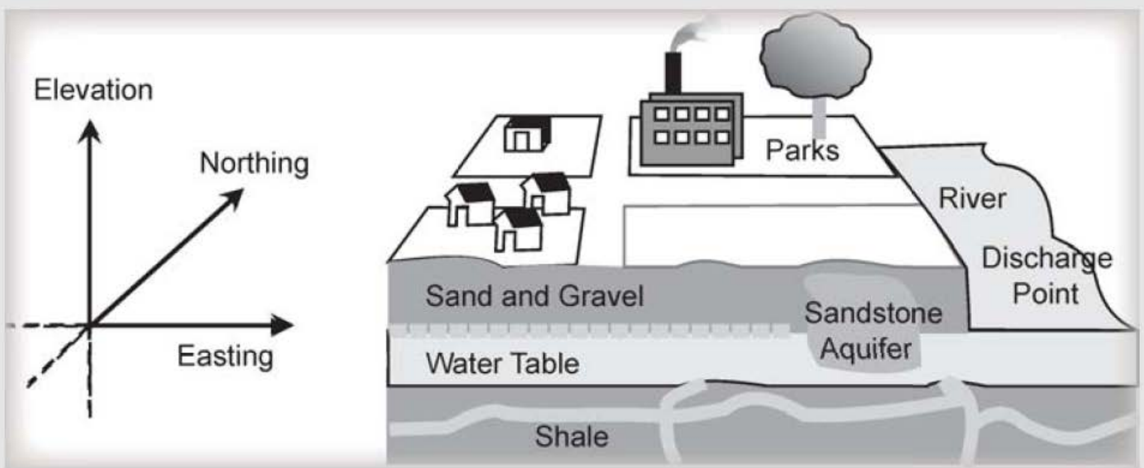
Chapter 6

GIS and the Real World Model



- The real world can be described 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 modelling.
  - The principles involved are illustrated in Figures 6.1 and 6.2.

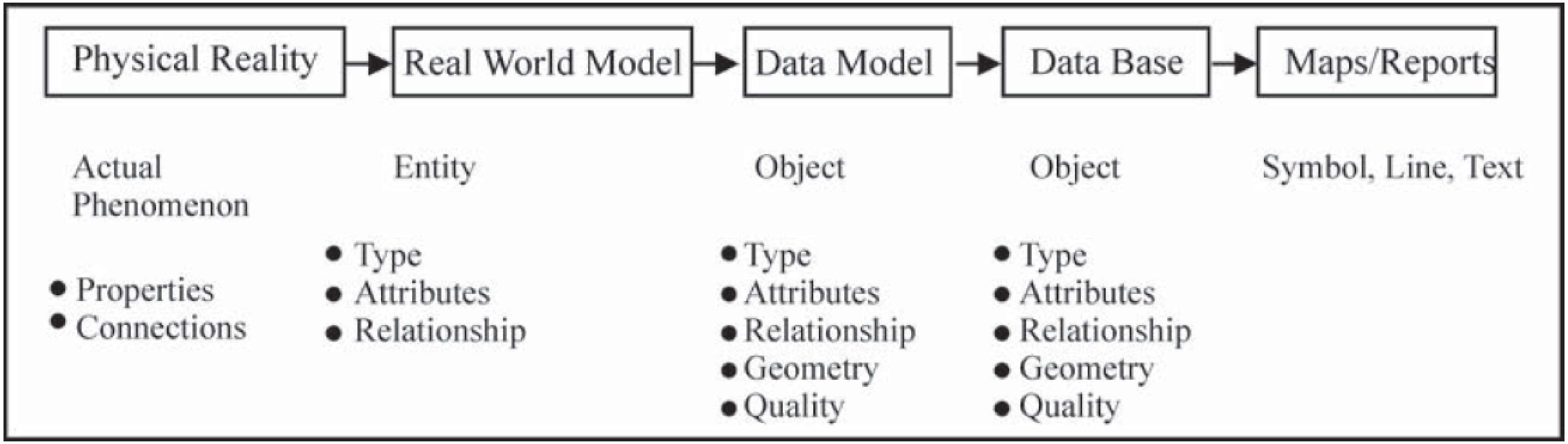
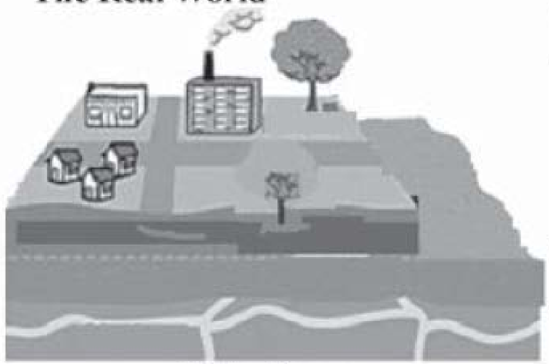


Figure 6.1: GIS makes simplified models to represent real world models. The data model is transferred to a database that can handle digital data, from which the data can be presented.

