

# UC Santa Barbara

## Core Curriculum-Geographic Information Systems (1990)

### Title

Unit 21 - The Raster/Vector Database Debate

### Permalink

<https://escholarship.org/uc/item/2rq7m7xm>

### Authors

Unit 21, CC in GIS  
Parson, Charles

### Publication Date

1990

Peer reviewed

# UNIT 21 - THE RASTER/VECTOR DATABASE DEBATE

## UNIT 21 - THE RASTER/VECTOR DATABASE DEBATE

Compiled with assistance from Charles Parson, Bemidji State University

- [A. INTRODUCTION](#)
  - [Raster or vector](#)
  - [Basic issues](#)
- [B. COORDINATE PRECISION](#)
  - [Raster precision](#)
  - [Location of coordinates in raster](#)
  - [Vector precision](#)
  - [Data precision](#)
- [C. SPEED OF COMPUTING](#)
- [D. MASS STORAGE](#)
  - [Raster storage](#)
  - [Vector storage](#)
- [E. CHARACTERISTICS OF PHENOMENA](#)
  - [Raster sampling](#)
  - [Vector sampling](#)
  - [Features, entities and objects](#)
- [F. SUMMARY](#)
  - [Combinations of raster and vector](#)
- [REFERENCES](#)
- [EXAM AND DISCUSSION QUESTIONS](#)
- NOTES

This unit specifically addresses the issue of raster versus vector by looking at the different dimensions of the debate. We have tried to stress the importance and value of both models by comparing specific technical aspects. The next unit looks at a related but more abstract debate.

## UNIT 21 - THE RASTER/VECTOR DATABASE DEBATE

Compiled with assistance from Charles Parson, Bemidji State University

### A. INTRODUCTION

#### Raster or vector

- Unit 4 introduces definitions of raster and vector
  - arguments about which was better have been commonplace since the earliest systems were created
  - raster databases are appealing
    - simplicity of organization
    - speed of many operations, e.g. overlay, buffers
    - especially appealing to the remote sensing community who are used to "pixel" processing
  - on the other hand, there are many situations in which the raster approach may appear to sacrifice too much detail
    - cartographers were appalled by the crude outlines of parcels that resulted in the "pinking shear" effect of diagonal boundaries represented by grid cell edges
- diagram
- surveyors were dismayed by the "inaccuracy" caused by the cells when portraying linear features and points
  - situations in which the raster approach sacrificed too much detail
- however, computing times for overlaying vector based information can be excessive
    - early polygon overlay routines were error-prone, expensive, slow
  - today, there are situations in which it is clear that one approach is more functional than the other
    - e.g. using "friction" layer to control width of buffer is only feasible in raster
    - e.g. viewshed algorithms to find area visible from a point are feasible with elevation grids (raster DEMs), not with digitized contours
    - e.g. land survey data can only be represented with precise lines
  - an important current trend involves linking raster and vector systems, displaying vector data overlying a raster base
    - raster data may be from a GIS file (perhaps a remotely sensed image) or from a plain scanned image file
  - therefore, the question has evolved from "Which is best?" to "Under what conditions is which best and how can we have flexibility to use the most appropriate approaches on a case by case basis?"

## Basic issues

- four issues to the discussions of raster versus vector:
  - coordinate precision
  - speed of analytical processing
  - mass storage requirements
  - characteristics of phenomena

## B. COORDINATE PRECISION

### Raster precision

- e.g. MLMIS (Minnesota Land Management Information System, see Unit 9)
- early version used cell sizes of 40 acres
- depended on a rectilinear public land survey
  - system "created" a state of perfectly square townships, composed of sections that were exactly one mile on each side
  - variability in the original survey lines were only addressed if mis-alignment was more than an eighth of a mile
- locational precision was limited by size of cells - 40 acres, or 1/4 mile on each side
  - all linear features had to be represented as 1/4 mile wide strips
  - point features "occupied" 40 acres on the map
- at scales smaller than 1:250,000 these conditions were not difficult to accept
  - for state-wide planning purposes, the database was adequate
- what about using smaller cells
  - five meter cells have been scanned into systems
    - this size is chosen on the basis that 5 m is the width of a #00 pen line on a 1:24,000 map
  - of course, still not possible to represent objects smaller than 5 m
    - e.g. fire hydrants, storm sewer grates, power poles
    - precision not adequate for facilities managers
  - on the other hand, at 5 m there is no appreciable loss of information for most natural phenomena, most occur at this scale or smaller

