



Online Ethics Center  
FOR ENGINEERING AND SCIENCE

# Topics: Data Management

## Author(s)

Michael Kalichman  
P.D. Magnus  
Dena Plemmons

## Year

2016

## Description

A guide that provides information and resources on teaching responsible conduct of research that focuses on the topic of data management. Part of the Resources for Research Ethics Education collection.

## Body

Introduction

[What is Research Ethics](#)

[Why Teach Research Ethics](#)

[Evaluation](#)

Topics

[Overview](#)

**Animal Subjects**

**Authorship**

**Biosecurity**

**Collaboration**

**Conflicts of Interest**

**Data Management**

**Human Subjects**

**Mentoring**

**Peer Review**

**Publication**

**Research Misconduct**

**Social Responsibility**

**Stem Cell Research**

**Whistleblowing**

**Educational Settings**

Descriptions of [educational settings](#), including in the classroom, and in research contexts.

**Discussion Tools**

**Case Studies**

**Debates**

**Literature**

**Textbooks**

## Videos

## Other Discussion Tools

### About the RCREC

Information about the [history and authors](#) of the Resources for Research Ethics Collection

## Summary

### What are data?

- Data can be defined as **measurements, observations, or any other primary products of research activity**.
- Data are the **empirical basis for scientific findings**. The integrity of research depends on integrity in all aspects of data management, including the collection, use, storage, and sharing of data.
- Data are **not just numbers in a lab notebook**. Depending on the research, data might include: images, audio or video recordings, genetically modified organisms, specialized software, ancient artifacts, or geological samples.

### Nominal “best practices”

- All researchers have an interest in, and responsibility to, **protect the integrity of the research record**.

*Questions to be asked to foster data integrity:*

1. How will the data be **collected**?
  2. How should records be **kept and stored**?
  3. How, if at all, will data be **backed up**?
  4. **How long** should data be kept?
  5. **Who owns the data**?
  6. When and with whom should data be **shared**?
- Research groups should at least be clear about how the above questions are answered for their particular circumstance.

- Research records should be sufficient to reconstruct what was done: Both for the purpose of **future research** and to **verify** that the work had been done as described in subsequent publications.

## Background

### Overview

Because of concern about many cases of research misconduct, the Department of Health and Human Services (1990) convened a workshop on data management. The workshop highlighted the many ways in which research depends fundamentally on responsible data management. Several good resources provide comprehensive reviews of good data management practices (e.g., Macrina, 2014; Mays and Macrina, 2014) and recordkeeping in particular (e.g., Kanare, 1985; NIH Office of the Director, 2008).

Responsible research:

- begins with **experimental design and protocol approval**;
- is supported by **recordkeeping** that ensures accuracy and avoids bias;
- is defined by **criteria for including and excluding data** from statistical analyses; and
- entails responsibility for the **collection, use, and sharing of data**.

### Responsibilities

- Everyone with a role in research has a responsibility to ensure the integrity of the data.
- The ultimate responsibility belongs to the principal investigator.
- However, the central importance of data to all research means that this responsibility extends to anyone who helps in planning the study, collecting the data, analyzing or interpreting the research findings, publishing the results of the study, or maintaining the research records.

## Regulations and Guidelines

### Overview

Data management in research is **rarely regulated**, except:

- most prominently for research subject to FDA requirements (FDA, 2016) and
- some instances regarding record retention and data sharing as noted below.

Otherwise, data management is **subject to professional** (e.g., Steward and Balice-Gordon, 2014; American Psychological Association, 2015; American Statistical Association, 2016), **institutional** (e.g., University System of New Hampshire, 2015; Michigan State University, 2016; UC San Francisco, 2016), **organizational** (Blum, 2012; Howard Hughes Medical Institute, 2006), or even **research group-specific guidelines**.

The summaries below present guidelines that might commonly be considered by most researchers.

## Data Collection

Because data collection can be repetitious, time-consuming, and tedious, there is a temptation to underestimate its importance.

However, **adequate planning and preparation** can:

- decrease the risk of wasted resources
- increase the likelihood of useful results
- assure that those responsible for collecting data are sufficiently trained and motivated
- generate research designs to limit or eliminate the effects of bias

## Recordkeeping

The best model for recordkeeping will not be the same for all areas of research.

However, nearly all types of research include records that could reasonably and usefully be kept in **bound lab notebooks**.

Nominal records would include:

- **date**
- **investigators**
- **what** was done, and

- **where** corresponding research products can be found.

Lab notebooks should be supplemented as needed by specialized methods of recordkeeping such as computer files, videotapes, and gels.

