

## Introduction

In today's world easy access to information is the life blood of commercial organisations. Financial data, *human resources* records, customer data, *product inventories*, form the backdrop to *operational* and *strategic decisions*. Misinformation can destroy companies, as happened with a multi-national company operating in Dublin in the 1970s when its manufacturing output was mistakenly based on sales to depots around the world rather than to customers. Modern computer systems make a repetition of this kind of error unlikely. Today the amount of data stored by an organisation can be enormous; the Ordnance Survey of Ireland (OSI), based in the Phoenix Park, holds more than 220 *terabytes* of data — possibly the largest database in Ireland. This lesson covers the basics of this technology, its origins, the types of database, the concepts of database management, database security, data transaction security and database design.

OSI's major data holdings comprise

- Spatial information
- Digital landscape models
- Height database
- Raw photos

Some of these are very *resource intensive* — hence OSI's requirement for a very large database.

## What is a Database?

A database is a computerised store of information (collection of data in *digital* form) organised so as to allow efficient access to the information for some defined purpose. For instance, a database might hold inventory information to allow stocktaking, identification of re-order levels, analysis of high usage items, etc. It is inextricably linked with (though distinct from) a database management system (DBMS), which is the software system that manages the efficient access to the database.

## Origins

In the early days of computing data was stored on punched cards, and later on tape. These were not easily accessible; reading and writing were very slow and cumbersome operations by modern standards. Like an audio or video tape, a computer tape needs to be wound and rewound to read a specific piece of information. However, the introduction of magnetic discs and drums in the mid 1960s allowed direct access to specific information locations; this innovation facilitated the development of electronic databases.

## Evolution of Storage Media

Punched cards:	1930's - 1970's
Magnetic tape:	1951 - present
Magnetic drums:	1950's - 1960's
Magnetic discs:	1960's - present
Solid-state 'disc':	2000 - present

## Types of Database

The first generation of database systems were *hierarchical* and *navigational*; applications typically accessed data by following *pointers* from one record to another. *Relational databases* were introduced in the 1970s and have become the norm for general purpose databases. Relational databases search for data by content, rather than by following links. This reduces the need to rewrite applications as content evolves. They require more *computing power* and have been dominant since the 1980s. In the relationship model data is kept in the form of tables using *keys* rather than pointers. This structure is suitable for most operational databases but is rather rigid when applied to databases for documents, images, multimedia or scientific and engineering purposes. Other approaches address this problem, two major ones being the *object database* and the *XML database*. OSI uses *spatial databases* and is moving towards *3D reality systems*, both of which are suitable for storing and querying data relating to objects in three-dimensional space.

## Database Management Systems (DBMS)

A database is not just any collection of data, but one that is ordered, accessible, and managed to defined levels of quality regarding accuracy, availability and usability. DBMSs have developed alongside database systems to ensure these functions. A general-purpose DBMS is a complex software system that meets these requirements for a wide range of database applications. The databases that it maintains are often large and complex. Well known DBMSs include *Oracle* Database, *Microsoft SQL* Server, *IBM DB2*, and the open source DBMSs *MySQL* and PostgreSQL as well as databases for individual users such as Microsoft *Access*, *Filemaker* or Open Office *BASE*.

Tools used in developing DBMSs include data definition languages (DDLs), data manipulation languages (DMLs), and query languages. The most widely supported standard database language is SQL, which has been developed for the relational database model and combines the roles of DDL, DML, and a query language. SQL became a standard of the American National Standards Institute (ANSI) in 1986 and of the International Organization for Standards (ISO) in 1987.

## Database Security

Database security deals with the protection of the database from software corruption, hardware malfunction or unauthorised access. Database security includes *access control*, covering such things as password control, reading and writing limitations; levels of authorised entry; use of private networks; *data security*, ranging from physical security to *encryption* of sensitive data such as credit card numbers; and *database auditing* to ensure that no breaches have occurred. Hardware malfunction can be protected against by *mirroring* or by Redundant Array of Independent Discs (*RAID*). In mirroring a full copy of the data is maintained on a similar device – this doubles the storage capacity required, but some of this can be used operationally and not just as back-up. RAID provides for one disc to provide restoration for any one of a group of discs. The number of discs in a group is often 5 or 6 (Mirroring is equivalent to RAID number equal to 1). RAID uses less memory but more processing and data can be lost if a second disc fails before the first is fully restored.



A sample Redundant Array of Independent Discs (RAID)

## Data Transaction Security

Every database transaction must follow the *ACID* rules: *atomicity* (a transaction is either fully completed or not at all), *consistency* (every transaction must follow the rules of the DBMS), *isolation* (one transaction must not interfere with another), *durability* (a completed transaction must endure through crashes, by being committed to a *non-volatile* memory).

