

This prepublication version of the report has been provided to the public to facilitate timely access to the committee's findings. Although the substance of the report is final, editorial changes may be made throughout the text prior to publication. The final report will be available through the National Academies Press later this year.

# **SHARING PUBLICATION- RELATED DATA AND MATERIALS:**

# **RESPONSIBILITIES OF AUTHORSHIP IN THE LIFE SCIENCES**

## **Executive Summary**

**Committee on Responsibilities of Authorship in the  
Biological Sciences**

**Board on Life Sciences  
Division on Earth and Life Studies**

**National Research Council**  
*OF THE NATIONAL ACADEMIES*

**THE NATIONAL ACADEMIES PRESS**  
Washington, D.C.  
[www.nap.edu](http://www.nap.edu)

COMMITTEE ON RESPONSIBILITIES OF AUTHORSHIP IN THE BIOLOGICAL  
SCIENCES

**THOMAS R. CECH**, (*Chair*), Howard Hughes Medical Institute, Chevy Chase, Maryland  
**SEAN R. EDDY**, Howard Hughes Medical Institute; Washington University, St. Louis, Missouri  
**DAVID EISENBERG**, Howard Hughes Medical Institute; University of California, Los Angeles  
**KAREN HERSEY**, Massachusetts Institute of Technology, Cambridge  
**STEVEN H. HOLTZMAN**, Infinity Pharmaceuticals, Inc., Boston, Massachusetts  
**GEORGE H. POSTE**, Health Technology Networks, Gilbertsville, Pennsylvania  
**NATASHA V. RAIKHEL**, University of California, Riverside  
**RICHARD H. SCHELLER**, Genentech, Inc., South San Francisco, California  
**DAVID B. SINGER**, GeneSoft, Inc., South San Francisco, California  
**MARY C. WALTHAM**, Independent Publishing Consultant, Princeton, New Jersey

*Project Staff*

**ROBIN A. SCHOEN**, Study Director  
**BRIDGET K. B. AVILA**, Senior Project Assistant  
**ELIA BEN-ARI**, Science Writer  
**NORMAN GROSSBLATT**, Editor

## Executive Summary

The publication of experimental results and sharing of research materials related to those results have long been key elements of the life sciences. Over time, standard practices have emerged from communities of life scientists to facilitate the presentation and sharing of different types of data and materials. But recently there is a concern that, in practice, publication-related data and materials are not always readily available to the research community. Moreover, in some fields questions have arisen about whether standard practices really exist, or whether putative standards are accepted by and commonly applied to all authors.

That uncertainty is driven by several factors, including the changing nature of the participants in the scientific enterprise, the growing role of large datasets in biology, the cost and time involved in producing some data and materials, and the commercial and other interests of authors in their research data and materials. These circumstances have engendered widespread interest in a reevaluation of the responsibilities of authors to share publication-related data and materials.

As interest in the topic of standard practices was growing, the National Academies approached the National Cancer Institute, National Human Genome Research Institute, National Science Foundation, and the Sloan Foundation with the idea of undertaking a study of the issues related to sharing publication-related data and materials. With their support, in October 2001, the Academies created the Committee on Responsibilities of Authorship in the Biological Sciences, whose members were chosen from academe and the commercial sector for their expertise in the life sciences and medicine, and their experience with issues related to scientific publishing, databases, software, intellectual property rights, and technology transfer. The committee was given the following charge:

To conduct a study to evaluate the responsibilities of authors of scientific papers in the life sciences to share data and materials referenced in their publications. The study will examine requirements imposed on authors by journals, identify common practices in the community, and explore whether a single set of accepted standards for sharing exists. The study will also explore whether more appropriate standards should be developed, including what principles should underlie them and what rationale there might be for allowing exceptions to them.

