Topic 2 - Tools for Information Processes

In order to understand and build information systems, information processes must be understood. This topic examines each of the information processes by focusing on some of the tools used to carry them out. The tools include information technology and non-computer procedures. In this topic, tools are categorised and presented according to a particular information process. In reality, however, one tool may overlap several processes. One tool can not operate in isolation: therefore, demonstrations of particular tools will involve additional tools and processes. Information processes and tools affect participants within the information system and people beyond it, giving rise to social and ethical issues. Additional tools for specific types of information systems will be examined in the HSC course.

Outcomes
A student must be able to -
• describe the nature of information processes and information technology
• classify the functions and operations of information processes and information technology
• identify the information processes within an information system
• recognise the interdependence between each of the information processes
• identify social and ethical issues
• describe the historical developments of information systems and relate these to current and emerging technologies
• select and ethically use computer based and non-computer based resources and tools to process information
• analyse and describe an identified need
• generate ideas, consider alternatives and develop solutions for a defined need
• recognise and apply management and communication techniques to project work
• use technology to support group work

Collecting
This is the information input process, that involves -
• deciding what to collect
• locating it
• collecting it

What to Collect
This sounds very straightforward, since if you are going to collect data you should know what you want, and often it may in general be just that. However, depending on the issue or person, event, etc. this may be much more involved - and unless the full range of necessary data is to be collected, and recorded correctly, then the process may not be very successful or may even have to be repeated.
Example - the November 1999 referendum on whether Australia should become a Republic and the necessary changes required to the constitution to enable this. The wording on the ballot papers required months to work out - so that each voter could -

- understand the questions being asked
- write just “Yes” or “No” to cast their vote

The YES vote gained about 46% and did not win (it needed 60%), yet many members of both main political parties were in favour of it, and the campaign in general had a lot of support in the community. It seemed that the way the questions were asked - particularly the part about the President being a political appointment rather than an elected one, seemed to put a lot of voters off.

If the referendum question had simply asked whether Australians wanted a President as head of a Republic rather than a Governor General as the Queen’s representative for the country, the result of the vote may have been quite different.

Therefore it can be seen that the decision on what to collect can be critical to - how you collect the data, the format of the data (Yes/No, numbers, sentences, etc.), the type and number of questions asked, etc.
Locating the Data
Again this process will depend on the type of data needed and the purpose behind the collecting of it. A nation-wide census requires every household to complete a large questionnaire on every member of every family in Australia. An opinion pole on the hottest top 40 song or the most popular video will need other forms of locating the data for it to be collected. Compiling a family history, possibly with data from all over the world would require a different set of strategies.

Data location techniques could include such things as - questionnaires, surveys, letters, faxes, telephone calls, personal visits, published statistics, database searches, Internet searches, etc.
If all relevant data is not located, the collection process is incomplete, and its accuracy will suffer. Example - Much of the history of ancient living things is added to every year with new finds and every few years with new techniques - thus the locating of such data is very incomplete and will be ongoing, probably for centuries to come.

Collecting the Data
The physical collection of data of course depends on the method(s) selected - and for many purposes will involve a range of such methods.
Example - surveys of people in the street, combined with phone questions and warehouse database searches to find the most popular washing powder.

Note - since there is always a range of data collecting alternatives, the most appropriate one or ones should be chosen and used.

Once the data has been collected it has to be organised, and how this is to be done needs to be worked out long beforehand - and the collection method(s) devised to make the organising as straightforward as possible. As stated above, the most appropriate method should be used, and this involves considering the organising method along with the collection method, rather than doing all the collecting and then wondering how to organise it later.
Exercise 2.1 complete the following -

The SRC at Westfields for the coming year want to have a much bigger say in what goes on around the school, including - starting and finishing times, subject choices, canteen food, after school detention, uniforms, hats, student parking, and sporting facilities, etc.

1. How should/could the SRC collect data about these issues? Name the best method (or two) and state why you consider it best. (/3)
2. From which groups of people should they collect data about these issues? Give reasons for selecting each group of people? (/2)
3. What should they do with all of the collected data, in whatever form(s) it is in? What are the alternatives? (/3)
4. Prepare a one page interview set of questions for a named group for data collection on this issue. Print it out to hand in. (/5)
5. Demonstrate how if you wanted one of these questions (from Q4) to be favourably or unfavourably answered, you would alter the wording of it - i.e. give two versions of the wording for that question. (/2)

To be completed and handed in by - .................................................................

Hardware Used For Collecting Data
A range of hardware devices may be used for collecting data, including the following - complete the table to indicate what type of data each device collects -

<table>
<thead>
<tr>
<th>Hardware Device</th>
<th>Data Type it Collects - complete these -</th>
</tr>
</thead>
<tbody>
<tr>
<td>scanners</td>
<td>..................................................................</td>
</tr>
<tr>
<td>digital cameras</td>
<td>..................................................................</td>
</tr>
<tr>
<td>microphones</td>
<td>..................................................................</td>
</tr>
<tr>
<td>video cameras and cards</td>
<td>..................................................................</td>
</tr>
<tr>
<td>keyboards</td>
<td>..................................................................</td>
</tr>
<tr>
<td>optical character readers</td>
<td>..................................................................</td>
</tr>
<tr>
<td>counters e.g. cars on a road</td>
<td>..................................................................</td>
</tr>
<tr>
<td>tablets</td>
<td>..................................................................</td>
</tr>
</tbody>
</table>
**Exercise 2.2** complete the following -

1. Each pair of students is to research and produce a one page (including diagrams) set of notes on the operation of the hardware device chosen for them. These will be laser printed and the best set copied for all members of the class. 

   Hardware device - ............................................................................. (/5)

2. Each group is to use the **digital camera** OR **flatbed scanner** to collect graphic data about a selected sport (you need to bring in a sporting magazine to scan). Combine the pictures with text (typed or scanned in) to produce a one page brochure about the selected sport. (/5)

   To be completed and handed in by - ..........................................................

---

**Software Used For Collecting Data**

The software in a computer system is the coded instructions or programs created to tell the computer system what to do.

For many years the challenge for software producers was to produce a chess program to defeat a world chess champion. IBM programmers finally achieved a win 4 games to 2 over Garry Kasparov the world chess champion in 1997. The program was called Deep Blue and it was running on a computer with 512 processors operating in parallel, allowing it to evaluate 200 million board positions per second - considering only plausible moves and countermoves. The program often went through 10 or more levels of moves before deciding what to do.

Some people thought that the victory showed an important breakthrough in machine intelligence and indicated great promise to come. Others thought that it merely demonstrated a great deal of concentrated effort on a situation that is limited by explicit rules of a type that do not exist in important real world problems.

To put this in perspective, consider the following information in 3 sentences, and what the
software in a computer could actually do with it, compared to the average 4 or 5 year old -

Billy was invited to Sally’s party. He asked his mother if she would like a kite. She said that Sally already had a kite and would return it.

Despite the fact that most small children would understand the meaning, no one knows how to program a computer to do this. The problem with the short story is that many of the aspects of the situation are unstated or implied, for example - we assume that -

• Billy and Sally are young children (but not infants)
• Billy thinks Sally would like a kite as a birthday present
• the she in the second sentence refers to Sally (and not his mother)
• the kite is a gift (since children normally take a gift to a birthday party)
• the party is for a birthday for Sally
• Billy’s mother understands that if Sally already has a kite she will not want another

All of these assumptions are easily made by people, even young ones, but no computer program ever devised could cope with any of the above points, let alone all of them.

Software will be required for a number of purposes -

• to interface with hardware collection devices
• to allow data entry into applications
• to directly allow the Internet as a collection source

Software Types

(i) Application - includes both specific programs e.g. a particular graphic scanning program, and the many general purpose programs e.g. ClarisWorks or Microsoft Office packages which contain a range of integrated programs. Web browser programs are another important group over the last few years. It is this type that covers most of the purposes listed above.

(ii) System Development - this is the software used by programmers and system analysts in the process of building and enhancing information systems.

(iii) System - this controls the internal operation of the computer system e.g. Mac OS9 or Windows 2000.

Utility programs such as Nortons Utilities can also be listed under this type.
Exercise 2.3 complete the following -

Use a type of presentation software (examples include PowerPoint, ClarisWorks or HyperStudio) to produce a presentation to show to Joe for upgrading the hardware and software, for his Video Store.

This presentation should cover the feasibility of putting a new computer system and the necessary software to upgrade the store to state of the art computerisation. Joe needs to be convinced of the benefits of such a system compared to its cost.

Costs include - 2 computers ($1500 each), barcode scanner ($650), printer ($350), software ($600), training ($150). Plus all other items you found in your survey of your local video store. You need to find some technical details in adds to show the quality of the hardware.

Your presentation should be of 5 to 10 screens, well planned and presented. (/10)

To be completed and demonstrated by - .................................................................

Software and Hardware

We all use devices to collect data, such as stopwatches for time, thermometers for temperature, microphones for sound and scales for weight. It would seem to most people that connecting any of these devices to a computer and having it do the collection of the data would be quite easy. However, most of these devices are what we term analog while typical computers are digital - this means that for them to communicate effectively with each other some type of interface must be provided.

Example -

The spoken word is analog as seen by the wave shape (at 1).

The speech recognition board (at 4) is the interface to convert the signal.

The digital or binary form is produced (at 6) for the computer.
A lot of effort and money must go into producing interfaces for our many devices to be connected to computers so that they will be both very fast and very accurate. Thousands of monitoring devices - from traffic sensors under the road, to automated doors and burglar alarms, all require this to occur.

The Internet is now a major data collection source, and most people can have access to it. The strengths of this huge international information and communication network include - the amount of information available, the ability to join groups of similar interests and to have questions answered, the chance to chat ‘live’ with people from all over the world. The weaknesses of the Internet are often due to the strengths - there is so much information out there it is often difficult to find just the bits you want, problems with accuracy and bias, unscrupulous and dishonest people, etc. mostly due to no one person or group actually being in charge of the Internet.

You will be using it regularly in this course - so make sure you can use good techniques to search the Internet as effectively as possible.

---

**Exercise 2.4 complete the following** -

1. Compare the appearance of a range of Search Engines available - for example - Yahoo, Excite, AskJeeves for Kids, AskJeeves, Altavista, Infoseek, Lycos, etc. Bookmark 2 which you think seem good (if not already done), and print out a summary of six search engines - what is good and bad about each, ease of use, etc. (/2)

2. What is different about the appearance of these search engines, and the way they have you perform your search? (/1)

3. Find the range of answers (and time taken) for each of the following - (/1 each)
   (i) what is a search engine?
   (ii) what do search engines do?
   (iii) how do icebergs form? (plus name of search engine used & number of pages found)
   (iv) what is a meta search?
   (v) who invented calligraphy? (plus name of search engine used & number of pages found)

4. Which 2 search engines do you consider were the best and why? (/2)

   To be completed and handed in by - ............................................................

---
Non-Computer Collection of Information

Obviously people collected information in a whole range of ways before we had computers, and many of these are still done at least to some extent. These methods include -

- literature searches - going through books and journals in a library
- surveys and interviews - questionnaires in verbal or written forms
- forms - forms designed to allow the applicant to complete the data
- manual recording of events - photographs, written accounts, etc. of what has occurred
- existing non-computing data - pieces of paper all around the place contain data

Assignment

Over recent years the collection of data has become a lot more automated, with the common use of such devices as -

- swipe cards - to show when a person has entered a building or a room e.g. Westfields sports students on late starts to show their time of arrival at school
- portable barcode type readers - for identifying and counting products on supermarket shelves, or log electricity use from household meter boxes.