## Process of Database Design

The process outlined here reflects a relational database that holds data in the form of tables. The steps include:

1. Describe the purpose of the database.
2. Find and organize the information required.
3. Divide the information into major categories (e.g. Products or Orders) that become tables.
4. Turn information items in each category (each item is a 'field') into columns.
5. Set up the table relationships - look at each table and decide how the data in one table is related to the data in other tables; add fields (columns) to tables or create new tables to clarify the relationships, as necessary. At least one field provides a unique key i.e. each row is uniquely identified (e.g. by a product number).
6. Check that the original purpose is met and refine the design using real data.
7. Apply the *normalisation* rules to validate the table structure.

## The Unknown Future

The amount of information held on databases is increasing daily at an amazing rate. The future of printed media, newspapers and literature, is in doubt as electronic media evolve. Governments and Corporations are daily acquiring detailed information about every one of us, which may be benign but can be abused. This financial turmoil through which we are now living is largely due to the enormous development of storage capacity for information and of rapid and ubiquitous ways to access it – i.e. of databases and database management systems.

**Ordnance Survey Ireland (OSI)** is the national mapping agency of the Republic of Ireland. It produces and sells a very comprehensive range of urban, rural, tourist and leisure maps at a variety of scales in digital and printed form. OSI also produces aerial photographs and digital terrain models.

Customers include:

- Individual members of the public
- Tourists
- Schools
- The construction industry
- Architects
- Engineers
- Property and legal firms
- Government Departments and local authorities.

Annual sales are currently about €24 million.

OSI also licences data for a wide range of computer based applications such as Computer Aided Design (CAD) and Geographic Information Systems (GIS).

OSI products are state of the art produced by the most up to date technology and consequently the company is a leader in the Irish geographic information market.

OSI owns a network of 16 GPS stations around Ireland continuously recording and streaming data from GPS satellites back to the agency's centre in the Phoenix Park in Dublin. That information is then processed in real time. In all, OSI stores a total of some 220 terabytes in 30 separate databases. This is delivered to users by 160 servers making OSI one of the largest database managers in the country.

All products are available directly from the OSI shop (+353 (1) 802 5300) at the OSI headquarters in the Phoenix Park (near Castleknock Gate) through the website and through a national network of retail outlets.

You can find this and other lessons on [www.sta.ie](http://www.sta.ie).

Find out more about the work of OSI on [www.osi.ie](http://www.osi.ie).



## Syllabus References

### Leaving Certificate Technology

#### Core: Introduction to Computer Systems

- The main components and specifications of a computer system.
- Management of a computer system; personal safety and the safe operation of the system.

#### Core: Skills Development, Applications and Software

- Development of basic skills in the use of information and communications technology (ICT).
- Use of the computer as an aid to investigation and research in a design process.
- Use of applications software.
- Use appropriate software such as word processing, graphics and spreadsheet applications.
- Use spreadsheet applications to store and manage data, and to produce output in a variety of forms.

#### Option: Information and Communications Technology (p. 30)

- Computer Architecture.
- Data Communications and Computer Networks.
- The Internet.
- Multimedia and Design.

## Learning Outcomes

On completion of this lesson, students should be able to:

- Describe the origins of databases.
- Describe some types of database.
- Distinguish a database from other types of information.
- Outline the concept of database management.
- Outline the basics of database security.
- List the ACID elements of data transaction security.
- Describe an approach to database design.
- Discuss the future uncertainty of electronic information.

## General Learning Points

The following information can be used to revise the lesson's main learning points and inform discussion.

**Spatial Databases:** A spatial database is a database designed to store and query data related to objects in space, including points, lines and polygons. The *Open Geospatial Consortium* created the *Simple Features* specification and sets standards for adding spatial functionality to database systems. The indices used in normal databases are not suitable for spatial databases. Spatial indices allow for specialised spatial queries such as measurements of distance and area.

**Face Recognition:** Brazil, which is hosting the World Cup in 2014, is testing glasses with mini-cameras attached that can scan faces and compare them with a criminal database. Off-the-shelf face recognition programs are available that together with a web camera and social sites such as Facebook can be used to identify complete strangers along with their interests. As these facilities develop there are complex implications both good and bad for society.

**Magnetic Tape:** Magnetic tape was first used to record computer data in 1951 on the UNIVAC-1 computer. IBM computers from the 1950s used ferrous-oxide coated tape similar to that used in audio recording. Tapes were widely used in the 1970's and '80s; compact cassette tapes were used for home computers. Modern tape systems use reels and have capacities of up to 5 terabytes. Tapes provide very fast transfer of streams of data, but are much slower than hard discs for random access (tens of seconds vs. tens of milliseconds).

**Clustering:** OSI has implemented Oracle Real Application Clusters (RAC) technology to consolidate its databases into one location, managed by a central Oracle server. As well as lowering management overheads this also reduces the number of Oracle licences OSI needs.

## Student Activities

### Set up a functional database

There are many easy-to-use database programs freely available on the Internet (if you do not already have database software).

1. Choose the subject (club members, school sports, books, music, contacts etc.)
2. Choose the fields that you require; you may need to specify the type of field (number, date, text, picture, notes...)
3. Select (or design) the kinds of output or report you require (screen, printed lists, forms, statistics...). Choose the fields to be included in each report. You may want to add some colour to fields, especially for screen output; it should look better and be easier to use.

