Chapter Seventy – Five

Data Management Practices in Educational Research

Valentine. J. OWAN
Department of Educational Administration and Planning,
University of Calabar, Calabar.

Bassey Asuquo Bassey
Department of Educational Foundations,
University of Calabar, Calabar.

ABSTRACT

Data is very important in any research experiment because it occupies a central place in making decisions based on findings resulting from the analysis of such data. Given its central role, it follows that such an important asset as data, deserve effective management in order to protect the integrity and provide an opportunity for effective problem-solving. The main thrust of this paper was to examine data management practices that should be adopted by scholars in maintaining the quality of their research data. In achieving this, various concepts related to this paper were clarified. Various data management practices were also discussed beginning from data generation to data shredding. Based on the underlying observations from the light of the discussions made in this paper, it was recommended among others that: tertiary institutions in every part of the world should endeavour to establish a data management unit that will be saddled with the primary duty of formulating research data management policies, and the hosting of research data; every Journal should as a matter of compulsion, require the submission of data set corresponding to empirical papers submitted by authors and scholars.

Keywords: Data, Management, Management practices, Data management, Educational Research, Management Practices, Research.

Reference to this paper should be made as follows:

Owan, V. J. & Bassey, B. A. (2019). Data management practices in educational research. In P. N. Ololube & G. U. Nwiyi (Eds), Encyclopedia of institutional leadership, policy, and management: A handbook of research in honour of Professor Ozo-Mekuri Ndimele (pp. 1251 – 1265). Port Harcourt, NG: Pearl Publishers International Ltd.

INTRODUCTION

Central to any research exercise is the need to solve problems and reach meaningful conclusions based on the evidence available. This usually requires the collection, analysis, presentation, and interpretation of data. Data are raw facts that needs to be processed into a form understandable by man. Data are usually collected and transformed using various methods in order to make decisions,

create awareness, modify existing knowledge with new evidence. According to Durosaro (2008) data are numerical measures of phenomenon and they are used in the process of scientific decision-making. The role of data is central to any piece of research or research project. The research data is meticulously collected, analysed, optimized, organized and made usable to carry out research studies. Research cannot happen and would not be complete in any area of study without authentic and objective data.

When data are processed through analysis, and interpretation, they become information. Thus, the processed data (information) becomes very useful in making inferences and deductions. For instance, data collected from a census exercise conducted in say, Cross River State, may not make sense in its actual form when it is collected. It might make more sense if such data is expressed in terms of number of children, adolescents and/or adults through a simple frequency distribution. On the other hand, it will be possible to determine the total number of males or females living in that region if the collected data were analysed to show the total population based on sex.

In education, data are usually collected using a wide-range of techniques, and the results are often applied in solving educational problems. The importance of data in educational research cannot be overstated, it is the yardstick with which inferences are drawn especially in empirical investigations conducted within education-related disciplines. Thus, it is pertinent that data collected in educational research be effectively managed to protect the integrity of such facts, and to offer a platform of subsequent retrieval.

Data management plays an essential role in any growing organization like staffing, placement of students, recruitment which are aided by use of computer. Research data are the core of any research study. The findings and conclusions of a study are entirely dependent on the research data (Patel, 2016). Traditional publishing did not focus on the presentation of data, along with the publications such as research monographs and especially journal articles, probably because of the difficulties involved in managing the research data sets (Patel, 2016). The current day technology, however, has helped in making this task easier. Information technology has generated advanced tools for analysing and managing data. Use of those tools like computer diskette, microfilm, and microfiche, CD-ROMs, DVD, World Wide Web, Flash Disk, among others which can improve the performance of almost any operation. Steps made in capturing mass data electronically have developed the need for effective management strategies (Ezeagu, 2013). It was based on this background that the focus of this paper was streamlined to explore conceptually the practices for managing data effectively in education.

Clarification of concepts

- *Data*

Data are raw, unorganized facts gathered through observation, testing, interview, questionnaires, check list and rating scales, and that needs to be processed to make them more meaningful. Data in its initial stage of collection, could be meaningless. Data are processed into information for proper understanding of the data. According to Soler, Ort, and Steckel (2016) research data are considered all information collected, observed or created for purposes of analysis and validation of original research results. Data can be quantitative or qualitative and comprises also photos, objects or audio files, resulting from as different sources as field experiments, model outputs or satellite data.

- *Management*

Management is a deliberate and systematic process of unifying human and material resources with the aim of reaching set goals. It involves regulating the activities and operation of men, materials, machines, money and time in order to attain set objectives. According to Skina as cited in Ezeagu (2013), management is the process of forming a strategic vision, setting objectives, crafting a strategy and then implementing and executing the strategy. Management therefore is the coordination of all the resources of an organization through the processes of planning, organizing, directing and controlling in order to attain organizational objectives.

