International Journal of Democratic and Development Studies (IJDDS), Vol. 5, No. 2, January, 2022.
Available online at http://journals.rcmss.com/index.php/ijdds ISSN: 2350-224X (E) ISSN: 2346-7223 (P)
Agba, Michael Sunday & James, Rijami, 2022, 5(2):31-45

Data Integrity in Research: A Fundamental Prerequisite for Credible Research Output across Disciplines and Professions

Agba, Michael Sunday, PhD
Department of Public Administration, Faculty of Management Sciences
University of Calabar, Calabar – Nigeria
ORCID: https://orcid.org/0000-0001-7907-1079

James, Rijami

B.Sc. Hons. (Port Harcourt); ACPA (Lagos); PGDPA, M.Sc., Ph.D. (in view) (Calabar)
Department of Public Administration, Faculty of Management Sciences,
University of Calabar, Calabar – Nigeria
ORCID: https://orcid.org/0000-0002-2129-973x
E-mail: rijamidchampion@gmail.com

Abstract
The use of data as a basic tool for carrying out a study has become absolutely essential in research and other academic engagements. The position of this paper is that, since data is indispensable in research, its integrity must also be indispensable to achieve research credibility. This paper views data integrity as a process and explains it as the measures taken to ensure accuracy and completeness of a data set in research. Data integrity in research is a fundamental requirement because; it ensures the use of correct parameters, hypotheses testing, data analyses, reporting and evaluation. Quality Assurance, Quality Control, Data Collection, Data Management and ALCOA are some ways to secure and ensure data integrity. Data integrity in research involves a great deal of time, efforts and cost but must be encouraged. This study opines that, academic scholars should inculcate the culture of transparency in research into the minds of young researchers through data integrity.
Keywords*: Data Integrity, Research, Research Credibility, Rigour of Research, Data Collection, Data Management.*

Introduction

One of the most dependable and reliable ways to ascertain the veracity of a correct decision is to use a scientific method or approach to arrive at the decision. It is for this reason that research is needed in the corridors of policy science, learning, government, management and business domains. Research is the search for generalizable knowledge (George and Arthur, 2009). It helps an investigator to obtain a more authoritative and objective information on any phenomenon or fact that aroused his study and curiosity. Research is conceived as "the process of finding out and arriving at a dependable solution to a problem, through a planned and systematic collection, analysis and interpretation of data"(Odu, 2007, p.10). Research is driven by the desire to solve an existing social problem, satisfy intellectual curiosity, expand the frontier of disciplinary and professional knowledge, professional growth and career advancement, and to accumulate data for futuristic purposes (Agba, 2014). The starting point of a research is when there is an interest to know something about a particular thing. It could also be conducted with the interest in gaining an in-depth knowledge and expanding our understanding on a phenomenon of interest. During research, a researcher has the opportunity to obtain relevant information and acquire the requisite knowledge needed to solve an

existing problem or make concrete decisions, which may have organisational, societal, national and global impact especially in a world that is highly globalized.

Research is a scientific tool for the advancement of professional/disciplinary knowledge and to solve the problems of the society. It empowers man to relate more effectively with his environment, accomplish his dreams and resolve social conflicts as they arise. According to Ndiyo (2015), the problems affecting man and society demands scientific research to facilitate the process of finding solution to the existing problems. Research is oriented towards the discovery of the relationship that exist among the phenomena of the world in which we live. It is a systematic, organized and controlled attempt to observe, describe, explain, examine and understand the world surrounding us. Scientific research is devoted to finding the conditions under which certain phenomena occurs and the conditions under which it does not occur in what might appear to be similar circumstances. The driving force behind research is the existence of a problem and the urge to solve it. Hence, without a problem, there will be no research (Ndiyo, 2015, p.3). Research is very essential in the acquisition of knowledge because, it brings about trust; which is usually gotten in the research process.

There are two categories of research: Pure (basic) or fundamental research and applied research. The pure or basic or fundamental research is that which is carried out to discover new truth, propound new theories, advance knowledge and develop general principles; while, Applied or Practical research is that which is carried out to solve or provide solutions to an existing problem, i.e., practical solutions to existing problems. In other words, it is the desire to know for the sake of being able to do something better. Applied research involves the application of pre- existing knowledge in solving real world problems. Basic and Applied research are inter – related. Both of them are very useful in the scientific world and they contribute to the development of knowledge. For instance, in Basic Research, theories are produced or postulated; while in Applied Research, the usefulness of those theories are tested and applied in practical terms or situations. The tested theories are later provided as feedback to the theory postulators or basic researchers, to enable them modify or refine those theories. The general doctrine and the enduring principles for conducting scientific investigation anchors on either the basic or applied research. But, in any study involving the application of any of the two categories of research explained above, the use of data as its fundamental tool for carrying out that study is *sine qua non* (meaning, it is absolutely essential or it is indispensable).

According to Kelly and Phillip (1986) the foundation of any research or study in modern science is data- the vital information we collect, manipulate, analyze and establish as facts to substantiate our investigation. Odu (2007, p.40) emphasized that, "a research problem can only be solved on the basis of data; and so, when carrying out a study, the onus is on the researcher to set up a research design that is capable of providing the needed data that can lead to solving the research problem. This is because, without data, the researcher has nothing to write about". This implies that, the entire process of modern scientific research revolves around data. Hence, the "backbone" of every research is data. However, the application of a well-defined procedure and the maintenance of integrity during data collection and management in research, determines the credibility of any research. This is what informs the focus of this paper.

Meaning of Data

Different fields of study or disciplines may have their various definitions or explanations of what constitute data in research and how it can be managed. In general, data simply means the information that we collect, store and process systematically, in order to meet the required objectives of a specific research work. They are the vital information that we collect, manipulate, analyze and establish as facts when carrying out a study on a particular area of interest (Kelly & Philips, 1986).

Putting it in statistical terms, Mario (2006, p.4) defined data as "those observations (such as measurements, genders, survey responses, etc.) that have been collected". Ndiyo (2015, p.204) further explained that, "statistical data are those enumerable (measurable, observable) characteristics of variables. Variables are quantities that are capable of assuming any form of specified set of values. Scientists tend to call the properties they study 'variables' (Ndiyo, 2015, p.114)". Zach and Andrian (2014) described data as those pieces of valuable or relevant information that we gather, store, and use when carrying out a particular study. To be more explicit and lucid, data simply refers to those facts, figures and items that describe, inform or represent ideas, objects, situations or conditions, that can be gathered for analysis (NRC, 2002, as cited in Vanessa, 2012, p.2). According to Henning (2017), data is nothing other than another word for information.

Meaning of Research Data and Types of Data

Research data are those set of useful information that a researcher obtains from credible sources that are relevant to his study. They could be in form of statistics, diagrams, equations, textual or numeric information, audio or visual recordings, instrumental readouts, text, semi-text, tables, graphs, charts, models etc. The relevant information that we gather when embarking on a particular study constitute what is called data in research.

Generally, there are two types of data, which are usually categorized into: Qualitative data or attributes and Quantitative data or variables.

(a) Qualitative Data or Attributes

These are the type of data that are sorted in groups or separated into different categories, with each possessing some characteristics or attributes that cannot be expressed in numeric form. That is, they possess non-numerical characteristics. For example: eye colours (green, brown, white, red); gender (male or female); marital status (single, married, divorced); interest, love, religion, intelligence, honesty, etc. They are said to be qualitative if they only possess observable characteristics called attributes (Ndiyo, 2015, p.205). In addition, Odu (2007, p.43) further explained that, "qualitative data are also used to express some sensitive important issues in the society, such as beliefs, attitudes to infertility, abortion, female circumcision, etc".

(b) Quantitative Data or Variables

Quantitative data are those data that possess some characteristics that can be measured or counted, i.e. the data sets consists of numbers. According to Odu (2007, p.44) "the characteristics of the items in quantitative data vary, and as such called variable or a variate". Examples of quantitative data include: Age, Intelligent Quotient (IQ), Height, Weight, Income, Distance, Number of Children, Examination Marks, etc.

Mario (2006, p. 5-6) opined that, "quantitative data can be further divided into Discrete data and Continuous data".

(i) Discrete Data or Variables are the types that result when the number of possible values is either a finite number or "a countable" number; i.e. it can only take on countable (integral) values such as 0, or 1 or 2 and so on. For example, the number of eggs that hens lay, age, last birthday, number of children in a family, number of goals scored in a football match, etc.

(ii) Continuous (Numerical) Data or Variables are those that result from infinitely many possible values that are said to be on a continuous scale, covering a range of values without gaps or interruptions. They are measurements that can assume any of all possible values within a specified interval or on a continuous span. For example: weight, length, distance, height, the amounts of milk from cows, etc.

Still on quantitative data or variables, Ndiyo (2015, p. 204-206) further added by explaining that, quantitative data also include: Random variables, Dependent variables, Independent variables and Dummy variables.

(iii) Random Variables or Data are those types of variables that, whatever value they assume at any point in time cannot be predetermined or known in advance. For example, the face that will show up on tossing a fair dice is a random variable.

(iv) Dependent Variables or Data are known to be response variables or the ones to be predicted by others. For instance, in an experiment, the variable that is measured to determine the effect of an experiment treatment is usually referred to as the dependent variable. It is usually denoted by Y and positioned on the left hand side of an equation.

(v) Independent Variables or Data. The variable to be manipulated in an experiment is often referred to as the independent variable. Independent variables are the predictor variables. They are usually denoted by X and are placed on the right hand side of an equation.

(vi) Dummy Variables or Data. For dummy variables, they are used to distinguish among several qualitative characteristics such as: educational level (literate or illiterate); occupation (lawyer or non-lawyer); ethnic group (Kalabari or non Kalabari); religion (Christian or Muslim); etc. In a case where there are more than two possible characteristics, several variables can be used. For instance, a dummy variable like religion (e.g. Christian, Muslim, Pagan and atheist), which can be represented by dummy variables such as R1, R2, R3, and R4. R1 may assign 1 standing for Christian and 0 for non –Christian; R2 assigns 1 standing for Muslim and 0 for non-Muslim; R3 assigns 1 standing for Pagan and 0 for non-pagan; while, R4 assigns 1 standing for atheist and 0 standing for non-atheist.

In some instances, data that are qualitative in nature can be put into quantitative forms. This is done by first of all classifying them by attributes and then later giving them ranks to the various classes, or counting the number of cases that belong to each class. For instance, individuals can be classified by various Blood Groups (A, B, AB and O), marital status, utility, and level of education.

It is very important to explain here that, the distinction between qualitative and quantitative data is usually made on the basis of process instead of the properties that are inherent in the phenomena. The reason is because, properties that are generally considered qualitative can be made quantitative by measuring them with an instrument designed to assign numerical values to the various degrees to which they exist (Ndiyo, 2015, p. 204). Odu (2007, p.44) also added by explaining that, "the distribution between qualitative factors (attributes) and quantitative factors (variables) is important because, variables lend themselves to mathematical treatment in the form of averages; while attributes do not".

Sources of Data

Basically, there are two main sources of data. They include: primary sources and secondary sources, which produces primary data and secondary data.

(a) Primary Sources of Data

According to Orji (2009), as cited in Adejoh (2013) primary sources of data refers to data source that contains direct accounts that are obtainable observation, direct participation and questionnaires. Ndiyo (2015, p. 209) opined that "primary data are those statistical information that an investigator collects for his own purpose. It refers to those data obtained directly by the researcher through personal interviews, questionnaires and direct observations". When a researcher decides to get the information he needs by first hand, then, the data so obtained are called primary data (Odu, 2007, p.40).

Ndiyo (2015, p.210-211) further explained that, "primary sources of data can be divided into Internal sources (personal documents, individual units of enquiry, etc); and External sources (observations, questionnaires, interviews, etc.)". Odu (2007, p. 41) averred that, "the advantages of primary sources of data are that the researcher has confidence in the data collected and the results are usually comprehensive, factual and original; while the major disadvantages of obtaining primary sources of data is the huge financial costs involved in field trips for personal observations, interviews and questionnaire distribution and collection".

(b) Secondary Sources of Data

If an investigator uses data that were not originally collected by him in his research work, such as data from published sources, then he is considered to be using secondary data. In other words, secondary data are those data that are obtained from various publications and documents of many organizations, especially offices of statistics and financial institutions (Ndiyo, 2015, p.213). Most published statistical data are categorized as secondary data. Secondary data are data obtained from critical studies of literatures from various publications and documents (Orji, 2009, as cited in Adejoh, 2013).

According to Ndiyo (2015, p.213), secondary sources of data may be divided into three classes which include: Regular data, Periodical data and Irregular data.

i. Regular Data- are statistical data published on daily, weekly, monthly, quarterly, or annual basis e.g. market prices, meteorological data, stock exchange transactions, etc.
ii. Periodical data- which are data that are published at long intervals e.g. population census.
iii. Irregular Data- are those that consist of statistical data published at irregular intervals e.g. reports from various committees and commission set up by the government.

Depending on the type of materials that they are needed for and the dimension of the research problem under investigation, some of the areas that may serve as major sources for secondary data include: Public health, Education, Population, Social Events, Economics, Business, Vital Events, Government Publications, University Press, Articles, Magazines, Newspapers, Insurance Companies, Transportation, Agriculture, Migration (immigration and emigration), External Trade, Works of Research Bodies, etc. Some data are collected for a particular purpose, but are later used for a different purpose, for reasons of statistical enquiry. Hence, they are regarded as secondary data. According to Odu (2007, p.43), "secondary data are very useful for literature reviews. They also provide theoretical frameworks for empirical studies".

The advantages of secondary sources of data are that they are: readily available, very cheap and conveniently used as existing records; and can be retrieved easily from the places they are stored; while the disadvantages are that: the information provided may not be available in a form that is suitable for the research problem under investigation; precision is also difficult to determine because the method of collecting the data is not usually known, and may be inaccessible (Ndiyo, 2015, p.215-216).

What is Integrity?

According to George and Arthur (2009), integrity is the strict and uncompromising adherence to strong moral principles, ethical values, standards, honesty, transparency and total avoidance of deception.

Data Integrity

Data integrity is the truthfulness, honesty and reliability of data, having the validity and sustainability of its characteristics in place throughout the period the data set is being used. It is also the trustworthiness and consistency of data throughout its life cycle (Henning, 2017). Data integrity is

also considered as good data quality, which simply means that the data collected is accurate, dependable and reliable. It implies maintaining and taking steps of ensuring the consistency and accuracy of data over its lifecycle. Regoniel (2012) averred that, data integrity is the accuracy and reliability of a well-structured data, which is consistent with its attributes over a period of time. It involves assuring the accuracy, consistency and reliability of data over its entire life cycle. According to the World Health Organization (WHO), "data integrity is the degree to which a collection of data is complete, consistent and accurate throughout its lifecycle. A complete, consistent and accurate data should be Attributable, Legible, Contemporaneously recorded, Original or true copy and Accurate (ALCOA), (WHO, 2016, Technical Report Series, Annex 05, p.996)". In essence, what constitute data integrity is the completeness, accuracy, reliability, and consistency of its characteristics, all through the period of its life cycle.

