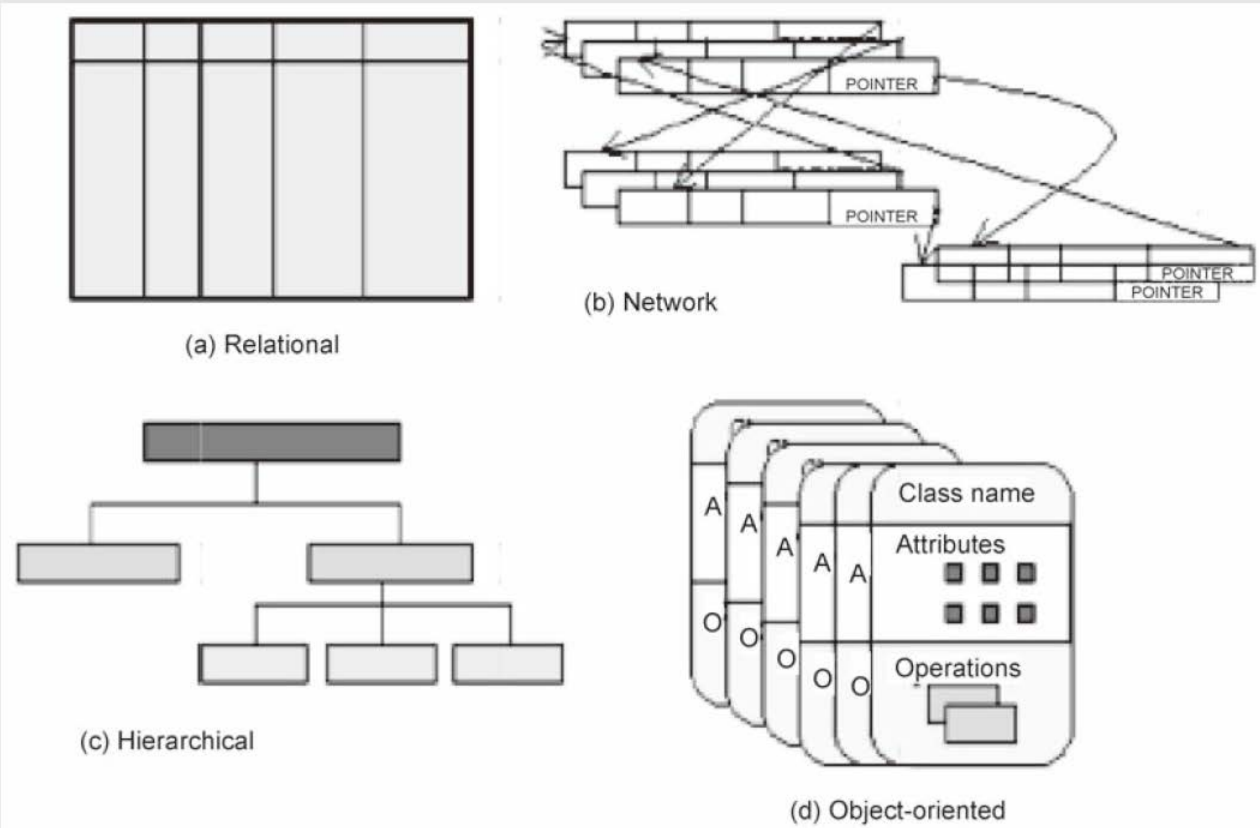
Chapter 4

Spatial Data Structure and Models



(a) Relational

(b) Network

(c) Hierarchical

(d) Object-oriented

INFORMATION ORGANIZATION AND DATA STRUCTURE

Data and information

- '*Data*'
  - A body of facts or figures, which have been gathered systematically for one or more specific purposes.
- Data can exist in the forms of:
  - linguistic expressions (*e.g.,* name, age, address, date, ownership)
  - symbolic expressions (*e.g.,* traffic signs)
  - mathematical expressions (*e.g.,* $E = mc^2$)
  - signals (*e.g.,* electromagnetic waves)
- '*Information*'
  - Data which have been processed into a form that is meaningful to a recipient and is of perceived value in current or prospective decision making.
- Although data are ingredients of information, not all data make useful information.

- Information is only useful to its recipients when it is:
  - relevant (to its intended purposes and with appropriate level of required detail)
  - reliable, accurate and verifiable (by independent means)
  - up-to-date and timely (depending on purposes)
  - complete (in terms of attribute, spatial and temporal coverage)
  - intelligible (*i.e., comprehensible by its recipients)*
  - consistent (with other sources of information)
  - convenient/easy to handle and adequately protected.

Information system

- The function of an *information system* is to change 'data' into 'information'.
- It uses the following processes:
  - conversion – transforming data from one format to another, from one unit of measurement to another, and/or from one feature classification to another
  - organization – organizing or re-organizing data according to database management rules and procedures so that they can be accessed cost-effectively
  - structuring – formatting or re-formatting data so that they can be acceptable to a particular software application or information system
  - modelling – including statistical analysis and visualization of data that will improve user's knowledge base and intelligence in decision making.

- ‘Organization’ and ‘structure’ are crucial to the functioning of information systems-without organization and structure it is simply impossible to turn data into information.

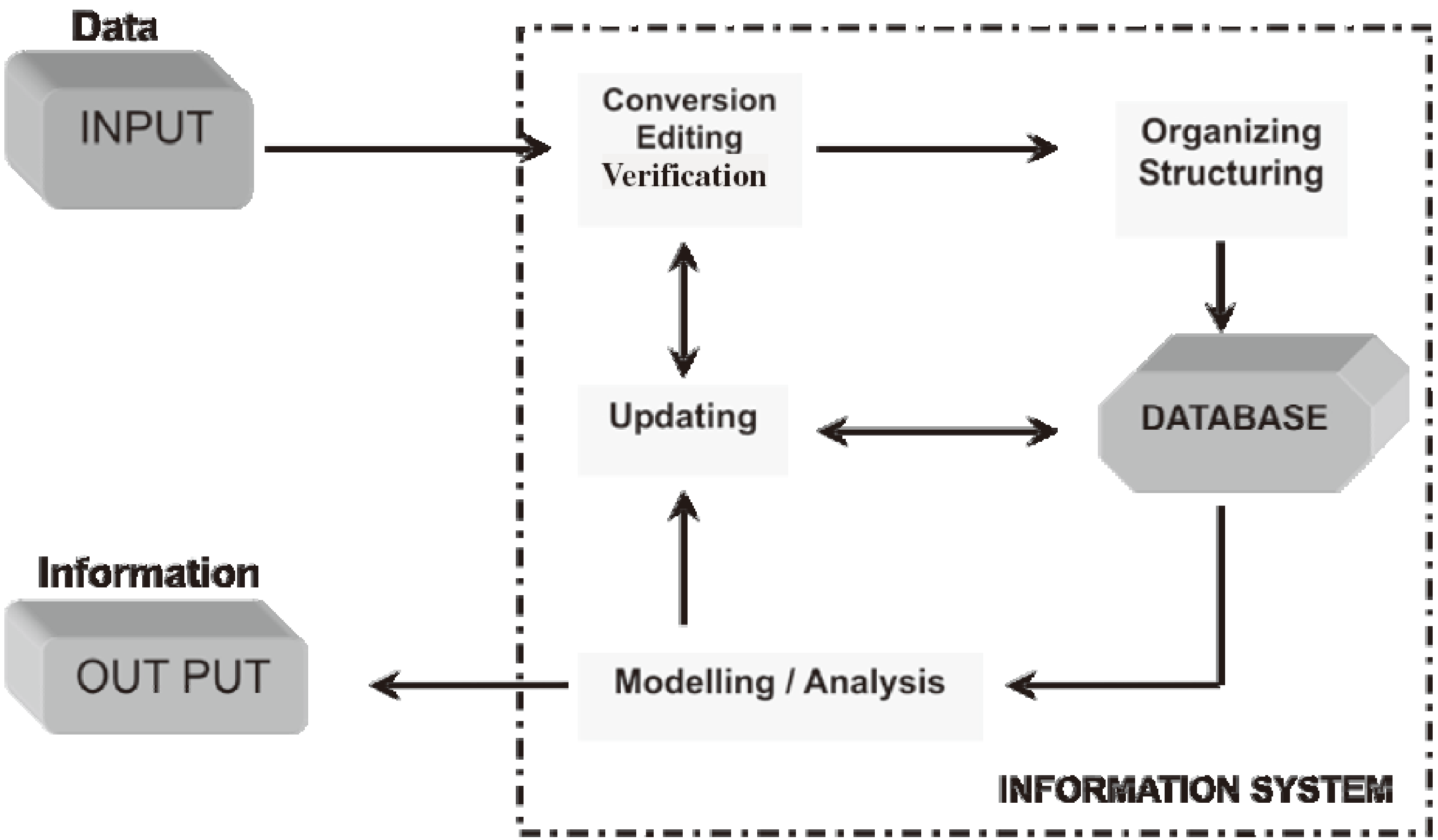


Figure 4.1: Changing data into information in an information system.