- *Data Management*

Data management is a conscious process of collecting, analysing, using, and preserving data with the aim of making decisions, reaching conclusions, creating new knowledge, adding to or modifying existing knowledge with new evidence, and providing a platform for easy retrieval and re-use in the future. It is the process whereby data are collected, sorted, coded, prepared, analysed (using appropriate statistical techniques), presented, interpreted, stored, and secured, for the purpose of using such data to solve practical problems, and for future purpose. To Ezeagu (2013), data management is the effective coordination of people, equipment and procedures together, to sort, analyze, evaluate and distribute needed, timely and accurate data for decision making. Data management is an administrative process that includes acquiring, validating, storing, protecting, and processing required data to ensure the accessibility, reliability, and timeliness of the data for its users (Ngdata as cited in Owan & Bassey, 2018).

- *Education*

Education is a process where someone acquires or modifies his or her knowledge, skills, attitudes, attributes, and competencies either through observation, experience, or from another person in the role of a teacher, from formal, informal or non-formal setting. The rationale behind education is to provide individuals with opportunities to become useful citizens and for societal development. According to Esu and Ntukidem (2015) education is a vehicle for social change and a powerful tool for social, economic, political and technological development of any country. It is not limited to formal schooling instruction but involves all life experiences.

- *Research*

Research is a problem-solving process of making decisions based on data systematically collected, presented, analysed and interpreted data, with the explicit aim of creating or improving man's knowledge. Walliman (2011) explained that research is a very general term for an activity that involves finding out, in a systematic way, things you did not know. A more academic interpretation is that research involves finding out about things that no-one else knew either. It is about advancing the frontiers of knowledge. Researches are conducted with one or more of the following objectives: to categorise, describe, explain, evaluate, compare, correlate, predict, and/or to control (Walliman, 2011)

As noted by Owan and Bassey (2018), the goal of every research is to solve a unique problem within the environment or to contribute to already existing knowledge by filling gaps that may exist in the literature. Research reaches its end through proper decisions that are made through inferences and deductions. Inferences are best made if there are empirical evidence to justify such decisions (Owan & Bassey, 2018). Research can be pure or applied depending on the outcome of such study. A research is pure if it was conducted to create, establish or test a theory; while applied researches are carried out to solve problems with a base on theoretical frameworks.

- *Educational Research*

Educational research is an aspect of research where studies are conducted in education-related disciplines to create new knowledge, modify existing knowledge that are obsolete, and for solving educational problems. Like other disciplines, education also relies on the findings from empirical investigations and a basis for testing educational theories, and for making decisions. Researches in education may be pure or applied depending on the goal of the researcher. If the researcher is attempting to develop an educational theory, or modify an existing theory, it is pure, basic or fundamental; on the hand, if the intent is to solve educational problems with the research results, it is applied research in education.

DATA MANAGEMENT PRACTICES IN EDUCATIONAL RESEARCH

Data in educational research can be effectively managed through several practices as shown in Figure 1 and discussed accordingly below

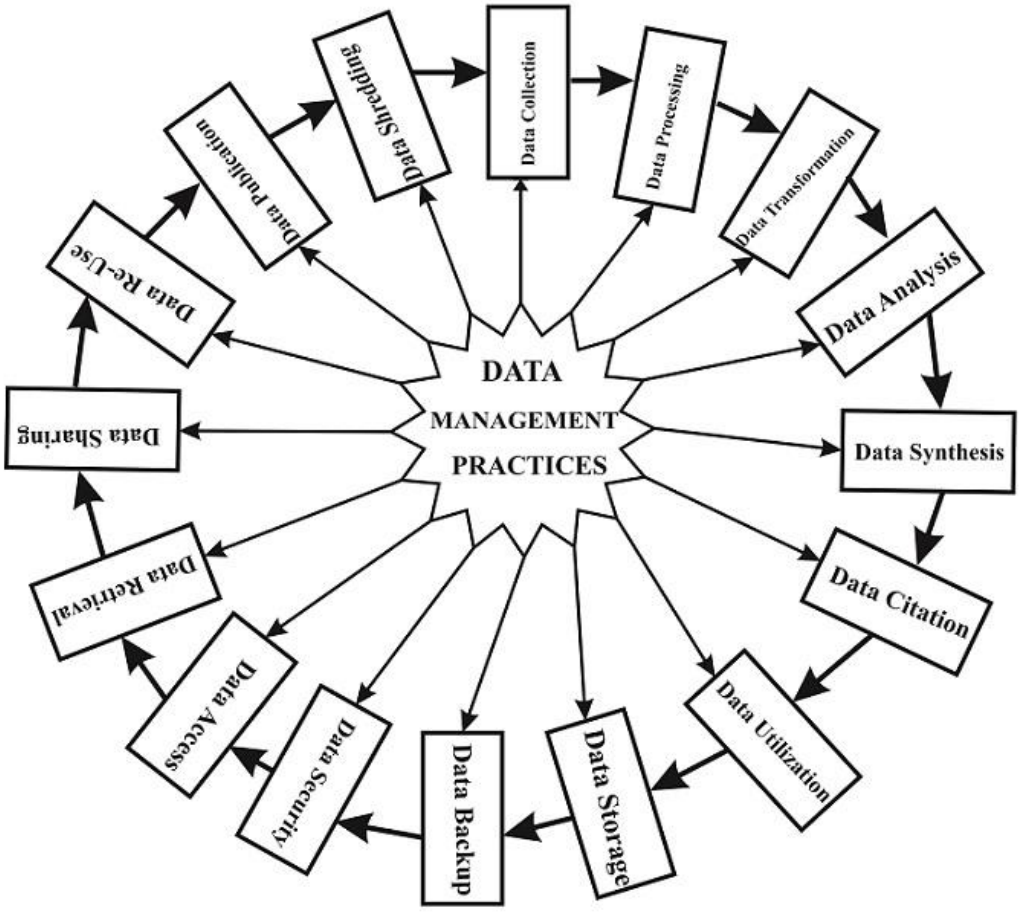

Figure 1: Data Management Cycle

- *Data collection/generation*

The generation or generation of data in educational research is as important as the research process itself. Educational research data needs to be generated, using appropriate instruments and methods. Proper generation of data provides a platform for validity of the findings of studies in education. Data

collected erroneously will result in misleading research results. In the process of collecting data, the researcher and/or his team must be equipped with the skills and knowledge of such data collection, the respondents or the phenomena of study must also ensure to provide details that are true. In the case of a phenomena, the data must be collected at a time when the behaviour desired to be observed is manifested.

Data collection refers not only to what information is recorded and how it is the recorded, but also to how a research project is designed. Data collection may seem tedious or repetitive, but the data produced must be reliable and valid to ultimately prove or disprove hypotheses and justify or counter a body of research (Pradal, Akullah, Bazarova, Devare & Gardiner, 2017). According to Soler et al (2016) data can be collected in many ways. Collection includes various procedures such as manual recordings of observations in the laboratory or field on hand-written data sheets as well as automated collection by data loggers, satellites or airborne platforms (Michener & Jones 2012).

- *Data processing/conversion*

Data processing is the second phase of the data management life cycle that takes a primary role in converting the data collected in the first stage of the life cycle to meaningful information. When data is collected, it may not be in a readily usable form. The process starts with discovering inconsistencies and other anomalies in the data into raw data, as well as data cleansing to improve the data quality. Users could then conduct analyses to produce meaningful information based on the data that may lead to a resolution of a problem or improvement of an existing situation (Miller, Miller, Moran, & Dai, 2018). Data conversion refers to the process of converting data into information or the process where textual contents in research instruments are converted into their assigned numeric values/codes in order to facilitate analysis.

Data processing or conversion enables the researcher to carefully sort out relevant data from irrelevant ones. When data are collected, they are raw and meaningless, consequently, they need to be processed into a form that meaning could be derived. Educational research data needs to be prepared and coded using a wide-range of techniques for easy interpretations, inferences and deductions to be made. Depending on the type of data obtained, quantitative and/or qualitative techniques may be employed in the processing of such data. For example, if quantitative data are collected, appropriate statistical methods may be used, and the choice of statistics depends on a lot of factors including: level of measurement (nominal, ordinal interval and ratio), purpose of the study (describe, explain, compare, correlate, predict, etc.) , instrumentation (test, questionnaire, rating scale, check list etc.), type of data collected (discrete or continuous), and so on.

Before processing or converting research data, it is pertinent to ensure that the collected data are without errors. Soler et al (2016) warned that; quality checks should be performed during collection, data entry and analysis and answer the following questions: are the data complete? are the data correct? is the format consistent throughout the data set? if it contains errors, which errors? are there missing values? In the validation process it should be checked whether data are incomplete, unreasonable or inaccurate. This can already be included in the data entry process. Statistical and graphical summaries (e.g. max/min, average, range) help to check for impossible values and outliers. After validation, the data set is cleaned. This means to check outliers, correct and fix errors.

Soler et al (2016) opined that; with technological innovations, computer programs are also used in processing data. Furthermore, spreadsheet software (like Excel) is very common for working with quantitative data, especially during data collection and entry. Its advantages are that many people already know working with it and first steps are quite simple. However, special care should be taken when working with spreadsheets as they can be quite error-prone.

- *Data transformation*

Data transformation occurs in a situation where a researcher changes the data used by an earlier study to suit the purpose of his /her study. Qualitative data could be transformed through paraphrasing an earlier data, or by analysing previous data to suit the context of another study. Quantitative data on the other hand could be transformed by switching the statistical method used, or by manipulating the sample size of a later study. Sometimes digital research data gets transformed due to the changes occurring in software versions as data are migrated from older versions of a software to newer versions or vice versa.

