CHAPTER 3

From the Real World to GIS

In many ways GIS presents a simplified view of the real world. Since the processes involved are seldom straightforward because realities are irregular and constantly changing, perception of the real world depends on the observer. For example, a surveyor might see a road as two edges to be surveyed, the roadwork authority might regard it as an asphalt surface to be maintained, and the driver will see it as a highway. Moreover, the real world may be described in terms of countless phenomena, from basic subatomic particles up to the dimensions of oceans and continents. The complexity of the real world, as well as the broad spectrum of its interpretations, suggests that GIS system designs will vary according to the capabilities and preferences of their creators. This human factor can introduce an element of constraint, as data compiled for a particular application may be less useful elsewhere.

The systematic structuring of the data determines its ultimate utility and consequently the success of the relevant GIS application. This aspect is also characteristic of the data available in traditional maps and registers. The real world can be described only in terms of models that delineate the concepts and procedures needed to translate real-world observations into data that are meaningful in GIS. The process of interpreting reality by using both a real-world and a data model is called data *modeling.* The principles involved are illustrated in Figure 3.1 and Figure 3.2.

The arrangement of the real-world model determines which data need to be acquired. The basic carrier of information is the *entity, which* is defined as a real-world phenomenon that is not divisible into phenomena of the same kind. An entity consists of:

- Type classification
- Attributes
- Relationships

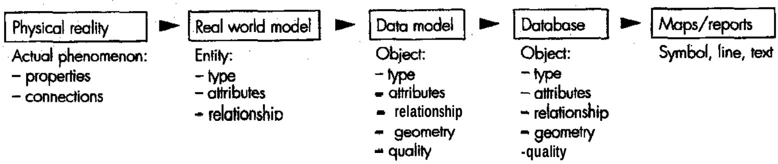


Figure 3.1
To bring the real world into GIS, one has to make use of simplified models of the real world. Uniform phenomena can be classified and described in the real-world model-which is converted into a data model by applying elements of geometry and quality. The data model is transferred to a database that can handle digital data, from which the data can be presented.

Entity types

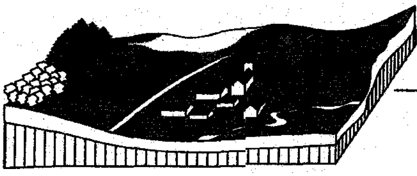
The concept of entity types assumes that uniform phenomena can be classified as such. During the classification process, each entity type must be uniquely defined to preclude ambiguity. For example, "house" must be defined in such a way that "detached house at No. 10 Church Road" is classified under "house" and not under "industrial building."

Some user organizations may need to classify entity types into categories as well as according to type. For example, national highways, country roads, urban roads, and private roads might come under the "roadways" category; alternatively, all entities within a specific geographical area might belong to a unique category of that area. In geographical data an entity type is also known as the nominal scale or qualitative data (Figure 3.3).

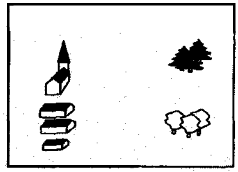
Entity attributes

Each entity type may incorporate one or more attributes that describe the fundamental characteristics of the phenomena involved (Figure 3.4). For example, entities classified as "buildings" may have a "material" attribute, with legitimate entries "frame" and "masonry" and a "number of stories" attribute with legitimate values of 1 to 10, and so on.

In principle an entity may have any number of attributes. For example, a lake may be described in terms of its name, depth, water quality, or fish population as well as its chemical composition, biological activity, water color, algae density, potability, or ownership. Attributes may also describe quantitative data, which may be ranked in three levels of accuracy: ratio, interval, and ordinal. The most accurate are ratio or pro*portional* attributes, such as length and area, which are measured with respect to an origin or starting point and on a continuous scale. *Interval* data, such as age and income category, comprise numerical data in groups and are thus less accurate. The least accurate are *ordinal* data of rank, such as "good," 'better," and 'best," which describe qualitative data in text form. These could also be characterized as quality data.



