Lecture 6 - Raster Data Model & GIS File Organization

I . Overview of Raster Data Model

Raster data models define objects in a fixed manner – see Figure 1. Each grid cell has fixed size (resolution). The raster approach is better for delineating continuous geographic fields of data, which can be thought of as surfaces. Some examples of raster geographic data include elevation, temperature, atmospheric pressure, airborne or satellite imagery.

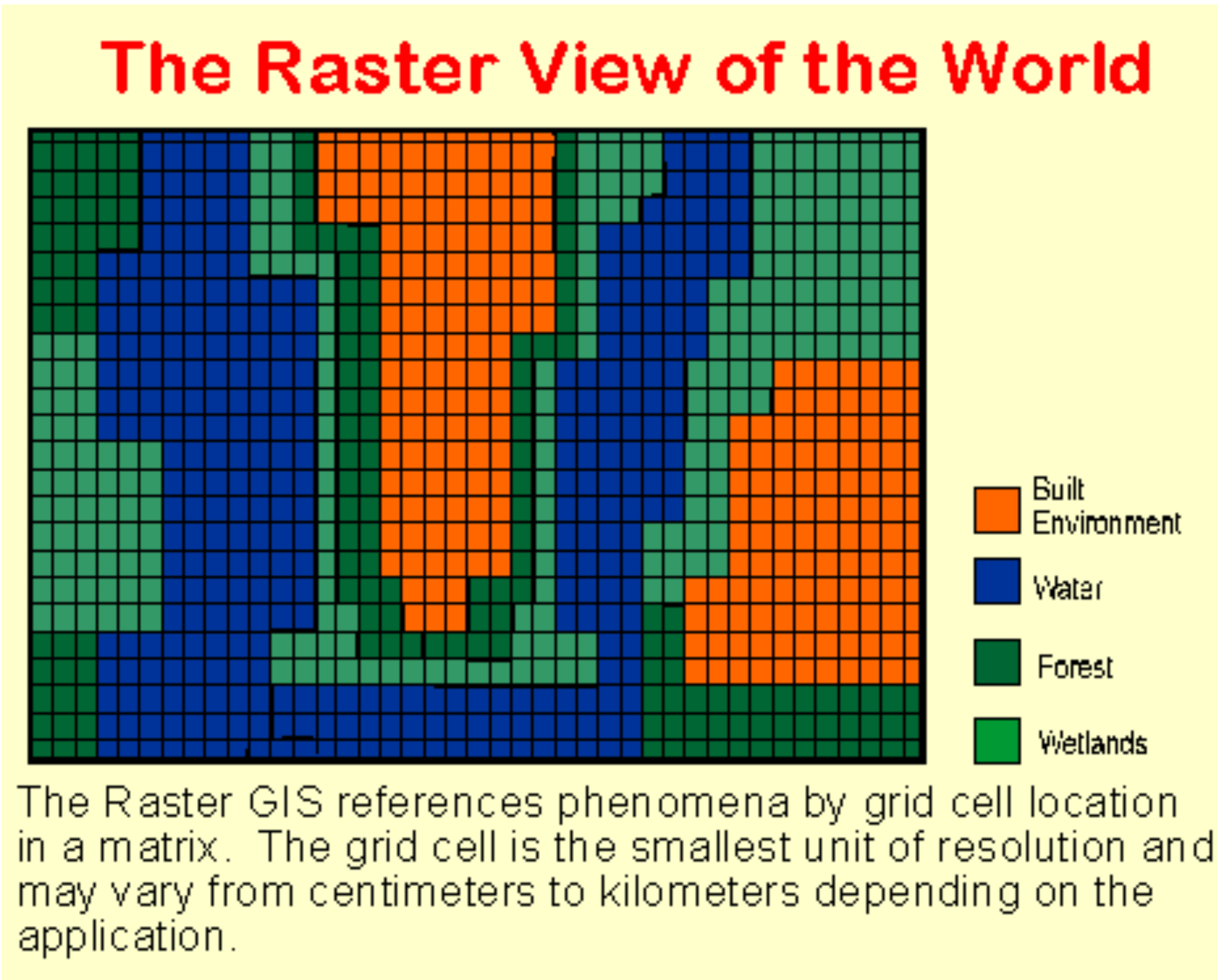


Figure 1. Raster Representation of different land use areas.

