Data and its Important

The collection and organization of data are an integral and critical part of the research process. This module will discuss data acquisition, its importance to research, management of data, and data management plans and tools.

Learning Objectives:
1. Define data and discuss the different types of data
2. Learn the importance of data and the importance of maintaining data quality

Definition of data

Data is the collected factual material commonly accepted in the scientific community as necessary to validate research findings (OMB Circular A-110). This information is used to develop an understanding of a natural phenomenon and in analyzing and predicting future events. "Research data is data that is collected, observed, or created, for purposes of analysis to produce original research results" (Boston University Libraries, n.d.a). In this module, the word "data" is used in reference to "research data".

Types of data

As discussed in the "Introduction to Research" module, there are different types of research. These research types may include quantitative data (e.g., statistical measures), qualitative data (e.g., interviews or narratives), and mixed methods research, which involves a combination of both quantitative and qualitative measures. Depending on the research type, the type of collected data and its source will change. Data may be intangible (as in measured numerical values) or tangible (as in physical research materials such as purified chemicals, modified cells, or samples of rocks, plants, or insects collected in the environment). Data can be generated or collected from simulations, interviews, observations, surveys, experiments, or even from previous literature (Boston University Libraries, n.d.a).

Observations can be conducted on humans, animals, plants, or sometimes even on inanimate objects such as soil. These observations may be conducted by human researchers or by optical or digital sensors. Many times, data are irreplaceable. For this reason, it is important for researchers to be accurate and precise with their data collection and storage/preservation techniques and processes. Simulation data is usually generated in laboratories on test modules. This type of data is reproducible but may be expensive and time consuming to replicate. Interviews and surveys are conducted on people, or human subjects, and this data can either be descriptive in nature or empirical. Data of this type may vary in the level of detail, and may or may not be generalizable to other populations, meaning that what you discover in one context may not applicable in another context. Depending on the source and methodology involved, data

collection may be time consuming and expensive. Derived and complied data are reproducible but may be costly because of data mining (Boston University Libraries, n.d.a). Data mining is the process of analyzing data from different perspectives and summarizing it in useful information (University of California, Los Angeles, n.d.).

Some examples of research data include measured values, responses to questionnaires, transcripts, algorithms, audio, video, spreadsheets, word documents, laboratory notes, and simulation output. The types of data that you collect will vary by field and are subject to the norms and practices of these fields.

Importance of data in research

Data are intended to represent facts and without proper preservation of the context of collection and interpretation, may become meaningless (Boston University Libraries, n.d.a). The collection of data and its analysis assists researchers with discovering answers to their research questions and hypotheses. In some cases, it even predicts future outcomes (Office of Research Integrity, n.d.a). This is why data is an integral part of research. The results of research and its methods are directly dependent on the collected data and its analysis.

It is extremely important to maintain integrity (accuracy and completeness) while collecting data. Accurately collecting data ensures that:
· Research questions are answered correctly
· Research/experiments are repeatable and validated
· Research resources are not destroyed or wasted
· Other researchers are not misled about results and can benefit from your research findings
· Human subjects and animals are not harmed while collecting data
· Public policy decisions are not compromised (Office of Research Integrity, n.d.a).

Data quality needs to be maintained before, during, and after data collection. It is important to develop detailed plans for data collection. Faculty advisors, or principal investigators (PIs), are often responsible for ensuring the integrity of the research process, including storage of data. For example, PIs who conduct human subjects research have to manage where and how the data will be collected and stored securely along with the recruitment of research subjects. To collect data, PIs may have to compensate participants for any costs or inconvenience to the participant and maintain confidentiality of the data once collected. It is good practice to document the details of data collection process (Office of Research Integrity, n.d a.). This practice not only helps in documenting the methods section of a research paper but also assists in the maintenance of the data quality. Once again, such practices may differ by field and by the types of research conducted.

References:

Boston University Libraries (n.d.a). What is research data? Retrieved from
http://www.bu.edu/datamanagement/background/whatisdata/

Office of Research Integrity (n.d.a) Responsible conduct in data management. Retrieved
from http://ori.hhs.gov/education/products/n_illinois_u/datamanagement/dctopic.html

University of California, Los Angeles (n.d.). Data Mining: What is data mining?
Retrieved from
http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm