IUBAT Review
A Multidisciplinary Academic Journal

Mozaffar Alam Chowdhury. 2016. "Research and Scientific Data Management in Academic Institutions." *IUBAT Review* 1 (1): 47-53. iubat.edu/journal

Research and Scientific Data Management in Academic Institutions

Mozaffar Alam Chowdhury
Assistant Professor, Finance
College of Business Administration
IUBAT-International University of Business Agriculture and Technology, Dhaka, Bangladesh
email:mchowdhury@iubat.edu

ABSTRACT: *The study of this paper is about research and scientific data management in academic institutions. Academic institutions are the creators of scientific research data, generated from both primary and secondary research. The objectives of the study are to identify research in academic institutions and identify how scientific raw data are managed, identify the data ownership in the research project, identify quality of raw data in research and identify the dissemination and publication process of the research results by academic institutions. The methodology of this study is based on secondary research that examines the theoretical framework of research in scientific disciplines. Data management addresses the key issues from raw data collection to recording in a hard and soft copy. Reputable academic institutions implement guidelines and policies for scientific data dissemination and publication. Finally, the suggestions with the concluding remarks have been made.*

KEYWORDS: *Research, Data Management, Record, Dissemination and Publication.*

Introduction

The topic of this study is research and scientific data management in academic institutions. Academic institutions are important creators of scientific research data in many disciplines. These data are generated from both primary and secondary research in all sciences and they are managed through collection, storage, validation, record, retention, protection and reporting/publication. Scientific data management varies institution to institution. Openness and dissemination of research results are very important in academic institutions. If scientific research data are not preserved and disseminated, future researchers interested in study of the subject cannot build on past knowledge and the general public cannot learn.

Quality of data, data ownership and data recording are the key issues that need to be addressed. Frequently funding agencies require a record of data to be kept for a certain period after the research project is completed. If data are not recorded in a digital format that allows access by others, then researchers will not be able to read/re-evaluate the results. This paper intends to address all these problems in a systematic manner. The paper has set some objectives and to serve the objectives, a methodology is designed to investigate the information in an orderly which is based on literature review/ secondary material. After having the current knowledge on the topic from the literature review, it will be determined whether there is any gap in the literature. Finally, findings will be determined from the detailed discussion followed by some suggestions and conclusion.

Objectives

The objectives of this article are to identify:

- the nature of research in academic institutions.
- how scientific raw data are managed.
- the ownership of data in research projects.
- the quality of raw data in research.
- the dissemination of research results.
- publication process of research results in academic institutions.

Methodology

The methodology is based on a review of the available literature/secondary material on research and scientific data management from recent articles, journals, books and reports.

Review of the Literature

New scientific research is often financed by funding agencies, which need to identify the people who will conduct the research. Among the group of people involved in the research project, one or more researchers are identified as the "principal investigator(s)" who are primarily responsible for issues related to data management. The issues include data ownership (who has the legal rights over the data?), data collection practice (collecting research data in a systematic and ethical manner), data storage (concerns for reconstruction of project

results and maintenance of confidentitality of human participants), recording (the physical process of collecting raw data in either a notebook/hardcopy and/or electronic copy in a computer), data protection (protecting written and electronic data from physical damage or theft), data retention (length of time project data need to be kept), data analysis (how raw data are chosen, evaluated, interpreted in a meaningful way with a significant conclusion so that other researchers and public can understand and use them), data sharing (whether data should be shared or not and how scientific data and research results are disseminated to other researchers and general public), and data reporting (publication of findings whether positive or negative) (Steneck, 2004).

The principal investigator, project director, relevant academic institutions and funding agencies may have the right to access and determine data where primary data should remain in the laboratory (University of Pittsburgh 2009).

There is no best method for data collection. Different types of research apply different data collection techniques. The important consideration that applies to data collection is integrity. For this, there should be an appropriate method, which makes data reliable. If the method is inappropriate or compromised by bias, any conclusion has little value. There is a common saying that 'garbage in, garbage out' (GIGO). In addition to the choice of method, researchers must pay attention to details to avoid mistakes. Appropriate consent must be obtained to collect data on human and animal subjects, to use biological specimens, to publish photographs and other copyrighted materials (Steneck, 2004).