The Real World



Data Base

Data Model

Buildings

- Probable categories: Houses, industrial buildings
- Location: Property no.
- Representation: Area (polygon)
- Geometric accuracy: ± 10 meters

Vegetation

- Probable categories: Mango, Neem
- Coverage: Hectares
- Representation: Point
- Geometric accuracy: ± 2 meters

ID	Type	Location	X	Y	Accuracy
10	Mango	North Avenue	325	654	± 10 meters
20	Neem	South Avenue	455	725	± 10 meters

ID	Type	Area	Coordinates	Accuracy
1	House	75	350, 540. 350, 400. 250, 540. 175, 400	± 2 meters
2	Industrial	50	750, 820. 750, 650. 250, 820. 175, 650	± 2 meters

Map with Symbol

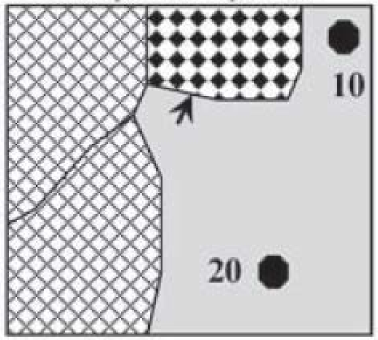


Figure 6.2: The transformation of the real world into GIS is achieved by means of simplification and models in the form of maps and reports.

REAL WORLD MODEL

- The arrangement of the real-world model determines which data need to be acquired.
- The basic carrier of information is the entity.
  - Entity 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

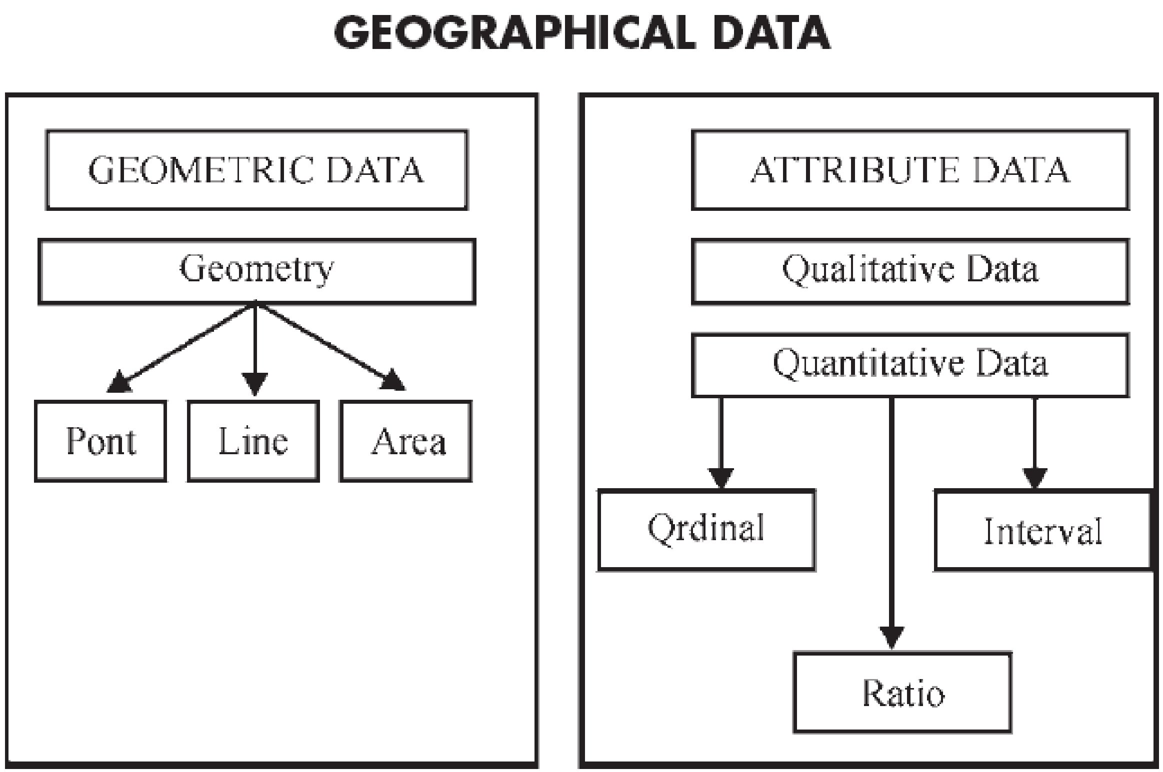


Figure 6.3: Geographical data can be divided into geometric and attribute data.

ENTITY TYPES

- The concept of entity types assumes that uniform phenomena can be classified as such.
  - Each entity type must be uniquely defined to preclude ambiguity.
- Some user organizations may need to classify entity types into categories as well as according to type.
- In geographical data an entity type is also known as the nominal scale or qualitative data (Figure 6.4).

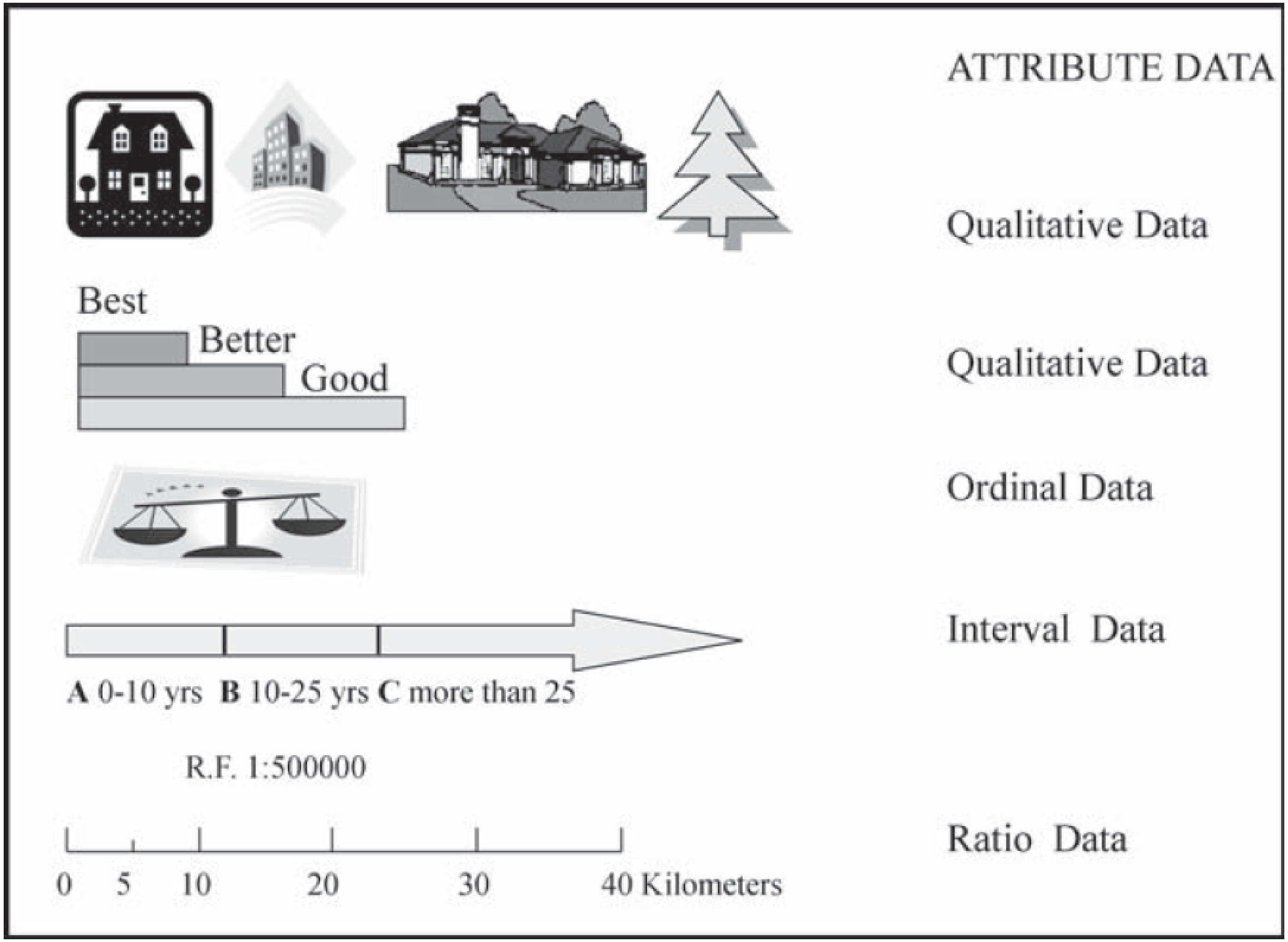


Figure 6.4: Attribute data consists of qualitative or quantitative data. Qualitative data specify the types of object, while quantitative data can be categorized into ratio data, data measured in relation to a zero starting point; interval data, data arranged into classes; and ordinal data, which specify quality by using text.

Entity Attributes

- Each entity type may incorporate one or more attributes that describe the fundamental characteristics of the phenomena involved.
  - Ex) ‘buildings’ with ‘material’ attribute, with legitimate entries ‘frame’ and ‘masonry’ and a ‘number of stories’ attribute with legitimate values of 1 to 10, and so on.
- An entity may have any number of attributes.
  - Ex) 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, 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 proportional attributes.
    - Such as length and area, which are measured with respect to an origin or starting point and on a continuous scale.
  - Interval data comprise numerical data in groups and are thus less accurate.
    - such as age and income category
  - 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.

ENTITY RELATIONS

- Relations often exist between entities, these include (Figure 6.5)

Relation	Example
Pertains/belongs	A depth figure pertains to a specific shoal, or a pipe belongs to a larger network of contiguous pipes.
Comprises	A state comprises districts, which in turn comprise townships.
Located in/on	A particular building is located on a land parcel.
Borders on	Two properties have a common border.

Figure 6.5: Examples of relations often exist between entities.

- Real-world models and entities cannot be realized directly in databases, partly because a single entity may comprise several objects.
  - For instance, the entity ‘Marris Road’ may be represented as a compilation of all the roadway sections between intersections, with each of the sections carrying object information.
  - The criteria for dividing a roadway in sections must be selected before the roadway can be described.

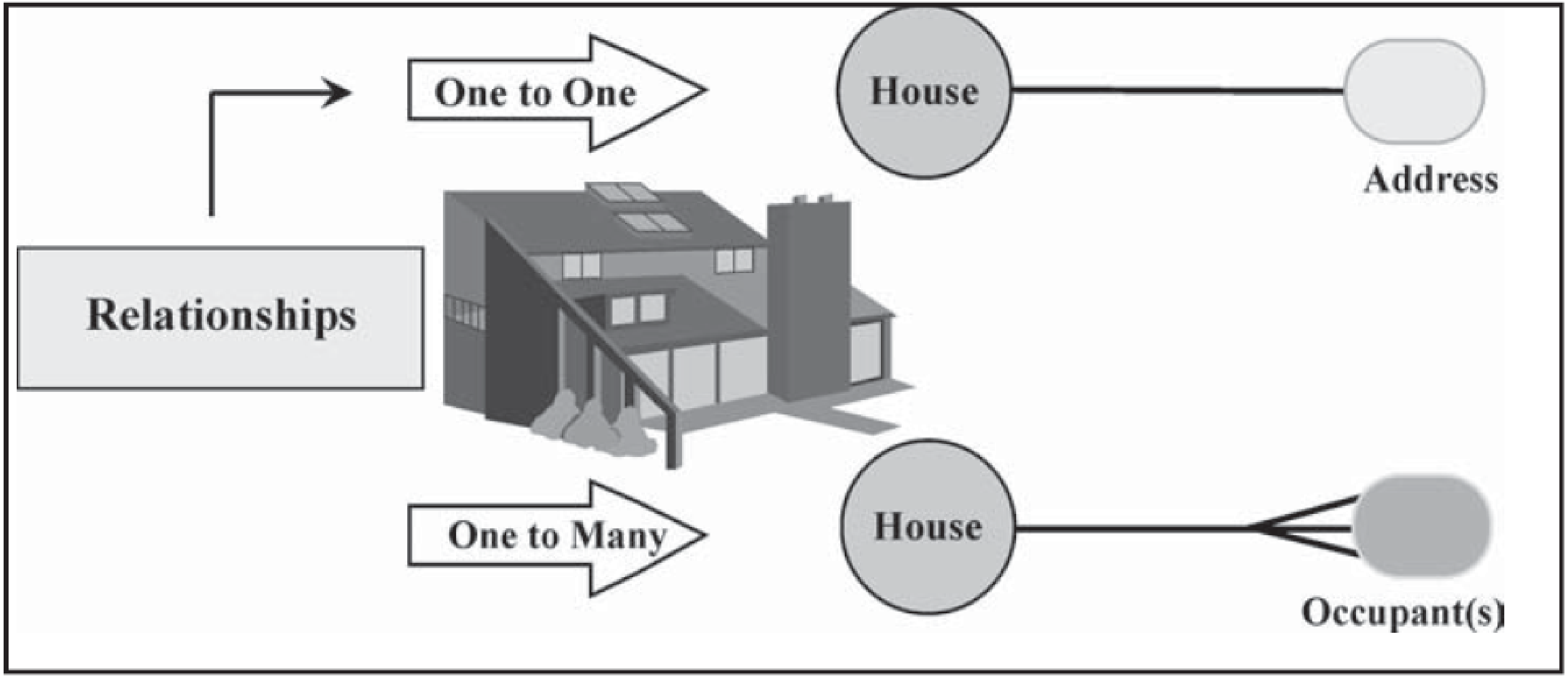


Figure 6.6: A single entity can be described by several objects (i.e., there are many relationships between entities).

Objects

- Objects are characterized by:
  - Type
  - Attributes
  - Relations
  - Geometry
  - Quality
- Identities, 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.

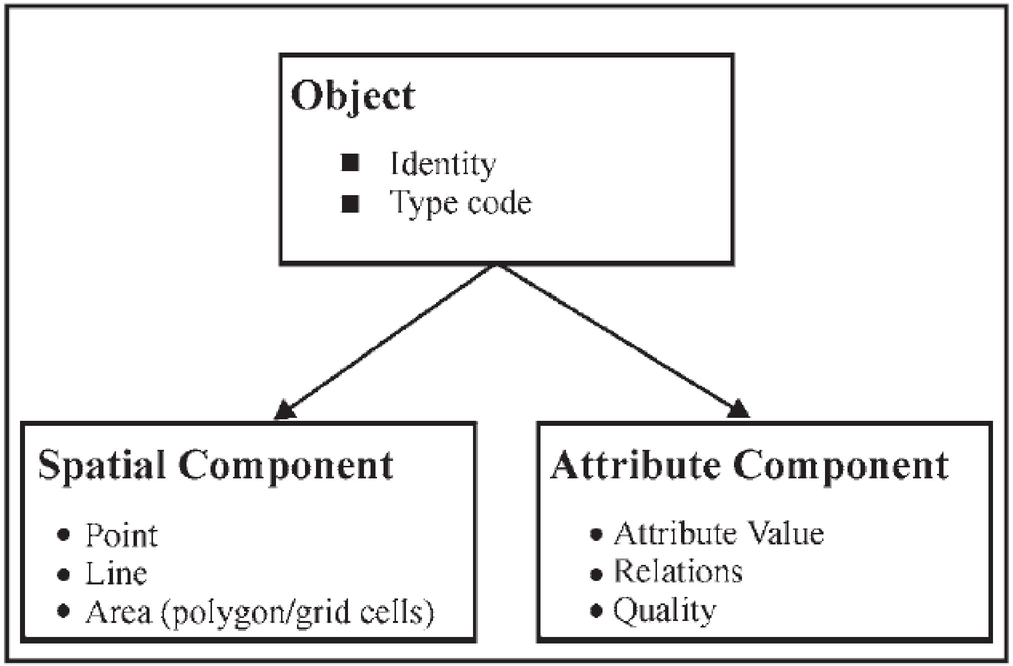


Figure 6.7: In a data model, objects are categorized as object classifications, geometric elements (point, line, area), attributes, relations between the entities and quality definitions of these descriptive elements.