To meet its charge and obtain a variety of perspectives on these issues, the committee organized a workshop, “Community Standards for Sharing Publication-Related Data and Materials”, that was held on February 25, 2002 at the National Academy of Sciences in Washington, DC. The participants included distinguished members of the life-sciences community—researchers and administrators from universities, federal agencies, and private industry; scientific-journal editors; and members of the legal and university technology transfer communities. Evaluation of the issues was stimulated by the group's analysis of several hypothetical situations (attached in an Appendix to the full report) that captured many of the difficult issues facing the community.

During the workshop, discussions about which data and materials related to a publication an author ought to provide and the precise manner in which they should be shared with others revealed how important those requirements are to the scientific community. Much of the analysis that took place in working groups was an effort to discern how an author (with individual competitive, commercial or other interests) could, by some minimum effort, meet the collective needs of the community. Regardless of the specifics of the hypothetical problem under discussion, the ability to resolve the situation satisfactorily depended ultimately on whether an author could meet the community's general expectations of getting what was needed to move science forward.

While largely unwritten, the community's expectations of authors are a reflection of the value of the publication process to the life-sciences community. The central role of publication in science also explains its value to scientists who want to publish their findings. For individual investigators, publication is a way of receiving intellectual credit and recognition from one's peers (and perhaps the broader public) for the genesis of new knowledge and the prospect of its conversion into beneficial goods and services. Publication also enhances a researcher's job prospects, ability to be promoted or gain tenure, and prospects for research support.

Companies whose scientists publish their findings also typically receive the intellectual credit, recognition, and prestige that come with such disclosures to the entire scientific community. Such nonfinancial benefits can translate into publicity and increased perceived value of a company to investors and business partners. They also strengthen the scientific reputation of the company in the eyes of potential collaborators, employees, and users of the company's products.

Regardless of the motivation, the arena of publication is where participants in the research enterprise share, and are recognized for, their contributions to science. Ultimately, this system benefits all members of the scientific community and promotes the progress of science. Although society encourages innovation in other ways (for example, through the patent system), the sharing of scientific findings, data, and materials through publication is at the heart of scientific advancement. A robust and high quality publication process is, therefore, in the public interest.

In this context, and informed by the views expressed at the workshop and its own subsequent deliberations, the committee found that the life-sciences community does possess commonly-held ideas and values about the role of publication in the scientific process. Those ideas define the responsibilities of authors and underpin the development of community standards--practices for sharing data, software, and materials adopted by different disciplines of the life sciences to facilitate the use of scientific information and ensure its quality. Central to those ideas is a concept the committee called "the uniform principle for sharing integral data and materials expeditiously (UPSIDE)", as follows:

**Community standards for sharing publication-related data and materials should flow from the general principle that the publication of scientific information is intended to move science forward. More specifically, the act of publishing is a *quid pro quo* in which authors receive credit and acknowledgment in exchange for disclosure of their scientific findings. An author's obligation is not only to release data and materials to enable others to verify or replicate published findings (as journals already implicitly or explicitly require) but also to provide them in a form on which other scientists can build with further research. All members of the scientific community--whether working in academia, government, or a commercial enterprise--have equal responsibility for upholding community standards as participants in the publication system, and all should be equally able to derive benefits from it.**

In addition to UPSIDE, the committee identified five corollary principles associated with sharing publication-related data, software and materials. The five principles further elucidate the common expectations of the life-sciences community of an author's responsibilities and form the basis of community standards tailored to the types of data and material integral to a particular field and the unique circumstances of research in a discipline. For example, the gene expression community is developing standards for sharing published microarray data, biological taxonomists are promoting a central repository for morphological images, and specialized distribution centers have arisen for many types of plant germplasm. Given the diversity of disciplinary communities in the life sciences, different standards are expected to arise. Nevertheless, the standards reflect a common basis in the principles identified in this report.

As noted in the full report, however, the details of community standards and the nuances of how the principles that underpin them should be interpreted are sometimes a matter of debate within disciplines. Some of these subtleties are discussed in the full report; the chapter in which they are addressed is indicated next to each of the five principles listed below.