Other such devices are in use today, so your task is to -

1. name and describe in some detail the operation of ONE other such data collection device
2. predict ONE example of new technology (not yet in use) for some type of data collection

To be completed and the printout handed in by - .................................................................

Social and Ethical Issues

These can both be quite important in the collection of data and include -

- bias in the choice of what and where to collect data
- accuracy of the collected data
- copyright and acknowledgment of source data when collecting
- the rights to privacy of individuals on whom the data is collected
- ergonomic issues for participants entering large volumes of data into an information system

Exercise 2.5

To gain an idea of the possible effects of such issues, read both of the following passages and complete the sections under them.

[1] Quatro Computer Company

The Quatro Computer Company recently announced the completion of a massive project. The project consisted of creating a data file on all adults over the age of 21 in the United States. The company worked on the project for two years with nearly 1000 people involved. The work was funded by many major companies representing conglomerates, loan companies, banks, employment agencies and government agencies.
Some of the collected data consisted of -
- Bank statements of deposits
- Credit references
- Complete educational background
- Family information - house owner/renter, single/married/divorced, no. of children, etc.
- Jobs held - complete job history

The companies and agencies that funded the project will be able to request the information from Quatro in order to help themselves meet their own particular needs. For example, a company with an opening could present a list of prospective employees and request information from Quatro. The company would not have to verify each prospective employee’s resume. They would receive information from Quatro and quickly narrow their choices for a position. A credit company could quickly decide whether to loan money to an individual. Obviously, it would be of great convenience to the lender, and would result in a quick decision for the individual requesting the funds. Although the system would offer certain conveniences to a company and perhaps an individual seeking either employment or services, some problems could still arise. This is the feeling of a group called Citizens for Privacy. Mr Wallace, head of the group, stated the following during a recent press conference:

Quatro’s system has no way of taking into account poor credit references, a series of jobs, or family problems. Thus, companies could eliminate people who would be dedicated employees but do not meet certain standards. A person who held many jobs could have been laid off due to the economy. The Quatro system does not include a job evaluation from a previous company. This could hurt the prospective company as well as the individual.

Mr Shand of the Quatro Company stated that his company simply had honoured a request from other companies. The information gathered would allow companies to obtain whatever they considered relevant to their particular needs. How each individual company used the information would be up to that company. Quatro provided the requested information, and nothing more!

To Do -
Analyse the above situation, and complete the following -
- state the Quatro Companies’ reason for gathering such information (/2)
- state why companies want such information (/1)
- state the concerns of the Citizens for Privacy (/2)
- choose a position (about whether such information should be assembled) and write a 200 word discussion paper (for or against) and how such information should be used. (/5)

Questions - answer the following -
1. How might this information be used by companies or public agencies? (/2)
2. How could such information be gathered? (/2)
3. What type of information should an independent company be allowed to obtain on an individual? (/2)
4. How might a computerised data-gathering centre like Quatro be beneficial to people? How might it be harmful? (/4)

[2] Cheating at School

It was certainly spite that made Christine do it, but she sure did it. It all began when Mrs O’Reilly caught Christine cheating on her Half Yearly examination. The teacher took Christine’s test, tore it up right there, and threw it into the bin. Before the whole class she announced, “You’re a cheat, and I am failing you on this exam.” Christine left the class in tears and spent half a day in the Principal’s office. Her parents were summoned to school, and there were a lot of lectures about honesty, dishonesty, stealing and cheating. A permanent entry was made on her record.

Two weeks later, Mrs O’Reilly brought in some computer software to show the class a program. She ran it, but when they were in the middle of the program, it crashed. Mrs O’Reilly didn’t know what to do. Somebody suggested that they look in the manual, but Mrs O’Reilly explained that she didn’t have the manual because a friend had bought the program and made a copy for her.
“I know what to do,” Christine said. She took the disk from the drive, put it in its jacket, and promising, “I’ll be back in a minute,” walked out the door with it. She went straight down to the Principal’s office to complain. He listened to her for two minutes and told her to go right back to Mrs O’Reilly’s class and return the disk. “I’ll never tell her you tried this,” he promised. Instead, she went straight to the police, and lodged a formal complaint. Moreover, she handed over about 50 pages that Mrs O’Reilly had photocopied from books and distributed to the class. They were further evidence of a pattern of illegal copying by Mrs O’Reilly.

After a discussion with the Principal and District Superintendent, the District Attorney decided not to prosecute. But he reckoned without Christine. First, she went to the local newspaper to be sure that it had the story, and then she wrote to each of the textbook publishers to tell them of her action. Under the copyright law, they had no choice but to bring suit or face the loss of their copyrights. Four brought suit immediately; two others gave the school two weeks to pay damages or face suit for actual and punitive damages. Furthermore, all the publishers involved are demanding that action be taken against the teacher. Some want the authorities to prosecute, others want the school authorities to discipline the teacher. All the publishers have made it clear that they will push the issue to a successful conclusion.

After sitting on the story for days, the newspaper carried a small piece. Seeing the publicity and hearing from Christine, half a dozen other students have come forward with examples of illegal copying of copyrighted software and books. Now, the accusations have become so widespread that the paper carries a story each day about the school. Faced with further evidence of infringement of copyright, the Association of Publishers has brought suit against the school.

To Do
Analyse the above situation, and as the lawyer for the school, complete the following - in 100 to 200 words each -

• a statement for presentation to the P&C meeting (/3)
• advise the Principal on the legal and moral position of the school (/4)
• advise the school executive on the actions it should pursue (/3)

Questions - answer the following -
1. Why did other students come forward with examples of illegal copying? (/2)
2. If you were the newspaper editor, how might you react to all the reports of illegal copying? (/2)
3. List three impulses which led the publishers to file suit. Are they all ethical considerations? (/4)
4. Give two solutions to the problem of ensuring that teachers will not violate copyright laws. How realistically effective are your solutions? (/2)

Major Project
For your group project you will be creating an information system for a Real Estate Agency.

You will be working in groups of 4, with each member having specific tasks to perform - to gain an individual mark, as well as the total project producing a group mark. These two marks will be added to produce the total mark per person.

If you choose to have people in your group who do not do the work properly, or fail to hand it in for marking, then your group mark will suffer - so choose your fellow workers carefully!

See the separate Project Sheet to find the range of tasks so that you can allocate jobs. Obviously the first task is collecting data for your project, so ensure that you follow all of the points set out in the preceding pages.
Organising

This is the process of arranging, representing and formatting of data for use by other information processes. Depending on the type of data collected, this arranging, representing and formatting will be dependent on the selection of the most appropriate software to do the planned organising. For example, you might decide that a spreadsheet is most useful for one organisational method (to make calculations and predictions), while a paint program might be best for another part of the organising of data (such as a picture from a digital camera).

Different methods of organising affect processing - for example -

- letters of the alphabet represented as images rather than as straight text (this could cause storage and/or memory problems)
- numbers represented as text rather than as numerals (such as nine rather than 9)

Depending on the hardware used in the collection process, the data organisation will vary e.g. images from a digital camera can be in a number of formats, resolutions, etc. and the most appropriate will need to be selected. All of the other hardware devices for collecting audio, video, numeric and text data will also have such considerations for the digitised format.

Exercise 2.6 compare the size of files created in -

(i) paint versus draw parts of ClarisWorks - create in draw and copy and paste into paint - save each separately and see file sizes, print out both on a page with file sizes as labels for each one - work out why they different. Why are they not the same size? Explain.

(ii) flatbed scanner files at different resolutions - scan the printout of (i) - is it different too?

To be completed by - .................................................................

Software for organisation includes -

- paint and draw software for images
- mixing software for audio
- animation software for video
- word processors and desktop publishing for text
- spreadsheets for numeric data
- organising software to put data into tables, hypertext and hypermedia
- software to convert data from one format to another

Exercise 2.7 complete the following -

1. design a spreadsheet to suit the data in the file 2.7 DATA, then enter the data and see what the program could be used for
2. design a database to suit the same data in the file 2.7 DATA, then enter the data and see what the program could be used for

Print out a half page comparison of the strengths and weaknesses of each of these application programs for organising the same set of data (bottom of part 2 page).

To be completed by - .................................................................
You may have found by now that files created in one program may or may not be opened in other programs. For example, a picture created on a Windows computer in a particular graphics program will open on a Macintosh in some programs but not others - and in fact will not open in some other programs on the same Windows computer. The common graphics formats include - TIFF, PICT, CGM, EPS, GIFF, BMP, JPEG, WMF, DXF, etc.

That is why such programs as Graphics Converter are so important in allowing the user to grab files from a whole range of formats and include them into documents or presentations wherever they may be required. However, even such programs cannot convert from all formats to all other formats.

**Non-Computer Tools for Organising**

We have used hard copy systems such as telephone books, catalogue cards in libraries and pen and paper forms in business, etc. for many years. Much of the historical record of the world is still on paper (or papyrus or stone, etc.) although many of the earliest movies for example have been transferred to digital media - so could now be distributed on DVD for instance.

Sometimes data stored in paper form is the most convenient for use e.g. recipes can be kept on computer, but since few kitchens have a computer in them, it makes little sense to do so. Since many people need a manual to refer to quite often, they need pages printed out (and sometimes pinned up). Models are still used to test things before full scale production begins.

If data on paper is worth keeping, then it is normally in book form, or as files in a cabinet or as listed above, in a set of cards or attached to the wall or notice board.

---

**Assignment**

1. Choose and describe 3 non-computer hard copy organisation systems (not ones named above), describing what they are for and how they achieve that result.

2. List 3 specific strengths of each hard copy system and also 3 weaknesses for each one, noting why they are in fact weaknesses.

3. Discuss what computerised systems have or could replace each one of these.

   To be completed by - ..............................................................

---

**Social and Ethical Issues**

Current trends in organising data include such techniques as Multimedia presentations and Internet WWW displays and access. Software available these days allows data to be organised and displayed in a variety of ways. e.g. spreadsheet programs normally have a charting facility which will produce a wide range of colourful graphs from a table of data - with little effort from the user.

That people now access the original work of others on the Net, copy all or some of it and incorporate it into their published work (often without asking or even acknowledging the creator of the original work) is a real copyright nightmare. Whether the published work is a business presentation, a Web presentation or a magazine article makes little difference to the ethics of the situation, though legally differences are apparent.

Companies and individuals regularly collect lists of people for specific purposes - social, ethnic, income, etc. and sell these lists to anyone prepared to pay for them. Thus richer people may be targeted for expensive items, while advertisements for high powered computer hardware will
only be sent to large companies, for example. The people do not know they are on these lists, or whether they are accurate, and generally cannot get themselves removed from them (since the list may be sold to 50 people and they may each sell to others).

One Internet individual has advertised (February, 2000) that he can sell people lists of all Internet users - all 57 million - for prices varying from $149 to $249 (depending on details).

If data is poorly organised (i.e. redundant or repeated data) in a database, then the database will be larger than needed and therefore slower to work with. If a mailout is based from such a database, then inaccuracies are more likely to occur.

The ‘millennium bug’, based on a two digit date instead of the full four digits is an example of such an inbuilt problem in a wide range of software. Even if the computer could handle the date problem, many of the pieces of software used (either singly or in combination) had real problems with scrambling valuable data. Macintosh computers, for example, had no trouble with going from 1999 to 2000 but some of the programs - such as FileMaker Pro, were found to have messed up some dates - whether in the short form 1/2/99 or the longer one 1st May 1998.

The stock market crashes of the late 80’s and early 90’s were a result of computer programs for analysing trends in the share trading area, being too finely set. Once one program reacted to an overly large sell or buy, this set off other programs and the whole of Wall St overreacted and large fluctuations and many millions of dollars lost, as a result.

**Analysing**

Analysing is the process that transforms data into information, and it is the use of computer hardware and software combining to analyse the huge amounts of data collected today. Because of the volume of available data, it is up to the interpretation of this data to yield the type and format of information required.

In large stores, for example, millions of data items will accumulate per month (all purchases, all sales, etc.) and this may be either simply kept stored (archived) or interpreted and made use of. Thus sorting, searching and producing tables and graphs are all parts of the analysing stage. It must be noted though, that real processing of the data has not yet occurred - as the data has not undergone any real manipulation or editing.

**Assignment**

1. (a) What is data warehousing? Why and how is it done?
   
   (b) Of what use is this warehousing of data to a company?

2. (a) What is data mining? Why and how is it done?
   
   (b) Of what use is this to a company?

3. (a) What is the SETI home project to do with? Why has it been set up?

4. (a) What is ‘fuzzy logic’?
   
   (b) How and why is it useful in analysing data/information?

   To be completed by - .................................................................
**Hardware Requirements**

For individual and small group information systems, the use of database queries and spreadsheet tool options will generally cover most of the analysis requirements. Large businesses with huge databases and millions of data items will struggle with spreadsheets, and even powerful databases will have problems with at least some queries.

The more data involved, the greater the requirements on CPU speed and speed of retrieval. The following list includes the most critical requirements when analysing data -

- CPU speed - higher speed means faster analysis
- Data transfer rate - to and from secondary storage, must be fast enough to match the CPU speed to deliver data fast enough
- Primary storage size - as large as possible to contain the data
- Secondary storage size - large enough for temporary storage of intermediate results
- Time limits - the faster the job must be done, the higher the needs for the 4 above
- Complexity of the analysis - is determined by the efficiency of the algorithm used

All of these requirements must be balanced, with the cost involved against the value of the analysis provided. Large volumes of data on underpowered computers will take a long time, while million dollar computers working on small amounts of data is wasteful.

**Software Features**

As previously mentioned, databases and spreadsheets provide a range of tools to allow -

- searching/selecting
- sorting
- querying
- modelling/simulations
- what-if scenarios
- charts and graphs.

Efficient analysis of data depends on the quality of the data collected, and the tools of analysis assume the data to be correct and error free.

**Searching** - as the amount of data increases, so the problem of locating any one piece increases. This ranges from a simple “Find” in a database program through to highly complex Search Engines used for locating resources on the Internet. Text and numeric data are generally quite simple to search for, with images, sounds and video much more difficult to locate.

**Selecting** - is normally done from within a database, and is done with a query - which selects certain data (or restricts the data displayed), according to the conditions of the query.

An example of query language -

```
SEX = “MALE” AND SUBURB = “FAIRFIELD” OR “WAKELEY”
```

The query language is based on - field name / s (and their data)
several operators (the = sign to give an exact match, AND to make two conditions, the OR to allow either of two alternatives).

**Operators** may be of two kinds -

- relational
- logical
(i) **Relational** - to select and display the part of the database to satisfy the query -

- **=** (equals) to give an exact match e.g. Age = 16 - selects those exactly 16
- **<>** (not equal to) to give the opposite e.g. Age <> 16 - selects all ages but 16’s
- **>** (greater than) to search above a set amount e.g. Age > 16 - selects 17’s on up
- **<** (less than) to search below a set amount e.g. Age <16 - selects 0’s to 15’s
- **>=** (greater than or equal to) to search up from a set amount e.g. Age >= 16 - selects 16’s on up
- **<=** (less than or equal to) to search down from a set amount e.g. Age <= 16 - selects 0’s to 16’s
- **CONTAINS** - the string of letters contains that wanted e.g Colour CONTAINS Blue - selects all colours with the term Blue in them (Sky Blue, Blueberry, etc.)
- **BEGINS** - allows searching for the first letter of a string of characters in a field e.g. Colour BEGINS ‘Bl’ - selects all colours starting with Bl (Blue, Black, Bloodred, Blueberry, etc.).
- **ENDS** - allows searching on the last letter of a string of characters in a filed e.g. Fruit ENDS ‘berry’ - selects all fruits ending in berry (blackberry, strawberry, boysenberry, mulberry, etc.).
- **EMPTY** - allows searching for empty fields e.g. Colour = ‘’ (the quotes together showing an empty string).

(ii) **Logical** - are used to link selections in the query to create a more detailed query.

- **AND** - links two queries so that both conditions must be satisfied e.g. Sex = ‘Male’ AND Height > 170 (this would select all males over 170 cm tall)
- **OR** - links two queries so that either condition must be satisfied e.g. Sex = ‘Male’ OR Hair = ‘Short’ (this would select all males, and those females with short hair).
- **NOT** - selects the opposite of the condition e.g. Sex = ‘Male’ NOT Hair =‘Red’ (this would select all males who do not have red hair).

**Sorting** - basically achieved easily with text or numbers (alphanumeric data), with both the files and the data within them being able to be arranged into a particular order. Data can be sorted into either -

- ascending (abcdef ... or 123456 ...)
- descending (... fedcba or ... 654321)

Sorted data is quicker to work with than unsorted, and the number of fields to be arranged will of course make the process longer (more fields to arrange = more time of processing). Some database applications have the ability to add indexes as an additional field, which indicates the sorted position of each record, while the simpler databases have to complete a physical sort every time. This takes time to process each time it is done.
Modelling/Simulations - if a prediction is needed for some purpose e.g. scientific research, town planning, etc. then computer software can be used to simulate the situation and predict the likely outcomes. If all of the possibilities are included and their likelihood weighted correctly, then the model should be quite a good predictor of what will happen in the future. Artificial intelligence programs are good for this, as to a simpler extent are the ‘what -if’ possibilities of the spreadsheet.

What-if scenarios - if a spreadsheet is constructed properly, then the changing of a few of the numbers will have an effect on many of the others. For example, in the Coffee Shop example following, if the tax rate changes, then the changing of one number will cause the program to recalculate all of the other figures for tax and take home pay automatically. Therefore the spreadsheet user can see what will happen if something alters or is planned to be changed, and rather than make a dramatic change and see what happens, the likely success of the changes may be predicted.

<table>
<thead>
<tr>
<th></th>
<th>COLOUR</th>
<th>OLD</th>
<th>NEW</th>
<th>TOTAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Yellow</td>
<td>23</td>
<td>46</td>
<td>69</td>
</tr>
<tr>
<td>3</td>
<td>Red</td>
<td>45</td>
<td>62</td>
<td>107</td>
</tr>
<tr>
<td>4</td>
<td>Blue</td>
<td>17</td>
<td>43</td>
<td>60</td>
</tr>
<tr>
<td>5</td>
<td>Orange</td>
<td>94</td>
<td>29</td>
<td>123</td>
</tr>
<tr>
<td>6</td>
<td>White</td>
<td>75</td>
<td>28</td>
<td>103</td>
</tr>
<tr>
<td>7</td>
<td>Green</td>
<td>35</td>
<td>89</td>
<td>124</td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Visualising - if you have ever tried to gain a full understanding of a full table of data, and tried to make comparisons, then you would understand how difficult it is to work out relationships. If a chart is used (line, column or sometimes pie) then the general trends can be seen at a glance, hence the use of charts and graphs particularly in presentations.

Exercise 2.8 complete the following -

1. For the database CD Now, complete the directions given and answer the questions on the file named CD Now.To Do - and print this page out.

2. For the spreadsheet Coffee Shop, complete the directions and answer the questions, add your name, then save the file - and print it out.

3. Using the Draw part of ClarisWorks (and some use of the Paint part), create an A4 title page for the unpublished novel The Clock (read Chapter One in the file), including a small hand drawn graphic (not clip art). These to be put together in ClarisWorks draw and printed out on the laser printer.

4. For many situations either a spreadsheet or a database may be used to analyse the data. List the strengths of each analysis tool, and suggest which would be most effective in your opinion for given data types. Give reasons for your choice.

To be completed by - ..........................................................
Non-Computer Tools
If we are to compare the effectiveness of computer databases and spreadsheets, we need to compare the old manual system of completing similar tasks. We may therefore compare -
1. searching a manual database versus to a computerised one
2. using manual simulations and models versus computerised ones

Note - you can try this by working out the time to complete parts 1 and 2 of Exercise 2.10 manually on a card system for CD Now and using a calculator and photocopied sheets for the Coffee Shop.

Social and Ethical Issues
Data analysis by computers is a huge business and a major area of computing, and the social and/or ethical problems generally arise from either -
- the unauthorised access or incorrect use of data,
- the incorrect analysis of data, or
- the use of database linking and cross checking

Data should only be used for the purpose it was collected for. For example, the Australian Shoppers Survey you completed in an Exercise in Topic 1 was designed to collect a range of data about buying habits for a variety of products (including pet food, holidays, reading material, newspapers, motor vehicles, mobile phones, etc.). It also contained questions about home loans, credit cards, health cover, illnesses, etc. and the people who collected the data could sell parts of it to Banks, Building Societies, Health care companies, etc. Such companies could sort through the data and target people by income, illnesses, etc. and deluge them with mail, phone calls or even call at the door to sell their product to a likely group of people.

Hopefully when such information is sold, it is analysed correctly, for entry or sorting mistakes have caused people in the past to go through a good deal of needless hardship. This includes missing out on getting loans - since a list of poor credit ratings was mixed up, or the person has the same name as another. Database linking and cross checking has been known to match up names with histories, addresses, and other data on people simply because one item in a database contained a same or similar item of data.
For example -

DB1 has  NAME: Peter Jones   POSTCODE: 2165   D.O.B. 17/05/82

DB2 has  NAME: Peter C. Jones  CREDIT RATING: Poor  D.O.B. 17/5/82

By linking and cross checking these two databases (and not carefully checking each entry) it is easy to see how Peter Jones can be given a poor credit rating.

Storing and Retrieving

This has two steps to the process, the first is the SAVING of the data/information(this is normally from primary to secondary memory), while the second RELOADS it (this is from secondary to primary memory) to allow for -
- other processing to take place
- a temporary halt in the system
- backup and recovery
- the transfer of data/information
You need to remember that - PRIMARY memory is the temporary storage of the data within the computer chips of RAM (random access memory). SECONDARY memory is permanent (or at least long term) storage of the data on devices connected to the computer.

**Assignment**

Each student (or pair on the computer) is to research the types of **secondary storage devices** available to personal computers in computing textbooks, in libraries, on the Internet, etc. and complete -

1. A table consisting of - names, storage capacity range, access time and range and approximate cost range for each storage device.
2. Select TWO from your table and describe how they work in some detail, including scanned in diagrams where necessary.

The printouts are due to be handed in by - ..................................................

**Hardware**

As you will find in the completion of the assignment above, the range of secondary storage devices (or memory in some books) is growing wider and is based on a number of different technologies - including - magnetic, optical and flash. You will discover the range of possibilities for many of these devices.

**Characteristics of Storage**

The three characteristics which distinguish storage technologies are -

- **Access**
- **Volutility**
- **Permanency**

**Access** - data is put onto the storage medium either directly (termed random) or **sequentially**.

Random access allows you to go to the data directly, while sequential access means you have to go in order through the data until the required part is located.

Examples -  Direct - chips, disks, etc.  Sequential - tapes

- Sequential is generally much slower (since you may have to go through all the tape, for example to locate the data) - so access time is measured in seconds or minutes.
- Direct access times are measured in in milliseconds (10^-6 seconds) or nanoseconds (10^-9 seconds).
- Generally RAM has access times measured in nanoseconds
- Secondary storage, like hard disks, have access times measured in milliseconds

It may be that it is 1000 times quicker to access data in RAM than in secondary storage, and 1000’s of times faster to access data from hard disks than from sequential storage such as tapes.
Volatility - this refers to the necessity to maintain power to keep contents stored. Thus RAM is volatile and the contents or data held in the chips is lost when the electricity is turned off. When the computer is running, RAM contains both the program running and the data it is working with, and both disappear as soon as the power goes off. If secondary storage is not used, then the work is lost forever.

Many recent programs have an automatic backup function (usually set to 10 minutes unless altered) so that the contents of RAM is copied to the disk, CD, etc. of secondary storage.

Permanency - paper has been the storage method for around 2000 years, though sheets of paper rarely have to last for more than a few hundred years. Floppy disks are not volatile, but will not reliably contain data for anywhere near hundreds of years (five to ten maybe), so the idea of ‘permanent’ needs considering. We now know that magnetic materials slowly lose their strength and the data will be lost over a period of years.

In the mid 1990’s a scare went around that the surface coating of CD’s was not going to last for more than a few years (even for the ones guaranteed for 99 years!). We now know that, so long as they are stored correctly, optical disks have a very long life - maybe hundreds of years.

Trends - The final characteristic of storage over time is that the trend is to faster and faster access times and greater and greater storage capacity. Every year the tendency is to double both the RAM and the secondary storage capacity, while the access times have decreased at a lower rate. Centralisation of the storage of data is another trend for the start of this century.

Software for Storage

• Hardware interface - Obviously software controls how data is stored on these hardware devices, as is the transfer rate to and from them. Control software is often stored in the ROM chips (termed firmware). Thus when the system software loads at bootup, it communicates with the software on the storage device.

• File management - operating system software usually provides basic file management procedures, which can be used to -
  - Browse - search through stored files
  - Search - for particular files - by a number of criteria e.g. content, creation date, etc.
  - Manipulate - delete, move, copy, compress, etc.
  - Display - preview, catalog, send to printer, etc.
  - Manage - error checking, connect to network, etc.

Exercise 2.9 complete the following -

1. (a) Expand the following file formats and
   (b) State whether each is for images, audio, video or alphanumeric data.

1. MP3 - 2. TIFF - 3. HTML - 4. WAV -
5. RTF - 6. PDF - 7. MPEG - 8. JPEG -
9. GIF - 10. ASCII -

2. Define and explain how data dictionaries are used in databases.

To be completed by - ..........................................................
Database management - databases may be structured in two main ways, as either a -

- **flat file database** - with a single table of data, or a
- **relational database** - with a number of tables linked in set ways

Simple databases are normally flat file, while larger more complex databases are usually the more powerful relational type.

Databases deal with a range of **data types**, namely
- **character**
- alphanumeric (or string) - numeric
- logical - binary

Each data type is kept in **fields**, each with a unique field name.

Flat file DB’s - such as ClarisWorks or Microsoft Works -
- are usually easier to use
- are easier to set up
- need less powerful hardware to run them

Relational DB’s use links between the tables to create a relation, using a primary key (or unique identifier) such as an ID number to allow the tables to be related. FileMaker Pro, Foxbase and 4D First are examples.
They are -
- harder to plan and set up
- harder to use and understand
- need more powerful hardware to run them

**The Internet**
As we saw in Topic 1, the Internet is a world-wide store of data/information and the WWW (World Wide Web) is the simple point and click browser part of it. This browser is **machine independent** - i.e. no matter what brand or model computer you use, it can be connected to the Web.

This means that files are all compatible and data can be retrieved and incorporated into user documents much more easily. Thus HTML (HyperText Markup Language) is the text-based language used, with GIF and JPEG the other formats, standardising the types.

Search Engines (searchable database software) are used to help wade through the enormous volume of data/information, gathering bits and pieces using software tools termed crawlers and bots. Therefore each search engine may contain different information and so a range of these may need to be used to find the full information.
Examples - Yahoo, Alta Vista, Excite, etc.

**Encryption and Password Protection**
These are software programs to safeguard data stored in places (networks, shared computers, etc.) where insiders or outsiders may want to either - copy the data for their personal use (theft) or alter the data to disadvantage the owner of that data (industrial espionage) or even destroy the data (sacked employee getting even or industrial espionage).

In most large companies data is transmitted along telephone lines from one computer to another, and it is not possible to prevent the tapping of these lines by thieves, hackers, competitors or anyone else determined to access the data.
Data **encryption** is the scrambling of data by using an algorithm, usually at one end of a transmission line (within a computer) and then **decryption** is the unscrambling process performed at the other end of the line so that the data may be used on the other computer. How successful the encryption is, is dependent on the sophistication of the algorithm used.

**Passwords** are required by virtually all information services (Yahoo, Excite, etc.) and are made up generally of a combination of User name and Password. If the user does not enter both of these correctly at login time, then access to the system is denied. Most such systems have minimum and maximum lengths for these, often 5 to 10 characters. Programs may be added to your home computer to password protect either your whole hard drive, your applications or your files (or some of one or more of these).

---

**Non - Computer Tools**

These systems include -

- **Paper based** storage systems which are bulky and require large storage areas, and also staff with library skills are needed to manage them. Thousands of filing cabinets are still in use, and who knows how many pieces of paper have been misfiled and effectively ‘lost’ in the system.

The number of paper copies need to be kept to a minimum and inevitably some are lost (sometimes the only copy), or damaged or destroyed (water, fire, pests, etc.). Control of paper files has always been something of a problem, with numerous people needing access to the same files, with confidential sections within some of them. Obviously only one person can use a paper file at a time, the file needs to be organised in some way (and how this is done can be a problem - by date, by purchases, by location, etc.)

- In order to reduce storage space, microfilm was used to photograph papers and store the data in a compact form. One such type was termed **microfiche**, which used sheets 15cm X 10cm, each containing around 200 pages of printed information. A reader, somewhat like an overhead projector, was needed to read the sheets (magnifying them up onto a vertical screen).

Finding the right page(s) was something of a problem since no automated search was possible, and no editing of the data could be done. Some companies still use this method - e.g. Peter Warren Fairfield spare parts department.

- **Libraries** continue to provide a range of information storage and retrieval tools. Some is done on computer in all libraries these days, but some are still paper based in the form of books - which will continue to be used because books -
  - are portable
  - need no special tools to read
  - are cheap to produce
  - last for several hundred years
Social and Ethical Issues
These issues mainly concern -
• **The security of stored data** - data must be protected from corruption, loss and unauthorised use (including privacy issues). The data belonging to a company, for example, should only be accessible by people allowed to see and use it - some of the data may not be available to other employees of that company. It should be saved into several different locations (say in case of a fire or a hard disk crash) and passwords used for different levels of entry. As mentioned above, illegal entry to stored data may be either from internal or external sources.

Companies or government departments selling or swapping data are of real concern to the privacy of each person - data matching could allow the pattern of a whole life to be discovered.

• **Unauthorised retrieval of data** - many crimes have been committed using computer technology, including the theft of data from NASA and other government agencies in the USA. Hackers have found such tasks a challenge in recent years - though fines, confiscations and prison sentences have made this a less attractive pastime recently. Military and banking institutions are still prime attractions for computer crime attempts.

Employees or ex-employees are also of concern, either innocently or maliciously trying to enter company files to ‘check on things’ or ‘fix up something’ etc. If you think about it, an employee could alter their own files to aid in getting a promotion, or increase their rate of pay. Similarly, a student could alter their grades and assessments if they could gain access to the school files.

**Exercise 2.10 complete the following** -

1. Computer crime has been considered ‘white collar’ crime and until recently, the penalties have been mild compared with say armed robbery or burglary. Give your opinions on the relative effects on society of these two forms of crime.

2. The main forms of computer crime have been -
   (i) theft or embezzlement
   (ii) vandalism or sabotage
   (iii) blackmail or extortion

Choose two of these three pairs and research what actually occurs in the computer crimes. Print out a report on each of the two you have chosen, including examples.

3. Outline the methods of prevention you would use to protect the valuable files of Daly Enterprises, an innovative IT company.

4. What does the Freedom of Information Act of 1989 allow the people of Australia to do with regard to data/information held on them?

   To be completed by - .................................................................
Processing

Processing is the manipulation of data, by editing and updating it. This editing might be updating a change of address, adding in latest sales figures or creating the final version of a multimedia presentation. Many of the exercises you have completed in this topic are of course processing.

As we have seen, the type and degree of processing needed should match up with the computing power available. The things to consider are -

- CPU speed
- Primary storage
- Secondary storage
- Data transfer rate

Examples -
1. A 200 page mainly text document would not put a strain on CPU speed and would require less than 1 megabyte of secondary storage.

2. A dozen A3 full colour posters could, in 24- bit colour, be 20 megabytes each, and could occupy 500 MB to 1 GB to complete all of the stages (scanning, versions, etc.). Demands would be very much higher on CPU speed, manipulating these large graphics.

Types of Processing - include -
- Centralised - with control of processing from a central computer
- Distributed - with work stations connected to shared resources at a local site or elsewhere
- Parallel - with simultaneous processing of instructions with multiple CPU’s

We can look at each of these as separate, but in reality they are generally combined - such as a centralised computer with parallel processing architecture. Each model has strengths and weaknesses and so the one(s) chosen will be determined by the current situation where it is to be used.

Centralised - early computers were terminals connected to a central computer performing all of the processing. The personal computer took control to the individual user, but then networking became important (to share data, software, technical expertise, etc.).

Thus organisations are now going back to a centralised model, with only a small amount of software on the desktop computer, so the individual user can only mess up to a limited degree - so fewer technical staff are needed. Only one copy of the software is needed and every user has the same version.

Advantages - single point of management
- high degree of physical and technical security
- fast installation of applications (one network version)
- easier remote access

Disadvantages - reduced user independence and initiative
- all work stops if central system fails
- expensive powerful CPU’s needed
Distributed - this method allows the splitting of the workload between computers, with subtasks to all contribute to the full task. This requires a well structured environment where hardware and software communicate and share resources.

Connecting computers via networks is typical of distributed processing today, with each computer having local input, output, storage and processing capacity. Data/information can be exchanged with other computers on the network, as well as central servers that hold centralised large data storage.

