INFORMATION SYSTEMS AND DATABASES

Information Systems

CHARACTERISTICS OF AN INFORMATION SYSTEM

- ***Organisation of data into info*** – Data must be **organised** (sorted, summarised, classified) **prior analysis**. If not done carefully, the data will become meaningless, so use data dictionaries.
- ***Analysing info to give knowledge*** –There are **varieties of tools available** for analysis by tables, queries and reports. Decisions are then made based on these decisions.

THE DIFFERENT TYPES OF AND PURPOSES FOR INFORMATION SYSTEMS

- ***Process transactions*** – **Transaction processing systems** collect, store, modify and retrieve the daily transactions of an organisation. It also provides data for other systems.
- ***Provide users with information about an organisation*** – **Management info systems** provide info on the organisation's performance. It takes data from TPS and organises it into info reports.
- ***Help decision making*** – **Decision support systems** assist people to make decisions by providing info, models and analysis tools. It takes data (from TPS, MIS and other external sources) to make decisions.
- ***Manage information used within an organisation*** – **Office automation systems** manage vast amounts of data. It provides ways to complete administrative tasks efficiently and effectively in an organisation.

Organisation

Organisation – Organising is the process of **preparing data for other info processes** by determining how the data should be arranged and represented. A database is an organised collection of data.

NON-COMPUTER METHODS OF ORGANISING – telephone book, paper filing system and card based apps.

Advantages	Disadvantages
Convenience in **acquiring info (no info tech)**	**Difficult to update**
Easily and **inexpensively** organised data	Takes up more **physical spaces**
No computer skills needed	Takes **more time** when searching and **only available to one person** at one time
Easier to **secure and private**	Not many display options and only contains **one view**

Computer based methods of organising – Flat-file systems, DBMS and hypermedia

Advantages	Disadvantages
Greater display options such as tables, forms and reports-printable reports.	More easily **prone to hackers**, **theft** of confidential information due to its greater availability over a network.
Large amounts of virtual storage space on disk	Data can be **wiped out if not backed up** by a virus, etc
Fast retrievals using searches and sorts	Requires **electricity**.
Easily edited, updated electronically	More **expensive** to set up and operate the I.T & training
Data can be **backed up**	**Training** for participants may be required
Data can be **accessed by several people** at the same time	
Confidential information can be restricted	
Arithmetic manipulation of data is possible	

FIELD DATA TYPES

- ***Numeric*** – integer, decimal, real/floating number, currency, Boolean and date+time
- ***Text/characters***

THE LOGICAL ORGANISATION OF FLAT-FILE DATABASES

- A flat file database organises data into a **2D table, using characters, fields, key fields, records and files**. The available data type, chosen by the programmer, is described in a data dictionary.
- ***File*** – A block of data.
- ***Record/tuple*** – All the data about a single entry.
- ***Attribute*** – A specific category of data in a database.
- ***Field*** – A particular attribute of a particular record.
- ***Character*** – Smallest unit of data. This includes letters, numbers and special symbols
- ***Key fields*** Used to **uniquely identify a record** in a file by using an item that only applies to one item.
    - Candidate key – Any single/combination of field(s) that uniquely identifies each record
    - Primary key – A new integer field, created specifically for this purpose, and can't be null.
    - Composite key – More than one field is used as primary key.