THE REAL WORLD



REAL-WORLD MODEL

Buildings	
– probable categories:	house, outbuilding, industrial building
– situated at:	property no./lot no.
– represented by:	single point
– geometric accuracy:	± 10 m

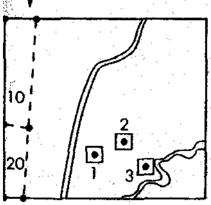
Vegetation	
– probable categories:	spruce, oak
– coverage/area:	hectares
– represented by:	area (polygon)
– geometric accuracy:	± 2.0 m

DATA MODEL

ID	Type	Property No.	X	Y	Accuracy
1	House	44 113	350	575	± 10.0
2	Outbuilding	45 6	375	600	± 10.0
3	Industrial	45 11	345	630	± 10.0

ID	Type	Area	Coordinates	Accuracy
10	Spruce	100	250,420 250,455 370,475 360,420 250,420	± 2.0
20	Oak	50	360,420 370,475 425,395 425,420 360,420	± 2.0

DATA BASE



MAP WITH SYMBOLS

Figure 3.2
Modeling process. The transformation of the real world into GIS products is achieved by means of simplification and models in the form of maps and reports.

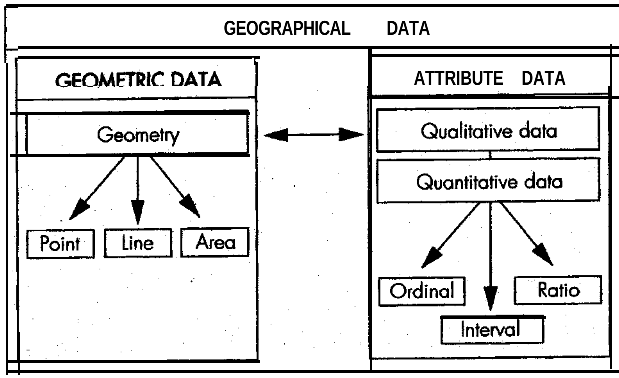


Figure 3.3
Geographical data can be divided into geometric data and attribute data. Attribute data can in turn be subdivided into qualitative data and quantitative data.

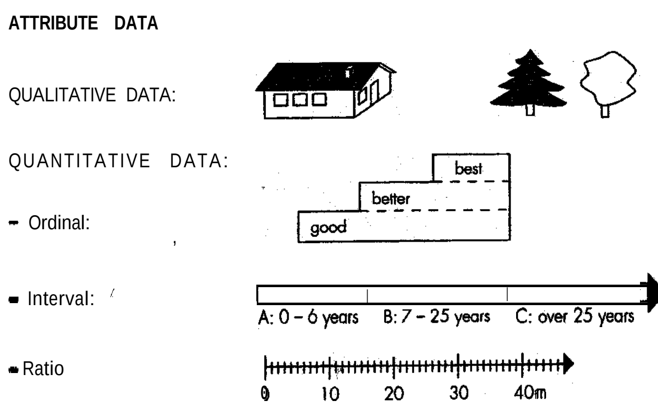


Figure 3.4
Attribute data consist of qualitative or quantitative data. Qualitative data specify the type of object, while quantitative data can be categorized into ratio data, data measured in relation to a zero starting point; interval data, data arranged into classes; and ordinal data, which specify quality by the use of text.

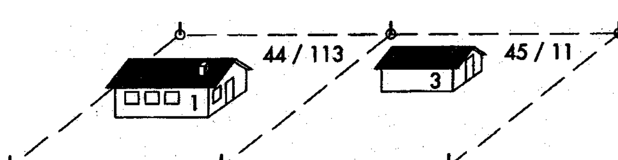


Figure 3.5
The computer cannot see the real world so it is necessary to specify the various relations between entities, such as belong *to, comprise,* are located *in/on,* and *border on.*

Entity relations

Relations often exist between entities. Typically, these include (Figure 3.5):

Relation	Typical example(s)
Pertains/belongs	A depth *figure pertains* to a specific shoal, or a pipe *belongs* to a *larger* network of contiguous pipes.
Comprises	A country or a state *comprises counties,* which *in turn comprise* townships.
Located in/on	A particular building *is located on* a specific property.
Borders on	Two properties have a common *border.*

Although such relations are intuitively obvious on ordinary maps, computers have no intuition. The computer processing or relations therefore requires either further descriptive information or instructions on how it may be compiled. This aspect of map reading-the human ability to see what a computer needs to be "told" to "see"-highlights pivotal differences between human and computer processing. As a sign in one international computer research laboratory states: Computers: Fast, Accurate, Stupid. Humans: Slow, Slovenly Smart.

