2.1 COLLECTING DATA

Collecting data includes all the activities involved in deciding what data to collect and the collection techniques used so that it can be entered into the computer or other information system.

Hardware

Input devices for collection of data:

- Pointing devices
- Scanners
- Digital cameras
- Video cameras
- Microphones
- Keyboards
- OCR devices.

Emerging trends in hardware collection devices:

- Data entry by voice
- Portable scanning and electronic communication
- Mobile phones connecting to computer networks
- Digital cameras capturing images for direct input into computer hardware.

Software

Operating systems:

- MS-DOS
- Win '95/2000
- Mac-OS
- UNIX
- Linux
- GUI interface and WIMP environment
- DOS interface.

Non-computer tools

Examples include:

- Data counters, eg. traffic monitors
- Literature
- Surveys
- Interviews
- Manual observation/data forms.

Social Issues Related to Collecting Data

When data is collected and entered into an electronic database, it becomes very easy to access and reproduce. If, for some reason, this data is inaccurate or duplicated the problem is escalated further than if the data were not so easily accessible.

Following are some of the social issues that can arise with the introduction of computerisation for data collection.

Data bias occurs when the attitude of the person collecting the data affects what data is included in the sample. For example, collecting data from mainly white, Anglo-Saxon males about a general population issue would result in biased data. The data would not be truly representative and thus not really accurate.

Data integrity refers to the accuracy of the data. Data may be misspelt, entered incorrectly or be inaccurate for some other reason.

Copyright is the right of an author or the originator of an idea or other intellectual property to make decisions on what happens to their work. No one else is able to copy it without the permission of the author. Electronic communication makes breaching copyright much easier and probably more prevalent than before computerisation.

Ergonomics is the study of the way in which people interact with their work environment, in terms of both their efficiency and comfort. This includes furniture, lighting, temperature, noise level and work practices.

Appropriateness of data collection forms refers to the adequacy of the layout and content of data collection forms to ensure that correct and meaningful data is collected. For example, if the right questions are not included on the forms, the necessary data will not be collected.

Error detection of inaccurate data is often difficult as data collection becomes more computer driven. Computerisation has meant that people rely more on computers and less on their own judgement.

Collection (input) device	How it works
Mouse/trackball	• Is a pointing device that moves a cursor around on the screen and when clicked can choose an object on the screen. • On the bottom of the mouse (or top of a trackball) is a moveable ball and its movement is translated into digital signals using sensors touching the ball. • Connected to the computer via a cable connecting to a serial port. • There can be up to four buttons and a scrolling wheel located on top of the mouse. The first button is for selecting and dragging. The function of the others will vary depending on the software being used.

Keyboard	• The most common form of data input device. • Each key (when pressed) sends a specific message to the computer in digital form, eg. when the letter 's' is pressed the ASCII code for lower case s is sent to the computer and the letter 's' will be displayed on the screen unless it is a hot key for another function. (*This presumes that caps lock or shift keys are not pressed.*)
Scanner	• Uses light beams and measures the intensity of the reflected light to translate images of text, drawings, photos and other graphics into digital form. Scanners usually do three passes of light, building up the image from three different primary light colours. • The reflected light is converted into a bit mapped image which is saved in the computer's memory. • Scanners offer a variety of resolution, eg. 9600 dpi. The higher the resolution of the original scanned image, the higher the quality of the processed output, but more memory is needed as the resolution increases.
ight pen	• The light pen is a pointing device made up of

	more memory is needed as the resolution increases.
Light pen	• The light pen is a pointing device made up of a small stylus connected to the computer by a cable. • The tip of the stylus contains light sensors which detect the intensity of light emitted from the screen. • The user places the light pen at a particular place on the screen and as the pen is wiped across the screen, the computer is able to detect its position. • The effect is that the light pen selects an icon or hot spot on the screen or can draw directly via the screen.
Graphics tablet	• The graphics tablet or digitising tablet consists of

	draw directly via the screen.
Graphics tablet	• The graphics tablet or digitising tablet consists of a specialised electronic pad and a stylus. • The graphic tablet is connected by cable to the computer. • The fine grid of sensors on the electronic pad detect the position of the stylus and send this information in binary form to the computer. • Some are pressure sensitive allowing for thicker lines to be drawn when greater pressure is exerted on the stylus.
Touch screen	• The touching device, eg. finger, interrupts the matrix of infra-red sensors on the screen. • Useful for the selection of menu items and icons but less effective than graphics tablets for the precision input of drawings or handwriting.

Digital camera	• Uses a light sensitive processing chip to capture photographic images in digital form. • Contains a view finder and a lens to focus the image as well as a disk or memory card to store images. • After a picture is taken it is transferred directly to the computer using a fire wire and can be manipulated as a bit mapped image.
Video cameras	• Conventional video cameras capture analog images on video tape which are then transformed into digital images using digitising cards such as frame grabber cards or full motion video cards. • Digital video cameras capture images and store them directly in a compressed digital form such as MPEG on video tape. • Digital video software is used to edit the video.
Microphone	•

	• Digital video software is used to edit the video.
Microphone	• Captures sound, including voices, in analog form. • The computer's sound card converts the analog sound into digital form. • The technique of changing voice sound into digital form is called voice recognition.
MIDI	• Musical instrument digital interface devices allow a musician to play an instrument where the notes are directly converted into digital form using the sound card, and then saved onto the computer.
OCR devices	

	sound card, and then saved onto the computer.
OCR devices	• Optical character readers are specialised scanners that read pre-printed character in a particular font. • The text is scanned and the computer matches it with a set of individual ASC code characters.
MICR	• Magnetic ink character readers are scanners that read characters that have bee printed using magnetic ink. • These magnetic characters are most commonly found at the bottom of cheque and show the account details.

SOFTWARE EXPLANATIONS (WIMP ENVIRONMENT)

Operating system software	• Also called system software, it is built into the motherboard when the computer is made. • Examples of operating systems are: ◆ MS DOS (Disk Operating System) uses a command line interface which was very popular before the GUI interface became common ◆ Apple Macintosh, eg. O/S 10, developed by Apple Corporation and uses the GUI interface ◆ Windows '95, '98, CE, NT and 2000 uses a GUI (graphical user interface) ◆ Novell Netware is a UNIX-like operating system especially designed for networking and is comparable to UNIX and Windows NT ◆ UNIX is one of the oldest operating systems used for multi-tasking by multiple users and has built-in networking capability ◆ LINUX is a more recently developed UNIX-like system that is freely available over the Internet. It is becoming very popular and is compatible with other operating systems such as Windows and Macintosh.

Application software	• Programs designed to perform specific tasks • Work in conjunction with an operating system • Can be bought as packaged (standard) or customised for a specific customer's needs.

Examples of application software	• **Word processor:**
	  ◦ Creates, edits and stores text-based documents
	  ◦ Is used for letters, reports, articles, assignments and books.
	• **Database:**
	  ◦ A collection of data organised into files, records, fields, and characters
	  ◦ Data structure is described in a data dictionary
	  ◦ Data can be sorted and searched using a database
	  ◦ Databases can appear in table format as rows and columns or record format with the information for one record.

with the information for one record.

- **Spreadsheets:**
  - Composed of a grid of cells made up of intersecting rows and columns
  - Contain values (number), labels (text) and formulas for doing calculations on the values
  - What-if predictions or simulations can be done using spreadsheets
  - Spreadsheet templates contain all labels, formulas and page layout. Values can be added each time the template is reused
  - Charts and graphs can be produced using spreadsheet data.
- **Desktop publishing:**
  - Used to combine text and graphics to produce a document
  - Data is organised and moved around the page using text frames and graphics frames for a variety of elements such as headings, columns of text and graphics.

graphics.

- **Multimedia:**
  - Used for the presentation of information using text, graphics, animation, audio and video
  - Graphics programs can be used to paint or draw
  - Animation software is used for video and sound software is used for audio.
- **Telecommunications:**
  - Manage the transmission of data between computers
  - Communications software supports on-line connections, eg. Internet; electronic mail (email); facsimile; remote access connection; file transfer; automatic dial-up and answering.

Utility software	• Used to support, enhance or expand application and system software • Can perform a range of tasks such as: ◆ Formatting a disk ◆ File management, file conversion, back-up and data recovery ◆ Disk fragmentation ◆ Virus protection ◆ Data compression ◆ Memory management and so on.

2.2 ORGANISING DATA

Organising data includes all the activities that get data ready to be used by the information system. These include formatting, digitising, sampling and other methods that enable data to be analysed, processed, stored and retrieved, transmitted and displayed.

Hardware	Software	Non-computer tools
Data including text, numbers, images, sound and video is digitised as it is converted to ASCII code and executed within the CPU.	• Word processor, spreadsheet, database, multimedia • Paint and draw • Operating system for ASCII conversion.	Books, magazines, filing cabinet, manual library catalogues, other hardcopy structures.