THE LOGICAL ORGANISATION OF RELATIONAL DATABASES

- A **collection of 2D tables** that **(with RDBMS)** organises data through **relationships** to provide flexibility in manipulating data. It allows you to **manage multiple DBs** and manipulate data by **searching and sorting**.
- ***Schemas*** (entity relationship diagram) – Illustrate the entire DB showing **entities** and **attributes, PK & FK** and **relationships** (in relational DBs). It underlines the primary and foreign keys and labels the relationship.
    - **Relationship** – The way the entities are related to each other.
        - 1:1 – **Seldom used** and possible to combine the attributes of both tables to one.
        - 1:M – Most **commonly used**
        - M:M – Join table, with two FKs, to create two 1:M relationships to the two other tables.
- ***Linking tables using primary and foreign keys***
    - A **primary key uniquely identifies** a record. It is the **same as the foreign key** (primary key of another table). The join between the keys connects the entities.
- ***User views for different purposes***
    - **Table view, form view and report views** – ***Pg. 6 IS&DBs***
- ***Referential Integrity*** – Ensures that **all PK has a FK**, except when **NULL values** are specified.
    - **PK is changed/update –** The **change isn't permitted** or the **FK will change automatically**
    - **PK is deleted** – **Deletion isn't allowed** or **FK is deleted** as well.

DATA MODELLING TOOLS FOR ORGANISING DATA

- Data modelling is the process of **identifying entities, the relationships and the attributes** of those entities. It is critical to creating an efficient database.
- ***Data dictionary*** – Describes the characteristics of data. It includes the **field name, data type, data format, field size, description and an example**. It is an example of metadata.
- ***Schematic diagrams that show the relationships between entities***
- ***Normalisation*** – **Reducing data redundancy** (wastes storage space and increases maintenance problems). It involves **designing an efficient schema** and by **splitting data into tables**, **linked by relationships** and **splitting fields into smaller fields**. Data is reorganised and **repeated fields are removed**.

THE LOGICAL ORGANISATION OF HYPERMEDIA

- Hypermedia is the **extension of hypertext that supports graphics, audio and video**. The most well-known application of hypermedia is the WWW. Each document is retrieved electronically using hypertext.
- ***Nodes and links*** – Nodes are a block info. Users follow links embedded in a node and is taken to other nodes.
- ***Uniform resource locator*** – **Unique address of a file or resource** on the Web, which can be retrieved by entering the URL into an address bar. A URL is made up of a **protocol, domain name** and **file path.**

    - **Protocol** – The **transfer method** used to **access the resource**. Can be **http**, https or ftp.
    - **Domain name** – **Name of the website**, which can be replaced with its **IP address**. It is used to locate the computer/webserver that hosts the domain's website. After the final full stop is the **top level domain name – gTLDs** and **ccTLDs**.
    - **File path** – The **full path** for the file to be retrieved.
- ***Meta data such as HTML tags*** – HTML is the method of **organising hypertext/media**. HTML documents give **instructions** (formatting, hyperlinks + other functions) with HTML tags. Includes **META tags** and **anchor tags**.
- `http://www.w3.org/protocols/overview.html`
  Protocol Domain name Subdirectory path Filename

TOOLS FOR HYPERMEDIA ORGANISATION

- ***Story boards to represent data organised using hyperlinks*** – ***Pg. 6 PM***. Each screen shows the **title, headings, content and navigational elements**. This includes **linear, hierarchical**, **non-linear** and **composite**. The type of storyboard is dependent on the **nature of info**. There is also a **navigation map**.
- ***Software that allows text, graphics and sounds to be hyperlinked*** – Simple **text editors** to **Web creation software. HTML tags are automatically inserted**. These software editors remove the need to understand technical knowledge and use a **WYSIWYG environment** (artistic layout).

Storage and Retrieval

DATABASE MANAGEMENT SYSTEMS (DBMS)

- A software app that allows the **entry, manipulation (search/sort) and S&R of data in a DB.** Data can be **organised into tables, viewed in forms, retrieved using queries and displayed in reports**. It uses **data dictionaries** and **schemas.** DBMS have **data validation** and **backup and recovery**.
- **SQL statements** are on DBMS, permissions checked, then process is performed and items are sent to DBMS.
- ***The role of a DBMS in handling access to a database*** – It allows users to access a database so they can enter, maintain and view the data. It restricts access to maintain and establish **data security.**
- ***Independence of data from the DBMS*** – **Separation of data and its management from the software app**. The DBMS **doesn't hold any data**, but **organises it for viewing**. Data organisation can be changed without affecting the software, which just needs to **adapt accordingly**.