Some relations may be unwieldy One cause may be complexity, as in networks where the states of switches or contacts, open or closed, determine which parts of the network may be viewed as comprising a logical entity. In such situations, one may differentiate between actual and potential relations.

Data Models

A real-world model facilitates the study of a selected area of application by reducing the number of complexities considered. Those outside the selected area are considered insignificant. However, if a real-world model is to be of any use, it must be realized in a database. A data model makes that possible.

Unlike humans, computers cannot "learn" the essentials of manholes, property lines, lakes, or other types of objects. What they can do is to manipulate geometric objects such as points, lines, and areas, which are used in data models. The carriers of information in data models are *known* as *objects* (Figure 3.7).* These correspond to entities

*The terms *entity, feature,* and *object* are often interchangeable. In this book we have chosen to use entity as a term for a real-world thing or phenomenon. The term *feature* is used very little in the book but is synonymous with the term *object* and is defined as a group of spatial elements which together represent a real-world entity.

Figure 3.6
A single entity can be described by several objects (i.e., there are many relationships between entities).

in real-world models and are therefore regarded as database descriptions of rear-world phenomena.

Objects are characterized by:

- Type
- Attributes
- Relations
- Geometry
- Quality

Real-world models and entities cannot be realized directly in databases, partly because a single entity may comprise several objects. For instance, the entity 'Church Road' may be represented as a compilation of all the roadway sections between intersections, with each of the sections carrying object information. Multiple representations produced by such divisions may promote the efficient use of GIS data (Figure 3.6). This means that information-carrying units and their magnitudes must be selected before the information is entered in a database. For example, the criteria for dividing a roadway in sections must be selected before the roadway can be described.

Objects

Objects in a GIS data model are described in terms of identity type, geometric elements, attributes, relations, and qualities. *Identifies*, which may be designated by numbers, are unique: no two objects have the same identity. Type codes are based on object classifications, which can usually be transferred from entity classifications. An object may be classified under one type code only.

Data models may be designed to include:

- Physical objects, such as roads, water mains, and properties
- Classified objects, such as types of vegetation, climatic zones, or age groups
- Events, such as accidents or water leaks
- Continuously changing objects, such as temperature limits
- Artificial objects, such as elevation contours and population density
- Artificial objects for a selected representation and database (raster)

3.3.1 Graphical representation of objects

Graphical information on objects may be entered in terms of (Figure 3.7):

- Points (no dimensions)
- Lines (one dimension)
- Areas (two dimensions)

Points

A point is the simplest graphical representation of an object. Points have no dimensions but may be indicated on maps or displayed on screens by using symbols. The corner of a property boundary is a typical point, as is the representative coordinate of a building. It is, of course, the scale of viewing that determines whether an object is defined as a point or an area. In a large-scale representation a building may be shown as an area, whereas it may only be a point (symbol) if the scale is reduced.

Lines

Lines connect at least two points and are used to represent objects that may be defined in one dimension. Property boundaries are typical lines, as are electric power lines and telecommunications cables. Road and rivers, on the other hand, may be either lines or areas, depending on the scale.

Areas

Areas are used to represent objects defined in two dimensions. A lake, an area of woodland, or a township may typically be represented by an area. Again, physical size in relation to the scale determines whether an object is represented by an area or by a point. An area is

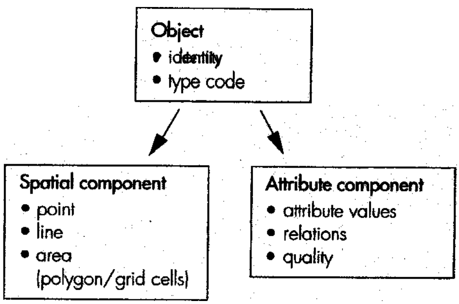


Figure 3.7
In a data model, objects are categorized as object classifications, geometric elements (point, line, area), attributes, relations between the entities, and quality definitions of these descriptive elements.

