Sponsored Programs Guidance "Cradle to Grave"

DIVISION OF AGRICULTURE
RESEARCH & EXTENSION
University of Arkansas System

Data Management

Data Management for Sponsored Programs
(Adapted from
Guidelines for Responsible Data Management in Scientific Research,
Clinical Tools, Inc.)

I. Introduction

The U.S. Department of Health and Human Services, Office or Research Integrity, identifies data management as one of several components of responsible scientific research. Data management oversight is a key responsibility of researchers, and some federal programs require a data management plan be included in funding proposals. This document provides an overview of data management issues and an outline of elements included in a data management plan.

II. Concepts of Data Management

Data is defined as "factual information used as a basis for reasoning, discussion, or calculation." Data include information and observations taken from scientific inquiry, questionnaires, and interviews, as well as the materials, means, and products used and obtained during the course of inquiry. In short, data encompasses all information obtained during an investigation as well as the means by which it is acquired.

In its 2004 publication, *Introduction to Responsible Conduct of Research,* the Office of Research Integrity identified several issues related to data management:

- Data Ownership: Legal rights and retaining of data, including transferring data between institutions.
- Data Collection: Reliable, valid collecting procedures with a system for evaluating and recording changes to protocol.
- Data Storage: Amount of data to be stored and storage costs.
- Data Protection: Ensuring data integrity and protection from physical damage, tampering, and theft.

- Data Retention: Length of time data must be stored and how data is to be destroyed.
- Data Analysis: How data are chosen, evaluated, and interpreted.
- Data Sharing: How and if project data and research results are disseminated to researchers and general public.
- Data Reporting: Publication of findings.

III. Data Ownership

Data ownership involves several concepts, including possession, rights to publishing, and questions of funding, affiliation, and sources of research. In the case of federal grants, ownership typically involves three parties: the institution, the principal investigator, and the funding agency.

A. Institution

Normally, data ownership rests with the institution in whose name funding for a study was obtained. As the grantee, it is the institution's duty to ensure responsible and ethical conduct of research. Effective data management is essential to meeting this responsibility.

B. Principal Investigator

While data ownership may lie with the institution, administrators often grant stewardship to the principal investigator (PI). The PI typically controls data and utilizes it for publication and copyright. In some cases, PIs retain partial right of ownership and take data with them if changing institutions. In these instances, institutional policy and funder guidelines dictate transfer of data.

C. Funding Agency

As the funding sources for research, agencies may stipulate data ownership. Sometimes the agencies specify how data is retained and disseminated, if publication is allowed, and how the data is used. When submitting a proposal, both the institution and principal investigator should be aware of restrictions or requirements on which funding is contingent.

In addition to ownership rights of those involved in conducting and funding the research, PIs must consider research subject rights. When human subjects are researched, investigators must ensure beneficence and dignity related to data collected. Informed consent requires investigators to fully explain how data will be used, and any deviation or expansion of the proposed use can result in legal issues related to ownership rights of research subjects.

IV. Data Collection

The data collected during the course of research represents the information necessary to reach conclusions. In order to be valid, researchers must collect data using reliable, consistent, and comprehensive methods. While collection methods are normally described in project proposals, investigators should take care to document collection and analysis techniques prior to beginning a project. In order for research to be disseminated, understood, and evaluated, investigators must adequately define the purpose of the data, collection methods, implementation processes, analysis techniques, unexpected results or errors, and implications of the conclusions.

Investigators must maintain written and electronic records in order to justify research results if questioned at a later date. PIs should consult institutional policy and funding agency requirements to ensure compliance with data collection and record maintenance guidelines.

V. Data Storage

Following collection and recording of data, an investigator must take steps to ensure that the data is protected in both scientific and legal terms. Appropriately stored data can be referenced in future research, safeguards all parties' investments in the project, and protects research subjects and investigators in the event of legal challenges. PIs need not retain all raw data collected during the course of a project, but they should keep statistics, analysis, notes, and observation for future use. Essentially, enough data should be stored to allow a study to be reconstructed if necessary.

