

Editors' Overview: Topics in the Responsible Management of Research Data

Joe Giffels · Sara H. Vollmer · Stephanie J. Bird

Received: 6 July 2010 / Accepted: 7 October 2010 / Published online: 28 October 2010
© Springer Science+Business Media B.V. 2010

Abstract Responsible data management is a multifaceted topic involving standards within the research community regarding research design and the sharing of data as well as the collection, selection, analysis and interpretation of data. Transparency in the manipulation of images is increasingly important in order to avoid misrepresentation of research findings, and research oversight is also critical in helping to assure the integrity of the research process. Intellectual property issues both unite and divide academe and industry in their approaches to data management. Central to the realization and promulgation of responsible data management is clear and careful communication of standards and expectations within the research community to trainees as well as among colleagues. These topics are examined and explored in depth in a special issue of Science and Engineering Ethics on responsible data management.

Keywords Data · Data management · Data integrity · Research integrity · Responsible conduct of research · Research oversight · Ethics in science · Research compliance

J. Giffels (✉)
University of Maryland Baltimore, 5th Floor, 620 W Lexington St,
Baltimore, MD 21201, USA
e-mail: jgiff001@umaryland.edu

S. H. Vollmer
Ethics & Values Center, Department of Philosophy, University of Alabama
at Birmingham, 900 13th St South, 414A, Birmingham, AL 35294, USA
e-mail: vollmer@uab.edu

S. J. Bird
PO Box 2007, Wrentham, MA 02093, USA
e-mail: sjbird@mit.edu

The goal of research is to expand understanding of the universe. The production of data is fundamental to that enterprise, whether the data are experimental or observational, and the quality of data is critical.

Consumers of research data depend on the data's accuracy. Other researchers depend on the reliability of data produced by their colleagues because previous results serve as the foundation for further basic or applied research. Research findings are also used to inform public policy and in the development of technical disciplines by those outside the research community. These individuals, too, depend on the quality of data management practices to assure the validity of scientific support for their work. Producers and distributors of research data, whether researchers or scientific publishers, rely on the integrity of the data that they communicate to one another and the world at large to uphold their professional reputations. Together, all share responsibility for assuring data integrity.

Data management is one of the core concepts in the responsible conduct of research paradigm. It encompasses a broad range of topics relevant to the collection, selection, interpretation, storage, and distribution of data. Producers of data and their colleagues must develop and maintain a working knowledge of current standard and best practices in data management in order to successfully collect, select and utilize those data. Achievement of this requirement is challenged by the rapid evolution of the field. Changes in information technology require frequent readjustment of data management techniques. In addition, public policy developments, such as protection of private health information and requirements to share research data obtained using public funding, have complicated matters. Ethical considerations, such as respect for research participants, fairness to scientific colleagues and balancing potential benefits and potential harms, place further demands on research in general and data management standards in particular.

Because data are central to the scientific method, sufficient accuracy and precision of measurement is essential for both scientific achievement and ethical practice. Methodology for representing the data is key to maintaining the integrity of the original measurements, whatever form the data take, e.g., tables, graphs or images, and whether or not the data are digitized. This methodology is important for representing the data at all stages in research, from data recording to processing, analysis and reporting. Good methodology makes recording and accessing descriptions of all processing to which published data were subjected a transparent matter, and helps researchers maintain the fidelity of the original measurements.

Recognizing the importance of data management in the responsible conduct of research (RCR) (US DHHS Office of Research Integrity 2010), the US Department of Health and Human Services (DHHS) Office of Research Integrity (ORI) co-sponsored two conferences related to this topic. One was with the University of Alabama at Birmingham on September 14 and 15, 2006, in Birmingham, Alabama, and the other was with the University of Maryland Baltimore on September 28 and 29, 2006 in Baltimore, Maryland. Each was attended by over 150 participants and presenters. The first conference was entitled, "Statistics, Images, and Perceptions of Truth: Detecting Research Bias and Misconduct" and the second, "New Capabilities, Emerging Issues and Responsible Conduct in Data Management." Both dealt with the manner in which data management practices are developing in ways

relevant to the responsible conduct of research. The conferences featured presentations and discussions of some of the most significant new developments as they relate to data integrity.

