## Data capture methods

### http://www.ixbt.com/short/images/paper_form_pv.jpgOnline and paper-based forms

Paper-based forms are a manual method of collecting data.

Paper-based surveys can be used to identify shopping habits or restaurant efficiency, for example.

Paper-based forms can be posted to customers for completion. Examples include disability, job seeker’s allowance and voting forms.

Paper-based forms are not popular as they are slow to collect, slow to process, and, if posted, cost more money. In addition, errors are not immediately detected and, in some cases, forms are not returned.

Paper-based forms are gradually being replaced by online forms. Online forms require a computer operator, or even the users themselves, to enter the data into a form template shown on a computer screen.

This type of form is used for online shopping surveys, VAT returns, ticket booking systems and hotel bookings. An advantage over the manually filled-in form (paper-based) is that validation software can be applied to the online form thus helping to reduce errors. In addition, because the data is being entered immediately into a computer system, the processing of this data is more efficient.

### Automated data capture systems

Automated data capture systems do not require humans to be present.

**Control system sensors** (e.g. heat, light, moisture sensors)

These measure analogue data and then use an ADC (analogue to digital convertor) to turn the analogue signals into digital signals which a computer can process. Once entered into the computer system, this data can be processed and graphs, etc. can be produced of the data. If the processed data has to be fed back into the system to control what happens next, this is regarded as **feedback**.



#### barcode readerBarcode readers

Barcode readers read barcodes which are attached to products.

A barcode is a series of thick and thin, dark and light lines which are read by a laser.

The barcode contains the product code, the manufacturer code, the country of origin and a check digit. It is normally accompanied by a 13-digit number which can be manually entered if the reader cannot read the code. Barcode readers are normally used in supermarkets where the barcode is scanned at the point of sale (POS) terminal and the data is compared to the overall stock database in the main computer system, and the price, etc. is returned to the POS to assist with the customer bill.



#### http://smallbiztrends.com/wp-content/uploads/2011/02/rfid-large.jpgRadio frequency identification device (RFID)

An RFID is a radio frequency identification system that consists of three components. An RFID tag/transponder microchip is electronically programmed with unique information and contains an antenna which uses radio signals to read and write to the tag when it is within reach of the RFID reader. It then passes the information to and from the host computer in digital form so processing can take place. The information contained on the tag can be for used for identification or location.

RFID tags are gradually replacing barcodes as they can hold varying types of information and do not need to be in line of sight of the reader. Paying for meals at drive-through restaurants and paying at road toll gates without stopping are some examples. Everyday uses include tracking animals (as shown on the right) and goods in a retail store (security).

The Baja Beach Club in Barcelona implant RFID tags into arms of patrons of the VIP section. The microchip has a barcode and it emits small radio frequency identification (RFID) signals.

Any information loaded on to the chip is displayed on a computer screen in the club. This means that the club can keep track of their VIPs and serve them accordingly. The tag not only identifies the person but functions as a debit tag. Many of the customers only wear bathing suits so this eliminates having to carry purses or wallets.



Near field communication (NFC)

NFC is a set of standards that establishes radio communications between devices, such as smartphones, when they are very close together or touching. This means the device does not have to be physically connected to a device in order to transfer information. For example, football stadiums are experimenting with allowing members entry to the grounds by waving mobile phones at an NFC reader. Payment for goods can also be made by waving a phone near the specified receiver.



## Factors affecting the choice of data collection method

The nature of the business or organisation affects the choice of the method of input. Where the business is based, where the customers are based, what processing has to take place, how much money the company has, how much security has to be applied to the data right down to how skilled the organisation staff are all need consideration. The following factors must be taken into account:

### Nature of information to be collected

#### Environmental conditions

 Paper-based methods of data capture are not environmentally friendly

 Whether the information to be collected is analogue or digital

#### Location of information

 Information collected centrally lends itself to online collection over manual collection

 Postal information needs to be collected

 Does the collected information have to be close to the processing site?

### Cost

#### Source document

 Time cost required to prepare the data, e.g. staff wages

 Financial cost in terms of paper, postage, etc.