In general, two primary data records are maintained in a laboratory: the methodology notebook and the experimental notebook. In the case of the University of California methodology and experimental notebooks and related data and records are the property of the university. The Principal Investigators (PIs) have final responsibility for the validity and quality of the data. All data must be stored in notebook and electronic copy form and retained in their laboratory for five years after the date when funding for a study ends (University of California, 2015). Protocol for recording data and preparing for publication of a research findings should meet these standards.

Data storage through electronic systems has some unique advantages relative to hard copies: rapid access to the data, fast read; low cost; ease of ability to archive the data, remove the data, and backup data by, for example, storing data on CDs (Straub 2004). Properly storing data is intended to safeguard data that may be needed in the future by other researchers who might wish to evaluate the results of your research and finally it is intended to protect researches in the event of legal allegations (US Department of Health and Human Services 2015).

Northwestern University has a policy on retention of university research data. Data must be retained for a minimum of three years after the financial report for the project period has been submitted. If research data are an intellectual property, then data must be kept for as long as necessary *(Northwestern University, 2012)*. There is no set amount of time for which data should be retained. Sponsoring institutions may have differing requirements. Some organizations may require that, after

the funding period, data should be retained a minimum of three years. When the decision has been made to end data retention, the data should be completely destroyed (US Department of Health and Human Services 2015).

Recording raw data should be done after data collection and validation. Recording should be durable, accessible and safe from tampering or falsification. After appropriate coding, electronic records allow researchers to access and compare information from different sources. There are numerous electronic data capture programs that allow researchers to enter, store and audit research data. There may be questions as to how recorded data are to be protected. For example, protection may involve use of unique user IDs and passwords and limited administrator access rights. The principal investigators should be required to update computer anti-virus protection and maintain up-to-date software. If the system is connected with the internet, use of firewalls and encryption may be required (US Department of Health and Human Services 2015).

Scientific research may employ a combination of hard copy and electronic record keeping balancing the risk and benefits of each. Guidelines and policies affect the validity of collected data that involve human and animal subjects and biological specimens. The record should include the information such as date and time, names of members who worked with the data, materials, instruments and software used and identification numbers to indicate the subjects. When transferring records from written to electronic format, use a double entry system to avoid error. To implement such a system, two researchers should enter all raw data into two different software programs, then cross-check the data to identify and remedy inconsistencies at the time of data entry (US Department of Health and Human Services 2015).

All data should be considered for data sharing and data should be made as widely and freely available as possible while safeguarding the privacy of participants, and protecting confidentiality and proprietary data (NIH 2003).

Researchers share the results of their work with colleagues and public in a variety of ways, such as laboratory meetings, seminars and professional meetings. Final results are communicated through scholarly journal articles and books. Public communication takes place through press releases, newspaper articles, public announcements and public testimony. All forms of publication should present full and fair description of the work, accurate report of the results and honest and open assessment (methods, results and discussions) of the findings (Steneck, 2004).

The Merriam-Webster Dictionary (2015) defines data as "factual information used as a basis for reasoning, discussion, or calculation". For example, scientific medical research data may include a patient's temperature reading, blood pressure reading, red-blood cell count etc. In recent developments, the sources of research data in digital form used by different institutions are human biological samples within the healthcare system, GIS data in climate and environmental research and employment information data used in social science research. The research data are variously owned by principal investigators, university and sponsoring organization/funding agency (US Department of Health and Human Services 2015).

Theoretical Framework

Research means detailed study of a subject, analysis, writing and reporting of results particularly in relation to scientific research in academic institutions. Research uses both primary and secondary data where primary data require collection of original research data through questionnaires, interviews and observation in social science, experiment and direct observation in engineering, medical and physical sciences. On the other hand, secondary research is literature/desk based and data are collected from published authorized documents and data sources. Research may be quantitative, qualitative or both. It depends on the researchers' approach – for example, choice of survey methodology and research questions, the type of claims researchers are willing to make, the topic/issue researchers are interested in and the skill of researchers. Quantitative research implies a systematic empirical investigation through statistical, mathematical, numerical data or computational techniques. On the other hand, qualitative research includes description, usually extracts from interviews, focus group discussion and observation *(Kenneth D. Bailey 1994)*.

