Lesson 6: Raster Data Quality

## INTRODUCTION

In this lesson you will review what raster data files are and explore different raster types as well as gain a better understanding of the elements related to raster data quality. The lesson explains ways to store raster data, managing rasters and their attributes. Resampling techniques were illustrated and explained as well as the four interpolation methods used in resizing and resampling. The lesson concludes as you learn about float and integer rasters and their significance.

## LESSON OBJECTIVES

By the end of this lesson, you will be able to:

1. Define key terms relating to types of raster data, storing data, managing raster attributes, resampling techniques, and integer and float grids.

2. Compare rasters and vectors.

3. Describe the methods of storing raster data and how attributes are managed within a raster data model, to include the differences between integer and floating point values

## LEARNING SEQUENCE

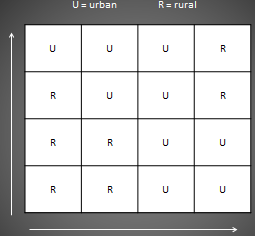
|  |  |
| --- | --- |
|  | |
| Required Reading | Read the following:  Spatial Data Quality   * Data Quality * Measuring Spatial Data Accuracy and Quality * National Standard for Spatial Data Accuracy (NSSDA) * Locational Errors * Data Aggregation |
| Assignments | Complete the following assignments:   * Lab 6: Raster Data Structure * Quiz 6: Raster Data Structure |

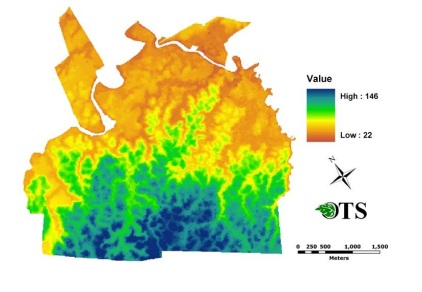
## INSTRUCTION

## Raster Data File Review

### What are Raster Data Files?

A raster is composed of cells in a grid pattern and each cell is square and adjacent to its neighbors. The cell dimensions define the cell area in the cell area defines spatial resolution. Phenomena of interest are represented by the attribute value stored at the center of each cell, and raster files best represent continuous data.

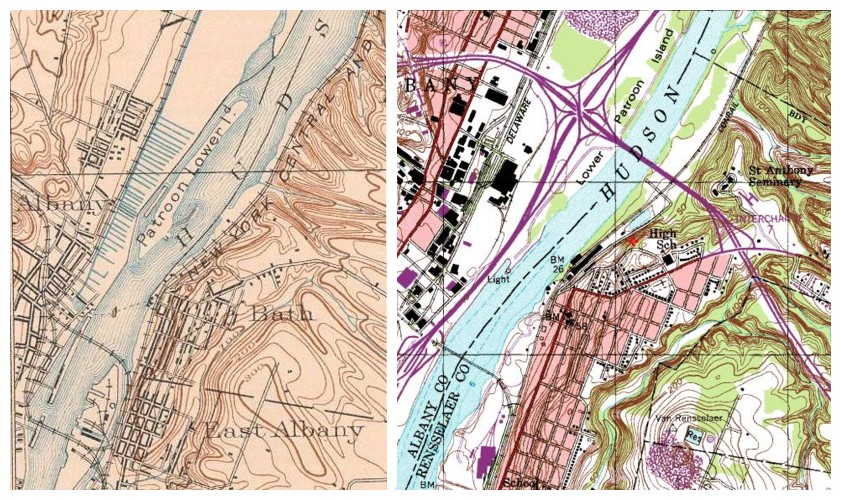
This is an example of a raster storing one of two variables: urban or rural. Note that the structure of a raster is a grid pattern of squares for each square adjacent, and each cell only stores a single value.

Rasters are created at many different resolutions in many different geographic scales. Rasters can also be assigned colors which reference an individual, or range of values in the raster’s cells.

## Common Rasters

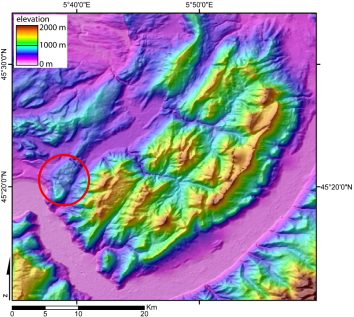
### Digital Raster

A digital raster graphic is a geo-referenced raster layer from a scanned map. Each cell in a digital raster graphic stores a color value representing the color shown at that location on the scanned map. Digital raster graphics are good for quick reference for consistency and data collection and may be used for general backdrop when navigating or displaying other geospatial data sets. Here are two examples of digital raster graphics. For both digital raster graphics, each cell represents a single color value.



### Digital Elevation Model (DEM)

A digital elevation model provides elevation data and a raster format which is typically a product of ground and/or aerial survey. In a digital elevation model, each cell stores a single elevation value.

This example of DEM shows a surface with its elevation represented in meters. While each cell of a raster represents in elevation value, colors have been assigned to the elevation values to make it look more interesting. It is important to note that a digital elevation model does not store colors, only elevation values and you should be careful and not assuming they particular color always represents a particular elevation.

### Satellite Imagery

## Satellite imagery is acquired by satellites orbiting the earth taking images of areas of interest and transmitting the images back to earth. Satellite imagery is often uses base maps or the basis for thematic maps and digitizing. Satellite images represent variations in color intensity of cover in each cell represents a single color value.

[](//upload.wikimedia.org/wikipedia/commons/6/60/CUET_NASA_Satellite_Imagery.jpg)This is an example of satellite imagery. This imagery shows colors in the visible spectrum of the electromagnetic spectrum. Each cell in the raster represents a single color value.

### Digital Orthophotos

A digital ortho photo is a photograph from which effects of tilt and relief are removed by the rectification process. Digital ortho photos have a uniform scale and serve well as a base map or for the basis of digitizing. Each cell in digital ortho photo stores a single color value.

This image is an example of an ortho photo. This imagery shows colors in the visible spectrum of the electromagnetic spectrum in each cell in the raster represent a single color value.

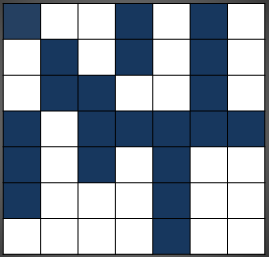
## Storing Raster Data

### Why Store Raster Data?

This section of the lesson will cover two methods of storing raster data in a computer. The two methods covered are: cell-by-cell, and run-length encoding. There are other methods of storing raster data in a computer, however these two methods are the easiest to understand and are sufficient in illustrating the reason you would want to choose different storage methods.

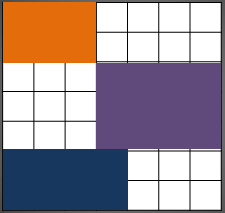
### Cell-by-Cell

The cell-by-cell storage method writes every individual cell value into a file by rowing column. This method is the simplest to implement, however uses the most amount of memory as every single cell in the raster is stored in memory. You may wish to use the cell-by-cell storage method when values change often and you need to update the raster quickly.

This is a visualization of how the computer would store a raster in memory using the cell-by-cell method. Notice that each cell is individually distinguishable and stored individually. Since is raster has seven rows and seven columns, the computer stores 49 individual values, one value for each cell.

### Run-Length Encoding

The next raster storage method is run-length encoding. Run-length encoding should be used when there is a large amount of contiguous redundant values. This method writes cell values into a file by row and group. For each stored growing group, it specifies the start and end cells that are contiguous and have the same value. This method is useful as it reduces the amount of storage required to save the raster to memory.

This is a visualization of how a computer would store a raster in memory using the run-length encoding method. Note that there are three large blocks of raster cells that are composed of multiple individual raster cells. Also note that in these large blocks, all the cells inside the block are contiguous. The way memory is saved in the computer is by storing a single value to represent all the cells inside of these large blocks instead of storing a single value for each cell inside the box. Since this raster has seven rows and seven columns, there are 49 individual cells, which would require 49 individual pieces of memory to store if we were using the cell-by-cell storage method.

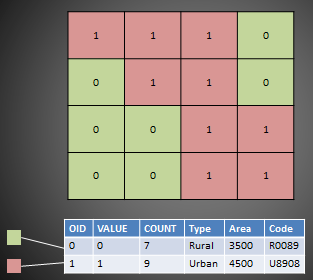
In the case of run-length encoding, 26 of the cells are broken into three blocks of contiguous redundant values, thereby only requiring three pieces of memory to store the value instead of 26. Therefore, in total for this raster, the white cells are stored in 23 individual pieces of memory, and the blocks are stored in three pieces memory, for a total of 26 pieces of memory required instead of 49 representing a 53% savings in memory space to store this raster compared to the cell-by-cell method.