GEOGRAPHIC DATA AND GEOGRAPHIC INFORMATION

- Geographic data are a special type of data; by 'geographic'.
- The data are pertinent to features and resources of the Earth, as well as the human activities based on or associated with these features and resources.
- The data are collected and used for problem solving and decision making associated with geography, *i.e.,* location, distribution and spatial relationships within a particular geographical framework.
- Geographic data are different from other types of data in that they are geographically referenced.
  - *i.e.,* they can be identified and located by coordinates.
  - They are made up of a *descriptive element* (*non-spatial data)* and *a graphical element* (*spatial data*).
- Geographic information is obtained by processing geographic data, the aim of which is to improve the user's knowledge about the geography of the Earth's features and resources, as well as human activities associated with these features and resources.

- Since the special nature and characteristics of geographic data, generic concepts of information organization and data structure cannot be applied directly to them.
- Geographic data have three dimensions:
  - a. Temporal – *e.g.,* 26th December 2004,
  - b. Thematic – *e.g.,* occurrence of tsunami in Indian Ocean,
  - c. Spatial – *e.g.,* affected area included south east coast of India.
    - GIS emphasizes on the use of the spatial dimension for turning data in to information.

INFORMATION ORGANIZATION

- Information organization can be understood from four perspectives:
  - a data perspective
  - a relationship perspective
  - an operating system (OS) perspective
  - an application architecture perspective

THE DATA PERSPECTIVE OF INFORMATION ORGANIZATION

Information organization of descriptive data

- Data item
  - The descriptive data, *data item* is the most basic element of information organization.
  - A data item represents an *occurrence* or *instance* of a particular characteristic pertaining to an entity (which can be a person, thing, event or phenomenon).
  - It is the smallest unit of stored data in a database, commonly referred to as an *attribute.*
  - In database terminology, an attribute is also referred to as a *stored field.*
  - The value of an attribute can be in the form of a number, a character string, a date or a logical expression.
  - Some attributes have a definite set of values known as *permissible values* or *domain of values.*

- Record
  - A group of related data items form a *record* (figure 4.2).
  - Related data items, means that the items are occurrences of different characteristics pertaining to the same person, thing, event or phenomenon.
  - A record may contain a combination of data items having different types of values.
  - In database terminology, a record is always formally referred to as a *stored record*
  - In relational database management systems, records are called *tuples.*

- Data file
  - A set of related records constitutes a *data file* (figure 4.2).
  - Related records, means that the records represent different occurrences of the same type or class of people, things, events and phenomena.
  - A data file made up of a single record type with single-valued data items is called a *flat file* (table 4.1).
  - A data file made up of a single record type with nested repeating groups of items forming a multi-level organization is called a *hierarchical file* (table 4.1).
  - A data file is individually identified by a *filename.*
  - A data file containing records made up of character strings is called a *text file* or *ASCII file.*
  - A data file containing records made up of numerical values in binary format is called a *binary file.*

- An *array* is a collection of data items of the same size and type (although they may have different values).
  - a one-dimensional array is called a *vector*
  - a two-dimensional array is called a *matrix*

- Table
  - A table is a data file with data items arranged in rows and columns.
  - Data files in relational databases are organized as tables.
  - Tables are also called *relations* in relational database terminology.
  - A *list* is a finite, ordered sequence of data items (known as *elements*).
    - An ordered list has elements positioned in ascending order of values.
  - Each element has a data type, in the simple list implementation, all elements must have the same data type but there is no conceptual objection to lists whose elements have different data types.

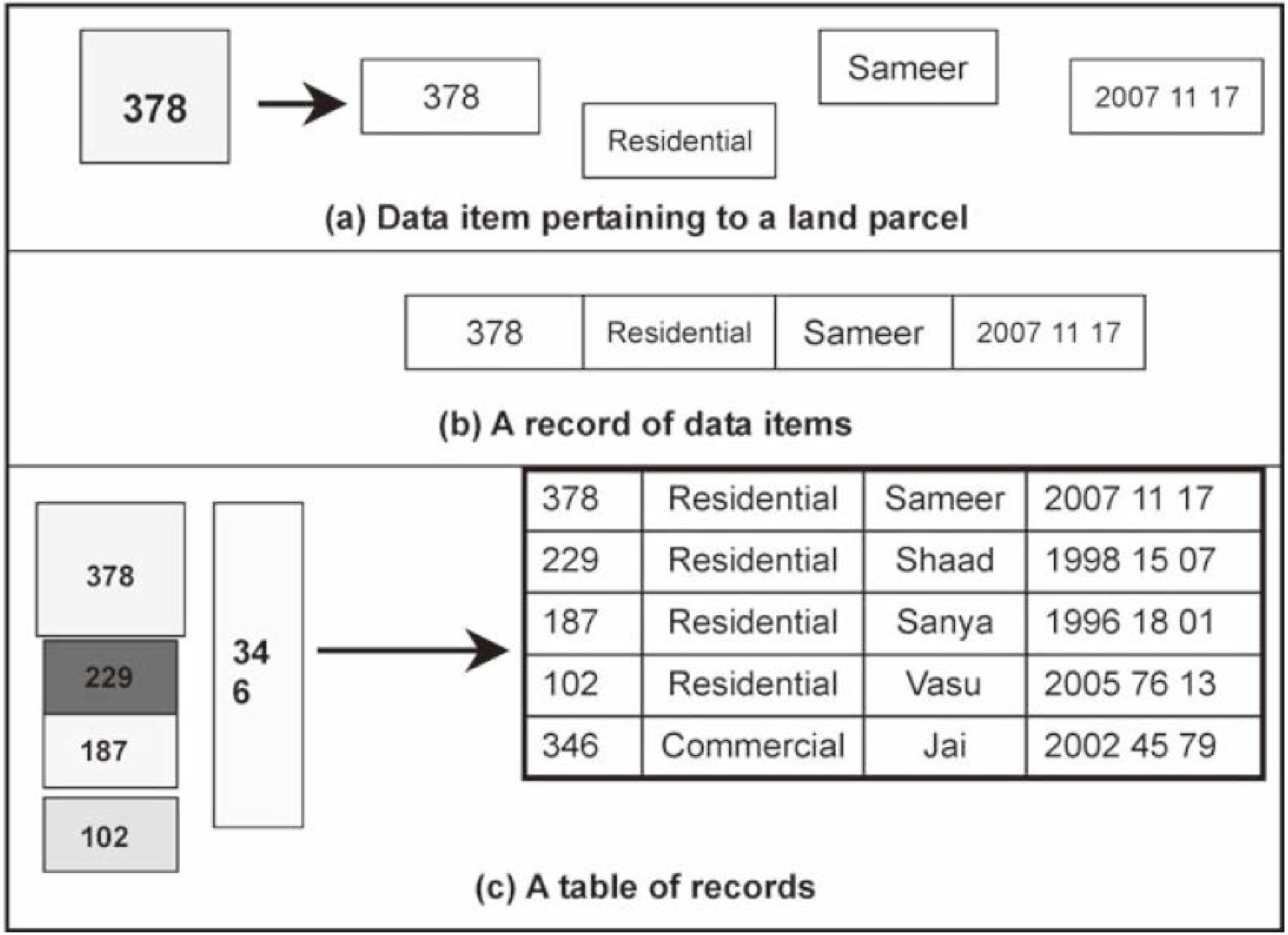


Figure 4.2: Data item, record, date file.

Table 4.1: Flat file and hierarchical file.

- **A Flat file**

Ward no.	Population	No. of households	Average monthly income
14	2431	654	Rs. 10,500
21	1740	389	Rs. 15,000
56	1985	557	Rs. 12,000

- **A Hierarchical file**

Ward no.	Population		No. of households		Average monthly income	
	1991	2001	1991	2001	1991	2001
14	1434	2431	568	654	Rs. 8,000	Rs. 10,500
21	1047	1740	307	389	Rs. 13,500	Rs. 15,000
56	1286	1985	489	557	Rs. 9,000	Rs. 12,000

- Tree
  - A *tree* is a data file in which each data item is attached to one or more data items directly beneath it (figure 4.3).
  - The connections between data items are called *branches.*
  - Trees are often called *inverted trees* because they are normally drawn with the root at the top.
  - The data items at the very bottom of an inverted tree are called *leaves*; other data items are called *nodes.*
  - A *binary tree* is a special type of inverted tree in which each element has only two branches below it.
  - A *heap* is a special type of binary tree in which the value of each node is greater than the values of its leaves.
  - Heap files are created for sorting data in computer processing - the *heap sort algorithm* works by first organizing a list of data into a heap.

