# GEOG 487 Lesson 3: Step-by-Step Activity

## Part I: Explore Publically Available Wetlands Data & Historical Imagery

In Part I, we will explore our study area and the time-series aerial photos used to digitize the vegetation data we will use in Part II. We will also look at two publically available datasets specifically related to wetlands: the National Wetlands Inventory from the U.S. Fish & Wildlife Service and a more detailed wetlands inventory from a regional public agency called the Great Lakes Commission.  In the process, we will explore several different data delivery options and sources in ArcGIS Desktop: Esri Map Packages, Esri Basemaps, ArcGIS Online Datasets, Web Map Services (WMS) from GIS Servers, and raw GIS files.

1. **Familiarize Yourself with the Study Area and Set Up Your Map**

    a. Open the L3Data folder you downloaded in the Lesson Data section. Double click on the "Lesson3.mpk" map package file. This will open a map inside ArcGIS Desktop with the study area boundary and historical imagery already loaded.

    > *i* You can share your own data and maps by creating a map package in ArcMap. Go to File > Share As > Map Package. The file can either be uploaded to ArcGIS Online or saved locally.

    b. Set your Current Workspace and Scratch Workspace to your L3 folder (Geoprocessing> Environments…>Workspace). Make sure to read the associated help topics about "Current Workspace" and "Scratch Workspace."
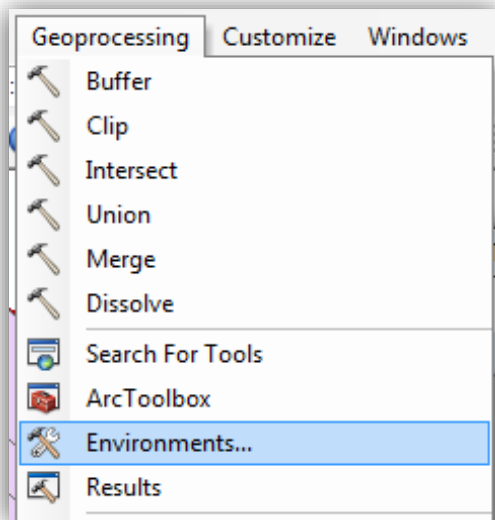
> Setting your Current Workspace allows you to customize the location of where output files created during geoprocessing steps are saved. By default, files will be saved at …My Documents\ArcGIS.
>
> There are two ways to easily access help files related to tools in ArcToolbox. The first is a button that says "Show Help>>," which provides a brief overview of the tool within the dialog box itself. The second is a button that says "Tool Help," which links to full help articles. It is a good idea to read the help information for tools you are not familiar with.

    c. Add the "Open Street Map" ArcGIS Online Service (File > Add Data > Add Basemap) so you can tell where the study area is located in relation to the other places. You may need to refresh your map for the Basemap to load. (You also need to be connected to the internet).

    d. Use the zoom and pan tools to explore the surrounding area. What is the nearest major city?  How far away is the site from the state border with Michigan?

    e. Turn off the Basemap, as this can slow down the drawing speed of your map in late steps.  Save your map.

2. **Explore Historical and Recent Aerial Photos**

    a. Zoom to the study area boundary by right clicking on it in the Table of Contents > Zoom to Layer.

    b. Turn on the "2005" layer in the Table of Contents. This is a Color Infrared (CIR) image. Notice the information on the edge of the scanned film showing the date, location, and scale of the original image.

    c. Turn on the "1973" and "1962" images. These images are black and white, a common format before 2000.

    d. Compare the three images by turning them on and off and viewing them at a number of different scales: 1:15,000, 1:10,000, and 1:5,000. Try panning to different locations within the study area.  What differences do you notice between the three images?

> The swipe and flicker tools available on the Customize > Toolbars > Effects Toolbar are useful for temporal change detection (especially of satellite images or air photographs taken at different times of the same location), data quality comparison, and other scenarios where you want to visually compare the differences between two layers in your map. Swipe allows you to interactively reveal what is underneath a particular layer; Flicker flashes layers on and off at the rate you specify.  You can read more about these tools in the Esri help.

e.  The photos that were used to create the detailed vegetation data we will use in Part II all show the study area in the past. Let's take a look at more current imagery and see if we notice any changes in the vegetation. We'll use an image from 2015 from the National Agricultural Imagery Program (NAIP).

f.  Connect to the GIS Server within ArcMap. File > Add Data > Add Data > Select "GIS Servers" from the drop down menu > Add ArcGIS Server > Use GIS Services > Type **https://gis.apfo.usda.gov/arcgis/rest/services** in the Server URL >Finish.

g.  You should see a new data service appear.
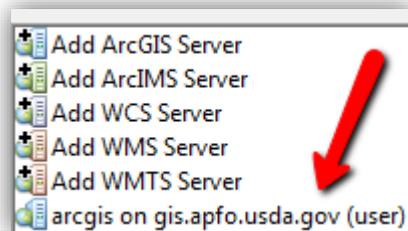


h.  Double click on the layer name to expand the folder options. Go to NAIP > Ohio > Add.
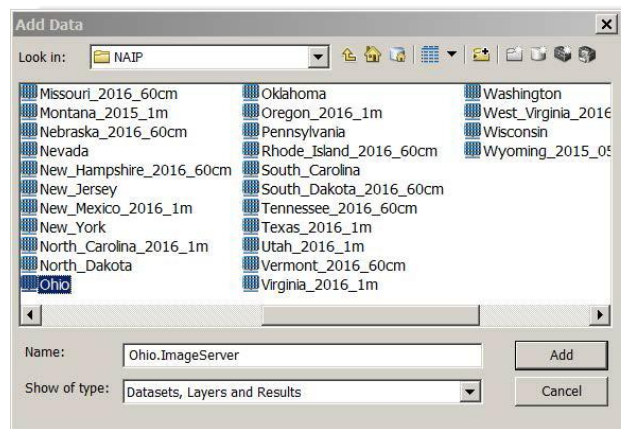


h.  You **ma**y see a geographic warning pop up box. Read the "About the geographic coordinate systems warning" and then close the pop up box.

j.  Compare the 2015 image with the historical aerials. What types of changes do you see? Notice that the NAIP image is natural color, that is a bit easier to interpret (as compared to color infrared and black and white formats).

k.  Turn off the imagery layers in the Table of Contents and save your map.

3. **Add the boundaries of the Fish & Wildlife Service Wildlife Refuges to your map**

   a. Click on the "AddData" dropdown link within ArcMap (to the right of the yellow plus sign). Click "Add Data From ArcGIS Online..."
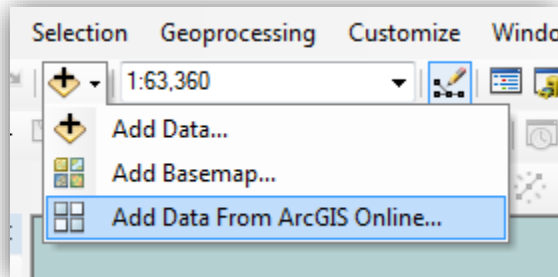


   b. In the ArcGIS Online search window type "wildlife service lands May 2010" and click on the magnifying glass to search the available data. Review the datasets that match the search. Locate the "U.S. Fish & Wildlife Service Lands – May 2010" dataset in the results. (You may need to scroll down awhile to find it).

   c. Click on the "Details" button to review the metadata.
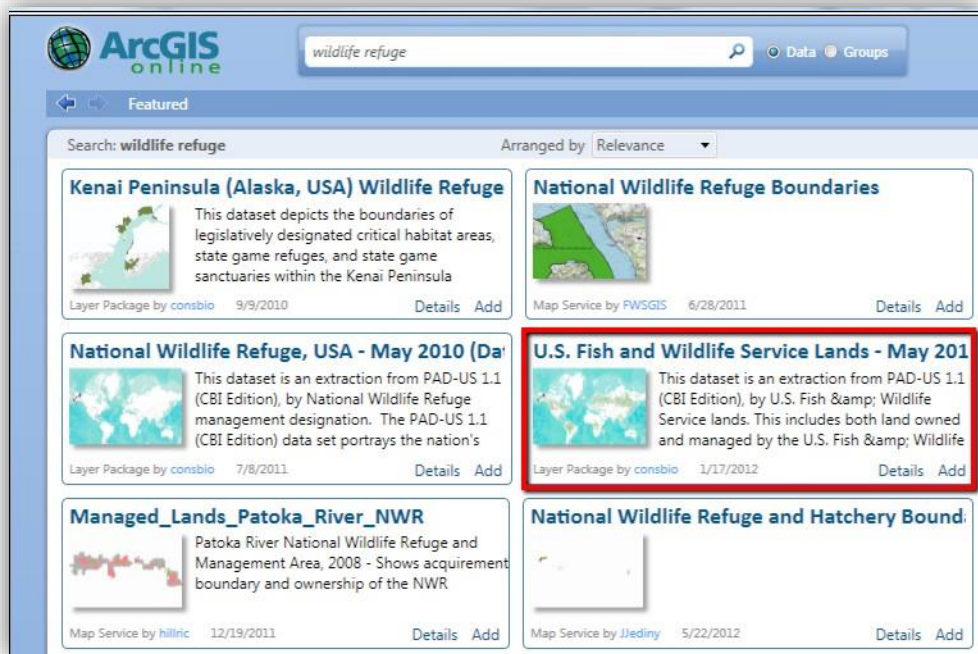


   d. Click on the yellow plus sign at the top of the description to add it to your map. Review the alignment warning message if it pops up.

e.  Drag the study boundary to the top of the Table of Contents so you can see the site boundary on top of the feature service.

f.  Zoom out to 1:250,000 so you can see the various federal lands in the vicinity.    The refuge is shown as a purple filled in polygon, which means it is a National Wildlife Refuge.   Use the identify button to find the name of the refuge.

g.  Notice how the lesson study area only covers a portion of the whole refuge. The study site boundary represents the wetland areas within Ottawa National Wildlife Refuge that are hydraulically connected to Lake Erie. Some of the wetlands in the refuge are excluded from this area because they are hydraulically separated from Lake Erie by dikes and are therefore not susceptible to fluctuations in Lake Erie water levels. Other areas of the refuge are not wetlands.

h.  Explore the attribute table.  Are there any abbreviations you don't understand?

i.  Right click on the "U.S. Fish and Wildlife Service Lands _ May 2010" layer > Properties > Source.  Look at the spatial reference information.  Is it the same as the study site? Does it match the data frame? (Right click on Layers in the Table of Contents > Properties > Coordinate Systems). **Note: You may experience errors editing a data layer if it does not match that of the data frame.**

4. **Create a new shapefile of the Ottawa National Wildlife Refuge Boundary**

a.  Right now, the refuge boundaries on our map are part of a layer package. We want to create a new shapefile from just a portion of the records so we can customize it for our study site.

b.  Open the U.S. Fish and Wildlife Service Lands _ May 2010 attribute table.  Click on the "Select by Attributes" icon.



c.  Use the wizard to select all of the polygons within the Ottawa National Wildlife Refuge ("p_des_nm" = 'Ottawa National Wildlife Refuge'). You should have two records selected.  Close the attribute table.

d.  Right click on the layer in the Table of Contents and click Data > Export Data. Export the selected features to a new shapefile called "OttawaNWR" in your L3 folder. **Make sure to use the same coordinate system as the layer's source data or you will not be able to edit this data later on in the lesson.**

e.  Add the "OttawaNWR" shapefile to your map when prompted. Remove the "U.S. Fish and Wildlife Service Lands _ May 2010" dataset from your map and save.

5. **Merge Ottawa National Wildlife Refuge Polygons**

   a. In step 4, we saw that there are two separate records for the Ottawa National Wildlife Refuge. We want our final dataset to have only one polygon, so we can symbolize it on our final maps as one outline.

   b. Use the Merge tool to collapse all of the polygons into one.

      i. Customize > Toolbars > Editor > Editor > Start Editing. Choose the "OttawaNWR" shapefile.

      ii. Select all of the polygons in the "OttawaNWR" shapefile.

      iii. Left click on the Editor Toolbar drop down menu > Merge. Choose either of the input polygons when prompted.

      iv. The shapefile should now only have one polygon. Confirm this by opening the attribute table. Save your edits, stop editing, and save your map.

   > *Some of the tools on the Editor Toolbar have the same names as tools available in ArcToolbox, but they are actually different. According to the Esri help, "the Merge and Union commands on the Editor menu create features in an existing feature class, whereas the Merge and Union geoprocessing tools create features in a new feature class." Another difference is that the "geoprocessing tools work on all features in a layer rather than just the selected features."*

6. **Download and Explore National Wetlands Inventory (NWI) Data**

   a. Connect to the National Wetlands Inventory data service. Add Data > GIS Servers > Add WMS Server > Copy and paste the URL:
   **https://www.fws.gov/wetlands/arcgis/services/Wetlands/MapServer/WMSServer?**

   b. > Press "Get Layers" > OK.

   c. Navigate to the "Wetlands" layer and add the data set to your map.

   d. Zoom to the study area boundary and explore the data.

   e. Use the Identify tool to see what information is included with the layer. Notice that you can't view an attribute table like you could with the NWI layer.

   f. According to the metadata, "the data are intended for use in publications, at a scale of 1:24,000 or smaller. Due to the scale, the primary intended use is for regional and watershed data display and analysis, rather than specific project data analysis."

   g. Unfortunately, the data does not contain enough detail to help us analyze vegetation changes in a specific wetland. It also does not allow us to study changes over time.

   h. Turn off the "WMS-Wetlands" layer in the Table of Contents and save your map.

7. **Explore the Great Lakes Coastal Wetland Inventory Data**

   a. Read a quick description of the data at <u>Great Lakes Commission</u> (scroll down to the "Links to Inventory and Metadata" heading). Notice the link to the product metadata provided on the site. We are going to use one file from this site called, "the complete polygon coverage in shapefile format."

   b. Download the data zip file from the site or from the Lesson Data page (<u>glcwc_cwi_polygons</u>.zip - 12.46 MB), save it in your L3 folder and unzip the file.

   c. Add the "glcwc_cwi_polygon" shapefile to your map.

   d. Open the attribute table and explore the available information. Notice how there is significantly more information than the previous wetland datasets we reviewed.

   e. The "HGM_CLS1" and "HGM_CLS2" attributes show the wetland classification codes. You can read detailed descriptions of these codes in the "Full Publication (enhanced metadata)" document. (Look in the Lesson Data tab for Lesson 3 under the "Great Lakes Coastal Wetland Inventory").

   f. Right click on the "glcwc_cwi_polygon" shapefile in the Table of Contents > Zoom to Layer.

   g. Zoom back to the study area. Do you see any wetlands inside the Ottawa National Wildlife Refuge that are not inside our study site boundary? You may need to move the study boundary and OttawaNWR layers to the top of your Table of Contents.

   h. Turn the layer off in your table of contents and save your map.

   i. The Great Lakes Coastal Wetland Inventory provides much more detailed attribute information than the National Wetlands Inventory. However, it still doesn't provide the time series information we need to answer our research questions.  For time series data, we need to look at yet another data set.

## Part II: Explore & Customize Site Level, Time Series Wetlands Data

The publicly available data sets we just explored are helpful for familiarizing yourself with your study area or for regional or other large-scale analyses. However, they do not contain the level of detail for the site level analysis we want to conduct. The highest resolution data you can usually find is 1:24,000 scale for vector data and 30 m cell size for raster data. Publicly available data sets also typically do not have time series information available. In Part II, we are going to explore a high resolution, time series dataset that was digitized from the aerial photos we reviewed in Part I. Often times, you will need to digitize information in this manner if you have a small study site or if you want to do an in-depth, time series analysis. The work required to create the data is significant. However, you can do a lot more with your data.

We want to explore how vegetation has changed over time in our study area. To answer our research questions, we need the following datasets: 1) polygons of vegetation species over time 2) polygons of vegetation groups over time and 3) polygons of invasive species over time. All of the files need to show just the region within our study site. We will create these custom datasets for three time periods using the Join, Union, Clip, and Dissolve tools. The workflow we will follow is illustrated in the diagram below using the data for the seventies time period. You may wish to consult this diagram after completing each step.
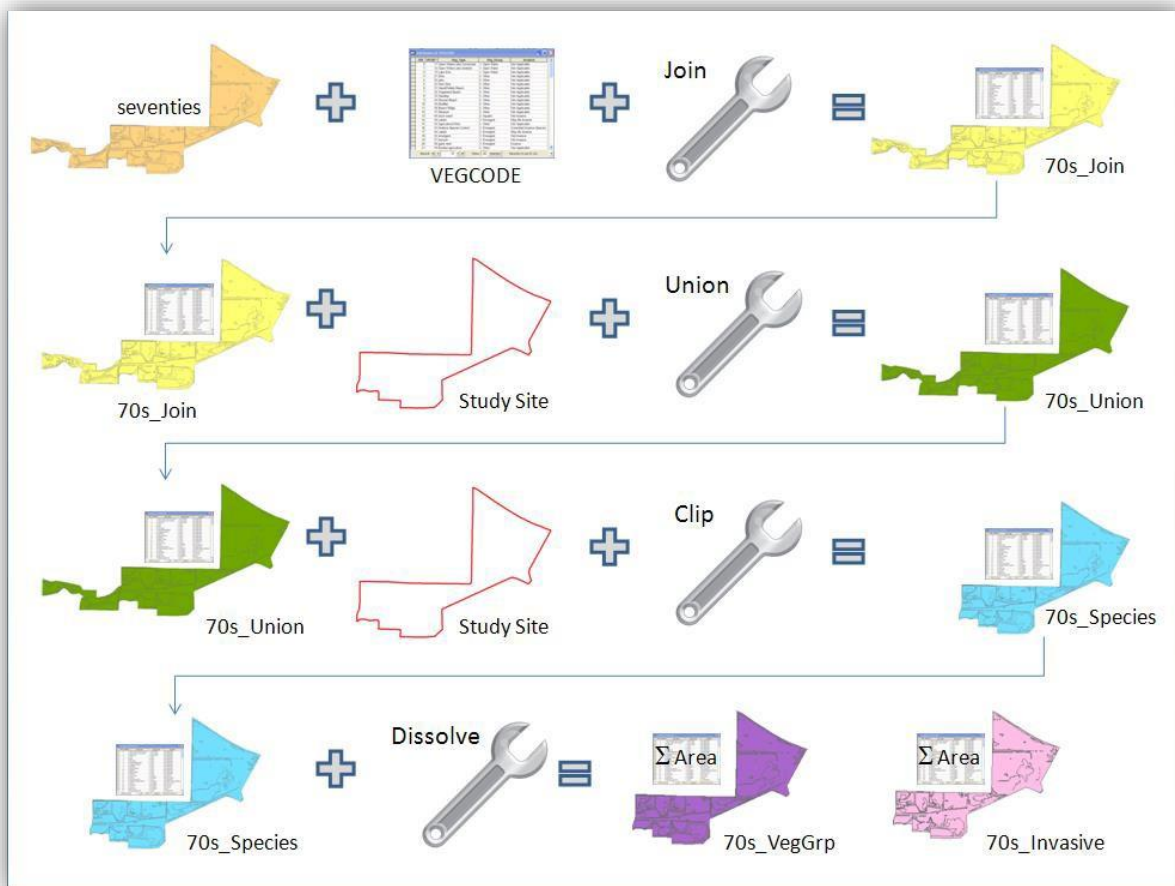
1. **Explore the Site Level Vegetation Data**