## Managing Raster Attributes

### Attribute Tables

Each raster data set can have its own attribute table. Typically, each attribute table reference is a cell value that defines a class, group, or category. The reason rasters may have associated entry table is because a raster can only store a single value per cell, however, that cell may represent complex data that requires multiple attributes to define appropriately. Therefore, by associating an attribute table to a cell value, we can store more information that a raster normally would be able to store.

Here is an example to consider that illustrates this point. A satellite image has undergone a classification method that created a raster data set defining urban versus rural areas. In the image, it may use colors or characters to represent each cell value. If each color or character represents either urban or rural unambiguously, then we can attach additional attributes to each of these two categories.

This is a visualization of a classified satellite image. The character zero and color green represent rural, and the color red and character one represents urban. This attribute table provides a summary of the two land cover classifications.

So for example, for value zero stored in the raster, we can see that there are seven cells in the raster having a value of zero which represents rural areas. The total area covered by the seven rural cells is 3500 acres and a coding value for rural areas is R0089. Having this attribute table associated with the value zero allows the raster to link to this more detailed information.

### Attributes

Consider some of the attributes stored in a raster’s attribute table. When a raster attribute table is first generated, there are three default fields: OID, VALUE, and COUNT. You cannot edit the content of these three fields as they are exclusively managed by the software. The OID, which is short for ‘Object ID’, is a unique object identifier for each row of the attribute table.

The OID is used by the computer so it can differentiate one row from all of the other rows in the attribute table. The value attribute is a list of each unique cell value in the raster data set with each row in the attribute table storing a single value in the list of unique values. The count attribute stores the number of cells in the raster with the same cell value stored in the value column of a row.

### Storage of a Raster Attribute Table

Raster attribute tables are stored in different ways depending on the file format of the raster. If the raster is file-based, then the raster attribute table is saved at the same directory level is the raster. If the raster is stored in a geo-database, then the raster attribute table is saved in the same feature set as the raster. Note that other formats may store the answer tables different ways, however as there are many different raster formats, it is always best to reference the file format documentation to determine how the attribute table is stored.

### Uses of Raster Attribute Tables

Raster data sets are much like vector data sets in terms of managing the data. Both raster and vector attribute tables can be deleted, previewed, viewed, exported, join to other tables, or used to symbolize the values stored in the raster or vector data set.

### Linking to the Attribute Data

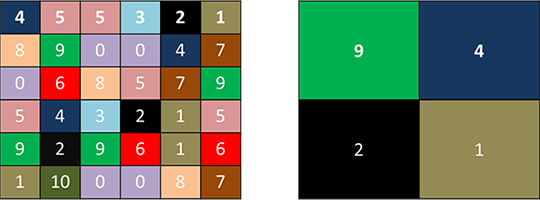
Much like with vector attribute tables, raster attribute tables can link to other tables. This may result in a large attribute table if you link multiple tables together, so you may wish to use caution or reduce the number of attributes you will link to.

## Resampling Techniques

### What are Resampling Techniques?

Resampling techniques are used to change the cell size of a raster and assign the new cells a value based off of the source cells from the original raster that is being resized.

For example, the raster on the left is our original, source raster which has 36 cells arranged in a 6 x 6 grid. The raster on the right is the new, resampled raster which has four cells arranged in a 2 x 2 grid. We will use this original, source raster as the input to all of our resampling techniques in this module.



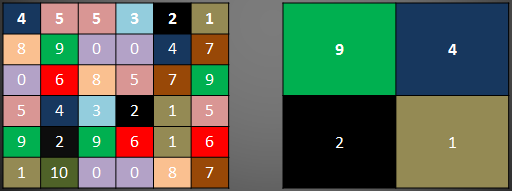
### 4 Interpolation Methods

## When we resample and resize a raster, we must use an interpolation method to assign the new cells values based off of the original source cells. This section covers the four most common interpolation methods used GIS for raster resampling. The four methods are: nearest neighbor, bilinear, cute convolution, and search.

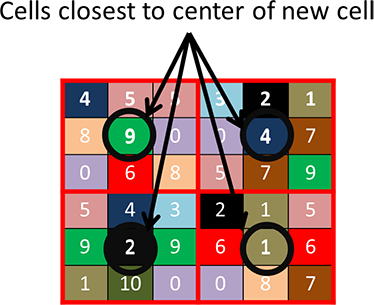
#### Nearest Neighbor

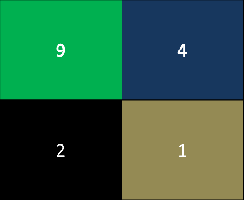
The nearest neighbor method performs a nearest neighbor assignment based on the closest source cell center to the resampled cell center. Of all for interpolation methods, this is the fastest interpolation method as it is the simplest to implement it does not require any complicated mathematics.

