Representing the Real World

• The **theory** of representing the real world in a GIS using digital data
  • The **nature** of digital data and binary notation
  • The **discrete object view** of the world
    • Entities, data objects, and data models
    • Topology and spatial relationships
  • The **field view** of the world
Representing the Real World

To map is to **transform** information from one form to another --- Mathematics

Earth surface \(\rightarrow\) map \(\rightarrow\) Paper

--- Geography
Representing the Real World

• Maps are a means by which to convey a lot of **spatial information** in a succinct package, because they allow us to gain knowledge of places we may never have seen ourselves.

• Because we get to observe a very small part of the world with our own eyes, we rely on a variety of methods of **representation** to convey information:

  • Photographs, spoken descriptions and written text (along with maps) are all familiar kinds of representation we use in everyday life to help us **assemble more knowledge** about the Earth.
Representing the Real World

• In a GIS, the representation of real world phenomena makes use of digital data formats.
Digital Representations

• You probably have a fairly broad experience of digital representations through the use of computers and the Internet for other (non-geographic) purposes.

• It is easy to be completely unaware of how digital representations function because technology has advanced sufficiently that their functioning is quite transparent to the consumer of information.

  • e.g. when you browse a web page, you see the formatted hypertext, not the html source code, nor the fundamental binary storage of the data.
Digital Representations

• The term **digital** derives from **digits** (which in turn is a reference to our fingers) and our system of counting based on the 10 digits of our hands.

• While our usual counting system makes use of the 10 symbols in the decimal counting system (0 – 9), digital computers use a system of counting that only uses 2 symbols that we represent with 0 and 1.

• Computers represent phenomena as binary digits. Every item of useful information about the Earth’s surface is ultimately reduced by a GIS to some combination of 0s and 1s (p. 61 of text).
**Binary Notation**

- Everything is represented as 0s and 1s in a computer. These two-state forms correspond to yes/no, on/off, open/closed.

<table>
<thead>
<tr>
<th>Binary</th>
<th>Decimal</th>
<th>One to one correspondence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 digit</td>
<td>0, 1</td>
<td>1 bit 0,1,2,…9</td>
</tr>
<tr>
<td>2 digits</td>
<td>00, 01</td>
<td>2 bits 00, 01,…</td>
</tr>
<tr>
<td></td>
<td>10, 11</td>
<td>97, 99</td>
</tr>
<tr>
<td>3 digits</td>
<td>000, 001</td>
<td>3 bits 000, 001,</td>
</tr>
<tr>
<td></td>
<td>010, 011</td>
<td>002, 003,</td>
</tr>
<tr>
<td></td>
<td>100, 101</td>
<td>…</td>
</tr>
<tr>
<td></td>
<td>110, 111</td>
<td>998, 999</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Binary Notation

Decimal:
72,479 = 70,000 = 7 × 10^4
2,000 = 2 × 10^3
400 = 4 × 10^2
70 = 7 × 10^1
9 = 9 × 10^0

Note: In binary
1010 + 110
= 10000

Binary:

<table>
<thead>
<tr>
<th>2^4</th>
<th>2^3</th>
<th>2^2</th>
<th>2^1</th>
<th>2^0</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

1 × 2^4 + 0 × 2^3 + 1 × 2^2 + 0 × 2^1 + 0 × 2^0
= 16 + 0 + 4 + 0 + 0
= 20
Electronic Representation of Bits

• Binary information can be represented by electrical pulses passing along wires with two different voltage levels as shown in the following sequence of pulses:
Bits and Bytes

1 bit = 1 binary digit
1 byte = 8 bits

8 bits = 1 byte

1024 bytes = 1 Kb
1024 Kb = 1 Mb
1024 Mb = 1 Gb
1024 Gb = 1 Tb
1024 Tb = 1 Pb
ASCII Encoding

• If computers store everything using 0s and 1s, then how are characters represented?

• The ASCII (American Standard Code for Information Interchange) code assigns the numbers 0 through 127 to 128 characters, including upper and lower case alphabets plus various special characters, such as white space etc.

• e.g. decimal 85 is assigned to represent upper case U. In binary, 01010101 = 85. Thus the computer represents U using 01010101.