DIRECT AND SEQUENTIAL ACCESS OF DATA

- ***Sequential access*** – Data must be access in a **linear sequence**. So it is must slower and impractical.
- ***Direct access*** – Data is **accessed without accessing other data**. It only requires its **location** and an **index**.

ONLINE AND OFFLINE STORAGE

- ***Online*** – The **peripheral device is under the control of the user**. E.g. hard disk of computer or internet.
- ***Offline*** – **Accessed when storage media is mounted into a drive.** Suited for **backup**. E.g. USB, centralised DB.

CENTRALISED AND DISTRIBUTED DATABASES

- ***Centralised database*** – **One DB** under the control of a **DBMS**. All users and client app connect to the DBMS. Problems may occur when there is an **increase in users** and **remote access**. **Communication lines** must **operate without failure** or loss of response times.
- ***Distributed Databases*** – **Multiple DBs** in multiple locations and controlled by a **DDBMS**. It has a **central** DB and other DBs that **regularly sync** to the central server. It **reduces data transmission costs** that would arise with centralised DB. Includes **fragmentation (horizontal/vertical), downloading** and **replication**.

	Advantages	Disadvantages
Centralised	Security (only need to look after one place) Easy to back up (on one server to deal with) Single DBMS	Dependent on Network connection Slow, unreliable for some users
Distributed	Generally faster Independent of network connection	Poorer security Prone to fragmentation redundancy

STORAGE MEDIA

- ***Magnetic storage*** – **Large storage** with **direct access** at **high speeds**. Digital data is written on the surface of a **magnetic medium**. The **strength of the magnetic force** determines a **0 (low)** and **1 (high).**
- ***Hard discs*** – **Metal or glass platters** covered with **magnetic material**. Data is arranged in **tracks** and **sectors**, and read and written by **read/write heads**. It is **direct access**.
- ***CD-ROMs*** – **Plastic disk** with a **reflective layer of metal** covering the surface. Data is read and written with **lasers** on a **spiral track** made of **pits** and **lands**. The **reflection of light** detected by **sensors** determines whether it is a **0 (constant)** and **1 (transition).** It allows **direct access** and holds approximately **650 MB.**
- ***Cartridge and tape*** – Long **strip of plastic** coated with a **layer of magnetic material**, wound inside a **cartridge**. It can **store large amounts of data** (back up) for a **cheap price**, outweighing **sequential access**.

ENCRYPTION AND DECRYPTION

- Encryption is the process of encoding data, making it unreadable, while decryption is the process of using a key to make the data readable. It is most effective for transmission of critical data.
    - **Single key (symmetrical)** – Requires the same key to encrypt and decrypt.
    - **Two key (asymmetrical)** – Requires a public key to encrypt and private key to decrypt.

Backup and Security Procedures

- Security is to **prevent data loss** and **unauthorised access to data**. Techniques are used in combinations. The techniques chosen are based on **sensitivity of data, how critical the data is to the organisation's continuation of operations and repercussions of data loss. Data loss (DL) /Unauthorised access (UA)**

Techniques	DL	UA	Techniques	DL	UA
Backup and recovery	Y	N	**Restricting access using DBMS views**	N	Y
Physical security measures	Y	Y	**Record locks in DBMSs**	Y	N
Usernames and passwords	Y	Y	**RAID (mirroring only)**	Y	N
Encryption and decryption	N	Y			

- ***Backing up*** – **A copy of data is stored at another location**, to be retrieved for recovery in case loss of data. Backups are made regularly with **full → weekly** and **partial → daily**. It includes **incremental** and **differential**.
- ***Physical security measures*** – **Climate controlled rooms** + **access controls (locks)** to stop unauthorised entry.
- ***Usernames and passwords*** – Secure files, directories, DBs by **identifying user** and their **assigned permission**.
- ***Restricting access using DBMS views*** – Restrictions are made through **assigning permissions**. Permissions give users access to a view, which is as a result of **SQL statements in DBMSs**, specifying **organisation of data**.
- ***Record locks in DBMS***
    - **Pessimistic lock** – Involves **locking a record** once editing, by a user, has started
    - **Optimistic lock** – Based on assumption that **conflict rarely occurs**. If there are changes, a **warning message** is shown to either overwrite the stored record of discard the user's current changes.