Migration refers to the transformation of data or other digital material from one format or technology to another (software or hardware). The ability to retrieve, display and use the contents is maintained by this action (DCC 2008b). Files can be migrated within one software product to a newer version or to other file formats when obsolescence occurs. Every migration changes data a little bit. If migrations are carried out multiple times, data can be subject to major changes (Soler et al, 2016). Soler et al (2016) also warned that; because functionality is lost, and integrity comprised as a result of migration, care must be taken to strict quality checking procedures, for example, to compare the original bit stream and the migrated bit stream. To check data integrity and detect bit errors, checksums can be used. A checksum algorithm calculates a value which can be compared to the value of the original file. A widely used algorithm is MD5.

- *Data analysis*

Data analysis is a technique in data management that involves breakdown the converted data into rudiments such that each component is understood respectively. Data analysis enables the researcher to understand the hidden truths behind the collected data. Once data is analysed, it gives room for decisions to be taken. In analysing data, the objective of the researcher must be given consideration. Just as in data conversion, data analysis can be carried out quantitatively, qualitatively or a mixture of both quantitative and qualitative techniques. According to Soler et al (2016) analysis comprises the actions used to derive and understand information from data. The types of analyses depend on the discipline and on the research questions to be answered. Furthermore, the software and hardware used to analyse data also vary (Soler et al, 2016).

Quantitative methods require that collected data be converted into numerical codes. Descriptive and/or inferential statistics can be performed on such numeric data for decisions to be taken. Soler et al (2016) disclosed that Statistics are one of the most common types of analyses used for quantitative data. They comprise for example (multivariate) analyses of variance or regressions. Conventional statistics tend to rely on assumptions such as random error and homogeneous variance. Descriptive statistics are traditionally applied to observational data. Descriptive data might include the distribution of organisms in space or community-habitat relationships. Statistics used to understand these types of data include diversity indices, cluster analysis, and principle components analysis, among many others. Statistical analyses might also include temporal or spatial analyses, and nonparametric approaches which do not rely on specific assumptions about the data's distribution. Other kinds of data analyses are simulations, model runs, parameter estimations and visualisations (soler et al, 2018).

Qualitative analysis on the other hand, requires that the data be analysed either in form of summary, on a thematic basis as they were observed. The choice on whether to use either methods or both methods of data analysis depends on the objective of the study, the instruments used for data collection, and the knowledge of the researcher.

In data analysis generally, care must be exercised to avoid committing type I or type II errors. Type I error occurs when a null hypothesis that ought to be accepted is rejected; type II error occurs when a null hypothesis that ought to be rejected is accepted. Sometimes these errors can occur when a wrong choice of data analysis technique is employed. It is possible for a statistical technique to yield significant results on a given set of data, while another statistical technique might yield insignificant results. Data are usually obtained at different scales of measurement thus; an appropriate technique must be selected that fits the type of data obtained. Computerized software or manual techniques could also be employed in analysing data. However, the most efficient technique of the two is computerized technique (Owan & Bassey, 2018).

- *Data synthesis/integration*

According to Soler et al (2016) data integration or synthesis is the merging of multiple datasets from different sources, like your recently collected data with former data from other owners, resulting in a new, bigger dataset. Integrating data from different sources is labour intensive and time consuming, because it requires understanding methodological differences, transforming data into a common representation, and manually converting and recoding data to compatible semantics before analysis can begin. Data integration for crosscutting studies is generally a manual process and can consume most of the time involved in conducting collaborative research (Michener & Jones 2012).

In data synthesis, the data broken into several components is again integrated to form a whole after having a thorough understanding of the different components that makes up the data. For instance, in studying the academic performance of SS2 students in 2018 second term examination in Cross River State, a researcher could collect 2018 mathematics results (data) of all SS2 students in Cross River State. Data analysis will enable the researcher to split the data either based on gender, school location, school type, or based on several other indices of interest. Data synthesis will require that the researcher integrate the separated data into a unified form that may be used to determine the overall performance of SS2 students in Cross River State. Data from different sources pertaining to the same project could be synthesized. Thus, data synthesis is the opposite of data analysis.

- *Data citation*

When data are collected and/or utilized from sources that are not of the researcher, it is ethical to cite the sources from where such data were collected. This is done to avoid plagiarism, and to shift the onus of explanation to the original author of such data if in any case, it is observed that such data were falsified or if there are errors. According to DataOne as cited in Soler et al (2016), when you use data from other sources than your own collection, it is important to indicate this. Citing allows for tracing the chain of use of data set and data elements, credit the creators of the data as well as the possibility that if errors or new information about the original data set or data elements comes to light, that any impact on your new datasets and interpretation of such could be traced

Benefits of data citation include: enables the sharing and re-use of data set, which in turn has been shown to increase the author citation; makes data legitimately citable; acknowledges data ownership; and allows for the use of data citation metrics (Pradal et al, 2017). Data must be cited like bibliographic resources have to be cited in publications (Soler et al, 2016).

The process of registration and citation depends on the type of data published and how it is published. Data can be considered as ‘published’ when it is generally discoverable (Pradal et al, 2017). Published data should be well-described (meta-data), citable, discoverable and re-usable wherever possible. As more journals have begun to require that data is made available to support research claims, research data has become an asset that requires citation.