The nearest neighbor interpolation technique is useful for categorical data as it does not change or average the values of the cells. In other words, the nearest neighbor method will assign the new cell a value that exists exactly in one of the source cells. Let’s walk through a nearest neighbor example. On the left is our source raster which is in a 6 x 6 grid. Our resampled raster will be a 2 x 2 grid. Its overlay the source raster and the new raster to see how they relate spatially.



With the two rasters overlaid, we can see that for each new cell, nine source cells lay within. To determine the nearest neighbor, you must identify which cell from the source raster is closest to the centroid of each cell in the new raster.

In this case it is easy to determine the nearest neighbor for ournew cells. Since the new cells contain nine of the source cells, and the cell most central of the nine source cells is closest to the center of the new cell. Before nearest neighbors, one for each of the four new cells, or circled in the graphic.

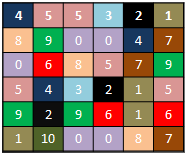
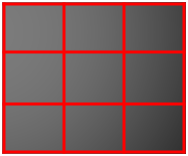


To complete the nearest neighbor interpolation, the values of the four closest cells from the source are signed to the values of the four new cells. In this case, the upper left cell gets the value nine, upper right cell gets the value for, lower left cell gets the value of two, and the lower right cell gets the value of one.

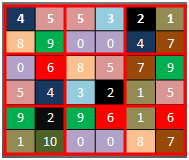
#### Bilinear

The second interpolation method we will cover is bilinear. The bilinear interpolation method determines the new value of the cell based on a weighted distance average of the four nearest source cell centers. Bilinear interpolation is great for continuous, numeric data as it takes a weighted average of the numeric data which is appropriate for continuous phenomenon. A byproduct and potential downside of a bilinear interpolation is that it causes smoothing of the data since it assigns new values based off of averages of the source values.

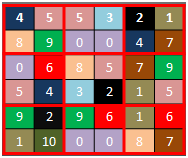
For this example of bilinear interpolation, we will again use the 6 x 6 input raster, and the new raster will be a 3 x 3 grid. Let’s lay them on top of each other to see the spatial relationship.

With the two rasters laid on top of each other, we can see that each new raster cell contains four source raster cells.



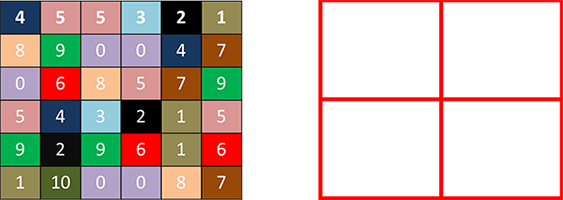
In this example case, the center of the new cell falls of the intersection of the four source cells, therefore, all for the source cells are selected for the weighted average and will be weighted the same.



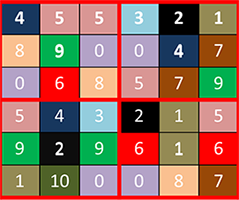
#### Cubic Convolution

The third interpolation method is cubic convolution. The cubic convolution interpolation method performs a weighted average of the 16 nearest input cell centers to the center of the new raster cell. Cubic convolution is good for continuous data, however, it requires significantly more processing time and bilinear interpolation and nearest neighbor.

Let’s see how cubic convolution works with a 6 x 6 input grid and eight 2 x 2 output grid.



Here we have the source and new rasters laid on top of each other. We see that each new raster cell contains nine source input cells, so this means that seven additional cells must be found outside of the new cell to reach 16 inputs.



The 16 cells closest to the center of the new cell center are first found. This is represented with the cells highlighted in yellow on the figure. The value of the new cell is the weighted average of the 16 source cells where the averages are based on the distance the input cells are from the center of the new cell. Once the top left cell has its value determined, we move on to the second, top right the cell.



For the top right new cell, the highlighted 16 source cells are used as inputs to the weighted average.



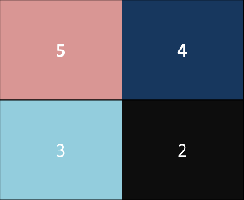
For the bottom left new cell, the highlighted 16 source cells are used as inputs to the weighted average.



And finally, for the fourth, bottom right new cell, the highlighted 16 source cells are used as input to the weighted average.



