

# UC Santa Barbara

## Core Curriculum-Geographic Information Systems (1990)

### Title

Unit 04 - The Raster GIS

### Permalink

<https://escholarship.org/uc/item/81k0v7n3>

### Authors

Unit 04, CC in GIS  
Tomlin, Dana

### Publication Date

1990

Peer reviewed

# UNIT 4 - THE RASTER GIS

## UNIT 4 - THE RASTER GIS

Compiled with assistance from Dana Tomlin, The Ohio State University

---

### **For Information that Supplements the Contents of this Unit:**

*[Links to the following resources have been omitted.]*

- [IDRISI Tutorial \(Lorup/Idrisi Project\)](#)
  - [Native American Research Information System \(NARIS\) \(AII/U of Oklahoma\)](#)
  - [Raster View of the World \(Foote and Huebner/Geographer's Craft\) -- Both illustrated and described.](#)
  - [Representation and Data Quality \(Chrisman/U of Washington\)](#)
  - [Scale, Accuracy and Resolution in GIS \(B.C. Environment\) -- Map and display scale; data accuracy, density, detail, resolution and uncertainty; raster data resolution; GIS analysis; separation of data and annotation; etc.](#)
- 

- [A. THE DATA MODEL](#)
- [B. CREATING A RASTER](#)
  - [Cell by cell entry](#)
  - [Digital data](#)
- [C. CELL VALUES](#)
  - [Types of values](#)
  - [One value per cell](#)
- [D. MAP LAYERS](#)
  - [Resolution](#)
  - [Orientation](#)
  - [Zones](#)
  - [Value](#)
  - [Location](#)
- [E. EXAMPLE ANALYSIS USING A RASTER GIS](#)
  - [Objective](#)
  - [Procedure](#)
  - [Result](#)
  - [Operations used](#)
- [REFERENCES](#)
- [EXAM AND DISCUSSION QUESTIONS](#)

- NOTES

Although most of the material in this Curriculum is designed to be as independent as possible from specific data models, it is necessary to deal with this basic concept early so that students can start hands-on exercises with a GIS program. Following Unit 5, we return to the more fundamental concepts and do not address specific vector GIS issues until Units 13 and 14. There are other several places these topics could be placed in a course sequence. We have tried to make Units 4 and 5 as independent as possible so that you can move them within the Curriculum relatively easily.

## UNIT 4 - THE RASTER GIS

Compiled with assistance from Dana Tomlin, The Ohio State University

### A. THE DATA MODEL

- geographical variation in the real world is infinitely complex
  - the closer you look, the more detail you see, almost without limit
- it would take an infinitely large database to capture the real world precisely
  - data must somehow be reduced to a finite and manageable quantity by a process of generalization or abstraction
  - geographical variation must be represented in terms of discrete elements or objects
- the rules used to convert real geographical variation into discrete objects is the data model
  - Tsichritzis and Lochovsky (1977) define a data model as "a set of guidelines for the representation of the logical organization of the data in a database... (consisting) of named logical units of data and the relationships between them."<sup>1</sup>
- current GISs differ according the way in which they organize reality through the data model
- each model tends to fit certain types of data and applications better than others
- the data model chosen for a particular project or application is also influenced by:
  - the software available
  - the training of the key individuals
  - historical precedent
- there are two major choices of data model - raster and vector
- raster model divides the entire study area into a regular grid of cells in specific sequence
  - the conventional sequence is row by row from the top left corner
  - each cell contains a single value \_\_\_\_\_<sup>1</sup>Tsichritzis, T.C., and F.H. Lochovsky, 1977. Data Base Management Systems, Academic Press, New York.
  - is space-filling since every location in the study area corresponds to a cell in the raster
  - one set of cells and associated values is a layer
    - there may be many layers in a database, e.g. soil type, elevation, land use, land cover
- vector model uses discrete line segments or points to identify locations
  - discrete objects (boundaries, streams, cities) are formed by connecting line

- segments
- vector objects do not necessarily fill space, not all locations in space need to be referenced in the model
- a raster model tells what occurs everywhere - at each place in the area
- a vector model tells where everything occurs - gives a location to every object
- conceptually, the raster models are the simplest of the available data models
  - therefore, we begin our examination of GIS data and operations with the raster model and will consider vector models after the fundamental concepts have been introduced.

## B. CREATING A RASTER

- consider laying a grid over a geologic map
  - create a raster by coding each cell with a value that represents the rock type which appears in the majority of that cells areas
  - when finished, every cell will have a coded value
- in most cases the values that are to be assigned to each cell in the raster are written into a file, often coded in ASCII
  - this file can be created manually by using a word processor, database or spreadsheet program or it can be created automatically
  - then it is normally imported into the GIS so that the program can reformat the data for its specific processing needs
- there are several methods for creating raster databases

### Cell by cell entry

- direct entry of each layer cell by cell is simplest
  - entry may be done within the GIS or into an ASCII file for importing
  - each program will have specific requirements
- the process is normally tedious and time-consuming
  - layer can contain millions of cells
  - average Landsat image is around  $7.4 \times 10^6$  pixels, average TM scene is about  $34.9 \times 10^6$  pixels
- run length encoding can be more efficient
  - values often occur in runs across several cells
  - this is a form of spatial autocorrelation - tendency for nearby things to be more similar than distant things
  - data entered as pairs, first run length, then value
    - e.g. the array 0 0 0 1 1 0 0 1 1 1 0 0 1 1 1 0 1 1 1 1 would be entered as 3 0 2 1 2 0 3 1 2 0 3 1 1 0 4 1
    - this is 16 items to enter, instead of 20
    - in this case the saving is 20%, but much higher savings occur in practice
- imagine a database of 10,000,000 cells and a layer which records the county containing each pixel
  - suppose there are only two counties in the area covered by the database
  - each cell can have one of only two values so the runs will be very long
- only some GISs have the capability to use run length encoded files
- note: Units 35 and 36 cover run length encoding and other aspects of raster storage in

more detail

### Digital data

- much raster data is already in digital form, as images, etc.
  - however, resampling will likely be needed in order that pixels coincide in each layer
- because remote sensing generates images, it is easier to interface with a raster GIS than any other type
- elevation data is commonly available in digital raster form from agencies such as the US Geological Survey

### C. CELL VALUES

#### Types of values

- the type of values contained in cells in a raster depend upon both the reality being coded and the GIS
- different systems allow different classes of values, including:
  - whole numbers (integers)
  - real (decimal) values
  - alphabetic values
  - many systems only allow integers, others which allow different types restrict each separate raster layer to a single kind of value
- if systems allow several types of values, e.g. some layers numeric, some non-numeric, they should warn the user against doing unreasonable operations
  - e.g. it is unreasonable to try to multiply the values in a numeric layer with the values in a non-numeric layer
- integer values often act as code numbers, which "point" to names in an associated table or legend
  - e.g. the first example might have the following legend identifying the name of each soil class:
    - 0 = "no class" 1 = "fine sandy loam" 2 = "coarse sand" 3 = "gravel"

#### One value per cell

- each pixel or cell is assumed to have only one value
  - this is often inaccurate - the boundary of two soil types may run across the middle of a pixel
  - in such cases the pixel is given the value of the largest fraction of the cell, or the value of the middle point in the cell
- note, however, a few systems allow a pixel to have multiple values
  - the NARIS system developed at the University of Illinois in the 1970s allowed each pixel to have any number of values and associated percentages

- e.g. 30% a, 30% b, 40% c

#### D. MAP LAYERS

- the data for an area can be visualized as a set of maps of layers
  - a map layer is a set of data describing a single characteristic for each location within a bounded geographic area
- only one item of information is available for each location within a single layer - multiple items of information require multiple layers
  - on the other hand, a topographic map can show multiple items of information for each location, within limits
    - e.g. elevation (contours), counties (boundaries), roads, railroads, urbanized areas (grey tint)
    - these would be 5 layers in a raster GIS
- typical raster databases contain up to a hundred layers
  - each layer (matrix, lattice, raster, array) typically contains hundreds or thousands of cells
- important characteristics of a layer are its resolution, orientation and zone(s)

#### Resolution

- in general, resolution can be defined as the minimum linear dimension of the smallest unit of geographic space for which data are recorded
- in the raster model the smallest units are generally rectangular (occasionally systems have used hexagons or triangles)
- these smallest units are known as cells, pixels
- note: high resolution refers to rasters with small cell dimensions
  - high resolution means lots of detail, lots of cells, large rasters, small cells