Examples of a distributed processing system are -
1. an organisation with a head office and branch locations e.g. Big W. Processing is performed both at the local branches and at head office.
2. the SETI system you looked at in Ex 2.9, where in February 2000, about 400 000 people are processing the SETI data.

Advantages - increased freedom to perform different tasks  
- branches can still work if the head office system is down
- greater flexibility, with local and/or centralised processing

Disadvantages - greater chance of redundant data  
- less security
- more technical staff needed

Parallel - with more computers processing alongside each other, the overall processing goes faster.

This can be of two types, at the -
• Instruction level - where parallel instructions are executed within a single processor - with duplicate internal logic circuits
• Thread level - where parallel instructions are executed within a single computer - with multiple processors

It is a challenge to actually split a task into equal parts matching the number of processors e.g. if 4 processors, the job 4 equal subtasks if the processing is to be anywhere near 4 times the rate of one processor. In fact, communication between processors, synchronisation, and other functions means that as the number of processors increases, so the speed increase drops off.

Advantages - greater processing power without special CPU’s
- many inexpensive computers can be linked to give very high processing

Disadvantages - needs a special operating system
- applications programs need to be specially written
Hardware

The CPU (Central Processing Unit)
If the computer was intelligent, and therefore had a brain, the CPU would be it. Since the computer is not intelligent, the CPU is really a large integrated circuit (IC) or microprocessor (or series of these in some computers), where most of the processing takes place. The basic components of a processor are -

- timing, controlling and co-ordinating devices
- device for calculating and making decisions
- several temporary storage devices
- connections to move data between the various parts

The CPU has a number of parts -

- **Control Unit** (CU) - this part carries out the program instructions, acting like a policeman directing traffic, in directing the movement of data and control signals between devices e.g. data to be stored in memory before it is processed. With different software instructions, and different hardware components, the processor has enormous flexibility and power to perform.

- **Arithmetic / Logic Unit** (ALU) - as implied by the name, this part is to perform arithmetic operations and make logical decisions. Thus it can add and compare numbers (equal, larger, smaller).

- **Registers** - these are temporary high speed storage devices, located within the CPU, to perform such things as -
  - program counter (stores the memory address containing the next program instruction)
  - instruction register (stores the instruction currently being decoded and executed)
  - accumulator (stores the intermediate and final results of calculations)
  - data register (holds the last block of data going from or to memory)
  - address register (hold the location co-ordinates of each piece of data or instruction held in main memory)
The Fetch - Execute cycle - the CPU executes instructions in a continuous cycle, with the parts carrying out the required operations.

This is also called the machine cycle in some books.

Cache (pronounced ‘caysh’) - this term refers to a small amount of fast chip memory holding recently accessed data. It is used to speed up further access to that same data - and the effectiveness of this is measured by what is termed ‘hit rate’. This is the % of memory requests that are satisfied by the cache, and the larger the cache the faster the processing.

Cache memory chips are faster than normal main memory chips. Cache located on the CPU is termed Level 1 cache (or primary) while that on a separate chip is Level 2 cache (or secondary), with primary being faster. Level 1 cache is normally 64 to 128 K, while Level 2 cache is 512K to 2MB.

Types of CPU
There are two main types - RISC (Reduced Instruction Set Computer)
CISC (Complex Instruction Set Computer)

CISC - have a larger set of more powerful instructions with more levels of logic and therefore are slower in execution per instruction.

RISC - have a less complex instruction set, with less levels of logic and hence a reduced instruction execution time.
Exercise 2.11 complete the following -
1. Why do we still have 2 quite different types of CPU’s in computers?

2. Why are computer chip manufacturers (like Intel and Motorola) continually trying to create faster processing CPU’s?

3. ‘Smaller, better, faster” is the cry normally associated with processor technology. What does this mean?

   To be completed by - .................................................................

**Processing Speed**

Increases in processing speed can be achieved in several ways, one of which of course is to increase the clock speed of the CPU (i.e. the number of pulses sent out by the CPU to co-ordinate its operations). This is measured in Megahertz (MHz) where 1 MHZ is 1 million cycles per second. Thus a CPU may have a clock speed of 400 MHz or be faster at 600 MHz, with 1000 MHz speeds soon possible - though the laws of physics will limit further increases.

The second method is to increase the bus capacity of the system (i.e. the main communication channels through which the CPU, memory and other components communicate and transfer data) thus allowing greater rates for the movement of data. Current fast systems use 200 MHz bus speeds.

Other techniques to try and improve speed have included reducing the number of clock cycles per instruction and increasing the size of data width processed internally - 64 bits instead of 32 for example. Performance is often measured in MIPS (millions of Instructions Per Second) or MFLOPS (Millions of Floating Points per Second).

**Non-Computer Tools**

The processing of data needs to follow some rules and procedures, otherwise when the people in charge of the initial operation are gone, then it is really difficult for someone else to take over. The planning of this procedure is generally human based with text and pictorial diagrams including -

- data flow diagrams
  - system flowcharts
  - decision tables
  - decision trees

**Assignment**

1. From pages 248 to 250 of the white Boyd textbook (or others) make notes on the 4 methods listed above, including any necessary diagrams.

2. Find out about the full computer network at Westfields and construct a diagram to represent it.

   To be completed by - .................................................................
Social and Ethical Issues
Due to the types of computers on networks and the different amount of flexibility they allow users, social issues of sharing data, remote as well as local access, etc. have created a number of issues.

The main issues at this stage of data processing are:
- security
- ownership of data
- bias from processing

**Security** - in a centralised model data security is quite high, while in a distributed arrangement security is low, since data is accessible to more people. Also in networks data can be processed into the system without approval, as well as increased chances to copy and sell data or damage or alter it in some way.

Distributing data and providing telecommunications links allows access to workers, but also opens the way for unauthorised access. Hackers may just look at data, but could copy it, alter it or plant viruses in it if they wished following illegal entry into the system.

**Data Ownership** - individuals sometimes feel that the data is theirs on large networks, and process it as they wish (which may not fit with the plan for it). This may lead to branches offering different deals, prices, etc. when head office has advertised sales or other programs for all of the branches.

**Bias** - if data contains bias at the time of collection or organising, then this generally produces flawed information. Similarly, biased assumptions made at processing time can produce incorrect results. For example, data collected on teachers could be linked to home ownership and processed from there. This would ignore all teachers who rent or lease and would therefore be biased. If processing is performed without preconceived notions, but clearly thought out and logical ideas, then the resulting information will be accurate and useful.

**Transmitting and Receiving**
This involves transferring information/data within and between information systems, thus it involves communication using computer-based systems. Over the last few years though, the difference between work and play using computers has overlapped somewhat so that at least some people regularly use communications to -
- telecommute or telework
- shop, bank and buy music
- read international newspapers
- exchange e-mails
- videoconference
- book holidays
- obtain information, technical help, sell stocks, etc

**Communication Concepts**
We all communicate with each other by speaking, listening, touching, tasting, etc. and some other methods like drawing, making gestures, looks, etc. All of these involve a sender of the message and a receiver of hopefully the same message, along with the message itself and the medium through which the message is actually carried.
When this is performed by devices instead of people, a range of technical concepts need to be considered for useful communication -
- message preparation - termed encoding
- message reading - termed decoding
- error handling - if the message is not understood by the receiver

Example -

![Message Flow Diagram]

Obviously, successful communication requires a common language and an agreed set of rules for that language - termed PROTOCOL for electronic communication. This is similar to two people working out that they can both speak a number of languages, though only one in common - and so deciding to speak French for their conversation.

The flow of data must also have some type of control, so that the transmitting device knows when the receiving one is ready to accept data, and when error checks will be done in the opposite direction. The signal used to achieve this control is termed HANDSHAKING. This is like two people talking on a two way radio (where only one can talk at a time, and they say ‘over’ to let the other one know they have finished their end of the conversation for the moment).

The two types of Handshaking are -
(i) hardware - where a dedicated extra wire connection is made to indicate whether data should be transmitted or stopped.
(ii) software - where codes are sent along the data carrying wires to do the starting and stopping of data flow.

**Data Transfer**

When computer systems are connected ready to transfer data, we term them ON LINE, and when we disconnect them, they are OFF LINE.

Data transfer can be achieved in two different ways -

(1) Parallel transfer - where all data bits arrive or leave at the same time. A wire is needed for each bit, so 8 wires are needed to transfer the 8 bits in a byte - see ribbon cable example.

![Parallel Data Transfer Diagram]

This is the method used in the data bus and to peripheral devices like disk drives, scanners and some printers. Parallel transfer only works effectively over a couple of meters distance, because the data bits can get out of order (data skew) on the way to the receiver if the cable is longer.
(2) Serial transfer - uses a single connection and the data bits move along one after the other. It is therefore not as fast as parallel, but is used in telephone lines, some printers and in networks.

Serial transfer can be done over long distances but the signal does fade with distance and needs to be amplified or boosted along the way.

There are 3 ways in which data can be transferred -

1. SIMPLEX - one way (sender to receiver) e.g. radio, TV, to keyboard or monitor

2. HALF DUPLEX - either direction, but only one way at a time e.g. two way radio, to/from disk drives

3. FULL DUPLEX - both ways at the same time, so both ends are transmitter and receiver. e.g. telephone line (but hard to understand), motorways, etc.

**Data Transmission Modes**

The transfer of data/information is either -

- **SYNCHRONOUS** - occurring at regular intervals. Communication within the computer is generally synchronous, with the synchronising done by the CPU clock. This method can be compared to a train with many carriages, all travelling at the speed, closely joined and exactly synchronised i.e. all the bits of data travel every now and then in the train carriages all together (until the next train).

- **ASYNCHRONOUS** - occurring at irregular intervals. Communication is intermittent say between two people sending e-mails. This method can be compared to cars all travelling in the same direction down a highway, all going with similar speeds, but with different distances between them and each car travelling on its own i.e. the bits leave at different times but all the data arrives in the cars at the destination.

**Synchronous Communication** - involves synchronising the devices so that accurate reception is made of the data sent out. This allows the stream of bits to be separated into blocks which can be processed (timing by the CPU clock). For example a 400 MHz CPU clock speed and a data bus speed of 200 MHz will mean that for every 2 clock cycles, the CPU can receive a parcel of data.

Data transfer within the computer must be synchronous if any sort of speed is to be achieved.
Asynchronous Communication - is done when typing characters on a keyboard, with the speed of transmission varying as the typing speed varies. A start bit is followed by the data bits of the character and then finishes with a stop bit(s) - so the computer can identify each character.

For asynchronous communication to take place, a number of parameters must be set with the software before establishing the data link, including - number of data bits (often 8), start and stop bits (usually one of each), parity bits (used for data checking - odd or even), rate of transfer (measured in bits per second ‘bps’) from 1200 bps to 56 Kbps.

Transfer Speed
This term refers to the maximum possible rate of data transfer, as measured in bits per second (bps) or in bytes per second (Bps). The other prefixes of Kilo (thousand), Mega (million) and Giga (thousand million) are also used with reference to data transfer speed or BANDWIDTH.

Parallel connections give higher transfer rates -
• up to 3 Gigabytes (GBps) per second for data buses
• up to 160 Megabytes (MBps) per second for storage devices

Serial connections achieve lower transfer rates but extend around the world -
• up to 1 Gigabit (Gbps) per second for fastest local area networks (LAN’s)
• only 10 Megabits (Mbps) per second for the slowest - like Westfields at present
• only 56 Kilobits (Kbps) for modems

Though serial connections are slower, it is possible to join several of these together (termed aggregating) to achieve faster rates - though not quite the sum of those added but closer to double the bandwidth for two aggregated. High bandwidth T-1 (1.544 Mbps - using 24 lines) and T- 3 (43 Mbps - using 672 lines) links are created from aggregating ISDN lines.

Direct Connection
Data may be transferred directly from one computer to another via a serial connection (null modem cable) or a parallel connection (over a short distance). The same communication software and protocols must be used for this, and special plugs called DB-9, DB-25, mini-DIN and USB are used on the ends of the cables.

Data sharing is often performed using direct connection, from desktop to desktop and especially to and from laptop computers these days. We can share files on the Macs in Computing at Westfields through the printing network cables.

Network cards in the computers connected as Intranets (like at Westfields) allow communication from computers on the network to file servers and printers around the school.

Using the cable and software, data can be transferred to or from the laptop to keep version of shared work up to date for each person working on the task.
**Exercise 2.12** complete the following -

1. Research the use of local area networks for computers - such as we are developing for Westfields. Answer each of the following -
   (i) What are they really good for?
   (ii) What speeds do they actually work at?
   (iii) How many users can be on the network at once?
   (iv) What are nodes and what are they for?
   (v) What are the weaknesses or drawbacks of such networks?

2. Describe how the following are done -
   (i) A couple of pages with diagrams are copied from the Internet.
   (ii) A fax machine prints out a message from another country.

3. List 3 advantages and 3 disadvantages of -
   (i) use of the Internet
   (ii) use of e-mail

   To be completed by - ......................................................

**Hardware**

The hardware for communication and data transfer includes -

1. **Buses** - these are a set of wires or tracks on a circuit board or connections in a chip. Most PC buses are 32 or 64-bit, though 128-bit are becoming available, and it is the bus width and speed that are becoming the limiting factors in the overall computer’s speed. Several bus types are used -
   - data or system bus - operates from RAM to the CPU and is generally 100 to 200 MHz in modern computers.
   - address bus - used to identify addresses in primary storage (RAM).
   - expansion bus - used to connect other devices like internal modems, SCSI devices, video digitising cards, etc.

2. **Modems** - this term is from the two words modulator and demodulator, where modulation is the conversion of digital signals to analog ones and demodulation is the reverse - this change needs to occur from digital computers so that analog phone lines can be used to carry the signal, then switch it back around to digital again for the receiving computer.
In the process of modulation (digital to analog) the analog signal, called the carrier wave, must be altered so that it is able to transmit the digital equivalent. There are 3 ways this alteration can take place by -

- **Frequency** modulation, where the base frequency represents 1 bit, while the change in frequency represents the alternate bit.
- **Amplitude** modulation, where the height or amplitude of the wave is altered, with the base height representing 1 bit and the altered height its alternative.
- **Phase** modulation, where the relative position of one complete cycle of the wave is reversed to provide the alternative bit.

Several types of modems are in use -

- **1. Acoustic coupler** - older style, where the telephone hand piece fits into 2 cups in the coupler and the signals still actually go through the mouth piece and from the ear piece. Problems associated with this system include corruption or deterioration of the signal as a tone is involved, and that handsets are not all the same size or shape to fit the cups.

- **2. Internal direct-connect** - transmits the signal directly through the channel. Is generally circuits on a card plugged into the motherboard inside the computer, but can be built in.

- **3. External direct-connect** - same as the internal version, but all the circuitry is housed in a separate box equipped with telephone outlet and lead to connect it to the computer.

Most modems have specialised circuits built in to perform such tasks as controlling data transmission speeds and automatic dialling/answering - the degree of these features leading to terms like ‘smart’ modems. To improve on the 56 Kbps maximum of analog phone lines, other methods must be used, including -

- **ISDN** - with 64 Kbps, then aggregated up to multiple Mbps
- **Broadband cable** - connects to cable TV network, with 1 to 3 Mbps speed
- **Satellite and radio links**, up to multiple Mbps
- **ADSL (Asymmetric Digital Subscriber Line)** and **SDSL (Symmetric)** - use the analog line in a digital manner, achieving up to 8 Mbps.

**Software**

Years ago many small telecommunication networks were in existence, each operating independently of each other - with no way of interconnecting them. Now with the Internet in common use, everyone can be connected.

**What is the Internet?**

The Internet was created in the early 1960s by the Department of Defence in the United States of America. Other organisations, such as universities, research groups, NASA (the National Aeronautics and Space Administration) and the National Science Foundation, joined the network to encourage the sharing of research findings. A vast array of information is available on the Internet in the form of text, graphics, sound, video or multimedia files. Information can be as diverse as medical imagery, previews of the latest movie releases, government policy, poetry, scientific data, library catalogues or someone’s favourite recipes.
The Internet is a worldwide connection of many thousands of computer networks. All of the networks that comprise the Internet use a common language, Transmission Communication Protocol/Internet Protocol, TCP/IP, to communicate. They are all connected to each other through communications channels, many of which remain permanently open. The Internet enables people who have access to these networks to share information and knowledge.

The Internet is a cooperative community of networks, and thus nobody owns it. It is made up of many small parts in many different countries, and within each country, there is an organisation that supports the Internet and provides the main communications channels. Within the world, there is a group that coordinates the overall network. It is a facility which anyone can use. People often comment that the Internet belongs to everyone and to no one.

A service that runs on the Internet, the WWW or World Wide Web provides (via browser software) access to a huge amount of all kinds of data/information. The Web has only been part of the Internet since the early 1990’s and before that text and numbers were used exclusively (through Telnet), rather than the images, animations, movies etc. common to the ‘point and click’ generation of today.

Sending Data Through the Internet

Data travels through the Internet in the form of packets containing parts of the message and the address of the computer to which it is being sent.

E-MAIL or electronic mail is a quick and easy way for people to communicate using networked computers. This network may be quite small (from just 2 computers) to say the 16 computers in rooms 18 and 19 at Westfields, right up to hundreds of Telstra computers or even millions of computers across the world. Each user has an address which is based on their name and location and can send and receive messages across the network.

e.g. John Macintosh of Apple might have the address jmacinto @ apple.inc

How e-mail works - on the network system, one computer acts as the main host and each user has an electronic mailbox on the hard drive of this computer. When a message is posted, it travels along the network to the host computer and the mailbox to whom it is addressed, and is stored there until the recipient checks the mail. The e-mail program downloads the selected messages from the host computer to the user’s computer where they are stored and shown as a list ready to read.

Attachments are generally allowed, so that images, audio or videos for example can be sent as well, along with the usual text.
Networks
These are of two main kinds - LAN or local area network - WAN or wide area network

Small LAN’s may not have a dedicated computer for sorting and sharing files, and this type is termed a Peer network. If there is a dedicated computer to perform these tasks, then this computer is called a file server and the network is referred to as being Server based - these are generally a larger network than the peer one, but still over a small area e.g on one floor of a building.

The Internet is an example of a WAN - in that it covers a wide geographical area.

Both LAN’s and WAN’s provide a sharing of devices such as printers, and data. This means that multiple users can access the same file at the same time, and both make changes and save these one after the other. This would result in the last save overwriting the previous one. To overcome this a range of options is available to be set up -
1. only one user at a time can open a particular file
2. multiple users can read a file but only one is allowed to write to that file
3. multiple users can read and write but only to different parts of that file

Network Topologies - these are the different shapes and configurations for the layout of networks, and they are - Star, Bus and Ring.

Using a Gateway or Bridge, the different topologies may be joined to form a hybrid arrangement.

The most common network protocol is Ethernet (the blue cables from the back of Westfields computers) and these connect to Network Interface cards plugged into the computers to allow them to send and receive transmissions. The wires all go to either a hub or a data switch (located in a data cabinet - in some cleaners storerooms) to allow the network to be wired in a star topology, and the speed is commonly either 10 Mbps or 100 Mbps.
Non-Computer Tools
We have transferred data/information for a lot longer than computer technology has been around.
- Mail has been a quite reliable method for hundreds of years - though automated sorting has made the process faster and easier.
- Telephone calls have been around for over a hundred years - though most exchanges are automated and digitally based.
- Radio and television have been in use for 50 to 100 years - but both are based these days on computer technology (satellites, outside broadcasts, etc.)

Thus we have utilised computer technology heavily in the communications field - probably more than in any other, but some methods still work to a degree.

Social and Ethical Issues
Due to the ease with which huge amounts of digital data is moved around the world, now more people have greater access to the stored data/information. As stated earlier in some instances, the issues involved with this include -
- security of data/information
- accuracy of that data/information
- privacy
- changing nature of work, communications, business operations and global issues
- copyright issues

With data communications any computer system connected to the Internet or to a modem, can potentially be accessed by anyone else with a similar connection. They may need to get past passwords or encryption, but sometimes these are in fact published on some part of the 'Net'.

Viruses have enjoyed a rapid and widespread dispersal due to the Net, with games, demos, and e-mail now all being suspect for infection. Networked databases are prone to misuse or attempts at sabotage, and so access control measures must be used. These include -
- hierarchical level access - different passwords allow lower up to higher amounts of data
- data logging - which records all transactions and reports abnormal activity
- firewalls - to protect unauthorised access to files
- encryption - to protect stored and particularly transmitted data

Privacy - is a problem when networking means that files are vulnerable, and if unauthorised access in not detected, then no action will follow - this would be like being robbed and not knowing that it has happened. The Federal Privacy Act applies to information held by federal government departments and credit agencies such as banks - but how do you find out if such information may have been illegally accessed?

Using the Internet may, by itself, give out information on you that you may not have considered- surfing the Net and e-mailing can easily disclose private information. To restrict this possibility users should consider -
- checking that online forms are secure - that the browser connection is secure.
- E-mail - employers are legally entitled to read any e-mail on a company computer, so a different e-mail system for private e-mails, or even encryption software (like Pretty Good Privacy - PGP) could be used.
- cookies - these can track the user online and can link a lot of information - reject them.
Accuracy of data/information - data files available on the Internet have been placed there by a whole range of people (some good, some bad) and since it is not checked, the data could be incorrect or out of date. Therefore the user should cross check a number of sources to see that the data seems accurate. Encyclopedias are quite reliable since they have been through a thorough checking process before being marketed.

Changing Nature of Work and Communication - some people are monitored via computers in networks for the time they are at work, including - work habits, time taken to go to the toilet, amount of work performed, etc. This is easily done using software to check on each user - like Telstra receptionists who can be monitored at any time.

The change in work may also include -
• checking e-mails as well as normal mail at the start of the work day
• telecommuting (or working from home) allows more flexible working

• videoconferencing allows people to meet in groups while staying in their office - though time zones differences and attire can present problems
• working longer, if more flexible hours
• language, date and currency rates may present problems

Copyright - issues involve a range of legal and ethical problems for the compensation for authors who have worked hard to produce their original pieces of work. With Internet publishing in particular, who knows (except maybe the author) which information - be it text, photograph, song, etc. has been illegally copied. Many such authors who have their work on the Net only want to be acknowledged, but others find their work has been put in many places, changed or unchanged without their permission or knowledge.

