

Intro to ArcGIS 10x: Data Handling

<http://guides.library.upenn.edu/gisdatahandling>

Prepared by Christine Murray, chrmur@upenn.edu

Adapted from “MANTRA ArcGIS Practical Introduction,” by James Crone, EDINA and Data Library, University of Edinburgh as part of the Research Data MANTRA Project, funded by JISC MRD Programme (2010-11): www.ed.ac.uk/is/data-library-projects/mantra



This work is licensed under a [Creative Commons Attribution-ShareAlike 3.0 Unported License](https://creativecommons.org/licenses/by-sa/3.0/)

Introduction

Goals

Through an exercise scenario in which a researcher needs to identify a set of fieldwork sites, this tutorial will provide a hands-on demonstration of best practices for research data management within the Esri ArcGIS 10.1 Desktop GIS software.

Learning outcomes.

Upon completion of this tutorial the student will have:

- gained practical experience of best practices for setting up a geodatabase in ArcGIS for the storage and processing of research project data.
- practiced documenting data processing steps carried out within ArcGIS.
- understand the strengths and weaknesses of different GIS data formats at different stages of the data lifecycle.
- gained familiarity with and practice in creating geospatial metadata records within ArcGIS.

Requirements.

- The tutorial is self-contained and should take no longer than 90 minutes to complete.
- All data used during the exercise will be included.
- It is assumed that the student is familiar with the ArcGIS 10x interface and has used geospatial datasets in the past.
- 100MB of disk space is required to store data during the exercises.

Datasets used during the exercise

The practical uses example geospatial datasets to demonstrate the principles of research data management. These datasets have been derived from a number of open data sources:

- PA DCNR and Rails-to-Trails Conservancy Explore PA Trails vector GIS datasets provided as a Shapefile.
- USGS Geographic Names Information System place name gazetteer provided in TXT format.

- 1 arc-second spatial resolution Digital Elevation Model from the National Elevation Dataset made available through the National Map service

Exercise scenario

A researcher needs to carry out fieldwork in Pennsylvania and wishes to use multi-criteria evaluation analysis within ArcGIS to determine the location of a set of fieldwork sites. Criteria for the selection of fieldwork sites is that they form a significant part of the natural landscape – that they appear as named summits in Pennsylvania and that they are both visible from and within close proximity to a section of the Mid-State Trail long distance trail. In identifying the sites from the provided data, best practices for research data management within ArcGIS will be explored.

Tutorial

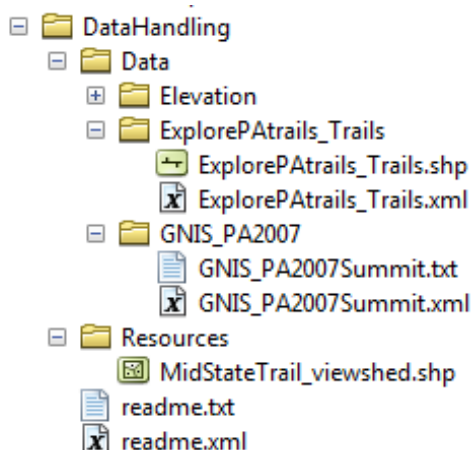
Explore project datasets.

Download the zip archive called **datahandling.zip** and unzip the contents of the **datahandling.zip** to a folder on your computer for which both you and ArcGIS are able to write to. After extracting the contents of the zip file there should be a new folder called **DataHandling** present.

Once you have extracted the contents of the **DataHandling.zip** archive to a folder start ArcMap. At the start-up dialog, select Blank map.

Once in ArcMap, access the Catalog window. If the Catalog tab is not visible on the right-hand side, you can access the Catalog via Windows > Catalog in the menu bar. Use the “Connect to Folder” button at the top of the Catalog window to navigate to the location of the “DataHandling” folder that you extracted in the previous step.

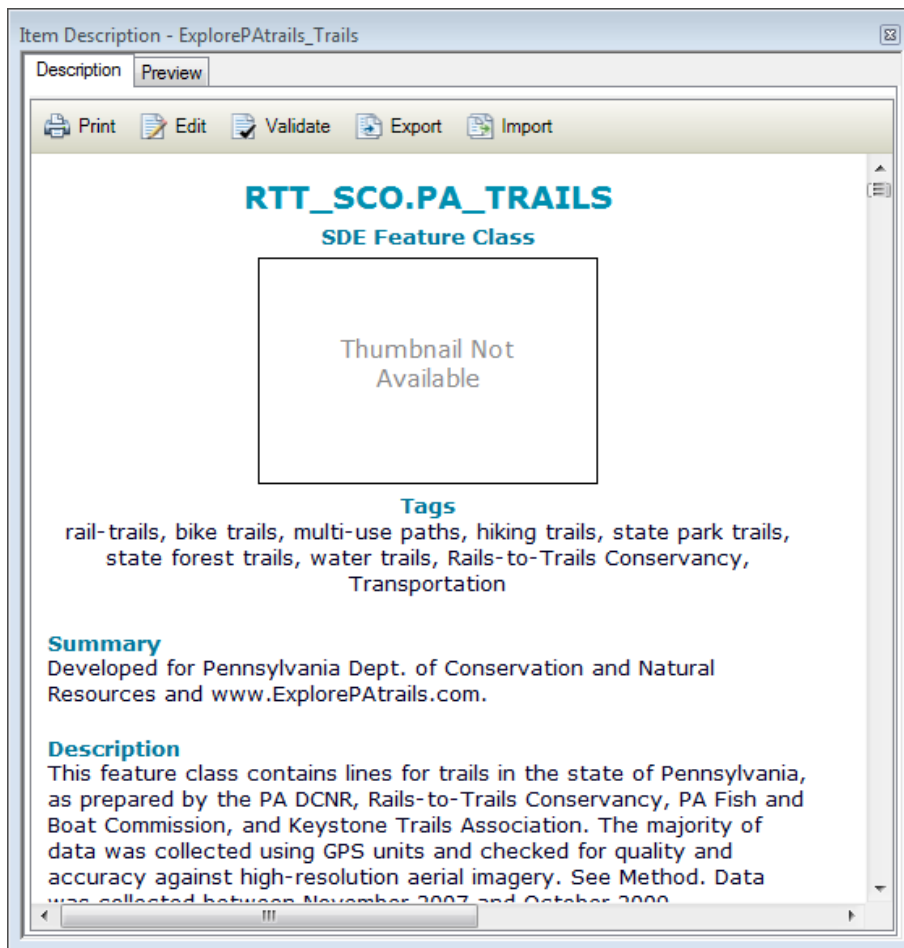
If you’ve never used Catalog before it can be thought of as a Windows Explorer for geospatial datasets. Expanding the DataHandling folder fully within the Catalog you should have something like the following:



Below the Data folder is an ExplorePAtrails_Trails and a GNIS_PA2007 folder. These contain different types of geospatial data that will be used during the practical as we go along. There is also a

folder called Resources. This contains supporting data which can be used during the course of the tutorial.

Before we go any further let's spend some time exploring the contents of the datasets that we have been provided with. Right-click on either GNIS_PA2007Summit.txt or ExplorePATrails_Trails.shp within the Catalog window and select Item Description. This allows you to view any geospatial metadata record that has been associated with the geospatial dataset. You should see something like the following:



Learning Point: Metadata.

Metadata is said to be data about data. By extension geospatial metadata is data about geospatial data. It has 2 main purposes – it provides a record of how a particular geospatial dataset has been created (and maybe from what other geospatial dataset it has been derived) and it provides a bibliographic record to facilitate the discovery and sharing of that geospatial dataset by others.

Create new geodatabase.

Let's create a new file geodatabase and then load our geospatial datasets into this.

Learning Point: The Geodatabase.

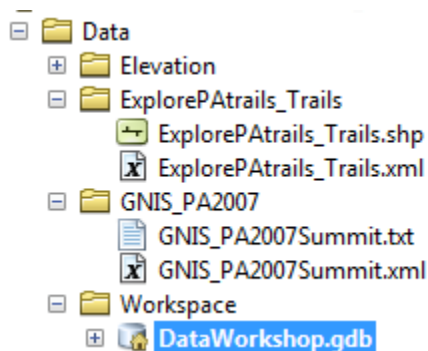
ArcGIS can use different type of geodatabases. A personal geodatabase stores all data in a Microsoft Access database whereas a file geodatabase stores all data in a local file structure. There is one

important difference between the personal geodatabase and the file geodatabase. A personal geodatabase may contain up to 2GB of data whereas a file geodatabase can contain up to 1TB of data. When using geographical information data volumes often become large quickly and it is recommended that you always create a file geodatabase.

Within the Catalog window click the top-level **Data** folder so that it is highlighted, right-click and from the pop-up menu select **New > Folder**.

Give the folder the name **Workspace**. Right-click on the **Workspace** and select **New > File Geodatabase** and give the file geodatabase the name **DataHandling.gdb**.

When done if you press the **F5** key on your keyboard to refresh the ArcCatalog Catalog tree, you should have something like this:



Load data into Geodatabase.

Now that the geodatabase is in place let's populate it with our datasets. First we'll create a new point feature class based on the records held in the GNIS_PA2007Summit.txt file. After which we'll add in the other 2 datasets. We will not load the elevation.tif dataset into the geodatabase.

Let's create a summit_names point feature class within DataHandling.gdb geodatabase from GNIS_PA2007Summit.txt delimited text file. To do this, in the Catalog window right-click on the GNIS_PA2007Summit.txt item, and then from the pop-up menu select Create Feature Class > From XY Table...

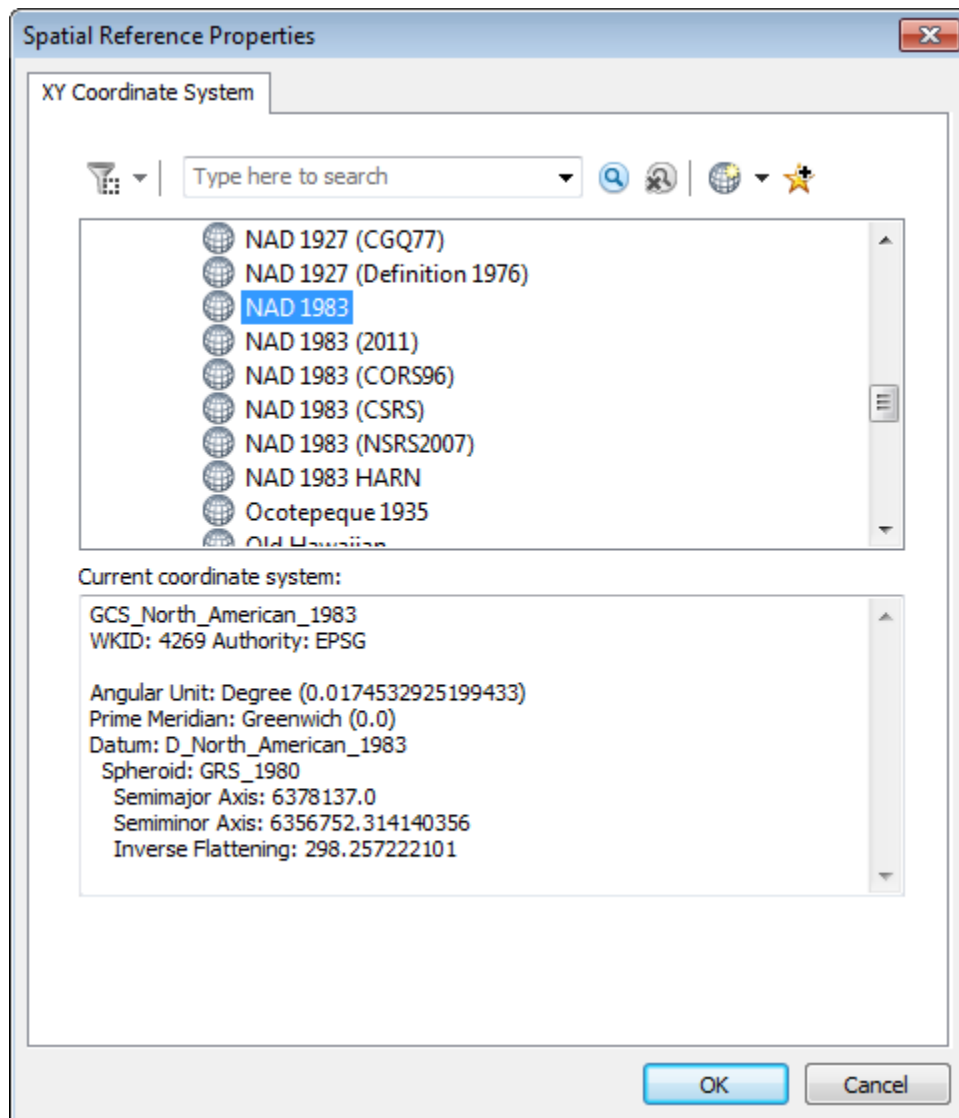
Doing so will bring up the Create Feature Class From XY Table dialogue.

The GNIS_PA2007Summit.txt contains a set of records. Each record contains a feature name, various information about the feature, plus the location of that feature in latitude (PRIMARY__2) and longitude (PRIMARY__1).

Set the X Field: to PRIMARY__2

Set the Y Field: to PRIMARY__1

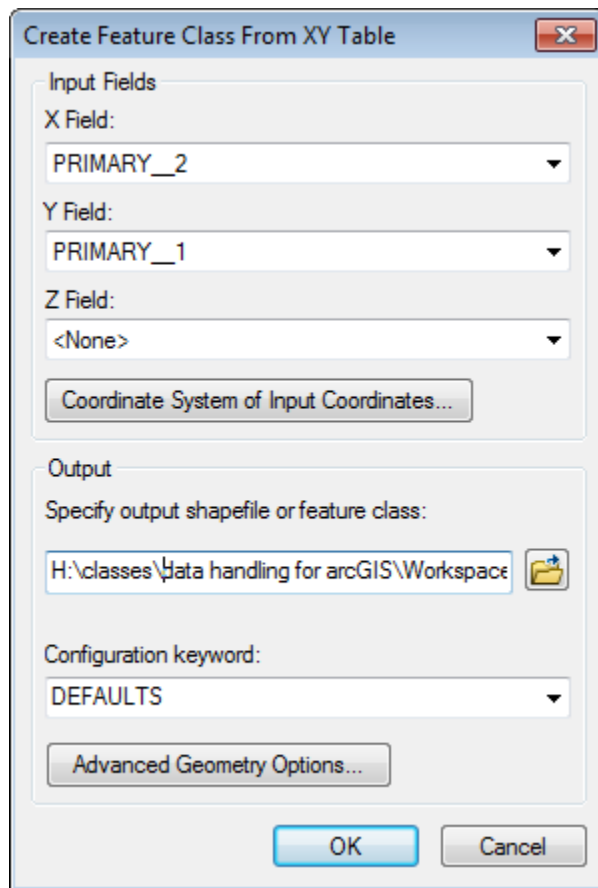
When we create a new feature class within the geodatabase we need to specify the coordinate system. Click the Coordinate System of Input Coordinates... button. The Spatial Reference Properties dialogue will be shown. In the dialog, use the folders to navigate to Geographic Coordinate Systems > North America, and locate NAD 1983. If you are using ArcGIS 10.1 or above, you can also search for NAD 1983 by name. Select this file like so:



Then click OK.

Lastly under Output, click the yellow folder button. This will open up the Saving Data dialogue. Ensure that Save as type is set to File and Personal geodatabase feature classes and then navigate to the location of the geodatabase that we just created in our workspace. Double-click the DataHandling.gdb geodatabase icon and then enter summit_names for the name of the file to be created. Then click Save

ArcGIS will return you to the Create Feature Class From XY Table, on which you should click the OK button.



Again the Item Description Preview tab at right can be used to explore the contents of this geodatabase feature class.