### **Data and Software**

**Principle 1. (Chapter Three) Authors should include in their publications the data, algorithms, or other information that is central or integral to the publication—that is, whatever is necessary to support the major claims of the paper and would enable one skilled in the art to verify or replicate the claims.**

This is a *quid pro quo*--in exchange for the credit and acknowledgement that come with publishing in a peer-reviewed journal, authors are expected to provide the information essential to their published findings.

**Principle 2. (Chapter Three) If central or integral information cannot be included in the publication for practical reasons (for example, because a dataset is too large), it should be made freely (without restriction on its use for research purposes and at no cost) and readily accessible through other means (for example, on line). Moreover, when necessary to enable further research, integral information should be made available in a form that enables it to be manipulated, analyzed, and combined with other scientific data.**

Because scientific publication is intended to move science forward, an author should provide data in a way that is practical for other investigators. The data might reasonably be provided on-line but should be available on the same basis as if they were in the printed publication (for example, through a direct and open-access link from the paper published on-line). Making data that is central or integral to a paper freely obtainable does not obligate an author to curate and update it. While the published data should remain freely accessible, an author might make available an improved, curated version of the database that is supported by user fees. Alternatively, a value-added database could be licensed commercially.

**Principle 3. (Chapter Three) If publicly accessible repositories for data have been agreed on by a community of researchers and are in general use, the relevant data should be deposited in one of these repositories by the time of publication.**

The purpose of using publicly accessible data repositories is a practical one--to expedite scientific progress and provide access to data in a manner that allows others to build on it. By their nature, these repositories help define consistent policies of data format and content, as well as accessibility to the scientific community. The pooling of data into a common format is not only for the purpose of consistency and accessibility. It also allows investigators to manipulate and compare datasets, synthesize new datasets and gain novel insights that advance science.

## Materials

**Principle 4. (Chapter Four) Authors of scientific publications should anticipate which materials integral to their publications are likely to be requested and should state in the “Materials and Methods” section or elsewhere how to obtain them.**

Consistent with the spirit and principles of publication, materials described in a scientific paper should be shared in a way that permits other investigators to replicate the work described in the paper and to build on its findings. If a Material Transfer Agreement (MTA) is required, the URL of a Web site where the MTA can be viewed should be provided. If the authors do not have rights to distribute the material, they should supply contact information for the original source. A frequently requested reagent can be made reasonably available in the commercial market or by an author’s laboratory for a modest fee to cover the costs of production, quality control, and shipping.

**Principle 5. (Chapter Four) If a material integral to a publication is patented, the provider of the material should make the material available under a license for research use.**

When publication-related materials are requested of an author, it is understood that the author provides them (or has placed them in an authorized repository) for the purpose of enabling further *research*. That is true whether the author of a paper and the requestor of the materials are from the academic, public, private not-for-profit, or commercial (for-profit) sector. Notwithstanding legal restrictions on the distribution of some materials, authors have a responsibility to make published materials available to all other investigators on similar, if not identical, terms.



During the workshop, it was recognized that the responsibility for creating, updating, and enforcing community standards for sharing publication-related data and materials lies with all members of the community who participate in the publication process and have an interest in the progress of science. This includes academic, government, and industrial scientists; scientific societies, publishers and editors of scientific journals; and institutions and organizations that conduct and fund scientific research. In addition to creating, implementing and enforcing standards, some workshop participants suggested that the scientific community should also confront the problems that contribute to uncertainty surrounding standards, for example by creating incentives to share data and materials, and addressing the costs, administrative barriers, and commercial issues related to sharing.

Reflecting these concerns, the committee developed a set of recommendations that describe possible actions by participants in the scientific enterprise to address issues concerning sharing publication-related data and materials. The committee puts these recommendations forward for further discussion and consideration as best practices by the life-sciences community, whose members have the ultimate responsibility to develop and implement community standards.