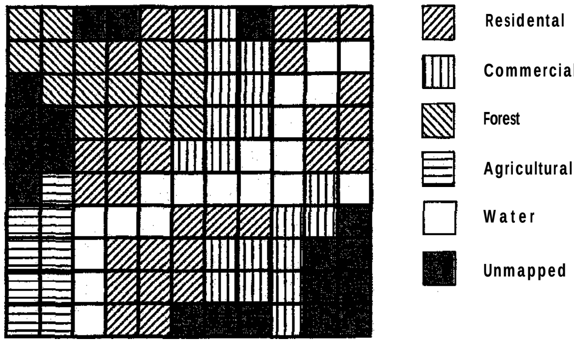
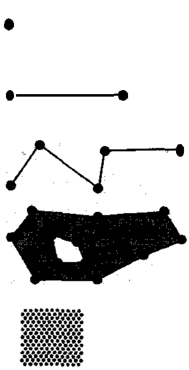


Figure 3.8
Land use in a town, in the form of a raster map. The land use is registered in a raster system with cells of varying size. Each category is given its own symbol on the map.

delineated by at least three connecting lines, each of which comprises points. In databases, areas are represented by polygons (i.e., plane figures enclosed by at least three straight lines intersecting at a like number of points). Therefore, the term polygon is often used instead of area.

Physical reality is often described by dividing it into regular squares or rectangles so that all objects are described in terms of areas (Figure 3.8). This entire data structure is called a *grid*. Population density is well suited to grid representation; each square or rectangle is known as a *cell* and represents a uniform density or value. The result is a generalization of physical reality. All cells of a grid in a data model or a database are of uniform size and shape but have no physical limits in the form of geometric lines.



Point: A &dimensional object that specifies geometric location specified througth a set of coordinates.

Line segment (vector): A one-dimensional object that is o direct line between two endpoints.

String: A sequence of line segments.

Area/polygon: A twodimensional object bounded by at least three onedimensional line segments.

Raster cell/pixel: A two-dimensional object [area) that represents an element of a regular tesselation of a surface.

Figure 3.9
Point, line, area.

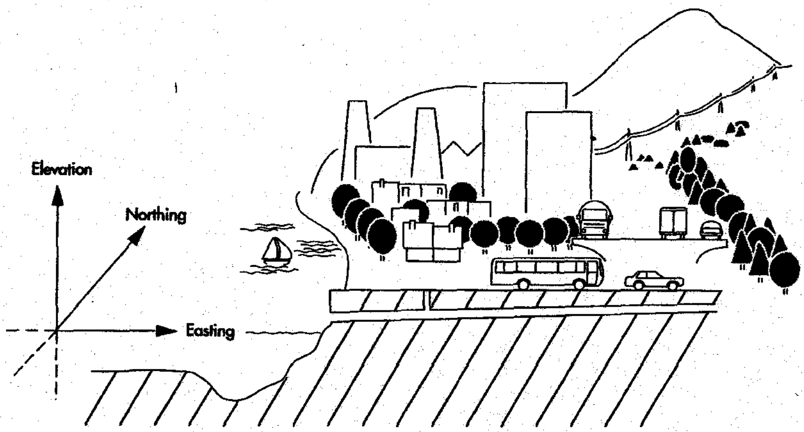


Figure 3.10
The world is three-dimensional with phenomena having a location and surface area in both elevation and ground plane.

In the traditional layer-based data model heights are treated as attributes to the objects, not as a part of the geometry But the real world is three-dimensional (Figure 3.10), and threedimensional data models are discussed in Chapter 5.

3.3.2 Object attributes

Object attributes are the same as the entity attributes of the real-world model. Attributes describe an object's features and may thus be regarded as a computer's "knowledge" of the object. In practice, object attributes are stored in tables (Figure 1.6), with objects on lines and attributes in columns. Theoretically, attribute values connected to grid data can be presented in the same way. Each grid cell corresponds to an object (Figure 3.11).