Eight papers based on presentations from these conferences are included in this Special Issue of *Science and Engineering Ethics*. Each is accompanied by a commentary that complements its companion paper, offering further insights on the topics presented. The conferences and this Special Issue invite consideration of data management as a holistic discipline. If all participants in the research enterprise were to exploit the synergy generated from addressing the many facets of responsible data management in developing methodologies, researchers would find it easier to both keep pace with evolving technology, and promote responsibility and integrity in the research arena.

Half of the papers observe how inadequate attention to recommendations in the areas of image manipulation, research design, and research oversight can negatively affect data integrity. These papers suggest how integrating contemporary methods for data management and oversight with more traditional research practices can increase the effectiveness of the oversight review process. The remaining papers emphasize best practices in some of the more global aspects of data management, such as data sharing, intellectual property management, and training in relevant policy. Attention to, and communication of, best practices can foster and promote improved data integrity overall. A survey of the papers provides the reader with some context for determining where, on a continuum from preferred research practice through acceptable and questionable practices to misconduct, a given practice resides.

Douglas Cromey's contribution, entitled "Avoiding Twisted Pixels: Ethical Guidelines for the Use and Manipulation of Scientific Digital Images," warns researchers against crossing the line between simply rendering an image suitable for publication using conservative adjustments reported in the figure legend, and giving it a general makeover that beautifies but misrepresents the data (Cromey 2010). Cromey notes that, while new digital imaging techniques offer enormous flexibility in processing images and have revolutionized the acquisition and publication of scientific data, this flexibility presents novel ethical challenges for data management. Cromey covers many technical aspects of digital image acquisition and manipulation, suggesting lines of demarcation between acceptable and unacceptable practices. He acknowledges that most image manipulation, even where it leads to misinterpretation of the actual data observed, is not intended to be fraudulent. To assist researchers, administrators, writers, editors, and readers in understanding why fidelity is lost in certain kinds of image processing, Cromey proposes ethical guidelines that promote the highest degree of accuracy and precision in the representation of a scientific observation as a published image, and explains the reasons for recommending them. The associated commentary (Benos and Vollmer 2010) suggests that the same advances in computer technology that make the need for guidelines for image reproduction a critical issue, also make detecting the misuse of images easier; the resulting ease of detection of falsified images has implications for the responsibilities of scientists and editors during the review process.

The paper by Sara Vollmer and George Howard, “Statistical Power, the Belmont Report, and the Ethics of Clinical Trials” (Vollmer and Howard 2010) emphasizes that, if there is a sound statistical plan for data management from the start, a researcher can be less likely to bow to the temptation to publish overstated experimental results. They argue that what may seem like a non-ethical issue in clinical trial design, the statistical power of a trial, if not attended to, can actually verge on misconduct and, in some cases, lead to it. Their article analyzes power from a mathematical standpoint, and argues that insuring that a clinical study is adequately powered is one way of meeting federal guidelines for the ethics of clinical trials. The associated commentary explores issues that arise as a result of differing expectations regarding responsible research practice (Bird 2010).

Two papers describe methodologies employed by federal agencies for detecting fabricated and falsified data and suggest that research institutions avail themselves of the expertise and related techniques. “Raising Suspicions with the Food and Drug Administration: Detecting Misconduct” by Michael Hamrell (2010) notes that, in addition to unintentional bias that results from inadequacy of data validation, intentional bias in collecting, analyzing and presenting data due to financial interests in the outcome of the research is a significant concern. When sponsors need research results that are publishable, patentable or readily accepted by the Food and Drug Administration (FDA), it is incumbent on researchers and the FDA as the cognizant government entity to be especially scrupulous in developing and utilizing methods to validate data, for example, when data are received in support of drug or device applications. Hamrell provides an overview of issues related to the inspection and monitoring of clinical study sites. The commentary by Patricia Spitzig (2010) suggests that federal regulations for clinical study monitoring and oversight are a minimum standard and that innovation, for example, in the areas of accepting patient (or participant) input, respecting whistleblowers, and implementing a systems-partnership approach would promote the ethics of the clinical environment while at the same time advancing research.