There are two dimensions in which data integrity can be viewed. Data integrity can be viewed as 'a process' and as 'a state'. Data integrity as 'a process', explains or describes the measures taken or used to ensure accuracy, validity and consistency of a data set in research; while, data integrity as 'a state', defines a set of data that is considered to be valid and accurate. This paper tilts to the first or initial view, which sees data integrity as a process that describes measures taken to make sure that the data so obtained and used in research are valid, accurate and consistent. The process employs the use of Error checking, Quality Assurance (QA), Quality Control (QC) and Validation methods during data collection and management to ensure integrity.

Characteristics of Good Quality Data

According to Nuno, Seyma and Jorge (2015) the degree to which data characteristics fulfills the requirements of its purpose (i.e. the quality of a set data) has a tremendous impact on the outcome of a study. Below are some characteristics of a good quality data:

(i) **Accuracy:** the degree to which the attributes of a data represents the true value of the intended object correctly. A good quality data must be correct, precise and reliable.

(ii) **Accessibility.** This the degree to which a data set can be accessed and would give a suitability representation. A quality good data must be accessible, retrievable and easy to understand.

(iii) **Consistency:** the degree to which a data set is presented in the same unit, format and is compatible with other data sets that are similar.

(iv) **Validity:** the extent to which the fields entered in a data set are correct and are within the acceptable range. E.g. 18/02/2021 = valid date.

(v) **Completeness.** This is the degree to which a data set has all the values that can be described as expected attributes. That is, the necessary attributes of the data are all present.

(vi) **Timely:** when a data set is up to date and is also available for immediate use.

(vii) **Currency.** This is the degree to which a data set holds the attributes of the right age.

Characteristics of Bad or Poor Quality Data

Bad or poor data quality consist of data that is opposite the characteristics of good quality data (i.e. inaccurate, inaccessible, non- consistent, invalid, incomplete, outdated and belated. It is the degree to which the characteristics of a data set does not fulfill the requirements of its purpose (Nuno, Seyma and Jorge, 2015). Such set of data could have wrong numbers, misspelled names, duplicate and outdated information.

Differences between Good Quality Data and Bad Quality Data.
The illustrations given below shows the difference between good and poor data quality.

(a) Good Quality Data
The illustrations given below are typical examples a good quality data that were obtained from a study in Brazil. Table 1 below shows, in absolute terms, the frequency of acne scars in 18- year old youngsters from a population- based study conducted in the city of Pelotas, Southern Brazil in 2010.

TABLE 1: Absolute and Relative Frequencies of Acne Scar in 18- year-old Adolescents (n = 2,414). Pelotas, Brazil, 2010

Prevalence	Absolute Frequency (n)	Relative Frequency (%)
No	1,855	76.84
Yes	559	23.16
Total	**2,414**	**100.00**

Source: Presenting data in tables and charts. *An Bras Dermatol*, 89(2): 280-285. April, 2014. Doi: 10.1590/abd1806-4841.20143388

The information shown in Table 1 above can also be presented in form of a Bar Chart and a Pie Chart, to further illustrate the Absolute Frequencies and Relative Frequencies of the categories respectively (Rodrigo, Joao, Renan, David & Jeovany, 2014). See the bar and pie charts below:

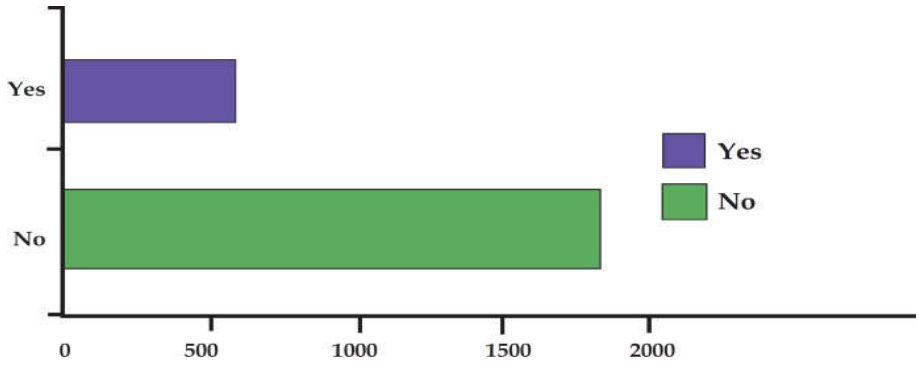

Figure 1: A Bar Chart showing the Absolute Frequencies of acne scar in 18-year-old Adolescents (n = 2,414). Pelotas, Brazil, 2010
Source: Presenting data in tables and charts. *An Bras Dermatol*, 89(2): 280-285. April, 2014. Doi: 10.1590/abd1806-4841.20143388

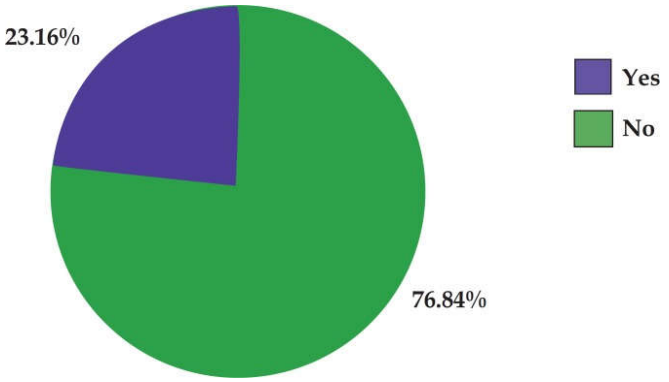

Figure 2: A Pie Chart showing the Relative Frequencies of acne scar in 18-year-old Adolescents (n = 2,414). Pelotas, Brazil, 2010
Source: Presenting data in tables and charts. *An Bras Dermatol*, 89(2): 280-285. April, 2014. Doi: 10.1590/abd1806-4841.20143388

The second example in Table 2 illustrates the distribution of educational levels among 18-year-old youngsters from a population- based study conducted in the City of Pelotas, Southern Brazil in 2010, with Absolute, Relative, and Cumulative Relative Frequencies.