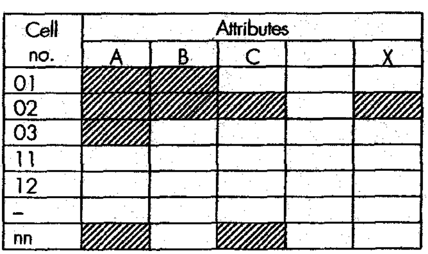


Figure 3.11
In principle, the difference between vector data and raster data is not that great. Raster data could well be arranged in tabular form with each cell number representing a line and each attribute (layer or raster values) a column.

3.3.3 Object relations

Object relations are the same as entity relations in the real-world model. Differentiation is made between:

1. Relations that may be calculated from:
   a. The coordinates of an object: for example, which lines intersect or which areas overlap
   b. Object structure (relation), such as the beginning and end points of a line, the lines that form a polygon, or the locations of polygons on either side of a line
2. Relations that must be entered as attributes, such as the division of a county into townships or the levels of crossing roads that do not intersect

3.3.4 Quality

The true value of any description of reality depends on the quality of all the data it contains: graphics, attributes, and relations. Graphical data accurate to ± 0.1 m obviously describe reality more faithfully than data accurate to. ± 100 m. Similarly, recently updated data are preferable to five-year-old data (which bring in temporal factors).

In the initial data modeling stage, the assessment of the data quality should include:

- Graphical accuracy (such as ± 1.0 m accuracy)
- Updating (when and how data should be updated)
- Resolution/detailing (e.g., whether roads should be represented by lines or by both road edges)
- Extent of geographical coverage, attributes included, and so on
- Logical consistency between geometry and attributes
- Representation: discrete versus continuous
- Relevance (e.g., where input may be surrogate for original data that are unobtainable)

Information on the quality of data is important to users of the database. Requirements for data quality are discussed in greater detail in Chapter 11.

3.4 From Database to GIS to Map

Once a data model is specified, the task of realizing it in a computer is technical and the task of entering data is simple and straightforward, albeit time consuming. A database need seldom be made to suit a data model, as many databases compatible with GIS applications are now on the market. The problem at hand is more one of selecting a suitable database with regard to:

- Acquisition and control
- ⬩♦□♦♍♦♦□♏
- **Storage ,**
- Updating and changing
- Managing and exporting/importing
- **Processing**
- Retrieval and presentation
- Analyses and combinations

Needless to say, a well-prepared data model is vital in determining the ultimate success of the GIS application involved. Users view reality using GIS products in the form of maps with symbols, tables, and text reports. The dissemination of information via maps is discussed further in Chapter 16.

3.5 Shortcomings of the Traditional GIS Data Model

3.5.1 Entities and fields

In the real world, one specific area or field may have many different characteristics; one area will in reality represent a number of entities or object types, such as coniferous forest, protected area, property no. 44/133, and so on. We experience on a daily basis that it is the area as an entity that carries the information. However, in our real-world model we split phenomena into entities (entity: a real-world phenomenon that is not divisible into phenomena of the same kind) and allow the entities to be bearers of information. This model will allow an entity to represent only one phenomenon (e.g., only coniferous forest or only protected area). To adapt the model to reality, overlapping phenomena (entities/objects) are separated into different layers (Figure 3.12). Reality is thus adapted to fit into a layer system, which is also traditionally used in map presentation. In the real world, areas are not

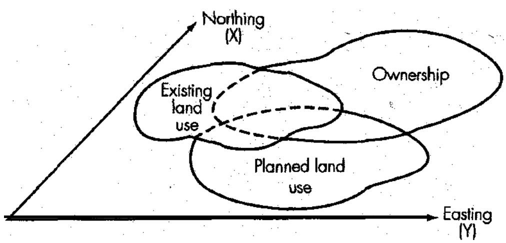


Figure 3.12
In the field model, the bearers of information consist of areas that often overlap. There are systems available today that can handle overlapping areas in the same layer.

divided into any form of horizontal two-dimensional physical layer— not even geological layers (strata) are presented in this way.

We can say that geometry-where coordinates define points, lines, and areas-is in many ways an artificial concept and an unnatural way to describe reality. Coordinates are not tangible and are never used in our everyday description of reality. Instead, we define a phenomenon's location in relation to other phenomena with which the recipient of the information is familiar. We can therefore establish that our model of reality is not perfect. During the 1990s, new models have been developed, *known* as *object-oriented models,* which to a certain extent can allow for the fact that the entity bearing the information can represent many phenomena. Object-oriented database systems are currently little used in commercial GIS but would appear to have many advantages over traditional database systems.