- 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)

GEOGRAPHICAL REPRESENTATION OF OBJECTS

- Graphical information on objects may be entered in terms of:
  - Points (no dimensions)
  - Lines (one dimension)
  - Areas (two dimensions)

Points

- The simplest graphical representation of an object.
- Points are displayed by using symbols.
- The corner of a property boundary is a typical point.
- The scale of viewing determines whether an object is defined as a point or an area.
  - Ex) In a large-scale: 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.
- 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.
- Physical size in relation to the scale determines whether an object is represented by an area or by a point.
- An area is delineated by at least three connecting lines, each of which comprises points.
- In databases, areas are represented by polygons
  - Therefore, the term polygon is often used instead of area.

- Grid
  - Physical reality is often described by dividing it into regular squares or rectangles so that all objects are described in terms of areas (Figure 6.8).
  - This entire data structure is called a grid.

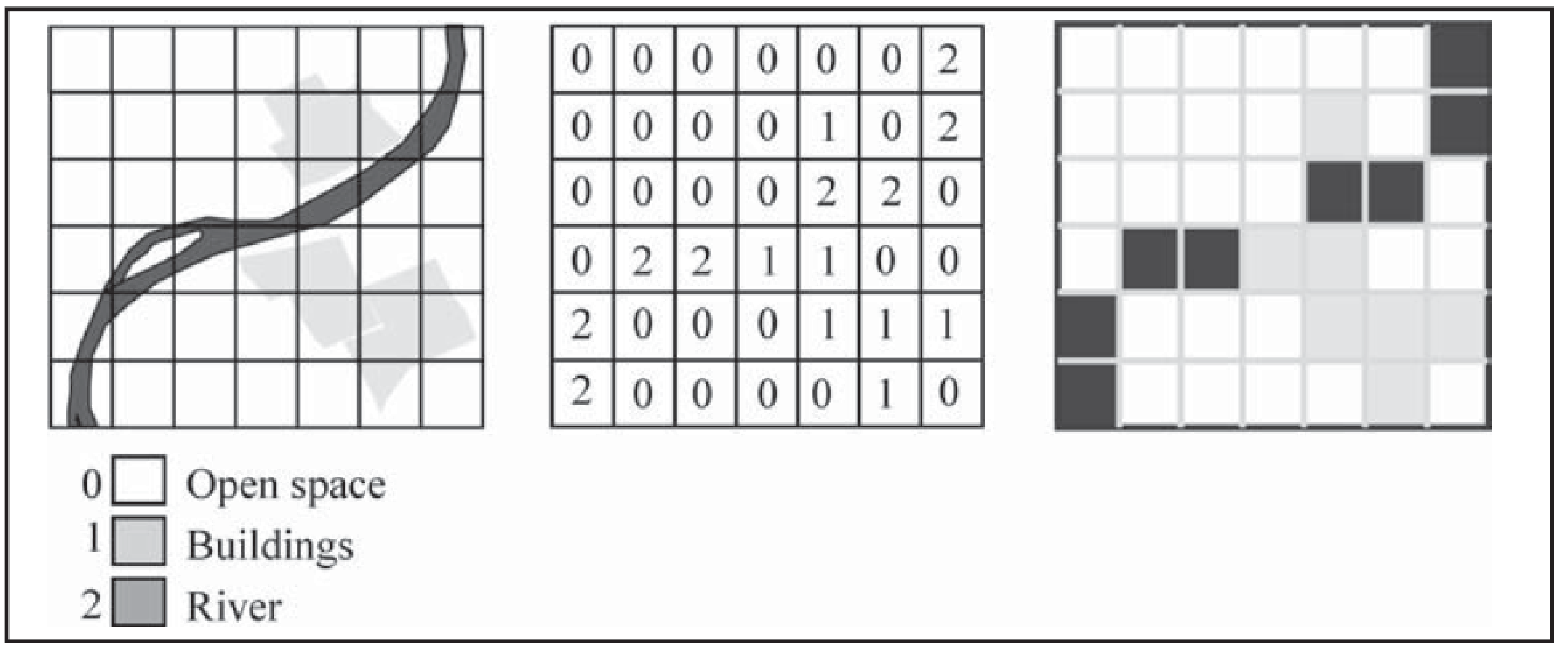
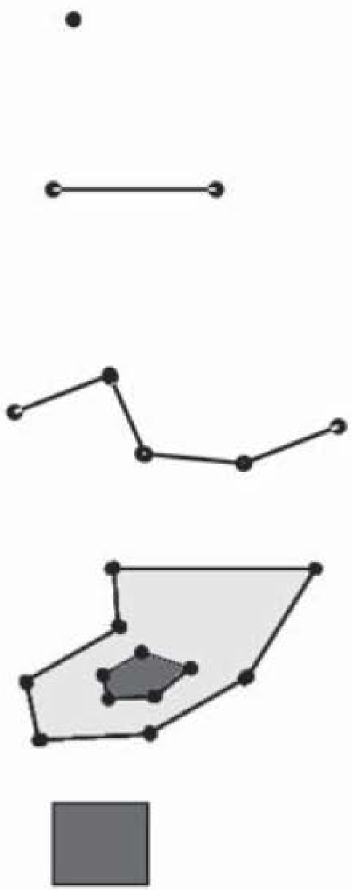