- *Data utilization*

Data utilisation refers to a stage where the results emanating from research data analysis and synthesis are used in making decisions and in solving problems. Simply put, data utilization refers to the use of research data to reach an end. Data utilisation implies making practical use of data to provide evidence or justification, back up claims and speculations made in formulation of hypothesis. For instance, a researcher may speculate or hypothesize that; there is no relationship between two variables of interest. Such claims usually lack any evidence to support or disfavour it. The results from data analysis will enable the researcher to determine whether the claims made earlier where true or false and will direct the researcher on what decision to make. However, the process of data analysis may also be considered as data utilization since at this point, the researcher is making use of primary or secondary data to reach an end.

From a transportation perspective, Miller et al (2018) averred that data is used in numerous ways to study, plan, design, construct, operate, and monitor our transportation system. It helps planners understand traveller behaviour and helps policymakers identify ways to make the system more efficient and cost-effective. It is also used to understand traveller behaviour. These different uses are what make data an asset. The potential for infinite possible uses of data also creates challenges throughout the data life-cycle, from data collection to data destruction. How data can and will be used is dependent on how it is collected, processed, and stored (Miller et al, 2018).

- *Data storage*

Data storage is the process of saving research data for future purpose. Data storage is very important because it enables the researcher to retrace back if errors are found in the results of his or her analysis. It also provides a platform for data sharing and reuse which are discussed below. Data storage is necessary for future investigations because many other researchers may find such data important.

In many sciences experiments or observations cannot be repeated making at least part of the data so valuable that it needs to be stored for a long time. In many cases the value of data can only be realized after many years by new generations (RDA Europe 2014).

- *Data backup*

Data backup refers to the activity of making sure that research data do not get lost. This is done by storing such data in various files, locations, and media. So that, a damage to media A will implies that there is the same data in media B. Thus, data backup prevents against data loss. Data could be stored or backed up manually using cabinets, files, drawers, cupboards etc; using the computer through files, folders, flash drives, hard drives, compact disc (CD)/digital video disc (DVD) ROMS, memory cards, and so on. The internet could also be used to store data through e-mails, survey monkeys, google drive, and other cloud storage systems.

Soler et al (2016) maintained that backup of data is not the same as preservation; backups are short-term recovery solutions whereas preservation includes measures taken for long term storage and archiving. For backups, there exist ideally at least three copies of a file, on at least two different media, with at least one offsite. It was noted by Pradal et al (2017) that; it is the responsibility of a researcher to ensure that their research data is regularly backed-up and stored securely for the life of the project and throughout the retention period. Where possible, it would therefore be more efficient to have the research data backup process managed by ICT units governed under service level agreements with Research units and Data management teams. Offsite backup storage is also encouraged to ensure that research data is sufficiently distant from the primary data centre in the

event the primary site is destroyed. These offsite facilities may be physical, or cloud based (Pradal et al, 2017).

- *Data security/confidentiality*

Data security refers to the protective measures applied to private data sets in order to prevent unauthorized access to IP addresses, whether it be through computers, databases, websites, mobile devices, or vehicles (Hibbard & Austin, 2008). According to Pradal et al (2017), physical security, network security and security of computer systems and files all need to be considered to ensure security of data and prevent unauthorized access, changes to data, disclosure or destruction of data. Data security arrangements need to be proportionate to the nature of the data and the risks involved. Attention to security is also needed when data are to be destroyed (Pradal et al, 2017).

Very important data needs very tight security due to the interest of so many people. Research data often needs to be secured against theft, alteration, hacking, viruses, and damages. A virus is anything that alters or destroys the quality or quantity of data. Viruses generally, are often characterized by malicious intents to destroy data for one reason or the other. It could take the form of a computer program, or it could be in human form. It must also be noted however that, all viruses originate from man since it is men who also design or write programs to cause damages.

A computer virus is a written program by computer experts or programmers with the intent of stealing data, altering data, or causing damages to other peoples' computers, to make them lose certain data of interest in their possession. A human can be called a virus when individuals intentionally access other peoples' data storage systems or computers to steal data, delete data, cause damages, and modify data. These humans with such intents are known as hackers.

Hackers are experts who specialize in gathering information using tricks and make use of such information to destroy computer networks and websites to gain unauthorize access to the data in such networks or websites. Hacking can also be practiced physically without a network of computers or websites. Once data is altered, it will yield a misleading or different result and could further lose its integrity.

Data can be secured through anti-viruses, firewalls, password, app lock, folder locks, username, fingerprint scanners, keys and padlock, and many others. In securing data, ensure that you do not share your username, password, network internet protocol (I.P) address, and keys with anybody. All folders, drawers and cupboards should always be locked. The password to your email address(es) should always be kept private and do no use the same password for everything. All computers handling data should be passworded with a very strong password that cannot be guessed by anyone. In setting passwords, avoid using phone numbers, date of birth, year of birth, children's name, and so on.