• Files which contain information encoded in ASCII are easily transferred and processed by different computers and programs. These are called “ASCII” or “text” files.
You now have some familiarity with how a digital representation works fundamentally. But now we encounter some key issues:

- What do we represent?
- What model should we use?
Representing the Real World

• We can think of geographic data as consisting of individual facts about the geographic world that link a place, a time, and some descriptive property, which we usually refer to as an attribute.

• A full and complete representation of the Earth would consist of an infinite number of geographic facts because

The world is infinitely complex, but computer systems are finite. Representation is all about the choices that are made in capturing knowledge about the world (p. 65 of the text)
Two Fundamental Sorts of Representations

• Regardless of what phenomena of interest we chose to include in our geographic representation, we first must choose between a model that either
  • represents geography as **discrete objects**
  OR
  • represents geography as **fields**
• Most GIS approaches focus on the discrete object view, and we’ll look at it in greater detail, although we will look at field representations as well
The Discrete Object View

- This view represents the world as objects with well-defined boundaries in empty space.
- We can refer to these objects as entities -- real world phenomena that can be represented in a database.
Entities → Data Objects

• In the discrete object view, we associate each and every entity we wish to represent with its own data object – a digital “building block” used in a data structure to represent an entity.
Collection of Data Objects → Data Model

• Taken all together, a set of data objects comprise a **data model** – a collection of data objects represented in a particular format
The Discrete Object View

• A **data model** is composed of **data objects**, each of which represents an **entity**

• Displayed schematically:
Data Objects and their Attributes

• Each data object has associated attribute data, comprised of one or more variables, that describes the entity (and we have selected which attributes because they are descriptive of our phenomena of interest)
Data Models

• We can consider two layers of data models:
  • **Conceptual data models** – these include sets of components and their relationships to each other, and are *independent* of a specific system
    These exist in your mind
  • **Logical data models** – these include the logical organization of the components of a data model in which the relationships among components are *explicitly* defined
    These exist in the GIS
Topology

• A *key component* of defining the rules of a spatial logical data model is defining the topological relations between objects

• **Topology** describes the relationship between objects, or entities, in space

• Do not confuse topology with topography, which describes the third dimension of relief on the Earth’s surface (elevation)
Conceptualizing Topology

- Topology defines certain **spatial relationships** between objects:
  - Adjacency
  - Connectivity
  - Containment
- With topology defined, it becomes a simple matter to assess other spatial relationships such as:
  - Direction
  - Proximity
• Supposing we had a data model that included towns as represented by polygons

• If two town polygons were next to each other, we could describe them as topologically adjacent
• We might define the connectivity of the polygons in our model as well (perhaps based on some attributes, or the nature of the shared boundary …)
• Our data model might also include point data that denotes the locations of businesses within the town polygons

• Using these we might assess whether a certain business was within a certain town using the containment relation.
• We can also the relative position of data objects in terms of direction:
  • Moe’s is North-East of the Kwik-E-Mart
  • The nuclear plant is South-East of the Kwik-E-Mart
Topology - Proximity

• We can also calculate the distances between objects to determine their proximity:
• Homer lives near Ned
• Homer lives far from Grandpa
(where near and far can be defined …)
Why We Need Topology

• Without topology, all you have is a “pretty picture”
• You can’t use a data model without topology to understand the relationships between the spatial objects
The Field View

• In contrast to the discrete object view, which models the world in terms of entities, the field view approaches the world as consisting of continuous phenomena:

  The field view represents the real world as a finite number of variables, each one defined at every possible position (p.69 of the text)

• This view is inherently problematic for digital representation, which necessarily has to discretize the world into some set of minimum units
Topography Lends Itself to the Field View

• Topography is measurable at any point on the Earth’s surface and changes in value across the surface.
Other Field-like Phenomena

• Anything that does not have a well-defined boundary is more easily represented as a field:

• E.g. When mapping soil, we do not find a crisp break between one soil type and another typically … they simply continuously vary

• Similarly, when we try to define the boundary of a lake, we might run into trouble where the lake drains into a river … where does the lake end and the river begin … we might prefer to define any given point’s ‘lakeness’