TOOLS FOR DATABASE STORAGE AND RETRIEVAL

- ***Indexes*** – Describes **particular records without actually ordering the records**. It enables the ability to **search through data quickly**. It should only be **specified for key fields** or fields that are commonly searched.

- ***Extracting relevant information through searching and sorting a database***
    - **Searching** – Examining DB to retrieve **data that fit the search criteria**
    - **Sorting** – Arranging data in an order (**alphabetical/numerical** or **asc/desc**) for **higher efficiency**.
- ***Selecting data from a relational database using Query by Example (QBE) and Structured Query Language (SQL) commands***
    - **QBE** – **Visual method of specifying a query** without SQL knowledge. Search criteria are entered into an "**empty record**". A query engine creates a **SQL statement**, **performs search & displays results**.
        - Relational operators

CONTAINS	LIKE	GREATER THAN	>
DOES NOT CONTAIN	NOT LIKE	GREATER THAN OR EQUAL TO	>=
EQUALS	=	LESS THAN	<
NOT EQUAL TO	< >	LESS THAN OR EQUAL TO	<=

        - Logical operators

True when both expressions are true	AND	True when at least on expression is true	OR
Opposite	NOT		

    - **SQL commands** – SQL is a **specialised language for searching** a (relational) DB and **manipulating data display**. Searching and sorting is done by a **SELECT statement**. Also includes **INSERT, UPDATE.**
        - SELECT: list of **attributes**. Using ***** will bring all attributes. E.g. Borrowers.LastName
        - FROM: **Tables** data will be retrieved from. E.g. Borrowers
        - WHERE: **Search criteria** + what the **PK** and **FK** are. All records are returned if there is no WHERE clause. Wildcards include **?** and ***.**
          E.g. Borrowers.BorrowersID=Book.BorrowersID, LastName="Shayshay"
        - ORDER BY: How the results will be ordered. It is automatically set **alphabetically ascending.**
          E.g. LastName DESC

TOOLS FOR HYPERMEDIA SEARCH AND RETRIEVAL

- ***Free text searching*** – **Searching a computer-based document or DB for characters or words**. Search engines search all the words and try to match search words supplied by the user.
- ***Operation of search engine*** – It is a **DB of websites**. **Specific info** is found based on **user's search criteria.**
    - **Crawling** the web to locate and retrieve web pages
    - **Indexing** and ranking each web page found
    - **Analysing** search criteria entered by users
    - **Retrieving** suitably ranked web pages.
    - **Indexing and search robot** – An index is a **table containing info about the location of data**. It is made through **search robots crawling the WWW**. URLs found are sent to **indexing software**, which finds the **page summary** and stores it in a DB. **Locations of specific words** are also stored alongside the word in the index database.
    - **Metadata** – Search engines search for metadata
- ***Use search engines to locate data on the World Wide Web*** – Users **enter search criteria** into a search engine. The criteria are **transmitted to its webserver**, then a **query engine**. The query engine analyses the search criteria and **transforms it into a logical expression**. It **performs the search**, **retrieves page references** (containing the webpage summary) and **ranks it** for the user.

REPORTING ON DATA FOUND IN HYPERMEDIA SYSTEMS

- ***Reporting*** –This includes web browsers, stand along apps and media players.

