

Data Management

By Michael Kalichman, 2001

Contributors: P.D. Magnus and Dena Plemmons

Originally published at www.research-ethics.net.

Republished with permission.

Summary

Data provide the factual basis for scientific work.

The integrity of research depends on integrity in all aspects of data management, including the collection, use, and sharing of data.

Integrity of the data is a shared responsibility.

All researchers have an interest in, and responsibility for, protecting the integrity of the research record.

Quality of data collection depends on thoughtful planning.

Adequate preparation for data collection helps to ensure that resources are not wasted and that significant results can be obtained.

Selection and analysis of data should be specified.

If the research is to be presented in a useful and significant way, critical decisions about selection and analysis must be made before the research commences, when possible.

Data should be shared.

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

Background

Although research misconduct examples like these are dramatic, conducting research responsibly is more than avoiding intentional fabrication or falsification of data. Because data provide the factual basis for scientific work, the value of research depends directly on integrity in and management of all aspects of the collection, use, and sharing of data.

Concern about research misconduct was a primary motivation for a 1990 conference on data management sponsored by the Department of Health and Human Services. One of the outcomes of that conference was a summary of the many ways in which the conduct of research depends on responsible data management. Responsible research begins with

experimental design and protocol approval; it involves recordkeeping in a way that ensures accuracy and avoids bias; it guides criteria for including and excluding data from statistical analyses; and it entails responsibility for collection, use, and sharing of data.

Data can be defined as measurements, observations, or any other primary products of research activity. These provide a factual basis for inference, conclusions, and publication. If data are defined in this way as research products necessary to validate the integrity of published or reported work, then “data” consist of much more than just measurements written in a lab notebook.

Everyone with a role in research has a responsibility to ensure the integrity of the data. The ultimate responsibility belongs to the principal investigator, but 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.

Discussion

Case Study 1ⁱ

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 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?

Case Study 2ⁱⁱ

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 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 assumed 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?

Case Study 3ⁱⁱⁱ

Dr. Roger Powers is the editor-in-chief of the *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 that was described and published in the journal has 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?

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-funded project must research records be retained? What rules or guidelines does your institution have for data retention?

Additional Considerations

Data Collection and Recordkeeping

Because data collection can be repetitious, time-consuming, and tedious, there is a temptation to underestimate its importance. Care should be taken to assure that those responsible for collecting data are adequately trained and motivated, that they employ methods that limit or eliminate the effect of bias, and that they keep records of what was done by whom and when.

The best model for recordkeeping will not be the same for all areas of research. However, nearly all types of research include records that should be kept in bound lab notebooks. At a minimum, such notebooks can provide a listing of the date of research, the investigators, what was done, and where the corresponding research products can

be found. The lab notebook should be supplemented as needed by specialized methods of recordkeeping such as computer files, videotapes, and gels.

Ownership of Data

Legally, data are the property of the institution and not the investigator. Because the products of research involve creative contributions to new knowledge, it is easy to assume that the resulting data are in some way different from the routine products of employees in any other private or public institution. However, the 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, but it is also an issue when someone moves from one institution to another. If the principal investigator is moving, then she or he can normally expect to take the data, although exceptions do occur and equipment transfer is nearly always a matter for negotiation. Absent some explicit agreement or ruling to the contrary, the principal investigator has primary responsibility for decisions about the collection, use, and sharing of data. Student or postdoctoral researchers should assume that their original data will stay with the principal investigator. However, most institutions and researchers have the expectation that graduating students may take copies of their research records. If regulations preclude researchers taking such copies, then the principal investigator is responsible for making this clear to members of the research group before work begins.

Retention of Data

The quality of data supporting published work becomes moot if the data are lost. Records of research are necessary not only for the purpose of research. Research records can have legal standing for a variety of purposes, including, but not limited to, demonstration of priority for claims of intellectual property, ownership or patent rights, and requests under the Freedom of Information Act. In addition, nearly all aspects of misconduct allegations hinge on the extent and quality of documentation of the research. These concerns raise questions about what should be retained, how it 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.

- **How should it 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 it be kept?**

Under current Health and Human Services 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. These formal requirements are minimal constraints. Decisions about retention of records should take into account the extent to which a line of research is still being pursued, the likelihood of ongoing interest in the research, continued assurances of confidentiality for any human subjects, and the space and expense necessary for storage.

Sharing of Data

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. It is reasonable to fear that sharing data before publication can result in loss of credit or opportunity.

Other concerns include exposure of data to the prejudiced scrutiny of competitors or detractors, risk of compromising confidentiality of human subjects, and 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.

Resources

Department of Health and Human Services. (1990). *Data Management in Biomedical Research, Report of a Workshop*, April, 1990, Chevy Chase, Maryland.

Engler, R.L., Covell, J.W., Friedman, P.J., Kitcher, P.S., Peters, R.M. (1987). Misrepresentation and responsibility in medical research. *New England Journal of Medicine*, 317: 1383-1389.

Kintisch, E. (2005). Scientific misconduct. Researcher faces prison for fraud in NIH grant applications and papers. *Science*, 307(5717): 1851.

Normile, C., Vogel, G., & Couzin, J. (2006). Cloning. South Korean team's remaining human stem cell claim demolished. *Science*, 311(5758): 156-7.

Endnotes

- ⁱ © ASM Press, 2000, Scientific Integrity by F.L. Macrina, used with permission.
- ⁱⁱ © ASM Press, 2000, Scientific Integrity by F.L. Macrina, used with permission.
- ⁱⁱⁱ © ASM Press, 2000, Scientific Integrity by F.L. Macrina, used with permission.