Social Issues Related to Data Organisation

Data has traditionally been collected and entered directly into computers using input devices. However, data is increasingly being accessed by means of the **Internet** and **hyperlinks** from remote locations typically using the Internet. Software has been developed that allows for improved methods of **accessing different types of data** such as voice, fingerprint and even using the image of the eye's iris.

to developments in LCD technology. Palm top computers and mobile phones are used to communicate with computer networks.

The population of today has become increasingly **dependent on electronic data**. If this data is not complete or stored appropriately there can be serious repercussions. One example of this is the occurrence of the '**Millennium Bug**' at the close of the twentieth century. Although none of the feared disasters came to pass, the fact that some microchips contained a date in two digits instead of four fuelled a concern that they would malfunction as the year changed from 99 (for 1999) to 00 (for 2000). Among the fears were that airplanes would crash, medical apparatus would fail, lifts would be stuck between floors and so on. The fact is the populace just did not know what was going to happen and spent many millions of dollars guarding against possible disasters.

Data Analysing Explanations

Learning requirement	Examples
• Identify hardware requirements to carry out a particular type of analysis.	• Large amounts of data and complex calculations require la RAM and secondary storage and fast processing. • Scientific analysis requiring complex calculations, such weather forecasting, may require a mainframe or supercomput
• Describe the best organisation of data for a particular type of analysis.	• **Searching** will require the use of a database application or at le a find or search command as in a word processor. • **Sorting** data alphabetically or numerically in ascending descending order can be done using a database application as v as in the table of a word processor and within a spreadsheet. • The use of programs incorporating graphics and animations, t can be used for **modelling** and **simulation**, are ways to m predictions about decisions. Examples of these programs are fli simulators and computer games. Other simulations can be d using spreadsheet and financial/statistical software. • More sophisticated modeling and simulations can be done mainframes and supercomputers. For example, stress testing rocket components. • **What-if predictions** can be done on spreadsheets where values changed within a template to reveal changing outcomes as in a l

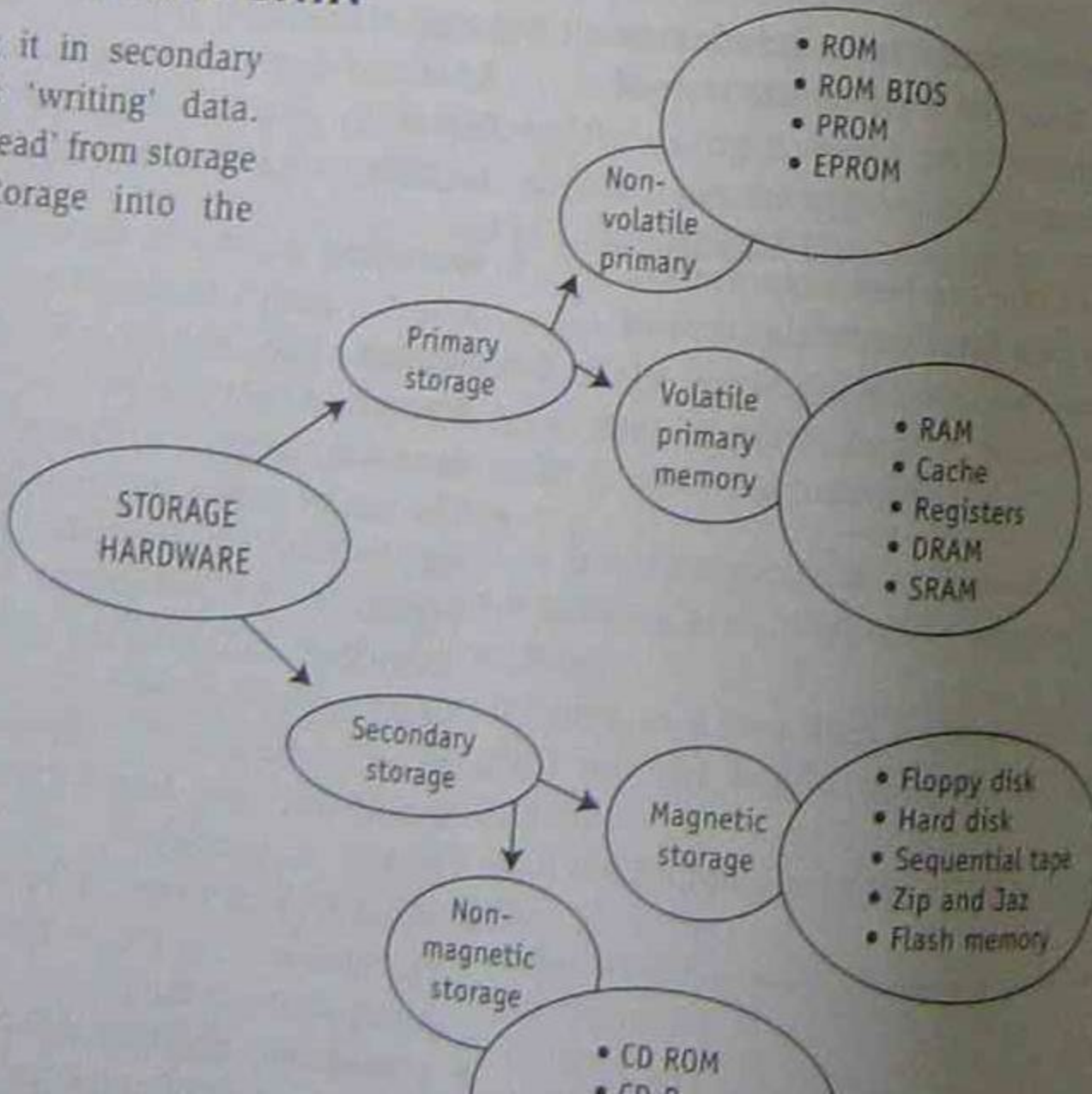
• Use software analysis features in a range of software applications to analyse image, audio, video, text and numeric data.	• Word processors can be used to analyse/process text and graphics. • Databases are use to store, search and sort data and create reports which manipulate the data in a specific manner. • Spreadsheets are used to store values, labels and formulae to create charts and answer what-if predictions using values entered. • Desk top publishers are used to store and manipulate text and graphics to create a required page layout. • Multimedia software is used to create, store and manipulate text, graphics, animations and sound.
• Compare and contrast computer	• Searching is done on a computer