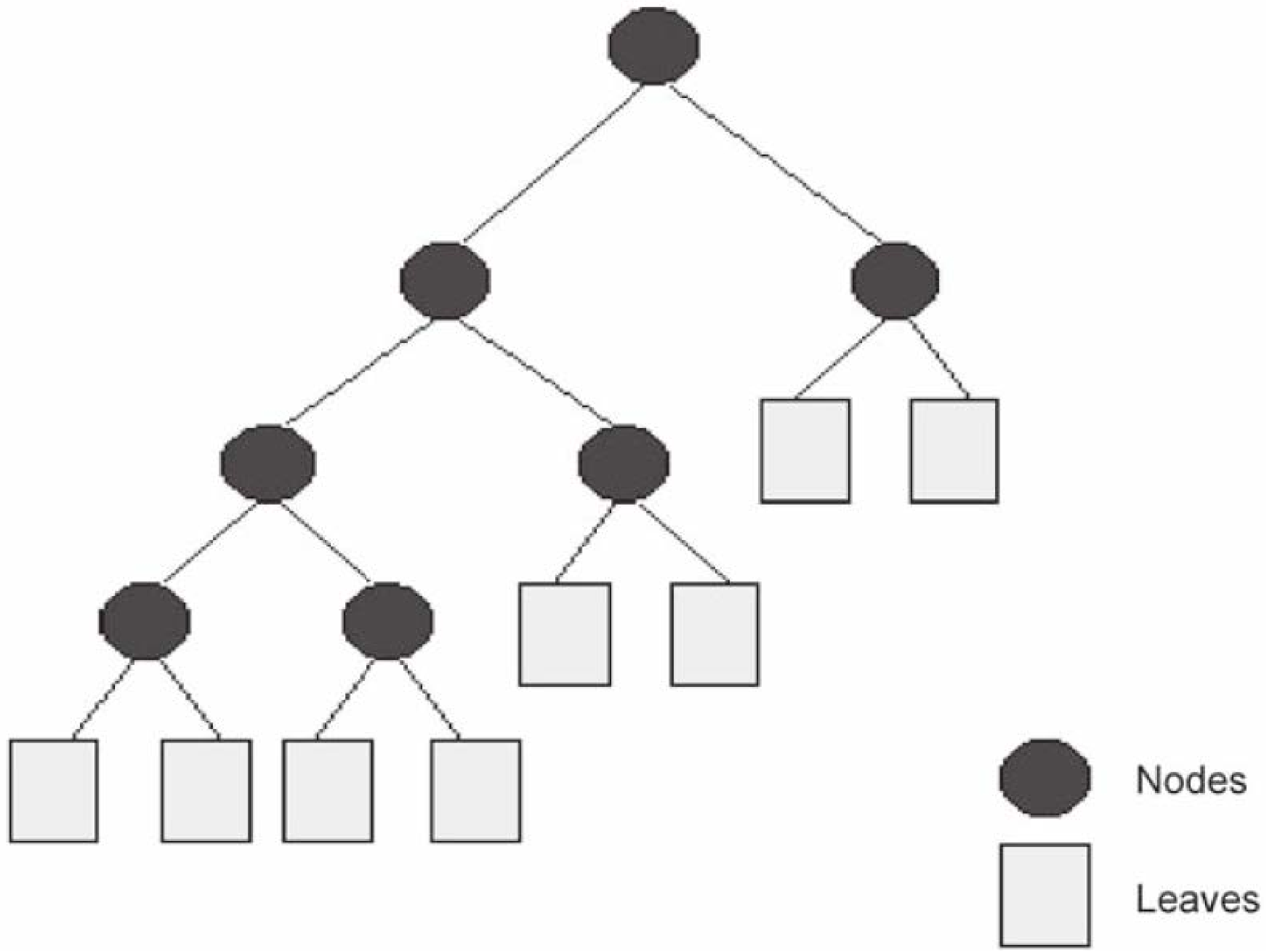


Figure 4.3: The tree data structure.

- Database
  - A *database* is defined as an automated, formally defined and centrally controlled collection of persistent data used and shared by different users in an enterprise.
  - A database is set up to serve the information needs of an organization and data sharing is key to the concept of database.
  - Data in a database are described as ‘permanent’ in the sense that they are different from ‘transient’ data such as input to and output from an information system.
  - Data in a database are still organized and stored as data files.
  - The use of database represents a *change* in the perception of data, the mode of data processing and the purposes of using the data, rather than physical storage of the data.

Table 4.2: Distinction between a data file and a database.

Characteristics of a data file	Characteristics of a database
A collection of records usually of the same data type and format description	A collection of interrelated records, organized in one or more data files, that may have different data types and format descriptions
Data file processing is usually associated with computer programming that aims at solving a particular problem, *i.e.*, it stops when an answer is obtained	Database processing is always associated with database management systems that aim at solving the operation or production needs of an organization, *i.e.*, it involves routine, largely repetitive applications executed over and over again
Mainly used in support of the information need of an *ad hoc* application	Mainly used in support of the day to day operation of business (transaction processing) but increasingly used in decision support (management decision making)

- Database models
  - Databases can be organized in different ways known as *database models.*
  - The conventional database models are: *relational, network, hierarchical and object-oriented* (figure 4.4).
  - a. Relational-data are organized by records in relations which resemble a table
  - b. Network-data are organized by records which are classified into record types, with 1: n pointers linking associated records
  - c. Hierarchical-data are organized by records on a parent-child one-to-many relations
  - d. Object oriented-data are uniquely identified as individual objects that are classified into object types or classes according to the characteristics (attributes and operations) of the object.

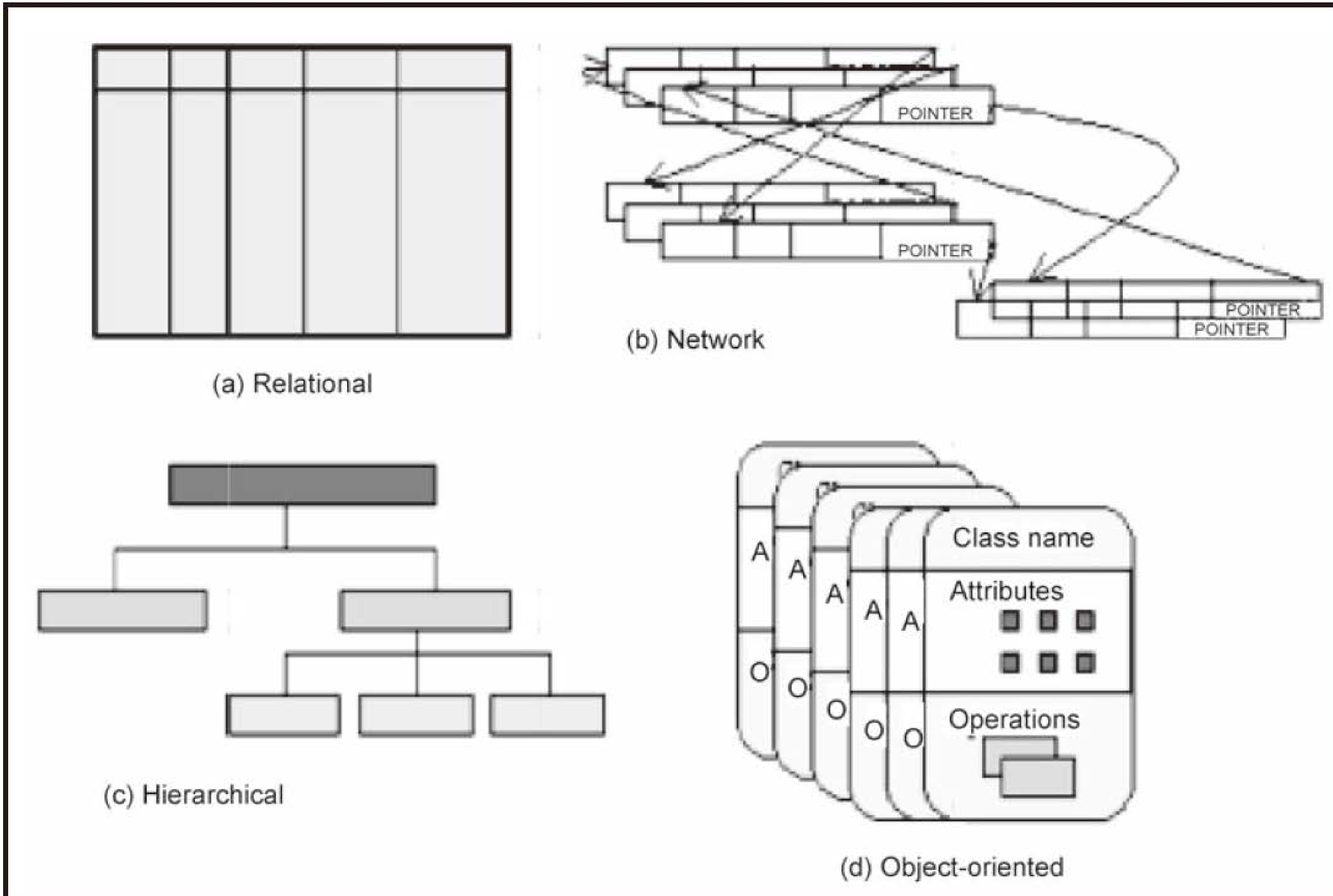


Figure 4.4: Database models.

Information Organization of Graphical Data

- The graphical data, where the most basic element of information organization is called as *basic graphical element.*
- There are three basic graphical elements (figure 4.5):
    - *point*
    - *line,* also referred to as *arc*
    - *polygon,* also referred to as *area*
- These basic graphical elements can be individually used to represent geographic features or entities.
    - For example, point for a well; line for a road segment and polygon for a lake.
- They can also be used to construct complex features.
    - For example, the geographic entity ‘India’ on a map is represented by a group of polygons of different sizes and shapes.