   a. Open the map from Part I. Make sure you are in the "Data View" instead of the "Layout View."
   b. Add the "sixties," "seventies," and "twothousands" vegetation shapefiles from the L3 folder to your map.
   c. The default symbology should show the vegetation polygons filled in. Do you see any gaps in coverage between the vegetation data and site boundary? Hint: turn some of the layers on and off and use the zoom tools. Look along the coast line along the northeast boundary.
   d. Compare the extents of the vegetation data and study site. Do you see any differences?
   e. Open the attribute tables. Do you notice any differences in the number of records in each dataset? Do you see any coded or missing values? Missing data may sometimes be coded as values of "0."
   f. Notice how the vegetation files contain a lot more spatial detail than the publicly available data sets we looked at earlier. At this point we do not know what the values in "VEG_ID" mean, though we can assume they correspond to different types of vegetation. Even without knowing what the "VEG_ID's" mean, we can still tell that the "twothousands" data has a lot more polygons than the other time periods. What do you think the VEG_ID code "11" means?

2. **Understand Coded Values**

   a. Now that we have a general sense of what our starting data looks like, we can work on customizing it for our purposes. Let's start by figuring out what the coded values in the "VEG_ID" fields mean.
   b. Add the "VEGCODE.dbf" table from your L3 folder to your map. Open the table. You may need to switch to the "List by Source" option in the Table of Contents to see the table.
   c. The VEGCODE table is a master lookup table that tells us what the coded values (VEG_IDs) mean. The VEG_IDs correspond with detailed vegetation types (Veg_Type).
   d. I have reclassified this information for you into 2 simpler categories: Veg_Group and Invasive. The numbers at the beginning allow us to sort the values based on the depth of water they prefer (e.g. most water (open water) to least water (upland vegetation) instead of alphabetically by name).

   Note: You may notice that some of the Veg_IDs are listed as "May Be Invasive" in the "Invasive" field. Two of the most common invasive species in the wetland (narrow-leaved and hybrid narrow/broad leaved cattails) look very similar to native species (broad-leaved cattails,) which makes them difficult to distinguish in aerial photos.

e. Join the VEGCODE table to the sixties vegetation shapefile. Right click on the "sixties" shapefile > Joins and Relates > Join. Base the join on the Veg_IDs. Keep all records.

Caution – watch out for similar attribute names like OID. This is not the same as Veg_ID.
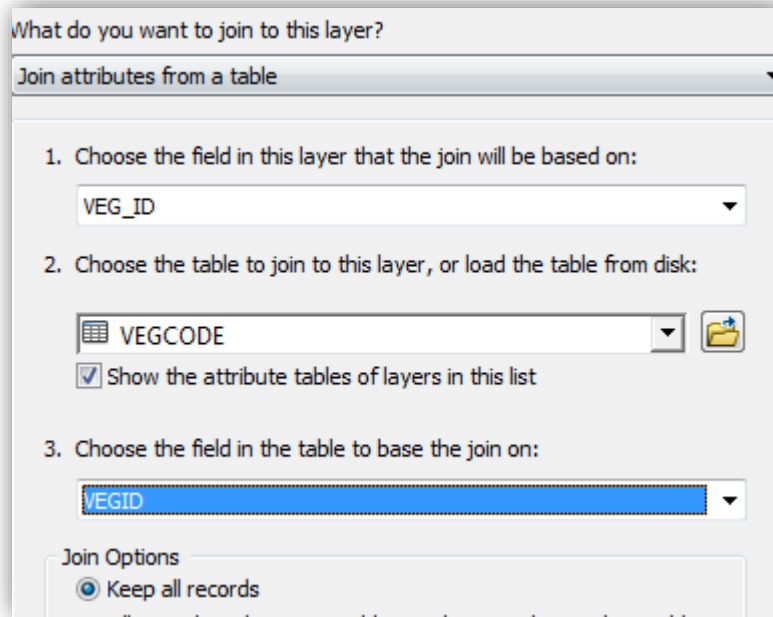


f. Open the attribute table of the sixties shapefile to make sure the join worked properly.
g. To make the join permanent, export it to new shapefile in your L3 folder called "60s_Join." Add it to your map when prompted.
h. Repeat steps e - g for the remaining two vegetation data sets. Name them "70s_Join" and "00s_Join." Remove the "sixties," "seventies," "twothousands," and VEGCODE table from your map and save.

> **Make sure you have the correct answer before moving on to the next step.**
>
> The 60s_Join, 70s_Join, and 00s_Join shapefiles should have the number of records and all of the attributes shown below. If your data does not match this, go back and redo the previous step.

70s_Join – 280 records.  00s_Join – 449 records.

3. **Modify Extents - Fill in Gaps**

   a. We want all of our input data sets to have the same total area so we can compare changes in the area of different vegetation types over time. This means we need to remove pieces from some study years and fill in gaps for other years so they all cover the same extent.

   b. First, we will fill in gaps. A quick way to fill in gaps is to use the union tool to create new polygons in areas that overlap within the extent of two datasets. We will union the vegetation shapefiles with the study boundary, since this file does not have any gaps and it covers the area we are interested in.

   c. Follow the steps below to union the two files:

      i. Geoprocessing > Union.
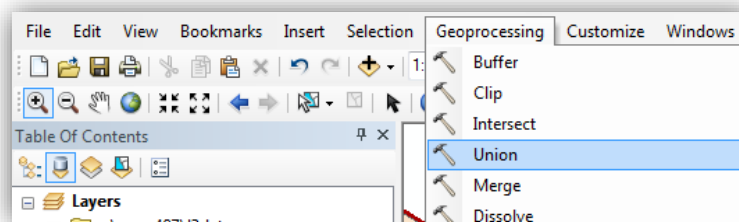


      ii. Click on the "Show Help" button for more information about what the tool does and what the different input criterion mean.
      iii. Input Datasets: Study Site; 60s_Join (make sure the study site is listed first).
      iv. Output Dataset: 60s_Union (save it in your L3 folder).
      v. Keep the defaults for Join Attributes (ALL) and XY Tolerance (leave blank).
      vi. Make sure the "Gaps Allowed" check box is <u>NOT</u> checked. Read the help topic about this so you understand why.

vii. Compare the output file to the original shapefile from the same time period to make sure the tool worked as expected. Notice the records that have an FID_Study_ value of -1. What do these mean? Hint: See step vi.

d. Repeat union for the remaining two vegetation shapefiles "70s_Join" and "00s_Join." Name them "70s_Union" and "00s_Union."
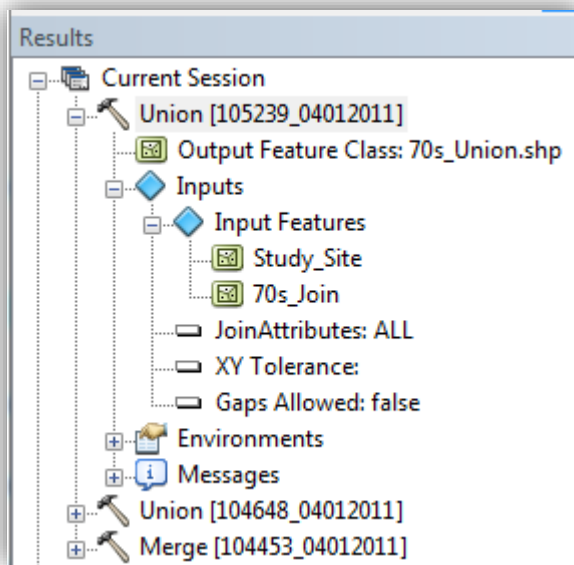
---

*Geodatabases may have naming restrictions for table and field names. For instance, a table in a file geodatabase cannot start with a number or a special character such as an asterisk (\*) or percent sign (%). Shapefiles do not have such restrictions and allow us to use names such as 60s_Join.*

*If you receive an Error 000361: The name starts with an invalid character during geoprocessing check to make sure you are saving your output as a shapefile.*

---

*It's very easy to make mistakes when using geoprocessing tools. For example, you can select the wrong input files by mistake. Another common error is running tools while unknowingly having records selected. Any output from geoprocessing tools will only contain the selected records. Comparing your results with your input datasets after using automated tools is a good habitat to get into.*

*If you want to double check the input files you used, the parameter settings, environment settings, etc., you can view them under Geoprocessing > Results.*

```
Results
└─ Current Session
   └─ Union [105239_04012011]
      ├─ Output Feature Class: 70s_Union.shp
      └─ Inputs
         └─ Input Features
            ├─ Study_Site
            └─ 70s_Join
         ├─ JoinAttributes: ALL
         ├─ XY Tolerance:
         └─ Gaps Allowed: false
      ├─ Environments
      └─ Messages
   ├─ Union [104648_04012011]
   └─ Merge [104453_04012011]
```