TABLE 2: Absolute, Relative, and Cumulative Relative Frequencies of the Educational Levels of 18-year-old Adolescents (n = 2,199). Pelotas, Brazil, 2010.

Educational Level (in years of education)	Absolute frequency (n)	Relative frequency	Cumulative relative frequency (%)
0	1	0.05	0.05
1	2	0.09	0.14
2	2	0.09	0.23
3	11	0.05	0.73
4	100	4.55	5.28
5	156	7.09	12.37
6	169	7.69	20.05
7	221	10.05	30.10
8	450	20.46	50.57
9	251	11.41	61.98
10	320	14.55	76.53
11	479	21.78	98.32
12	31	1.41	99.73
13	6	0.27	100.00
Total	**2,199**	**100.00**	**-**

Source: Presenting data in tables and charts. *An Bras Dermatol*, 89(2): 280-285. April, 2014. Doi: 10.1590/abd1806-4841.20143388

It is clear from the examples that the quality of the above data characteristics fulfills the requirements of its purpose.

(b) Bad or Poor Quality Data

(i) The table below is an example of poor quality data presentation

Region	% adults taking a holiday
East Anglia	50
East Midlands	64
Greater London	56
Humberside and Yorkshire	64
North	54
North West	59
South east	60
South west	61
West midlands	56

The data presented in the above table is bad or poor quality because:

1. The table lacks a title
2. The source of the information is not provided
3. Row titles straddle two lines
4. Each cell is bounded as if in a spreadsheet
5. The alphabetical listing of regions results in a non-numerical ordering of data down the columns

(ii) The table below is another example of poor quality data presentation with symbols within the cells, considered as footnotes.

The Scores

	Maths	English	History	Government #	Science
Class 1	32	43	34	32	65
Class 2	53	76	44	21	91 @
Class 3	87	44	33	36	58
Class 4	38	52	46	27 $	17
Class 5	76	68	93	12	11
Class 6	36	40	39	48	24
Class 7	23	34	43	33	56
Class 8	53	76	45	22 *	91
Class 9	78	45	36	33	85
Class 10	83	25	64	72	71
Class 11	67	86	39	21	12
Class 12	63	50	93	84	42

Footnote:

Numbers procured from the Central Government Resource room
*The I have checked 4 times and my colleague have checked 5 times
$ Quoted in the National Scholastic Journal Volume 2
@ They are included for effectiveness

The data presentation made in the second table above is also of poor quality because:

1. The table lacks a clearly defined title
2. The source of the information is not known
3. The symbols in the table do not have a well-defined heading
4. Some cells in the table have symbols that are considered as footnotes.

Data Integrity in Research

Data integrity in research simply refers to the level of validity, honesty and consistency of a data set in research. It defines the dimension in which the accuracy of a set data contributes to the rigour or trustworthiness of a research. It also describes the level of confidence, reliability and trust invested in the information or data obtained and used in a particular study. Data integrity in research is the foundation for research credibility. It is essential in every area of scientific research and in all categories of investigation. In research, it is commonly said that, "no research is better than the quality of its data". Therefore, "quality data is key to a good research". This implies that, ensuring accurate, valid, and honest data collection in research is important for maintaining the integrity and quality of the research. When the data obtained in a study is faulty, corrupt, deceptive and compromised, it affects negatively on the nature of the investigation and influences the outcome of the research. That was why Krishan (2016) contended that, the quality of data that a researcher observes, identifies and collects for his study determines the level of integrity and credibility of his research. Therefore, data integrity is key, if the credibility of research is to be achieved. This paper is opining that, observing truthful and reliable measures that will enhance data integrity in research is a rudimentary requirement for a good research. In other words, data integrity is the basic fundamental principle for research credibility.

The preliminary measures usually observed to ensure validity, reliability and integrity during data collection in research include: Quality Assurance (QA), Quality Control (QC), Error Checking, Validation methods, etc., depending on the kind of research. The Quality Assurance and Quality Control measures are often applied at different stages involving the collection and use of the data in research. For example, QA activities are usually observed before data collection. QA employs the establishment of Standard Operating Procedures (SOPs), Quality Systems, Standardization of Protocols, Training of Personnel, etc., when collecting data for research, so as to instill transparency and integrity in the process. While QC activities often take place after data collection. It involves monitoring, detecting and correcting any error or actions related to it when the research data has been collected. Also, the World Health Organization (WHO) prescribed the use of the ALCOA principles (ALCOA is an acronym, which simply means Attributable, Legible, Contemporaneous, Original and Accurate data) as research guidelines to ensure data integrity in some regulatory bodies. It recommended that, "observing the ALCOA principles in a study, validates the attributes of the data used during investigation and ensures that the information or data collected is complete, consistent, enduring and available throughout the required period that it is retained for usage (WHO Guideline on Data Integrity, 2020, p.6)". Ahmed, Kumar and Hafeez (2019, p.307) explained that, "the use of basic ALCOA principles to maintain data integrity in pharmaceuticals, research and manufacturing, serves as evidence to show that the quality of data collected is in compliance with a defined regulatory guideline".