In dealing with electronic storage methods, investigators should take care to ensure ease of access and adaptability to future computer hardware and software. Rapid access, fast read/write rates, low cost, archival ability, and backup storage methods must all be considered when data is maintained in electronic format.

VI. Data Protection

Responsible data management requires that data be protected from physical damage, loss, tampering, or theft. Limiting access only to those directly involved in the project is generally considered the best method of protecting data, and it should be stored in a safe, secure location out of public view. Privacy and anonymity of research subjects must also be ensured. In some cases, researchers can develop data protection methods; however, it is reasonable to employ a third-party if necessary.

The Office of Research Integrity recommends the following protections related to data:

- Access
  - Use unique user IDs and passwords
  - Change passwords on a regular basis
  - Provide access to data through centralized process
  - Evaluate and limit administrator access rights
  - Ensure outside devices cannot access the network
- Systems
  - Update anti-virus protection
  - Maintain software and media storage devices
  - Use firewalls
  - Employ intrusion detection software
- Integrity
  - Record original creation dates of files
  - Use encryption, electronic signatures, or watermarking
  - Back up electronic data and maintain hard copies
  - Destroy data when appropriate

VII. Data Retention

The length of time data should be retained varies based on the nature of the research, funding guidelines, and institutional policy. In developing retention guidelines, officials must consider the cost of continued storage as well as the ongoing potential for violation of anonymity and confidentiality of research subjects.

When data storage ends, all data must be destroyed to the point that it can no longer be extracted, reconstructed, or analyzed. Shredding services and software can be utilized for the destruction process.

VIII. Data Analysis

Data analysis techniques should be appropriate for the research conducted and in compliance with funding agency guidelines. Research team members and third-party consultants should be designated specific duties related to analysis to ensure continuity and reliability of results. Research ethics and misconduct of science must be considered during any analysis. For additional information, see the appropriately titled document.

IX. Data Sharing and Reporting

The scientific process requires that data be shared and reported to the extent the funding agencies' guidelines allow. Sharing of data encourages additional research and collaboration that may have direct impacts on individuals or society as a whole. Data should be made available to appropriate individuals as soon as the results of a study are published or released. In some cases, sharing of preliminary data is encouraged as further analysis may be of immediate benefit. In all cases, however, investigators should consult institutional policy and funder requirements prior to sharing data, as Homeland Security, Freedom of Information Act, and general confidentiality laws may require limitations.

X. Data Management Plans

The University of Michigan has developed an outline of some elements that can be included in a data management plan. It should be noted that some funders place limits on the length of a plan that can be included in a proposal, so it may be necessary to highlight aspects of the plan rather than include the plan in its entirety. In any event, a complete data management plan should be kept on file.

- Data Description: A description of information to be gathered, including the nature and scale of the information.
- Existing Data: A survey of existing data relevant to the project and how it will be integrated.
- Format: Formats in which data will be generated, maintained, and shared, including justification for chosen methods.
- Metadata: Description of metadata to be provided and a discussion of standards used.
- Storage and Backup: Methods by which data will be maintained, including physical and electronic resources.
- Security: Technical and procedural protections including confidentiality, permissions, restrictions, and embargo methods.
- Responsibility: Names of individuals responsible for data management.
- Intellectual Property Rights: Individuals or entities holding rights to data and copyright constraints.
- Access and Sharing: Description of how data will be shared, including access procedures, technical mechanisms for dissemination, and approved user groups.
- Audience: Secondary users of data.
- Selection and Retention Periods: Description of how data will be selected for storage, how long it will be stored, and plans for eventual destruction.
- Archiving and Preservation: Procedures for long-term preservation including succession plans if data owners or stewards are no longer capable of holding the data.
- Ethics and Privacy: Methods of informed consent and privacy protection.

- Budget: Costs of preparing and archiving data.
- Data Organization: Description of how data will be managed during the project.
- Quality Assurance: Procedures for ensuring data integrity during the project.
- Legal Requirements: Listing of relevant laws or policies impacting data management.