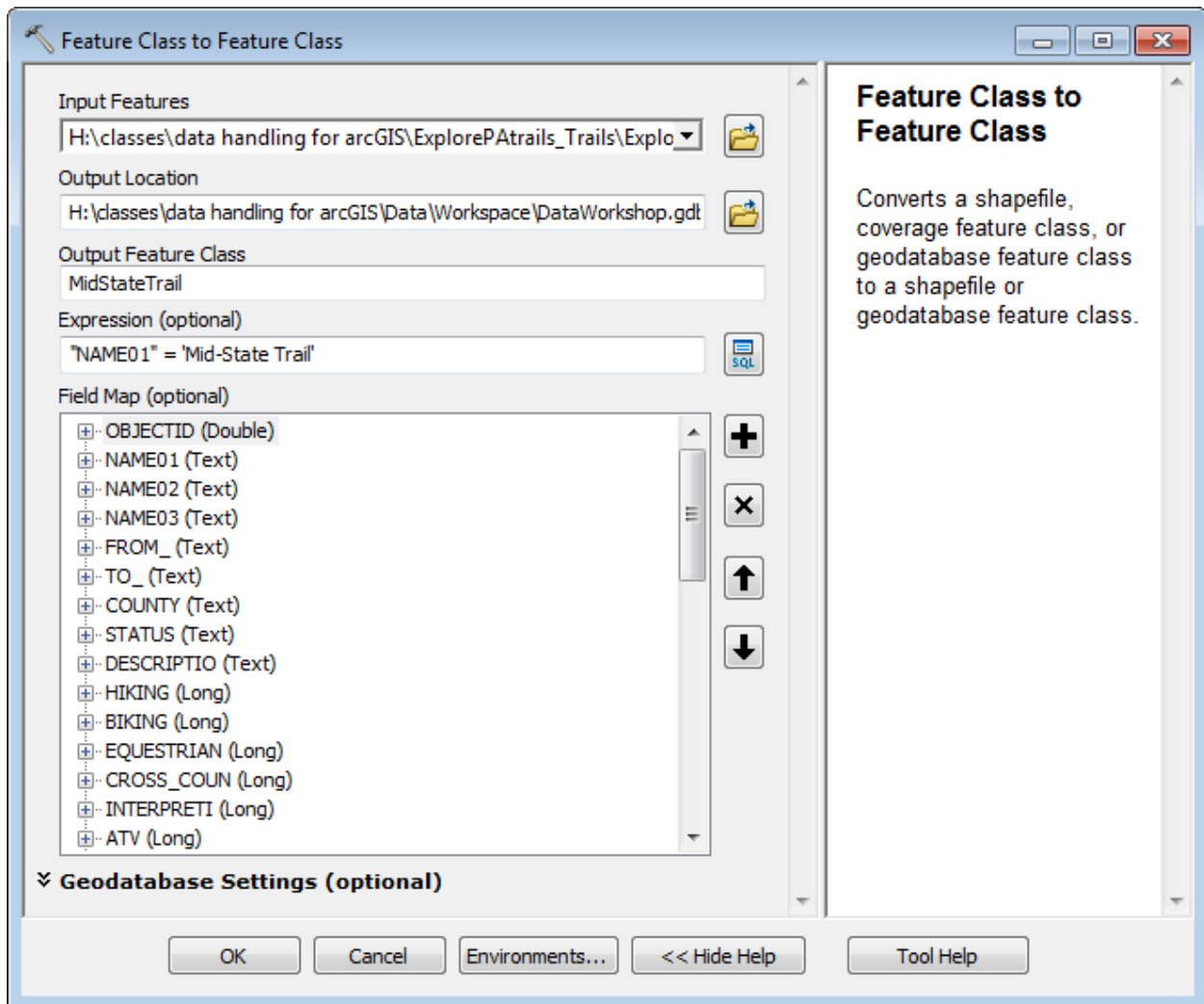
Learning Point: Fieldnames.

Click on the summit_names item in the geodatabase so that it is selected, ensure that the Preview tab is visible and then set the Preview dropdown to Table rather than Geography. Doing so will allow us to view the contents of the dataset as a table. Unfortunately the column headings used in the table are from the supplied text file and are not very meaningful. ArcGIS will allow us to provide aliases for the attribute names which will be used when the data is used within ArcMap. To set an alias on an attribute right-click on the summit_names item in the geodatabase and select Properties from the pop-up menu. This will bring up the Feature Class Properties dialogue. Click the Fields tab. Click on the button to the left of FEATURE_NA to select this entire row, then under Field Properties at the bottom section give the field the alias Feature Name. If you want you can do something similar for PRIMARY__1 and PRIMARY__2. Appropriate aliases might be Latitude and Longitude. After you have finished assigning the alias click OK to close the Feature Class Properties dialogue.

Let's now create a MidStateTrail polyline feature class within the DataHandling.gdb geodatabase from the ExplorePATrails_Trails.shp Shapefile. Right-click on the ExplorePATrails_Trails.shp item in the ArcCatalog Catalog Tree, and select Export > To Geodatabase (single)...

Doing so will bring up the Feature Class to Feature Class dialogue. Clicking the Show Help >> button will display ArcGIS contextual Help for this tool. The dialogue which should be completed

like so:



We specify our geodatabase as the Output Location and provide a new name for the Output Feature Class of MidStateTrail.

The ExplorePAtrails_Trails.shp Shapefile contains features for across Pennsylvania. However we only need features that correspond to the Mid-State Trail long distance trail. During the import process we can restrict the features that are brought in to the geodatabase from the shapefile using the optional Expression parameter. Clicking the SQL button next to the Expression text box brings up the Query Builder dialogue which allows us to build up a query string through which we limit the features imported based on their attributes.

The scrollable window at the top of the dialogue lists all of the fields available in the dataset that we want to make selections from. Clicking on the name of a field so that it is selected in this dropdown and then clicking the Get Unique Values button will list all of the unique values that appear within a specific field.

In the Query Builder dialogue we specify that we want to import only those features that have a NAME01 attribute with value of Mid-State Trail. So we enter the following query:

```
"NAME01" = 'Mid-State Trail'
```

Note: choice of quote mark is significant. Double-quotes should be used for field names, single-quotes for textual values to search for within that field.