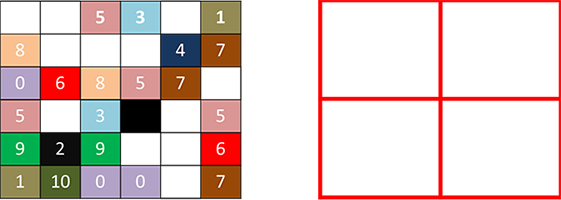
This is the resulting new raster based on the weighted values of 16 closest cells for each of the four new cells.



#### Search

The final method is search. The search method assigns the value for each new cell by searching for the closest cell from the source raster that has a valid value. The search method is appropriate for resampling categorical data to a coarser resolution when it is important to preserve cells with values that may be surrounded by cells with no data.

For this example of the search method, the 6 x 6 input raster of the left has been modified. A selection of cells has been changed from having valid data to having no data. The new raster is a 2 x 2 grid.



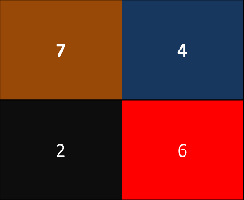
As the overlay the source and new rasters, we see that each new raster cell contains nine source cells.



For each of the four new cells, it searches for the nearest cell that has a valid value and chooses that value as the value of the new cell. In the case where two source cells are the same distance from the center of the new cell, the decision must be made to determine what to do with two input values.

This can typically be set by the software program and may include solutions such as averaging, choosing the most common, choosing either, or choosing a value that has a higher priority. In this case, in the upper left-hand new cell, both eight and six have been determined be closest to the center of the new cell and the value the new cell will be the average of eight and six.

This is the result of the search interpolation method.

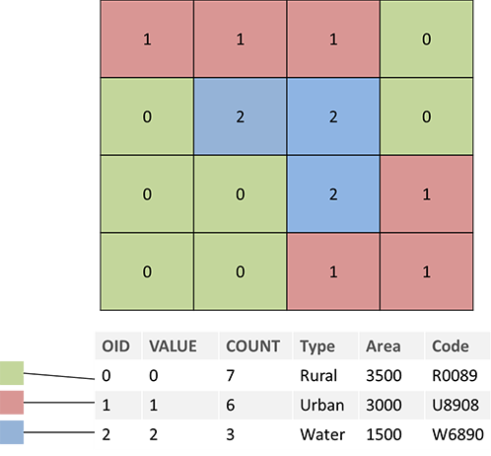


## Float & Integer Rasters

### Integer Raster

An integer raster stores a single integer value in each cell. Typically, attributes are also stored in a value attribute table attached to the raster. In the attribute table, one record is stored for each unique value found in the raster.

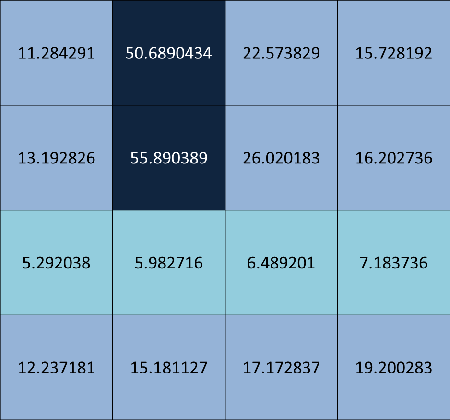
This is an example of an integer raster. Note that each cell contains a single integer number and is associated with an attribute value where there is one record for each unique value found in the integer raster. In the attribute table, it generates and stores the object ID, value, and count for each unique integer value. Additional attribute categories can then be added by the user to further describe the injuries in the raster.



### Float-Point Raster

A floating-point raster stores a single float number in each cell. A floating-point raster does not have a value attribute table as a can assume any value within a given range of values and the value of the cells do not fall neatly into discrete categories. Therefore, the value in each cell is the descriptor.

This is an example of a floating-point raster. In this case, this raster is a digital elevation model because each cell represents a single elevation value represented as a floating-point number. The raster is symbolized using a gradient the blue with a lighter blue color represents a lower elevation value and the darker blue color represents a higher elevation value.



## SUMMARY

In this lesson you reviewed raster data files and explore different raster types to gain a deeper understanding of the elements related to raster data quality. The lesson explained ways to store raster data, manage rasters and their attributes. Resampling techniques were illustrated and explained as well as the four interpolation methods used in resizing and resampling. The lesson wrapped up with you learning about float and integer rasters and their significance.

## ASSIGNMENTS

1. Lab 6: Raster Data Structure
2. Quiz 6: Raster Data Structure