3.5.2 Uncertainty

To regard the real world as consisting of geometric constructs (points, lines, areas) means viewing objects as discrete data model representations. That is, all objects have clearly defined physical limits. These limitations are most obvious on maps, where lines imply sharp demarcations with no smooth, continuous transitions.

A discrete data model does not always suit reality Difficulties arise in depicting phenomena that lack clear physical demarcation, such as soil types, population densities, or prevailing temperatures. There can also be uncertainty in the attribute values to be retained. In the traditional discrete model, entities or objects are defined as being either within specific classes or outside them and thus operate only with areas that are homogeneous-with respect to limitation and classification. In reality, phenomena will often vary even within small, limited areas. For example, coniferous forest often contains deciduous trees, population density' is variable, and terrain surface changes continuously Once again, we have established that our real-world model is not perfect and that it is closely linked to traditional mapping concepts. Some of these problems can be partly solved by using the *fuzzy set theory,* which allows an object to belong only partially to a class. The fuzzy theory has as yet been little used in commercial GIS software; thus the significance of this type of deficiency in the data model is left to the person interpreting the results (maps and reports) of the GIS process.

3.5.3 Conceptual generalization

When points, lines, and polygons are selected as the geometric representation of objects, this very often results in a *generalization* of the real world; a town can be represented by a point rather than a polygon, and a road will frequently be represented by a center line and not two road verges. The need to divide objects into classes also results in a generalization. For example, an area of forest that is mainly coniferous, with some deciduous, will often be generalized and classified as coniferous,' not as a combination. Thus conceptual generalization is also a method for handling uncertain elements.

It will always be necessary to make choices about such generalizations in relation to the real world when making data models. This may be seen as a problem, but generalization is also a technique that makes it possible to obtain an overview of our complex reality. It can also be difficult to create data models that have a uniform and clear definition of the objects' classes. For example, does a pedestrian area that is accessible to emergency vehicles classify as a road?

3.5.4 Role of maps in data modeling

Maps are, in general, good sources for describing objects and their attributes. However, maps always represent particular models of the real world, and GIS shouldrepresent the real world, not the maps that depict it. For instance, ferry routes are often shown by dotted lines on maps, whereas in transport planning data models should form integral parts of a contiguous road network. As a rule of thumb, therefore, always look at a map as a data source, not as a data model.

[illegible]

The traditional model for transformation from the real world to GIS, as described above, has its obvious faults. In addition, it only describes flat and unchanging reality. Models for describing objects in three-dimensional space and terrain have not yet been discussed, nor has the fourth dimension-time-and its inroads into a geographical data model. The same applies to models for dealing with objects (traffic) moving along defined networks.

Here it is also most practical to use the same basic concept: a geometry consisting of points, lines,, and polygons, and attributes that describe the objects or phenomena. Elevation values can be linked to points, lines, and polygons and thereby give the objects a position in space. The surface of the terrain can be described with the help of sloping areas or with the help of horizontal surfaces with an elevation value linked as an attribute. Elevation values can also be linked to objects such as towers, wells, and buildings as attributes. The time factor can be accommodated by storing *all* historical data for the objects, such as changes in the geometry or attribute values. The movement of objects (traffic) along a road network can be simulated by assigning attribute values to elements in the network. These should be values that are relevant to the transfer speeds, and the sum of attribute values for different routes will be the measurement of passage in time or distance. A technical description of how this can be realized in GIS is given in Chapter 5, while the basic data models are described in Chapter 4.

Undoubtedly the traditional data model concept has definite drawbacks when describing these new real-world elements. We must accept that the real world is too complex to be described in full at present, although researchers are continuously engaged in developing improved models.

Geographic Information Systems

An Introduction

Second Edition

Tor Bernhardsen

Asplan Viak
Arendal, Norway

John Wiley & Sons, Inc.
New York • Chichester • Weinheim • Brisbane • Singapore • Toronto