Topic 2: Data and Information

Data

Anything that a computer system can store and process is considered to be data. In practice this means that data could be

a fact	such as a person's surname or the price of a good.
an event	such as a change of address or the arrival of new stock.
a transaction	such as a sale, order or reservation.

Data is produced as a result of an event or transaction. In most situations data occurs as a series of values relating to the event or transaction. Data can be defined as the raw values that are input to a computer system, which have been generated by an event or transaction or which represent some fact. By 'raw values' we mean values that have been stripped of their context and which therefore, on their own have no meaning.

If an item of data had a value of 040389, for example, we could not attach any meaning to it unless we knew the context that it has arisen from. It could be a date of birth, a map reference, a catalogue number, a telephone area code or any one of a number of other things. The computer system will be able to process this data correctly because it will be supplied in a structured way. In other words, if the data value is a date of birth then it will be input as part of a sequence at the point where the computer is expecting a date of birth to be input.

When a customer takes goods to the check out in a supermarket a series of values is generated. The first value might be the customer's loyalty card number. This might be followed by a series of bar code values representing the items bought. The final data in this sequence might be a code signal indicating that there are no more goods, followed by an amount tendered. This whole sequence of data has been generated by an event – the customer buying some goods.

Data values are not necessarily numerical; they could be numbers, letters, dates, true/false values or anything else, such as moving or still images or sound. For any particular type of data there will be a set of possible values. Surname data, for example, will take values like Smith, Brown etc.

Coding data

Coding is when a short combination of characters and or numbers is used to represent the actual data. For example eye colour might be coded as:

Bu – Blue, Br – Brown, Gn – Green, Gr – Grey.

Data is often coded when it is collected or when it is being prepared for input to the computer. This is done to reduce the amount of storage space needed and also to speed up both the entering of data and data searches.

It will take less time to type in the data as long as the person preparing or entering the data can remember the codes. This is because there are fewer characters to enter. The data will take up less storage space for the same reason.

When the data is searched it will take less time for the contents of each field to be processed, again because there are fewer characters to compare. Matching, when looking for particular data, will be exact because both the person who entered the original data and the person who is entering the search criteria will have used the same code.

Coding is often useful, particularly when the set of possible data values is small, but does have some disadvantages. Data may not fit the list of available codes exactly and so it will lose precision. For example, when coding eye colour, blue–grey eyes could not be coded accurately if only codes for blue and grey existed.

Another problem is that anyone wanting to use the data will need to know the codes that have been used. Using memorable and obvious codes will help overcome this.

Encoding Data

Digital computers can process only numerical data. This means that other types of data such as text, sound or pictures must be converted to numbers before they can be input to a computer system.

The process of converting data values to a suitable format for use in a computer system is called encoding.

Text characters are usually encoded using the internationally agreed ASCII codes where A is represented by the number 65, B by the number 66 and so on. Standard encoding methods exist for various other data types such as video, sound and pictures. In each case a file extension will usually indicate the type of encoding that has been used so that jpg, bmp and tif represent three different methods (Joint Photographic Expert Group, Bitmap and Tagged Image File Format) of encoding picture data.
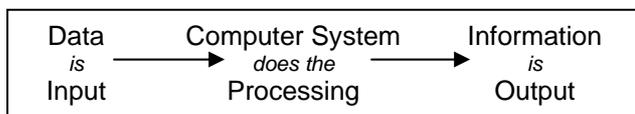
Processing Data

Processing is the actions that the system carries out on the data once it has been entered. It can involve sorting, searching, calculations and comparisons.

Sorting	Puts data in some order – usually alphabetical or numerical. An example would be printing out a list of company creditors sorted in order of amount owed.
Searching	Finds data that matches one or more criteria. An example would be searching a library database to find all the books by a particular author.
Calculation	Various different types of calculations can be performed. For example calculating the total bill for a customer at a supermarket checkout.
Comparison	Comparisons such as *less than*, *equals* and *more than* can be performed to determine what processing should be performed. For example to find all customers who have spent more than a certain amount in the past year so that they can be sent a special offer.

When a bar code data value is entered at the supermarket checkout, the computer system searches its database to find a matching entry in the stock file. One item of processing that then takes place is that the 'number in stock' value for the stored record representing the item is decreased by one.

Information

The purpose of processing data is to provide information. So data is input into a system and the resulting output is information. Information can be used by people to help decision making.

Data → Computer System → Information
is — *does the* — *is*
Input — Processing — Output

Information = data + context

A company selling various products may feed data about each sale into the computer. This data could be processed to provide information about which items were the best sellers or how much money was taken in at different times of the day. Managers can use information provided by the computer to help them plan and make decisions.

One definition of information is to say that it is data that has been processed and given a context that makes it understandable to a human.

Quality of Information

It is not sufficient to provide information. If the information is to be acted on then it must be seen as 'good' information. A managing director arriving for work is unlikely to believe a passing tramp who calls out that the company is going bust. The same information passed informally by the company president would be more believable. For information to be considered good it must be relevant, accurate and complete. It must be provided to the right