

DATA, INFORMATION AND KNOWLEDGE

What are data?

Any set of symbols (alphabetical, numeric, graphical, etc) that has been consciously selected and organized for the purposes of expressing an idea and information.

Among such symbols are:

- A set of numbers arranged as a list or in a table;
- The sentences in a textbook;
- The words displayed by a computer on its monitor;
- The words and images on an advertisement billboard;
- Words exchange between two persons on a telephone line;
- The photograph of a car accident scene;
- Music transmitted on a radio channel.

Data versus Precepts

Precepts refer to the sensations we get when we use any of our five senses - hearing, sight, smell, taste and touch. However, such sensations are not data themselves. But they can be transformed into data when we use symbols to describe the sensations.

Data versus information

Information is usually extracted from the knowledge possessed by someone or some community, and then communicated through data to other people or communities. Which can then be interpreted to serve as knowledge.

Information and Knowledge

Knowledge is the extent of conversant or familiarity by particular individuals, communities, or mankind as a whole, with certain facts, truths, principles or subjects. Knowledge serve as a pool from which people can extract specific truth, facts and ideas (i.e., information)

Data sources as information sources

A data source is any medium that has been used to convey data symbols.

DATA, INFORMATION AND KNOWLEDGE MANAGEMENT

Data organization

Data organization is concerned with the selection, combination, arrangement and formatting of symbols such as words, numbers and images to form data.

Data management

Data management refers to the various activities that must be performed by individuals or organizations to ensure that only useful data are created for conveying information.

The data management cycle

- (i) Data policy making
- (ii) Data definition and structuring
- (iii) Data collection and creation
- (vi) Data validation and quality control
- (vi) Data analysis and summarization
- (v) Data storage and retrieval
- (ix) Data archiving and protection
- (vii) Data communication and transfer
- (viii) Data display and presentation

Data policy-making: refers to the process of making decisions about what data should be created and for what purpose.

Data definition and structuring: refers to the process of determining the formats in which data will be created, acquired, stored and transferred.

Data collection and creation: refers to the process of actually creating or collecting data in line with the data policies in (i) above, and the defined data definitions in (ii) above.

Data validation and quality control: refers to the process of verifying and certifying that the data that have been created, collected or acquired from other sources meet certain minimum standards of accuracy and consistency.

Data analysis and summarization: refers to the different processes for comparing and analyzing data, and for computing, aggregating and summarizing data into other data such as statistics, charts, maps, abstracts and summaries.

Data storage and retrieval: refers to the physical storage and retrieval of data from different types of media, such as paper and computer media.

Data archiving and protection: refers to the process of providing either special protection or secure long-term storage for extremely valuable data.

Data communication and transfer: refers to the process of transporting data across space and time through different media such as documents and computer and telephone networks.

Data display and presentation: refers to the process of displaying and presenting data for direct human perception and interpretation.

Information management

Information management refers to activities and processes for facilitating the flow of

information from various information and knowledge sources to people who desire the information.

Knowledge management

knowledge management entails activities and processes that facilitate the growth, integration and application of ideas from various subjects and disciplines to solve human problems.

INFORMATION SYSTEMS FOR DATA MANAGEMENT

Information systems

What is an information system? An information system may be described as a set of interdependent activities designed to provide data and/or information to people.

All systems share the following features:

- Goals or objectives: These give direction to the activities of a system.
- Sub-systems: These are the mutually dependent components of a system.
- Environment: This comprises all other systems that do not belong to a particular system, and with which the system shares boundaries.
- Activities: A system usually must perform some physical activity in order to achieve its goals and objectives.
- Products. Communication and information activities are those that involve the creation, storage, communication, etc., of data and/or information.
- Resources: These are required for performing physical and information activities. The resources usually include tangible and intangible things, such as space, equipment, energy, finance, time and human effort.

Data management activities of information systems

information systems provide the settings within which data and information management activities are performed. These activities are often grouped into five major kinds:

- (a) Input;
- (b) Processing;
- (c) Storage;
- (d) Output;
- (e) Communication.

Data input activities involves the inflow (i.e., input) of data to the system from other systems.

Data processing activities involves the transformation of data into more informative data, by aggregating, desegregating, summarizing and computing the data.

Data storage entails storing the data temporarily or permanently before, during and after processing.

