DCN Planetears

Geospatial Data Curation:
an introduction
Module: Common GIS Data Types

Common GIS Data Types

Module Objectives: Common GIS Data Types

This module has four objectives. At the end of the module, learners should be able to:

1. Differentiate between raster data, vector data, and geodatabases;
2. Identify one common file format for vector data;
3. Identify one common file format for raster data;
4. Gain an understanding of complex GIS file structures.

The module has Lecture, Activity, and Quiz components to help reinforce new information.

Common GIS Data Structures

File/Data Structures

1. Tabular;
2. Vector;
3. Raster;
4. Databases; and,
5. GIS Projects.

Learn to:

1. Recognize;
2. Open; and,
3. Assess

All Geospatial Data Types: Assess

cURATED

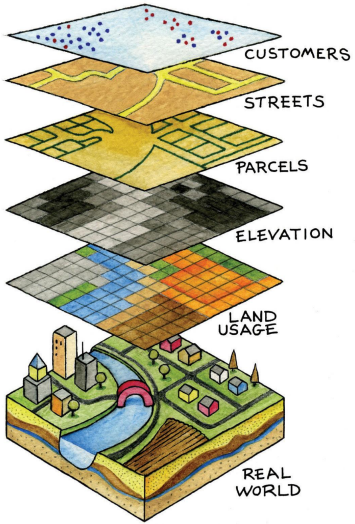
C
U
R
A
T
E
D

Questions to ask about all Geospatial Data file types:

1. Are the filenames descriptive enough to ascertain what they display?
2. Do files open in GIS or other software properly?
3. Are the relationships between the data layers clear?
4. Is the documentation robust and complete enough for re-use?

Vector and Raster Data Structures

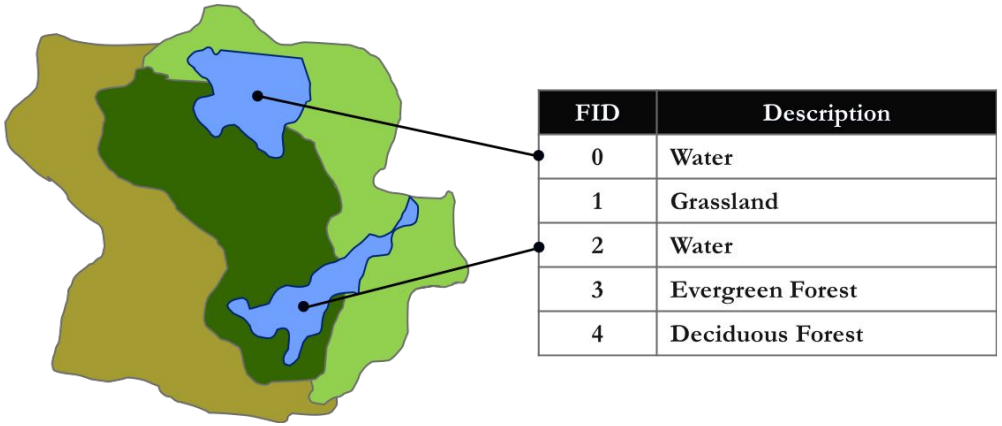
	Raster Structure	**Vector Structure**
Physical Model (files on disk)	geotiff, grid	shapefile, geodatabase, json
Logical Model (vector and raster)	Raster Model	Vector Model
Conceptual Model (discrete or continuous)	Object View Continuous	Object View: Discrete
Reality (the world out there)	Entity: Temperature, Topography	Entity: Trees, Houses, Streets



adapted from: https://saylordotorg.github.io/text_essentials-of-geographic-information-systems/index.html

Vector Data Model
Points, Lines, and Polygons

- all based on x,y,(z) coordinate pairs of geographic data
- lines and polygons are built from groups of points
- attribute data is linked to points, lines, or polygons (features)
- each feature is associated with a unique record in an attribute table



FID	Description
0	Water
1	Grassland
2	Water
3	Evergreen Forest
4	Deciduous Forest

Image adapted from original by; Barry Nickel

Vector Data : Recognize 01

CURATED

ATTRIBUTES

FEATURES

UID	Name	Type	County	State
001	Oak Tree	Natural	St. Louis	MO
002	Bus Stop	Transport	St. Charles	MO
003	Fire Hydrant	Emergency	Jefferson	MO

