

Data Types and Data Creation

Goal

Learn about GIS Data Management and best practices.

Outline

- How does a Computer Work?
- Different Places Data can Live
 - Where do I Store Data for This Class?
- GIS Data Formats
- Working with ArcGIS Pro Projects
- Download GIS Data
- ArcGIS Pro DEMO
 - Adding Data
 - Using ArcGIS Catalog
 - Saving in Different Formats
 - Fixing Broken Files

Sections Covered in Learning ArcGIS Pro 2

- Chapter 5: Creating and Working with Projects

Today's goal is to understand GIS Data Management, introduce our next assignment, and familiarize ourselves with the terminology we will be using during the class.

The Basics: About Our Computers

Lets start with the basics. It's not really about GIS, but it's good to know.

How Your Computer Works

How Computer Memory Works

Our computers work on two levels: we have harddrives and the CPU. The CPU is typically controlled by an operating sytem whcih loads a bunch of programs to a User Experience GUI for the purpose of running various operations. Additionally, this can give permissions to accounts which can lock down access to programs, files, etc.

So for me, where do I store this?

The H Drive.

The H Drive is a Network Drive, so it is accessible from machines that are on the University of Puget Sound Network. This is a computer hard drive on a server, which your accounts are given permission to access via some server protocols <-(whatever this may be).

This network allows us open your drive from every UPS computer. However, because it is over the network, we have to pass those files over the internet before we can actually access them.

If you work on the same computer, you can work in My Documents or another folder on the C Drive. This is a local harddrive on your computer, so it is able to grab data quickly and might give you all a better experience for working with Arc.

Additionally, you may want to work on an external harddrive. You likely would need something with at least 32 GB, but these are so cheap now you may want to just buy something like a TB.

You could either work directly off your External Harddrive or just use it to backup your work.

All this to say, I am anticipating our labs to take some of you up to four hours to complete. Additionally, it is likely that your final research project will be something like 25-30 hours of work. We don't want to lose that, so this is important!

Just to cover our bases we can also talk about cloud storage and other services.

Cloud Storage and Data Services

Cloud Storage is similar to a network drive, but the actual harddrive is hosted by a third party that you pay to keep your software running. Google Drive, Microsoft One Drive - all of these are providing the same services. For more technical implementations, you use some utility to connect as a user with certain privileges.

Data can also be hosted on a server, then shared publically via Services and Application Programming Interfaces.

Examples in GIS:

- Feature Service
- Map Service

These services are sharing data on the web using an agreed protocol which allows different applications to display, query, and even in some cases analyze different datasets.

Here is some documentation for Feature Services in ArcGIS. ArcGIS Feature Service

SO, WOW!!! THIS IS ALL COOL BUT WHAT DO I ACTUALLY DO NOW?

Well, let's get into the data we are actually going to be managing.

Vector Data Formats:

- Shapefile
- FileGeodatabase
- Geopackage
- GeoJSON

Shapefiles

A shapefile is a file-based data format native to ArcView 3.x software (OLD!!!)

- A Basic Feature Class
 - collection of features that have the same geometry type (point, line, or polygon), the same attributes, and a common spatial extent.
 - At least three different files, but up to eight

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.

HOLD UP!!!

A Relational Database Format is a pretty loaded term, and is used pretty frequently in GIS Data Management.

I wasn't going to cover this yet, or is it necessary for this class; but I would highly recommend that you familiarize yourself with this if you are going to work with large datasets in the future. (Possibly even next week!?)

Relational Databases IBM

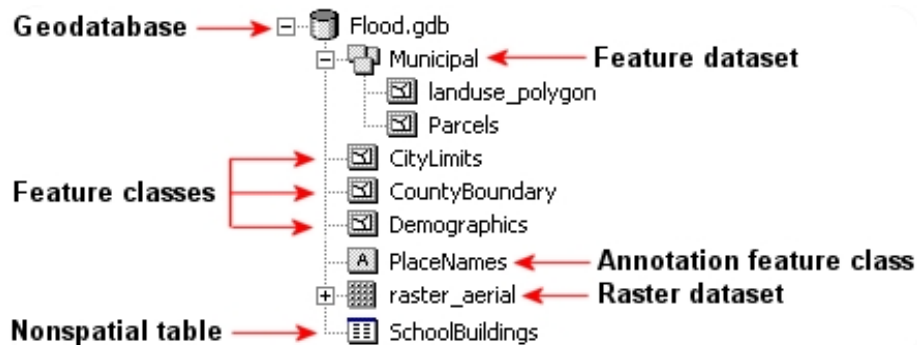
Back to our previous programming...

Basically, a File Geodatabase is a .gdb folder.