Data and information output entails sending out processed data and information from the system to other systems

Data and information communication entails the sending and receiving of data and information among the components of a system, as well as between the system and other systems in its environment

Information systems resources for data management

- **Data**, as raw materials for producing information, and also sometimes as output;
- **People**, such as data collector and data entry clerks, supervisors, managers, etc.
- **Physical facilities and equipment**, such as buildings and offices, computers, telephone lines, fax machines, etc.
- **Organizational structure and policies**, to link and control interactions between people, facilities, activities, data, etc.
- **Energy**, such as electricity to drive the machines, and human energy and motivation to drive people to work for the system;
- **Time**, because activities take time.

LANGUAGES FOR DATA ORGANIZATION

Natural languages for data organization: Natural languages are the standard vehicles for social communication within societies.

Natural language symbols, rules and usages

Accordingly, written natural languages often have:

- (a) **Alphabets** and other special symbols (e.g. comma), which are the symbols that can be used for creating and recording data to express information.
- (b) **Spelling rules** that guide how alphabetic symbols may be combined to form words;
- (c) **Grammar rules** that guide how words could be combined to form sentences;
- (d) **Rules, conventions or styles** of structure that guide how sentences can be strung together to form paragraphs, and how paragraphs can be arranged when writing for different audiences;
- (e) **Conventions** or rules as to the likely meanings that can be implied by different combinations of symbols, words, etc.
- (f) **Rules and conventions** as to how to pronounce combinations of symbols, such as words, in the language.
- (g) **Dictionaries** specifying the recognized symbols, spelling, grammar and usages in the language, as well as the meanings of the symbols and combinations of symbols.

The scientific and cultural disciplines and professions have the following additional language symbols, rules and usages:

- (a) Numeric symbols (0, 1, 2 ...), as well as rules on how the symbols can be combined to express quantitative information
- (b) Pictorial and graphic symbols, such as lines, dots, and the like, that can be used to create image data, such as maps, cartoons, line graphs, drawings, paintings, etc

(c) Conventions as to the best way to arrange numerical and pictorial symbols in data tables, charts, mathematical formula, maps, dictionaries, indexes and abstracts, etc.

(d) Conventions as to how to arrange or organize data symbols, such as words, numbers, pictures, etc., its various formats, such as resumes, technical reports, application letters, etc.

Scientific methods for representing quantities

1. decimal number and counting system
2. scientific notation for expressing numbers.
3. binary number system

DATA REPRESENTATION IN THE COMPUTER

How data are input and represented in computers

Data are input to a computer through input devices connected to the computer's central processing unit (CPU). Among such devices are:

- **Keyboard**, for entering alphabetical, numerical and special characters;
- **Mouse** (for indicating and selecting options from menus displayed on a computer monitor);
- **Scanners** (for taking pictures of paper documents);
- **Digital cameras** (for taking photographs of objects); microphones (for sound and voice), etc.

Binary numbers and signals are represented with a sequence of electrical impulse and signals that can be used for storing and processing sound and image or textual data.

DATA PLANNING AND POLICY MAKING

Planning for data

Planning invariably aims to answer, before the commencement of any activity, the following five basic questions:

- Why should the activity be done?
- Where should the activity be done?
- When should the activity be done?
- What tasks must be done in the activity?
- Who should do which tasks in the activity?

Planning for data management requires that one finds prior answers to such questions as:

- **What** information should be **produced** and why?
- **What, How, When** data should be **collected**, stored, analyzed, **communicated**, etc., toward producing the required information?
- **Who** should be responsible for the different activities in the data management cycle?

Information resource management

The main ideas championed by IRM professionals are:

- That information resources include information and data, as well as the technologies

for creating, acquiring, storing, processing and communicating data, and for producing and using information.

- That information and data are very valuable because they cost money to acquire, process and use, and because they can provide value.
- Managerial principles must be devised for managing information resources.
- That information resources must be managed by qualified information professionals.

Data and information policies

A policy is a set of statements that describe the circumstances, objectives and constraints to be followed by people in an organization or information system concerning a particular activity or issue.

Data policies for data collection and acquisition will often provide guidelines for answering the following questions:

- (a) **Why** should certain types of data be collected, acquired or created?
- (b) **What** specific **types** of data should the organization or information system collect, acquire or create?
- (c) **What symbols** should be used to create data, and/or in what format and from what media should data be collected or acquired?
- (d) **From what sources** should data be collected or acquired?
- (e) **When** and **how frequently** should data be acquired, collected or created?

Data policy manuals

A data policy is a basic guide to action that sets the boundaries within which data management activities are to take place.

The following are some of the advantages of policy manuals. Policy manuals

- **Provide plans of action** in respect of most data management issues, activities, events and problems;
- **Forces** people within an organization or information system to **think** carefully through whatever they want to do with data;
- **Enable** people within an information system to **perform** their respective data management **functions** and tasks in a standardized, consistent manner;
- Often **establish** data **quality control points** which often help to ensure data quality;
- Can be used as **material** for **training** and job **orientation** of new employees;
- Help **improve** data management activities within an organization and information system.

Data planning and policy-making requirements

demand that people in the organization or information system

- (a) Accept the importance of data, information and knowledge activities.
- (b) Establish specific policies to guide and regulate various data management activities within the organization or information system;

- (c) Use the policies as guide for determining what data to collect acquire or create;
- (d) Evaluate either continuously or periodically whether the established policies are being followed

DATA DEFINITION AND STRUCTURES

Unstructured versus structured data

Structured data are data that have been uniformly or logically defined and arranged to facilitate human and machine processing and understanding. On the other hand, unstructured data are those that have not been so defined and arranged.

Importance of data structure

Data that are logically structured are more likely to convey better information than data that are not logically structured.

Guidelines for pre-defining data to be created/collected

- (i) Make sure you understand the reason why data are being collected.
- (ii) List all the pieces of data to be created or collected, such as name, address, birth date, etc.
- (iii) Provide adequate space for recording each of the listed pieces of data,
- (iv) Provide guidelines to ensure that all those from whom data are to be collected provide data in the same way.

Entities, data records and data fields

A class of persons, objects or activities e.g. is often referred to as an entity. Entities are described in terms of their attributes, thereby creating data, the data on each member of an entity is a record (of data) on the member. each separately meaningful piece of a record is often referred to as a data element or a data field.

DATA ARRANGEMENT, GROUPING AND MODELING

Data modeling and data models

A specific arrangement or grouping of data or data records is usually referred to as a 'view' or 'model' of the data, or a data model.

Types of data structures or models

Hierarchical data structures/models

A hierarchical data model is one whereby data records are grouped or interrelated in terms of being super-ordinate or subordinate to one another.

Network data structures/models

The network approach is based on the idea that data are, or can be, related in many different ways at the same time.

DATA CAPTURE, ACQUISITION AND COLLECTION

Primary and secondary data

Primary data are data that are being created or recorded for the first time. Secondary data, on the other hand, are data that someone had previously created, collected, stored or published.

Data creation, collection, acquisition and capture

Data may be either created, collected, acquired, or captured.

- (a) **Creation** of primary data by or within the organization or system itself,
- (b) **Collection** of primary data from outside the system, usually through the use of various instruments, forms and questionnaires;
- (c) **Acquisition** of secondary data from outside the system, from either paper-based or computerized sources;
- (d) **Capture** of data automatically by a system through various computer input devices such as scanners, cameras, microphones, etc

Data creation

Data is created when people write or record data by hand or with devices such as typewriters, keyboards, scanners, cameras, etc.

Data collection

Data collection is often undertaken by a system or organization during laboratory or field studies.

Data capture methods and devices

Data capture is a term often used to refer to the automatic capture and recording of data by mechanical and computer input devices.

Data acquisition

Data acquisition is usually undertaken to obtain secondary data from various sources. Among the major sources of secondary data are:

- (a) Books;
- (b) Newspapers;
- (c) Technical reports;
- (d) Statistical bulletins
- (e) Annual reports of organizations;
- (f) Computer files or databases
- (g) The Internet

Data collection activities and instruments

Data collection entails four stages of activities:

- (a) Defining what data to collect.
- (b) Selecting or designing the instrument or device to use for collecting the data.
- (c) Using the instrument of device correctly to measure the variables.

(d) Recording correctly the measurements made with the instrument.

DATA QUALITY CONTROL - FUNDAMENTAL CONCEPTS

What is data quality control?

Data quality control refers to the processes and methods by which the accuracy, validity and reliability of data is ensured at the different stages of the data management cycle.

Accuracy of data

Data accuracy refers to the preciseness or exactness with which data describes or expresses information about an entity.

Validity of data

Data validity refers to the extent to which a set of data expresses accurate and true information about an entity.

Accuracy and reliability of instruments

An instrument must, when used correctly, be able to do what had been designed to do accurately and reliably.

Accuracy of data collection instruments

A data collection instrument is considered accurate if it can be used to measure a variable with little or no error.

Reliability of data collection instruments

A data collection instrument will be considered reliable if it reports the same data every time it is used to mention the same thing.

Data collection instruments and procedures

A good instrument for collecting data on an entity is an instrument:

- (a) That is accepted by experts in a discipline or profession as the standard instrument for collecting valid data on the entity; or
- (b) That, although not yet considered standard, had previously been used by different people to obtain valid data on the entity; or
- (c) That experts agree will collect valid data on the entity when used; or
- (d) That has been tested to be accurate and reliable.