According to Miller et al (2018), when data is secure and appropriately regulated, there is greater trust and confidence in its use. Data must be trustworthy and safeguarded from unauthorized access, whether malicious, fraudulent or erroneous. There are certain data types that does not need to be exposed for safety and security reasons. Once data is secured properly as in 3.7. above, data confidentiality is guaranteed. However, not all types of data need to be kept confidential due to their nature, importance and usefulness. Data that need to be kept confidential usually, are data pertaining to the identity of the respondents or sources where such data was collected. Data may sometimes be kept confidential due to the implication that exposing it will have on those affected by it.

- *Data access*

Public research data should be easily accessible by those interested in obtaining them. Data stored in large data repositories should be made available so that they could easily be searched and discovered. Secured data should also be accessible when the need arises. Thus, password used for data security should always be remembered. Data access gives room for verification of results and for data reuse. Most often, findings from studies are only published without the data that yielded such results. This makes it difficult for other researchers to verify the results of such studies due to the inaccessibility and/or unavailability of data. In the 21st century, data from studies could be accessed if they are stored using appropriate media. The rationale behind making data available is to promote accessibility and discovery.

Soler et al (2016) maintained that discovering data means to search and find data collected by other researchers. This data can be used for different purposes like long-time analysis, modelling or comparative studies. Data and metadata are made accessible through submission on any kind of shared environment. The pre-requisite for discovering data is that the authors are willing to share their data with the research community. Michener and Jones (2012) revealed that the problem with data access is that; many valuable and relevant data products are not readily available as they are stored on laptops and computers in the offices of individual scientists, projects and institutions. So, only submitted and published data can be discovered. Specialized data archives, centres and repositories offer search functions to discover relevant data. Another way to discover data is via journals, when data are appended to articles. Often it is necessary to register at a repository or archive to get access to the data (Soler et al, 2016).

- *Data retrieval*

Data retrieval is a data management technique that allows individuals to regain access to stored data. It enables individuals to make use of data that were previously stored. Secured data could be retrieved by those granted the access to utilize such data. Data in research may be retrieved for several reasons, including verification of results, data reuse, checking errors in research findings, for observation, and for reference purposes.

- *Data sharing/distribution*

Most times it is necessary to share research data with those conducting similar researches, or with those interested in the data. Data sharing gives room for comparison of results and to track changes in a population of interest. It reduces cost and time in conducting empirical investigations as many researchers will prefer to utilize secondary data (if available) than collect primary data. Educational research data are a valuable resource, usually requiring much time and money to be produced. Many data have a significant value beyond usage for the original research as is being increasingly recognized by donors. Some reasons for sharing data include the fact that it: encourages scientific enquiry and debate; promotes innovation and potential new data uses; leads to new collaborations between data users and data creators; maximizes transparency and accountability; enables scrutiny of research findings; encourages the improvement and validation of research methods; reduces the cost of duplicating data collection; increases the impact and visibility of research; promotes the research that created the data and its outcomes; can provide a direct credit to the researcher as a research output in its own right and it provides important resources for education and training.

Soler et al (2016) explained that the growing awareness for the importance of data results in the conclusion that research data should be made accessible. The DFG states in its “Guidelines on

the Handling of Research Data in Biodiversity Research" (2015) that data management should assure a re-use of data also for purposes other than those they were collected for.

Sharing data can bring advantages for individual researchers as well as for the scientific community in general. There are three dimensions of sharing data (Soler et al, 2016). First, data can be shared among researchers within the project team. Therefore, data is submitted to a shared drive. Second, data can be shared with researchers outside the core research team, e.g. when there is cross-institutional cooperation. In this case, data is submitted to collaboratively used drive. Third, data can be made publicly available. This is referred to as publishing data. Data centres like GFBio preserve data and make them discoverable (Soler et al, 2016). A study showed that the publication of data may increase citations (Piwowar & Vision 2013).

Another advantage is the prevention of unnecessary duplication of data collection. Furthermore, a key factor for science is replicability, so researchers can collect data and analyse them in order to produce similar results or assess previous work in the light of new (UK Data Service 2012-2015). However, if that information is not available or poorly documented and difficult to understand, re-use or replication is difficult (Heidorn 2008). Besides the voluntary sharing of data, many journals already request the submission of data underpinning a paper.

- *Data re-use*

Data re-use is the process where an already-used is put into use by researchers either to solve another problem or to validate the results of an earlier study in which the data was first used. Data are often re-used by other researchers conducting similar studies in other locations. Data stored in public repositories are often reused by data due to the issues they can address. According to Pradal et al (2017), data often have a longer lifespan than a research project that creates them. Researchers may continue to work on data after funding has ceased, follow-up projects may analyse or add to the data, and data may be re-used by other researchers. Therefore, well organized, well documented, preserved and shared data are invaluable to advance scientific inquiry and to increase opportunities for learning and innovation (Pradal et al, 2017).