Computer software can be in a number of categories - freeware, shareware, commercial, etc. but if located on the Net, it is all treated by some people as ‘fair game’. Other people may then get copies of this and the amount of legal software sold decreases to the point where all prices must go up to pay for production costs - since so little is actually paid for.

Many copyright owners are now embedding codes within the digital media - stored in graphics, photographs or text, and invisible to the casual observer. The proof of ownership can be shown and the source of the theft often tracked back to the person or company. This is one way of protecting intellectual property.

Exercise 2.13 complete the following -
1. Give a number of reasons why you think e-mails and chat rooms are so popular with people your age?

2. Send an e-mail to a friend - with an attachment if possible (picture, tune, etc.)

3. What network topology does Westfields have? Why?

4. Do you consider it fair that a person is monitored during the hours they are employed? If ‘yes’, why? if ‘no’ why not?

5. Should all computer software have some cost to the user? When is software piracy not a real problem, and when does it become one? Give reasons.

To be completed by - .................................................................
Displaying

Output to the screen or printer for display purposes is sometimes the final process for the information, that is, to meet it’s purpose. Text, image and number can be in a printed form, while audio and video cannot really be recorded on paper. The output from the computer can be sent to a variety of devices. Where this output is printed it is called hard copy, when it is stored or displayed it is called soft copy.

THE MONITOR

The monitor is also known as the video display unit (VDU) and is the main output device of the computer.

Most consist of a grid of 25 rows by 80 columns and each of these can display one alphabetical character - made up of little dots called pixels - the letter B is seen from the pixels turned on.

The most common type of monitor screen is the Cathode Ray Tube which is the same type as used by a TV. As shown in the diagram, the CRT creates an image on a phosphorus coated screen by bombarding it with a controlled beam of electrons. To determine the intensity of the beam, a negatively charged grid allows through only the required amount of electrons. This concentrated beam is directed at the individual dots on the screen (pixels). When an electron strikes a phosphor atom on the surface of the screen it releases its energy in the form of light, though since this only lasts a short time, it needs to be struck regularly to maintain the picture - this is termed the refresh rate. This refreshing generally needs to be done at least 30 times per second, and if it is not quick enough will cause flickering and the image will begin to fade.

Monitors are of 3 types - monochrome, grey-scale and colour. The choice of which of these, should be based on the intended use for the computer system. They vary in their quality of resolution and their ability to reproduce colour.

Internally CRT’s function as analog devices, and therefore the video controller needs to convert digital image and video data to analog before being sent to the CRT.

Liquid crystal displays - these use similar technology to digital watches, mobile phones and calculators, though if a higher resolution. These are normally found in palmtops, notebooks and some laptop computers, though some high quality ones are now becoming available for desktop models. The controller works much like the CRT, but some manufacturers have backlit the screens to make them easier to read.

Monochrome ones are now less common, and it is predicted that LCD screens will take over from CRT’s since - the quality is about the same, they consume less power, they are thinner and therefore lighter in weight, they emit only tiny amounts of electromagnetic radiation and they take a digital signal directly.
Plasma Displays - these use a sealed tube of inert gas, somewhat like a neon light to display the pixels - though using a lot of electricity and weighing more than other displays they are good for well lit areas and can be quite large (150 cm by 110 cm).

Data Projector - these are used to project images up onto screens or walls for large audience presentations. Newer models do not even need the room lighting to be reduced as they are provided with powerful lights. They are connected to the computer and simply display whatever is on that screen.

THE PRINTER
All printers communicate with the computer using either a serial or a parallel data link. Software is needed to provide the driver for that particular printer. The most common method of printing involves the applying of dots of ink to the paper - called bit mapped printing, where the more dots per inch (dpi) the better the quality of the image. This may be achieved with either an impact or non-impact technique.

(a) Dot Matrix - these were the most common and least expensive impact printers e.g. Imagewriter printers in both computer rooms. They produce both text and graphics, these being formed from a collection of dots stamped onto the paper. As seen in the diagram, the print head contains a group (matrix) of pins or wires - generally 9 or 24, which strike the ribbon as the print head moves in small amounts across the page.

The closer the dots produced, the higher the quality of the output, so the 9 pin type are generally referred to as near letter quality (NLQ) as the text is inferior to that from a typewriter. Printers with 24 pins are sometimes called letter quality(LQ) but are still not as good as the print from a typewriter.

All printers have their print quality expressed either as its resolution in dots per inch (dpi) they produce or as degree of letter quality - near letter quality dot matrix printers produce 72 to 144 dpi and up to 300 for letter quality.

Dot matrix printers have two possible types of interfaces for connection to the computer - serial or parallel. Serial involves transmission of data 1 bit at a time, while parallel transmits 8 bits simultaneously. Parallel transmission is faster but serial may be sent greater distances. Dot matrix printers have dual in-line package (DIP) switches which must be set correctly for the printer to print correctly.

(b) Daisy wheel - these are also impact printers, and are similar to typewriters, with each of the characters placed at the end of of the daisy petals and the daisy petal wheel rotates until the character to be printed is found. The hammer then hits that petal and impresses it against the ribbon onto the paper. These print letter quality, but are slow, expensive, cannot print graphics, and the daisy has to be changed to change fonts or styles.
(c) **Inkjet Printers** - these are capable of printing in black or in colour (adding a colour cartridge) at relatively high resolution. These special inks are sprayed as individual tiny drops onto the page from nozzles, with the drops being directed by electrical fields onto the paper so that text and graphics are built up as a pattern of tiny drops. The spraying is achieved in one of two possible ways - either by electrical heating of the ink, or mechanically.

They have between 80 and 304 nozzles, and care must be taken not to smudge the page while the ink is still wet. They are quiet and somewhat slow, but give good quality printing of 360 dpi (or even higher) so long as good quality paper is used (otherwise blurring can occur).

d) **Bubble jet Printers** - these employ a print head with a heating element around the nozzle to vaporise part of the ink (around 100 times per second), so that the resulting bubble ejects ink onto the page.

Otherwise these work similarly to the inkjets, though colour bubble jets reproduce multiple colours by dithering (partial scattering of dots so they mingle to produce new colours), but this reduces the resolution to 180 dpi.

(e) **Thermal Printers** - are mainly used for colour printing, where a heated (thermal) head melts dots of coloured wax stored on a ribbon, one line at a time. The paper is mechanically pressed onto the wax and then peeled away, leaving the unmelted wax still attached to the ribbon. Each colour is printed on a separate run through, taking 2 or 3 runs.

(f) **Laser Printers** - are quite fast, quiet, give good quality (300 - 1200 dpi) and are becoming relatively inexpensive. They use photocopier technology and the same printing method, thermally fusing a plastic toner to the paper. The laser is used to draw the page on the printing drum as it is described to it by a page description language such as Postscript. The toner is attracted to the drum and forms the image, then as the paper is passed over the drum, the image is transferred to the paper and heated to fuse the toner to the paper permanently.

Very high resolution laser printers (1000 - 2560 dpi) are termed typeset, or in printing terms camera-ready quality. Colour laser printing is becoming cheaper as they become more common. Printing rates are commonly 6 to 24 pages per minute (though up to 100 ppm can be achieved). Though the print appears continuous, it is still made up of dots - though at 600 to 1200 dpi they are hard to see.

(g) **Plotters** - these can do large complex drawings, often with inkjet print heads now being used instead of pens, though still moved continuously as the older pen models did. Two types are found -
- flatbed - where the paper is attached to a flat table and a movable arm moves the pen(s) drawing the images as it goes as would a person.
- drum - where the drawing arm moves back and forth in just one direction over the paper on the drum or roller, with the paper moving forward and back. This takes up less space than a flatbed plotter and therefore generally takes larger sheets of paper - for large posters, etc.
(h) **Film output** - many printer types have the ability to print to overhead transparencies, either in monochrome or colour. Special output devices are also available to expose photographic film for use by commercial printing presses or to make slides - though using a data projector would be better.

**Audio** - output of sound is quite simple, with speakers or headphones connected to the sound card or audio out of the computer. The quality of the sound produced will depend on a number of factors - speaker or headphone quality, sound card quality, and resolution of the sound being produced.

The quality of the speakers or headphones is a straightforward match up, but the quality of the sound file itself is the critical factor - for example - a 16 bit sample at 44KHz will be much better than an 8 bit 11KHz digital wave file (though the first will sound terrible if played via an 8 bit sound card with cheap speakers). MP3 files are now a source for users wanting to add music to presentations, for example.

**Software for Display**

With the increase in the number of people using computers over the 1990’s, the amount of programs produced for the amateur to produce cards, messages, advertising sheets, assignments, and more recently Web pages, etc. (i.e. to display information) has meant that professionals do less of the designing and producing of such work. However, having a program on a computer that can do desktop publishing and a printer that can make colourful pages, does not make the user skilled at the art of desktop publishing.

In fact, some people put so many ‘bells and whistles’ into many of the documents that they produce, that the message or information contained in it are almost lost by being overwhelmed by the borders, point sizes, fonts, colours, etc.

**Display Features** - two of the features used in desktop publishing need to be considered if a document is to be understandable.

1. **Typography**

This is the art of designing typefaces which are easy to read and aesthetically pleasing, with the goals of - legibility, invisibility and beauty.

- A **Typeface** is the design for a family of characters and symbols in a particular style.
  - e.g. Palatino, **Courier**, **Chicago**, New York, Avant Garde, Mac Humaine

- A **Font Family** is a set of fonts of different sizes, weights and postures that all have the same typeface.
  - e.g. Helvetica 12, Helvetica 10 **Black**, Helvetica 10 compressed

- A **Font** is a complete collection of printable characters, including letters, numbers, punctuation marks and other special characters designed with a consistent appearance.
  - e.g. Palatino 11, **Palatino 12 italic**, **Palatino 14 bold**
Serif and Sans Serif Typefaces
Typefaces are basically grouped into two categories -

**Serif** - which have fine cross-strokes (termed serifs) across the ends of the main strokes of each character.

This type is used for **body type** because they are more legible (particularly under 14 point), and they are also used as **display type** for headlines (as in newspapers).

Serif fonts include - **Times**  **Courier**  **Bodoni**

**Sans Serif** - which do not have the fine cross-strokes (sans means without) and are suited to font sizes over 14 points, especially when their are few words e.g headings

Sans serif fonts include - **Helvetica**  **Avant Garde Geneva**

Fonts
The effective use of fonts is a science, since the huge variety available makes the choice important for getting your message across. The right choice enhances the message, the wrong one can seriously detract from it.

Main points -
• limiting the number of fonts - generally 3 is a maximum in one publication (heading, body and highlighting)
• serif are best for body text, sans serif for headings
• try to match the typeface with the message (formal for business, imaginative for advertisements)
• actively use white space

Letter Size
Once the typeface has been chosen for a document, the size has to be worked out. This will depend on the number of columns required - single column, generally has between 9 and 12 words (55 - 65 characters). This will depend on individual taste and intended audience, but most formal documents stick to the 9 to 12 words per line.

6 point
9 point
10 point
12 point
14 point
18 point
36 point
48 point
Type Style
Is used to highlight words within the text, such as \textbf{bold} or \textit{italics} for example.
- **Bold** - is used to add weight to a word, but due to the heaviness it cuts down on legibility, and should not be used for more than headings or a few words in the body text.
- *Italic* - is a good way to add emphasis to a single word
- Underline - is viewed as less professional these days (probably since in typewriter days it was the only highlighting method)
- **Outline** and **Shadow** should only be used occasionally
- All upper case (or caps) must only be used for short headings as legibility is poor - TEXT ALL IN CAPITALS IS REALLY HARD FOR ME TO READ.

SPACING

- **Spacing between words and letters**
The spacing between letters in each word or between words themselves (the \textbf{space band}) may be adjusted by DTP software. This may be done to -
  - improve the readability of the font in use
  - slightly increase or decrease the length of a line so it fits neatly into a space
  - produce a special effect e.g making a heading wider or taller to fit a required space in a document

Methods of adjusting spacing -
1. manual letter and word spacing allows the user to specify the percentage increase or decrease on default settings - both for letters and words independently.
2. \textit{tracking} adjusts both letter and word spacing together - with tracking being a character attribute - so it can be 'tightened' or 'loosened'.
3. \textit{kerning} is used to adjust the spacing between letter pairs that do not fit well together.

- **Spacing between lines**
The units of measurement most commonly used in typography are -
  - **points** - with about 72 points to the inch (2.54 cm to the inch), and
  - **picas** - with 12 points to the pica

Thus 1 point is 0.353 mm and 1 pica is 4.233 mm.

Note the point size (or font height) for a font is measured from the bottom of the \textbf{descent line} to the top of the \textbf{ascent line}. The point size of a font gives only an approximate size for the letters, as different typefaces may have longer or shorter ascenders and descenders relative to their \textbf{x-height}.

\begin{center}
\begin{tikzpicture}
\node[above] at (0,0) {font height};
\node[below] at (0,0) {base line};
\node[above] at (0,0) {descent line};
\node[above] at (0,0) {ascent line};
\node[above] at (0,0) {point size};
\node[above] at (0,0) {x-height};
\node[above] at (0,0) {p M};
\end{tikzpicture}
\end{center}

- **Spacing between paragraphs**
This improves the readability of the document and gives a lighter feel to the text - more white space.
Justification

Justification is of course the alignment of lines of text along the left, right or both margins.

- **Left justified** - the ragged right edge is more readable text, as the reader can keep track of their position on the page - so most of your text should be left justified.
- **Right justified** - used mainly for special effects, columns of numbers, etc.
- **Fully justified** - should not be used with narrow columns, as it creates rivers in the text.
- **Centred text** - used for centring headings and other special effects.
- **Forced justified** - this is where even a short last line is forced to extend to the column width, spreading the words over the space.

2. Page Layout and Design

When you design the layout of the page you are producing the skeleton and when you decide on the typography, you are adding the flesh and blood of that page. Some aspects of page layout apply to all of the pages of the document (page size, margins, columns, etc.) and these are termed **global decisions**. Most other decisions will effect just the page concerned (e.g. placement of a graphic).

**Step 1** - Number of columns and space between them - for this you need to know the size of the paper for printing, the orientation of the paper (portrait or landscape), and whether it is to be folded or not.

**Step 2** - Set gridlines to guide the placing of text and graphics on the page (most DTP programs have this facility, otherwise a ruler must be used). These gridlines are both vertical and horizontal (similar to the grid of the drawing module of ClarisWorks).

Good page design has a range of elements -

- consistency
- simplicity (or restraint)
- variety
- white space
- headlines
- graphics
- proportion and balance

**Exercise 2.14** complete the following -

1. Open the files named Ex2.14 (both pages 1 and 2) on the floppy disk, and follow the directions found at the bottom. Save as Ex2.14 fixed.

2. Make summarised notes and print them out - on each of seven elements listed above (from pages 192 to 196 from the red Chivers book, and the white Boyd book). This will be marked for both the -
   (i) use good page layout and typography skills in each of these, including any necessary diagrams or pictures, and
   (ii) quality and accuracy of the notes

To be printed out on the laser printer and handed in by - ..........................................................
Tables
Sometimes the simplest and most effective way of arranging text and/or numbers for display is to put them in columns and rows - a table. Comparisons are generally much easier this way, rather than searching through paragraphs of text looking for details - so including a table in a passage of text is economical as well as clear.

Which information goes in the columns and which goes into the rows needs to be planned, and point size is generally a little smaller than for the body text (maybe 10 point). Tables may have lines dividing each part into boxes or not - at least some lines make the information being displayed easier to follow.

Example -

<table>
<thead>
<tr>
<th>Region</th>
<th>Ninety1</th>
<th>Ninety2</th>
<th>Ninety3</th>
<th>Ninety4</th>
<th>Ninety5</th>
<th>Ninety6</th>
<th>Ninety7</th>
<th>Ninety8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Northern</td>
<td>1200</td>
<td>1185</td>
<td>1221</td>
<td>1157</td>
<td>1228</td>
<td>1198.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>North West</td>
<td>1064</td>
<td>1108</td>
<td>1167</td>
<td>1229</td>
<td>1284</td>
<td>1170.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Western</td>
<td>1009</td>
<td>998</td>
<td>1102</td>
<td>1056</td>
<td>1078</td>
<td>1048.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South West</td>
<td>1285</td>
<td>1358</td>
<td>1422</td>
<td>1589</td>
<td>1668</td>
<td>1464.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Southern</td>
<td>1210</td>
<td>1224</td>
<td>1199</td>
<td>1289</td>
<td>1142</td>
<td>1212.8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>South East</td>
<td>1286</td>
<td>1229</td>
<td>1207</td>
<td>1287</td>
<td>1254</td>
<td>1252.6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Charts
Even the best design and produced tables of information containing numbers are more difficult to follow than a simple chart. Trends can be noted immediately, at least if the right type of chart has been selected and produced accurately.

The easiest method for making a chart is to enter the data into a spreadsheet and have it make the chart (using one or several of the optional types available. Colour and 3D are often options, as well as some special features (like arrows, and labels, etc.) though again, the objective of conveying information can be made difficult if the chart is tarted up too much. Simple is generally better.

Note - Pie charts are often popular with students, but show only very general trends, not clear accurate details if the segments are close in size to each other.

The common types of charts used include - • bar (horizontal column)
• stacked bar
• column
• line
• pie
• combination

Charts can be made to emphasise certain aspects of the data, rather than others - and some would consider this making the chart tell a lie. For example, the chart on page 309 of the Boyd textbook with trees giving a false impression. You can alter the look of a graph or chart by changing say the vertical scale - either to make a line look like it is going up steeply (for say yearly profits being shown the shareholders meeting) or hardly dipping at all (for say losses on a certain project).

See the difference between charts 1 and 2 for scale difference, and with number 3 for clarity, on the next page.
• Which chart shows the trends most clearly?
• Why is 1992 entered as Ninety2?
Exercise 2.15 complete the following -
1. Using ClarisWorks spreadsheet, enter the Sydney rainfall data (from the previous page) and save the file as Rainfall SS.

2. You are to produce a 1-2 page report to be submitted to your superior (you work as a government analyst and planner) about the capability of the Sydney drainage system to handle rainfall and prevent flooding. Money is limited and you need to work out which regions need the most work done. Your report must contain a table and a chart and recommendations based on these. Desktop publishing skills should be displayed.

To be handed in by - .......................................................
Reporting
When we use databases, we are generally searching for certain data and when we want to print out report we do not want the whole database in hard copy. We therefore go to the reporting section and choose the format to display the selected data in.

If the database program does not allow enough reporting options then the user may need to transfer the output to another application program (spreadsheet, desktop publisher, etc.) and use it’s features to present the information.

Merging
Mail merge is a much used facility in databases these days, allowing the user to combine a form type word processed letter with a list of data from the database. All the personalised letters people receive from Time Life or Art Union groups are produced in this way -

Form letter
Mr X. Baldock
34 Road St
Wakeley 2176
Dear Mr Baldock,
Your son Greg has just become the top computing salesperson and has won a holiday to Hawaii.
Yours sincerely,
J. Smith

Final letter from mail merge
Mr X. Baldock
34 Road St
Wakeley 2176
Dear Mr Baldock,
Your son Greg has just become the top computing salesperson and has won a holiday to Hawaii.
Yours sincerely,
J. Smith

Section of database

Merge (and not mail merge) is also used for a range of text, such as - envelopes, invoices, labels, credit cards, certificates, and e-mails to simply move data from one application into another for better display possibilities.
Exercise 2.16  complete the following -
1. Prepare a database for 4 of the class members, with fields - Title, First Initial, Last Name, Address, Suburb, Postcode. - as in the section of the database on page 48. Save this.

2. Prepare a form letter - stating that the son/daughter has won a new computer in a competition. Save this.

3. Mail merge these - see ClarisWorks Help on how to do this. Print these 4 pages out.

To be handed in by - ......................................................

Non-Computer Displays
Visual displays either for a single viewer or for a room full of people have been done for several generations. Traditional methods include the use of -
• pen and paper
• blackboard or whiteboard
• 35 mm slides
• overhead projector transparencies
• verbal presentation

Since paper presentations have been around for a long time, all other methods have adopted some of the features of this method - like title pages, page numbering, table of contents, etc. Visually displayed information is still evolving its set of accepted expected parts, though this is occurring - but note, computer screens have a lower resolution than a printed display so less can be shown per page (and generally at a much larger point size - around 18).

Storyboarding
This is a good planning method for all forms of presentation. Presenters normally begin with an outline containing the main areas to be covered, generally just in text format.

A man takes a walk in a park one day. The sun is shining brightly and the sky is a clear blue. A dog runs by him.

A storyboard is a pictorial representation of the screens used to make up a project. It is the visual interpretation of the outline at various stages and is developed to show the important parts of the presentation.

Storyboards are made from rectangles or grids with each representing a screen. The contents of a display can be represented by pictures or writing and sometimes both. The rectangles can be extremely detailed or very brief, often depending on whether or not the information regarding the display needs to be communicated to others.

See the example of the storyboard on the next page.
Exercise 2.17 complete the following -

1. Prepare an outline for a 10 screen presentation entitled “Use of the Internet”, this is to be typed and printed out.

2. Prepare a storyboard for the presentation - either hand drawn or by computer.

3. Using the program HyperStudio, produce this presentation to be demonstrated by the set date (get the skeleton done first, the fancy bits last).
   
   To be completed by - .................................................................

Social and Ethical Issues For Displays
The communication skills of the presenter used to be the main persuading aspect of the presentation with verbal skills, personality and charisma contributing to win over the audience. These days, display technology is becoming increasingly important, with audiences expecting a high degree of sophistication and technical skills. Someone who just gets up and speaks in front of an audience may no longer be viewed too highly.

Presenters should be aware that there may be poorly sighted people or even blind ones in an audience and therefore large point size and even Braille handouts should be provided. The following points should all be considered in such a presentation -

- audience literacy level - complicated and very technical language will be wasted on a non-technical audience
- audience age group - suitability for the expected age group, young, old, etc.
- visual capabilities - point size, colours used, audio tape or Braille alternatives
- aural capabilities - amplifiers for those with hearing difficulties