Lab 2: GIS Data Models

Site:	GauchoSpace
Course:	GEOG 176B - TECHNICAL GIS - Winter 2021
Book:	Lab 2: GIS Data Models

Printed by:	Ian Hinds
Date:	Saturday, January 30, 2021, 11:52 AM

Table of contents

1. Purpose

- To gain a clear understanding of what a data model is, and why data models are important.
- To learn the data models Esri supports in ArcGIS, and the similarities and differences between them.
- To learn the advantages and disadvantages of using certain data models for different tasks.
- To reinforce basic ArcGIS skills.

2. Introduction and Background

2.1 Basic Concepts

You will notice some diversity in the following definitions, as they are in the context of different companies, software, times, and degrees of specificity. For this lab, focus on the hierarchy described in the main body of the lab.

There are three basic spatial *data types* used with GIS (points, lines, and areas):

- *Points* represent anything that can be described as a discrete x, y location
- *Lines* represent anything having a length
- *Areas*, or polygons, describe anything having boundaries

These data types comprise the vector model, which is the model you will deal with most often in GIS.

Vector data model:
Discrete features, such as customer locations, are usually represented using the vector model. Features can be discrete locations or events, lines, or areas. Lines, such as streams or roads, are represented as a series of coordinate pairs. Areas are defined by borders, and are represented by closed polygons. When you analyze vector data, much of your analysis involves working with (summarizing) the attributes in the layer's data table.

Raster data model:
Continuous numeric values, such as elevation, or discrete categories, such as vegetation types, are represented using the raster model. The raster data model represents features as a matrix/lattice of cells in continuous space. A point is one cell, a line is a continuous row of cells, and an area is represented as continuous touching cells.

Tabular data:
Contain information describing a map feature in the form of a table or spreadsheet. For example, a GIS database of customer locations may be linked to address and personnel information. GIS links this tabular data to associated spatial data.

Question 1:
Give an example of one way you could represent a continuous phenomenon using the vector data model. Be specific about how the data might be captured.

2.2 Geographic Data Modeling: An Introduction

Data Model - An abstraction of the real world that incorporates only those properties thought to be relevant to the application at hand, defines specific groups of entities and their attributes and states relationships between these entities. A data model is independent of a computer system. [Data Models for GIS]

Data models are a crucial concept for GIS users to understand. Data models describe how geographic features will be represented in the GIS. Any time you wish to deal with geographic data in a computing environment, you must choose a geographic data model by which to do it. The choice of data model will yield benefits in terms of simplifying real-world features enough to deal with them easily, but will also incur costs in terms of oversimplifying or misrepresenting different aspects of them in the process.

A paper map is an example of an analog data model -- it is a formalized framework that cartographers use to capture and represent information about a landscape on a sheet of paper. The same sort of thing is also needed to capture and represent geographic information when the medium is digital rather than ink-and-paper. In a GIS, abstractions of real-world features must therefore be formalized into a data model that defines how the computer will represent and manage the geographic information (geometry and attributes).

Bernhardsen (1999) diagrams the data model formalization process along these lines:

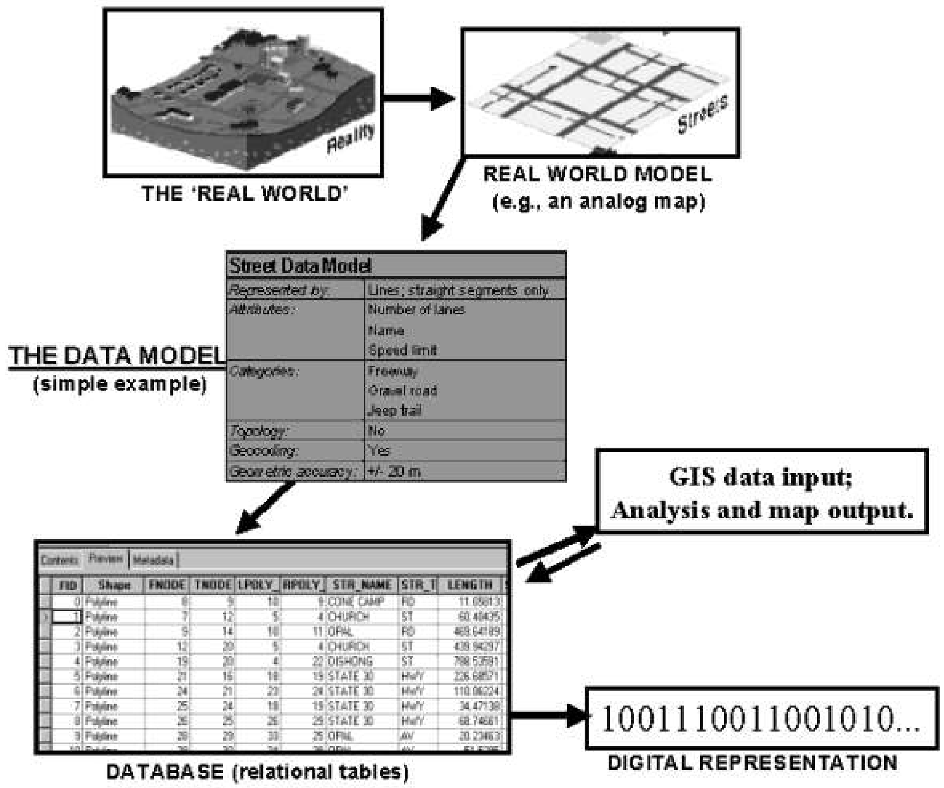


Figure 1: The modeling process. (after Bernhardsen 1999, p.39.)

Most of the confusion about data models arises from their diversity. Some data models are more abstract/theoretical while others are made with specific database frameworks in mind. For example, the vector data model and the raster data model are very general, whereas the georelational data model and geodatabase data model are made to fit specific categories of database software. Furthermore, a given data model may belong to more than one category: a coverage is both a vector data model (general) and a georelational data model (database-specific).

The many types of data models are easier to think about if one pictures of them as being part of a general hierarchy. Below is a figure showing the hierarchy of ArcGIS's data models:

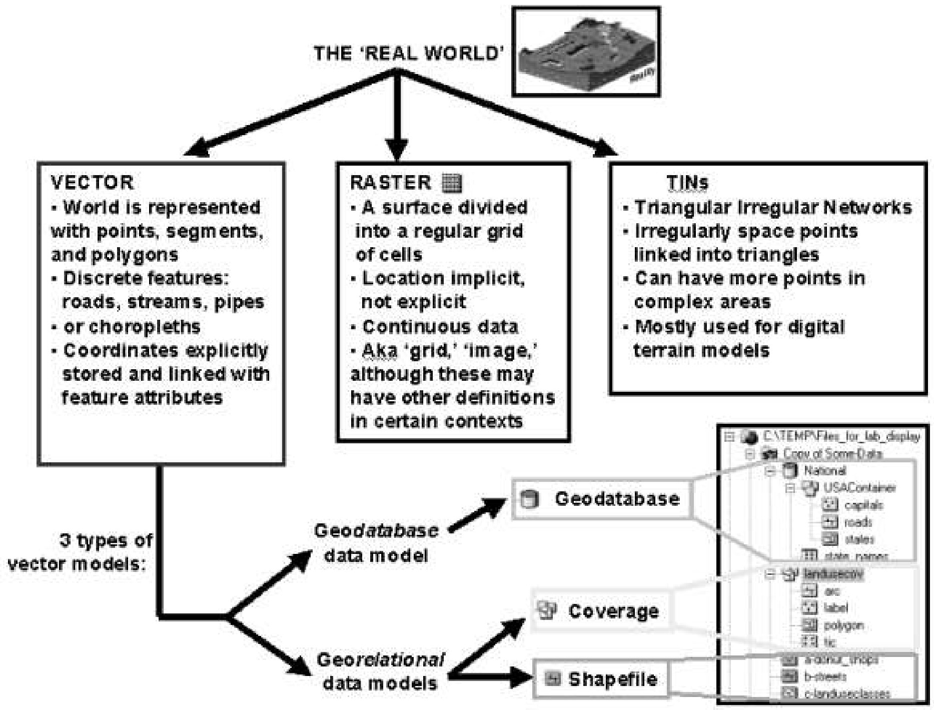


Figure 2: Hierarchy of ESRI's ArcGIS data models.

The data models go from most general at the top level (vector, raster, TIN) to most specific at the bottom level (shapefile, coverage, geodatabase). It is important to note that a geodatabase can handle all three general models, not just the vector model.

Geographic data models have evolved under the influences of technology (e.g., increasing storage space and processing power, networking, or software evolution) and even history (e.g., ESRI introduced the "coverage" data model in 1980).

Every GIS software package will be capable of supporting a number of data models. The capabilities of the data models may change with new versions of the software, and compatibility issues may arise between different GIS software, and even between different versions of the same software. Certain functions will be accessible using data in the form of one data model but not another.

2.3 Data Structures vs. Data Models

Once we have decided on a data model to use, there remains the question of how to actually store this model in the computer. The specific format to be used for storing it is known as the ***data structure***. To illustrate, consider a basic vector data model. The vector model represents features as consisting of lines which individually link together a start node, vertices in between, and an end node. To draw and analyze features represented this way, the computer needs information on the locations of each node and vertex of the lines. This could be provided in the form of a table listing the coordinates of these points, and indicating which line(s) go through them. This table would be the basic ***data structure***. Shapefiles use this type of structure.

In Figure 1 above, the lower left box titled "DATABASE (relational tables)" represents the *data structure*. In it you can see numbered rows and columns with labels, this is the 'structure' of the data. Some columns have only numbers, some have only text and some have both.

Several different types of data structures can potentially be used to represent the same data model. For example, you could represent a vector data model using shapefiles or geodatabases. Although these all take the same basic approach in representing the model, there are still significant differences between them. We will discuss what these differences are later, but for now keep in mind that 1) data models do not necessarily imply any particular data structures; and 2) data structures can represent the same data model while still being very different from one another.

2.4 GIS Information and Resources

National Center for Geographic Information and Analysis (NCGIA) has its Core GIScience curriculum online. Some resources relevant to Data Structures and Data Models: Fundamentals of Data Storage, Information Organization and Data Structure, and Non-spatial Database Models.

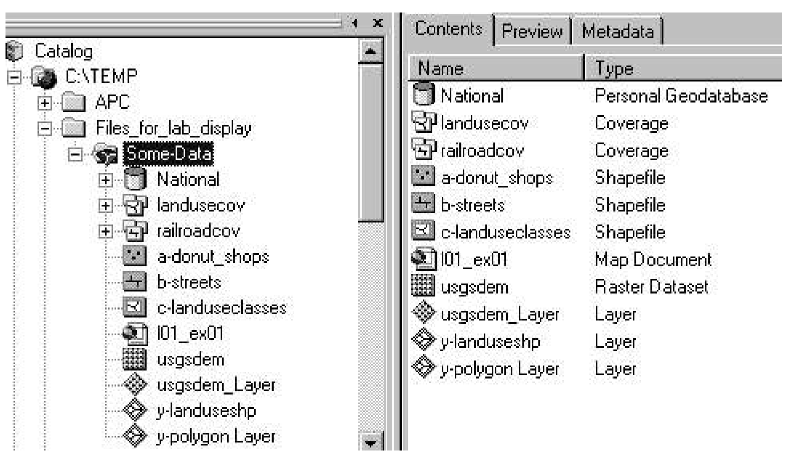
Question 2:

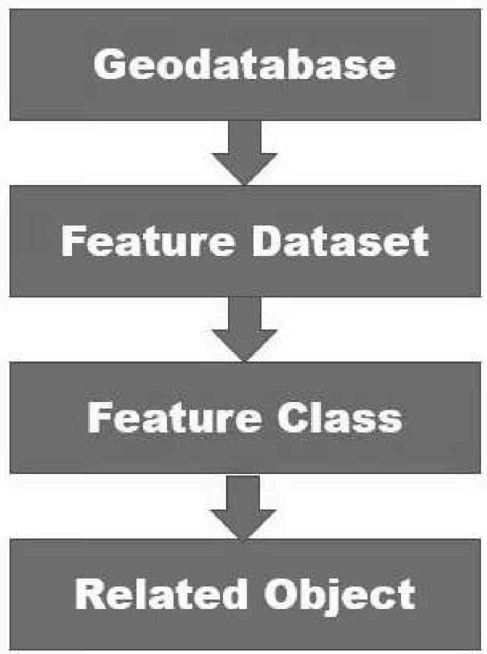
What (general) data models can be stored in a geodatabase?

Question 3:

What is the difference between a data structure and a data model? Answer in your own words.

2.5 Databases, Datasets, and Feature Classes: Data Models in ArcGIS





Source *SSP Innovations*

At the top, the Geodatabase contains the necessary background data structure which enables the functionality of ArcGIS. Within the Geodatabase, there may exist one or multiple Feature Datasets. A Feature Dataset is an organizational container of similar Feature Classes (i.e. Gas, Electric, Telecom) of the same Projection/Coordinate System. Although not mandatory, we recommend all Feature Classes be located within a Feature Dataset because Feature Datasets can maintain topological relations. Lastly, Related Objects leverage Esri Relationships to link to feature classes. Because they have no spatial component (i.e. they are just a table) they can't be shown on a map or contained within a Feature Dataset. [SSP Innovations]

Feature classes are the lowest level that the user accesses.

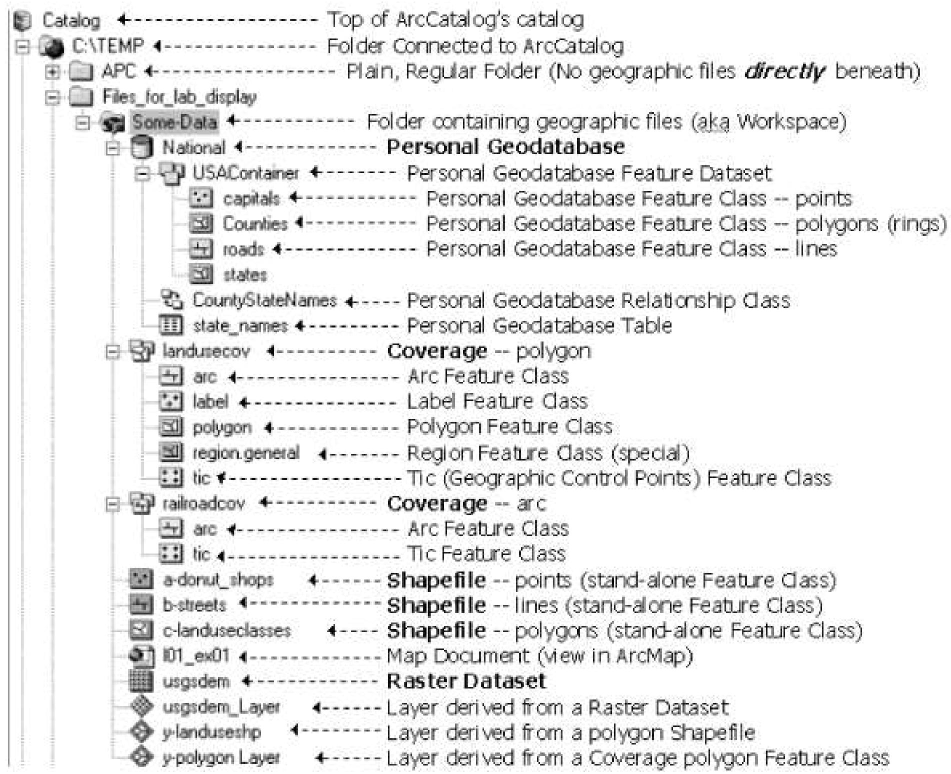


Figure 3: Icons and hierarchy

2.6 Some Bottom Level Data Models in ArcGIS

Shapefiles: A single geographic feature type (counties, roads, capitals, etc.) will be contained in a shapefile, and each shapefile corresponds to a feature class. The geometric information (stored in hidden binary files) will be displayed in Catalog View's "Details" and the attribute information (stored in dBASE tables) will be displayed in the "Table Preview". This linkage of geometric files to separate attribute tables is common to shapefiles and is called a ***georelational data model*** by Esri. Shapefiles have a size limitation of 2 GB.

Coverages: A coverage is a georelational data model that stores vector data—it contains both the spatial (location) and attribute (descriptive) data for geographic features. Coverages use a set of feature classes to represent geographic features. Each feature class stores a set of points(tics), lines (arcs), polygons, or annotation (text). Coverages can have topology, which determines the relationships between features.

A coverage is stored as a directory within which each feature class is stored as a set of files. For example, a coverage appears in ArcCatalog with the icons as shown below. In this example, you can see that the streams coverage is a line coverage containing an arc (line) file, annotation for the line, and a tic file. There are also two versions of coverage files.

Geodatabases: A single geographic feature type corresponds to a feature class, as with shapefiles. Multiple feature classes can be grouped into a *feature dataset*(symbolized as three overlapping grey tiles) which specifies a common geographic framework for all its constituent feature classes contained within it. All feature classes grouped inside of a feature dataset must have the same spatial reference i.e. projection and coordinate system information (In Figure 3, for example, the "USA container" contains information about the USA, capitals, counties etc.). Unlike shapefiles, geodatabases employ a *geodatabase data model* that stores each feature, complete with its geometry, as a row in a relational database table. A number of feature datasets can be stored in a geodatabase.

It is important to note that there are two types of personal geodatabases supported in ArcGIS Pro. The File Geodatabase stores each feature dataset as a file in a shared geodatabase folder. There are some limitations on file sizes (1TB, configurable to 4TB or 256 TB) and editors (single editor, multiple readers) with file geodatabases. Enterprise geodatabases, on the other hand, offer almost unlimited sizes and number of users. Feature datasets are held as individual tables in a relational database.

- <u>Note:</u> Do NOT use spaces in file/folder names! Use an underscore ("_") instead. Certain ArcGIS software components seem to need an "unbroken path" to function correctly -- if you use spaces, you may run into problems. This is a holdover from older Arc software even though most Windows-based software can now handle spaces in names. An example of a location that contains spaces in the path name is C:\Documents and Settings\student\Desktop\176B_Lab2. Notice that there is a space after the word "Documents" and another space after the word "and".

Question 4:
The fact that you can't use spaces in filenames or folders has to do with what? (the software, data model, data structure, or something else)

Question 5:

What is a "feature class"? Answer in your own words.