#### Orientation

- the angle between true north and the direction defined by the columns of the raster

#### Zones

- each zone of a map layer is a set of contiguous locations that exhibit the same value
- these might be:
  - ownership parcels
  - political units such as counties or nations
  - lakes or islands
  - individual patches of the same soil or vegetation type
- there is considerable confusion over terms here
  - other terms commonly used for this concept are patch, region, polygon
  - each of these terms, however, have different meanings to individual users and different definitions in specific GIS packages
  - in addition, there is a need for a second term which refers to all individual zones that have the same characteristics
  - class is often used for this concept

note that not all map layers will have zones, cell contents may vary continuously over the region making every cell's value unique

- e.g. satellite sensors record a separate value for reflection from each cell
- major components of a zone are its value and location(s)

### Value

- is the item of information stored in a layer for each pixel or cell
- cells in the same zone have the same value

### Location

- generally location is identified by an ordered pair of coordinates (row and column numbers) that unambiguously identify the location of each unit of geographic space in the raster (cell, pixel, grid cell)
- usually the true geographic location of one or more of the corners of the raster is also known

## E. EXAMPLE ANALYSIS USING A RASTER GIS

### Objective

- identify areas suitable for logging
- an area is suitable if it satisfies the following criteria:
  - is Jackpine (Black Spruce are not valuable)
  - is well drained (poorly drained and waterlogged terrain cannot support equipment, logging causes unacceptable environmental damage)
  - is not within 500 m of a lake or watercourse (erosion may cause deterioration of water quality)

### Procedure

- recode layer 2 as follows, creating layer 4
  - y if value 2 (Jackpine)
  - n if other value
- recode layer 3 as follows, creating layer 5
  - y if value 2 (good)
  - n if other value
- spread the lake on layer 1 by one cell (500 m), creating layer 6
- recode the spread lake on layer 6 as follows, creating layer 7
  - n if in spread lake
  - y if not
- overlay layers 4 and 5 to obtain layer 8, coding as follows
  - y if both 4 and 5 are y
  - n otherwise
- overlay layers 7 and 8 to obtain layer 9, coding as follows
  - y if both 7 and 8 are y
  - n otherwise

## Result

- the loggable cells are y on layer 9

## Operations used

- recode
- overlay
- spread
- we could have achieved the same result using the operations in other sequences, or by combining recode and overlay operations
  - e.g. overlay layers 2 and 3, coding as follows
    - y if layer 2 is 2 and layer 3 is 2, n otherwise
    - this would replace two recodes and an overlay
  - e.g. some systems allow layers to be overlaid 3 or more at a time
- the names given to operations vary from system to system, but most of the operations themselves are common across systems

## REFERENCES

Star, J.L. and J.E. Estes, 1990. Geographic Information Systems: An Introduction, Prentice Hall, Englewood Cliffs, NJ. An introduction to GIS with a strong raster orientation.

Further references can be found following Unit 5.

## EXAM AND DISCUSSION QUESTIONS

1. What types of geographical data fit the raster GIS data model best? What types fit worst?
2. Review the issues involved in selecting a resolution for a raster GIS project.
3. What resolutions would be appropriate for the following problems: (a) determining logging areas in a National Forest, (b) finding suitable locations for backcountry campsites, (c) planning subdivisions to take account of noise from an airport?
4. Review the methods of planning described in Ian McHarg's classic book Design with Nature (1969, Doubleday, New York). In what ways would they (a) benefit and (b) suffer from implementation using raster GIS?
5. Using the documentation for the raster GIS program you have, determine how that program uses (a) the concept of "zone" as a contiguous group of cells of the same value, and (b) the concept of several groups of cells that all have the same value. Is there any ambiguity in the way your program deals with these two concepts?