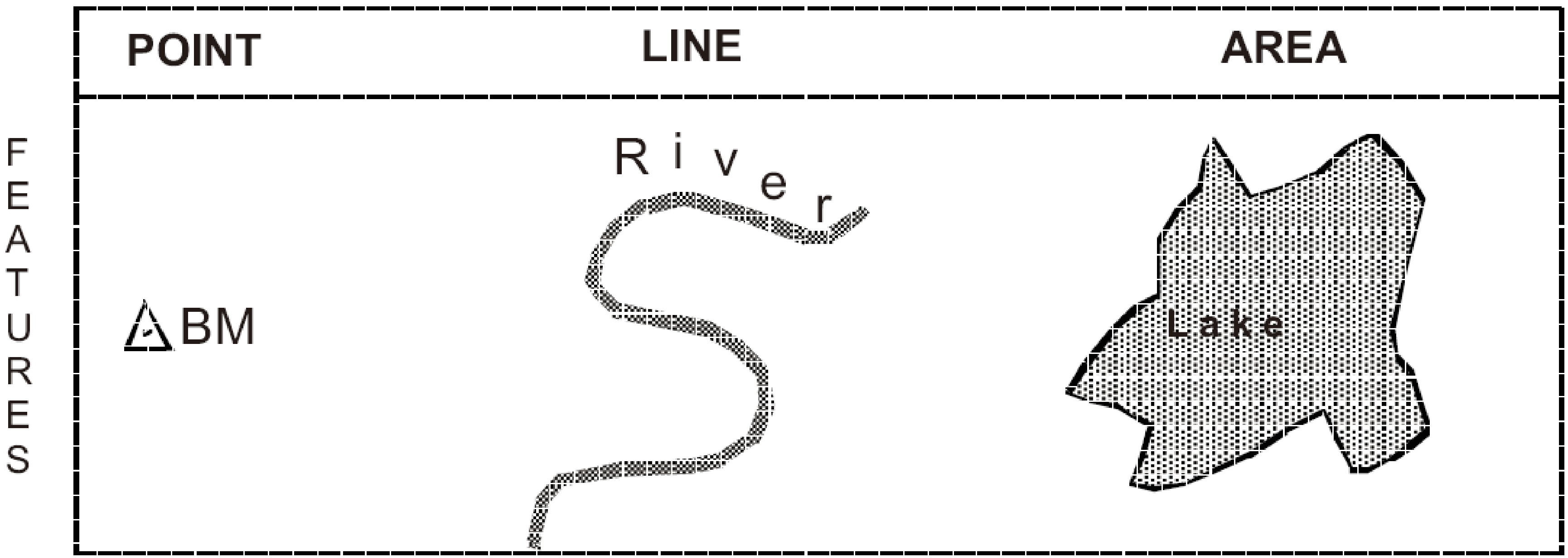


Figure 4.5: The feature model: Examples of a point feature (elevation bench mark), a line feature (river) and an area feature (lake).

- Point feature
  - A point has neither length nor breadth and hence is said to be of dimension 0.
  - A point feature represents as single location.
  - A point is the simplest graphical representation of an object.
  - Points 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.

- Line feature
  - Lines have length, but not breadth hence is of dimension 1.
  - They are used to represent linear entities such as rivers, roads, pipelines, and cables etc.
  - A line feature is a set of connected, ordered coordinates representing the linear shape of a map object that may be too narrow to display as an area or feature with no width.
- Area feature
  - Area objects have the two dimensions of length and breadth.
  - An area feature is a closed figure whose boundary encloses homogeneous characteristics, such as a state boundary, soil type or lake.
  - 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.

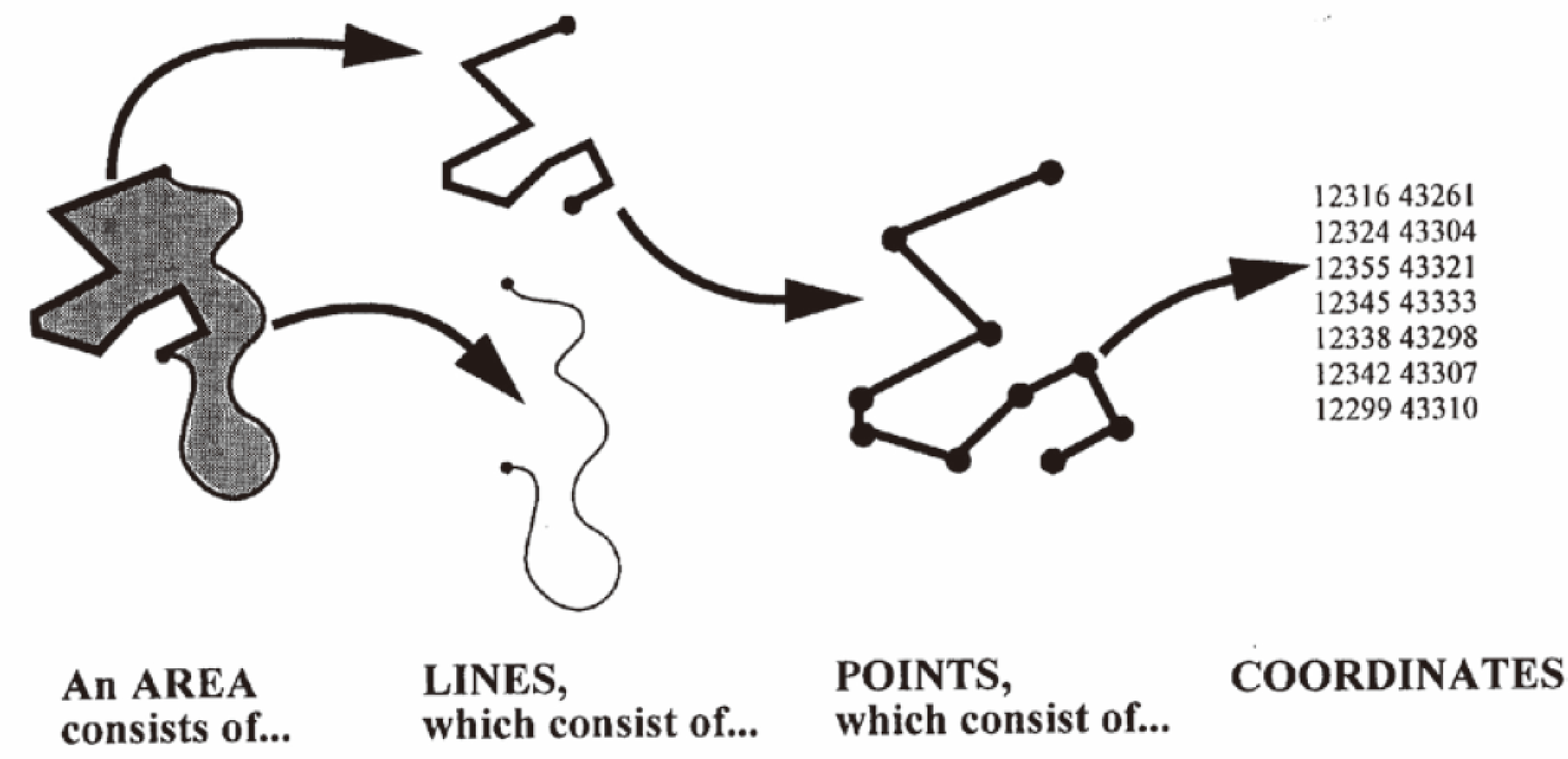


Figure 4.6: Geographic information has dimensions, areas are two dimensional and consists of lines, which are one dimensional and consists of points, which are zero dimensional and consist of a coordinate pair.

- Vector data model
  - The method of representing geographic features by the basic graphical elements of points, lines and polygon is said to be the *vector method* or *vector data model*, and the data are called *vector data*.
  - Related vector data are always organized by *themes,* which are also referred to as *layers* or *coverages.*
  - The data may be divided into *tiles* so that they can be managed more easily, a tile is uniquely identified by a file name.
  - A collection of themes of vector data covering the same geographic area and serving the common needs of a multitude of users constitutes the *spatial component* of a *geographical database.*
  - The vector method is based on the concept that the features can be identified as discrete entities or objects, this method is therefore based on the *object view of the real world* (Goodchild, 1992).

- Raster data model
  - The method of representing geographic features by pixels is called the *raster method* or *raster data model,* and the data are described as *raster data.*
  - A raster pixel represents the generalized characteristics of an area of specific size on or near the surface of the Earth.
  - The actual ground size depicted by a pixel is dependent on the resolution of the data.
  - Raster data are organized by themes, which are also referred to as layers.
  - Raster data covering a large geographic area are organized by *scenes* of by *raster data files.*
  - The raster method is based on the concept that geographic features are represented as surfaces, regions or segments, this method is therefore based on the *field view of the real world* (Goodchild, 1992).