Question 6:

What is the main difference between the geodatabase data model and other bottom level data models?

3. Get the Data

Under current circumstances, please save the data and the map project on Google Drive (if you use AWS) or your own computer (if you have ArcGIS Pro installed locally).

*You need to copy the data for Lab 2 to your flash drive. To copy t*he data, visit the relevant week in Gauchospace and select Save Link As or Save Target As depending on which browser you are using. Save the zipped data file to an appropriate location such as your Zip disk or a folder you created in E:\John176BL\JohnLab2. Unzip the data.

You can view the data in Catalog (View or Pane) to verify that the following files and folders have been extracted:

/mystery --Contains 8 data layers of several features in different data models. You will be figuring out what these are in the lab.

/sb
- *roads* -- Santa Barbara county roads gdb feature dataset, clipped to the Goleta-Santa Barbara region
- *sbdem*-- digital elevation model of Santa Barbara county
- *sbtin*-- TIN derived from *sbdem*
- *sbcontour* -- Contour geodatabase (gdb) feature dataset derived from *sbdem*
- *cacounty*– gdb feature dataset of counties of California, from the GDT dataset

(The street data we are using in this class was provided by Tele Atlas.)

4. Procedure

4.1 Understanding data models

As you work through the lab, fill out Tables A and B below based on information from the lab introduction, exercises, course text, and lecture.

Question 7: Table A

	Vector	Raster	TIN
Briefly describe the essential characteristics of each.			
Include the types of data generally represented (i.e. continuous or discrete)			
Give an example of a likely geographic feature that would be represented.			

Question 8: Table B

B	File Geodatabase	Enterprise Geodatabase	Shapefiles
Historic Software Origin:	ArcInfo8	Earlier ArcInfo	ArcView

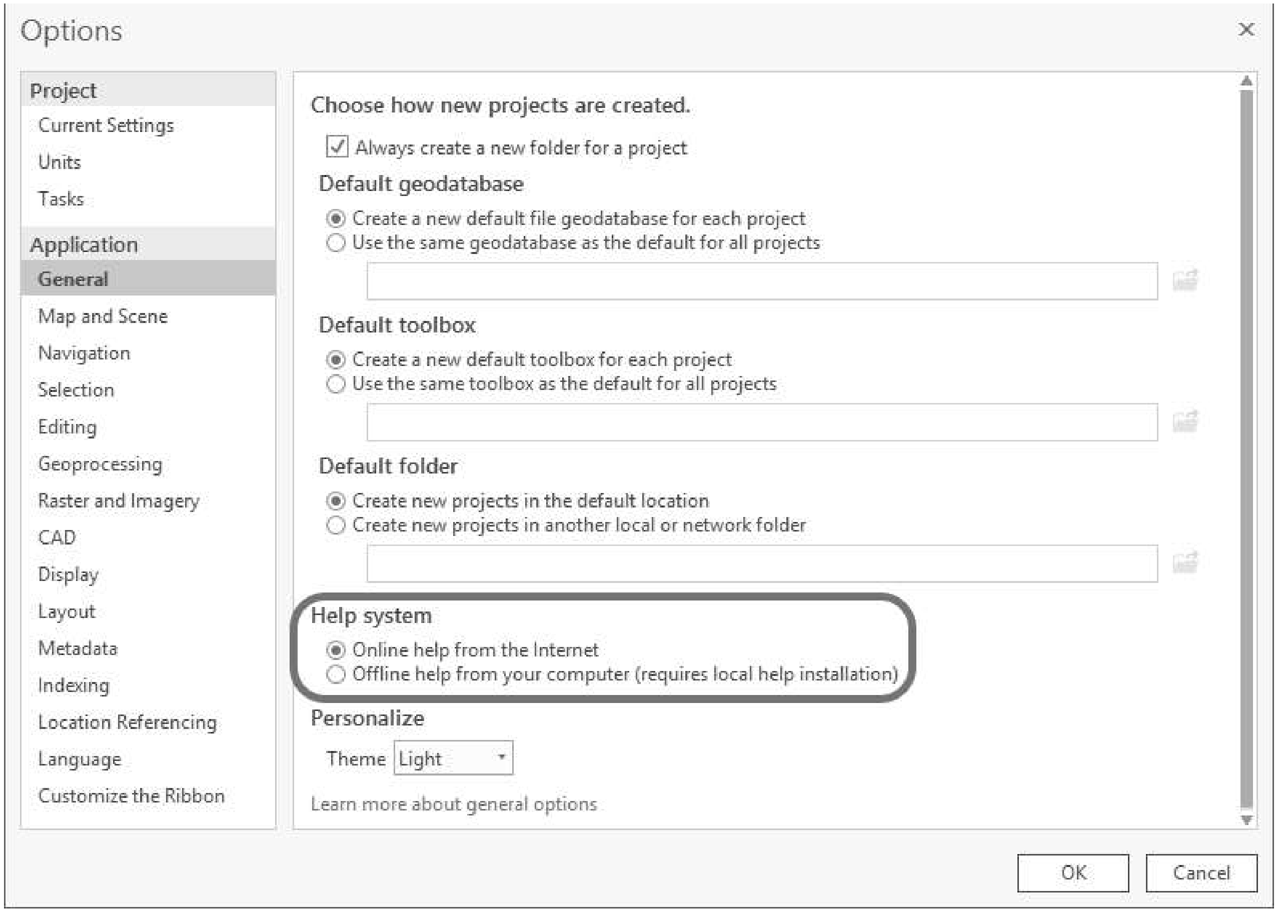
In what type of files is the data stored in the computer? How the data is stored in the computer (i.e. does the data need to be in a special type of folder? What files are required for the data model?)			No special folder for storage. Three files containing spatial and attribute data are required, there may be other files with index information.
Describe the topological features in each data model	*Allows for topological feature classes, geometric networks. Polygon topology implemented through on-the-fly topological editing.*	*Allows for topological feature classes, geometric networks. Polygon topology implemented through on-the-fly topological editing.*	
What type of data can be created in each data model?			
What is the size of files that you can store?			