Pixel Resolution - The spatial resolution of a pixel varies depending upon the specific type of data. For example, aerial photography can be used to generate a raster layer with an extremely fine resolution such that each pixel can represent as small as 3 inches of the earth's surface. Conversely, remotely sensed data obtained from satellite platforms can represent a very large area. For example, the Advanced_Very_High_Resolution_Radiometer (AVHRR) has a misleading name. This satellite mission has the ability to discriminate between different wavelengths of light with a very high degree of spectral resolution but there is a trade-off in that the spatial resolution over which it is sensing this light has an extremely coarse resolution. Each pixel monitored by has a spatial resolution of 2.4 km!

- *Zoom in on the campus aerial photograph layer in Assignment #3. What is the spatial resolution of this dataset?*

File Size - With raster data there is a fundamental trade-off between spatial resolution and file size. Basically, the finer the spatial resolution the more cells will be required to cover a given area as shown in Figure 2 and the greater size of the file. Additionally, the relationship between spatial resolution and number of cells is based on a factor of two. So that if you decrease the spatial resolution of a raster image by a factor of two the number of cells required to cover a given area grows by a factor of four. High spatial resolution raster layers can commonly be as large as GB in size. The larger the raster file the longer it will take to load into a GIS program. A common practice when dealing with raster files is to clip the raster so that it represents only the area of interest. The Night Lights layer in your packet actually covers the whole world.

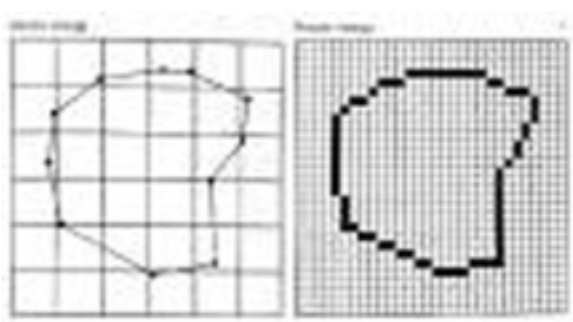


Figure 2. Spatial resolution of raster data is indicated by the pixel size. Left image depicts low resolution whereas right image indicates fine resolution.

Data Compression - Because of the inherently large size of raster data various data compression schemes have been developed that allow GIS software load a raster layer faster. You may have noticed when you first added the raster layers in your data packet that it asked you to build pyramids. This is a common way of saving space in a raster file. I will not go into the details of the compression algorithms - if you want to learn more about them take a computer science class. An example of compression that your probably have interacted outside of this class with is the jpeg format for an image, which dramatically reduces the size of an image file.

Data compression falls into two categories. Compression can be lossless in which none of the original information is altered or less ideal is lossy compression where some of the pixel values are altered. Lossy compression in raster data can be thought of as generalization where some of the detail is given up. Losslees compression is used when it is critical that none of the original characteristics of the earth's surface are changed - like for raster layers that depict land use or elevation. Lossy compression is used on data layers that are not used for analysis but for basic visualization such as aerial photographs. Two of the most common algorithms are based on run-length coding and Quad Tree representations (see Campbell and Shin; Chp. 4.1)

Attribute Information - Raster data tends to be more limited in terms of attribute information. Geographic objects are depicted by pixel intensity. In grayscale images (black and white) pixel intensities can vary between 0 (black) and 255 (white). With a color image there are several options; a common format is RGB (red, green, blue) where millions of colors can be shown by 0 to 255 intensities of R, G, and B. A real advantage and power of raster layers is that mathematical operations can be performed between multiple raster images facilitating GIS analysis. Additionally, raster layers can support numeric analysis such as least-cost path modeling - i.e. routing of water within a watershed.

- *Figure 3 shows an example of a raster attribute table. Note you only get the total of the number of pixel with a certain value. The values reflect a given attribute type. Check out the attribute table for your Land Cover layer in your data packet.*

Section II - Raster Data File Formats

Some common stand-alone raster data formats include:

- ESRI Grid - ESRI's main raster type of file that can support either pure binary format (0, 1) or ascii format - Ascii format is based on the numbers 0 to 255 with each number representing either text or a special character - Commonly, unformated text is saved in an ascii format
- GeoTIFF - A tiff is a common image file format. A GeoTIFF has auxiliary files that allow a GIS program to perform georeferencing of any position on the image
- GeoPDF - A PDF that can be georeferenced - You need to download a plug-in for Adobe Acrobat
- USGS DRG - Is basically a scanned raster image of a USGS topographic map.
- MrSID - Multi-Resolution Seamless Image Database, which is based on a compression format that can have both lossy and loseless compression -this is a good format for aerial photos
- Jpeg - Common format for images - Jpegs can be loaded into a GIS program and in some cases can even be georeferenced. It can support both lossy and loseless compressions formats.

