PEOPLE: International Journal of Social Sciences
ISSN 2454-5899

El Hacen Moulaye Ahmed, 2019

Volume 5 Issue 1, pp. 718-726

Date of Publication: 2nd May 2019

DOI-https://dx.doi.org/10.20319/pijss.2019.51.718726

This paper can be cited as: Ahmed. E. H. M., (2019). The Nature and Types of Data. PEOPLE: International Journal of Social Sciences, 5(1), 718-726.

THE NATURE AND TYPES OF DATA

El Hacen Moulaye Ahmed
Modern University of Nouakchott, Nouakchott, Mauritania
hacentwo@hotmail.fr

Abstract

The current paper lays out the nature and types of data. It is underpinned by models presented in two chapters from two different books: Walliman's (2011) "the Nature of Data" and Knight and Parsons' (2005) "What Kind of Data Do I Need and How Do I Get Them?" The chapters are selected because they represent heuristic and representative models of the nature and types of data. Walliman's model represented data as a fugacious and fractional sight of the events, opinions, or beliefs. Based on diverse criteria and situations, Walliman governed his typologies. Based on type of source, for example, they were described as either primary or secondary, and based on their characteristics, data can be seen as quantitative or qualitative. Knight and Parsons' model listed only two categories, primary and secondary. In addition, they explained in depth their advantages and disadvantages which were hesitantly referred to in Walliman's chapter.

Keywords

Data, Information, Types of Data, Nature of Data, Advantages and Disadvantages of Data

1. Introduction

We no longer suffer from the weight of history, as Marx suggested, but from the burden of an ever increasing density of information (Rutsky & Cohen, 2005, p. 1). Indeed, with the tremendous quantity of data that saturates today's society, we can get answers to almost all

questions in nanoseconds no matter where we are (Mackall, 2004, p. 9). Because answers are so effortlessly available, Tensen (2013) stated that it is understandable why many members of the "Millennial Generation" have begun to question the relevance of the research paper and the purpose of an assignment (p. v). Accordingly, students want a rationale to the value of their time and money spent on conducting such research studies, and so this paper thereby seeks to address their questions and reservations.

Broadly speaking, writing a research paper is a way of remembering, understanding, evaluating, and developing skills (Williams, Colomb & Booth, 2005, p. 13). The effectiveness of fulfilling these purposes will be determined by the nature and the types of data in which the research paper is grounded simply because not all data are created equal. However, before talking about the types of data, it is worth clearing up an issue that crops up a lot with data. Time (2010) noted that the origin of the word "data" is from the Latin word "datum," which means "given" or a "piece of information" (p. 1). Data is simply the plural form of datum, observed Coleman (2013, p. 3). Although many research studies are dotted with the use of the term "data" as a synonym to the term "information," Oz (2009) argued that they differ from each other (p. 9). According to Rob, Morris, and Coronel (2013) "data are raw facts. The word raw indicates that the facts have not yet been processed to reveal their meaning…. information is the result of processing raw data to reveal its meaning" (pp. 5-6). This means that information is represented by data. In Ratzan's words, "information is a coherent collection of data organized in a particular way that has meaning" (2004, p. 5).

To reveal the nature and types of data, the paper reviews two chapters, each of which will be addressed in separate sections. The first chapter is entitled "The Nature of Data" and is taken from *Research Methods: The Basics* by Walliman (2011). The second chapter is entitled "What Kind of Data Do I Need and How do I Get Them?" and it is taken from *How to Do Your Dissertation in Geography and Related Disciplines* by Knight and Parsons (2005). These two chapters were chosen because they are representative of different divisions of data. For example, researchers divide data into two main categories primary and secondary data based on their sources (Ruddick, Sherwood, Wrenn, & Stevens, 2006, p. 90). Other researchers, such as Stead (2001), extend the division to include other categories that are based on their characteristics, basically data reduced in numbers and others presented in words (p. 40).

N.B. The term "data" will be used interchangeably throughout this paper with the term "information."