### Location of coordinates in raster

- in most cases it is unclear whether the center of the cell or one of its corners is the precise location of the coordinate
  - e.g. top left corner of the grid may be referenced to a specific UTM coordinate, but it is unclear whether that location is in the middle of the top left cell or at the top left corner of that cell
  - locational precision is thus 1/2 the cell's width and height

### Vector precision

- can be encoded with any conceivable degree of precision
  - precision is limited by the method of internal representation of coordinates
  - typically 8 or 16 decimal digits are used ("single" or "double" precision)
  - this limits precision to 1/108 or 1/1016 of the size of the study area respectively
  - for equivalent raster precision we would need 108 by 108, or 1016 by 1016 cells respectively, neither of which is feasible, even with run length encoding
- however this argument may be artificial
  - real vector data accuracy may be much worse than one line width
  - e.g. digitizing from a 1:24,000 quadrangle map may appear to allow points to be recorded to the nearest 2 m, from a map that has common errors of 12 m

### Data precision

- vector precision is true for certain classes of data
  - data captured from precision survey (Coordinate Geometry - COGO)
  - plat maps created from land surveyors' coordinates
  - political boundaries defined by accurate survey
- few natural phenomena have true edges which can be accurately represented as mathematical lines
  - soils, vegetation types, slopes, wildlife habitats, all have fuzzy boundaries
  - due to the methods used to record the spatial information
  - due to the transitional nature of variation in the phenomenon
- it can be argued that the fine lines from the vector system gave a false sense of precision
  - lines on maps are typically 0.5 mm wide and are often assumed to represent the uncertainty in the location of the object
  - in a raster system uncertainty is automatically reflected in the cell size
- true comparison in terms of precision is between raster cell size and the positional uncertainty of a vector object, not the coordinate precision

### C. SPEED OF COMPUTING

- raster data can be processed very quickly to answer most analytical questions involving overlays, proximity, and

Boolean queries, no calculations are required to determine relative positions between layers

- in most cases, analysis requires cell-by-cell comparison of the contents of layers
- little arithmetic computation is required beyond simple conditional statements
- thus raster systems are suitable for small computers and have low program development costs
- the same questions require considerable computation in vector systems

- vector topology is complex
- complex geometrical problems must be solved, e.g. to find line intersection points
- complex geometrical algorithms are needed in polygon overlay to avoid generating spurious polygons
- calculation of distances may be complex depending on the projection/coordinate system used
- these requirements slow the response times of vector based systems
  - may prevent large scale work, or multiple users on small computers
- on larger machines, the speed factor will still be evident:
  - may limit the number of concurrent users
  - may require that complex work to be done in batch mode at non-peak times

## D. MASS STORAGE

### Raster storage

- simplest raster data storage method requires one memory location (e.g. one or two bytes) per cell
  - this is not at all efficient, but is used by several systems
  - such systems severely limit the maximum numbers of rows and columns that can be used
- file compression is possible through a variety of approaches - see Unit 35
  - most common are forms of run length encoding
  - degree of compression depends on spatial variability of data
  - for very complex data the benefit of run length encoding can be negative - use of run length encoding should be optional
  - there is a small overhead in packing and unpacking data compared to cell-by-cell storage

### Vector storage

- use very little storage for simple polygons
  - memory requirements depend on complexity of objects
  - also on precision of coordinates (i.e. single or double)
- volume also depends on which relationships between objects are stored in the database
  - some systems store few relationships, require small amount of storage, compute other relationships as needed
  - other systems offer more elaborate database models, store more relationships, require larger amounts of storage
  - e.g. system A required 150 Kbytes to store 700 lots, system B required 5 Mbytes to store the same 700 lots, both using vector database models
- generally, vector systems should use less mass storage than a raster based system of high enough resolution to emulate the vectors
  - assuming that the required resolution is defined by the line width on the input

document, rather than by the width of the transition zone in reality

## E. CHARACTERISTICS OF PHENOMENA

### Raster sampling

- raster is a regularly spaced sampling of phenomena
  - reflects lack of knowledge of spatial variation
  - if we knew where the complex variation occurred, we would sample there more heavily - not wasting samples in areas of little variation
  - e.g. knowing that population in California is concentrated in Los Angeles basin and San Francisco Bay area, we would collect more data in those areas than in the Mojave Desert
  - raster is appropriate for remote sensing as the satellite is not intelligent enough to vary its sampling in response to variation on the earth's surface
- data typically collected in raster format is satellite imagery (raw and classified) and elevation data