Clicking OK having entered this in the Query Builder dialogue followed by OK on the Feature Class to Feature Class dialogue will load the data. When done if you press the F5 key on your keyboard to refresh the Catalog window, the geodatabase should now contain both a summit_names item and a MidStateTrail item.

Part of our criteria for the selection of the fieldwork sites is that the site is visible from the trail. This visibility factor will be captured as a viewshed. A viewshed is a region of visibility observable from one or more observation points. In our case the viewshed will be the region of visibility of the landscape which is observable from the route of the Mid-State Trail. Viewsheds are calculated by comparing locations with an underlying terrain surface held in a Digital Elevation Model. This calculation can be computationally intensive. To save time a viewshed corresponding to the Mid-State Trail has been created for you. This dataset is provided as MidStateTrail_viewshed.shp in the Resources folder of your DataHandling folder.

Let's create a viewshed polygon feature class within DataHandling.gdb geodatabase from MidStateTrail_viewshed.shp. In the Catalog window expand the Resources folder and click on the MidStateTrail_viewshed.shp item so that it is selected, right-click the file and then select Export > To Geodatabase (single)...

Again this will bring up the Feature Class to Feature Class dialogue. All we need to provide is the name of the Output Feature Class and the Output Location. Set Output Location to our geodatabase and Output Feature Class to viewshed. Clicking OK will tell ArcGIS to load the data.

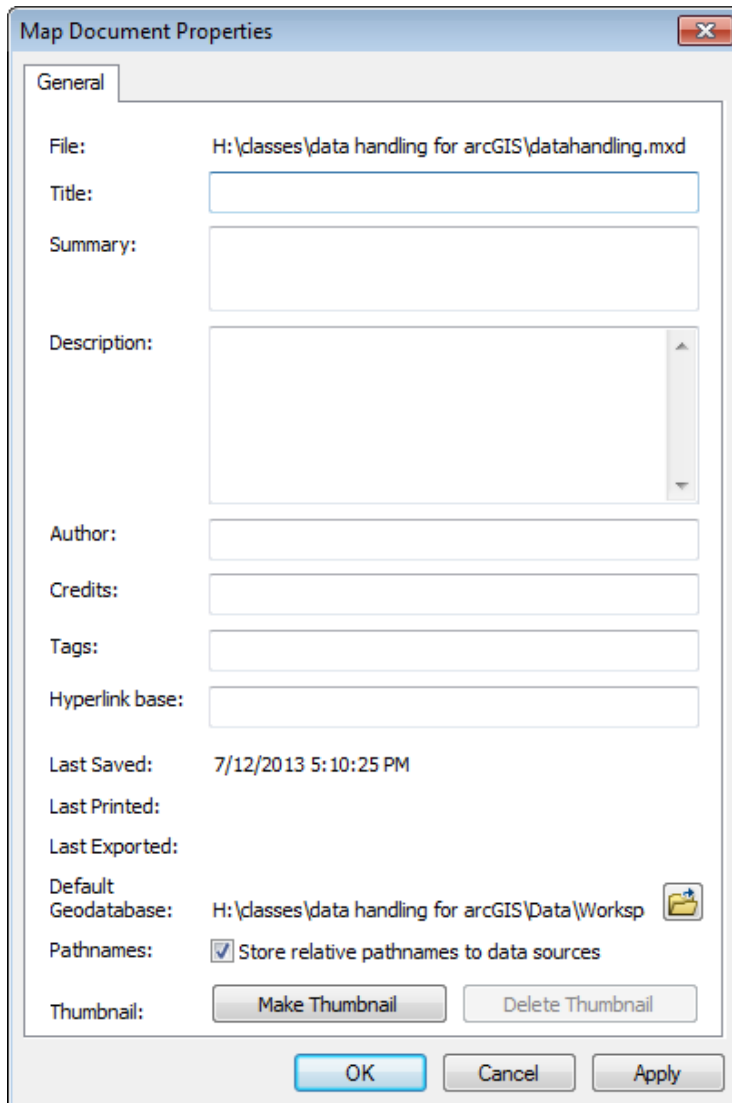
We've learned how to create a new geodatabase in which to carry out our research data work within. We've understood the differences between different types of geodatabase and seen how we add data to the geodatabase. In the initial exploration of our data we also saw the benefit of geospatial metadata as a mechanism of understanding the contents and source of data. Before we go any further we'll create a new ArcMap map document for the purposes of this practical and add to it the datasets from our geodatabase.

Add data from Geodatabase to ArcMap project.

Learning Point: Absolute and Relative paths in ArcGIS.

Before we go any further let's set an important property of the ArcMap project held in the .mxd file. Files stored within the local directory structure on your computer are expressed as pathnames. Pathnames can be expressed in absolute or relative terms. An absolute path is a path that points to the same location on one file system regardless of the working directory and is usually written in reference to a root directory such as C:\. A relative path by contrast is a path relative to the working directory of ArcGIS (e.g. same folder, one folder up, or in a folder two down called /mapdata/features).

ArcGIS will allow us to use either relative or absolute paths within our ArcMap project. Relative paths means that the ArcMap project should be able to be transferred between different file systems (different computers) without problem. To check whether the ArcMap project is using absolute or relative paths go to File > Map Document Properties and click the Data Source Options... button. On the Data Source Options dialog which is displayed ensure in our working project that the option is set to Store relative path names to data sources.



You can also enter descriptive information about the map, as well as set a default geodatabase. Use the Browse Folder icon to set the default geodatabase as DataHandling.gdb. Click OK to return to ArcMap.

Now let's add the data from our geodatabase to ArcMap. Either use File > Add Data... or use the Add Data button on the main ArcMap toolbar.