The two main ways to secure data and observe integrity in research are:

(a) Data Collection and (b) Data Management.

(a) Data Collection

According to Chukwuemeka (2006) data collection can be described as a process that involves the assembling, gathering and measuring information on variables of interest, in a pattern that is systematic and well defined, which enables an investigator to give answers to the stated research questions under investigation, test the hypotheses and evaluate the outcome of his research. Denga and Ali (1998, p. 168) pointed out that, "data collection is a very important stage of any study; hence, during the process of data collection, it is advisable to make sure that the data collection instrument is carefully selected or constructed in such a way that it can be used to obtain valid, standardized and relevant information from all the subjects that constitute the sample". For purpose of emphasis, it is

very expedient to note that the quality of data collected to a large extent determines the credibility of the research outcome and the mode of the statistical technique used, depends also on the nature of the data (Odu, 2007, p.49).

It is important to note that, the components for data collection in research may be similar in most fields of study such as Social Sciences, Management Sciences, Humanities, Physical Sciences, Engineering, etc.; but the methods differ by the discipline. However, the point of emphasis here is that a researcher should ensure that the data he is collecting for his study is honest, accurate and consistent, so as to have a credible research outcome. The reason is because, the level of acceptance and reliability of the data collected, largely depends on the soundness of the method used and the measures taken to ensure a valid and standard data collection (Odu, 2007, p.41). Therefore, the validity of data collected during research, determines the quality of a research outcome. In research, different forms of data are usually collected using different forms of data collection tools or instruments. The type of data needed for a research determines the kind of research instrument or tool to be used for data collection. This also influences the type of research method and statistical technique used for data analysis in the study in question. In Social Sciences and Humanities, some data collection instruments or tools used in research include: the Questionnaire, Observation, Interviews, Rating Scale, Attitude Scale and Interest inventories.

(i) The Questionnaire- is any research tool comprising carefully designed questions for the respondents to answer. Basically, there two types of questionnaires: the structured or closed form questionnaire and the open ended questionnaire (Denga and Ali 1998, p. 169). The structured or closed form questionnaire contains already written items, for which respondents are to only fill the blanks spaces with a tick, one word or phrase or sentence; while the open-ended questionnaire contains direct questions for which enough space is provided for the respondents to write their answers.

(ii) Observation- is a direct means of studying the overt behaviour of people. It provides an opportunity for a researcher to have a direct record of a particular behaviour the way it happens. Observation is a primary instrument that is very important in scientific research. Goode and Hatt (1982) opined that, “science begins with observation for its final validation”.

(iii) Interviews- involves collecting information from others in a face to face contact. It produces data through verbal communication between the researcher and the respondent.

(iv) Rating scales- are used to discriminate human behaviour into levels or categories to give an appropriate indication of where an observer belongs. For instance, to scale the value of honesty, a researcher may ask people to rate a person’s level of honesty into a 5- point scale: very honest, honest, neutral, dishonest, very dishonest.

(v) Attitude scale. An attitude scale is the tendency of an individual to think or act in a particular way under a given circumstance. It is a response pattern that constitute an individual’s inclinations, feelings, ideas, fears, preconceived notions, threats and convictions on the basis of experience. The instrument used to measure attitude is the Likert Scale. E.g. Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), and Strongly Disagree (SD).

(vi) Interest inventories- is an instrument for measuring a person’s like and aversions (Odu, 2007, p.70). That is, the likes and dislikes of a person.

(b) Data Management

Kenneth (2002, p.191-205) defined data management as "the administrative process which ensures that the needed data is acquired, scrutinized, stored, guided, processed and presented in such a way that it meets the expectations of data users, in terms of reliability, accessibility and timeliness. It is the neglected, but a very important twin to scientific methods". In the real sense of it, data management covers almost all stages of research; starting from conceptualization to accomplishment of the research work; as well as the disposal of research materials. Zach and Andrian (2014) emphasized that "responsible and effective data management helps to improve the integrity of research, thereby preventing gaps, errors and inaccuracies. The way in which data is managed has an impact on the legitimacy and credibility of a research or scholarly record".

The type of processes used by different fields of study and the variability of their research methods makes it impossible to adopt a universally recommended form of data management. However, basically, there are six areas or stages of data management. They include:

(i) Data Selection - which involves choosing the observable facts or data or information that suits the variables of interest in the research.

(ii) Data Gathering - the assemblage of data and measurement of information that are relevant to the study.

(iii) Data Processing- the application of those procedures, techniques and methods that are used to manipulate or convert raw information (data) into a form that is suitable for analysis and interpretation (Ndiyo, 2015, p.222).

