

Coding Data

What is coding of data?

Any system will need to have data collected, entered and stored. One method of storing data is to **assign codes** to it. This usually means shortening the original data in an agreed manner.

Example 1

Original data: Monday; Tuesday; Wednesday; Thursday; Friday

Coded data: Mon; Tues; Wed; Thurs; Fri

Example 2

Original data: Xtra Large; Large; Medium; Small

Coded data: XL; L; M; S

Reasons to code data:

It is common for much of the data collected and entered into a system to have some degree of repetition and redundancy i.e. extra information that does not add anything. And this pattern or repetition is why it is efficient to code the data in some way.

- **Speeding up data entry**

Let's take the example of collecting data about a person's gender. People can be either 'Male' or 'Female'.

Whilst these two options are easily understood by all, imagine having to enter the word 'Male' and 'Female' into a system many hundreds of times. It is a waste of time and effort because no extra information is contained in the full words compared to a single letter.

- **Increase accuracy of data entry**

The other issue is that no matter how accurate a person is at data entry, at some stage they are likely to make a mistake and might spell 'Male' as 'Mail' or 'Female' as 'Femal'. This type of mistake will make any results from your database queries unreliable.

Instead of entering 'Male' or 'Female' you could code the data and instead enter it as 'M' or 'F'.

Simply having to enter one letter instead of a possible six will speed up data entry. It will also cut down on the risk of mistakes being made with spelling.

- **Use of validation**

When data has been coded it makes it easier to use validation to check if the data entered is sensible. With the example above, the person entering the data could still make a mistake and enter 'S' instead of 'M' or 'F'.

But if you set up validation so that the field will only accept the letters 'M' or 'F' and absolutely nothing else then that should further cut down on possible mistakes.

- **Less storage space required**

Every letter that you store in your database system will take at least one byte of storage. If you store 'Female' as 'F' then you will save five bytes of storage space. If the system belongs to a large organization, there might be many thousands or millions of records stored - simply by coding one field, a huge amount of hard disk storage can be saved.

- **Faster searching for data**

The smaller the size of your database, the faster it will be to search and produce results. Thus by coding data and keeping the size of the system to a minimum the more time you can save in the long run when running queries.

Problems caused by coding data:

Whilst coding data can bring many benefits it can also lead to some problems.

- **Coarsening of data**

This means that during the coding process some of the subtle details in the data are lost.

The colours could be classed as:

Light pink, pale blue, black and mid blue

However, when these colours are coded they may become:

PK (pink), B (blue), BK (black), BE (blue)

In this case, no allowance has been made for shades of colours. The fine detail has been lost. This is what is meant by 'coarsening of data'.

- **Coding can obscure the meaning of the data**

A reader seeing the 'gender' data as M/ F is pretty likely to know that it means Male/ Female.

But some codes are more obscure, for example the country code for Switzerland is CHE. Many people might not recognize what this code represents.

- **Coding of Value Judgments**

When you are collecting data about people's opinions it might be difficult to code their answers with accuracy. The code they give will depend on their individual opinion. Coding of value judgments will inevitably lead to coarsening of the data since there will be a wide range of opinions that could be held and only a limited number of codes available.

Further examples of data coding:

In our everyday lives we come across many examples of how coding is used to represent data. Here are just a few more ideas:

- **Country names**

The name of a country can be represented by two letters. For example:

Great Britain - GB

France - FR

Canada – CA

- **Airline flight codes**

When you fly you may have noticed that your flight is given a code.

This code consists of two letters to identify the airline that you are flying with. The letters are usually followed by numbers to represent a particular route.

Examples:

So for example, a British Airways flight from Heathrow to Oslo might be coded as BA766.

A flight operated by the airline company Emirates which depart from Dubai and arrives at Heathrow might be coded as EK029.

Advantages of coding:

- Data entry can be faster
- Data entry can be more accurate
- Validation can further improve accuracy
- Less storage space required

- Faster searching for data
- Coded data can be more secure if people don't know what it means

Disadvantages of coding:

- Coarsening of data
- Meaning of data can be obscured
- Value judgments are difficult to code
- If people don't know the code it can slow down data entry
- If codes are complicated they might be entered incorrectly
- Might run out of code combinations