2. "The Nature of Data"

Generally speaking, in this chapter, Walliman focuses on the nature of data, their level of measurement, and their typologies. To start with, he stated that data is not only elusive, but also, ephemeral and corruptible. For instance, a survey conducted to measure people's voting intentions in a forthcoming election will not come up with the same results as another one conducted in another time even if the samples, people who are engaged in the phenomenon (Piepenburg, 2011, p. 52) [*reference added*], are the same. They are corruptible in the sense that second hand reports and biased views are often exhibited as facts. In other words, the recording methods are not one hundred percent trustworthy and a distortion of interpretation often occurs.

In the nature of data still, Walliman states that data is related to knowledge as a whole through a hierarchical system, going from general to particular, from abstract to concrete. This hierarchy is expressed as the following:

- **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. (p. 66)

This hierarchy shows that each theory contains several concepts, each concept several indicators, each indicator several variables, and each variable several values. It also shows that the level of abstraction is expressed from the overly theoretical to the extremely practical. Putting this hierarchy into a practical form, Walliman gave the following 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 areas of dwellings, numbers of dwellings per hectare, etc. (p. 66)

Having defined the nature of data as elusive, ephemeral and corruptible, and that it can take two forms abstract and concrete, Walliman argues that there are four types of data, primary data, secondary data, quantitative and qualitative data, and assesses the advantages and disadvantages of each one of them. Based on their sources, they are either primary or secondary. On the one hand, primary data refer to those data that have been observed, experienced, or collected close to the event. In other words, these data are the first and the most immediate recording of a situation. These data are the cornerstones of any research project. Without them it is difficult to make any depth study to the targeted phenomenon. However, recording primary data is time consuming and not always possible, in the sense that some historical events have left no direct evidence.

Furthermore, Walliman states that primary data can be divided into four basic types, 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. 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 experiences 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. (p. 70)

Willman's classifications reveal that they are done in accordance with their closeness to the event recorded. It is noticeable also that the researcher is directly involved in the collection of these data.

Secondary data, on the other hand, refer to the interpretation of primary data, or data that are already interpreted and recorded. This type of data takes many forms, news bulletins, books, magazines, newspapers, documentaries, advertising, the internet etc. In comparison with primary data, these data are less reliable; however, they are not necessarily unreliable. Instead, caution is needed once it comes to collecting this type of data. To assess the quality of data, the researcher should review the quality of evidence that has been presented to confirm the validity of the data which can be done by comparing it with data from different sources.

In addition, referring to their characteristics, Willman states that data can be divided into two other categories, quantitative and qualitative data. Quantitative data are presented in the form of numbers, and they can be measured, more or less accurately. Even though some forms of data can be recorded in numbers, scientific measurement and population counts, other forms of data can only be converted into numbers, peoples' opinions and beliefs. Census figures, economic data, performance data and all measurement in scientific endeavor, are typical examples of quantitative data. However, some information cannot be reduced into numbers, essentially human activities and attributes, people's feelings, emotions, ideas, and beliefs. Rather, this type of data can only be described in words. Hence, they are called qualitative data. They are presented in different forms: observation notes, documentaries, interview transcripts, literary texts, and historical records to name a few. Some of these data are closely related to the phenomenon, while others are only remotely connected. This type of data is based on interpretation. Thus, assessment of reliability "must" be made.

In short, this chapter walks the reader through the nature of data and their typologies. Walliman argues that data do not present facts simply because this will imply a status of solidness, permanence, and truthfulness, but rather they present a fugacious and fractional sight of the events, opinions, or beliefs. Moreover, in relation with knowledge as a whole, Walliman capitalizes on the nature of data as abstract and concrete. Furthermore, four types of data were distinguished. According to their type of source, primary data refer to data that are directly collected from the actual circumstances to which they pertain such as transcripts of legal proceedings and government documents, whereas secondary data refer to data that are collected from pre-recorded sources like registers, newspapers, and books. Walliman also states that primary data are more reliable than secondary data. This does not mean, however, that credibility cannot be obtained when dealing with secondary data. Walliman zeroes on another pair of distinction. Based on data characteristics, data can be either quantitative or qualitative. Unlike quantitative data which are expressed in numbers as in pressures, and cost indices, qualitative data are expressed in a natural language description like peoples' judgments, emotions, and ideas.

