Data verification and validation

Inaccurate data can result in embarrassment and economic loss to individuals and organisations. If you put incorrect data into a computer system then the output will be incorrect. It is therefore very important to ensure that data entering a system has integrity, meaning it is accurate and that it is input to the computer without mistakes.

The causes of inaccuracies in input data are:
- **Mistakes or inaccuracies in collecting the data.** In the case of a greenhouse the sensors used to collect data such as temperature and humidity may not be working properly and as such the wrong data will be fed to the computer controlling the environment in the greenhouse.
- **Software and hardware errors.** Software errors may occur due to a virus erasing or corrupting previously entered data. Hardware errors include:
  - Transmission errors – data sent from one device to another within the computer or from one computer to another is changed due to a hardware failure.
  - Read errors – this occurs when an input device is unable to read the input medium correctly.
- **Errors in preparing the data.** This may be due to:
  - Typing errors, such as transposition errors – typing digits or letters in the wrong order. For example if you had to enter the digits 50673 but entered 50763 you have made a transposition error.
  - Misreading characters in a coding sheet, such as replacing a zero (0) with the letter O or vice versa.

People who develop computer systems use two main techniques to ensure that the data entered into a system has integrity. These techniques are:
- verification
- validation.

Verification

Verification is a check to ensure that data from the medium it was originally stored on is accurately transcribed or copied into a computer. This may be achieved by visual checks and dual inputs.

Visual checks

This involves carefully checking what has been typed against the original document. Another method is the use of screen prompts. In this case the user enters data into the computer. After a predetermined amount is entered it is redisplayed on the monitor. The user is prompted to read the data and confirm that it has been entered correctly. If the user has entered any data incorrectly he/she simply retypes the incorrect parts. The accuracy of this method depends on the thoroughness of the individual checking the document.

Dual input

This method is used when data is entered at the keyboard. The data to be entered is typed in twice by two different operators and saved. The two copies of the data are then compared. Any differences are detected by the computer. The operators will be prompted to retype the sections that differ until both copies agree. When the two copies agree it is assumed by the
computer that the data has been entered correctly. This method is much more accurate but is more time-consuming and expensive as two individuals have to be employed to do the same task.

**Validation**

Validation is the checking of data for errors before the actual processing. It is used to determine if the data entered is incomplete, inaccurate or unreasonable. The following are some of the validation checks that may be carried out by some or all validation programs.

**Presence check**

Presence checks are used to check that data has been entered into a field and that it has not been left blank. For example a database storing the names of students would require the surname field for each student to contain data. However, this check would not ensure that the correct surname was entered for the student.

**Data-type check**

Data-type checks are used to check that an entered value is of a particular type. When a table in a database is created each field is given a data type. Data may be of type text, memo, number, date/time etc. When data is entered into a specific field the database will check to ensure the data is of the correct type. If not, an error message will be displayed and the data will have to be re-entered.

**Length check**

Length checks are used to check that the number of characters entered in a field does not exceed the amount allocated for the field or is not less than specified. Each field in an Access database has a preset length. This length can be adjusted to reduce storage space and to try to ensure accuracy of the data entered. For example a field that is used to store the sex of someone can be set to hold the letters M for male and F for female. To ensure that a person would enter only one letter the length of the field can be set to 1. Another example may be a password which has to be a specified length.

**Range check**

Range checks are used to check whether data is within a range of possible values. For example when entering the months of a date of birth the range of acceptable numbers lies in the range 1 to 12. Also the range of acceptable numbers for number of normal hours worked for a week may be in the range 0 to 40.

**Format check**

Format checks are used to check that the data is entered in the format specified by the software. For example a car registration number should consist of one to three letters followed by one to four numbers.

**Check digit**

A check digit is an extra digit attached to the end of a string of digits to ensure that if any of the digits are changed by mistake, the error will be detected. It is calculated from the digits of the number. Barcode and ISBN numbers in books both contain check digits. When a barcode number is printed on a product or label of a product the computer calculates and adds a check digit to the end of the number. When the number is input into the computer a calculation is performed to check whether the check digit at the end of the number is valid. If it is not then the number has been misread and must be re-inputted.
There are many methods of calculating a check digit, one simple method is as follows:

Given an ISBN number such as 1-4058-2043

1. Starting from the left and moving to the right multiply the first number by 10, the second by 9, the third by eight and so on. We will therefore get
   \[1 \times 10 + 4 \times 9 + 0 \times 8 + 5 \times 7 + 8 \times 6 + 2 \times 5 + 0 \times 4 + 4 \times 3 + 3 \times 2 = 155\]
2. Divide 155 by 11 = 14 remainder 3.
3. The remainder is then subtracted from 11 to give the check digit:
   \[11 - 3 = 8\]. The check digit is therefore 8. Whenever the remainder is 10 it is represented by an X.
4. The number printed on the book will therefore be 1-4058-2043-8.

Parity check

A parity check is used in data communications to ensure that data is not corrupted during transmission. A parity bit is a bit added to a bit string to adjust the parity. When data is transmitted each character is encoded as a 7-bit binary number. An eighth bit is added to make a byte. This bit is called a 'parity bit' and is used to ensure that the data received is not corrupted in any way. A system can use either even or odd parity. In an even parity system the receiver checks that each received byte contains an even number of 1s. In an odd parity system the receiver checks that each received byte contains an odd number of 1s. If this is not the case then an error must have occurred. A request will be sent to the transmitter to ask it to send the byte again.

For example if a system is using even parity and the letter C has to be sent, the letter must first be encoded. The ASCII code for the letter C is 1000011. Since this code contains an odd number of 1s a 1 is added to the left hand side (e.g. 11000011) so the total number of 1s in the byte would be even. If the computer was using odd parity then a 0 would be added so the total number of 1s would be odd (e.g. 01000011).

Typical input devices

Typical input devices include the following peripherals. In this next section, we will look at each of these in turn.

<table>
<thead>
<tr>
<th>Keyboard</th>
<th>Image scanner (OCR)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mouse</td>
<td>Flatbed/drum scanners etc</td>
</tr>
<tr>
<td>Joystick</td>
<td>Optical Mark Reader (OMR)</td>
</tr>
<tr>
<td>Digitising tablet</td>
<td>Barcodes</td>
</tr>
<tr>
<td>Gaming wheel</td>
<td>RFID</td>
</tr>
<tr>
<td>Gaming pad</td>
<td>Magnetic reader</td>
</tr>
<tr>
<td>Light gun</td>
<td>Smart cards</td>
</tr>
<tr>
<td>Dance pads</td>
<td>Voice data entry</td>
</tr>
<tr>
<td>Motion Sensor control</td>
<td>Sound capture</td>
</tr>
<tr>
<td>Touch-sensitive screen</td>
<td>Video capture</td>
</tr>
<tr>
<td>Light pen</td>
<td>Biometric systems</td>
</tr>
<tr>
<td>Digital stills camera</td>
<td>Remote control</td>
</tr>
<tr>
<td>Magnetic Ink Character</td>
<td>Sensor</td>
</tr>
<tr>
<td>Recognition (MICR)</td>
<td>Interactive Whiteboard</td>
</tr>
</tbody>
</table>