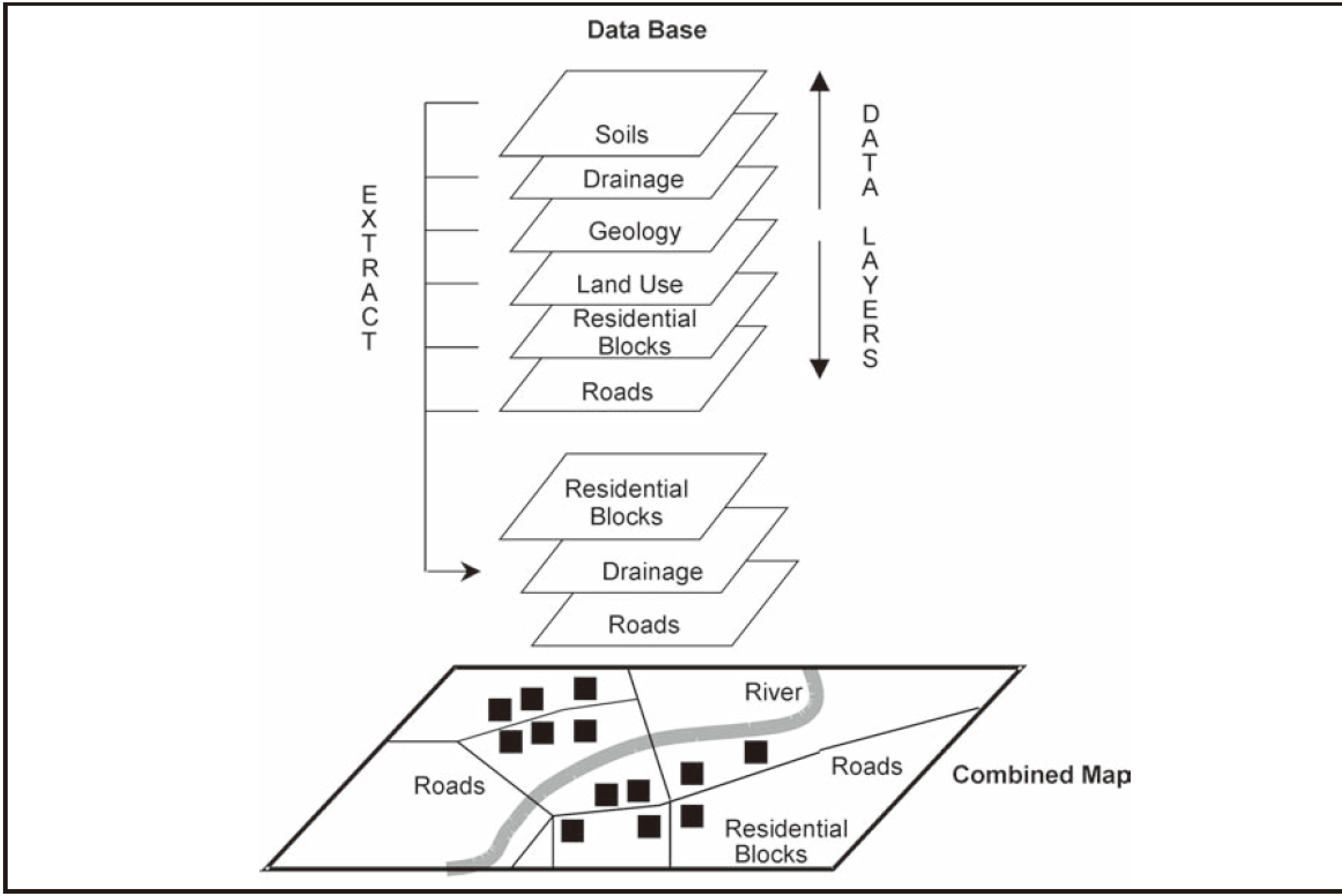


Figure 4.7: The layer based approach.

Levels of data abstraction

Information organization is concerned with the internal organization of data. It represents the user's view of data, *i.e., c*onceptualization of the real world. It is the lowest level of data abstraction, which can be done with or without any intent for computer implementation and it is expressed in terms of *data models* (Peuquet, 1991).

The difference between "data models" and "database models" is:

The vector and raster methods of representing the real world are "data models" and,

The relational, network, hierarchical and object-oriented databases are "database models" --- they are the software implementation of data models

Data structure represents a higher level of data abstraction than information organization in the sense that it is concerned with the design and implementation of information organization. It represents the human implementation-oriented view of data and expressed in terms of database models, this implies that data structure is software-dependent but hardware is not yet a consideration. Data structure forms

Contd...

Box 6: Data structure

the basis for the next level of data abstraction in information system: *file structure* or *file format*. File structure is the hardware implementation-oriented view of data, which reflects the physical storage of the data on some specific computer media such as magnetic tapes or hard disk. This implies that file structure is hardware-dependent.

Descriptive data structures

Descriptive data structures describe the design and implementation of the information organization of non-spatial data. As most commercial implementations of information systems today are based on the relational and object-oriented database models.

Relational data structure: The relational data structure is the table which is formally called a relation.

Object-oriented data structure: Unlike the relational data structure, there is not a formalized object-oriented data structure, this means that different object-orientation implementations have different data structures.

Graphical data structures

Raster data structure: In the raster data structure space is subdivided into regular grids of square grid cells or other forms of polygonal meshes known as picture elements (pixels). There are several variants to the regular grid raster data structure, including: *irregular tessellation* (*e.g.,* triangulated irregular network (TIN)), *hierarchical tessellation* (*e.g.,* quad tree) and *scan-line* (Peuquet, 1991)

Vector data structure: there are many implementations of vector data structures, including: spaghetti - a direct line-for-line unstructured translation of the paper map, hierarchical - a vector data structure developed to facilitate data retrieval by separately storing points, lines and areas in a logically hierarchical manner and topological - a vector data structure that aims at retaining spatial relationship by explicitly storing adjacency information.

The georelational data structure

The georelational data structure was developed to handle geographic data. It allows the association between spatial (graphical) and non-spatial (descriptive) data. Both spatial and non-spatial data are stored in relational tables and entities in the spatial and non-spatial relational tables are linked by the common FIDs of entities.

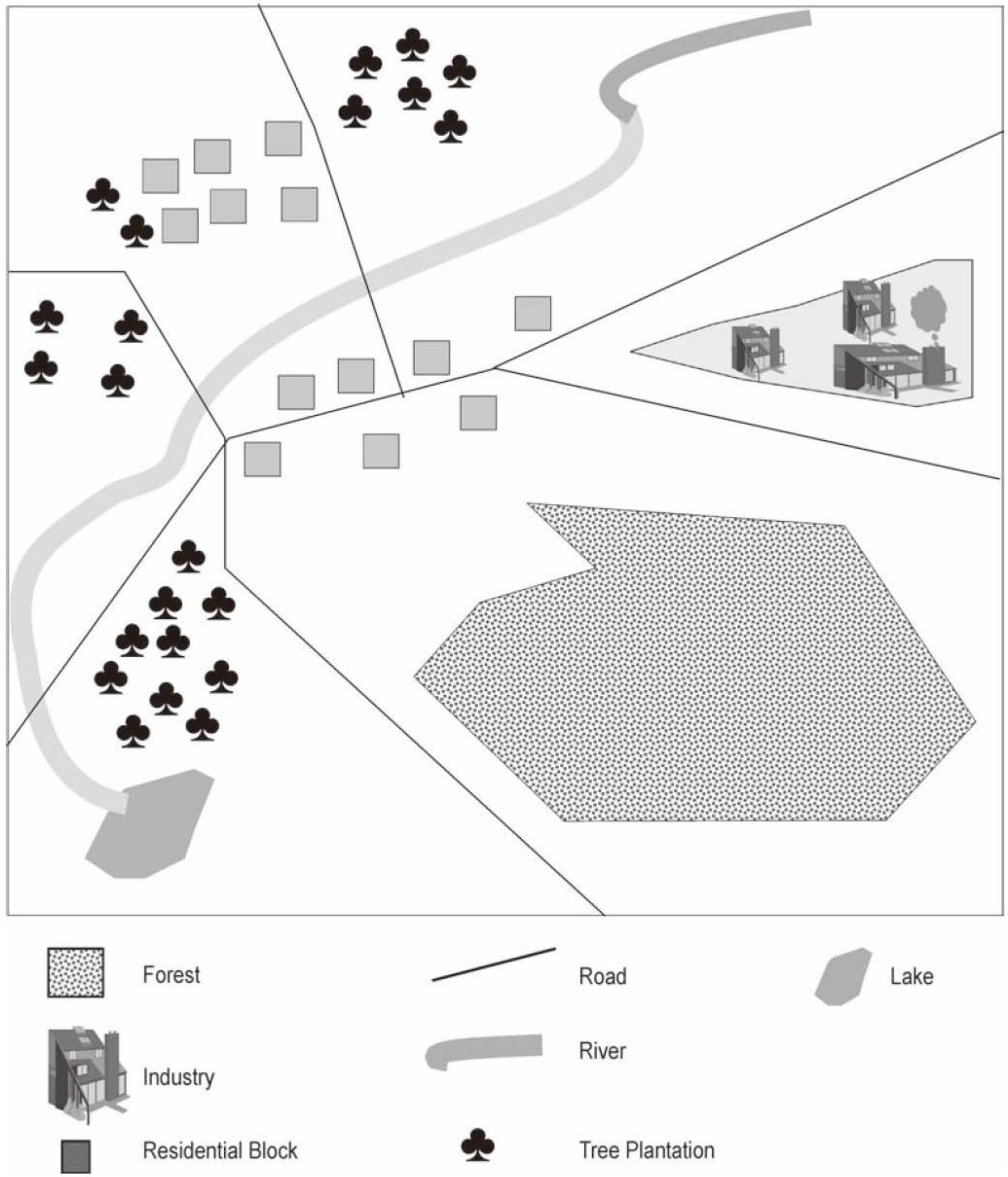


Figure 4.8: The object-oriented approach.