(iv) Data Analysis- the ordering and breaking down of data into constituent parts, which may involve the performance of statistical calculations using the raw data, so as to provide answers to the research questions (Odu, 2007, p.130). It involves a meticulous and careful review, breaking down in details and giving explanations on critical parameters that constitute the data.

(v) Data Presentation/Reporting and Publishing- which involves giving general descriptive analysis, showing systematic illustrations, making inferences and drawing conclusions based on the findings of the study and presenting it either in the form of text, semi-text, tables, graphs, charts, or models etc.

(vi) Data Storage and Ownership- which has to do with the preservation of the captured data in a special device or medium that is considered safe for use or for future referencing and having a complete legal right and control over the data so acquired.

The primary rationale for data integrity in research is to support research credibility, legitimacy, and validity. Responsible data collection and management are the rudimentary principles that can compel a genuine pursuit for a study of interest and can also create an atmosphere of trust and reliability on a research outcome. Like it is commonly said "nothing good comes easy"! Although managing data is not an easy thing to do, but, for a research to be legitimately good, valid and credible data must be well managed.

Reasons why Data Integrity in Research is a Fundamental Requirement for Research Credibility

Data integrity is one of the processes that forms the rudimentary structure of a good research. It helps an investigator to think, be more organized and well-coordinated when embarking on his study. It is a fundamental requirement in research because:

(a) It addresses issues of data validity, which leads to research credibility and reliability.

(b) It ensures that the data produced or obtained are raw, undistorted, verified and complete.

(c) It provides opportunities for comprehensive documentation and data preservation.

(d) It serves as a source for reproducibility of research components when needed and for easy accessibility.
(e) It ensures the use of correct parameters for a particular study.
(f) It provides a good template for data analysis in research.
(g) It brings about data security, thereby preventing data loss.
(h) It informs the researcher to be more diligent, meticulous, detailed, zealous, enthusiastic and passionate about his study.
(i) It brings about stable performance in research, enables the proper testing of hypotheses, good analyses of results, provision of answers to complex research questions and the evaluation of research outcome.
(j) Data integrity in research ensures reliability of the data obtained.

Since empirical studies are concluded with recommendations flowing from the data used or gathered. In addition, the translation of the recommendations of the study into policy options/decisions and programme of actions for public and private consumptions implies that the data used must possess the integrity elements.

Research Credibility

Credibility can be described as having the quality to inspire belief or trust. It is one of the criteria for trustworthiness. Credibility in a study refers to the rigour, honesty and transparency of a study conducted in compliance with scientific methods. Trustworthiness or rigour refers to the level or degree of confidence that can be invested on a substance or source of information or data. Research credibility simply describes the extent to which the outcome of a research conducted is trusted and relied on; as a result of the approaches used in the research, which guarantees conformity with the expected standards, ethics and procedures usually observed in scientific research. Credibility in research defines the level of confidence and trust that can be invested in the findings of a study. According to Irene and Moser (2018, p.121), "research credibility establishes whether the research findings represents plausible data drawn from the respondents' original information and is the true interpretation of the respondents' actual views in a study". It questions the level of authenticity of the source of information or data used in an investigation. That is why, when data sources are not reliable in a study, it makes the outcome to be misleading and deceitful. What makes a data source to be reliable or credible in research is the level of accuracy, currency, objectivity, authority and coverage of the data source.

The rigour of research is based on the degree of confidence imposed on the data used and the procedures followed to ensure integrity in the research. This implies that, data integrity is the anchor of credibility in research. Therefore, in any scientific investigation, the researcher is expected to show and establish that his study followed the lay down rules, protocols and procedures that ensured honesty, transparency and consistency of the data collected, managed and reported; in compliance with the highest standards expected of a scientific research. The question that follow after a study has been conducted is: did the research follow the required standard and procedures involving data collection to ensure integrity? An offence of "research misconduct" can be committed, if the required standard and ethics involving data collection procedure in the research process are ignored or abused. The reason is because, a credible research is expected to have followed the required standard, using the most recent, correct and reliable data, so that the claims made from the outcome of the research can be trusted, sound convincing and plausible to members of the public. That is why, when the integrity of data is lacking in a study, the outcome of that study will be disastrous to the scientists community.

Conclusion

The science of data integrity in research has emerged because, it is getting harder to conduct a transparent and credible research than to frame a proposal. The concerns raised by scientists all over the world about research integrity preempted the need for a critical look into the issue of data integrity in research; reasons being that, data has become one of the most valuable assets of any study and it is a constituent of the rudimentary structure of a good research work. Therefore, its accuracy, consistency, completeness and originality is very important because, it increases the chances of stability, confidence and reliability in the quality of a research. This dimension of study has proven to be that, the key to research validity and credibility anchors on data integrity. Hence, data integrity in research cannot be over emphasized because, it is very expedient; it is a constituent of the rudimentary structure of a reliable investigation and it is also the backbone of a good research. Although data integrity in research involves a great deal of time, efforts and cost, it should be encouraged among researchers because, without credible data, there can be no credible research. Trust and reliability can be invested into the outcome of a study, if it observes the fundamental rules, protocols and procedures for data collection and management, in relation to data integrity during the research process.

The academic community must emphasize the fundamental need for responsible data collection and management in research, so as to support research integrity. It is the opinion of this paper that, observing truthful and reliable measures that will enhance data integrity in research is a fundamental requirement for a credible research. Hence, scholars should inoculate and inculcate the culture of transparency in research into the minds of young researchers through data integrity.

References

Adejoh, E. (2013). An assessment of the implication of contributory pension scheme to Nigerian economic development. *Global journal of management and business research,* 13 (2) 1. Retrieved from: https://journalofbusiness.org/index.php/GJMBR/article/view/957.

Agba, M.S. (2014). *Fundamentals in Research Methodology in Social Sciences and Humanities*. Calabar: University of Calabar Press

Ahmad, S., Kumar, A., & Hafeez, A. (2019). Importance of data integrity and its regulation in the pharmaceutical industry. *The Pharma Innovation Journal*, 8 (1): 306-313, January, 2019. E-Retrieved from: www.thepharmajournal.com . Accessed September 29, 2020.

Chukwuemeka, E. O. (2006). *Research methods in thesis writing*. Enugu: HRV Publishers.

Denga, D. I. & Ali, A. A. (1998). *An introduction to research methods and statistics in education and social sciences*. Calabar: Rapid educational publishers.

George, M. B. & Arthur, E. K. (2009). *Ensuring the integrity, accessibility, and stewardship of research data in the digital age*. Washington, DC: National Academics Press.

Goode, W. J., & Hatt, P. K. (1982). *Methods in social research*. London: McGraw- Hill Kogakusha Limited.

Guidance on good data and record management practices. In: WHO Expert Committee on Specifications for Pharmaceutical Preparations: fiftieth report. Geneva: World Health Organization; 2016: Annex 5. *WHO Technical Report Series, No. 996*. Retrieved from: https://www.who.int/medicines/publications/pharmprep/WHO_TRS_996annex.05.pdf?ua=1. Accessed September 28, 2020.

Guideline on Data Integrity. World Health Organization (WHO) Draft Working Document for Comments: Working documents QAS/19.819/Rev.1, page 1-29, June, 2020. Retrieved from: http://www.who.int/medicines/areas/quality_safety/quality_assurance/guidelines/en/

Henning, L. (2017, February 15). Data integrity: what is it and why is it important? *RAPIDI: DATA INTEGRATION IS A PEOPLES'S BUSINESS.* Retrieved from: https://www.rapidionline.com/blog/data-integrity-what-and-why

Irene, K., & Moser, A. (2018). Series: Practical guidance to qualitative research. Part 4: Trustworthiness and Publishing. *European Journal of General Practice*. 24 (1), 120- 124. DOI: 10.1080/13814788.2017.1375092

Kelly, P. C., & Phillips, M. J. (1986). Better data for decision –making: Implications for the geosciences. In E. P. Shelley (Ed.), *Proceedings of the 3rd International Conference on Geoscience Information*: *Vol. 1 (pp. 51-61)*. Adelaide, SA: Australian Mineral Foundation.

Kenneth, P.D. (2002). Six domains of research ethics: A heuristic framework for responsible conduct of research. *Science and Engineering Ethics,* 8 (191-205). Doi: 10.1.1457.1918.pdf. http://www.poragen.co.uk

Krishan, K. (2016). Data integrity in pharmaceutical industry. *Journal of Analytical and Pharmaceutical Research (JAPR)*. 2 (6): 00040. DOI: 10.15406/japlr.2016.02.00040

Mario, F. T. (2006). *Elementary statistics (annotated instructor's edition).* New York: Pearson Education Inc.

Ndiyo, N. A. (2015). *Fundamentals of research in behavioural sciences and humanities*. Calabar: Excel publishers.

Nuno, L., Seyma, S., & Jorge, B. (2015). *A survey on data quality: Classifying poor data*. A Conference Paper presented at the 21st IEEE Rim International Symposium on Dependable Computing (PRDC, 2015). At: Zhangjiajie, China. Retrieved from: Doi:10.1109/PRDC.2015.41

Odu, E. N. (2007). *Introductory statistics and research methods in education and social sciences.* Calabar: Ojies Ojies production.

Regoniel, P.A. (2012, December 6). The importance of data accuracy and integrity for data analysis. *Simplyeducate.me.* Retrieved from: https://simplyeducate.me/2012/12/06/the-importance-of-data-accuracy-and-integrity-for-data-analysis/

Rodrigo, P. D., Joao, L. B., Renan, R. B., David, A. G., & Jeovany, M. (2014, April). Presenting data in tables and charts. *An Bras Dermatol*, 89(2): 280-285. Retrieved from: Doi: 10.1590/abd1806-4841.20143388

Vanessa, E. J. (2012). The role of information professionals in geoscience data management: A western Australian perspective. *Library and Information Science Research e-journal,* 22 (2). ISSN: 1058-6768. Retrieved from: https://www.libres-ejournal.info/508/ https://pdfs.semanticsscholar.org.

Zach, C. & Andrian H. (2014). Ensuring research integrity: The role of data management in current crises. *College & Research Libraries News: Association of College & Research Libraries*, 75 (11). P-ISSN: 0099-0086, E-ISSN: 2150-6698. Retrieved from, https://crln.acrl.org/index.php/crlnews/article/view/9224