6

THE NATURE OF DATA

We are all aware that we live in an information age. It is said that the amount of information in the world is doubling every month, or is it every week? Who knows? One thing is for sure – we are deluged with the stuff every day. Data is another word for bits of information (singular – datum). Research uses data as the raw material to reach conclusions about an issue. The issue being investigated dictates what data need to be collected.

Although many data appear to be solid fact and permanently represent the truth, this is not the case. Data are not only elusive, but also ephemeral. They may be true for a time in a particular place, as observed by a particular person, but might be quite different the next day. Take, for example, a daily survey of people's voting intentions in a forthcoming general election. The results will differ each day, even if the same people are asked, because some change their minds due to what they have heard or seen in the interim period.

Data are not only ephemeral, but also corruptible. Hearsay, second-hand reports and biased views are often paraded as facts. The further away you get from the event the more likely it is that inconsistencies and inaccuracies creep in. Memory fades, details are lost, recording methods do not allow a full picture to be given, and distortions of interpretations occur.

Because it is dangerous for a researcher to insist that his or her data – and findings derived from them – are infallible, the outcomes of research are often couched in 'soft' statements, such as 'it seems that', 'it is likely that', 'one is led to believe that', etc. This does not mean, however, that the knowledge gained is useless, only that it is not absolutely certain – like most things in life.

LEVELS OF ABSTRACTION

How do data, the raw materials of research, relate to knowledge as a whole? They are part of a hierarchy of information, going from the general to the particular; from abstract to concrete. Understanding this hierarchy makes it possible to break down research problems expressed in theoretical language to more practical components that can be measured in some way. This hierarchy can be expressed as follows:

- **theory** – abstract statements that make claims about the world and how it works. Research problems are usually stated at a theoretical level;
- **concepts** – building blocks of the theory which are usually abstract and cannot be directly measured;
- **indicators** – phenomena which point to the existence of the concepts;
- **variables** – components of the indicators which can be measured;
- **values** – actual units of measurement of the variables. These are data in their most concrete form.

You can see that the briefest statement of a theory will be the most general and abstract, while the detailed components of the statement will become increasingly particular and concrete. Each theory will contain several concepts, each concept several indicators, each indicator several variables, and each variable several values. For example:

- **theory** – poverty leads to poor health;
- **concepts** – poverty, poor health;
- **indicators** of poverty – low income, poor living conditions, restricted diet, etc.;
- **variables** of poor living conditions – levels of overcrowding, provision of sanitary facilities, infestations of vermin, levels of litter, etc.;
- **values** of levels of overcrowding – numbers of people per room, floor area of dwellings, numbers of dwellings per hectare, etc.

Theory

Although the meaning of the term 'theory' is rather imprecise, in research it refers to a statement that makes a claim about a phenomenon. Theories

can range from complex, large-scale, well-researched and substantiated claims developed through academic research, to informal guesses or hunches about specific situations. Our understanding of how the world works is based on theories, and much research is concerned with challenging, refining and extending existing theories, or developing new ones.

As theories, statements tend to be expressed in abstract terms. It is necessary to break them down into their constituent parts in order to examine them. Statements are usually made up of concepts and how they relate.

Concepts

A concept is a term for a phenomenon, often quite abstract, such as alienation, socialism, equilibrium, society; but it can also be quite concrete, such as animal, town, income.

We use concepts all the time as they are an essential part of understanding the world and communicating with other people. Many common concepts are shared by everyone in a society, though there are variations in meaning between different cultures and languages. For example, the concept of 'respect' will mean something different to a streetwise rapper than to a noble lord. Some concepts can only be understood by experts in certain disciplines, e.g. dermatoglyphics, milfoil, parachronism, anticipatory socialization, etc.