### Vector sampling

- vector representation permits more spatial variability in some areas than in others
  - e.g. rapid variation at edge of area objects, none in the middle
  - e.g. census tracts are small in urban areas and large in rural
- this is appropriate for social, economic, demographic variation which is much more intense in some areas than in others
- also appropriate for some natural phenomena
  - e.g. variation in vegetation cover is more rapid near the Nile than in the Sahara
  - e.g. variation in geology is instantaneous across a fault
- some objects are vector by definition
  - variation in ownership is instantaneous at edge of lot
  - variation in county is instantaneous at boundary
- data typically collected in vector format are coordinate geometry (surveyor's records) and legal boundaries

### Features, entities and objects

- difficult to group cells together as an object with attributes in raster
  - e.g. connect cells along a road
  - e.g. connect cells as a numbered forest stand
- raster "sees" the world as populated by cells of uniform size
  - raster arranges geography in fixed sequence - gives "sequential access" to world

vector "sees" the world as populated by entities, represented in the database model as objects

- vector arranges geography in any sequence - gives "random access" to data
- operations on objects are easier in vector, e.g. analysis on a network - routing vehicles through a road network
- a point object must occupy a full cell in raster, this creates some problems:
  - locations of water wells, emergency call boxes should not be indicated by raster cells, of some arbitrary cell size, in which they lie somewhere
  - for environmental modelling of water quality, we may need to know precisely where the wells are
  - to estimate costs of connecting wires to call boxes, we may need to be able to determine distances accurately
- on the other hand, some data can only be presented in aggregated form:
  - e.g. census data is typically aggregated by census tract, both vector and raster representations may be appropriate depending on the application
  - e.g. sensitive natural resource-related data, such as the location of rare plants or archaeological sites, may be presented in large cells to allow preservation of the phenomena by indicating its presence without identifying its location precisely

## F. SUMMARY

- the raster/vector debate can be summarized as a series of decision rules, for example: handout - Recommendations for the use of vector and raster structures

## Combinations of raster and vector

- is the best of both worlds available?
  - to an extent it is, in two ways
- 1. can store data in one form and process it in another
  - needs an efficient algorithm to convert from raster to vector and vice versa
  - possible to capture and store data in vector mode, yet analyze it in raster form
  - may save computing time and mass storage
  - especially important for small machines
- 2. use combinations of systems which run raster and vector analytical systems in parallel
  - e.g. install raster and vector systems on the same PC, use conversion functions in one or both systems
  - e.g. overlay a vector based landuse parcel map over a Landsat image to improve the interpretation of the satellite image
  - the image might then be used to correct a vector based vegetation parcel map

## REFERENCES

Burrough, P.A., 1986, Principles of Geographical Information Systems for Land Resources



Assessment, Clarendon, Oxford. See raster/vector summary on p. 169.

Gahegan, M.N., and S.A. Roberts, 1988. "An intelligent object-oriented geographical information system," *International Journal of Geographical Information Systems* 2(2):101-110. Discusses an interface to a spatial analysis systems which allows the underlying geographical domain to be represented using a high-level, feature- oriented model.

Star, J.L. and J.E. Estes, 1990. *Geographic Information Systems: An Introduction*. Prentice Hall, Englewood Cliffs NJ. Chapter 4 summarizes both sides of the raster-vector issue.

### EXAM AND DISCUSSION QUESTIONS

1. Summarize the dimensions of the debate between raster and vector database models.
2. Describe the design of a study to compare the amounts of storage used by raster and vector models in creating a digital representation of a specific map? If you feel this cannot be done, explain why not.
3. Summarize the arguments for raster GIS and the application areas in which it has distinct (a) advantages and (b) disadvantages.
4. What factors would determine the appropriate cell size for a raster GIS being developed for a power transmission corridor study?
5. In physics, light appears to behave sometimes as particles, sometimes as waves. Discuss whether a useful analogy exists between this debate in physics and the raster/vector debate in GIS.

---

*Last Updated: August 30, 1997.*