files and are often managed in ArcGIS workspace folders. Click the Metadata tab to preview and work with metadata.

.prj   A file containing coordinate system information to provide spatial referencing of data in ArcGIS.

.shp   Shapefiles are a simple, nontopological format for storing the vector-based geometric location and attribute information of geographic features. A shapefile is one of the spatial data formats that you can work with and edit in ArcGIS. (see entry below for more details on shapefile architecture)

.txt   In ArcGIS you can directly access data in delimited text files and work with them as tables. ArcCatalog and the Add Data dialog box in ArcMap list files with .txt, .asc, .csv, or .tab extensions and assign them a file type of Text File.

Files with a .txt, .asc, or .csv extension are interpreted as comma-delimited, while files with a .tab extension are interpreted as tab-delimited by default. Any file with one of these extensions will be interpreted as a text file table even if it doesn't contain tabular data. If you attempt to display a text file that doesn't contain tabular data, the software will either produce an error or attempt to display the data as a table. To avoid this problem, give your delimited text files a .csv or a .tab extension. This will help differentiate text files with delimited data from unformatted text files.

*.xls   You can open Microsoft Office Excel tables directly in ArcGIS and work with them like other tabular data sources. For example, you can add them to ArcMap, preview them in ArcCatalog, and use them as inputs to geoprocessing tools.

Excel files are added to ArcMap like other data, through the Add Data dialog box. When you browse to an Excel file, you will need to choose which table you want to open. For example, if you have an Excel workbook called sales_figures.xls that contains three worksheets—Sales, Month, and Year to date— each worksheet is a separate table in ArcGIS. Any name references to cells or ranges defined in Excel are preserved in ArcGIS.

.xml—Metadata for ArcGIS—stores information about the shapefile.

Shapefile file extensions (vector data format)

Shapefiles are a simple, nontopological format for storing the geometric location and attribute information of geographic features. A shapefile is one of the spatial data formats that you can work with and edit in ArcGIS.

The shapefile format defines the geometry and attributes of geographically referenced features in three or more files with specific file extensions that should be stored in the same project workspace. They are:

.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.**

NOTE: 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. The record number forms the key field that connects all geometry/map elements to the attribute tables.

.sbn and .sbx—The files that store the spatial index of the features.

.fbn and .fbx—The files that store the spatial index of the features for shapefiles that are read-only.

.ain and .aih—The files that store the attribute index of the active fields in a table or a theme's attribute table.

.atx—An .atx file is created for each shapefile or dBASE attribute index created in ArcCatalog. ArcView GIS 3.x attribute indexes for shapefiles and dBASE files are not used by ArcGIS. A new attribute indexing model has been developed for shapefiles and dBASE files.

.ixs—Geocoding index for read-write shapefiles.

.mxs—Geocoding index for read-write shapefiles (ODB format).

.prj—The file that stores the coordinate system information. Used by ArcGIS.

.xml—Metadata for ArcGIS—stores information about the shapefile. "xml" stands for extensible markup language, similar to "html", hyper-text mark-up language. Both xml and html are codes that are used widely in creating formatted web-page environments. Browser tools such as Firefox and Internet Explorer read the xml and html files, and render their codes into formatted displays that you see on "web sites".

.cpg—An optional file that can be used to specify the codepage for identifying the characterset to be used.

Each file must have the same prefix. For example: roads.shp, roads.shx, and roads.dbf.

When viewing shapefiles in ArcCatalog (or any ArcGIS program), you will only see one file representing the shapefile; however, you can use Windows Explorer to view all the files associated with a shapefile. When copying shapefiles, it is recommended that you do so in ArcCatalog or by using a geoprocessing tool. However, if you do copy a shapefile outside ArcGIS, be sure to copy all the files that make up the shapefile.

_____

Projects and map documents

ArcView GIS 3.x (old desktop version) stores maps, charts, and tables you create in a project. A project is a file that organizes all the information you need to do your work. A project has an .apr file extension.

ArcMap in ArcGIS9.x stores maps, graphs, and tables in a map document. A map document is the disk-based representation of a map. Map documents have an .mxd file extension, and are project, pointer files that organizes all of the data and workspace variables that allow you to save and recall your work.

_____

Auxiliary Files

An auxiliary (AUX or AUX.XML) file accompanies the raster in the same location and stores any auxiliary information that cannot be stored in the raster file itself. An auxiliary file can store the following:

- Color map
- Statistics, histogram, or table
- Pointer to the pyramid file
- Coordinate system
- Transformation
- Projection information

_____

Georegistered Image Files

*.BIL image files

Band interleaved by line (BIL), band interleaved by pixel (BIP), and band sequential (BSQ) are three common methods of organizing image data for multiband images. BIL, BIP, and BSQ are not in themselves image formats but are schemes for storing the actual pixel values of an image in a file. These files support the display of single and multiband images and handle black-and-white, grayscale, pseudo color, true color, and multispectral image data.