	A	B	C	D	E	F
1	OBJECTID	Value	Count	Raster1	Raster2	NewValue
2	1	1	1316356	42	42	0
3	2	2	299180	42	41	113
4	3	3	373006	52	41	115
5	4	4	468178	52	43	145
6	5	5	1080865	42	43	143
7	6	6	370235	52	42	130
8	7	7	818029	43	43	0
9	8	8	252788	43	42	129
10	9	9	269347	42	81	203
11	10	10	1377739	90	43	149
12	11	11	1375924	90	41	119
13	12	12	413666	71	43	146
14	13	13	245665	71	41	116
15	14	14	216226	71	42	131
16	15	15	619037	52	81	205
17	16	16	224591	43	41	114
18	17	17	167698	22	43	138
19	18	18	206789	43	81	204
20	19	19	834195	90	42	134
21	20	20	85898	22	42	123
22	21	21	6821	52	31	70
23	22	22	16618	71	31	71
24	23	23	116579	22	41	108
25	24	24	635108	90	81	209
26	25	25	758485	71	81	206
27	26	26	19818	31	31	0
28	27	27	21694	31	43	141
29	28	28	42068	31	81	201
30	29	29	405423	22	81	198
31	30	30	224870	23	81	199
32	31	31	49809	52	82	220
33	32	32	14249	31	42	126
34	33	33	18664	31	41	111
35	34	34	7286376	81	81	0
36	35	35	659904	81	82	222
37	36	36	80702	71	92	281
38	37	37	569509	81	41	117
39	38	38	5613	42	82	218
40	39	39	78278	71	82	221

Figure 3. Raster data has a limited ability to encode attributes. Typically limited to the numeric value for each pixel.

Section III - Vector versus Raster - Which Data Model is Most Appropriate?

Data Structure - Vector data represents discrete geographic entities and tends to be relative simple compared to raster data. Additionally, vector files tend to be much smaller (KB to MB) than raster data (MB to GB), and therefore, vector data is always uncompressed; whereas raster data is generally compressed.

Coordinate, Datum, Projection Conversion - Generally, vector data can be easily converted "on the fly" by most GIS programs. Raster data is much more difficult to convert because the location of every pixels has to be determined. Conversion generally involves resampling of pixels - *i.e.* stretch pixels so they fit into the new system.

Analysis - You can perform mathematically operations on raster data layers - *i.e.* you can add, subtract, multiply, and divide two raster layers together and get a meaningful result. Vector data, if it is topologically correct is more suited for network analysis; although there are other types of analyses that can be performed with vector data, which we will discuss later in the course.

Positional Precision - With raster data the size of geographic entities represented can be no smaller than the spatial resolution of a pixel. And generally, it takes at least several pixels to represent a geographic entity. With vector data there is no limit to the precision to which you can define entities. If you have a millimeter grade GPS data then you can represent entities to this level of precision.

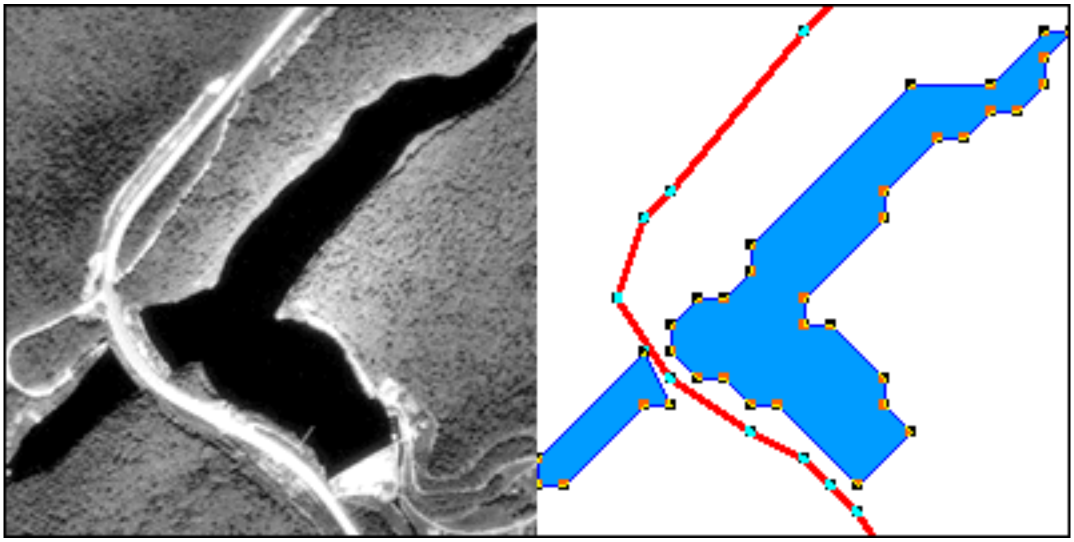
- *Show a comparison of spatial accuracy associated with vector and raster data of a road and reservoir (Figure 4)*

