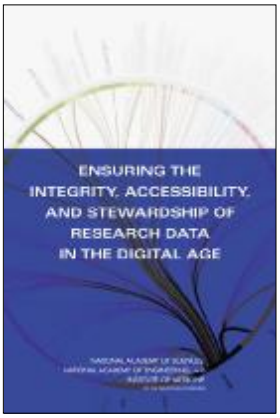
THE NATIONAL ACADEMIES

REPORT IN BRIEF

Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age (2009)



Digital technologies have transformed every facet of research, from the questions asked and the methods used, to the ways in which researchers interact. Large, openly accessible data collections are enabling the secondary use of data and transforming fields as diverse as astronomy (Sloan Digital Sky Survey) and the life sciences (GenBank). Researchers are accessing and combining data from multiple sources to predict the consequences of climate change and better understand the functions of the brain. New approaches based on the ability to widely share digital data, such as open notebook science, crowdsourcing, and citizen science, hold the potential for engendering greater transparency and efficiency in the research enterprise.

This transformation has raised many questions and concerns about the methods traditionally used to ensure the integrity and utility of research data. For example, what role does peer review or evaluation of data play in verifying research results when the underlying datasets are extremely large and have undergone automated collection and analysis? Who decides when and how research data should be made available to peers and to the broader public? Who is responsible for the stewardship of data that might have value long after the completion of the research that generated them?

The report provides a framework for dealing with the challenges of maintaining the integrity of scientific research data in this age of revolutionary digital technologies. It examines the impact of these technologies on acquiring, sharing, and storing data across scientific, engineering, and medical research, and recommends actions for researchers, universities, journals, and funding agencies to maintain data integrity, provide data access, and ensure future availability. The purpose of the report is to move the research community forward, collectively, to consider the integrity of data in ways that best support the advancement of scientific, engineering, and medical knowledge, as well as the maintenance of public trust.

It is not possible to provide specific recommendations that apply across disciplines. Each research field has its own conventions and procedures for acquiring, analyzing, and disseminating data. In some fields, for example, there are already strong incentives for researchers to make data widely available at the time when results are reported, or even prior to publication, while in others the propensity to share is currently much weaker. Nevertheless, there are three broad principles that extend across all fields of research, each with recommendations for specific action.

THE NATIONAL ACADEMIES
Advisers to the Nation on Science, Engineering, and Medicine

National Academy of Sciences • National Academy of Engineering • Institute of Medicine • National Research Council

Data Integrity Principle

Ensuring the integrity of research data is essential for advancing scientific, engineering, and medical knowledge and for maintaining public trust in the research enterprise. Although other stakeholders in the research enterprise have important roles to play, researchers themselves are ultimately responsible for ensuring the integrity of research data.

Recommendation 1: Researchers should design and manage their projects so as to ensure the integrity of research data, adhering to the professional standards that distinguish scientific, engineering, and medical research both as a whole and as their particular fields of specialization.

Recommendation 2: Research institutions should ensure that every researcher receives appropriate training in the responsible conduct of research, including the proper management of research data in general and within the researcher's field of specialization. Some research sponsors provide support for this training and for the development of training programs.

"Today, the actual implementation of training varies greatly from field to field and institution to institution. The National Institutes of Health (NIH) requires that graduate and postdoctoral students who are supported by NIH training grants receive instruction in the responsible conduct of research. The Office of Research Integrity at the Department of Health and Human Services supports . . . [the development of] education and training programs in the responsible conduct of research."

Recommendation 3: The research enterprise and its stakeholders—research institutions, research sponsors, professional societies, journals, and individual researchers—should develop and disseminate professional standards for ensuring the integrity of research data and for ensuring adherence to these standards. In areas where standards differ between fields, it is important that differences be clearly defined and explained. Specific guidelines for data management may require reexamination and updating as technologies and research practices evolve.

Recommendation 4: Research institutions, professional societies, and journals should ensure that the contributions of data professionals to research are appropriately recognized. In addition, research sponsors should acknowledge that financial support for data professionals is an appropriate component of research support in an increasing number of fields.

Data Access and Sharing Principle

Research data, methods, and other information integral to publicly reported results should be publicly accessible.

Recommendation 5: All researchers should make research data, methods, and other information integral to their publicly reported results publicly accessible in a timely manner to allow verification of published findings and to enable other researchers to build on published results, except in unusual cases in which there are compelling reasons for not releasing data. In these cases, researchers should explain in a publicly accessible manner why the data are being withheld from release.