When data are appropriately handled, it can have a long life with many uses beyond its original one and serve projects yet to be planned. Data reuse refers to using the same data more than once for the same purpose; data repurpose means using the same datasets to serve a new purpose that is different from the original purpose of the datasets (Miller et al, 2018). The repurposing of data enables the continuous extraction of value from data and leverages the data to solve new problems. This could also help to justify the expense of accumulating and managing huge volumes of data when organizations are monetizing or productizing their information assets. Mining existing data for potential value creates the opportunity to turn some of the cost of backup into a resource. There is no end in the data life cycle as far as data being continually reused and repurposed, creating new data products that may be processed, distributed, discovered, analysed, and archived (Miller et al 2018).

Despite the advantages of data re-use and repurpose, there are barriers such as data quality and perceived risk of reusing others' data. While secondary data analysis entails reusing data created from previous projects for new purposes, trustworthiness of data sources could be an issue. Oftentimes, there is lack of documentation of what has been done to the data set, which becomes a significant disincentive to reusing data. Not knowing how the data was collected, coded, validated and cleaned poses a potential risk of generating invalid results. Standardization of procedures and formats could help to address the problem.

- *Data Publication*

The publication of research data is also very important in the data management cycle. It enables the sharing of data used in publishing an empirical paper. The publication of research data alongside its corresponding manuscript should be encouraged to enable verification of results, to foster the conduct of similar studies, and to provide a basis for future researches that might require the use of such data. Soler et al (2016) stressed that it is necessary to publish datasets linked with the publication of related academic papers. It is because datasets are a fundamental part of the research process, as important as discussions and conclusions derived from them (American Geophysical Union (AGU), 2012). AGU recommended that the scientific community should recognize the professional value of such [data] activities by endorsing the concept of publication of data, to be credited and cited like the products of any other scientific activity and encouraging peer-review of such publications.

Furthermore, publication of datasets promotes transparency in the research cycle and facilitates verification and reproducibility. Therefore, it is important to establish mechanisms which allow sharing datasets while at the same time promoting rewards mechanisms for authors, as journals do for academic publications (Soler et al, 2016).

- *Data shredding*

Data shredding simply implies complete destruction to data that are no longer necessary. It refers to the process of destroying data that are obsolete/outdated and cannot be further used for any other meaningful activity. According to Pradal et al (2017), having a strategy for reliably erasing data files is a critical component of managing data securely and is relevant at various stages in the data cycle. During research, copies of data files no longer needed can be destroyed. At the conclusion of research, data files which are not to be preserved need to be disposed of securely (Pradal et al, 2017). The destruction method is normally selected based on the underlying sensitivity of the data being destroyed, or the potential harm they could cause if they are recovered or inadvertently disclosed (Miller et al, 2018).

It must be noted data shredding does not mean the same thing as deleting data. It is possible to delete data and such a deleted data can still be assessed by individuals who were not supposed to use them. Deleting data is like throwing away unnecessary data in a dustbin. Unauthorized persons may still pick up such data from the dustbin and use them as they wish without your knowledge. In a computer data management system, deleting data or formatting a storage device does also mean data shredding. Data may be deleted into the recycle bin, a computer expert may still have access to such data by visiting the recycle bin and restoring the data to its original location for unauthorize access. Even data deleted permanently from a computer hard drive can still be assessed by interested individuals by using third party software such as “recover my files”, file recovery”, “quick recovery”, “data recovery” and so on.

Thus, the most effective ways of managing unnecessary physical data is to shred them into unidentifiable forms such as burning, tearing booklets containing such data into very tiny particles and throwing or burying same in far locations. According to Miller et al (2018), physical media can be shredded or shattered using various physical destruction methods to keep the data from being recovered. For very low risk information, this may mean simply deleting electronic files or using a desk shredder for paper documents. However, these types of destruction methods can be undone, making these methods inappropriate for more sensitive data. For more sensitive data, stronger methods of destruction at a more granular level are needed to assure that the data are truly irretrievable. On the other hand, physical destruction can provide the highest assurance of absolute destruction of the data since it is impossible to reconstruct or recover the data from a disk or drive

that has been physically destroyed. But this involves high capital expenses and is considered an unsustainable and a costly way to dispose of data.

For computerized data, file shredding software may be used to destroy the data completely to a form that even when it is recovered using such programs mentioned above, they will be in an unhealthy state and cannot be useful. For instance, the CBL Data Shredder Program was designed to eliminate the chances that information stored on your hard drive may be retrieved by anyone when it, or the computer containing it, is disposed of. Experience has shown that files thought to have been deleted years previously, containing personal details, bank account details, credit card numbers, correspondence, etc., can be recovered all too easily, and simply formatting the drive is not an effective means of rendering this data inaccessible. This situation is made worse by the availability of off-the-shelf products that will automate the recovery process in some cases. When used in accordance with the instructions, the CBL Data Shredder Program will do what file deletion and partition formatting cannot: erase the entire contents of a treated hard drive, rendering them irretrievable to existing and future software-based recovery tools.