#### Data entry processing

 If it has to be done quickly then staff costs will be higher

 If edits are required to be made to the data, the whole process will take longer

 The amount of data that has to be keyed in – the more data, the longer it takes

 The legibility of the document, e.g. handwriting, etc., if it is illegible then it takes longer and so more cost is involved

 If data is to be scanned, cost is affected by:

o size of the document

o uniformity of the pages

o document condition

o arrangement of the pages, e.g. double-sided

o quality of the paper

o whether batch processing of documents can be used

### Availability

 Paper is readily available for manual collection whereas as barcodes and magnetic cards have to be specially produced

 Availability of staff to prepare and collect the information – e.g. offshore employees, teleworkers

#### Ease of use

 What ability has the person giving the information, e.g. a child

 Form design

 OMR lends itself to the collection of marks in predefined places

 OCR lends itself to the collection of typed text

 Handwritten data capture lends itself to manual forms

 Complexity of the data collected

 Keyboard suitable for low volumes of text

 Automatic data capture means less/no human intervention

### Data security

 No human intervention then data is more secure

 Consideration has to be given to where the source documents are held

 How many people have to use the source material before it gets processed

During 2006 approximately 780,000 people used online techniques, such as assistive technology and on-screen reader software, to complete the Australian census. The software had to allow the user to ‘save the data’ and be able to return to it to ‘alter’ and ‘check’ the data. Security methods employed included encryption of the data during transmission, decryption for processing and decryption when sending back to the user for altering. Only authorised ABS (Australian Bureau of Statistics) officers could view the data. Some forms still had to be collected manually and, therefore, the data collectors needed to identify those who had completed the form online. They did this by ensuring that the software sent an SMS message to the data collectors when an online form had been completed.

The reasoning behind using the e-census was:

 better for the disabled

 more convenient, more flexible

 data was kept private

 needed fewer data collectors to give out and collect forms

 more efficient processing

 fewer printed materials

 easy for the people to complete

 easier for the government to collate the statistics

However, all did not go smoothly:

 Not all the population were IT literate.

 The software had to be compatible with the wide variety of web browsers.

 The software had to perform well over slow dial-up connections.

 The system had to allow for the fact that the majority of the people submitted on the night the census was due in.

## How to design data capture forms to obtain specified information

The successful collection of the data, whether manual or online, relies on the efficiency and design of the data capture form.



## Coding information for use in a spreadsheet or database

Businesses need to code information for various needs:

 Privacy – coded information is more private

 Ease of entry – it is faster and easier to enter data

 Less memory is needed in the computer system

 Validation can be applied more easily

However, they need to be careful that the codes are understandable, easy to remember and any used do not reduce accuracy.

Below are some examples of coded data.

**M/F** Male/Female

**Y/N** Yes/No

**12/02/12** 12th February 2012

**NY**  New York

Once the data is collected it will need to be inserted into a spreadsheet and, therefore, any text will need to be converted to numerical information as spreadsheets cannot manipulate text very well.

Take the popular types of data capture questions:

Male/Female?

Yes/No?

What is your opinion of the service? Excellent/Satisfactory/Unsatisfactory

How do you travel to school? Bus/Car/Cycle/Walk

The idea when coding this information is to turn it into numbers.

Male/Female? Y = 1 and N = 0

Yes/No? Y = 1 and N = 0

What is your opinion of the service? Excellent (1) / Satisfactory (2) / Unsatisfactory (3)

How do you travel to school? Bus (1) / Car (2) / Cycle (3) / Walk (4)

This would mean a data capture form can be turned into something resembling the following:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Name | Male/Female | Are you under 14 (yes/no)?  | How do you travel to school?  | Are you happy with the school dinners? |
| Mary | 0 | 1 | 2 | 1 |
| Bill | 1 | 0 | 3 | 3 |
| William | 1 | 1 | 1 | 3 |

A **database** is also a popular method for processing data. Like in a spreadsheet, the data needs to be coded to save memory and speed up data entry/processing. Coding methods can be similar to those discussed above but, as the database can manipulate text much easier, it is also possible to enter the data not only as number codes but also as text codes. The rule, however, is still to make the codes easy to remember.

Examples:

**BB** Bed and Breakfast **NYK** New York

**FB** Full Board **LON** London

**HF** Half Board **BEL** Belfast