FACTFILE:
GCE DIGITAL TECHNOLOGY

AS2 DATA AND INFORMATION

Data and Information

Learning Outcomes

Students should be able to:
- distinguish between data, information and knowledge by using examples;
- describe how the quality of the information produced can be affected by the following factors: accuracy, relevance, up to date, completeness, presentation and reliability ;
- describe and apply the following data validation methods: range, type, length, format, presence, Modulus 11 check digit and lookup, check digits and batch totals (hash and control);
- understand the purpose of data verification methods: double entry, proofreading: and
- explain the limitations of data validation and data verification;

Content in Data and Information Fact File

- ✔ Data and Information
- ✔ What is knowledge?
- ✔ Factors affecting the quality of information
- ✔ Data Validation
- ✔ Data Verification
- ✔ Limitations of Data Validation and Verification

Data and Information

Difference between data and information
Data consists of raw or unprocessed facts and figures which are not set in a particular context. Therefore it does not have any meaning until it is processed. When data is processed the result is information which means that it has a context.

Information could be summarised as "data with meaning". To help understand the difference between data and information we will consider a large clothing shop which consists of a number of departments. The target sales and the actual sales for each department are shown below.

Department	Target Sales (£)	Actual Sales (£)
Men	15000	12341
Women	20000	21304
Children	10000	10312

In the above table 15000 on its own which means it does not have a context is just a sequence of digits. When we give 15000 a context which in this case would be £15000 is the Target Sales for the Men's Department it becomes information.

What is Knowledge?
Knowledge is the result of applying rules to information to allow decisions to be made or to allow you to interpret the information. In the example above knowledge could be making a decision to plan for further advertising of Men's clothing due to the sales target not being met as it is below the actual sales whereas the other departments are above Sales Target.

Questions
A health centre stores each patient's medical history in a database. A report of each doctor's appointments is produced at the start of each day.

Date 12th June			Doctor Brown
Time	Patient ID	Name	Notes
09:30	321	Green	Cancelled
09:45	362	White	
10:00	289	Grey	

(a) By referring to the shaded cell, explain what is meant by Data and Information

....

....

....

(b) By referring to the report, explain what is meant by knowledge.

....

....

....

Factors affecting the Quality of Information
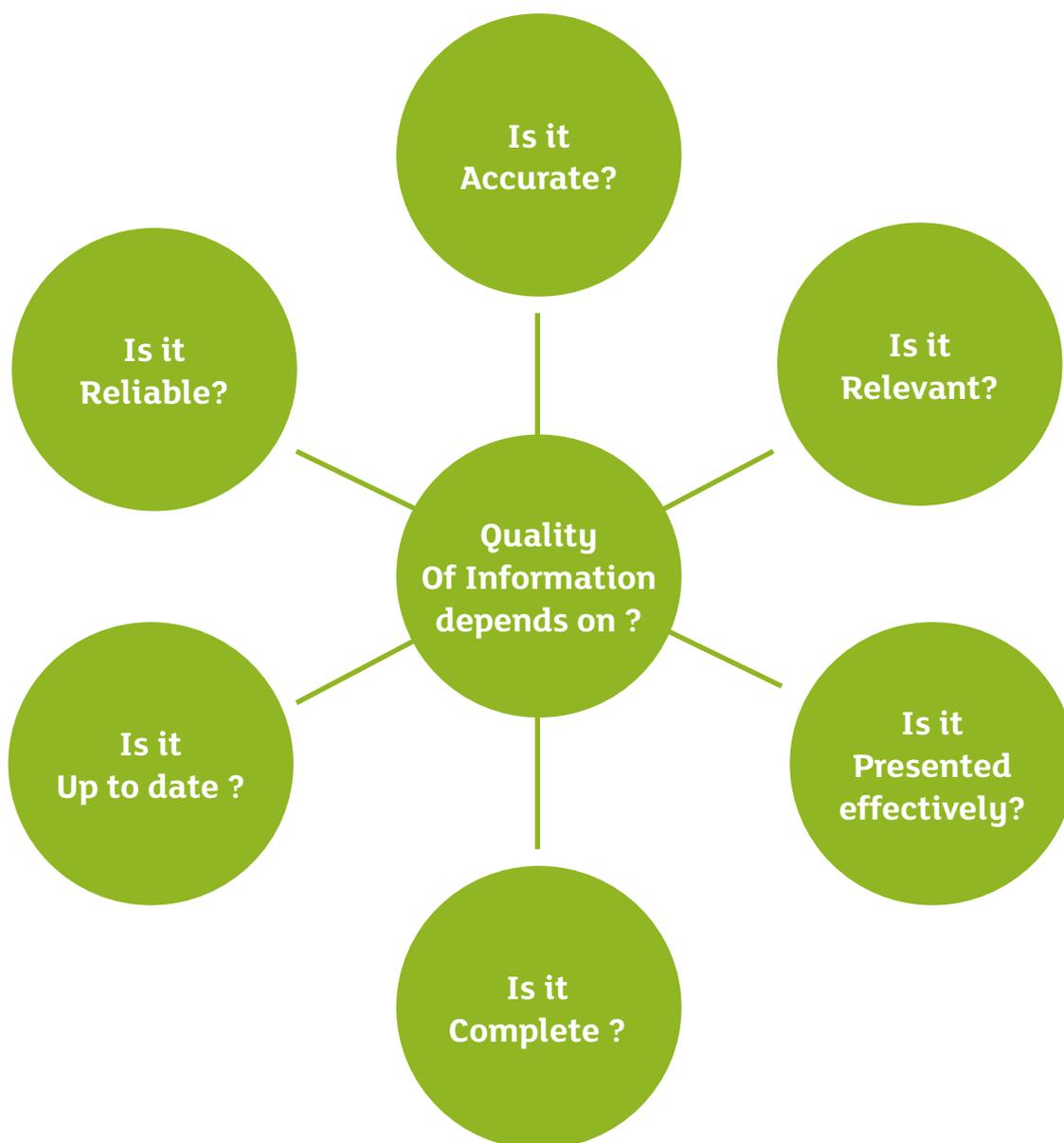
Accuracy