**Recommendation 1. (Chapter Three) The scientific community should continue to be involved in crafting appropriate terms of any legislation that provides additional database protection.**

Some companies have identified the lack of commercial protection for databases as the key reason why they need to require investigators who want publication-related data to sign an agreement about their use of the data with the company. Database protection is important to the publication process because it could affect how and whether the community can use and recombine data held in databases. In the past, legislative proposals for increased database protection have been perceived by the community as having potentially negative consequences for sharing and using scientific data. It is in the interest of the life-sciences community to be an

active participant in ensuring that any proposed database protection is consistent with the principles of publication and enables researchers working in companies to publish on the same terms as other authors.

**Recommendation 2. (Chapter Four) It is appropriate for scientific reviewers of a paper submitted for publication to help identify materials that are integral to the publication and likely to be requested by others and to point out cases in which authors need to provide additional instructions on obtaining them.**

Most journals today explicitly or implicitly require that authors provide enough detail about their materials and methods to allow a qualified reader to verify, replicate, or refute the findings reported in a paper. Members of the scientific community support the publishing process by participating as peer-reviewers, often requesting additional supporting information. Identifying materials likely to be requested is consistent with that practice.

**Recommendation 3. (Chapter Four) It is not acceptable for the provider of a publication-related material to demand an exclusive license to commercialize a new substance that a recipient makes with the provider's material or to require collaboration or coauthorship of future publications.**

Authors should enable others to build on their findings. To build on the author's work, a recipient might need to assemble materials from multiple providers, and they cannot all be granted exclusive licenses. Demanding an *exclusive* license to a new substance made by another investigator using the author's material will effectively block the recipient from assembling the materials needed to conduct research. In addition, although collaborations and coauthorship often arise naturally when materials are shared (to the mutual benefit of the scientists involved) it is unacceptable to require collaboration or coauthorship as a condition of providing a published material, because that requirement can inhibit a scientist from publishing findings that are contrary to the provider's published conclusions.

**Recommendation 4. (Chapter Four) The merits of adopting a standard MTA should be examined closely by all institutions engaged in technology transfer, and efforts to streamline the process should be championed at the highest levels of universities, private research centers, and commercial enterprises.**

The purpose of sharing publication-related materials is to enable research--that is, to allow the recipients of material to replicate and build on the work of the authors--and the terms of MTAs and their negotiation should not create a barrier to this goal. Because there are so many nuances in the negotiation of MTA related issues, there is a potential for delay in reaching agreement, and sometimes there is an impasse. The proliferation of MTAs with idiosyncratic requirements set by multiple institutions is, in the end, an impediment to sharing publication-related materials.

**Recommendation 5. (Chapter Four) As a best practice, participants in the publication process should commit to a limit of 60 days to complete the negotiation of publication-related MTAs and transmit the requested materials or data.**

Such a commitment would eliminate uncertainty for the requestors of materials and remove what is currently perceived as a substantial barrier to the ability of investigators to move forward with their research plans. If sharing publication-related materials in a timely fashion is important to participants in the publication process, authors and others should encourage their institutions to commit to achieving that goal.

**Recommendation 6. (Chapter Six) Scientific journals should clearly and prominently state (in the instructions for authors and on their Web sites) their policies for distribution of publication-related materials, data, and other information. Policies for sharing materials should include requirements for depositing materials in an appropriate repository. Policies for data sharing should include requirements for deposition of complex datasets in appropriate databases and for the sharing of software and algorithms integral to the findings being reported. The policies should also clearly state the consequences for authors who do not adhere to the policies and the procedure for registering complaints about noncompliance.**

Many journals do not specify policies about sharing data and materials in their instructions to authors. By incorporating transparent standards into their official policies (including a statement of consequences for authors who do not comply), journals can encourage compliance. It is not known how many instances of noncompliance are ever brought to the attention of journal editors or other external authorities; however, a letter from the editor in chief or managing editor is often sufficient to resolve problems. Although some journal editors would consider denying a noncomplying author further rights to publish in their journals, on rare occasions, public opinion might be the most influential way to obtain an author's compliance. A journal might choose to declare an author's noncompliance (after all honest attempts were exhausted) in a specific section dedicated to this purpose.