## Ownership of Data

Research data belong to the institution

Research ownership typically passes from the funder of the research (e.g., a federal agency or a private funder) to the University or institution, **not to the research investigators**.

Although the products of research involve creative contributions to new knowledge, the resulting **data are in effect no different from the routine products** of employees in any other private or public institution.

Equipment, materials and reagents, and the resulting data all belong to the institution in which they are purchased or produced, despite the language and practice of science.

The issue of institutional ownership becomes especially salient if:

- a marketable product is produced
- someone moves from one institution to another

If a principal investigator moves, then she or he can normally expect to take the data, but exceptions do occur and equipment transfer is nearly always a matter for negotiation.

### **Ownership by Principal Investigator**

In practice, even though the University or institution has legal standing to make decisions about what can or will be done with research data, it does not typically do so.

Absent an explicit agreement or ruling to the contrary, the **principal investigator**

**(PI) has primary responsibility for decisions about the collection, use, and sharing of data.**

## Retention of Data

The quality of data supporting published work is moot if the data are lost or discarded.

Retaining records of research is necessary not only **for the purpose of research**, but to:

- validate priority for claims of **intellectual property**
- demonstrate **ownership or patent rights**
- respond to requests under the **Freedom of Information Act**
- document the validity of **allegations of misconduct**

These concerns raise questions about what should be retained, who keeps the records, how they should be stored, and for how long.

*What should be retained?*

This depends in part on the nature of the products of research.

Some materials, such as thin sections for electron microscopy, cannot be kept indefinitely because of degradation.

It is also impractical to store extraordinarily large volumes of primary data.

At minimum, **enough data should be retained to reconstruct what was done.**

Who keeps the records?

Original data are the responsibility of the **principal investigator** (PI) and should be kept in her or his lab or office.

Although most researchers have the expectation that graduating students may take

copies of their research records, student or postdoctoral researchers should assume unless told otherwise that their **original data will stay with the PI**.

If regulations or other considerations preclude researchers from taking copies, then **the PI has a responsibility to make this clear** to the research group before work begins.

How should records be stored?

Any stored data will be rendered useless if there are insufficient records to locate and identify the material in question.

Ease of access must be **balanced against security**, for instance, if the study involved human subjects with a reasonable expectation of confidentiality.

Although the institution is the legal owner of the data, it is usually the **responsibility of the principal investigator** to ensure that records are stored in a secure, accessible fashion.

How long should records be kept?

Under current National Institutes of Health (NIH, 2015) and National Science Foundation (NSF, 2005) requirements, research records must be maintained for **at least three years after the last expenditure report**.

Federal regulations or institutional guidelines may require that data be retained for longer periods. However, these formal requirements are minimal constraints.

Decisions about retention of records should take into account:

- extent to which a line of research is **still being pursued**
- likelihood of **ongoing interest** in the research
- continued assurances of **confidentiality for any human subjects**
- **space and expense necessary for storage**.

## Sharing of Data



Federal agencies, particularly the NIH (2003) and NSF (2010), have made **funding contingent on plans to share research data and products**, particularly after publication.

An open data policy reflects positively on those who share and benefits science by **increasing the likelihood for new insights, collaboration, and reciprocal sharing**.

Although sharing of data is generally in the best interests of science and the individual, it is clear that such sharing can place an individual scientist at risk:

- Sharing data before publication could result in **loss of credit or opportunity**
- Exposure of data to the **prejudiced scrutiny of competitors or detractors**
- Risk of **compromising the confidentiality of human subjects**
- Expense of **time and resources** to meet requests for sharing of data

However, reasonable strategies to minimize potential problems should make it possible to choose sharing over secrecy. **Before publication**, it is best to maintain an open data policy with appropriate caution. **After publication**, be prepared to grant reasonable access to the raw data; that is, honor requests that are in the interest of scientific inquiry and can be accomplished without inordinate expense or delay.

In 2003, the National Institutes of Health put out a Final NIH Statement on Sharing Research Data. This document addresses some of the concerns listed above and makes clear that data sharing is a crucial and necessary part of responsible conduct in research.