4.  **Modify Extents - Remove Excess Area**

    a.  Notice that some of the study years still cover a larger area than others. For example, compare the 60s_Union shapefile with the 70s_Union shapefile. Let's change the extent of all data sources to match the study area boundary by using the clip tool. Note: The Clip tool is only for vector data sets. In later lessons, we will look at tools to clip raster data sets.
        i.   Geoprocessing > Clip.
        ii.  Click on the "Show Help" button for more information about what the tool does and what the different input criterion mean.
        iii. Input Features: 60s_Union
        iv.  Clip Features: Study_Site
        v.   Output Feature Class: 60s_Species (save it in your L3 folder)
        vi.  Keep the defaults for XY Tolerance
    b.  Compare the output file to the joined shapefile and unioned shapefile from the same time period.  What differences do you notice?
    c.  Clip 70s_Union and 00s_Union using the directions above. Name them "70s_Species" and "00s_Species."

    d.  Remove the original vegetation, unioned, and joined shapefiles from your map and save.
    e.  All of our study years should now have the same extent. Let's confirm this by calculating the area of each study year. Add a new double field to each year called "sqm" with the default scale and precision values. I find it helpful to name fields by their units so I remember what they mean later on.

Specifying a specific precision and scale when adding a field to a shapefile gives you the option to limit the number of digits (precision) and decimal places (scale) of values within number fields. There are many situations where you would want to do this. However, there are also occasions where it is best to keep all of your options open. Accepting the default value of 0 for both properties gives you the most versatility. It may seem counterintuitive, but the value of 0 acts somewhat similar to the value of infinity in this case.
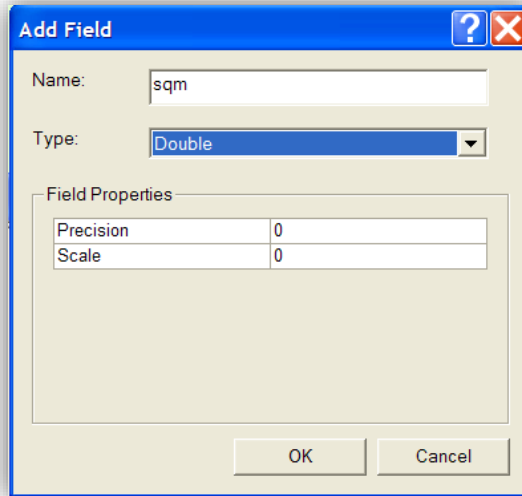
Setting custom precision and scale values is only relevant to data stored in an enterprise or workgroup geodatabase. Default values are always enforced when editing data in a shapefile or file geodatabase. Refer to the ArcGIS field data types for more information.

I recommend using values of 0 when you are in the preliminary stages of data exploration.  That way you won't unknowingly exclude values in your results. For example, if you are calculating area values for the first time you probably won't know how many digits you will need to store your calculated values (precision) until after you've made the calculation. If you estimate a number to use for precision that ends up being too low, you will not be able to store the full range of values. For example, a precision of 2 would limit your values to two digits, whereas a precision of 4 would limit your values to four digits.

f.   Right click on the "sqm" field and click "Calculate Geometry." Choose area, and units of square meters. Use the coordinate system of the data source. Repeat this step for the remaining two shapefiles.

g.   Use the statistics tool to find the total area for each year by right clicking on "sqm" field and choosing "Statistics".  All of the study years should have the same "SUM" value. You may notice very small differences between the layers (3310948.183528, 3310947.900263, and 3310947.900555). This is due to tiny topology errors such as overlapping sliver polygons. We could have corrected these with the XY Tolerance settings during our union and clip operations if we needed this level of precision for our analysis. In this case we don't, but I wanted to point out this issue in case you come across it in other projects.  For more information about XY Tolerance, see the Esri Help.

5. **Explore Attributes & Missing Data**
   a. Now that we've fixed the geometry of our input data, we can start to work with the attributes. Before inputting data into an analysis, you should have a good understanding of the distribution of your values. You should also be aware of any missing values or outliers that can skew your results, so you can exclude or recode these if necessary.
   b. One way to quickly get a sense of the distribution of data values and missing data is to change the symbology to be categorical based on each attribute.
   c. Right click on the "60s_Species" shapefile in the Table of Contents > Properties > Symbology > Categories> Unique values > Value Field "Veg_Type." Click "Add All Values." You may need to left click on the "count" field to populate the values. This is essentially a histogram of each variable.
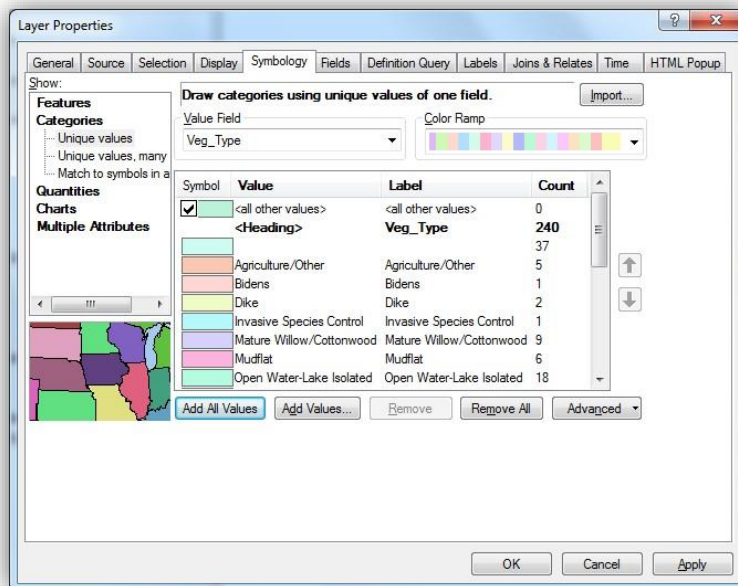