Concepts should be clearly defined so that they can be understood in the same way by everyone. This is relatively easy in the natural sciences where precise definitions of concepts such as radio waves, acceleration and elements are possible. In the humanities and social sciences, this may be much more difficult – e.g. concepts such as beauty, honour, motivation, kinship, etc. – as their meanings are often based on opinions, emotions, values, traditions and so on.

Indicators

Concepts that are abstract in nature can be difficult to detect, evaluate or measure. Take 'anxiety' as an example. How could you detect this in a person? The answer is to look for indicators – those perceivable phenomena that give an indication that the concept is present. In this example, the signs that might indicate anxiety could be trembling,

worried facial expression, pacing up and down, sweating, shortness of breath, etc. In most scientific and technical subjects, indicators are usually well defined and universally accepted. In the humanities and social sciences they often need to be carefully defined in each research project, as this consensus is frequently lacking.

Variables

In order to gauge the extent or degree of an indicator, you will need to find a measurable component. In the case of anxiety, it would be very difficult to measure the level of worry in a person's expression, but you could easily measure their rate of breathing.

In the natural sciences, the identification of variables is usually relatively simple. Area, temperature, speed and velocity are some examples. Some of these may be appropriate to the social sciences, particularly in quantitative studies, e.g. the number of people in a demonstration, type of occupation, income, etc.

Values

The values are the units of measurement used to gauge the variables. The level of precision of a measurement depends on the nature of the variable and the type of values that are appropriate. Certain scientific studies require that variables are measured incredibly accurately, while some social variables might only be gauged on a three-point scale such as 'agree', 'neutral', 'disagree'.

Data, seen as bits of information, can be used at any level of abstraction. Research projects usually start at the more abstract end of the spectrum, move to the more concrete information during the investigation, and return to the abstract in the conclusions. Data that can be manipulated, measured and analysed tend to be more at the values level but, in many subjects in the humanities and social sciences, the variables may be difficult or even impossible to measure with precise values.

You can relate these levels of abstraction to how to structure your research. Your title and main research question will be expressed at a theoretical level, and your sub-questions will be about the separate concepts. To investigate these, you will need to find out what type of measures can be used to assess the existence and scale of the concepts, then the scales that can be used in the measures, i.e. the type

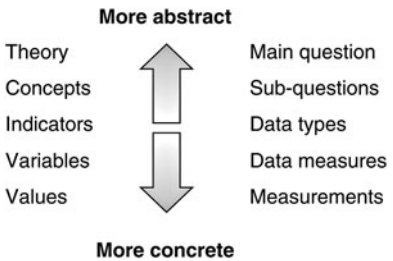

Figure 6.1 Diagram of levels of abstraction

of measurements, and finally the actual measurements that provide the basic data for analysis. Figure 6.1 provides a simple diagram to illustrate the levels of abstraction in your research structure.

PRIMARY AND SECONDARY DATA

Data come in two main forms, depending on their closeness to the event recorded. Data that have been observed, experienced or recorded close to the event are the nearest one can get to the truth, and are called **primary data**. Written sources that interpret or record primary data are called **secondary sources**, which tend to be less reliable. For example, reading about a fire in your own house in the newspaper a day after it occurred will probably give you less accurate information than what you gained by experiencing the event yourself. You will be more informed about the facts and these will not be distorted by someone else's interpretation.

Primary data

We are being bombarded with primary data all day. Sounds, sights, tastes, tactile things are constantly stimulating our senses. We also have instruments to measure what we cannot so accurately judge through our senses, such as clocks, barometers, business accounts, etc.

There are four basic types of primary data, distinguished by the way they are collected:

1 **measurement** – collections of numbers indicating amounts, e.g. voting polls, exam results, car mileages, oven temperatures, etc.;

2 **observation** – records of events, situations or things experienced with your own senses and perhaps with the help of an instrument, e.g. a camera, tape recorder, microscope, etc.;
3 **interrogation** – data gained by asking and probing, e.g. information about people's convictions, likes and dislikes, etc.;
4 **participation** – data gained by the experience of doing things, e.g. the experience of learning to ride a bike tells you different things about balance, dealing with traffic, etc. rather than just observing.