Image credit; Jennifer Moore

Vector Data: Recognize 02

CURATED

There is only one geometry type per feature class.
Geometries don't mix in a dataset, but they can be layered in a project.

points lines polygons

Image credit: Jennifer Moore

Vector Data: Common File Types

Extension	Description	QGIS	ArcGIS Pro	R, Python	Text Editor
.shp .dbf .shx .sbx .sbn .prj .xml .cpg .qmd	shapefile	✓	✓	✓	
.json .geojson	geojson	✓	✓	✓	✓
.kml .kmz	keyhole markup (google)	✓	✓	✓	✓ (kml)
.csv	comma separated values	✓	✓	✓	✓
.gdb	Geodatabase (vector)	✓	✓	✓	

Note that this list is not exhaustive, see the cheat sheet for more extensions and formats.

Tabular Data: Recognize

CURATED

Common ways to represent geospatial information in a spreadsheet

1. Latitude and Longitude
2. Address (Street Address, City, Zip Code, State, Country)
3. “Join” field

warblerNest	latitude	longitude	AW_prod	GW_prod
1	47.017853	-95.5818	2.00266	1.04543
2	47.012411	-95.6164	0.833307	2.00133
3	47.014676	-95.5626	0.619742	2.42882
4	47.012044	-95.6176	0.842108	1.58087
5	47.012215	-95.6177	0.834223	1.55911

Tabular Data: Open

cURATED

C
U
R
A
T
E
D

How to open a spreadsheet with lat/long coordinates in QGIS:

1. Go to Layer → Add Layer → Add Delimited Text Layer
2. Select the AMWO_nests_P.csv file and choose which fields to use for the longitude values (X) and the latitude values (Y)
3. Check that the correct CRS is selected
4. Review the spreadsheet headers and data slice in the preview and then click “Add”
5. (If needed) Add a basemap by going to Web → QuickMapServices → and choosing an option (e.g. OSM Standard)
6. Open the attribute table and check column headers and values.

Tabular Data: Assess

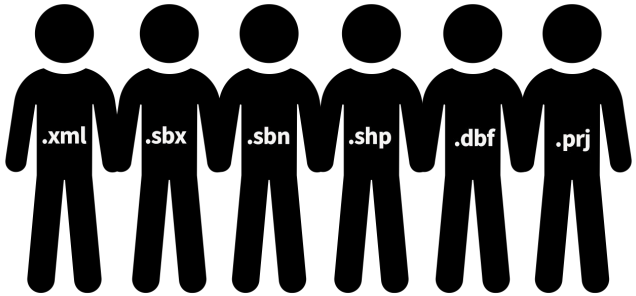
C
U
R
A
T
E
D

Questions to Ask about Tabular Files:

1. When you inspect the spreadsheet, does anything seem strange?
2. When opened in GIS software, does map location information match the textual description of dataset? (e.g. Do points show up within the expected extent?)
3. Is the projection / coordinate reference system documented somewhere?

Vector Data: Shapefiles

C
U
R
A
T
E
D



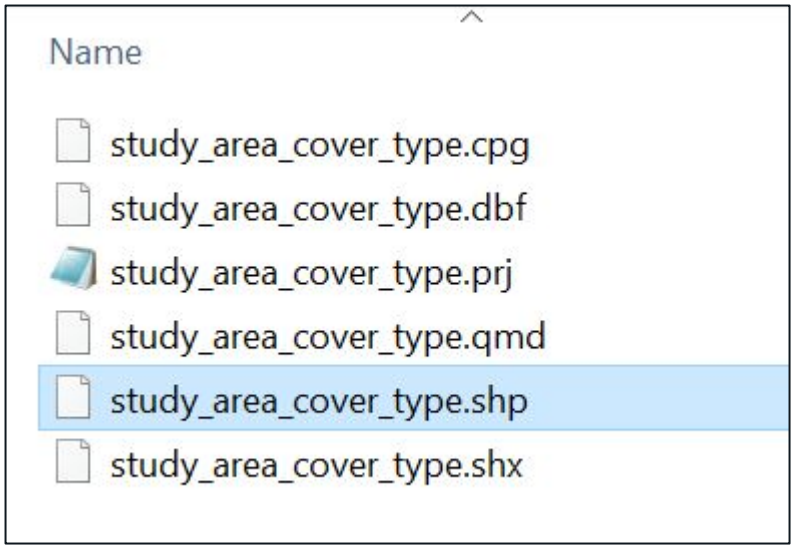
Example Vector Data File Format: Esri Shapefile