## Discussion

### Case Studies

#### Case Study 1

Dr. Apple, a researcher working under a National Science Foundation (NSF) grant is studying the replication of bacteriophage in *E. coli*. Dr. Apple attends a lecture where world-renowned scientist, Dr. Ball, discusses her studies on the replication of a particularly useful bacteriophage that infects *E. coli*. Dr. Apple requests a sample of Dr. Ball's bacteriophage. Dr. Ball declines to provide a sample, even after several persistent and strongly-worded telephone calls from Dr. Apple. Dr. Apple, obsessed with securing Dr. Ball's bacteriophage has a plan. Dr. Apple writes a letter to Dr. Ball and again requests the material. At the conclusion of the letter, Dr. Apple pleads, "If you insist on denying me this virus, at least give me the courtesy of a written response to this letter." Dr. Ball quickly responds with a one-page, one-sentence response: "Forget it!" After receiving Dr. Ball's letter, Dr. Apple knowing Dr. Ball's propensity for performing all her working tasks at the lab bench, takes the letter, places it in a blender making a slurry using sterile buffer and spreads the slurry on lawns of bacteriophage recipient strains of *E. coli*. Soon, Dr. Apple isolates the long-sought strain of the bacteriophage.

Questions:

1. What are the implications of a refusal to share data?
2. Were Dr. Apple's actions appropriate?
3. Should Dr. Apple's actions give rise to an investigation of possible scientific misconduct by the NSF?
4. If the bacteriophage were used in a commercial pharmaceutical process and Dr. Ball was employed by the pharmaceutical company, did Dr. Apple illegally obtain a trade secret from Dr. Ball?

© ASM Press, 2000, Scientific Integrity by F.L. Macrina, used with permission.

## Case Study 2

Dr. Bernard Sears is the principal investigator on a grant from the National Institutes of Health. The primary aim of the research is the chemical modification of specific natural products for use as anti-cancer agents. Part of this research involves phase I clinical trials in humans, testing these chemically modified derivatives for toxicity. During the fourth year of this five-year grant, Dr. Sears develops health problems and is diagnosed as being terminally ill. During the first several months of his illness, he is able to work from his home office. He has his research technician bring all of the relevant data books from the chemical synthesis phase of the research to his home. He begins analyzing and organizing the data with the intention of preparing

manuscripts for publication. When Dr. Sears becomes too ill to work, the institution contacts the NIH and requests the appointment of an interim principal investigator. NIH approves of this request. The phase I clinical trials move ahead smoothly during the last year of the proposal and the results are promising. Unfortunately, before the grant funding period ends, Dr. Sears dies. The financial books on the grant are closed at the end of the five-year funding period. A final financial report is filed with the NIH several months later. During this time Dr. Sears' household, has been liquidated and his assets are subsumed into an estate. The NIH sends a request for a final scientific report on the project. The Director of Sponsored Programs gives this request to the interim principal investigator. After several weeks of investigation, no information can be gained on the location of the data books that Dr. Sears took home with him in order to write his manuscripts. Following an investigation by the University legal office and the campus police, it is concluded that the data books were destroyed when Dr. Sears's household was sold. Is there any recourse to salvage this important study? Who should be held accountable for these events? Should anyone be punished? Has scientific misconduct been perpetrated?

© ASM Press, 2000, *Scientific Integrity* by F.L. Macrina, used with permission.

### **Case Study 3**

Dr. Roger Powers is the editor-in-chief of *Infectious Agent Sciences*, a journal published by the North American Society of Infectious Diseases. He recently received a letter from Dr. William Ernst, the head of a laboratory outside of the United States. Dr. Ernst complains bitterly that his request for a bacterial strain described and published in the journal have been refused. This request has been made to Dr. Stanley Fields, an expert in the field of bacterial antibiotic resistance. The strain being requested by Dr. Ernst contains a novel combination of genetic elements making it resistant to most families of commonly used antibiotics. Dr. Fields, like many microbiologists, has been following world-wide developments about bioterrorism and the weaponizing of infectious agents. Fields knows the institution that employs Dr. Ernst was once engaged in biological warfare research during the Cold War era. Fields fears that the antibiotic resistance traits found in his strain could be put to harmful use by engineering them into disease-causing bioweapons, and this is his rationale for refusing to send a culture of it to Dr. Ernst. However, Dr. Ernst points out in his letter that the Instructions to Authors of *Infectious Agent Sciences* explicitly states that any materials reported in the journal must be made available to interested researchers wishing to employ them in non-commercial uses.

This policy is stated as a condition of publication in the journal. Dr. Ernst affirms this is in keeping with his request. Obviously, Dr. Rogers' uneasiness leaves him feeling quite differently. What should Dr. Powers do to resolve this problem?

© ASM Press, 2000, Scientific Integrity by F.L. Macrina, used with permission.

