3 Representing Geography

OVERVIEW

- This chapter introduces the concept of representation, or the construction of a digital model of some aspect of the Earth's surface.

- The geographic world is extremely complex, revealing more detail the closer one looks. So in order to build a representation of any part of it, it is necessary to make choices, about what to represent, at what level of detail, and over what time period.

- Generalization methods are used to remove detail that is unnecessary for an application, in order to reduce data volume and speed up operations.

LEARNING OBJECTIVES

After reading this chapter, your students will be able to explain:

- **The importance of understanding representation in GIS;**

- **The concepts of fields and objects and their fundamental significance;**

- **What raster and vector representation entails and how these data structures affect many GIS principles, techniques, and applications;**

- **The paper map and its role as a GIS product and data source;**

- **The importance of generalization methods and the concept of representational scale;**

- **The art and science of representing real-world phenomena in GIS.**

Digital, binary, representation, Tobler's First Law of Geography, attributes, the fundamental problem (the world is infinitely complex), discrete objects, attribute tables, continuous fields, raster, vector, pixel, polygon, polylines, TIN, representative fraction, generalization, simplification, weeding.

OUTLINE

CHAPTER SUMMARY

3.1 Introduction

- This section mentions the notions of
  - *spatial concepts* such as containment and proximity
  - *data models*, the structures and rules that are programmed into a GIS to accommodate data
  - *ontologies*, the frameworks that we use for acquiring knowledge of the world
- Figure 3.1 is a time-space *representation* of the daily journeys of several people
- Representations are reinforced by the rules and laws that we humans have learned to apply to the unobserved world around us, such as using spatial interpolation to guess the conditions that exist in places where no observations were made.
- Introduces *Tobler's First Law of Geography*: Everything is related to everything else, but near things are more related than distant things.

3.2 Digital representation

- This section introduces the notions of *digital* and *binary.*

- Technical Box 3.1 outlines fundamentals about the binary number system, including short (16-bit) and long (32-bit) storage, ASCII, floating point numbers, and BLOBs (binary large object).
- Benefits of digital representations of geography include: can be handled in ways that are independent of meaning; easy to copy and transmit; stored at high density; easy to transform, process, and analyze.

3.3 Representation for what and for whom?

- *Geographic* representation is defined as a representation of some part of the Earth's surface or near surface, at scales ranging from the architectural to the global
- Short history of maps and the significance of their use historically
- Like maps, any application of GIS requires attention to questions of *what* should be represented, and *how*.

3.4 The fundamental problem

- Geographic data are built up from atomic elements, or facts about the geographic world.
- At its most primitive, an atom of geographic data (strictly, a datum) links a *place*, often a *time*, and some descriptive *property*.
- *Attribute* refers to this descriptive property and may:
  - vary slowly or rapidly
  - be physical/environmental or social/economic
  - identify a place or entity
  - provide a measurement of something at that place
- Technical Box 3.3 defines the types of attributes: nominal, ordinal, interval, ratio, and cyclic
- The *fundamental problem* is "the world is infinitely complex, but computer systems are finite".

3.5 Discrete objects and continuous fields

3.5.1 Discrete objects

- In the *discrete object* view, the world is empty, except where it is occupied by objects with well-defined boundaries that are instances of generally recognized categories.
  - Objects can be counted

- Objects have dimensionality: 0-dimension (points), 1-dimension (lines), 2-dimensions (areas, polygons)
- 3-dimensions are problematic in GIS
- Associating attributes to discrete objects can be expressed in a table (see Table 3.1):
    - each row corresponding to a different discrete object,
    - each column corresponding to an attribute of the object.
- A table does not look like the real world

3.5.2 Continuous fields

- The continuous field view represents the real world as a finite number of variables, each one defined at every possible position.
- Continuous fields, on the other hand, can be distinguished by what varies, and how smoothly.
- The challenge of counting the lakes in Minnesota provides an elegant illustration of various ways to think about the conceptual transition from objects to fields
- Technical Box 3.4 explains the concept of 2.5 dimensions

3.6 Rasters and vectors

- raster and vector are two methods that are used to reduce geographic phenomena to forms that can be coded in computer databases
- In principle, each can be used to code both fields and discrete objects, but in practice there is a strong association between raster and fields, and between vector and discrete objects.

3.6.1 Raster data

- In a raster representation geographic space is divided into an array of cells, each of which is usually square, but sometimes rectangular (Figure 3.9).
- All geographic variation is then expressed by assigning properties or attributes to these cells.
- The cells are sometimes called pixels (short for *picture elements*).
- A few fundamentals of remote sensing are introduced
- When information is represented in raster form all detail about variation within cells is lost, and instead the cell is given a single value based on a rule such as largest share or value at the central point.