---

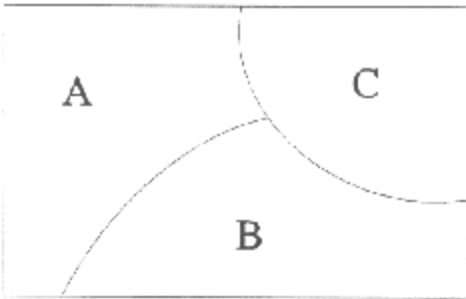
*Last Updated: August 30, 1997.*



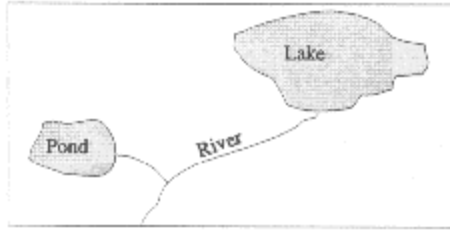
# UNIT 4 IMAGES

A	A	A	A	C	C	C	C
A	A	A	A	C	C	C	C
A	A	A	B	B	C	C	C
A	A	B	B	B	B	B	B
A	B	B	B	B	B	B	B

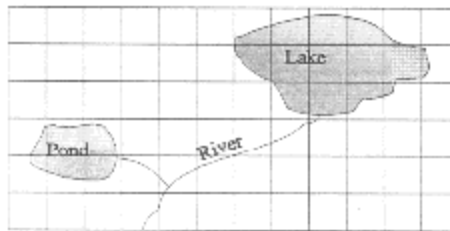
Raster



Vector



Reality - Hydrography



Reality overlaid with a grid

0	0	0	0	0	0	0	1	1	1	0	0
0	0	0	0	0	0	1	1	1	1	1	0
0	0	0	0	0	0	0	1	1	1	0	0
0	1	1	0	0	0	2	2	0	0	0	0
0	1	1	2	2	2	0	0	0	0	0	0
0	0	0	2	0	0	0	0	0	0	0	0

0 = No Water Feature  
 1 = Water Body  
 2 = River

Resulting raster

**ROW BY ROW**

1	1	1	2	2
1	1	1	1	2
1	1	1	1	1
3	3	1	1	1
3	3	3	1	1

**EACH VALUE ON AN INDIVIDUAL LINE**

1
1
1
2
2
1
1
1
1
2
1
1
1
1
1
3
etc.

**VALUES ENTERED CONTINUOUSLY**

1	1	1	2	2	1	1	1	1	2	1	1
1	1	1	3	3	1	1	1	3	3	3	1

0	0	1	1	1	2	3	INTEGER VALUES
0	1	1	1	2	2	3	e.g. soil class
0	0	1	1	1	2	2	or # of farms
0	0	0	1	1	1	2	
0	0	0	0	0	0	1	

2.1	2.3	2.4	2.7	2.9	3.2	3.4	REAL (DECIMAL) VALUES
1.9	2.1	2.4	2.6	3.0	3.3	3.4	e.g. elevation
1.8	2.0	2.3	2.5	2.8	3.1	3.2	
1.8	1.9	1.9	2.3	2.5	2.8	3.0	
1.8	1.8	1.8	1.9	2.1	2.3	2.7	

a	a	b	b	c	c	a	NON-NUMERIC VALUES
a	b	b	b	c	c	a	e.g. vegetation class
a	b	b	b	b	c	c	
a	a	b	b	b	c	c	
a	a	a	b	c	c	c	

## Raster Data Values

## Raster description

- resolution is 500 m
- entire array is 2.5 km by 2.5 km
- columns are aligned North-South

## LAYER 1: LAKE

1 = lake, 0 = not lake

```

1 1 0 0 0
1 1 0 0 0
1 0 0 0 0
0 0 0 0 0
0 0 0 0 0

```

## LAYER 2: FOREST SPECIES

1 = Black Spruce, 2 = Jackpine, 0 = no forest

```

0 0 2 2 2
0 0 2 2 2
0 1 1 1 2
1 1 1 1 1
1 1 1 1 1