John Dahlberg and Nancy Davidian of the US DHHS Office of Research Integrity (ORI) describe how scientific forensic methods are used by their Office to generate and develop evidence that will strengthen institutional findings of fabrication or falsification of data relating to federally-funded research (Dahlberg and Davidian 2010). Citing among others the case of the former University of Vermont nutrition researcher Eric Poehlman, Dahlberg and Davidian describe analytical methods that can improve the review of questioned data. They emphasize that using forensic methods to present clear and compelling evidence is important not only when research misconduct has been committed, but also when accusations are groundless.

In the associated commentary, Samuel Tilden (2010) explains how, following a determination of a finding of misconduct by ORI, the remedies pursued (whether administrative and thus designed to protect the integrity of the research program, or civil and criminal and designed to deter and punish wrongdoers) depend on the evidentiary findings. Tilden analyzes the Eric Poehlman case with regard to the findings and legal consequences.

Together, the Hamrell and the Dahlberg and Davidian papers make several cogent points. First, methodology is playing an increasingly important role in

assuring ethical practices, not only by preventing inadequately researched data from remaining uncorrected in the scientific record, but also perhaps by discouraging fabricated and falsified data from entering the scientific literature in the first place by proving cases of misconduct which lead to the punishment of culpable actors. Second, the authors invite others in the research enterprise (research grant and manuscript reviewers, scientific journal editors, and research institutions investigating misconduct) to adopt some of the methodologies and other technical assistance available from federal agencies in their mission to promote the responsible conduct of research and maintain the integrity of the scientific record.

Taking a wider view, Margi Joshi and Sharon Krag (2010) survey the topic by explaining the importance and implications of fundamental issues such as what qualifies as data, who is responsible for data management, and how context influences the attributes that qualify an instance of data management as responsible. Joshi and Krag stress that data are both the basis for communication in science, and a defense against allegations of misconduct. In their commentary, Julie Richardson and Diane Hoffman-Kim (2010) examine the definition of data for the institution or organization that produces it, particularly with regard to the relationship that exists between the definition and the production, retention, and informational needs of the particular institution.

Based on their collective experience as scientists and teachers, Julia Frugoli, Anne Etgen and Michael Kuhar (2010) offer suggestions for delivering training in data management and other RCR topics. In “Developing and Communicating Responsible Data Management Policies to Trainees and Colleagues,” they offer practical approaches for successful communication of RCR principles, policies, and practices, such as grouping didactic instruction by scientific discipline and exploiting the “teachable moment,” in the window of time early in the trainee’s career before bad habits have become ingrained. They also recommend teaching general principles as contrasted with specific rules as a way of helping trainees to keep up with the continuing, rapid evolution of data collection, storage and distribution methods.

The associated commentary by C.K. Gunsalus (2010) suggests that when RCR instruction includes the teaching of skills related to being a contributing member of the community, the entire community of scholars benefits. Skills that can improve the research climate in this way include those related to collaborating across disciplines, approaching disputes professionally, and dealing with issues that arise when implementing any suggested best practices.

An intellectual property perspective is taken by Lisa Geller (2010), who explores how, especially in academic settings, the demands of academic freedom and property rights often conflict. Geller describes how appropriately conceived data management practices can meet the needs of academic research while observing the property rights of both institutions and sponsors. In her commentary, Ramona Albin (2010) addresses the question of whether recent increased privatization in the form of expansion of patent protection of biotechnology has caused the system to lose effectiveness in balancing the protection of innovation and the avoidance of monopolies.

Finally, Beth Fischer and Michael Zigmond (2010) recognize “The Essential Nature of Sharing in Science.” Although sharing has costs, such as reducing the

uniqueness of the resources available to a scientist, as suggested by the title of their paper, the operational and strategic success of the scientific enterprise is dependent upon sharing among researchers, especially of data. Joe Giffels' (2010) commentary makes the case for robust data-sharing plans supported, both financially and operationally, at the institutional level.

While the adoption of sound data management practices would seem an immediately obvious component of any program to assure the integrity of research, too often it is overlooked. This may be because, as many of the papers in this volume suggest, such practices require thoughtful application of often highly technical considerations. It may also be due to the rapidity with which the technical aspects as well as norms within the scientific community evolve. Time and resource considerations undoubtedly play a role, as do researchers who must decide whether to spend time meeting ongoing data management requirements of published research, or moving on to new research projects.