"This principle may seem to apply only to publicly funded research, but a strong case can be made that much data from privately funded research should be made publicly available as well . . . . Note that this principle covers data underlying publicly reported results. When a researcher working at a corporate lab seeks to publish results, patent applications can be filed in advance of publication, so that making data accessible at the time of publication will not compromise commercialization of the invention in question."

Recommendation 6: In research fields that currently lack standards for sharing research data, such standards should be developed through a process that involves researchers, research institutions, research sponsors, professional societies, journals, representatives of other research fields, and representatives of public interest organizations, as appropriate for each particular field.

Recommendation 7: Research institutions, research sponsors, professional societies, and journals should promote the sharing of research data through such means as publication policies, public recognition of outstanding data-sharing efforts, and funding.

Recommendation 8: Research institutions should establish clear policies regarding the management of and access to research data and ensure that these policies are communicated to researchers. Institutional policies should cover the mutual responsibilities of researchers and the institution in cases in which access to data is requested or demanded by outside organizations or individuals.

"The development of standards . . . can be led by journal editors, professional societies, ad hoc bodies of researchers established to solve particular problems, or permanent institutions charged with overseeing data management issues. National Academies committees and advisory committees to federal agencies can play constructive roles. . . . . Input and participation from international stakeholders will often be needed."

Data Stewardship Principle

Research data should be retained to serve future uses. Data that may have long-term value should be documented, referenced, and indexed so that others can find and use them accurately and appropriately.

Recommendation 9: Researchers should establish data management plans at the beginning of each research project that include appropriate provisions for the stewardship of research data.

"At a minimum, data management plans for research projects should provide for compliance with the relevant legal and policy requirements . . . . In other cases, the data management plan would specify whether the data would be deposited in an institutional and/or disciplinary repositories, and annotation and metadata specifications . . . .This recommendation does not imply that individual researchers are responsible for ensuring indefinite preservation of their own data, only that they ensure that that it is prepared and transferred to the appropriate archives or repositories."

Recommendation 10: As part of the development of standards for the management of digital data, research fields should develop guidelines for assessing the data being produced in that field and establish criteria for researchers about which data should be retained.

Recommendation 11: Research institutions and research sponsors should study the needs for data stewardship by the researchers they employ and support. Working with researchers and data professionals, they should develop, support, and implement plans for meeting those needs.

These core principles and accompanying recommendations describe the steps that need to be taken to respond to the rapid changes being created by digital technologies. They do not represent a radical departure; they reinforce and extend the traditional openness and collaborative nature of science.

The research enterprise and its constituents are clearly moving in the right direction, albeit in a decentralized and uneven way. Continued progress will require short-term actions and persistent longer-term efforts, as well as dialogue and collaboration across sectors and fields. Adapting the fundamental principles that have guided researchers in the past to new realities is the surest way for the research enterprise to accelerate the advance of knowledge while maintaining the public trust.

COMMITTEE ON ENSURING THE UTILITY AND INTEGRITY OF RESEARCH DATA IN A DIGITAL AGE: Daniel Kleppner (Co-Chair), Massachusetts Institute of Technology; **Phillip Sharp** (Co-Chair), Massachusetts Institute of Technology; **Margaret Berger**, Brooklyn Law School; **Norman Bradburn**, The University of Chicago; **John Brauman**, Stanford University; **Jennifer Chayes**, Microsoft Research; **Anita Jones**, University of Virginia; **Linda Katehi**, University of Illinois at Urbana-Champaign; **Neal Lane**, Rice University; **W. Carl Lineberger**, University of Colorado at Boulder; **Richard Luce**, Emory University; **Thomas McGarity**, The University of Texas at Austin; **Steven Paul**, Eli Lilly and Company; **Teresa Sullivan**, University of Michigan; **Michael Turner**, The University of Chicago; **J. Anthony Tyson**, University of California, Davis; **Steven Wofsy**, Harvard University; **Thomas Arrison**, Study Director, National Research Council; **Deborah Stine**, Study Director, National Research Council (until 7/31/2007).

For more information

Copies of *Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age* are available from the National Academy Press (NAP); call (800)624-6242 or (202)334-3313 (in the Washington metropolitan area), or visit the NAP website at www.nap.edu. For more information on the project, contact staff at (202)334-3755 or visit the Policy and Global Affairs website at www.nationalacademies.org/pga.