- Hosts hundreds of different files - Can store multiple feature classes - Increased complexity allows for topological definitions to be set at the folder rather than individual file.
- For example, projections

But its format is proprietary. This means it works best with only ESRI stuff and typically you need some other knowledge about how to transform the file format.

An Example of a File Geodatabase.



(src: ESRI)

Geopackage

A Geopackage is a relatively new file format (I've never used one) using Open Data Standards. Its built on another relational database called SQLite which is a standalone file.

- .gpkg
 - Stores all data as one file
 - * coordinate value
 - * metadata
 - * attribute data
 - * projection information

Relatively new, so have to have used newer programs to use this format.

GeoJSON

GeoJSON is a file format used in the web. Often data from APIs is delivered in JSON, a readable text format which is easily parseable in many different code languages. GeoJSON is an extension of that JSON format to store geographic data.

Here is the GeoJSON Spec.

However, some things to note.

- By the Spec, you should only store data in EPSG:4326
- Can store different geometry types in a single file.
- Mostly used in the web.

```
{
  "type": "FeatureCollection",
  "features": [
    {
      "type": "Feature",
      "geometry": {
        "type": "Point",
        "coordinates": [0, 0]
      },
      "properties": {
        "name": "null island"
      }
    }
  ]
}
```

Raster Data Formats

Rasters are defined by pixel depth.

1 bit (binary)



2 bits



4 bits



8 bits



16 bits



Examples:

Aerial Imagery: Stored in three bands - Red: 255 pixels - Green: 255 pixels - Blue: 255 pixels

These three bands together are what make a clear image.

For a video on RGB, take a look at this. This will also be helpful later in the quarter when we think about styling our data and how to chose the right color.

RGB

Image Files

The Imagine file format was originally created by an image processing software company called ERDAS (I've mentioned this before).

Sometimes contains an xml file for metadata

GeoTiff

Good easy open file data format.

File Geodatabase

Again, Rasters can be stored in your File Geodatabase

Additional Benefits:

- Can create image mosaic structures
 - “stitched” images from multiple image files stored in the geodatabase
- Processing very large raster files can be computationally more efficient than other file formats

Cool lets do something now

Seatbelts everyone. *Please let this be a normal fieldtrip.* With Arc? No way!

ArcMap is the replacement for ArcGIS Desktop, which was a system of software used to do the same thing we are doing here in ArcGIS Pro. ArcGIS Desktop has an End of Life of March 01, 2026, and there is not future development anticipated for the software until that time.

Also, this means agencies are phasing out its use. For example, DFW is no longer supporting ArcGIS Desktop in Decemeber 2022.

So, we are using ArcGIS Pro in this class.

Opening ArcGIS Pro.

We covered this in class last week, but I figured that we would revisit working in ArcGIS Pro.

You have already opened an existing ArcGIS project, but lets start from scratch.

If we open ArcGIS without a template, it opens up a blank document.

As discussed in the assignment, Opening and Saving an ArcGIS Pro Document creates a folder with a specific structure.

- **Future_Project**
 - **Some Other Folders** <- Backups

- **Future_Project.gdb** <- Geodatabase
- Future_Project.aprx <- ArcGIS Pro Project
- Future_Project.tbx <- ArcGIS Toolbox

Typically we will want to keep all data related to our project within the same folder.

Before ArcGIS Pro...

There were many applications used to work together to do GIS.

- ArcMap was used to create maps and do analysis
- ArcCatalog was used to manage GIS data
- etc.

Now ArcGIS Pro handles all of that for us.

We can open a Blank Template, and Map Template, or a Catalog Template. However, its easy enough to be able to add any of these panes/views in any project, and I expect you to be familiar with each of them to be able to complete a project.

But since we are talking about data, lets start talking about the Catalog Pane/View.

We want to use the Catalog Pane to manage most of our GIS Data. One, it reads information in File Geodatabases, and is smart enough to keep all our file together.

Lets Download some data:

DNR Roads
National Parks
Shaded Relief

Looking at these files from Windows File Explorer vs ArcGIS Catalog.

DEMO Examples:

- Create a Map Pane/View and lets look at some data.
- See how there is now a map in the Catalog Pane.
- Right click layer in catalog, and add to map.
- Click and drag another layer and add to map.

Break Stuff Example:

- Move a shapefile to a different location or rename.
 - Fix broken links to the file.
 - Add folder to ArcCatalog project.

DEMO Examples:

Lets add our data from our Survey123.

- Login to ArcGIS Online.

- Go to Portal Content
 - My Orgs or My Groups
 - * ENV250 My First Survey

DEMO Examples:

- Coping Data into our own GeoDataBase
 - Converting Formats
 - Export Data