The accuracy depends on the use made of the information by a person or organisation. For example the information on a bank statement must be exact to the nearest penny otherwise a customer may complain whereas the weather forecast could be given to the nearest one degree. To ensure information stays accurate thorough error checking must take place and regular updates must be made.

Up to date

If information is out of date then it can reduce the quality of information. For example if a school does not update pupil addresses for 5 years then some pupils may not get their school report if they have recently moved house. Real time processing needs data inputted to be processed immediately otherwise it is not up to date. For example when you purchase a flight ticket the bookings file must be updated before the next transaction takes place.

Complete

This would refer to situations whereby information is missing. For example a school may ask students leaving school to complete a questionnaire with details of their third level course and their student address so as information about the past pupils reunion can be sent next year. If the questionnaires are incomplete such as a missing e-mail address then the student maybe not receive an invitation.

Relevance

Having information that is not relevant can be a disadvantage as it adds to the volume which can increase the time taken to find relevant information. Information that is essential in one situation may have no use in another. For example Key Stage three grades may not be relevant for a University application form.

Presented effectively

If Information is not presented in such a way that is difficult to understand then it loses its value.

For example information about sales trends in a business would be more effective if a graph is used rather than a text document.

Reliable

The information has been verified and deemed trustworthy. If information comes from a direct source it is more likely to be reliable than information from an indirect source. For example if information is used for a specific purpose it could be more reliable than if it is used for a different purpose or than originally intended.

A number of these factors may be considered collectively to ensure the quality of information is good, such as a Car Navigation System

© SergeiKorolko / iStock / ThinkstockPhotos

Satellite Navigation Systems provide motorists with ***Accurate, up to date and relevant information*** about roads based on data received from satellites.

Question
The quality of information can be influenced by how up-to-date it is. By referring to a typical invoice, describe **two** other factors which can influence the quality of information.

....

....

....

Data Validation

Data validation is a check performed by computer software on data as it enters a computer system at the input stage for processing. Its purpose is to trap any data that does not conform to certain rules. It cannot prove that the data entered is the actual value the user intended. However it does allow the computer to use a number of techniques to ensure that the data entered will be:

- ✔ Sensible
- ✔ Reasonable
- ✔ Within acceptable boundaries
- ✔ complete

Data is not processed until the validation check has been successful.

Range check
This will check an input value against an upper limit boundary and a lower limit boundary. If the value falls outside the limits then it is invalid. An example could be month of the year whereby the data entered must be an integer between 1 and 12.