## Discussion Questions

1. What products of your research might reasonably be classified as data and/or necessary to verify the integrity of your work?
2. In your field of research, what are some of the steps an investigator can take during the planning stages to help ensure the integrity of a research project?
3. How are research records maintained in your research group? Does this approach meet the proposed goal of documenting what was done, when the work was done, who did the work, and the location of the corresponding research products?
4. Under what circumstances is it acceptable in your field of research to exclude an anomalous data point from analysis? If data were excluded from an analysis, then how should the published manuscript reflect that not all data are reported?
5. Is it unethical to choose a statistical test only after seeing which of several tests provide a statistically significant result? Why or why not?
6. When someone leaves your research group, what restrictions, if any, are imposed on what research records he/she takes with them?
7. If two people work together on a research project that is not yet published, and then decide to stop working together, who has the right to use the data in a future publication (both, the more senior of the two investigators, or neither)? In cases where this is not clear, what could be done in your institution to resolve the dispute?
8. In your area of research, what advantages might be gained by sharing your data and findings with other research groups?
9. In your area of research, what disadvantages might result from sharing your data and findings with other research groups?
10. What rules or guidelines does your institution have for data sharing?
11. How long after the final expenditure report for a Public Health Service- or National Science Foundation-funded project must research records be retained? What rules or guidelines does your institution have for data retention?

## Resources

- [OEC Data Management Subject Aid](#)  
A good starting point for individuals interested in learning more about the ethics of data management, including a definition, guidelines, and key articles and readings.
- [OEC Data Management Bibliography](#)  
Policies, books, and articles on data sharing and data management ethics.
- [Big Data in the Life Sciences](#)  
A bibliography looking at data management in the life sciences.
- [Data Management Bibliography](#)  
A bibliography developed by Oxford University Computing Services

## Cited Resources

1. American Psychological Association (2015): [Data Sharing: Principles and Considerations for Policy Development](#).
2. American Statistical Association (2016): [Ethical Guidelines for Statistical Practice](#).
3. Blum C (2012): [Access to and Retention of Research Data: Rights and Responsibilities](#). Council on Governmental Relations, Washington, DC.
4. Department of Health and Human Services (1990): Data Management in Biomedical Research, Report of a Workshop, April 1990 Chevy Chase, Maryland.
5. FDA (2016): [Regulations](#). U.S. Food and Drug Administration. U.S. Department of Health and Human Services.
6. Ferguson AR, Nielson JL, Cragin MH, Bandrowski AE, Martone ME (2014): [Big data from small data: data-sharing in the 'long tail' of neuroscience](#). Nature Neuroscience 17: 1442-1447.
7. Howard Hughes Medical Institute (2006): [Chapter 8. Data management and Laboratory Notebooks](#). A Practical Guide to Scientific Management for Postdocs and New Faculty. Pp. 143-152.
8. Kanare HM (1985): Writing the Laboratory Notebook, American Chemical Society, Washington, DC.
9. Macrina FL (2014): Chapter 10. Scientific Recordkeeping. In (Macrina FL, au): Scientific Integrity, 4th ed., ASM Press, Washington, DC, pp. 329-359.
10. Mays TD, Macrina FL (2014): Chapter 9. Research Data and Intellectual Property. In (Macrina FL, au): Scientific Integrity, 4th ed., ASM Press, Washington, DC, pp. 287-357.

11. Michigan State University (2016): [Life Cycle Data Management Planning](#).
12. NIH (2003): [Final NIH Statement on Sharing Research Data](#).
13. NIH (2015): [8.4.2 Record Retention and Access](#). Monitoring, Administrative Requirements, NIH Grants Policy Statement.
14. NIH Office of the Director (2008): [Guidelines for Scientific Record Keeping in the Intramural Research Program at the NIH](#).
15. NSF (2005): [Records Retention and Audit](#). Chapter III - Grant Administration. NSF Grant Policy Manual.
16. NSF (2010): [Data Management and Sharing Frequently Asked Questions \(FAQs\)](#).
17. Steward O, Balice-Gordon R (2014): Rigor or Mortis: Best Practices for Preclinical Research in Neuroscience. *Neuron* 84(3):572-581.
18. UC San Francisco (2016): [Data Sharing & Data Management](#).
19. University System of New Hampshire (2015): [UNH Policy on Ownership, Management, and Sharing of Research Data](#).

## Notes

The Resources for Research Ethics Education site was originally developed and maintained by Dr. Michael Kalichman, Director of the Research Ethics Program at the University of California San Diego. The site was transferred to the Online Ethics Center in 2021 with the permission of the author.

## Rights

Use of Materials on the OEC

## Resource Type

Instructor Materials

## Parent Collection

Resources for Research Ethics Education

## Topics

Data Management

## Discipline(s)

Research Ethics

Teaching Ethics in STEM