1. Primary files: .SHP, .DBF, .SHX
2. Possible supplemental files: .PRJ, .XML, .CPG, .SBX, .SBN, .QMD

Vector Data: Open

cURATED

C U R A T E D



How to open a shapefile in QGIS:

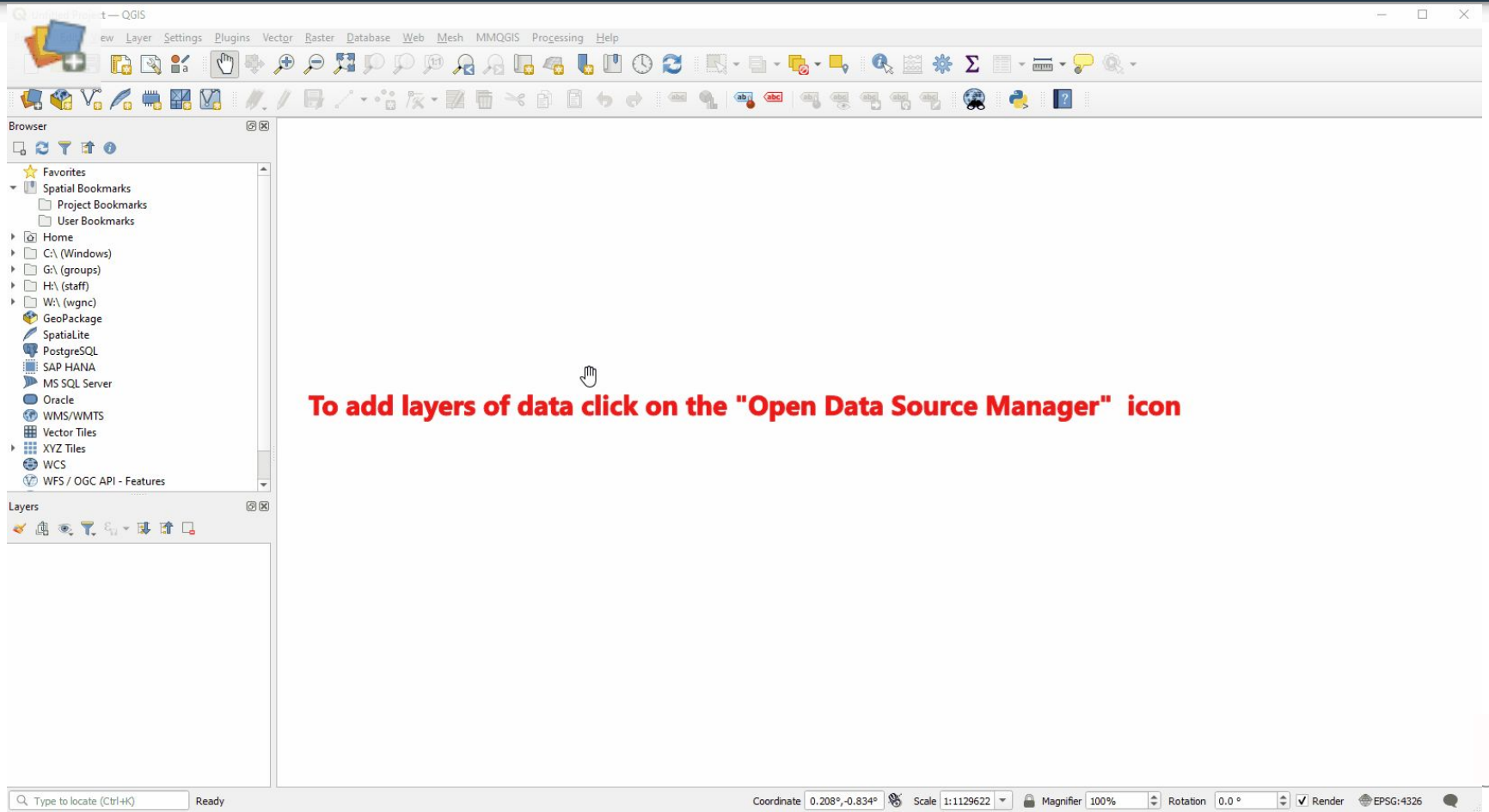
1. Go to Layer → Add Layer → Add Vector
2. Navigate to the shapefile and select the .shp file. Click “Add.”