Primary data are the first and most immediate recording of a situation. Without this kind of recorded data, it would be difficult to make sense of anything but the simplest phenomenon or to communicate the facts to others.

Primary data can provide information about virtually any facet of our lives and surroundings. However, collecting primary data is time-consuming and not always possible. Although more data usually mean more reliability, it is costly to organize large surveys and other studies. Furthermore, it is not always possible to get direct access to the subject of research. For example, many historical events have left no direct evidence.

Secondary data

Secondary data are data that have been interpreted and recorded. Just as we are bombarded with primary data, we are cascaded with secondary data in the form of news bulletins, magazines, newspapers, documentaries, advertising, the Internet, etc. These data are wrapped, packed and spun into pithy articles or digestible sound bites. The quality of the data depends on the source and the methods of presentation. Refereed journals containing papers vetted by leading experts and serious journals, such as some professional and trade journals, will have authoritative articles by leading figures. Magazines can contain useful and reliable information or be entirely flippant. The same goes for books – millions of them! They range from the most erudite and deeply researched volumes to ranting polemics and commercial pap. Television and radio programmes vary likewise, as does information on the Internet.

A major aspect of using secondary data is the assessment of the quality of information or opinions provided. This is done by reviewing

the quality of evidence that has been presented in the arguments, and the validity of the arguments themselves, as well as the reputation and qualifications of the writer or presenter. It is also good practice to compare data from different sources. This will help to identify bias, inaccuracies and pure imagination. It will also show up different interpretations that have been made of the event or phenomenon.

QUANTITATIVE AND QUALITATIVE DATA AND LEVELS OF MEASUREMENT

Data are also divided into two other categories, referring not to their source but to their characteristics: basically, whether they can be reduced to numbers or presented only in words. This affects the way that they are collected, recorded and analysed.

Numbers are used to record much information about science and society, for example pressures, bending forces, population densities, cost indices, etc. This type of data is called **quantitative data**. Numbers can be analysed using statistical techniques. However, a lot of useful information cannot be reduced to numbers. People's judgements, feelings of comfort, emotions, ideas, beliefs, etc. can only be described in words. These record qualities rather than quantities, hence they are called **qualitative data**. Words cannot be manipulated mathematically, so require quite different analytical techniques.

Quantitative data

Quantitative data can be measured, more or less accurately, because they contain some form of magnitude, usually expressed in numbers. You can use mathematical procedures to analyse the numerical data. These can be extremely simple, such as counts or percentages, or more sophisticated, such as statistical tests or mathematical models.

Although some forms of data are obviously expressed as numbers – e.g. population counts, economic data, scientific measurements, etc. – other forms that seem remote from quantitative measures can also be converted to numbers. For example, people's opinions about the performance of political parties seem to be difficult to quantify. But if a set choice of answers is given in a questionnaire, then you can then count the numbers of the various responses. The data can then be treated as quantitative.

Table 6.1 Set of quantitative data

Name	*Test 1 %*	*Test 2 %*	*Test 3 %*	*Average %*
Adams, Rolf	35	64	47	49
Boulter, Helen	55	74	63	64
Carter, Jim	45	68	55	56
Durrant, Chris	63	47	64	58
Escaliente, Laura	36	68	37	47
Fuente, Karl	47	57	57	54
Gardiner, Rachel	53	49	70	57

Census figures (population, income, living density, etc.), economic data (share prices, gross national product, tax regimes, etc.), performance data (e.g. sport statistics, medical measurements, engineering calculations, etc.) and all measurements in scientific endeavour are typical examples of quantitative data. Table 6.1 provides a simple example of a set of quantitative data, in this case referring to scores achieved by seven individuals in three tests.

