The Raster Data Model

Llano River, Mason Co., TX
Rasters are:

- Regular square tessellations
- Matrices of values distributed among equal-sized, square cells

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>565</td>
<td>573</td>
<td>582</td>
<td>590</td>
</tr>
<tr>
<td>575</td>
<td>580</td>
<td>595</td>
<td>600</td>
</tr>
<tr>
<td>579</td>
<td>581</td>
<td>597</td>
<td>601</td>
</tr>
<tr>
<td>580</td>
<td>600</td>
<td>620</td>
<td>632</td>
</tr>
</tbody>
</table>
Why squares?

- Computer scanners and output devices use square pixels
- Bit-mapping technology/theory can be adapted from computer sciences
- 1-to-1 integer mapping to grid coordinate systems!
Cell Location Specified by:

- Row/column (R/C) address
- Origin is upper left cell (1,1)
- Relative or geographic coordinates can be specified
Registration to “World” Coordinates

Unregistered

Registered

510400
6005300
Registration to “World” Coordinates

 Requires “world file”:

- Specify coords. of upper left corner
- Specify ground dimensions of cell, in same units
World File – DRG Example

- 2.4384000000000000: CELL SIZE IN X DIRECTION (m)
- 0.0000000000000000: ROTATION TERM
- 0.0000000000000000: ROTATION TERM
- -2.4384000000000000: CELL SIZE IN Y DIRECTION (m)
- 487988.64154709835000: UTM EASTING OF UPPER LEFT CORNER (m)
- 3401923.72301301550000: UTM NORTHING OF UPPER LEFT CORNER (m)

/* UTM Zone 14 N with NAD83

/* This world file shifts the upper left image coordinate to the corresponding
/* NAD83 location, resulting in an approximated NAD83 image.

/* Map Name: Art
/* Map Date: 1982
/* Map Scale: 24000
Spatial Resolution

- Defined by area or dimension of each cell
  - Spatial Resolution = (cell height) \times (cell width)
  - High resolution: cell represent small area
  - Low resolution: cell represent larger area
- Defined by size of one edge of cell (e.g. “30 m DEM”)
- For fixed area, file size increases with resolution

![Spatial Resolution Diagram]

<table>
<thead>
<tr>
<th></th>
<th>Low Res.</th>
<th>Hi Res.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Area</td>
<td>100 m²</td>
<td>25 m²</td>
</tr>
<tr>
<td>Dimension</td>
<td>10 m</td>
<td>5 m</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Resolution of 30 m data is 9 times better (9X as many pixels) than 90 m data.

(50 m contours, vector data layer)
Resolution Constraint

- Cell size should be less than half of the size of the smallest object to be represented ("Minimum mapping unit; MMU")

Cell size = MMU

Cell size ~ ½ MMU
Resolution is size of sampled area on ground, not MMU

- "1 m resolution" raster data
- MMU = 2 m

- "1/3 m resolution" raster data
- MMU = ~ 2/3 m
Raster Dimension:

- Number of rows x columns
  - E.g. Monitor with 1900 x 1200 pixels

Dimension = 4 x 4

<table>
<thead>
<tr>
<th></th>
<th>565</th>
<th>573</th>
<th>582</th>
<th>590</th>
</tr>
</thead>
<tbody>
<tr>
<td>575</td>
<td>580</td>
<td>595</td>
<td>600</td>
<td></td>
</tr>
<tr>
<td>579</td>
<td>581</td>
<td>597</td>
<td>601</td>
<td></td>
</tr>
<tr>
<td>580</td>
<td>600</td>
<td>620</td>
<td>632</td>
<td></td>
</tr>
</tbody>
</table>
Raster Attributes

- Two types:
  1. *Integer codes* assigned to raster cells
     - E.g. rock type, land use, vegetation
     - Codes are technically *nominal* or *ordinal* data
  2. Measured "real" values
     - Can be integer or "floating-point" (decimal) values; technically *interval* or *ratio* data
     - E.g. topography, em spectrum, temperature, rainfall, concentration of a chemical element
Integer Code Attributes

- Code is referenced to attribute via a “look-up table” or “value attribute table” – VAT
- Commonly many cells with the same code
- Different attributes must be stored in different raster layers

Nominal Coded Raster

<table>
<thead>
<tr>
<th>Value</th>
<th>Count</th>
<th>Rock Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>21</td>
<td>Marble</td>
</tr>
<tr>
<td>5</td>
<td>37</td>
<td>Gneiss</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>Granite</td>
</tr>
</tbody>
</table>
Mixed Pixel Problem

- Severity is resolution dependant
- Rules to assign mixed pixels include:
  - “edge pixels”: not assigned to any feature—define a new class
  - Assign to feature that comprises most of pixel
Coded Value Raster Types

- **Single-band:** *Thematic data*
  - **Black & White:** binary (1 bit) (0 = black, 1 = white)
  - **Panchromatic** (“Grayscale”) (8 bit): 0 (black) – 255 (white) or graduated color ramps (e.g. blue to red, light to dark red)
  - **Colormaps** (“Indexed Color”) (8 bit): code cells by values that match prescribed R-G-B combinations in a lookup table

Figures from: Modeling our World, ESRI press
Single Band Examples – Black & White (Grayscale)