Adding Vector Data in QGIS: Activity

1. Click on the “Open Data Source Manager” icon 
2. Select “Vector” from the side menu
3. Click on the three dots next to the Source box
4. Locate where the data layers are saved and select the SHP file extension
    a. Select the file *0_study_area_cover_type.shp*
5. Click the “Add” button

CURATED

Adding Vector Data in QGIS: Example Video



Video credit: Kelly Grove

Vector Data: Assess

cURATED

C
U
R
A
T
E
D

Questions to Ask about Vector Files:

1. (For shapefiles) Are the mandatory files present (e.g. shp, shx, dbf)
2. When opened in GIS software, does map location information match the textual description of dataset?
3. When you inspect the attribute table, does anything seem strange?

Dixson, Nadia; Milliken, Genevieve; Mukunda, Keshav; Murray, Reina; Starry, Rachel. (2019). GeoJSON Data Curation Primer. http://hdl.handle.net/11299/210208

Vector Activity

Download and unzip (if needed) the provided example GIS datasets.
Link to example dataset: 1-4_Excercise_Dataset

1. Look at the files in the subfolder labeled study_area_cover_type, what tool or tools can you use to open this data?
2. Is there a specific file extension you would load into the tool to view the data?

Raster Data: Recognize

CURATED

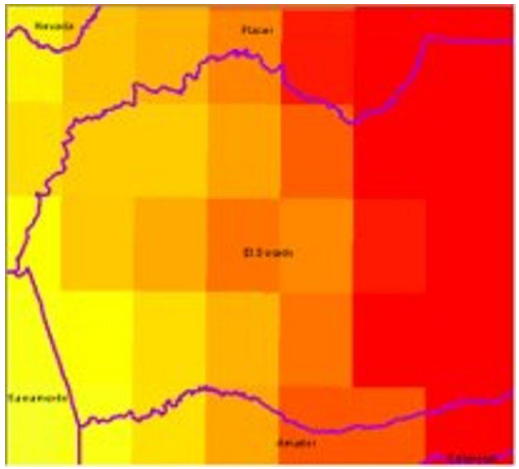
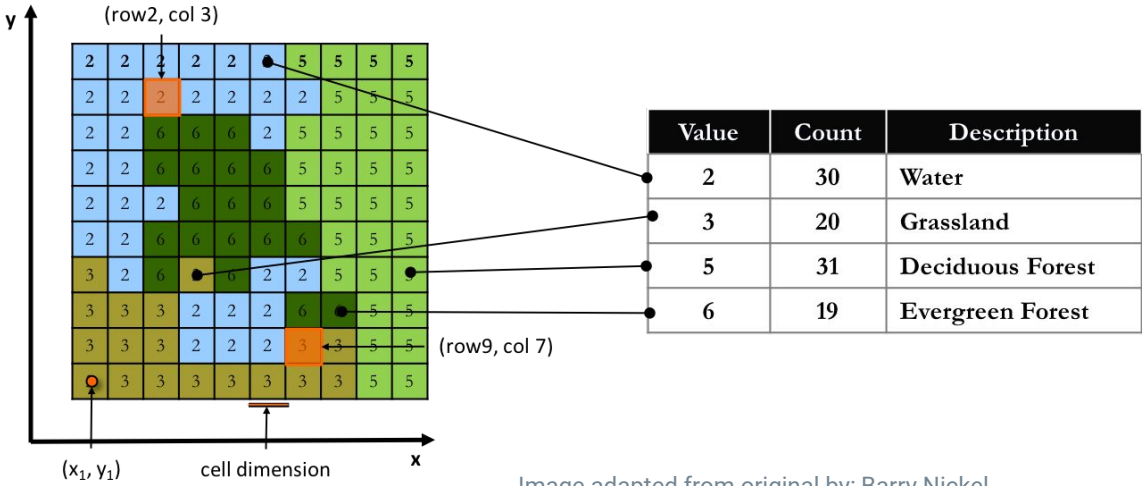


Image credit: Melisa McClean

Raster Data Model
Grids and Images



Value	Count	Description
2	30	Water
3	20	Grassland
5	31	Deciduous Forest
6	19	Evergreen Forest

Image adapted from original by; Barry Nickel

- each cell represents an x,y coordinate
  - each cell has a specific size on the surface of the earth (scale)
  - cell scale is based on the resolution of the image
- each cell has only one value (color, numeric, or categorical)

Raster Data: Common File Types

Extension	Description	QGIS	ArcGIS Pro	R, Python
.tiff, .tif	geoTIFF	✓	✓	✓
.img	ERDAS Image	✓	✓	✓
.adf	ESRI Grid (legacy, in a folder)	✓	✓	✓
.gdb	Geodatabase (raster)		✓	

Note that this list is not exhaustive, see the cheat sheet for more extensions and formats.