ArcGIS Help

Help is THE MOST important resource you will have for this class. Read it, learn it, use it.

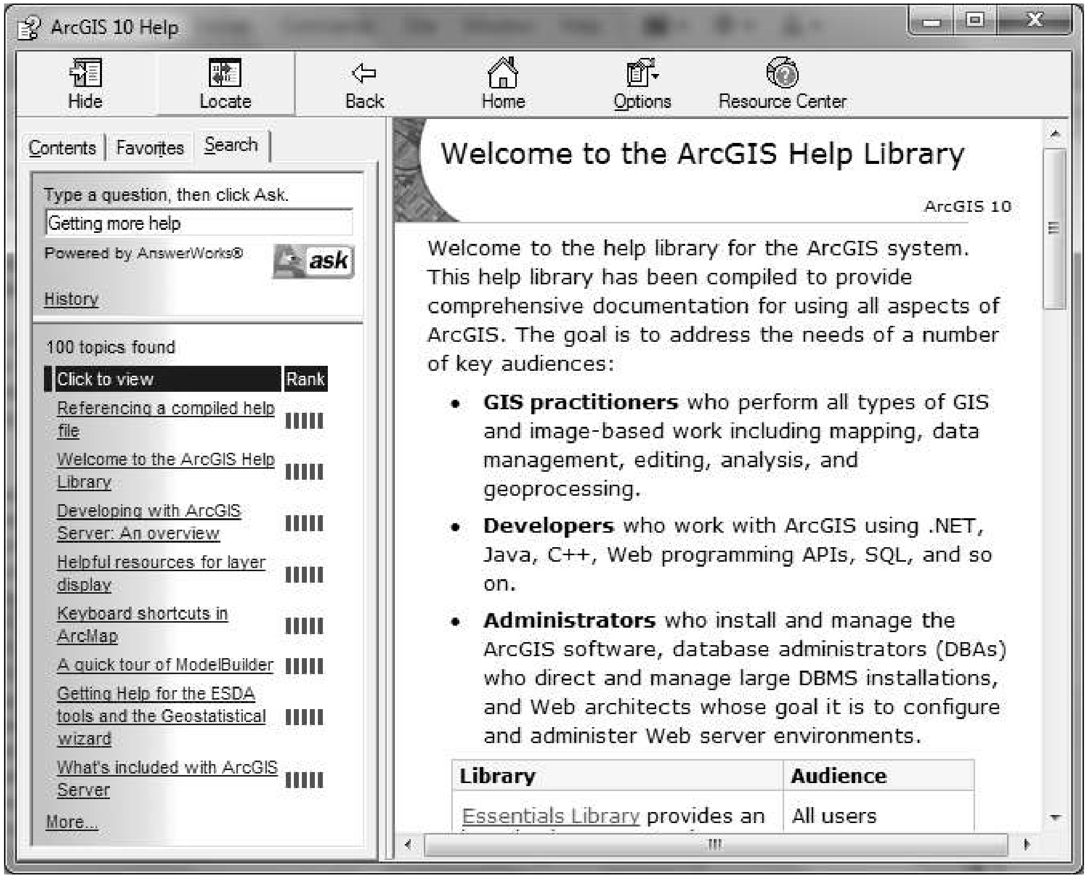
There are two ways to access native Help resources for ArcGIS Pro – one online and the other offline, local to your computer. Under the Project tab, click on "Help" to visit this.