```

## LAYER 3: SOIL DRAINAGE

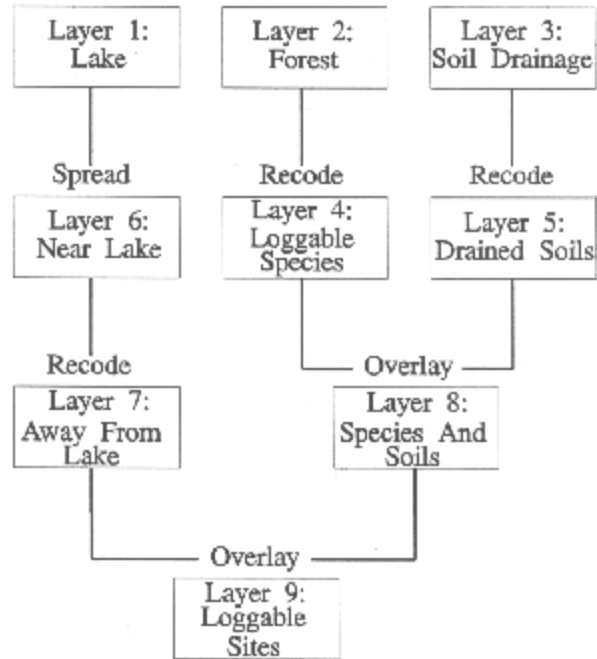
0 = waterlogged, 1 = poor, 2 = good

```

0 0 2 2 2
0 0 1 1 2
0 0 1 1 2
0 1 1 1 2
1 1 1 2 2

```

## Example Raster Database



### Example Project Steps

**OBJECTIVE:** Identify areas suitable for logging

An area is suitable if it satisfies the following criteria:

- is Jackpine
  - Black Spruce are not valuable
- is well drained
  - poorly drained and waterlogged terrain cannot support equipment, logging causes unacceptable environmental damage
- is not within 500 m of a lake or watercourse
  - erosion hazard

**PROCEDURE:**

Recode layer 2 - y if value 2 (Jackpine), n if other value

LAYER 2: FOREST SPECIES					LAYER 4: LOGGABLE SPECIES				
0	0	2	2	2	n	n	y	y	y
0	0	2	2	2	n	n	y	y	y
0	1	1	1	2	n	n	n	y	y
1	1	1	1	1	n	n	n	n	n
1	1	1	1	1	n	n	n	n	n

Recode layer 3 - y if value 2 (good), n if other value

LAYER 3: SOIL DRAINAGE					LAYER 5: DRAINED SOIL				
0	0	2	2	2	n	n	y	y	y
0	0	1	1	2	n	n	n	n	y
0	0	1	1	2	n	n	n	n	y
0	1	1	1	2	n	n	n	n	y
1	1	1	2	2	n	n	n	y	y

## Example Project Details

(Page 1 of 3)

Spread the lake on layer 1 by one cell (500 m)

LAYER 1: LAKE

```

1 1 0 0 0
1 1 0 0 0
1 0 0 0 0
0 0 0 0 0
0 0 0 0 0

```

LAYER 6: NEAR LAKE

```

1 1 1 0 0
1 1 1 0 0
1 1 0 0 0
1 0 0 0 0
0 0 0 0 0

```

Recode layer 6 - if in spread lake, y if not

LAYER 6: NEAR LAKE

```

1 1 1 0 0
1 1 1 0 0
1 1 0 0 0
1 0 0 0 0
0 0 0 0 0

```

LAYER 7: AWAY FROM LAKE

```

n n n y y
n n n y y
n n y y y
n y y y y
y y y y y

```

Overlay layers 4 and 5 - if both 4 and 5 are y, n otherwise

LAYER 4: LOGGABLE SPECIES

```

n n y y y
n n y y y
n n n n y
n n n n n
n n n n n

```

LAYER 5: DRAINED SOIL

```

n n y y y
n n n n y
n n n n y
n n n n y
n n n y y

```

LAYER 8: SPECIES AND SOILS

```

n n y y y
n n n n y
n n n n y
n n n n n
n n n n n

```

## Example Project Details

(Page 2 of 3)



Overlay layers 7 and 8 - if both 7 and 8 are y, n otherwise

**LAYER 7: AWAY FROM LAKE**

```
n n n y y
n n n y y
n n y y y
n y y y y
y y y y y
```

**LAYER 8: SPECIES & SOILS**

```
n n y y y
n n n n y
n n n n y
n n n n n
n n n n n
```

**LAYER 9: LOGGABLE SITES**

```
n n n y y
n n n n y
n n n n y
n n n n n
n n n n n
```

**RESULT:** The loggable cells are y on layer 9

## Example Project Details

(Page 3 of 3)