_____

ArcInfo Export Files

*.E00 Files

An ARC/INFO interchange file can be used to transport coverages; INFO tables; text files, such as AML macros; and other ARC/INFO files between various machine types. An interchange file contains all coverage information and appropriate INFO table information in a fixed-length ASCII format.

_____

ASCII Text Files vs. Binary File Formats

ASCII Text Files - A text file (sometimes spelled "textfile": an old alternate name is "flatfile") is a kind of computer file that is structured as a sequence of lines. A text file exists within a computer file system. The end of a text file is often denoted by placing one or more special characters, known as an end-of-file marker, after the last line in a text file. In MS-Windows, text files are commonly given a *.txt file extensions.

Because of their simplicity, text files are commonly used for storage of information. A disadvantage of text files is that they usually have a low entropy, meaning that the information occupies more storage than is strictly necessary. A simple text file needs no additional metadata to assist the reader in interpretation, and therefore may contain no data at all, which is a case of zero byte file.

ASCII Grid The ArcInfo ASCII Grid format is an ArcInfo Grid exchange file with a single file extension *.asc

Binary File Formats: A binary file (commonly, but not necessarily, with the extension .bin) is a computer file which may contain any type of data, encoded in binary form for computer storage and processing purposes; for example, computer document files containing formatted text (e.g. MS Word *.doc)

Binary files are usually thought of as being a sequence of bytes, which means the binary digits (bits) are grouped in eights. Binary files typically contain bytes that are intended to be interpreted as something other than text characters. Compiled computer programs are typical examples; indeed, compiled applications (object files) are sometimes referred to, particularly by programmers, as binaries. But binary files can also contain images, sounds, compressed versions of other files, etc. — in short, any type of file content whatsoever.

Some binary files contain headers, blocks of metadata used by a computer program to interpret the data in the file. For example, a GIF file can contain multiple images, and headers are used to identify and describe each block of image data. If a binary file does not contain any headers, it may be called a flat binary file.

_____

ArcInfo Grid Format

BND table - A grid BND contains the boundary of the grid. The boundary is a rectangle that encompasses the cells of a grid; it is stored in map coordinates. All grid BNDs are stored in double precision.

The minimum coordinates in the BND are for the lower left corner of the lower left cell in the grid. The maximum coordinates are for the upper right corner of the upper right cell in the grid.

HDR file - The HDR is a binary file. Information stored in the file includes the cell size, type of grid (integer or floating point), compression technique, blocking factor, and tile information.

STA table - The STA table is an INFO table that contains statistical data about a grid. The minimum, maximum, mean, and standard deviation for the grid are stored as floating-point values in the STA table. You should not attempt to alter these values directly.

Because NoData represents an unknown value, NoData is not used in calculating the statistics in the STA table.

VAT table - The VAT is an INFO table that stores attributes associated with the zones of a grid. Only integer grids have a VAT associated with them. Every VAT has at least two items, VALUE and COUNT. The VALUE item contains integer values that are used to distinguish the characteristics of one location from the other locations in a grid. All cells that are assigned the same value contain the same characteristics and, therefore, belong to the same zone. COUNT is the number of cells in a zone.

New items can be added to the VAT. The VALUE and COUNT items should not be changed, and the VAT must be kept sorted on the VALUE item. Never add new items before VALUE or COUNT.

Google Earth KML Files

Keyhole Markup Language (KML) is an XML-based language provided by Google™ for defining the graphic display of spatial data in applications such as Google Earth™ and Google Maps™. KML enables these applications to support the open integration of custom data layers from many GIS users.

Each KML file is composed of a collection of graphic elements, images, and settings. KML is used to:

Symbolize and display GIS data as elements within Google Earth™ and Google Maps™ using symbols, color, images, and balloon-style information popups.

Provide access to attribute information about geographic features -- for example, by presenting attribute information when you click on a feature's placemark.

Define the user's interaction with those features -- for example, to control fly-to and camera location settings in Google Earth™.

KML files have either a .kml file extension or a .kmz file extension (for compressed and zipped KML files).

In addition, map services published using ArcGIS Server can be served using KML. The KML content is retrieved each time the map service is accessed from ArcGIS Server. This enables KML content to be served dynamically using ArcGIS Server.

Microsoft Access Database Files

Access .mdb, you should not open or modify the tables in a personal geodatabase inside Microsoft Access. This can corrupt the geodatabase and cause data loss.