Accessibility - Raster data is relatively easy to modify on a global basis (throughout the entire layer) through simple mathematical operations. Conversely, vector data editing can be quite tedious involving spending hours searching for small gaps, overshoots, and dangles (see Figure 2 from lecture 5 notes).

Visualization - Raster data is good for visualizing continuous surfaces. It can also depict discrete entities in a stair-step fashion, where it is obvious that the edge of the entity consists of pixels especially at large scales. Vector data can provide a more realistic representation of discrete geographic entities at any scale.

Encoding Attributes - Finally, raster data can only encode one attribute, typically the pixel value, while vector data is nearly unlimited in the number of attributes it can support.

A



B

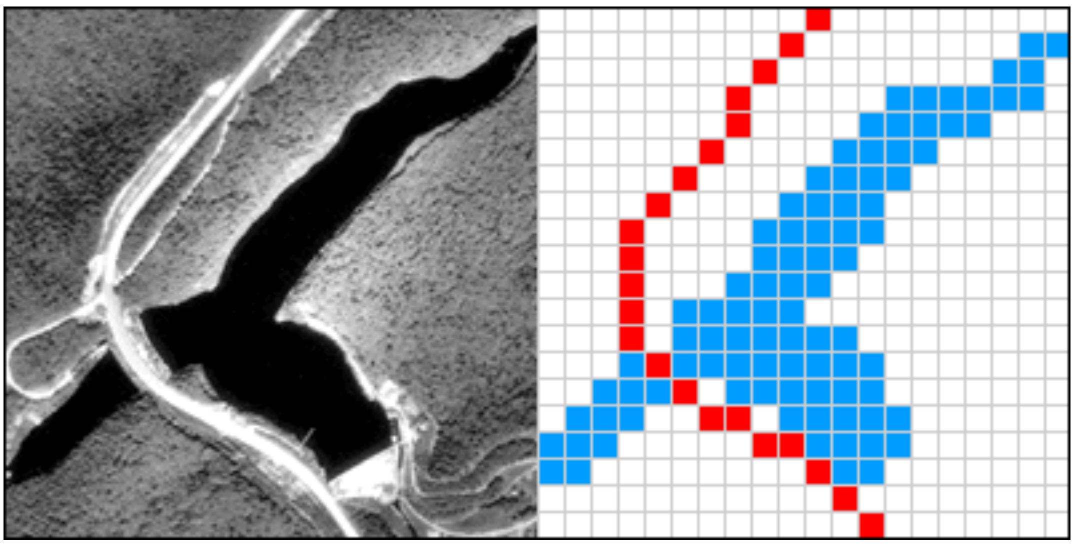


Figure 4. Representation of road and reservoir, (A) Vector approach and (B) raster approach.

Section IV - GIS File Organization

All GIS systems and data layers are created and maintained within a database management system (DBMS). A DBMS is a software tool that manages the database and ensures the integrity, consistency, and security of the data stored within the database environment.

There are two file structures for organizing vector files within a DBMS associated within a GIS that are shown in Figure 5.

Georelational File Structure - where spatial data is stored separately from the database of attributes and the two sets of files are linked with a Feature ID number (typically the left most attribute or field in an attribute table). Most common vector file formats with this type of file structure includes shapefile and coverage formats.

Geodatabase File Structure - where both spatial and database information are stored in the same file structure so that the file can be used for both GIS and more traditional database applications. In ArcGIS (9.x to 10.x) the database management system (DBMS) the file type is referred to as a **geodatabase**. Within a **geodatabase** multiple layers and tabular data can be stored. Basically, you can think of a geodatabase as a container that can store many types of data including vector and even raster layers. The power of a geodatabase is that all the files associated with a project can be stored in one database making easier to share your data with co-workers. See link on next page for more about geodatabases.