Miller et al (2018) also revealed that using software or hardware appliances to overwrite data is one of the most common ways to address data remanence. The advantage of this approach is that it is relatively easy and low-cost. It can be used selectively on part or all of a storage medium. using a device to remove to reduce the magnetic field of a storage disk or
Drive is also advantageous because it makes data completely unrecoverable.

CONCLUSION

The importance of data management educational research cannot be overemphasized. Data constitute the frame upon which research findings and decisions are built. It is imperative that educational research data be effectively managed in order to ensure that results emanating from empirical studies are reliable in solving educational problems. The data management practices or cycle as presented herein should be followed by researchers and scholars in the creation through to the shredding of research data.

Recommendations

Based on the position of this paper, it was recommended that:

i. Tertiary institutions in every part of the world should endeavour to establish a data management unit that will be saddled with the primary duty of formulating research data management policies, and the hosting of research data.
ii. Every Journal should as a matter of compulsion, require the submission of data set corresponding to empirical papers submitted by authors and scholars.
iii. Data should be stored/backed up using both manual and computerized devices, with adequate security measures in place to guard against unauthorized access.
iv. There should also be proper description of the data cleaning process(es) before data are uploaded to Journals or other data repositories. This will facilitate data re-use by reducing the complexities in the understanding of secondary data.
v. Obsolete and irrelevant data should be destroyed or shredded using both manual and computerized techniques, with the choice depending on the nature of the data and/or the implication of unauthorized access to such data.

REFERENCES

American Geophysical Union. (2012). Data position statement. Retrieved from http://sciencepolicy.agu.org/files/2013/07/AGU-Data-Position-Statement_March-2012.pdf

Digital Curation Centre. (2008b). Digital curation 101 migrate. Retrieved from http://www.dcc.ac.uk/sites/default/files/documents/DC%20101%20Migrate.pdf

Deutsche Forschungsgemeinschaft. (2015). Guidelines on the handling of research data in biodiversity research. Working group on data management of the dfg senate commission on biodiversity research. Information for Researchers No. 36. Retrieved from http://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_biodiversity_research.pdf

Durosaro, D.O. (2008). *Problems confronting schools' personnel in educational data collection analysis and storage*. Department of Educational Management, University of Ilorin, Ilorin. http/www.google.com.ng.

Esu, A. E. & Ntukidem, E.P. (2015). Primary education: Its origin. In A. E. Esu & E.P. Ntukidem (Eds). *Fundamentals of elementary education* (pp. 1 – 11). Calabar, Nigeria: Stiffaith Prints

Ezeagu, M.N. (2013). Data management practices of secondary school administrators in Nsukka Education Zone of Enugu State. M.Ed. Thesis Submitted to Department of Educational Foundations, University of Nigeria, Nsukka.

Heidorn, B. (2008). Shedding light on the dark data in the long tail of Science. *Library Trends, 57*(2), 33 – 42.

Hibbard, E. A., & Austin, R. (2008). *Storage security professional's guide to skills and knowledge*. SNIA, www.snia.org/ssif.SNIA

Michener, W. K. (2006). Meta-information concepts for ecological data management. *Ecological Informatics, 1*(1), 3–7.

Michener, W. K., & Jones, M. B. (2012). Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution, 27*(2), 85–93.

Miller, K., Miller, M. Moran, M. & Dai, B. (2018). *Data management life cycle*. Texas A&M Transportation Institute, PRC 17-84 F

Ngdata (2018). What is data management? Retrieved on 20th Nov., 2018 from *https://www.ngdata.com/what-is-data-management/*

Owan, V.J. & Bassey, B. A. (2018). Comparative study of manual and computerized software techniques of data management and analysis in educational research. *International Journal of Innovation in Educational Management (IJIEM), 2*(1), 35 – 46.

Patel, D. (2016). Research data management: a conceptual framework. *Library Review, 65*(4/5), 226-241.

Piwowar, H. A., & Vision, T. J. (2013). Data reuse and the open data citation advantage. Peer J, 1(e175). http://doi.org/https://dx.doi.org/10.7717/peerj.175

Pradal, P., Akullah, B., Bazarova, M., Devare, M. & Gardiner, P. (2017). Research data management: Good practice note. CGIAR Internal Audit Unit, SMB-ARC 13/9/17.

RDA Europe - Research Data Alliance. (2014). Report on the RDA-MPG Science Workshop on Data. Retrieved from https://europe.rd-alliance.org/sites/default/files/report/RDA-Europe-Science-Workshop-Report_final_April2014.pdf

Soler, A, S., Ort, M. & Steckel, J. (2016). An Introduction to Data Management. BEFmate, GFBio project. Unpublished preprint

UK Data Service. (2012-2015). Why share data? Retrieved May 22, 2015, from http://ukdataservice.ac.uk/manage-data/plan/why-share.aspx

Walliman, N. (2011). *Research methods: The basics*. New York, USA: Routledge