Data and Information
INFO 2

Oliver Boorman-Humphrey

www.oliverboorman.biz

14 April 2013

What is Data?

What is data?

Data

Raw facts and figures collected together, before they have been processed.

Data can come in a number of forms:

- ➔ Text
- ➔ Numbers or Statistics
- ➔ Images
- ➔ Moving images
- ➔ Sound

Data Sources

Data Source

A location from which data can be found.

They come in different forms:

- ➔ Direct (or Primary)
- ➔ Indirect (or Secondary)
- ➔ Static
- ➔ Dynamic

Data Sources

Direct Data

This is data is collected from its original source

Indirect Data

Data that is used for a different purpose to that that it was meant for. The people involved in collecting the data are different to those who use it.

Static Data Source

A source of data that's data remains the same over time.

Dynamic Data Source

A source of data that's data changes or is updated over time.

Data Sources

Match the data source to it's type.

Dynamic Data Source	Telephone directory
	Internet
Static Data Source	Online Survey
	TV program
Direct Data Source	CD-ROM
Indirect Data Source	Newspaper

Data Sources

Advantages of direct data:

Disadvantages of direct data:

Data encoding

Sometimes we use data encoding to represent data in an information system.

Encoding

Putting data into a code or shorthand notation - by taking the original data and converting it in a different representation.

Examples of Encoding include:

- ➔ Jan, Feb. Mar ... for months of the year
- ➔ M or F ... for male or female
- ➔ Y or N ... for yes or no

We need more than just data

Example

What does 190813 mean?

We can not tell what data is supposed to mean with some prior knowledge or context.

Knowledge

Application of Information in a Situation

If we know that our example is a date in the form `ddmmyy`, then the meaning is obvious. Our data represents 19th August 2013.

Information

Information

The result of taking data and processing it. This involves giving the data meaning.

Information = Data + [Context] + [Structure] + Meaning

Example:

- ➔ Data : 67
- ➔ Context: *ICT Exams*
- ➔ Structure: *Percentage*
- ➔ Meaning: *Average*

The information we get from this is that the average score in the ICT exam was 67%,

Quality of Information

Computers are not intelligent, they are dumb! They only respond to commands and instructions.

They don't have any way of knowing whether data you enter is *correct* or *accurate*.

GIGO

Garbage In, Garbage Out

GIGO means that if the user inputs the wrong or inaccurate data, the computer will output the wrong data.

Quality of Information

The *Quality* of Information is affected by six factors:

- ➔ Accuracy
- ➔ Relevance
- ➔ Age
- ➔ Completeness
- ➔ Presentation
- ➔ Level of Detail

Quality of Information

Accuracy:

If data is not accurate, it cannot be trusted.

If you have a database which stores people's birthdays and you have one birthday entered wrongly, that data is useless!

Relevance:

If you have some information but it does not relate to the topic, it is worthless!

If you have a list of telephone numbers but you want a list of fax numbers, you have no information.

Quality of Information

Age:

Information can change over time. If you know information is from the past, it may not be relevant now.

If you have some statistics about the number of accounts in a bank, but it was collected 3 years ago , it's almost certainly incorrect now!

Completeness:

If you only have part of the information then it is worthless!

If you have a list of books in a library but half the books aren't on it, we have no useful information.

Quality of Information

Presentation:

If the information is not presented in a way that you can understand it loses value.

If you have just a page of statistics but they are have no labels, then it has no meaning!

Level of Detail:

The volume of data determines whether you have enough to make a decision or too much

If you have a recipe but only the ingredients are listed with no quantities, then you don't know how to bake the cake!

Data Checking

Although it can be tricky checking quality, we should be able to check if data is *valid*, though, using data validation and verification techniques.

Data Validation

A check of entered data that is carried out by the computer to stop data that does not conform to pre-set rules being entered.

Data Verification

The process of ensuring that data entered into a computer matches the original paper version.

Data Checking

Match the problem to the checking type which wish identify it.

Data Validation

Data Verification

Entering 24/5/2090 instead of 24/5/2009

Entering false when you meant to click true

Entering 13 for a month number

Entering TNE4 0BH instead of TN34 0BH

Entering -1 for age