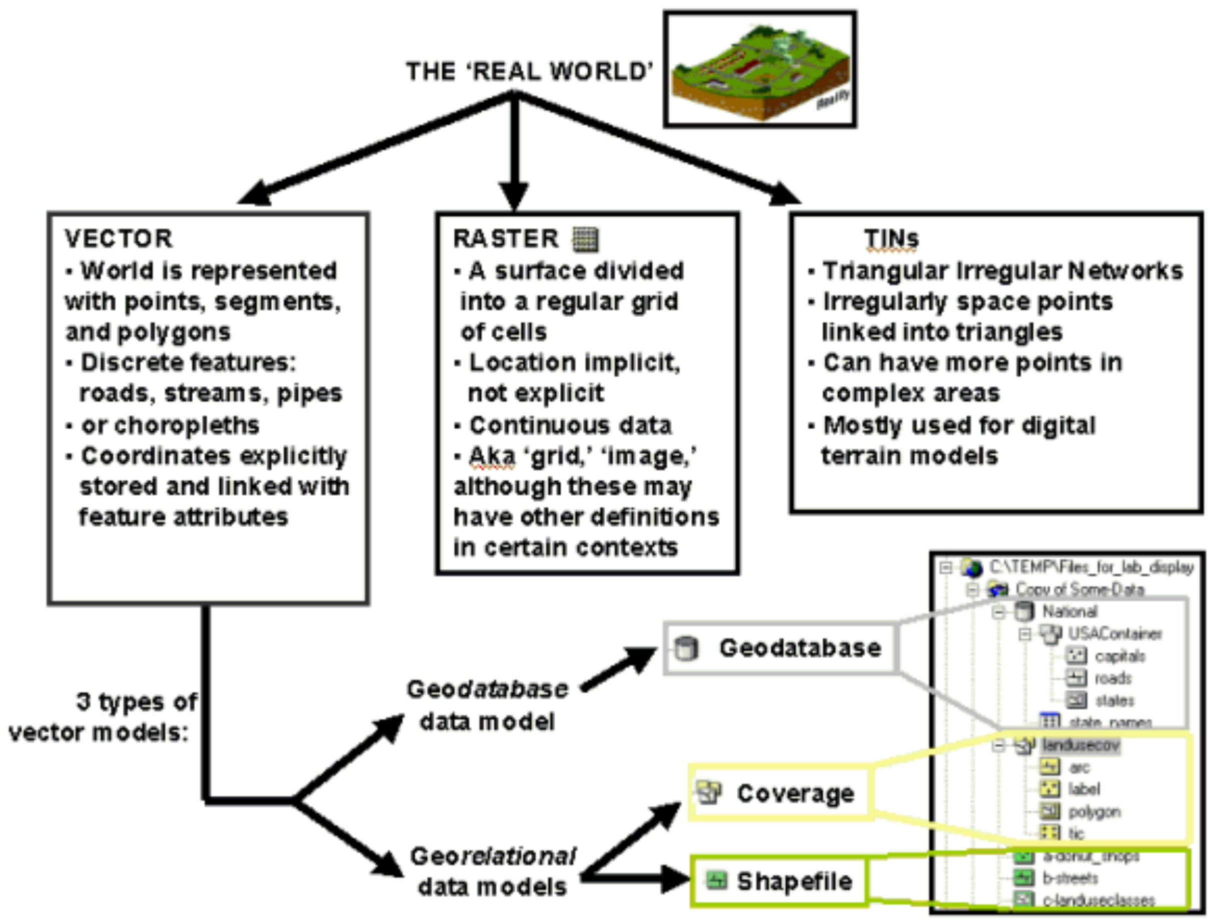


Figure 5 Hierarchy of ESRI's ArcGIS data models. From http://www.geog.ucsb.edu/~jeff/gis/data_models1.html

Readings
GIS Commons webpage; Chapters 1, 4.
Campbell, J and Shin, M., 2011, Essentials of Geographic Information Systems. Chp. 4.1.

Terms

Pixel Resolution	Lossy compression	Loseless compression	RGB
ESRI Grid	GeoTIFF	GeoPDF	DBMS
Georelational File Structure		Geodatabase	

Concepts
For what type of features is the raster data model more appropriate? And the vector data model?
What are the different characteristics that define a raster data layer?
For raster data what is the general relationship between file size and pixel resolution.
Make two lists. What are the advantages of the raster data model? Disadvantages? Do the same for vector data

HOMEWORK
1. Describe two geographic entities that are best defined by the vector data model.

2. Describe two geographic entities that are best defined by the raster data model.

3. In your own words contrast the vector and raster data models.

4. Discuss some of the advantages of a geodatabase.