Qualitative data

Qualitative data cannot be accurately measured and counted, and are generally expressed in words rather than numbers. Essentially human activities and attributes such as ideas, customs, mores, beliefs, that are investigated in the study of human beings and their societies and cultures cannot be pinned down and measured in any exact way. These kinds of data are therefore descriptive in character. This does not mean that they are any less valuable than quantitative data; in fact, their richness and subtlety often lead to great insights into human society.

Qualitative research depends on careful definition of the meaning of words, the development of concepts and variables, and the plotting of interrelationships between these. Concepts such as affluence, happiness, comradeship, loyalty, etc., are real and detectable, even if they are difficult to record and measure.

Observation notes, interview transcripts, literary texts, minutes of meetings, historical records, memos and recollections, documentary films, are all typical examples of qualitative data. Some are recorded very close to the events or phenomena, while others may be remote and highly edited interpretations, so assessments of the reliability of

these must be made. Also, qualitative data rely on human interpretation and evaluation and cannot be dispassionately measured in a standard way. Checks on the reliability and completeness of qualitative data can be made by consulting a variety of sources of data relating to the same event – this is called triangulation. Table 6.2 illustrates an example of qualitative data, in this case personal reactions to a change of work practices in a factory of various individuals in different roles.

Research, particularly when about human beings, often combines the examination of both qualitative and quantitative data. In fact, there are many types of data that can be seen from both perspectives. For example, a questionnaire exploring people's attitudes to work may provide a rich source of qualitative data about their aspirations and beliefs, but might also provide useful quantitative data about levels of skills and commitment. What is important is that appropriate analytical methods are used for the different types of data you are dealing with.

MEASUREMENT OF DATA

Data can be measured in different ways depending on their type. These are commonly referred to as levels of measurement – **nominal**, **ordinal**, **interval** and **ratio**.

Nominal level

Nominal measurement is very basic – it divides the data into separate categories that can then be compared with each other. By sorting out the data using names or labels, you can build up a classification of types or categories. This enables you to include, or exclude, specific cases in the types and compare them. For example, buildings may be classified into many types, e.g. commercial, industrial, educational, religious, etc. Some definitions allow only two types – called **dichotomous** – e.g. sex (male or female), while others fall into a set number such as marital status (single, married, separated, divorced or widowed). What is important is that every category is distinctive and that there is no overlap between them which makes it difficult to decide where to place a piece of datum. Ideally, it should be possible to categorize all the data, though sometimes you will need a 'remainders' category for those that cannot be so ordered.

Table 6.2 Set of qualitative data

Role	*Reaction*				
	Ease of work	*Speed of work*	*Duration of work sessions*	*Level of efficiency*	*Remuneration*
Unskilled factory worker	More complicated work sequence	Takes less time to complete each task	Prefer more tea breaks	Mistakes made because difficult to concentrate for long periods	More difficult to achieve bonus payments
Skilled operative	More logical production stages	Overall reduction in time taken to produce each component	Longer sessions lead to tiredness	Too much time needed preparing and clearing up	Less overtime payments not offset by increase in wages
Supervisor	Easier to control quality	Less time needed in supervision	Less downtime	More productive working	No extra pay for time needed in making changes
Middle management	Fewer consultations required	Greater output per month	More difficult to plan staffing due to long shifts	Fewer wasted resources	Less complicated pay structure
Managing director	Easier to achieve consistent output	Improved overall productivity	Better utilization of staff and machinery	Shorter manufacture times	Greater profit margins

Nominal data can be analysed using simple graphic and statistical techniques. Bar graphs, for example, can be used to compare the sizes of categories and simple statistical properties, such as the percentage relationship of one subgroup to another, or of one subgroup to the total group, can be explored.

Ordinal level

This type of measurement puts the data into order according to a property that they all share, such as size, income, strength, etc. Precise measurement of the property is not required, only the perception of whether one is greater or lesser than the other.