The Relationship Perspective of Information Organization

- Relationships can be *categorical* or *spatial,* depending on whether they describe location or other characteristics.

Categorical relationships

- Categorical relationships describe the association among individual features in a classification system.
- The classification of data is based on the concept of *scale of measurement:*
  - Nominal – a qualitative, non-numerical and non-ranking scale that classifies features on intrinsic characteristics
  - Ordinal – a nominal scale with ranking which differentiates features according to a particular order
  - Interval – an ordinal scale with ranking based on numerical values that are recorded with reference to an arbitrary datum
  - Ratio – an interval scale with ranking based on numerical values that are measured with reference to an absolute datum

- Categorical relationships based on ranking are hierarchical or taxonomic in nature which means that data are classified into progressively different levels of detail.
- The classification of descriptive data is typically based on categorical relationships.

Table 4.3: Example of a classification scheme of descriptive data.

Level I	Level II
1. Built-up Land	1.1 Residential 1.2 Commercial 1.3 Industrial 1.4 Services 1.5 Transportation
2. Agricultural Land	2.1 Crop Land 2.2 Orchards, Vineyards, Nurseries 2.3 Pastures
3. Forest Land	3.1 Mixed Forest 3.2 Evergreen Forest 3.3 Deciduous Forest
4. Water Bodies	4.1 Rivers 4.2 Pond/Lake 4.3 Water Logged Area

Spatial relationships

- Spatial relationships describe the association among different features in space.
- There are two types of spatial relationships (figure 4.9)
  - topological – describes the property of adjacency, connectivity and containment of contiguous features.
  - proximal – describes the property of closeness of non-contiguous features.
- Spatial relationships are very important in geographical data processing and modelling.
  - The objective of information organization and data structure is to find a way that will handle spatial relationships with the minimum storage and computation requirements.

Table 4.4: Point-line-area relationship matrix

	Point	Line	Area
Point	Is nearest to Is neighbour of	Ends at Is nearest to Lies on	Is within Outside of Can be seen from
Line		Crosses Joins Flows into Comes within Is parallel to	Crosses Borders Intersects
Area			Overlaps Is nearest to Is adjacent to Is contained in

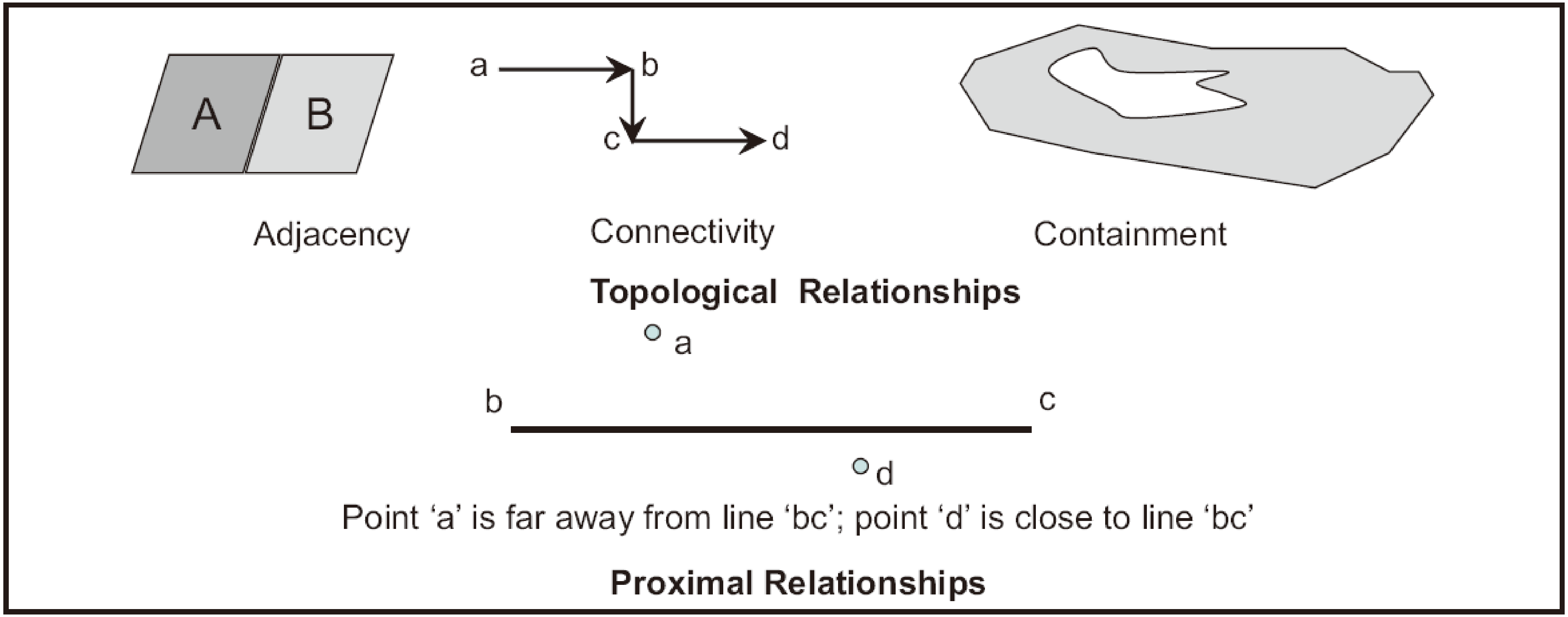


Figure 4.9: Topological and proximal relationships.

THE OPERATING SYSTEM (OS) PERSPECTIVE OF INFORMATION ORGANIZATION

- From the operating system perspective, information is organized in the form of *directories.*
- Directories are a special type of computer files used to organize other files into a hierarchical structure (figure 4.10).
- Directories are also referred to as *folders,* particularly in systems using graphical user interfaces.
- A directory may also contain one of more directories:
  - the topmost directory in a computer is called the root directory
  - a directory that is below another directory is referred to as a *sub-directory*
  - a directory that is above another directory is referred to as a *parent directory*
- A directory is identified by a unique directory name.
- A data file can be accessed in a computer system by specifying a *path* that is made up of the device name, one or more directory names and its own file name.