3.6.2 Vector data

- Usually, in a vector representation, all lines are captured as points connected by precisely straight lines (some GIS software also allows curves).
- Features are captured as a series of points or *vertices* connected by straight lines
    - areas are often called polygons
    - lines are sometimes called polylines
    - the choice between raster and vector is often complex, as summarized in Table 3.3.

3.6.3 Representing continuous fields

The six approximate representations of a field used in GIS are (Figure 3.12)

- Regularly spaced sample points
- Irregularly spaced sample points.
- Rectangular cells.
- Irregularly shaped polygons.
- Irregular network of triangles
- Polylines representing contours

3.7 The paper map

- Key property of a paper map is its *scale* or *representative fraction*, defined as the ratio of distance on the map to distance on the Earth's surface.
    - when a scale is quoted for a digital database it is usually the scale of the map that formed the source of the data.
    - while the paper map is a useful metaphor for the contents of a geographic database, we must be careful not to let it limit our thinking about what is possible in the way of representation.
- Digital representations can include information that would be very difficult to show on maps.
- Digital databases can represent all three spatial dimensions, including the vertical, whereas maps must always show two-dimensional views.

3.8 Generalization

- ways of simplifying the view of the world include:
    - describe entire areas, attributing uniform characteristics to them, even when areas are not strictly uniform;

- identify features on the ground and describe their characteristics, again assuming them to be uniform;
- limit our descriptions to what exists at a finite number of sample points, hoping that these samples will be adequately representative of the whole.
- some degree of generalization is almost inevitable in all geographic data.
- A map's specification defines how real features on the ground are selected for inclusion on the map.

3.8.1 Generalization about places

- Geometry generalization methods, see Figure 3.14
- Attribute generalization methods include
  - *Simplification*
  - *Smoothing*
  - *Aggregation*
  - *Amalgamation*
  - *Merging*
  - *Collapse*
  - *Refinement*
  - *Exaggeration*
  - *Enhancement*
  - *Displacement*

- Weeding is the process of simplifying a line or area by reducing the number of points in its representation
- The operation of the Douglas-Poiker weeding algorithm is shown in Figure 3.16.

3.8.2 Generalization about properties

- Many representations in GIS bring together multiple properties of places as composite indicators of conditions at particular locations
- Representation may not be fit for its purpose if it is based upon inappropriate indicators of real-world conditions

3.8.3 Representation using VGI

Generalizability of representations based on VGI is questioned as follows:

- Are all places equally accessible to volunteers?
- Are volunteers more likely to provide data on places or properties that interest them than those that do not?

- Where volunteers supply information about themselves, is this representative of other non-volunteers?
- Are volunteers at liberty to collect locationally sensitive data?
- Should volunteers supply data that are open to malevolent use?
- Is it socially acceptable to make available any observations of uniquely identifiable individuals?
- Is the information date-stamped, as an indicator of its provenance and current reliability?

ESSAY TOPICS

Again, these are in rough order of sophistication of necessary response:

1. What do you understand by the terms raster and vector? How would you decide which to use in any specific given project? Describe the situations encountered in this chapter where the distinction between raster and vector becomes blurred.

2. Why is digital map generalization necessary and why does it involve so much more than line simplification?

3. GIS professionals and cartographers represent and understand 'geography' in certain ways. Compare and contrast with your own everyday experience of spaces and places.

4. Distinguish between nominal, ordinal, interval, ratio, and cyclic data types, and explain how this classification does not cover all the attribute types that might be recorded in a GIS

5. Table 3.3 summarized some of the arguments between raster and vector representations. Expand on these arguments, providing examples, and add any others that would be relevant in a GIS application.

6. "No representation of geographic phenomena can ever be perfect". Is this true, are there exceptions, and what implications does this statement have for users of GIS?

7. A map is defined by the International Cartographic Association as 'A conventionalized image representing selected features or characteristics of geographic reality designed for use when spatial relationships are of primary relevance' (Board, 1991). Explain the reasoning behind every element of this definition.

8. An argument has been made that representing geography in a GIS can be improved by understanding human perception and cognition. Review and discuss this argument. In what ways do you agree or disagree with this idea?

9. To what extent do you agree with the two propositions expressed by Goodchild in 1995 that 'It is GIS's supreme conceit that one can structure a useful representation of