Other Information Processes for Database Information Systems

DISPLAYING

- ***Reporting on relevant information held in a database*** – It is the **formatted and organised presentation of data**. A DBMS controls the design of a report. E.g. invoices and sales summaries
    - The purpose of the report determines its **content, format** and **style**. A typical report produced by a DBMS includes a **report header, page header**, **the details, page footer** and **report footer**.
- ***Constructing different views of a database for different purposes*** – A **form designer** is a standard tool in DBMS. It gives you control over the layout and appearance of a database form.
    - **Table view** – **All records displayed** in tables **without any images**. **Overall view** of the data.
    - **Report view** – Used as a **summary format** and **output of data for printing**
        - Design elements – White space, legibility of text (serif & san serif), colour & graphics, consistency, grouping of info and justification.
    - **Form view** – Each **record is displayed separately** and is used to **enter, view and edit data**

Issues Related to Information Systems and Databases

ACKNOWLEDGEMENT OF DATA SOURCES – **Permission is needed from the source** to use their data before publication. It is concerned with the **work and expense** used in gathering the data. People should acknowledge work for **credibility, justifying data**, mechanism for **tracking data** and as a **requirement** by the source organisation.

- ***Copyright Act 1968***

THE FREEDOM OF INFORMATION ACT – **Only governments** and **not to commercial organisations**. It **enforces access to info**, ability to **correct wrong info** and **appeal against decision** to deny access. It is the right to access data that relates to the individual and does not invade another's privacy.

- ***Freedom of Information Act 1982 (Cth)***
- ***NSW Freedom of Information Act 1989***

PRIVACY PRINCIPLES –This **protects an individual's personal info** from **unauthorised access** and **corruption**. It enforces organisations to state **why personal info is collected**, **how it will be managed**, provide individuals **access to it** (to correct wrong info) and **divulge organisations** that may have access to the data.

- ***Privacy Act 1988 (Cth)***
- ***Privacy and Personal Information Protection Act 1998***
- ***Health Records and Information Privacy Act 2002***
- ***National Privacy principals*** – collection, use and disclosure, data quality and data security, openness, access and correction, identifiers, anonymity, transborder data flow and sensitive info.

QUALITY OF DATA – Data integrity is the **reliability and accuracy of data**. This can be improved by **data verification** and **data validation**. Data validation can be done by **range check, list check** and **type check**. **Self-validation** includes **radio buttons** and **list boxes**. This also includes **effective prompts**.

ACCURACY OF DATA AND THE RELIABILITY OF DATA SOURCES – Systems must be able to **resist user mistakes**, **system malfunctions** and **intentional alterations**. Errors may arise from mistakes during gathering, entering data, or simply out-dated. Info from internet should be checked with **accuracy, authority, objectivity, currency and coverage.** Also, **compare the data with other sources**.

ACCESS TO DATA, OWNERSHIP AND CONTROL OF DATA – **Access** of data is concerned with the privacy and FOI act. **Ownership** should be defined in policy statements or terms and conditions. **Control** is implemented to restrict data access for only authorised persons.

DATA MATCHING TO CROSS LINK DATA ACROSS MULTIPLE DATABASES

CURRENT AND EMERGING TRENDS IN THE ORGANISATION, PROCESSING, STORAGE AND RETRIEVAL OF DATA

- ***Data warehouse*** – A database that **contains all the data from all the organisation's databases**. It is analysed to assist in **making decisions**. The content of the data is usually **historical** and **read only**. Therefore, it acts like a **backup**. Processes are performed on it to find techniques to improve performance.
- ***Data mining*** – It is the process of **looking for trends in the data stored** in databases. It **discovers patterns that predict future behaviour based on past trends**. Patterns may be coincidental or have no real world significance. This is able to occur as a result of improvement in technology.
- ***Online Analytical Processing (OLAP)*** – OLAP is a technique to **provide business decision makers with statistical evidence visually, online and quickly**. It **optimises organisation of large data stores** and combines DBs into multidimensional structures – **data cubes**.
- ***Online transaction processing (OLTP)*** – OLTP systems are databases that **allow transactions of remote users** to be **processed immediately** (in real time). Completing a **transaction online** is an example of a transaction performed by an OLTP system. A transaction is a sequence of operations that must be completed successfully or it will fault. E.g. ATMs, online banking.