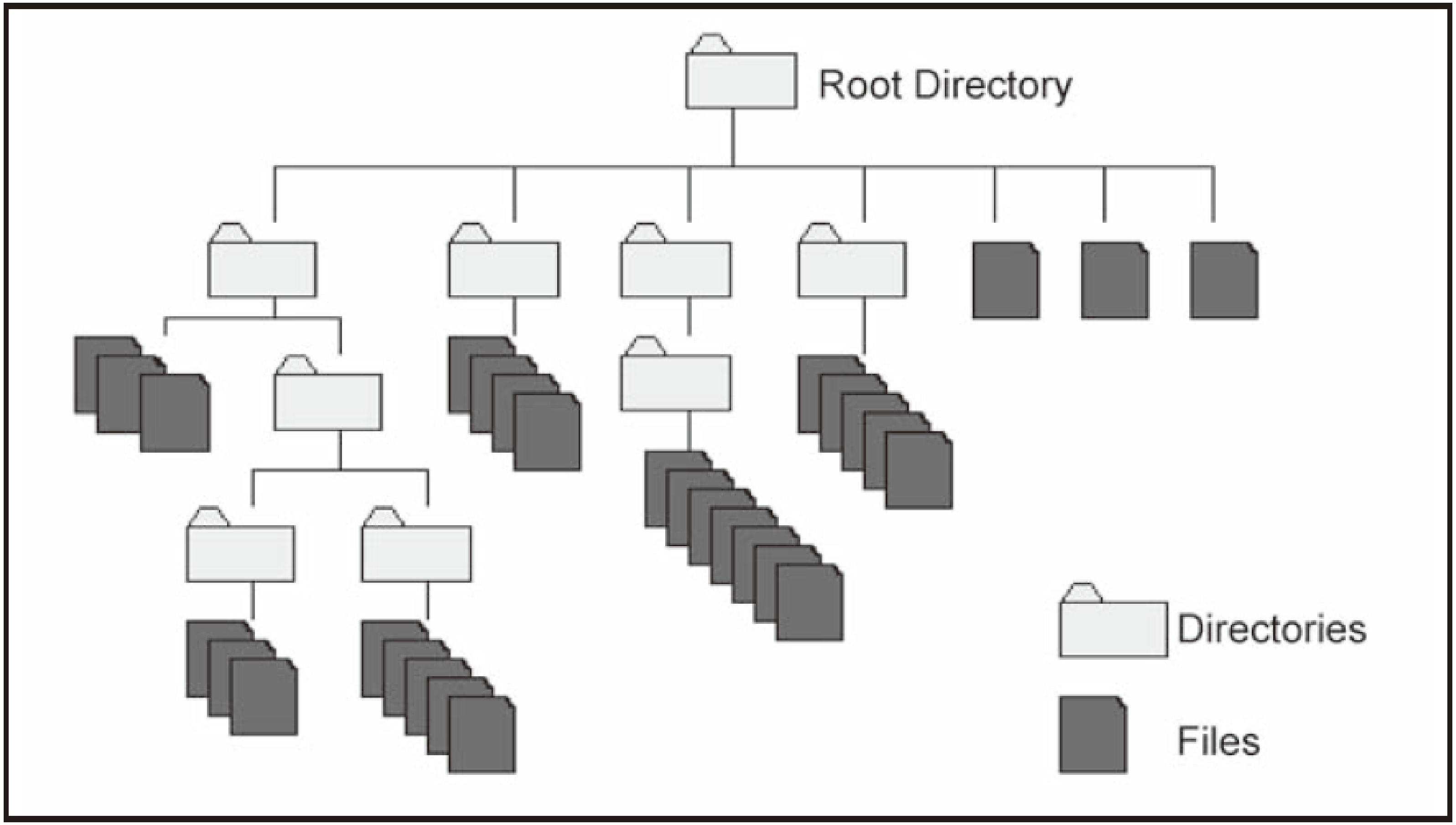


Figure 4.10: Information organization by directories.

- The concept of *workspace* used by many geographic information system software packages is based on the directory structure of the host computer.
- A workspace is a directory under which all data files relating to a particular project are stored (figure 4.11).

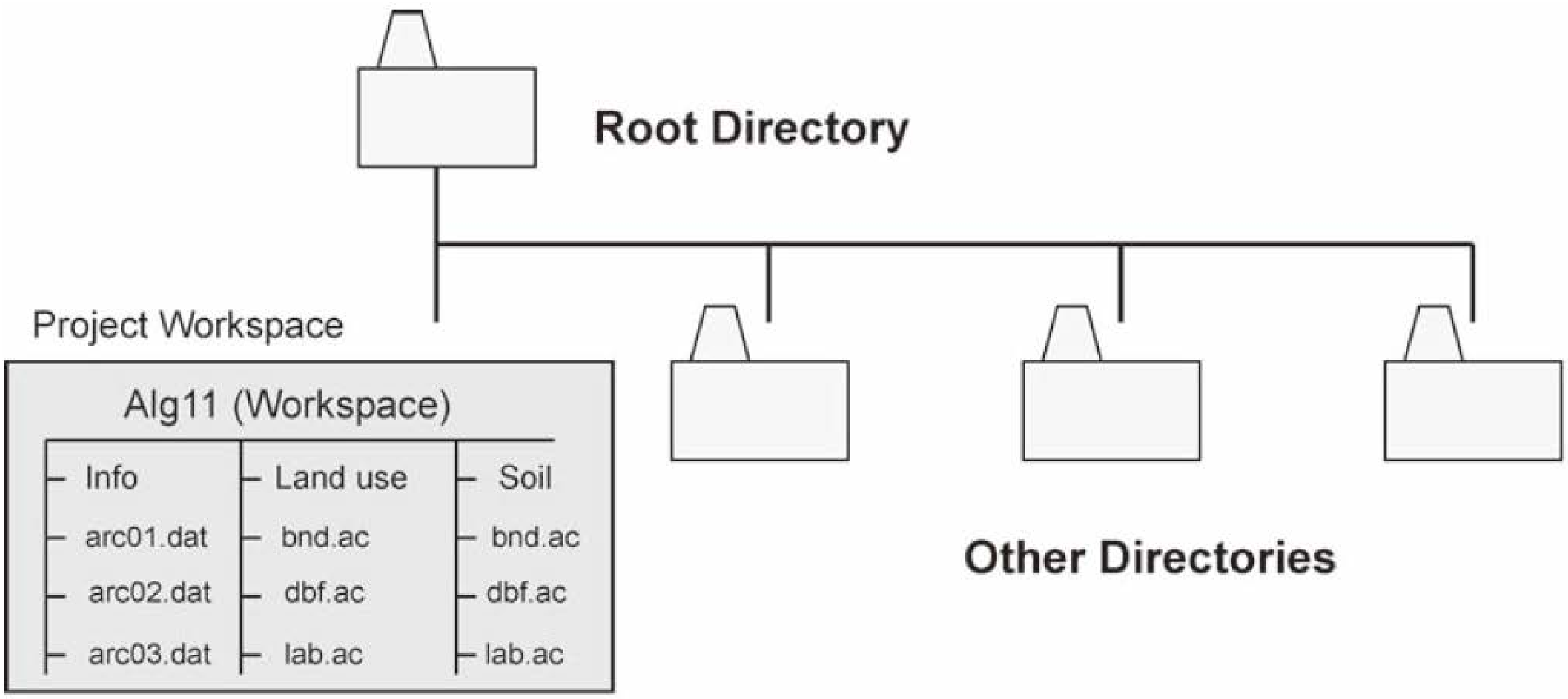


Figure 4.11: Example of a GIS project workspace.

lab

THE APPLICATION ARCHITECTURE PERSPECTIVE OF INFORMATION ORGANIZATION

- Client/server is primarily a relationship between processes running in the same computer or, more commonly, in separate computers across a telecommunication network.
- *Client*
  - The client is a process that requests services.
  - The dialog between the client and the server is always initiated by the client.
  - Client can request services from many servers at the same time.
- *Server*
  - The server is a process that provides the service.
  - A server is primarily a passive service provider.
  - A server can service many clients at the same time.

- From the perspective of information organization, the following five are most important:
  - *file servers* – the client requests specific records from a file; and the server returns these records to the client by transmitting them across the network
  - *database servers* – the *client* sends *structured query language* (SQL) requests to the server; the server finds the required information by processing these requests and then passes the results back to the client
  - *transaction servers* – the client invokes a remote procedure that executes a transaction at the server side; the server returns the result back to the client via the network
  - *Web server* – communicating interactively by the *Hypertext Transfer Protocol* (HTTP) over the Internet, the Web server returns documents when clients ask for them by name

  - *groupware servers* – this particular type of servers provides a set of applications that allow clients (and their users) to communicate with one another using text, images, bulletin boards, video and other forms of media.
- From the application architecture perspective, the objective of information organization and data structure is to develop a data design strategy that will optimize system operation by balancing the distribution of data resources between the client and the server.

DATA – FUNDAMENTAL CONCEPTS

- Data are facts.
- ‘Data’ is a plural and is a broad concept that can include things such as pictures (binary images), programs, and rules.
- Informally, *data* are the things we want to store in a *database.*

SPATIAL – NON-SPATIAL DATA

- Spatial data includes location, shape, size, and orientation; spatial data includes spatial relationships.
- Non-spatial data (also called *attribute* or *characteristic data)* is that information which is independent of all geometric considerations.
- Fundamental differences between spatial and non-spatial data
  - spatial data are generally multi-dimensional and auto-correlated.
  - non-spatial data are generally one-dimensional and independent.

Databases for spatial data

- Difficulties of using standard database system in spatial data.
  - Many different data types are encountered in geographical data, *e.g.,* pictures, words, coordinates, complex objects, but very few database systems have been able to handle textual data.
  - Standard database systems assume the order of records is not meaningful. In geographical data the positions of objects establish an implied order which is important in many operations and often need to work with objects that are adjacent in space.
  - Standard database systems do not allow linkages between objects in the same record type (class).
  - The integrity rules of geographical data are too complex, *e.g*., the arcs forming a polygon must link into a complete boundary.

- More recently, GIS have been built around existing database management systems (DBMS).
- The DBMS handles many functions which would otherwise have to be programmed into the GIS.
- Any DBMS makes assumptions about the data which it handles and to make effective use of a DBMS it is necessary to fit those assumptions.
- There are two ways to use DBMS in a GIS:
  - I. Total DBMS solution: All data are accessed through the DBMS, so must fit the assumptions imposed by the DBMS designer.
  - II. Mixed solution: Some data (usually attribute tables and relationships) are accessed through the DBMS because they fit the model well, while some data (usually locational) are accessed directly because they do not fit the DBMS model.

Repository

- A repository is a structure that stores and protects data.
- Repositories provide the following functionality:
  - add (insert) data to the repository
  - retrieve (find, select) data in the repository
  - delete data from the repository
- Repositories are like a bank vault.
  - *Security:* Repositories are typically password protected, many have much more elaborate security mechanisms.
  - *Robustness:* Accidental data loss is safeguarded against via the *transaction* mechanism.

- Transaction
  - Transaction is a sequence of database manipulation operations.
  - Transactions have the property that, if they are interrupted before they complete, the database will be restored to a self-consistent state, usually the one before the transaction began.
  - Transactions protect the data from power failures, system crashes, and concurrent user interference.