By default, you will be taken to the online collection (http://pro.arcgis.com/en/pro-app/help/). The same content is also available offline. In the Options under the Project tab, visit General under Application. Choose the "Offline help" radio button. Now, when you select Help (under the Project tab), you will be taken to an ESRI-built native Help application.



support.esri.com is another excellent resource as well as the GIS Dictionary.

Please note: If you prefer to use Google to search for online Help resources, you may accidentally access ArcGIS Desktop Help sites. Don't get sucked in! ArcGIS Desktop, having been around for far longer and having had far more visitors on its Help pages, is often cached to show up at the top of a Google search. If you are using Pro, click on the ArcGIS Pro Help links.



Use GIS Dictionary or ArcGIS Help Online to answer the following questions about a **geodatabase.** * Do not copy and paste. Try to answer within 1-3 sentences.

Question 9: List at least 4 feature classes that a geodatabase can contain.

Question 10: What are tic points?

Question 11: What is topology?

Use GIS Dictionary or ArcGIS Help Online to find **shapefiles** to answer the following questions. * Do not copy and paste. Try to answer within 1-3 sentences.

Question 12: How many feature classes can a shapefile use?

4.2 Mystery Models

1. Copy the *mystery* and *sb* directories onto your removable disk or to your Lab_2 folder in E:\John176BL\JohnLab2 if you have not done so already.

2. Connect to this folder in Catalog View or Pane (under the Insert Tab Ribbon) and examine the layers in the folder *mystery*.

Please only focus on the arc files (e.g. mystery1_arc) for question 13.Mystery 1,2,3,4,5, and 8 can be added from the geodatabase "mystery.gdb".

Question 13:

What are the data models for each of the layers? What geographic feature does each layer seem to represent?

mystery1 (arc, in gdb) --

mystery2 (arc, in gdb) --

mystery3 (in gdb) --

mystery4 (arc, in gdb) --

mystery5 (in gdb) --

mystery6 --

mystery7 --

mystery8 (in gdb) --

3. Once you have identified the layers and their data models, convert *mystery5* into the same data model as *mystery2*. You will have to figure out how to do this yourself, but here are some hints:

Converting Between Data Models

- You will have to use Toolbox to accomplish this task.
- Find the toolbox menu that would contain the appropriate tools. Find the appropriate sub menu for converting data in *mystery5*'s data model, then find tool that will let you convert to *mystery2* 's data model. Recall that you can drag-and-drop from Catalog Pane instead of typing or browsing.
- **Be sure to Uncheck the "Simplify" option when using tools that offer this option**. Use the defaults for everything else.

4. Specify the output directory, give the file a name you will remember, and run the conversion.

5. Take your resulting layer and display it in Map, along with *mystery5.*

An example conversion (filename: convert_mystery5) is provided in the geodatabase. You can compare your result to that.

Question 14:
How similar are *mystery5* and your converted layer? Briefly describe the major differences between the two. What is the cause of them?

6. Go to the directory *sb*.

7. Now, add *sbcontour, sbdem*, and *sbtin* from your Lab2 sb folder (*sbcontour is in* sb.gdb). Ensure that *sbcontour* is on top of *sbtin,* and *sbdem* is on the bottom. To make the display intelligible, you will have to change the properties for the two layers.

8. Go to the Appearance tab.

9. Change the transparency of *sbtin* so that the DEM raster can be seen underneath it, and click OK.

You will be repeating these steps to change a layer's properties many times throughout the quarter. You will probably find the Properties functions very useful. Map's Style Manager is an easier way to manipulate layer properties that we will learn about later on in the quarter but feel free to experiment with it.

Question 15:

Which of the three layers (*sbdem*, *sbtin*, *sbcountour*) do you think was the original data layer? Which is "second generation" and which is "third generation"?

Question 16:

Please provide the reasons supporting your choice in question 15.

4.3 Data Models and Toolbox

ArcGIS continues to draw on a variety of data models and formats for its functionality. The Toolbox tools reflect this, requiring different data types for input depending on the analysis and data management tasks at hand.

Finding and Examining Tools

Again, recall that you can open the Toolbox by clicking on the Toolbox button in Catalog or Map. 

If you can't find a particular tool in Toolbox, use the Search for Tools button under the Geoprocessing drop-down menu near the top of the Map.

Note: The engine is set up for individual tools. Do not use the search toolbar to search for entire toolsets.

Important: If you click on a tool and get an error message referring to a license problem, you may just need to enable the extension that manages the tool. To do this, click the Customize drop-down menu near the top of the Map, choose Extensions, and then make sure all of the extensions are checked. Each extension requires a license and you may or may not have access to all of them. UCSB has licenses for almost all extensions.