Raster Data: Recognize

C
U
R
A
T
E
D

Example Common Raster File Format: GeoTIFF

1. Primary files: .TIF or .TIFF
2. Supplemental files: .TFW, .OVR, .XML

Image credit: Melisa McClean

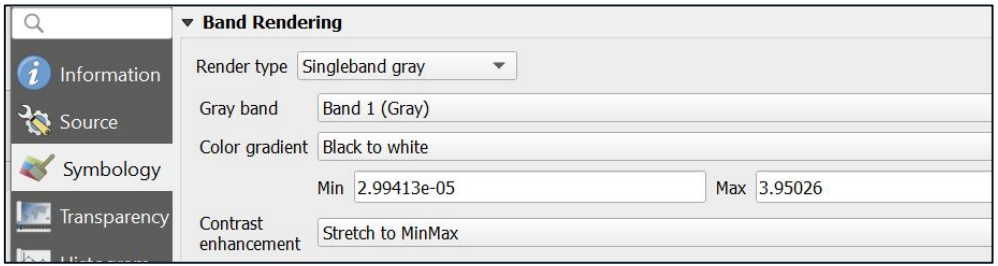
Raster Data: Open

cURATED

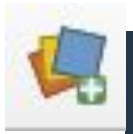
C U R A T E D

How to open a GeoTIFF in QGIS:

1. Go to Layer → Add Layer → Add Raster. Select the .TIF or .TIFF. Click “Add.”
2. Open the Properties. Look at the Information and Symbology tabs to see whether there are multiple bands or options to edit how the raster displays