Summary Observations and Findings

In many disciplines, such as medicine, mathematics, engineering, physical and social sciences scientific research projects are usually conducted by academic institutions under a funding agency. The principal investigators address complex issues related to data management.

The principal investigators and the academic institutions have rights to access and retain the data in research projects where primary data remain in the laboratory.

Some academic institutions have policy on retention of research data for a specified minimum number of years after the project period has been ended. If research data are intellectual property, then data are kept for as long as necessary.

The two primary forms of data records maintained in a laboratory are the methodology notebook and the experimental notebook, which are the property of the academic institutions. Principal investigators have final responsibility for the validity and quality of the data. Publication of a research work should meet various data management standards.

Data collectors should be concerned with integrity of the process, although there is no single best way to collect data. If the collection method is compromised by bias, the results will be of limited value ("garbage in, garbage out").

Research data can be disseminated and made freely available whle respecting privacy, confidentiality and proprietary issues. Dissemination takes many forms: laboratory meetings, seminars, professional meetings, scholarly journals, books, press releases, newspaper articles, public announcements and public testimony.

Suggestions

Principal investigators should address issues related to data management before conducting scientific research.

Principal investigators, academic institutions and funding agency should keep the legal rights and ownership of the work done.

Academic institutions should have a clear and written guideline and policy related to right of access, ownership and reporting/ publication of the data and results of scientific research.

Assessment of the results should be checked for research integrity to allow for publication.

Standards for publication of scientific research should be communicated by journals to authors wanting to publish results of their research.

Conclusions

Data management in social and scientific research raises varied issues. If they are not treated adequately, the reliability of research cannot be assured. As Bangladesh acquires more universities and more faculty members become interested in research, it becomes more important to assure responsible data management in Bangladesh universities. The responsibility of the principal investigator becomes more important. So too, the role of the university in determining adequate ethical procedures becomes more important. Typically, universities enjoy access and ownership rights with respect to research undertaken by their faculty members.

Over time, Bangladesh universities will have to prepare appropriate guidelines and policy related to research data management, and require all faculty members and other researchers to respect the guidelines and policy. This article has indicated the requirements of high quality data management in American universities and scholarly journals. While Bangladesh universities cannot copy these requirements precisely, they serve as a model of best practice for Bangladeshi universities.

References

Anderson et al. 2011. *Statistics for Business and Economics*. USE: South Western.

Bailey, Kenneth D. 1994. *Methods of Social Research*. 4th ed. NY: The Free Press.

Arovelius, R. et al. 2010. *Management and Preservation of Scientific Records and Data*. International Council on Archives. http://www.1tu.se/cms_fs/file/Handbook.pdf

Merriam-Webster Dictionary. 2015. Accessed on 27/01/2015, http://www.merriam-webster.com

Northwestern University (2012). Policies and Guidelines for Investigators in Scientific Research, at http:www.research.northwestern.edu/policies/documents/research data.pdf, accessed on 23/01/2015.

National Institutes of Health (2003). NIH Data Sharing Policy and Implementation Guidance, available at http://grants.nih.gov/grants/policy/data_sharing.htm, accessed on 25/01/2015.

Straub J. 2004. "The Digital Tsunami: A Perspective on Data Storage." *Information Management Journal*. http://connection.ebscohost.com/c/articles/digital-tsunami-perspective-data-storage

Steneck, Nicholas H. 2004. *Introduction to the Responsible Conduct of Research.* Office of Research Integrity, http://ori.hhs.gov/sites/default/files/rcrintro.pdf

US Department of Health and Human Services. 2015. "Guidelines for Responsible Data Management in Scientific Research." Accessed 23/01/2015, http://ori.hhs.gov/images/data.pdf

University of California – San Francisco, Department of Neurological Surgery. 2015. "Guidelines on Research Data and Reports." Accessed 23/01/2015, http://neurosurgery.ucsf.edu/research_guidelines.html

University of Pittsburgh. 2009. "Guidelines on Research Data Management." Accessed 23/01/2015, http://www.provost.pitt.edu/documents/RDM_Guidelines.pdf