Section 5 Information Processing

# Objective 5.1 distinguish between data and information

**Data**

This is raw unprocessed facts e.g. 32, 28, 36. This data can represent anything and therefore has no meaning.

**Information**

This is processed data e.g. Jane scored 32, Phillip scored 28 and Barry scored 36. This information tells us how well or poorly the student performed. Information has meaning.

# Objective 5.2 explain the characteristics and functions of Information Processing

Information Processing

Basically information processing is using computers to process data into information. Information Processing takes place when a machine or processor interacts with the outside world and acts on the inputs its receives e.g. an automatic electric kettle which senses when the water temperature reaches boiling point and switches off; or a car production line sense when a car body is in a certain position to weld together the relevant part of a car. Information processing is only valuable if the information can be stored and retrieved quickly, accurately and efficiently and cannot be changed accidentally. The basic sequence of events in processing information includes (1) input, (2) processing, (3) storage and (4) output.

**Forms of Information Processing**

**Commercial data processing**

This refers to the use of computers in businesses to automate or perform administrative tasks such as:

* Calculate payroll
* Record and track orders made by customers
* Process sale transactions
* Preparing and sending invoices to customers
* Making payments for goods and services
* Keeping track of income and expenditure
* Making budgets
* Sending correspondence to suppliers and customers
* Designing packaging and advertising
* security

**Industrial data processing**

This refers to the use of computers to carry out all the tasks listed above in addition to:

* controlling the processes used to manufacture products using control systems
* using CAD to design products
* using CAE and CAM to develop and test products

**Control systems**

This is the use of computers to control machinery automatically. Some examples are:

* using computers to manufacture chemicals or pack goods in a factory
* using robots to manufacture goods such as cars
* using computers to control signalling systems e.g. on railways
* using computers to control transport vehicles e.g. autopilot on aircrafts, driverless trains, fuel injection systems for car engines
* using computers to control electronic equipment in the home e.g. video recorders, digital cameras, washing machines, microwave ovens, cell phones, etc.

There are two main types of control system:

1. **Automation systems**: these are used to carry out straightforward tasks automatically such as running through a program on a washing machine. They do this without human involvement. Once it is switched on it follows a set of predetermined actions to the end.
2. **Process control systems**: these are used for complex operations. They constantly check the state of play and change the actions in response e.g. in manufacturing actions are changed depending on feedback from sensors. The data is captured automatically through input devices called sensors. Sensors measure things like temperature, humidity, light, sound, movement and pressure. It converts the measurement into electronic signals that it sends to the controller (processor). Signals can be sent as a steady stream or it can take readings at intervals. Software in the controller processes the data from the sensor and determines what action to take. It then sends instructions to other special devices such as motors and switches, which are part of the machine that is being controlled.

**Scientific data processing**

This is the use of computers to collect data during experiments. Sensors are used to capture data over a period of time. The data collected is used to make connections between the data. Statistical methods are used to analyze the figures collected and the results are display as charts.

**Information retrieval**

Information retrieval is when data stored on a computer system is searched and the relevant information is found. To ensure that information retrieval is done efficiently, the data should be

* accurate
* up to date
* structured in such a way that makes it possible to search for specific data
* stored on a suitable storage medium.

**Information Management**

Information Management is the task that ensures that all data is input accurately, stored correctly, retrieved easily and output in a way that provides useful information.

# Objective 5.3 identify the sources of data in specified application areas;

**Documents used for information processing**

**Source documents**

A source document is one that contains data that is going to be input into a computer system. Once the data has been entered it should be filed away safely because the data may have to be checked again.

**Human-readable documents**

A human-readable document is any document that needs to be read by humans e.g. forms that are filled in by hand and has to be read before the data is entered on a keyboard.

**Machine-readable documents**

A machine-readable document is one on which the data can be ‘read’ by an input device, such as a scanner or mark reader, directly into the computer system. Examples of machine-readable documents are multiple choice answer sheets and barcodes on labels of products.

**Turnaround documents**

A turnaround document is one that is printed by a computer system but is later used to input new data into the same system. E.g. a tax return form with the employee’s personal information pre-printed on it. After the rest of the form is completed by the employee the new data is entered so that the tax return can be processed.

# Objective 5.4 describe methods of validation and verification of data;

# Objective 5.5 identify appropriate verification and validation checks given a particular scenario;

**Data verification** is a process whereby a person checks that data has been correctly entered into a computer system.

**Data validation** is any check carried out by a computer system when data is entered, to identify data that cannot be correct.