## True/False Questions

- |   |   |   |
|---|---|---|
| 1. A Terabyte is 1,000 Gigabytes.   | T | F |
| 2. A database is any collection of data in digital form.                    | T | F |
| 3. Discs became popular because they provided fast access to specific data. | T | F |
| 4. A Relational database searches for data by content rather than links.    | T | F |
| 5. Relational databases were chosen to handle increasing content.           | T | F |
| 6. A standard Relational database is ideal for Ordnance Survey purposes.    | T | F |
| 7. SQL combines Data Definition, Data Manipulation and Query languages.     | T | F |

- |  |   |   |
|--|---|---|
| 8. RAID stands for Random Access to Independent Discs.                             | T | F |
| 9. Mirroring is equivalent to a RAID number of 2.                                  | T | F |
| 10. Atomicity requires that a transaction is either fully completed or not at all. | T | F |
| 11. In a relational database a column in a table is called a field.                | T | F |
| 12. Every database must have an associated DBMS.                                   | T | F |
| 13. A key uniquely identifies a row on a table.                                    | T | F |
| 14. ISO stands for International Service Obligation.                               | T | F |

Check your answers to these questions on [www.sta.ie](http://www.sta.ie).

## Examination Questions

### Leaving Certificate Technology (HL, paper B) 2010, Q. 3

- (a) (i) In the management of prescriptions and general patient welfare, the medical profession has recognised the increased use of ICT as a positive development. Outline three ways in which ICT could be used by doctors and other medical professionals.
- (ii) Many multi-national companies make extensive use of ICT to support the sharing of ideas and resources. Give two advantages of using video conferencing rather than email or telephone communication.
- (b) (i) An international computer games company, HARPOON® Ltd, have installed a LAN (Local Area Network) with connections for the network server, 15 office computers and 5 wireless laptops in their new office. Outline three functions of the server on the network.
- (ii) HARPOON® Ltd has appointed one of its employees as a system administrator who works closely with the branch manager. Describe two important tasks for which the system administrator would have responsibility.
- (iii) How could HARPOON® Ltd make use of an intranet?
- (c) HARPOON® Ltd decided to make the first edition of a popular game available to play online. In order to access the game, the registration form shown had to be completed and submitted.
- (i) Outline two reasons why HARPOON® Ltd collected such information from game users.
- (ii) Parents complained about the nature of the information collected. Discuss two issues that might have arisen.



The image shows a web-based registration form for HARPOON® Ltd. The form has a title bar that says "HARPOON® Ltd Online Registration". It contains several input fields: First Name, Surname, House No., Town, Postcode, Date of birth, Gender (with radio buttons for Male and Female), and Email. There is a "Submit" button at the bottom right of the form.

## Did You Know?

- In 2010 people stored enough data to fill 60,000 Libraries of Congress. YouTube claims to receive 24 hours of video every minute; manufacturers have embedded 30 million sensors into their products, converting them into data-generating nodes. The number of smartphones is increasing by 20% a year and the number of sensors by 30%. Tesco collects 1.5 billion elements nuggets of data every month and uses them to adjust prices and promotions. Williams-Sonoma, an American retailer, uses its knowledge of its 60 million customers, including income and houses value to develop its catalogue. Amazon, has claimed that 30% of its sales are generated by its recommendation engine ("you may also like"). Companies such as America's Placecast are developing technologies that allow them to track potential consumers by their mobile phones and send them offers when they get close to a Starbucks. (*The Economist*, May 26th 2011).
- The largest database in the world may be the World Data Centre for Climate (WDCC) Operated by the Max Planck Institute for Meteorology and German Climate Computing Centre. It holds 220 terabytes of web data and 6 petabytes of additional data. A petabyte is 1,000 terabytes or 10<sup>15</sup> bytes.

## Biographical Notes

### Michael Hart (1947-2011)

Michael Hart, a man with no interest in money, who died on September 6th, 2011, was the founder of *ebooks* and of *Project Gutenberg*. He believed everyone should have free access to the world of books. He dreamed that by 2021 he would have provided a million ebooks for free. He started on 4th July 1971 by making the text of the American Declaration of Independence available for download. From the 1980s he had a growing number of volunteers scanning books. By 2011 there were 36,000 books, accumulating at a rate of 200 a month, with translations into 60 languages. The Gutenberg site states that "over 100,000 free ebooks are available through our Partners, Affiliates and Resources". His work was made harder by changes in copyright laws. During the lifetime of Project Gutenberg alone, the average time a book stayed in copyright in America rose from 30 to almost 100 years.

## Revise the Terms

Can you recall the meaning of the following terms? Reviewing terminology is a powerful aid to recall and retention.

3D realities, access control, atomicity, computing power, consistency, data security, database auditing, digital, durability, encryption, hierarchical, human resources, isolation, key, mirroring, navigational, non-volatile, normalisation, object database, operational decision, pointer, product inventory, RAID, relational database, resource intensive, spatial database, strategic decision, terabyte, XML database

Check the Glossary of Terms for this lesson on [www.sta.ie](http://www.sta.ie).