   d. How many polygons have missing data (blank entry in the "Value" column? Do you see any values with typos?
   e. Repeat this process for the "Veg_Group" and "Invasive" variables.
   f. Repeat steps c and d for the remaining two time periods ("70s_Species" and "00s_Species").

6. **Generalize Data**

   a. For our analysis, we are particularly interested in two attributes, "Veg_Group" and "Invasive." Right now, each polygon represents vegetation clusters of the same species, which is more detailed than we need. We want to create two new data sets in which the polygons represent clusters of vegetation groups and clusters of invasive types over

time. We will use these customized data sets in Lesson 4, where we will discuss how to interpret and present results from several datasets.

b. First let's create the time series shapefiles of vegetation groups.
   i. Geoprocessing > Dissolve
   ii. Click on the "Show Help" button for more information about what the tool does and what the different input criterion mean.
   iii. Input Features: 60s_Species
   iv. Output Feature Class: 60s_VegGrp (save it in your L3 folder)
   v. Dissolve Field: Veg_Group
   vi. Statistics Field: Select "sqm" as the field and "SUM" as the statistics type
   vii. "Create multipart features" should be checked
   viii. "Unsplit lines" should NOT be checked

c. Compare the output file to the input file from the same time period.

d. Repeat the dissolve for the remaining two time periods. Name them "70s_VegGrp" and "00s_VegGrp."

> The Summary Statistics tool (ArcToolbox > Analysis Tools > Statistics > Summary_Statistics) is another option you can use to calculate statistics for your data. This tool is similar to the "Summarize" option available by right clicking on a field in an attribute table. The advantage of the Summary Statistics tool is that it allows you to create statistics based on multiple fields. For example, you could use it to find the total area for every unique combination of vegetation type and invasive classification. You could interpret the results to find out which plant type makes up the majority of invasive species for each time period.
>
> Multipart polygons are features that have more than one polygon for each row in the attribute table. If you want to explode these into individual records at a later time, there is a tool available on the Advanced Editing Toolbar.

e. Now let's create the time series shapefiles by invasive type.
   i. Geoprocessing > Dissolve
   i. Input Features: 60s_Species
   ii. Output Feature Class: 60s_Invasive (save it in your L3 folder)
   iii. Dissolve Field: Invasive
   iv. Statistics Field: Select "sqm" as the field and "SUM" as the Statistics Type
   v. "Create multipart features" should be checked.
   vi. "Unsplit lines" should NOT be checked

f. Compare the output file to the input file from the same time period.

g. Repeat step c for the remaining two time periods. Name the output files "70s_Invasive" and "00s_Invasive."

h. Remove the "60s_Species," "70s_Species," and "00s_Species" from your map and save.

That's it for the required portion of the Lesson 3 Step-by-Step Activity. Please consult the Lesson Checklist for instructions on what to do next.

After experimenting with online data services in Lesson 2 and raw data in Lesson 3, which do you think is easier to work with? What are the pros and cons of each one? Can you think of any scenarios in which one is preferable over the over?

Do you have a good understanding of why we completed each step in Part II? If not, compare the starting vegetation files and final outputs (XX_Species, XX_VegGrp, XX_Invasive) in terms of extent, area, gaps, spatial detail, and attributes.

**Try This**

A) In Lesson 3 we familiarized ourselves with the study site using the "Open Street Map" layer from Esri. As we discussed in Lesson 1, Google Earth is another excellent application for this purpose. Try one or more of the activities listed below.

- Open the KML file of the Study Area (in the L3 Data folder) in Google Earth, zoom to the study boundary, and explore the area around the study site. For example, look for Street View images or other sources of imagery (Layers > More > DigitalGlobe Coverage). If you receive an error message (about outdated versions) by clicking directly on the KMZ file, open the file from within Google Earth > File > Open > Study_Boundary.kmz. You can also right click on the KMZ file from your desktop > Opens With > Google Earth.
- Experiment with the "view historical imagery" tool by clicking on the clock icon to see if you can find these. How far back in time do the images go?



B) Explore the Great Lakes Phragmites Collaborative. Can you find any other Phragmites projects along the Lake Erie shoreline? (If this link misses check under the "Management" tab for "Programs and Projects")

Note: Try This Activities are voluntary and are not graded, though I encourage you to complete the activity and share comments about your experience in Canvas Lesson 3 Discussion.