Black & White - 1 bit

Grayscale – 8 bit; black, white & 254 shades of gray
Each pixel contains one of 12 unique values, each corresponding to a prescribed color (Red, Green & Blue combination).
Measured, “Real Value” Attributes

- Commonly stored as floating point values
- Different attributes must be stored in different layers, e.g. spectral bands in satellite imagery
- Compression techniques for rasters of integer-valued cells, but not floating point (see below)
Multiband Image Raster Attributes

- **Multi-band Spectral Data**

- **Band = segment of Em spectrum**
- **Map intensities of each band as red, green or blue.**
- **Display alone or as composite**

Figures from: Modeling our World, ESRI press
Multiband Image
8 bits/Band, 3 Band RGB

E.g. Austin East 7.5’ Color Infrared Digital Orthophotograph (“CIR DOQ”)
Cell Values Apply To:

- Middle of cell, e.g. Digital Elevation Models (DEM)
- Whole cell, e.g. most other data

Source: Modeling our World, ESRI press
Digital Elevation Model
Airborn Magnetic (TFI) Map

TFI Pixel Values
How Are Rasters Projected?

- Problem: Square cells must remain square after projection.

- Solution: Resampling (interpolation); add, remove, reassign cells to conform to new spatial reference.
Raster File Size

- fixed by dimension, not information

At 1 bit/cell, file size = 8 x 8 x 1 = 64 bits (8 bytes)
Raster File Size

File Size = Rows x columns x bit-depth

- Bit depth: number of bits used to represent pixel value
- “8-bit” data can represent 256 values ($2^8$)
- “16-bit” data ($2^{16}$) allows 65,536 values
- “32-bit” data allows ~4.3 billion values
File Structure

Header: (dimension, max. cell value) + resolution, coordinate of one corner pixel, etc.

8, 8, 8

5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5
5 5 5 5 5 5 5 5

8, 8, 8

8 x 8 raster

Data File (linear array)
File Compression

- E.g. Run-length encoding

<table>
<thead>
<tr>
<th>Row, Run</th>
<th>Freq., Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1,1</td>
<td>8,5</td>
</tr>
<tr>
<td>2,3</td>
<td>4,5 2,2 2,5</td>
</tr>
<tr>
<td>3,3</td>
<td>4,5 2,2 2,8</td>
</tr>
<tr>
<td>4,3</td>
<td>4,5 2,2 2,8</td>
</tr>
<tr>
<td>5,2</td>
<td>6,2 2,8</td>
</tr>
<tr>
<td>6,2</td>
<td>2,2 6,5</td>
</tr>
<tr>
<td>7,2</td>
<td>4,2 4,5</td>
</tr>
<tr>
<td>8,3</td>
<td>1,5 3,2 4,5</td>
</tr>
</tbody>
</table>

Before: 64 characters
After: 46 characters
(28% reduction; ratio of 1.4: 1)
File Compression

E.g. Block encoding

<table>
<thead>
<tr>
<th>Block Size</th>
<th>Value</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>5,1 6,1 3,6 4,6 8,6 8,7</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>7,5 6,5</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>3,5 4,5 1,7 2,7 2,8</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>7,1</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>5,2 5,4 1,5 3,7</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>7,3</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>5,6</td>
</tr>
<tr>
<td>16</td>
<td>5</td>
<td>1,1</td>
</tr>
</tbody>
</table>

Before: 64 characters

After: 61 characters
(5% reduction ratio of 1.05: 1)
MrSID or ECW (Wavelet) Compression

- **Multi-resolution Seamless Image Database** – commercialized by [LizardTech](https://www.lizardtech.com/)

- Compression ratios of 15-20:1 for single band 8-bit images

- Ratios of 2-100:1 (!) for multiband color images

- Also ECW by ER Mapper Ltd. (now Intergraph/ERDAS)

  *** Enormous raster data sets now manageable on PCs and across web with this technology ***
“Lossy” vs. Lossless Compression

- Techniques that combine similar attribute information to reduce file size are “lossy” e.g. JPEG, GIFF, PNG, MrSID

- Lossless formats; TIFF, BMP, GRID
Raster Pyramids

- Store reduced-resolution copies of a raster for rapid display – e.g. ArcGIS, Google, many others
- Often combined with image tiling and compression for rapid rendering of images

Source: ESRI ArcGIS Help file
Image “Tiling”

- Split raster into small contiguous rectangles or squares = Tiles
- Display only the tile required upon zooming

Level 0 = 100% of image = 16 low res. tiles
Level 1 = higher res. (parts of 4 med. res. tiles)
Level 2 = highest res. (1+ high res. tiles)
Supported Raster Formats

- See ArcCatalog>Tools>Options
- Each explained in Help
- 24 supported formats
Vector or Raster?

- Spatially continuous data = raster
- Modeling of data with high degree of variability = raster
- Objects with well defined boundaries = vector
- Geographic precision & accuracy = vector
- Topological dependencies = vector or raster
Raster or Vector?

**Raster**
- Simple data structure
- Ease of analytical operation
- Format for scanned or sensed data – easy, cheap data entry

But.......  
- Less compact
- Query-based analysis difficult
- Coarser graphics
- More difficult to transform & project

**Vector**
- Compact data structure
- Efficient topology
- Sharper graphics
- Object-orientation better for some modeling

But...  
- More complex data structure
- Overlay operations computationally intensive
- Not good for data with high degree of spatial variability
- Slow data entry