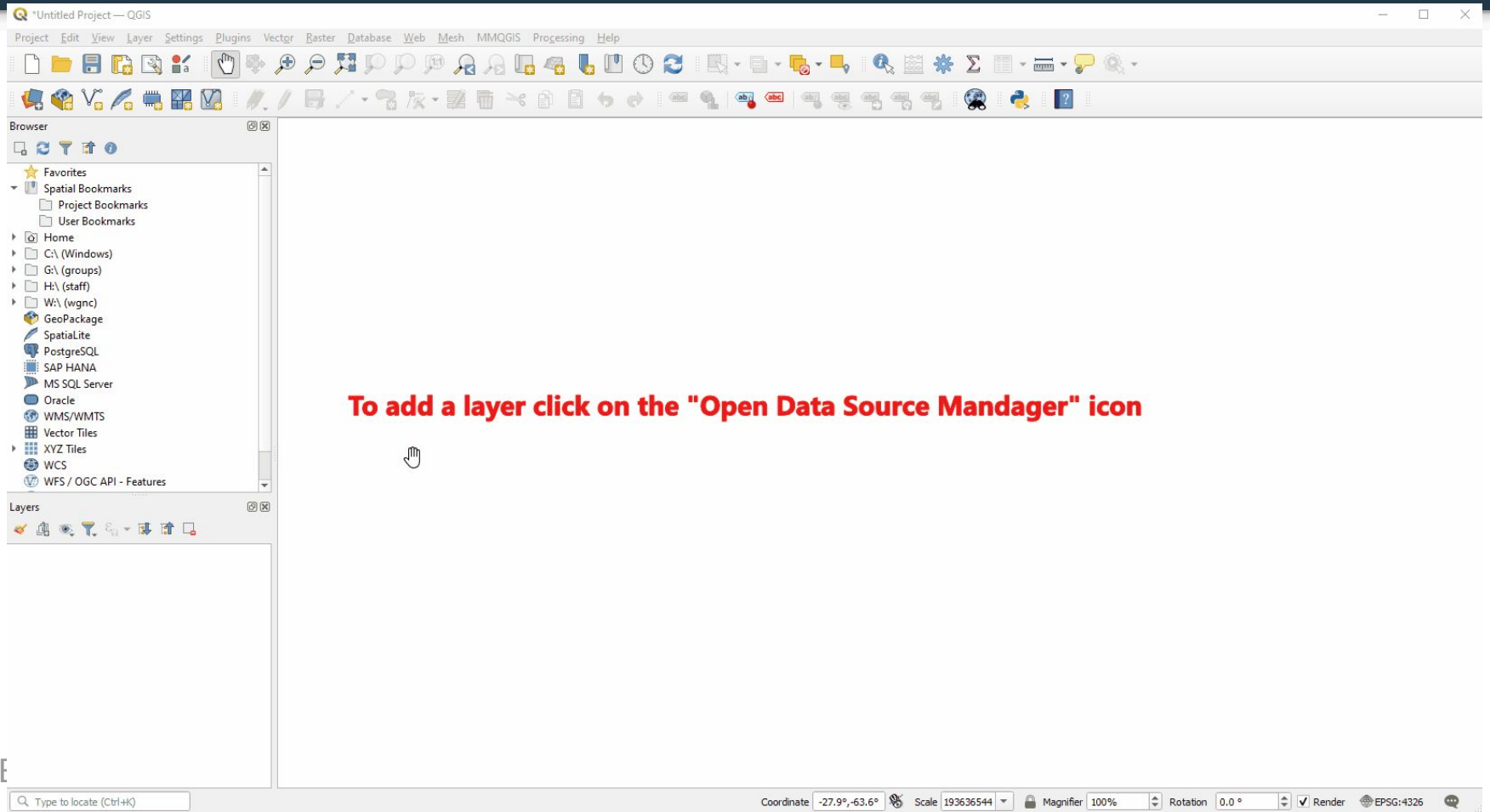


Adding Raster Data in QGIS: Activity

1. Click on the “Open Data Source Manager” icon 
2. Select “Raster” type from the side menu
3. Click on the three dots next to the Source box
4. Locate where the data layers are saved and select the .tif file type
    a. Select the files *warblerProductivity.tif* and *woodcockProductivity.tif*
5. Click the “Add” button

CURATED

Adding Raster Data in QGIS: Example Video



Raster Data: Assess

cURATED

C
U
R
A
T
E
D

Questions to Ask about Raster Files:

1. Are measurement units, temporal information, and spatial information recorded?
2. Does data contain multiple layers or bands?
3. Are the files georeferenced? (i.e. does the raster “snap” to the right spot on the map?)

Kearney, Courtney; Ruhs, Nick; Sedlins, Mara; Tien, Tracy; Trelogan, Jessica; and Watts, John. (2020). GeoTiff Data Curation Primer. http://hdl.handle.net/11299/216574

File Type Activity

Download and unzip (if needed) the provided example GIS datasets.
“Dataset_No_ReadMe.zip” can be found in the folder:
1-4 Exercise_Dataset

1. Identify which files are a vector type and which are a raster type.

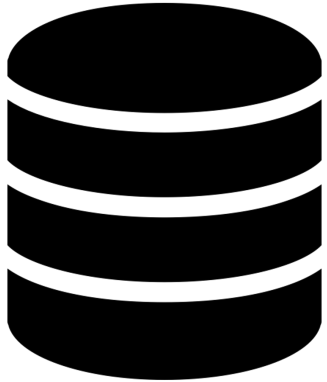
Database: Recognize

C U R A T E D

A data structure that can contain multiple tabular, vector, and raster data layers

Common GIS Database File Formats:

1. ESRI File Geodatabase .GDB
2. OGC GeoPackage .GPKG



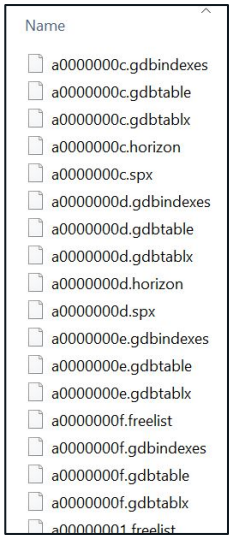
Created by araeva from Noun Project

Databases: Open in QGIS

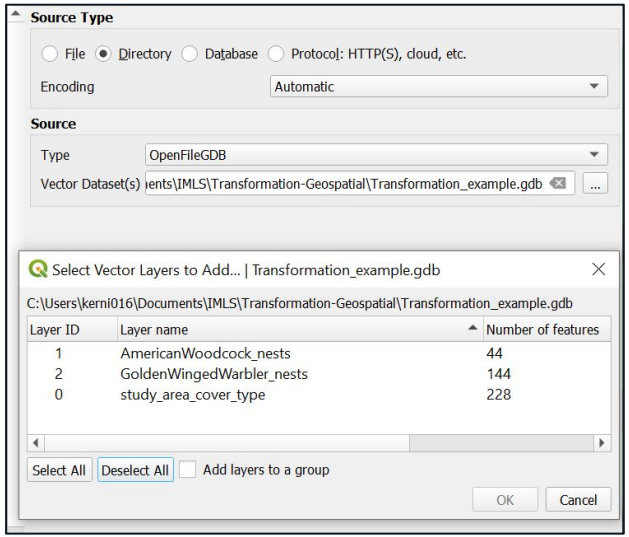
cURATED

C U R A T E D

In a geodatabase directory folder in File Explorer



In QGIS: **Can only open .gdb vector layers**

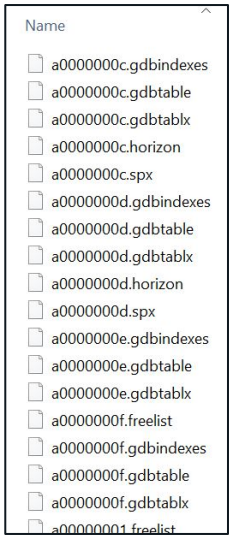


Databases: Open in ArcGIS

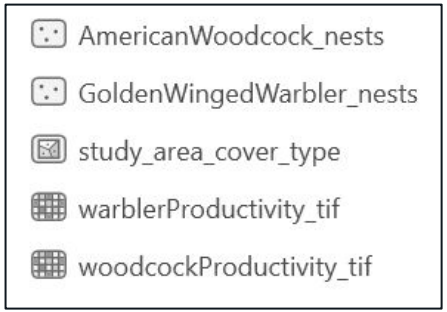
cURATED

C U R A T E D

In a geodatabase directory folder in File Explorer



In ArcGIS Pro:



Databases: Assess

cURATED

C
U
R
A
T
E
D

Questions to Ask about Database Files:

1. Are there any files stored in the geodatabase that are not in the typical geodatabase format? (i.e. do not begin with the letter a followed by a series of numbers and letters)
2. How many files are contained within the geodatabase? Are any of them repetitive?
3. Is there enough data provided to understand how the data files within the geodatabase were created and what they are intended to display?

GeoDatabase (.gdb) Data Curation Primer. Battista, Andrew; Brittnacher, Tom; Garrett, Zenobie; Moore, Jennifer; Pirmann, Carrie (Data Curation Network, 2019). https://hdl.handle.net/11299/202823

GIS Projects: Recognize

C U R A T E D

Description: GIS project files are used in GIS applications. They hierarchically store layers and then display them in a layout. They retain symbology, queries, labeling, and other properties for building maps. They may also contain tools, models, or workflows and non-geospatial files.

Common GIS Project File Formats:

1. QGIS Project Files: .QGS (2.X), .QGZ (3.X)
2. ArcGIS Pro Project Files: .APRX; .PPKX
3. Legacy ArcGIS Map: .MXD

Reviewing Common GIS Data Types: Visual Aid

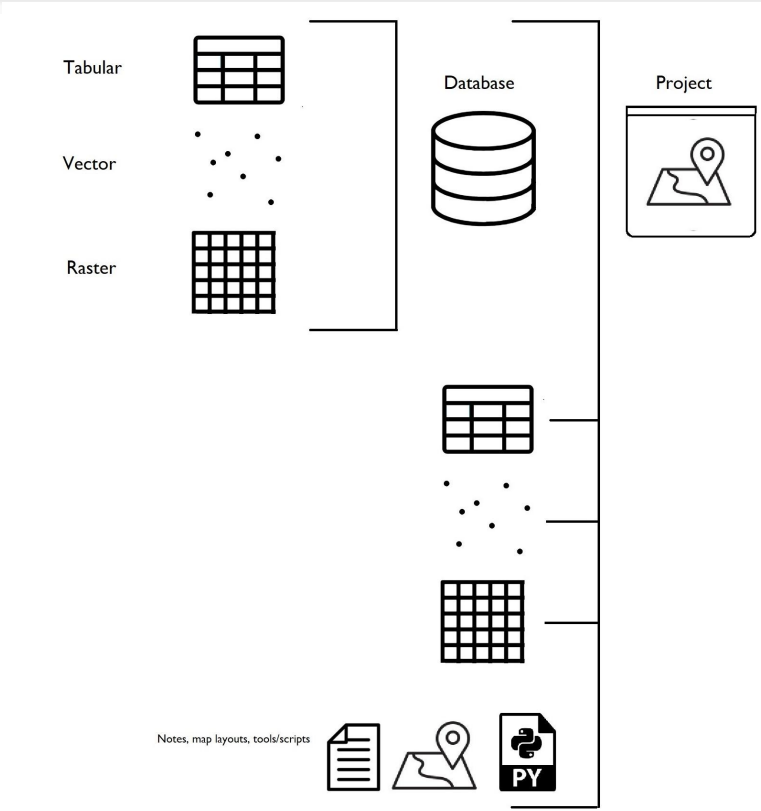


Image credit: Melinda Kernik

GIS Projects: Open

C
U
R
A
T
E
D

Project files are software (and version) specific:

- Use QGIS for .QGS or .QGZ
- Use ArcGIS for .APRX, .PPKX, and .MXD

GIS Projects: Assess

cURATED

C
U
R
A
T
E
D

Questions to Ask about GIS Project Files:

1. Does the project open correctly (are the links relative)?
2. Are there any files stored in the project folder that are not in typical GIS formats?
3. How many data layers are contained within the project? Are any of them repetitive?
4. Is there enough documentation provided to understand how the data files within the project were created and what they are intended to display?
5. Could you make sense of the data (ie. how the files are related) without the project file?

How to Approach Unfamiliar GIS File Formats

If you don't recognize a file format:

- It is okay! This is very common when curating GIS data
- Look it up in the Library of Congress digital formats list
- Attempt to open the file with QGIS
- Ask the researcher for information about what software was used to create and view the data

Common GIS File Types Activity

1. Download and unzip (if needed) the provided example GIS datasets with an appropriate tool.
   a. Link to example dataset: Kramer_2019_Birds.zip
2. Based on the file extensions you find there, identify which tool you will use to open the dataset.
   a. Answer:
3. Still in the file folder, identify the file types you find there by Type (Vector, Raster, etc.) and File Extension.
   a. Answers:
4. Now open “Working_01” with the tool you identified above and make yourself familiar with how the data displays and the menu layout of the tool.
5. Now make a copy of your “Original” dataset and label it “Working_02”.
6. Open Working_02 with software tools that are NOT indicated for that data type.
   a. For example, many Vector and Raster files can be opened with image processing tools, such as PhotoShop, Paint, etc.
      i. Try this, and note the differences in how the data displays differently and/or similarly to the appropriate GIS tool.
   b. Now try various text editors, spreadsheet tools, etc.
      i. Note how each tool impacts the data display.
      ii. This could have an impact on how you transform and preserve the data!
   c. Record your observations:

Common GIS File Types Quiz: 01

Quiz Questions 1 to 5 (of 8)

1. Describe, in your own words, the characteristics of raster data. Include at least 1 difference between raster and vector data.
    a. Answer:
2. Describe, in your own words, the characteristics of vector data. Include at least 1 difference between vector and raster data (different from the one you used above).
    a. Answer:
3. Describe, in your own words, the characteristics of GIS database.
    a. Answer:
4. Name a common raster GIS file type.
    a. Answer:
5. Name a common vector GIS file type.
    a. Answer:

Common GIS File Types Quiz: 02

Quiz Questions 6 to 8 (of 8)

6. Name 2 common GIS database file types.
    a. Answer:
7. Name 2 GIS Project file types.
    a. Answer:
8. Name 2 common GIS software tools
    a. Answer:

Common GIS Data Types Review

C U R A T E D

Common GIS Data Types Review:

1. Vector: A coordinate-based data model. Scales well.
   a. .SHP; .GEOJSON; GML; .GPX; .OSM; .CSV
2. Raster: A spatial data model that defines space as an array of equally sized cells arranged in rows and columns.
   a. GeoTIFF: .XML; .ASC; .IMG
3. Databases: Includes data about the spatial locations and shapes of geographic features.
   a. .GDB; .GPKG; .MBTILES
4. GIS Projects: All hierarchically store layers and then display them in a layout.
   a. .QGS; .QGZ; .APRX; .MXD; .MXT; .WOR

Tools You may Use:

- QGIS
- ArcGIS
- Open Street Map
- Text editors
- Many Other tools