Figure 6.8: Land use/ Land cover in the form of a raster map. The land use is registered in a raster system with cells. Each category is assigned its own symbol on the map.

- Each square or rectangle is known as a cell and represents a uniform density or value.
- Population density is well suited to grid representation.
- 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 zero dimensional object that specifies geometric location specified through a set of coordinates.

Line:
A one dimensional object that is a direct line between two end-points.

String:
A sequence of line segments.

Area/Polygon:
A two dimensional object bounded by at least three one-dimensional line segments.

Raster Cell/Pixels:
A two dimensional object (area) that represents an element of a regular tessellation of a surface.

Figure 6.9: Point, line, area objects in GIS.

- Heights
  - 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.

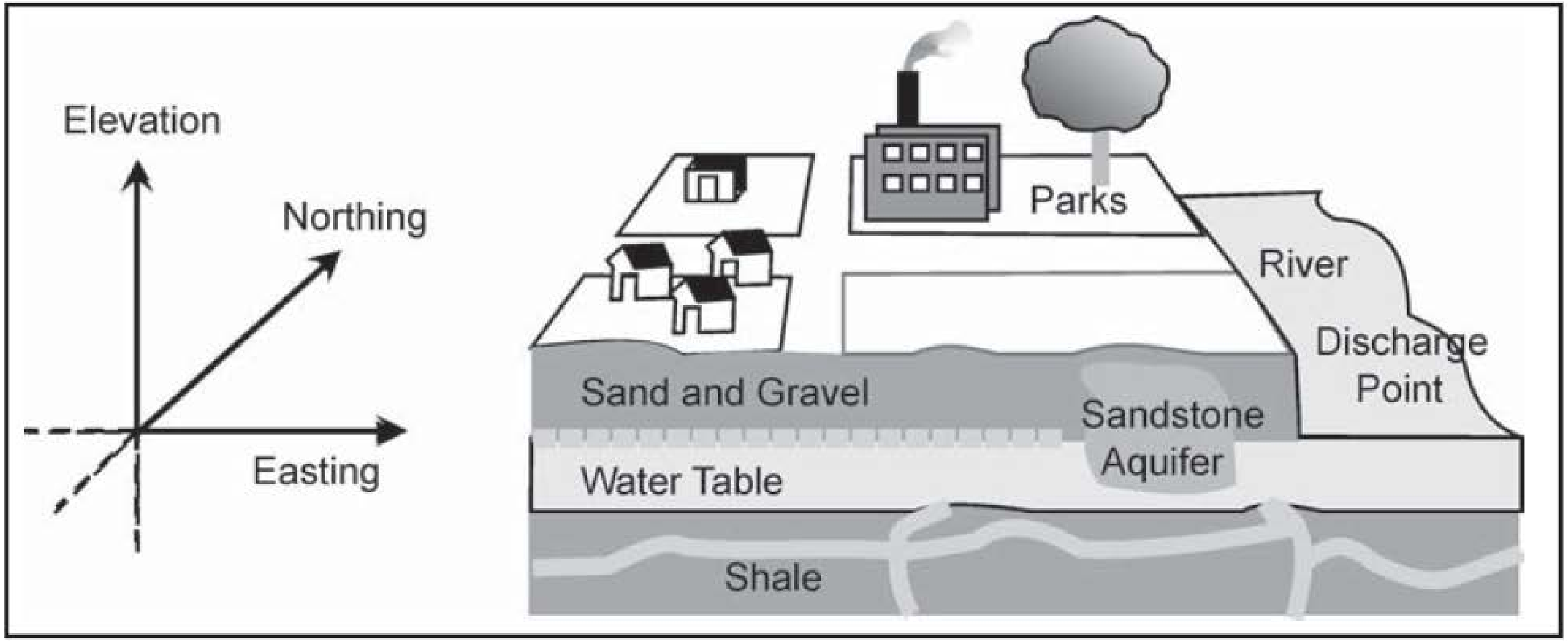


Figure 6.10: The world is three-dimensional with phenomena having a location and surface area in both elevation and ground plane.

OBJECT ATTRIBUTES

- Attributes describe an object’s features and may thus be regarded as a computer’s ‘knowledge’ of the object.
- Object attributes are stored in tables (Figure 6.11), 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.

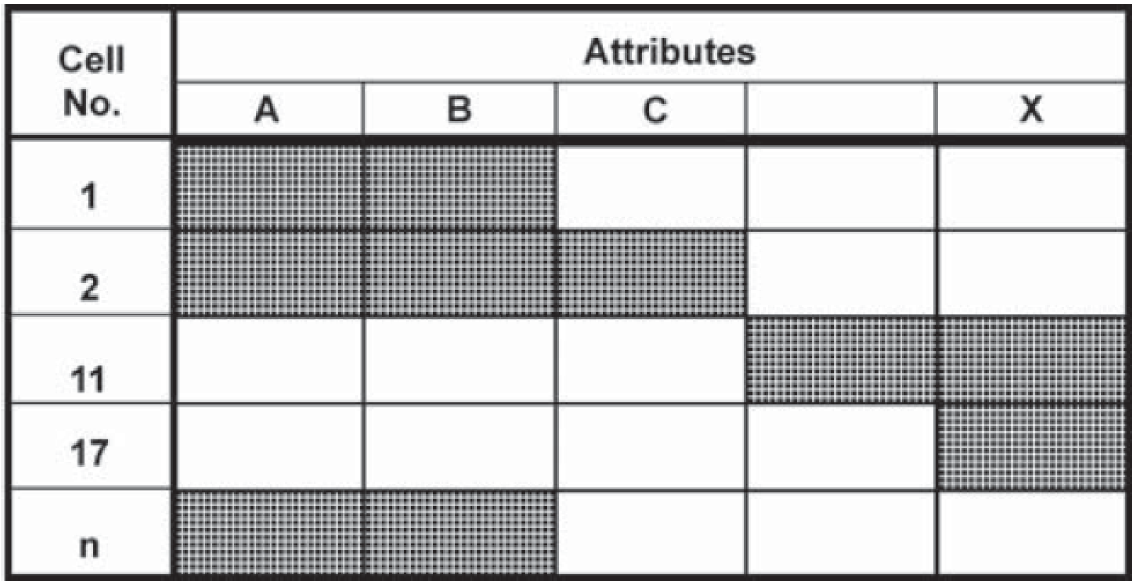


Figure 6.11: In principle, the difference between vector data and raster data is not that great. Raster data could well be arranged in tabular form with each cell number representing a line and each attribute (layer or raster) a column.

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 townships in to different wards or the levels of crossing roads that do not intersect.

Quality

- The true value of any description of reality depends on the quality of all the data it contains:
  - Graphics
  - Attributes
  - Relations
- Graphical data accurate to ± 0.1 meter obviously describe reality more faithfully than data accurate to ± 100 meter.
- 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 meter 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.

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.
- 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
  - Structure
  - Storage
  - Updating and changing
  - Managing and exporting/importing
  - Processing
  - Retrieval and presentation
  - Analyses and combinations

- 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.

SHORTCOMINGS OF THE TRADITIONAL GIS DATA MODEL

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.
- However, the traditional data model allows an entity to represent only one phenomenon.
- 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.

Uncertainty

- A traditional 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; For example, coniferous forest often contains deciduous trees, population density is variable, and terrain surface changes continuously.
- 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.

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.
  - An area of forest that is mainly coniferous, with some deciduous, will often be generalized and classified as coniferous, not as a combination.
- Generalization may be seen as a problem, but it 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?

ROLE OF MAPS IN DATA MODELLING

- Maps always represent particular models of the real world, and GIS should represent 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.

EXTENSION OF THE REALITY CONCEPT

- The traditional data model concept has definite drawbacks when describing new real-world elements : elevation, time, traffic.
  - It only describes flat and unchanging reality.
- 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 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.