Investment in an integrated data management process, through the combined efforts of institutions and individual laboratories, will undoubtedly improve the overall integrity of research. Just as importantly, thoughtful data management practices and their careful implementation will support the logistics and execution of research itself as the output of research, data, are better managed. These are dividends which surely warrant such investment.

References

- Albin, R. (2010). Patents, innovation, and privatization—commentary on: “Data management in academic settings: an Intellectual Property perspective.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9234-2](https://doi.org/10.1007/s11948-010-9234-2).
- Benos, D., & Vollmer, S. (2010). Generalizing on best practices in image processing: A model for promoting research integrity—commentary on: “Avoiding twisted pixels: Ethical guidelines for the appropriate use and manipulation of scientific digital images.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9226-2](https://doi.org/10.1007/s11948-010-9226-2).
- Bird, S. J. (2010). Responsible research: What is expected?—commentary on: “Statistical power, the Belmont Report, and the ethics of clinical trials.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9248-9](https://doi.org/10.1007/s11948-010-9248-9).
- Cromey, D. W. (2010). Avoiding twisted pixels: Ethical guidelines for the appropriate use and manipulation of scientific digital images. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9201-y](https://doi.org/10.1007/s11948-010-9201-y).
- Dahlberg, J. E., & Davidian, N. M. (2010). Scientific ethics: How the Office of Research Integrity can assist institutional investigations of research misconduct during oversight review. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9208-4](https://doi.org/10.1007/s11948-010-9208-4).
- Fischer, B. A., & Zigmond, M. J. (2010). The essential nature of sharing in science. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9239-x](https://doi.org/10.1007/s11948-010-9239-x).
- Frugoli, J., Etgen, A. M., & Kuhar, M. (2010). Developing and communicating responsible data management policies to trainees and colleagues. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9219-1](https://doi.org/10.1007/s11948-010-9219-1).
- Geller, L. N. (2010). Data management in academic settings: An Intellectual Property perspective. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9240-4](https://doi.org/10.1007/s11948-010-9240-4).
- Giffels, J. J. (2010). Sharing data is a shared responsibility—commentary on: “The essential nature of sharing in science.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9230-6](https://doi.org/10.1007/s11948-010-9230-6).

- Gunsalus, C. K. (2010). Best practices in communicating best practices—commentary on: “Developing and communicating responsible data management policies to trainees and colleagues.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9227-1](https://doi.org/10.1007/s11948-010-9227-1).
- Hamrell, M. R. (2010). Raising suspicions with the Food and Drug Administration: Detecting misconduct. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9232-4](https://doi.org/10.1007/s11948-010-9232-4).
- Joshi, M., & Krag, S. (2010). Issues in data management. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9223-5](https://doi.org/10.1007/s11948-010-9223-5).
- Richardson, J., & Hoffman-Kim, D. (2010). The importance of defining ‘data’ in data management policies—commentary on: “Issues in data management.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9231-5](https://doi.org/10.1007/s11948-010-9231-5).
- Spitzig, P. (2010). Clinical study reflections: Another view—commentary on: “Raising suspicions with the Food and Drug Administration: detecting misconduct.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9235-1](https://doi.org/10.1007/s11948-010-9235-1).
- Tilden, S. J. (2010). Incarceration, restitution, and lifetime debarment: Legal consequences of scientific misconduct in the Eric Poehlman case—commentary on: “Scientific forensics: How the Office of Research Integrity can assist institutional investigations of research misconduct during oversight review.” *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9228-0](https://doi.org/10.1007/s11948-010-9228-0).
- US DHHS Office of Research Integrity (2010). <http://www.ori.dhhs.gov>. Accessed 6 Oct 2010.
- Vollmer, S. H., & Howard, G. (2010). Statistical power, the Belmont Report, and the ethics of clinical trials. *Science and Engineering Ethics*, 15, doi:[10.1007/s11948-010-9244-0](https://doi.org/10.1007/s11948-010-9244-0).