The procedure for using the instrument to measure and record data should be:

- (i) Should be specified as a sequence of tasks that can be easily understood and performed;
- (ii) consistent each and every time that the instrument is used;
- (iii) Written so that people who want to use the instrument can confirm the steps of the procedure.

Data quality control - context and strategies

Planning of data quality control

Three main tasks are involved when planning for data quality control, as follows:

- (a) Establishing data quality policies and standards
- (b) Adopting or developing the language to be used for creating and recording the data
- (c) Developing data collection instruments and procedures.

Data quality control before, during and after

Data quality can be controlled before data are actually collected by implementing strategies for producing valid, accurate and reliable data collection instruments and procedures

Data quality can also be controlled during data collection by ensuring that the instruments and procedures are used consistently as had been designed

Finally, data quality can be controlled after data had been collected by cross checking the data for errors.

Textual data

There are two main methods for cross-checking textual data:

- (a) A spelling check
- (b) A check of the grammatical structure of the sentences and paragraphs

Numeric data

For numerical data, checks can be performed to determine whether the values of the data are likely to be inaccurate.

Structured data

checking for data accuracy can be done by finding out if the data matches the given structure.

DATA STORAGE MEDIA AND ORGANIZATION

Data storage and retrieval

Data storage media

- (i) Human brain
- (ii) Paper
- (iii) Microforms: Microforms is the general word used to describe all miniature but non-computerized storage media such as film rolls, film slides, microfiche, etc.
- (iv) Computer: Includes media tapes, disks, diskettes, compact disks (CDs), smart cards, mobile phone recharge cards, etc.
- (v) Audio and video media:
- (vi) Databases

DATA STORAGE IN COMPUTER DATABASES

What is a database?

A database is as a set of computerized and linked files that have been created with a database management system for storing the data of an organization or information system.

Software for creating and managing databases

Among them are: Microsoft Access, dBase V for Windows, Visual FoxPro, Paradox, Oracle, DB2, etc.

Form: This refers to a pre-defined format for entering data into one or more data tables in a database.

View: A view is a pre-defined way of viewing or displaying some or all the records and fields in a data table.

Query: A query is a statement that instructs a DBMS to find and display from a database all data that meet some criteria.

Report: A report is either a selective or a comprehensive list or summary of the data in records of one or more data tables in a database.

Related database tables: These are two or more tables the data from which are linked together such that records and fields in the tables can be viewed, processed or reported together as if the records and fields belonged to a single table

Database system: A database system comprises of a DBMS, a database created with the DBMS, other computer programs that are used with the database, and the people who operate or manage the database

CREATING AND USING DATABASES: COMMON TASKS

Common tasks in creating and using databases

Five of the most common tasks:

- (a) Creating a database.
- (b) Creating data tables.
- (c) Updating records in data tables.
- (d) Sorting the records in the tables.
- (e) Creating and using indexes to tables.
- (f) Displaying records and fields from tables.

DATA AND INFORMATION RETRIEVAL CONCEPTS

Context of information retrieval

Information retrieval is the process by which a person searches through a data store for some data or information.

DATA PROTECTION AND ARCHIVING

DATA ANALYSIS AND SUMMARIZATION

Data analysis and summarization

Data analysis is the process of investigating and identifying the attributes or characteristics of a particular set of data.

Data summarization is the process of reducing some voluminous data into less voluminous data.

Data aggregation and disaggregation

Data aggregation refers to the process of merging two or more originally separate sets of data into a single set of data

Data disaggregation refers to the process of breaking or decomposing some data into two or more separate data sets

Summarizing and analyzing textual data

(i) **Extract important sentences:**

(ii) **Write an abstract:** This method involves writing new but fewer and shorter sentences to convey as much information as possible from the original paragraph.

(iii) Select some or all key terms or phrases

(iv) Select some or all of the key words:

(v) Coin an appropriate subject phrase or topic

Analyzing structured and numeric data

(a) Frequency counts and tables of data

(b) Bar and pie charts of data

(c) Sums and averages of numeric data

DESIGNING AND IMPLEMENTING INFORMATION SYSTEMS FOR DATA MANAGEMENT

The system development cycle comprise of the following seven stages of activities:

(1) Ascertaining users' information requirements

(2) Feasibility study of candidate new systems

(3) Detailed analysis of the old system

(4) Designing the new system

(5) Implementing of the new system

(6) Operating of new system

(7) Monitoring and evaluating the new system