3. "What Kind of Data Do I Need and How do I Get Them?"

Knight and Parsons (2005), in this chapter, cover the types of data and the advantages and disadvantages of each. Data are classified as either primary data or secondary data. Primary

data refer to any data that the researchers themselves collect directly from the actual circumstances and for the specific research project at hand. Typical examples of primary data include measurements and questionnaires that are administered in the field. The main advantage of collecting primary data is that they are specific to the phenomenon. Moreover, it is easy to assess their credibility because the researcher is familiar with them.

However, compiling primary data consumes a lot of time and money and may result in a small data set. This means that because this data set depends highly on the sample and response rate, it may provide only a snapshot of the phenomena examined. Furthermore, collecting this type of data is tedious, and it can be both slow and inaccurate. Respondents, for example, may give answers that are not what they believe, but rather what they think they should believe. In addition, compiling these data requires experience. Furthermore, mistakes in writing the questions may lead to miscommunication between the researchers and the respondents. The researcher can overcome these disadvantages by undertaking a pilot study, "pretested early draft to assess the validity of the survey" (McNabb, 2010, p. 105) [*reference added*]. Again this, of course, is time taking, but it makes time worth spending. It would also be helpful to make a preliminary analysis of the gathered data. Of course these steps are time consuming, but they can prevent a lot of errors that might otherwise be made.

Unlike primary data, secondary data are data that are collected from pre-recorded sources and that the researcher will use for his or her own purposes. For instance, surveys that have been conducted by another researcher, written books, documentaries, and literary texts are typically secondary data. However, these examples are not only limited to secondary data. In effect, the nature of the discipline under which the researcher is carrying his or her own research study determines the type of data produced. In history, geography, and culture, for example, public records, written books, etc. are regarded as primary data. In this case, the term "secondary data" refers to data that are obtained from other studies that are contemporary with the data. Knight and Parsons stated that this distinction is not as straightforward as it might appear.

Like most things, using secondary data has its pros and cons. Their advantages and disadvantages are the converse of those for primary data. Secondary data have the advantage of being easily accessed. There is potentially a large set of data available, especially with the recent increase in accessibility of the academic journals and other sources available on the Internet. Accordingly, secondary data are both time and money saving. Nevertheless, like primary data, in some instances, collecting secondary data is time consuming because of the sheer volume of data

available and time required to check the reliability of the data. Moreover, secondary data may have restricted access and use. Yet, even if they are more readily accessible in a digital format, there is a hefty price to be paid for data in that format. In comparison, the distinction between primary data and secondary data is that primary data provide first-hand information which will help in assessing the reliability of collected data; while, for secondary data, the task is "difficult." Thus, with the lack of information on their quality, several measures such as the date and the relevance to the phenomenon examined should be taken to gauge the creditability of information.

Finally, this chapter is an essential reading for any student embarking on a research assignment. In addition to explaining the types of data, Knight and Parsons explained in depth their advantages and disadvantages which were hesitantly referred to in Walliman's chapter "nature of data." They limned that data fall into two broad categories, primary and secondary data. The data that are observed, experienced or recorded by the researcher(s) are referred to as primary data. However, the data that are collected from pre-existing sources are referred to as secondary data. Both types of data have advantages and disadvantages. On the one hand, primary data are specific to the research questions and their credibility is easily assessed because the researcher has an intimate knowledge of them; nevertheless, they exist in a small set and are time consuming. On the other hand, secondary data are quick to obtain and are available in a large set. Nonetheless, they may not answer the research questions and may have restricted access and use.