Every time you move your cursor near the tool name, a short description displays in the Toolbox window. For more information on a tool, click to open it and and click Show Help near the upper right of the dialog box if it is not already showing. You can now click your mouse in any of the input boxes to read a description of what the tool is looking for.

Question 17:

Find each of these tools and determine what data model type(s) (or perhaps other file types) it takes as input:

a) Clip, Select, Buffer, Intersect, & most other Analysis Tools (all the same answer)

b) Add Spatial Index
c) Feature Class to Geodatabase
d) Raster to Other Format
e) Join field

4.4 AATs & PATs

As discussed above, coverages have been the standard vector data model for previous releases of Arc. With the release of ArcGIS 8, all of the modules of Arc/INFO (Arc, ArcEdit, Grid, Tables, ArcPlot, INFO, etc) were fully integrated, and the new geodatabase model was launched and promoted. However, coverages have been in use for such a long time, you will undoubtedly encounter them.

Recall that coverages employ the ***georelational database model***. The INFO part of Arc/INFO was the relational database manager for earlier versions of the Arc software (Arc was the name given to the mapping component). An INFO file is a table that stores the information associated with the geographic features of a spatially referenced dataset. This gives a GIS the ability to manipulate information both spatially and via standard tabular database functions. An example relational model is when two tables share a common column. In a georelational model the individual records in two or more tables are related through their location in space. The polygon coverage below serves as a simple example of this concept. The common column is often called the KEY (of ID) column and is used to relate the tables and features.

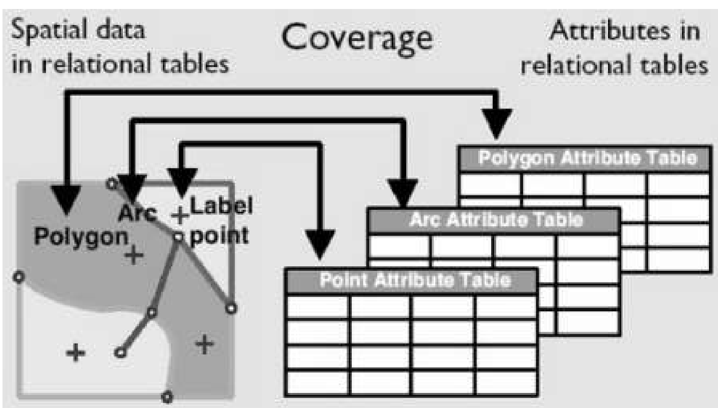


Figure 4. Diagram showing the coverage data structure for storing vector data.

The coverage format is no longer supported in ArcGIS Pro. But we can still explore the data since it is now represented in multiple files (e.g. the *cacounty* coverage is composed of *cacounty_tic, cacounty_label, cacounty_arc, cacounty_region_cty, and cacounty_polygon*).

Let's explore an attribute table that is part of the *roads* coverage. Go to the Map view and explore the arc attribute table in roads.

Question 18:

How many records are there in roads_arc?

Question 19:

Explore the coverage attribute tables help page and find out what do FNODE# and TNODE# mean?

Question 20:

What other attribute information can you recognize or guessed at in the table (pick 3 columns)?

To take a look at polygons and Polygon Attribute Tables (PATs), open *cacounty*. Explore the tables for the *arc*, *polygon*, and *region.cty* coverage feature classes.

Question 21:

How many counties are there in California?

Question 22:

Why do the AAT (Arc Attribute Table), PAT (Polygon Attribute Table), and RAT (Region Attribute Table) have different numbers of records?

Question 23:

Explain the relationship between *arc*, *polygon*, and *region.cty* in this coverage.

Question 24:

What are the *label* and *tic* feature classes for?

Hints: To figure out the answers, you will need to examine several of the tables.

Question 25:

Map for Lab 2: Make a map of the greater Santa Barbara metropolitan region with the roads coverage overlaid on the contour coverage. You will have to choose appropriate properties for the two themes so that map readers can tell them apart. Also, make sure you follow the **basic principles of cartography** outlined in Lab 1. Export your map.

5. Conclusion

In this lab, you have gained a basic understanding of geographic data models and data modeling, and the primary data models used in ESRI's ArcGIS Pro software. You have seen how the ESRI data models are similar and different from each other, and how each has advantages and disadvantages for certain purposes. You have gained further experience with some basic ArcGIS skills, such as changing properties and using the help functions.

6. Lab Responses

This is NOT a "quiz". This is your weekly lab assignment, therefore, you should have your lab text open alongside this quiz so that you can fill the answers as you go along.

To save and come back to your responses, submit an attempt. Later, you can re-attempt the quiz and your previous responses will be available there. You have an unlimited number of attempts.