Navigate to the location of the DataHandling.gdb geodatabase and add all three items that we just populated the geodatabase with. That is:

summit_names
viewshed
MidStateTrail

In addition, to give context to the viewshed we can add the DEM which was used to create the viewshed. This is provided for you in the Data\Elevation folder as elevation.tif. This is a GeoTiff format dataset. Before adding the elevation.tif layer to ArcMap, click on the item so that it is selected in Catalog, right-click and select Build Pyramids. The Build Pyramids dialogue will be shown. Just click the OK button. Then ensure the elevation.tif item is selected, right-click and select Calculate Statistics. The Calculate Statistics dialogue will be shown. Accept the defaults and click the OK button. You can now add the elevation.tif layer to ArcMap.

Having added all of the data the best way to zoom into the extents of the data is to click on the MidStateTrail layer, right-click and choose Zoom To Layer. You might at this point want to change how ArcMap displays any of the four layers from the default symbologies. It would be useful to change the symbology (from View >> Layer Properties) of the viewshed layer so that features are shown with a transparent fill. That way we can view the underlying terrain described by the elevation.tif through the viewshed layer.

Either way we should save our ArcMap project. From the main ArcMap menu bar select File > Save As... to save the current ArcMap project. Give the project the filename DataHandling.mxd, ensure that Save as type is set to ArcMap Document (*.mxd) and that you save the file to the DataHandling/Data/Workspace folder.

Use Spatial Analysis to find sites visible from Mid-State Trail.

Let's remind ourselves of our objective. We are interested in identifying field study sites in Pennsylvania that are visible from and falling within distance of a section of the Mid-State Trail. This analysis will be undertaken under a methodology known broadly as Multi-Criteria Evaluation (MCE). MCE is an analytical means of making selections based on specific criteria. MCE does not need to be applied within GIS. Here though, our criteria are that the fieldwork sites:

- appear as a named summit in GNIS, the rationale being that such features are culturally significant and represent important landscape features.
- are visible from the path of the Mid-State Trail long distance footpath.
- are within a certain distance (1km) of the Mid-State Trail long distance footpath. We do not want to have to walk far from the footpath if we are carrying specialized and perhaps cumbersome fieldwork equipment as part of the research, especially as we are carrying this equipment on the leeward slopes of hills and mountains.

We will first carry out this analysis interactively using tools within ArcMap.

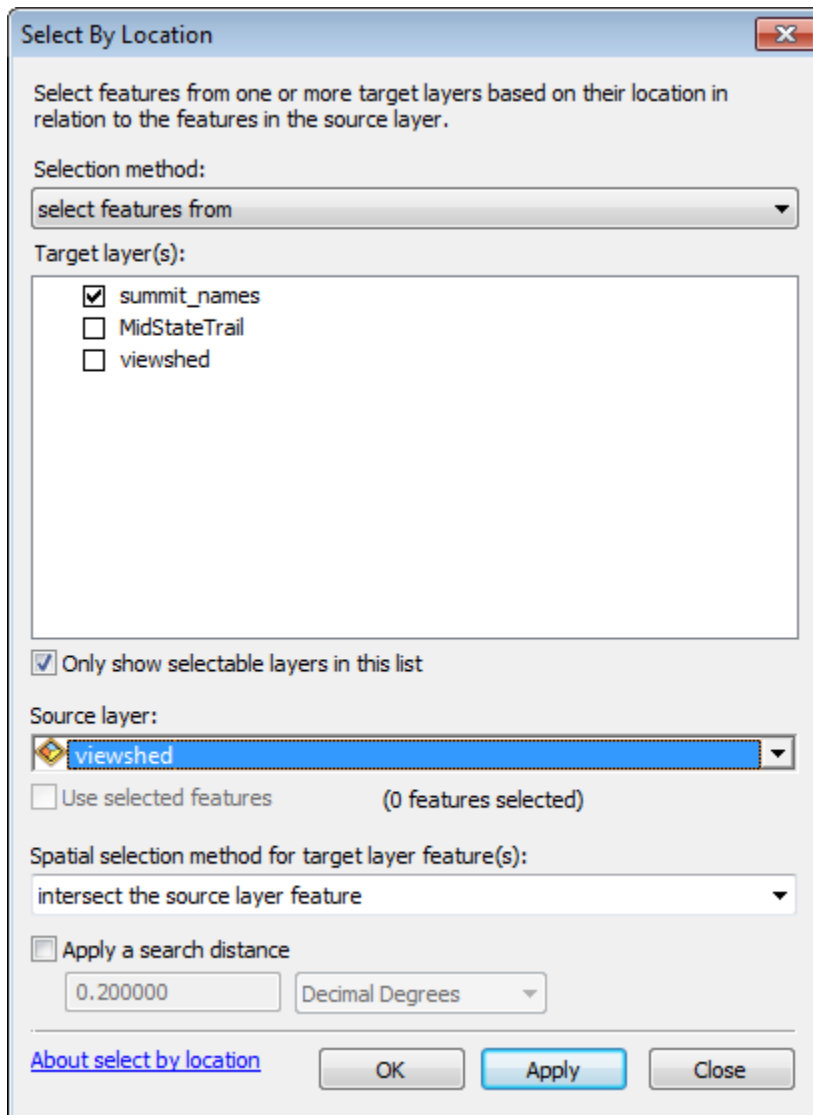
First we need to use ArcGIS to select all features from the summit_names feature class which fall within the viewshed. Remember that the summit_names feature class represents candidate locations for the fieldwork sites as these are locations which appear as named locations in GNIS. After applying this first selection ArcMap will reduce the candidate sites to only those that are visible from the Mid-State Trail. We then need to further refine the candidate sites – to include only those sites that are within one kilometer of the trail.

Click on the summit_names item so that it is selected in the ArcMap layer list. Then from the ArcMap menu bar choose Selection > Select By Location... Doing so will bring up the Select By Location dialogue. If you haven't used this ArcMap tool before, it allows us to define which layers to make selections from and how these selections should be made. For our purposes set the dialogue so that it reads:

Selection method: select features from Target layer(s): summit_names

Source layer: viewshed

Spatial selection method for target layer feature(s): intersect the source layer feature



After you have done this, click the OK button.

ArcMap will make the selections and highlight those features within the summit_names layer that fall within the viewshed and thus are visible from the Mid-State Trail footpath.

Use Spatial Analysis to sites which are also within 1km of Mid-State Trail.

We now want to limit this set further to only those selected features which are also within 1km of the trail. In order to do this we'll create a buffer geometry around the Mid-State Trail polyline and then use the features in this buffer geometry to further refine the selected features. To create the buffer geometry we'll use ArcToolbox. If it's not already shown click on the ArcToolbox button on the ArcMap menu bar.

Once open the Buffer tool is to be found within the Proximity toolset of the Analysis Tools toolbox. Double-click the Buffer tool to open up the Buffer dialogue.

We need to set the features that the buffer geometry will be based on. This is the MidStateTrail feature class in our geodatabase. We also need to define the name of the buffer geometry that will be created where this will be created. We want to create a dataset called MidStateTrail_1km_buffer within our geodatabase. The only other mandatory parameter that we need to supply to ArcGIS is how big the buffer should be and the unit of measure that this is provided in. We want the buffer to extend 1km either side of the Mid-State Trail. We can thus enter the value of 1000 ensuring that meters is selected. (You can also enter a distance in different units and specify the unit type in the buffer dialog.)

Buffer

Input Features: MidStateTrail

Output Feature Class: H:\classes\data handling for arcGIS\Data\Elevation\MidStateTrail_1km_b

Distance [value or field]: 1 Kilometers

Side Type (optional): FULL

End Type (optional): ROUND

Dissolve Type (optional): ALL

Dissolve Field(s) (optional):

- OBJECTID_1
- OBJECTID
- NAME01
- NAME02
- NAME03
- FROM_
- TO_
- COUNTY
- STATUS

Dissolve Type (optional)

Specifies the dissolve to be performed to remove buffer overlap.

- NONE—An individual buffer for each feature is maintained, regardless of overlap. This is the default.
- ALL—All buffers are dissolved together into a single feature, removing any overlap.
- LIST—Any buffers sharing attribute values in the listed fields (carried over from the input features) are dissolved.

OK Cancel Environments... << Hide Help Tool Help

Once you've filled in the form click the OK button. When finished, ArcGIS will add a new polygon layer called MidStateTrail_1km_buffer to the map. Again it's a good idea to change the symbology

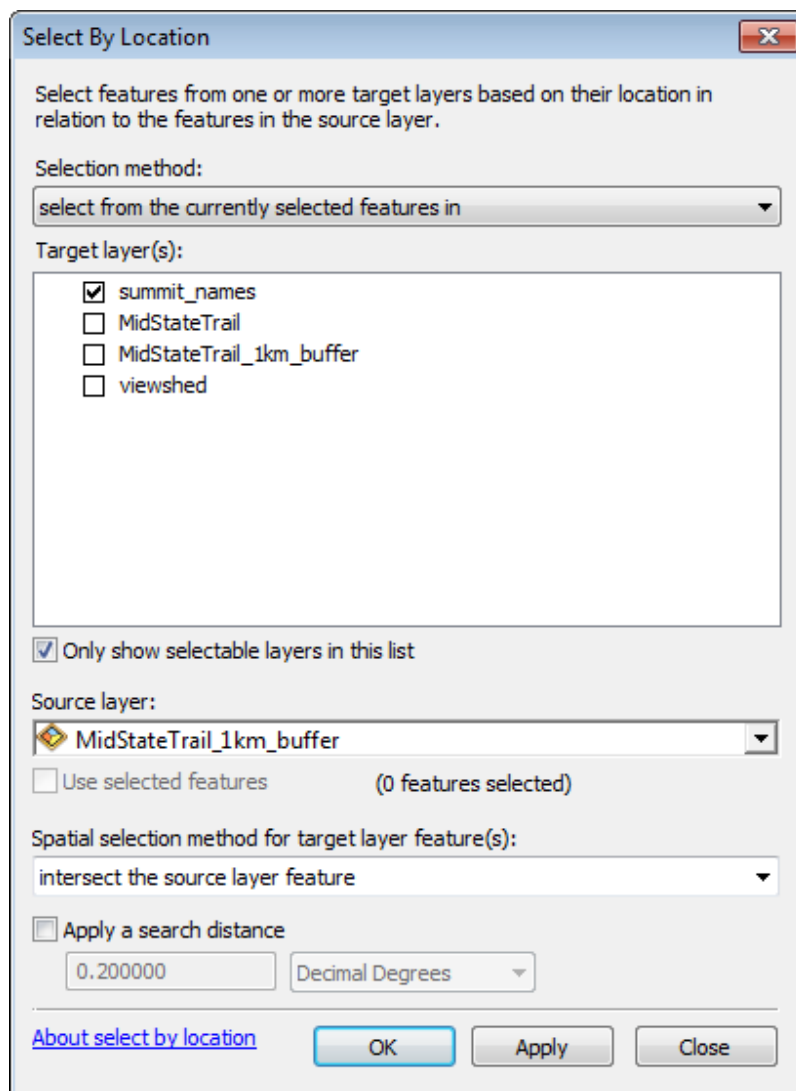
of this polygon layer so that it is displayed with a transparent fill. We can close the ArcToolbox panel now by clicking on the X at the top-right corner.

Now that we have a buffer geometry of 1km around the route of the path we can use this to refine further our fieldwork sites. Within the summit_names layer ArcMap has already selected those features that fall within the viewshed. We can compare those selected features with the buffer geometry to find those that are within 1km of the path. Click the summit_names layer so that it is selected in the layers list and then select Selection > Select By Location... again. This time set the dialogue so that it reads:

Selection method: select from the currently selected features in Target layer(s): summit_names

Source layer: MidStateTrail_1km_buffer

Spatial selection method for target layer feature(s): intersect the source layer feature



Once we click OK ArcGIS will carry out a further selection on the data and change those features that are highlighted:

Since this selection represents the answer to our research question we better save it as permanent dataset. If it's not already selected, click the summit_names item, right-click and select Selection > Create Layer From Selected Features.

This will add a temporary item called summit_names selection to the layers list. To create a dataset from this, click the summit_names selection item so that it is selected, right-click and select Data > Export Data... This will bring up the Export Data dialogue.

Under Output shapefile or feature class: ensure that Export_Output is changed to fieldwork_sites before clicking OK.

ArcMap will ask if you want to add the exported data to the map as a layer. We don't, so click No. You should already have seen the benefit that geospatial metadata has in helping you understand the contents of and how datasets were created. In the last part of the workshop we will look at creating a metadata record for the fieldwork sites so that we and others can always be clear how the dataset was created.

Export dataset from Geodatabase as a Shapefile.

Use of a geodatabase is great for managing our and others geospatial data as part of our research project as it allows us to keep data in one place within the ArcGIS environment. However it is not always easy to interact with the contents of data stored in a geodatabase outside ArcGIS. This is problematic if we want to share the data from our project, so that others can build on the results of our research or just explore the output datasets in systems other than ArcGIS. Where a user has ArcGIS they will be able to open our ArcMap project and view the contents from the geodatabase. A de-facto standard for the exchange of vector GIS datasets is the shapefile. If we create versions of our vector datasets in shapefile format and versions of raster datasets in GeoTiff format we don't second-guess what applications other people may be using.

Within the DataHandling project folder, use the Catalog window to create a project outputs folder. Click on the top level DataHandling/Data folder so that it is selected, right-click and select New > Folder... Give the new folder the name Outputs

We want to create a version of the fieldwork_sites item within the geodatabase as a Shapefile. Select the fieldwork_sites item so that it is selected in the DataHandling.gdb geodatabase, right-click and select Export > To Shapefile (single)...

The Feature Class To Feature Class dialogue will be shown. Clicking the Show Help >> button will display ArcGIS contextual Help for this tool. Earlier in the practical we used this dialogue to load data into our geodatabase to create a workspace in which to process our data. Here we'll go the other way to export data from our geodatabase into a shapefile.

We need to fill in the details for Output Location and Output Feature Class. To provide Output Location use the button at the right of the Output Location textbox to navigate to and select the Outputs folder we created above. Now we need to provide the name of the Output Feature Class to be created. In the Output Feature Class textbox enter fieldwork_sites.shp. After doing this click OK.

Create Geospatial Metadata record.

If you now right-click on the fieldwork_sites.shp item in the Catalog window within the DataHandling\Data\Outputs folder and click on Item Description, it will have inherited metadata from the GNIS2007Summits file.

We can change this. Since we are the creator of the fieldwork_sites.shp dataset we are now responsible for creating metadata associated with that data. Click on the Edit button to enter your own metadata.

On the Item Description page, under Enter a title for the dataset enter the text: DataHandling fieldwork sites

In the Summary box, enter the text: Point dataset that shows the location of fieldwork sites identified during the Data Handling ArcGIS practical to find sites accessible from the Mid-State Trail.

Click on Citation on the left-hand side of the window. Under Date use the dropdowns to select today's date as the Created date.

Scroll down to Lineage, under Resource on the left-hand side of the window. In the Statement box, enter the text: Candidate fieldwork sites were initially identified as being locations from GNIS that corresponded to a hill or mountain name. Selections were made from these candidate sites on the basis that they were both visible from and within distance of the Mid-State Trail. A viewshed was used to represent visibility. A buffer geometry was created to test proximity of candidate sites to the trail

Under Points of Contact, add a new Contact for yourself, and delete the current contact.

We have now edited our metadata. Click Save to retain these changes and exit the window.

Create documented data archive.

Great, we have now exported the output from our research as a Shapefile and ensured that we have a valid geospatial metadata record for this dataset. The only thing that we have left to do is to package our data up.

First we'll create a document describing what all of the data within the DataHandling folder is. There is a placeholder Readme.txt file under the top level DataHandling folder ready for you to fill out.

Open the Readme.txt document in notepad and at a minimum add the following text:

This archive contains geospatial datasets that were created as part of the Mantra ArcGIS practical. The Data\ExplorePAtrails_Trails and Data\GNIS_PA2007 folders contain input datasets which the practical was based on.

The Data\Workspace\DataHandling.gdb is an ArcGIS file geodatabase

The Outputs folder contains a single shapefile fieldwork_sites.shp which has been created through analysis of the dataset. A geospatial metadata record provides information on how this dataset was created.

Then save the file and close Notepad. The idea is that by reading this file anyone will be able to understand what the various files are and how they were created.

Having done this we can then collect all of our project files together to create a (zip) archive to store all of the project data.