For example, a class of children can be lined up in order of size without measuring their heights; the runners in a marathon can be sorted by the order in which they finished the race. Likewise, we can measure members of the workforce on an ordinal scale by calling them unskilled, semi-skilled or skilled.

The ordinal scale of measurement increases the range of statistical techniques that can be applied to the data.

Interval level

With this form of measurement, the data must be able to be measured precisely on a regular scale of some sort, without there being a meaningful zero. For example, temperature scales: in the Fahrenheit, Celsius and Réaumur scales, the gradation between each degree is equal to all the others, but the zero point has been established arbitrarily. They each precisely measure the temperature, but the nought degrees of each are different. Another example is the calendar date – compare the Chinese and Western calendars.

In the social sciences, some variables, such as attitudes, are frequently measured on a scale like this:

Unfavourable −4 −3 −2 −1 0 +1 +2 +3 +4 Favourable

Despite appearances, you must be cautious to interpret this as a true interval scale, as the numbers are not precise measurements and indicate preferences on an essentially ordinal scale.

The interval level of measurement allows yet more sophisticated statistical analysis to be carried out.

Ratio level

The ratio level of measurement is the most complete level of measurement, having a true zero: the point where the value is truly equal to nought. Most familiar concepts in physical science are both theoretically and operationally conceptualized at a ratio level of quantification, e.g. time, distance, velocity, mass, etc.

A characteristic difference between the ratio scale and all other scales is that the ratio scale can express values in terms of multiples of fractional parts, and the ratios are true ratios. For example, a metre is a multiple (by 100) of a centimetre, a millimetre is a tenth (a fractional part) of a centimetre. The ratios are 1:100 and 1:10. There is no ambiguity in the statements 'twice as far', 'twice as fast' and 'twice as heavy'. Of all levels of measurement, the ratio scale is amenable to the greatest range of statistical tests.

In summary, you can use the following simple test to determine which kind of data measurement you impose on the values of a variable. If you can say that:

- one value is different from another, you have a **nominal scale**;
- one value is bigger, better or more of anything than another, you have an **ordinal scale**;
- one value is so many units (degrees, inches) more, or less, than another, you have an **interval scale**;
- one value is so many times as big, or bright, or tall or heavy as another, you have a **ratio scale**.

Figure 6.2 provides a summary of the levels of measurement.

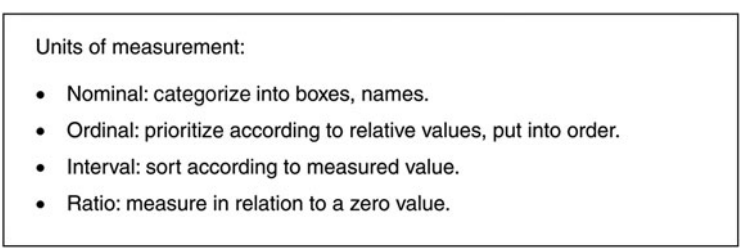

Figure 6.2 Levels of measurement

WHERE TO FIND OUT MORE

What counts as data, and what to do with it, is a big subject in research and gets dealt with exhaustively in most books about academic research. Below are some useful other ways of looking at this aspect, without getting too deeply into technicalities.

Seale, C. (ed.) (2012) *Researching Society and Culture* (third edition). London: Sage.
A well explained section on theories, models and hypotheses appears in chapter 5.

Cooper, D.R. and Schindler, P.S. (2013) *Business Research Methods* (twelfth edition). New York: McGraw-Hill.
Chapters 11 and 12 deal in detail with data measurement and scales, with some useful examples given to bring them to life.

Leedy, P.D. and Ormrod, J. (2015) *Practical Research: Planning and Design* (eleventh edition). Harlow: Pearson.
Chapter 4 provides a rather nice philosophical approach to the nature of data.

Blaxter, L., Hughes, C. and Tight, M. (2010) *How to Research* (fourth edition). Buckingham: Open University Press.
The first part of chapter 8 provides another angle on data and its forms.