	graphics, animations and sound.
• Compare and contrast computer and non-computer tools for analysis on the basis of speed, volume of data that can be analysed and cost.	• Searching is done on a computer using query language in database or the 'find' facility of a word processor. • In a non-computerised system, searching can be done on files a filing cabinet or cards in a library catalogue. • The computerised search is usually faster and more accurate. • File storage takes up less space in a computerised system and consequently, depending on the cost of hardware and volume of data, is less expensive to store. • Privacy can be more easily eroded using computer databases due to their ease of access.
• Analyse data on individuals for the purpose it was collected.	• Incorrect analysis or the poor organisation of data can lead to output information being inaccurate. • If data is analysed for purposes other than those for which it was collected the resulting information may be inaccurate.

the resulting information may be inaccurate.

2.4 STORING AND RETRIEVING DATA

Storing of data involves placing it in secondary memory and is referred to as 'writing' data. Retrieving data occurs when it is 'read' from storage and moved from secondary storage into the working memory RAM.



Storage/Retrieval Tools and Issues

Software	Non-computer tools	Social and ethical issues
Utility • Hardware interface • File management • Database management • File formats **Internet browser** • Machine independent browsers • Search engines	• Paper-based storage • Microfiche • Libraries	• Data security • Unauthorised access • Advances in storage and retrieval with data matching

Data Storage and Retrieval Explanations

Writing data	Is the process of saving data to secondary storage.
Reading data	Is the process of retrieving data from secondary storage.
Back-up copy	Data storage should always include a second copy, called a 'back-up copy', which is stored separately for security purposes.
Primary storage	Primary storage, often called 'memory', refers to either: • RAM (random access memory), the temporary, volatile or working memory. • ROM (read only memory), the non-volatile memory that is found on the motherboard and includes the operating system.
Secondary storage	

	motherboard and includes the operating system.
Secondary storage	Secondary storage is non-volatile or permanent memory. Information can be saved to secondary storage using peripheral devices. Examples of secondary storage are: • Floppy disk (magnetic) • Hard disk (magnetic) • Removable cartridges such as Zip disks, Jaz disks • Sequential tape (magnetic) • CD ROM (optical) • DVD (digital versatile disk).
Volatile data	Volatile data is temporary data and is lost if the computer is switched off before it is saved to secondary storage.

Non-volatile data	Non-volatile data is not lost if the computer is switched off. Secondary storage and ROM are non-volatile. RAM is volatile.
Magnetic disks	Magnetic disks are the most widely used form of data storage. They consist of a circular piece of metal covered with a plastic casing. The metal disk is coated with a thin layer of magnetic material and the disk is divided into a series of circular tracks and radial sectors. The data are stored in between the tracks and sectors. Floppy disks, hard disks and sequential storage tapes are magnetic.