Advantages of a database approach

- reduction in data redundancy
- shared rather than independent databases, which reduces problem of inconsistencies in stored information, *e.g.,* different addresses in different wards for a postman
- maintenance of data integrity and quality
- data are self-documented or self-descriptive, where information on the meaning or interpretation of the data can be stored in the database, *e.g.,* names of items, metadata
- avoidance of inconsistencies, which means data must follow prescribed models, rules, standards
- reduced cost of software development
- security restrictions, which means database includes security tools to control access, particularly for writing.

DATABASE MANAGEMENT SYSTEM (DBMS)

- A *database management system* is a data repository along with a user interface providing for the manipulation and administration of a database.
  - A phone book is an example of a DBMS.
  - A DBMS is like a full-service bank, providing many features and services missing from the comparatively Spartan repository.
- It is a software system, a program (or suite of programs) that is run on a digital computer.
- A few examples of commercially available DBMSs include Codasyl, Sybase, Oracle, DB2, Access, and dBase.

Queries

- Many DBMSs provide a user interface consisting of some sort of formal language.
  - *A data definition language* (DDL) is used to specify which data will be stored in the database and how they are related.
  - *A data manipulation language* (DML) is used to add, retrieve, update, and delete data in the DBMS.
  - A *query* is often taken as a statement or group of statements in either a DDL or a DML or both. Some researchers view queries as read-only operations, no data modifications are allowed.
  - *A query language* is a formal language that implements a DDL, a DML, or both. Examples of query languages include SQL (Structured Query Language), QUEL, ISBL, and Query-by-Example.

DATA MODELS

- A *data model* is mathematical formalism consisting of two parts:
  - A notation for describing data
  - A set of operations used to manipulate that data.
- A data model is a way of organizing a collection of facts pertaining to a system under investigation.
- Data models provide a way of thinking about the world, a way of organizing the phenomena that interest us.
- A major benefit we receive by following a data model stems from the theoretical foundation of the model.
  - From the theory emerges the power of analysis, the ability to extract inferences and to create deductions that emerge from the raw data.

- DBMSs are seen to be composed of three levels of abstraction:
  - *Physical:* This is the implementation of the database in a digital computer. It is concerned with things like storage structures and access method data structures.
  - *Conceptual:* This is the expression of the database designer's model of the real world in the language of the data model.
  - *Logical(User View):* Different user groups can be given access to different portions of the database. A user groups portion of the database is called their view.

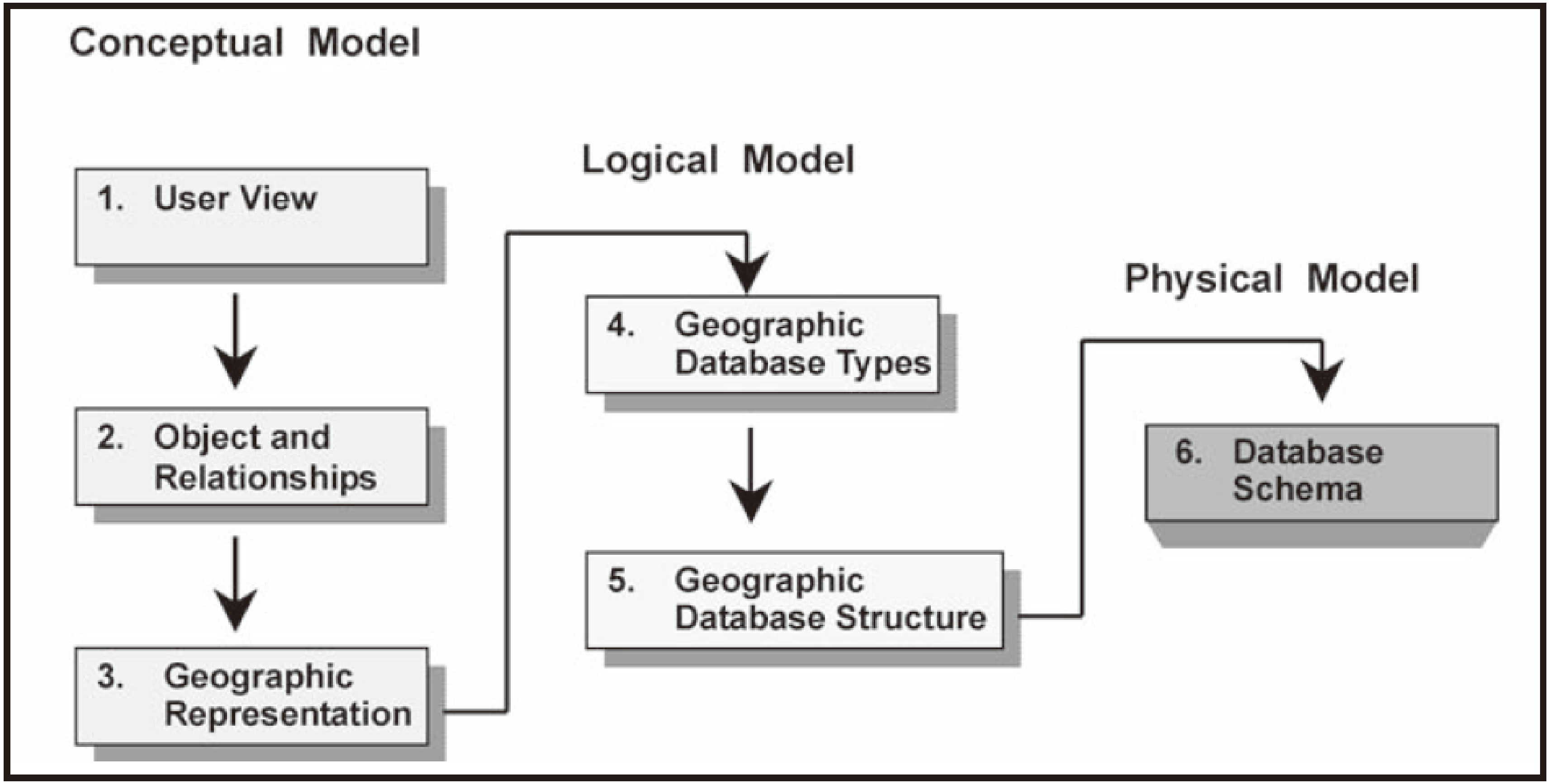


Figure 4.12: Stages in database design.

DATA MODELLING

- Data modelling is the process of defining real world phenomena or geographic features of interest in terms of their characteristics and their relationships with one another.
- Three steps of the data modelling process
  - *Conceptual data modelling*–Defining in broad and generic terms the scope and requirements of a database.
  - *Logical data modelling* – Specifying the user's view of the database with a clear definition of attributes and relationships.
  - *Physical data modelling* – Specifying internal storage structure and fi le organization of the database.
- Data modelling is closely related to the three levels of data abstraction in database design:
  - conceptual data modelling $\Rightarrow$ data model
  - logical data modelling $\Rightarrow$ data structure
  - physical data modelling $\Rightarrow$ file structure

a. Conceptual data modelling

- *Entity-relationship* (E-R) modelling is probably the most popular method of conceptual data modelling.
- It is sometimes referred to as a method of *semantic data modelling* because it used a human language-like vocabulary to describe information organization, involving four aspects of work:
    - identifying entities – defined as a person, a place, an event, a thing, etc.
    - identifying attributes
    - determining relationships
    - drawing an *entity-relationship diagram* (E-R diagram)

b. Logical data modelling

- Logical data modelling is a comprehensive process by which the conceptual data model is consolidated and refined.
- The proposed database is reviewed in its entirety in order to identify potential problems such as;
  - irrelevant data that will not be used; omitted or missing data; inappropriate representation of entities; lack of integration between various parts of the database; unsupported applications; and potential additional cost to revise the database.
- The end product of logical data modelling is a logical schema.
  - It is developed by mapping the conceptual data model (such as the E-R diagram) to a software – dependent design document.

c. Physical data modelling

- Physical data modelling is the database design process by which the actual tables that will be used to store the data are defined in terms of:
  - data format – the format of the data that is specific to a DBMS.
  - storage requirements – the volume of the database.
  - physical location of data – optimizing system performance by minimizing the need to transmit data between different storage devices or data servers.
- The end product of physical data modelling is a *physical schema.*
  - It is also variably known as *data dictionary, item definition table, data specific table* or *physical database definition.*
  - It is both software – and hardware specific, this means the physical schemas for different systems look different from one another.

d. Process modelling

- Process modelling is the process-oriented approach, as opposed to the data-oriented approach, of information system design.
- The objective is to identify the processes that the information system will perform.
- It also aims at identifying how information is transformed from one process to another.
- The end product of process modelling is a *data flow diagram*(DFD).
- In the context of information system design, process modelling is one of the methods of *structured business function decomposition* used to determine user requirements in conceptual modelling.

- The components of data flow diagram
  - process – it represents the transformation of data as they flow through the system: data flow into a process, are changed, and then flow out to another process or a data store.
  - entity – the basic definition of an entity is similar to that for E-R modelling and it represents the initial source and final destination of data in a DFD.
  - data store – a temporary or permanent holding area for data.
  - data flow – the connection between processes and data stores along which individual entities or collection of entities flow.