Type check
This ensures that the data item is of a particular data type. For example the item price in a stock file is of currency type.

Length Check
This is used to check that the data entered contains a certain number of characters.
For example, a mobile telephone number contains 11 digits.

Format Check
This is used to ensure a data item matches a previously determined pattern and that particular characters have particular values such as letters or digits. For example a date of birth field maybe required to be entered in a predefined format using the pattern DDMMYY.

Presence check
When entering data into a database some fields maybe optional. This check will not allow certain fields to remain blank. For example a mobile telephone number may be optional as the person may not own one whereas National Insurance number is required (must be present) on each record in the employee file.

Check digit
This involves using an extra digit which is added to the numeric data item. These are used on bar codes placed on products in a typical supermarket or ISBN numbers placed on books. It is calculated using the digits from the number and then added to the end of the number. When the number is inputted the same calculation takes place and the result is compared to the check digit. If the results are the same processing continues. If they are not the same, an error has occurred and the numerical value needs to be re-entered. The main way of calculating a check digit is by using Modulus-11 arithmetic.

How do you calculate a check digit?

Consider the following product number 91167 and calculate the check digit.

1.Apply weights – beginning at the right hand digit apply weights. The Check digit has a weighting of 1

Number	9	1	1	6	7	**Check Digit**
Weight	6	5	4	3	2	1

2.Calculate the products

Number	9	1	1	6	7	**Check Digit**
Weight	6	5	4	3	2	1
Product	54	5	4	18	14	

3.Sum of products
   54+5+4+18+14 = 95

4.Divide the total by 11 and write down the remainder
   95 / 11 = 8 remainder 7

5.Use the rule **Check digit = 11-R**
   11-7 = 4 therefore the check digit is 4

Therefore the product number including the check digit is 911674

It is important to note that there are special cases to the rule Check digit = 11-R. If R=1 then check digit = x and If R=0 then check digit = 0. A check digit system is very good at detecting transposition errors because if you interchange 2 digits then they will have different weights and the sum of the products will be incorrect leading to a data entry error.

Look up check
This method uses a lookup table. The data value entered is compared against a stored list of data values looking for a match. If a match is found the data value is valid otherwise it is invalid. For example when a product code is looked up in a stock file to confirm such an item exists.

Batch Totals
This refers to the total value of one or more fields in a batch of data. They are calculated in advance, normally by humans, and then compared with the total as calculated by the computer. We referred to a batch total as either a control total or a hash total. If the batch total calculated is meaningful such as the total monetary value of all orders during a transaction period then it is referred to as a control total. If the batch total calculated has no clear meaning such adding all the dates of orders together to produce a total of dates this is referred to a hash total

Question

When purchasing a product from a website, a customer completes the following on-screen form.

Name _____

email address _____

delivery address _____

Product Code eg. 12345 _____

Quantity _____

State and justify the most suitable field for each of these validation checks.
Character/type, Range, Format, Check Digit

Data Verification

Data verification is used to confirm the integrity of data entered into the system. It ensures the data is consistent and that it has not been corrupted. The following methods are used:

Double entry
This involves entering the data twice into the computer system and the computer will check both copies to ensure there are on differences. Any differences will be manually corrected.

Proof reading
This is also known as visual checking. This involves the user checking data entered into the computer system perfectly matches the source of the data. If the user sees that the data is correct then they confirm this to the computer.

Limitations of Data Validation and Verification

Data validation and verification do not guarantee that all data entered will be correct and error free. There are limitations in using either of these methods. When using a range check for the day of a typical month, it is possible that the day is valid but incorrect such as the day was entered as 13 and should be 31. This is referred to as a transposition error. When using a presence check if data exists in the given field it will be valid but the actual data could be incorrect such as a forename rather than a surname. In using a lookup table a data value may be valid but is not stored in the lookup table and therefore it will be classified as invalid by the computer. This will require the lookup values stored in the table being updated regularly to ensure the data validation check is not limited.

When using data verification a limitation could be the data source. If the data source is unreliable and the data entered is checked against this source it will cause a problem. The user may proofread and seeing a match with what is on the source document with what is displayed on the screen and then they will confirm this to the computer as being correct when the actual source is incorrect.