The difference between validation and verification is that verification is manual i.e. carried out by humans whereas validation is automatic; it is carried out by software.

Verification methods

When data is being entered, two types of errors can occur:

**Typographical errors**: typing errors that are made when the wrong key is hit.

**Transposition errors**: when numbers or characters are entered in the wrong order e.g. entering 2345 instead of 2435.

To reduce these errors the person entering the data can check that the data on the screen matches the data on the source document.

Another method used for verification is **double entry verification**. In this method, a second person enters the same data and the computer system checks each data item entered the second time against the first data entry. If they do not match a warning message is displayed and the person has to confirm which is correct the first entry or the second.

Validation methods

There are many validation checks and suitable ones must be chosen depending on what data is being entered.

**Range checks**

These check whether numerical data is within expected limits e.g., if you are asked to enter someone’s age, the number should be between 0 and 110 years; or if you are asked when you joined an organization, the date cannot be a date in the future.

**Data type checks**

These check whether the data is of the correct type, such as number, date or text. E.g. age is numerical, so a data check would notice if a letter was entered.

**Inconsistency checks (consistency checks)**

These check one piece of data against another. E.g. if data includes both gender (M or F) and title (Mr., Mrs. or Miss). If someone has entered M and Miss, then the two data items are inconsistent.

The main purpose of data validation is to check that the data is reasonable. It should identify some errors that will not be picked up by data verification. It will not spot errors such as incorrect spelling of names. To make sure that data entered is as accurate as it can be, both data verification and data validation checks should be used.

# Objective 5.6 describe various methods of file organization and access;

# Objective 5.7 select appropriate file organization for particular applications;

**File organization**

Files stored on storage media can be organized in a number of ways. There are advantages and disadvantages to each type of file organization, and the method chosen will depend on several factors:

* How the file is to be used;
* How many records are to be processed each time the file is updated;
* Whether individual records need to be accessed quickly.

**Types of file organization**

**Serial files**

In a serial file, the records do not follow each other in any particular order. If another record needs to be added, it can be added to the end of the file.

**Sequential files**

Sequential files are like serial files except that the records are held in a sequence e.g. if you decide to order pupil files in ascending or descending order of student number. One of the fields is used for the ordering. This field is called the sort key e.g. it is very common to sort a file of personal details using “surname” as the sort key. This is also known as **sequential field ordering**.

**Random files**

Random files are stored in any order on the disk surface. The OS keeps a map of the disk surface and the read/write heads use this map to find the data. The data is found without the whole disk having to be read. Random access allows data stored on disk to be found extremely quickly, and the records are accessed regardless of the order in which they were stored.

**Indexed-sequential File Organization**

Each record of a file has a **key field** which uniquely identifies that record.

An **index** consists of keys and addresses (physical disc locations).

An **indexed sequential file** is a sequential file (i.e. sorted into order of a key field) which has an index.

A **full index** to a file is one in which there is an entry for every record. e.g.

|  |  |
| --- | --- |
| Key | Address |
| 2351 | 781040 |
| 2352 | 781041 |
| 2353 | 781042 |
| **2354** | 781043 |
| 2355 | 781044 |
| 2356 | 781045 |
| 2357 | 781046 |
| 2358 | 781047 |
| 2359 | 781048 |
| 2360 | 781049 |

Indexed sequential files are important for applications where data needs to be accessed.....

* sequentially
* randomly using the index.

An indexed sequential file allows fast access to a specific record.

Example: A company may store details about its employees as an indexed sequential file. Sometimes the file is accessed....

* sequentially. For example when the whole of the file is processed to produce payslips at the end of the month.
* randomly. Maybe an employee changes address, or a female employee gets married and changes her surname.

An indexed sequential file can only be stored on a random access device  
e.g. magnetic disc, CD.

**Accessing files**

**Serial access**

This means the records are accessed in the same order as they were stored. To read a serial file, a computer has to read each record until it reaches the one required.

**Sequential access**

This means reading or writing data consecutively. With sequential access the records are in order. Therefore if just one record is required, the method is slow but if all the records are required, it is very fast. Magnetic tape is accessed via this process. A batch processing system, such as a payroll system uses sequential access where the record for each employee is accessed in sequence and the appropriate pay calculated.

**Direct access or random access**

With a direct or random access files it is possible to go directly to a record without having to look at any other records. You can access random access files directly. Magnetic disks and optical disks use direct or random access.