4. Conclusion

Why are students asked to write research papers when technology has made information so readily accessible? Although it is true that one of the purposes of writing a research paper is to get information, the urgent question one must ask is how. How, then, is an oxymoron, in the sense that it becomes an answer-question to the previous question raised by many members of the "Millennial Generation." Generally speaking, just as food nourishes the body, knowledge nourishes the intellect, and just as all food is not edible, all information is not relevant. Thus, conducting research studies typically requires both knowing the nature of data and the skills to evaluate them. The task is taxing, but the prize, is well worth it, since the evaluation skills ultimately are the measures upon which the difference between an adequate research paper and an excellent one will be made.

Walliman (2011) and Knight and Parsons (2005) introduce the nature of data, from their types to their pros and cons, in an understandable and well organized manner, yet their chapters

suffer in varying degrees. In terms of credibility, both of them introduced types of data as traffic lights with red, yellow, and green light. This division was made within a larger division they adopted in their chapters. Walliman, for example, divided types of data into four types basing his typology on both sources and characteristics. In effect, his division was inaccurate simply because he could have grouped his division under two types, thus have said that primary and secondary data can be characterized as quantitative or qualitative data.

Walliman states that data have two forms abstract and concrete form. Besides, unlike Walliman, who explored the elusive, ephemeral, and corruptible nature of data, Knight and Parsons, started their article with a hollow definition- "data is the plural of datum." However, they break their data into two types, based on their sources. Moreover, Knight and Parsons clearly state the relativity of this division in their statement that the classification of data as either primary or secondary is determined by the nature of the study, the nature of the problem, and the nature of the discipline under which the study is carried out. Furthermore, they do not only mention the advantages and disadvantages of data, but also offer ways through which the disadvantages can be minimized.

References

Coleman, L. S. (2013). *Measuring Data Quality for Ongoing Improvement: A Data Quality Assessment Framework*. Massachusetts: Ingenix, Inc.

Knight, P. G & Parsons, A. J. (2005). *How to Do Your Dissertation in Geography and Related Disciplines*. New York: Routledge.

Mackall, J. (2004). *Research and Information Management*. New York: An Imprint of Facts on File, Inc.

McNabb, D. E. (2010). *Research Methods for Political Science: Quantitative and Qualitative Approaches*. New York: M. E. Sharpe, Inc.

Oz, E. (2009). *Management Information Systems*. Massachusetts: Cengage Learning, Inc.

Piepenburg, K. (2011). *Critical Analysis of Hofstede's Model of Cultural Dimensions: To what Extent his Findings are Reliable, Valid and Applicable*. Norderstedt Germany: Grin Verlag.

Ratzan, L. (2004). *Understanding Information Systems: What They Do and Why We Need Them*. New York: The American Library Association.

Rob, Morris, S. A. & Coronel, C. (2013.) *Database Systems: Design, Implementation, and Management*. Massachusetts: Cengage Learning Customer & Sales Support.

Ruddick, M. E., Sherwood, P. K., Wrenn, B. & Stevens, R. E. (2006). *The Marketing Research Quide*. New York: Best Business Books, an Imprint of the Haworth Press, Inc.

Rutsky, R. L. & Cohen, S. (2005). *Consumption in an Age of Information*. Oxford: Berg.

Stead, G. B. (2001). *Planning, Reporting & Designing Research*. Cape Town: Maskew Miller Longman, Ltd. https://doi.org/10.1068/b2677

Tensen, B. L. (2013). *Research Strategies for a Digital Age*. Massachusetts: Wadsworth.

Time .(2010). *Data Interpretation and Logical Reasoning*. New Delhi: Magic International Pvt. Ltd, Greater Noida, UP.

Walliman, N. (2011). *Research Methods: The Basics*. New York: Routledge. https://doi.org/10.4324/9780203836071

Williams, J. M., Colomb, G. G. & Booth, W. C. (2005). *The Craft of Research*. Chicago: The University of Chicago Press.