**Recommendation 7. (Chapter Six) Sponsors of research and research institutions should clearly and prominently state their policies for distribution of publication-related materials and data by their grant or contract recipients or employees.**

NSF, NIH, and other funding organizations, such as the Howard Hughes Medical Institute (HHMI), have policies that reinforce and in some cases extend the standards set by the research community for depositing data in public databases. NIH has also issued a set of principles and guidelines on obtaining and disseminating biomedical research resources, although these are not tied specifically to publication. Universities and private sector sponsors should consider adopting policies that facilitate the distribution of publication-related data and materials.

**Recommendation 8. (Chapter Six) If an author does not comply with a request for data or materials in a reasonable time period (60 days) and the requestor has contacted the author to determine if extenuating circumstances (travel, sabbatical, or other reasons) may have caused the delay, it is acceptable for the requestor to contact the journal in which the paper was published. If that course of action is not successful in due course (another 30 days), the requestor may reasonably contact the author's university or other institution or the funder of the research in question for assistance. Those entities should have a policy and process in place for responding to such requests for assistance in obtaining publication-related data or materials.**

Few universities, research institutions, or funding organizations have published procedures for resolving problems of noncompliance by their employees or grantees. Although a telephone call to an author from a program director or other representative of an organization can be effective in achieving compliance, funding organizations and research institutions, like journals, can encourage compliance earlier in the process by developing and enforcing transparent policies that encourage sharing of research resources.

**Recommendation 9. (Chapter Six) Funding organizations should provide the recipients of research grants and contracts with the financial resources needed to support dissemination of publication-related data and materials.**

One reason that researchers have cited for not sharing published materials is the time, effort, and cost involved in doing so. This is a legitimate concern that research sponsors should

address. By supporting the development of repositories, allowing grantee institutions to recoup the costs of distribution, and through other mechanisms, funding organizations can help to assist scientists in meeting their obligations as authors. Authors should take advantage of existing ways to facilitate and minimize the costs of sharing publication-related research resources, including the deposition of research materials in existing public repositories. Some researchers have established their own "cottage industries" for producing and distributing commonly requested materials.

**Recommendation 10. (Chapter Six) Authors who have received data or materials from other investigators should acknowledge such contributions appropriately.**

Authors often fail to acknowledge those who have provided materials, data, or other information that helped in obtaining the findings they are publishing. Sharing should be recognized by citing a relevant publication of the donor of the material, and in the acknowledgement section of a paper. Another idea is to create a public database for acknowledgments. Such approaches would make it easier to recognize and reward those researchers who have been generous in sharing publication-related materials, data, software, or other information.



Community standards are not federal regulations; rather, they are self-imposed by members of the community and are sometimes incorporated in the official policies of journals. During its deliberations the committee became convinced that most arguments for making exceptions to standards could not be rationalized without sacrificing the integrity of the principles of publication. Such arguments include making exceptions to accommodate commercial interests, the original costs of producing data and materials, the vulnerability of young investigators to competition, the existence of contractual agreements with industrial sponsors, and an investigator's right to mine his or her data before others. In considering these arguments, however, the committee found that participants in the publication system were just as likely to benefit as to be hurt by sharing their data and materials. In some instances, avenues other than publication are available for investigators who want to publicize their findings while maintaining control of the related data. In other cases, reasonable and innovative ways can be found to overcome the problems of costs, contractual restrictions, and competition.

At the same time, it is expected that community standards respect laws that protect human subjects or restrict access to radioisotopes, explosives, controlled substances, and certain pathogens. The expectation that an author share publication-related materials is superseded, for example, by prohibitions imposed by many nations on the distribution of biological materials and organisms collected in those countries.

Aside from situations such as those, exceptions unfairly penalize the community, which would have otherwise had access to the data, information or material being withheld. Furthermore, granting a special exception to certain categories or particular researchers is problematic for a variety of reasons, including the difficulty of deciding who qualifies for the exception. Considering that community standards maintain quality and facilitate the work of the community in moving science forward, the committee observed that exceptions are likely to weaken the effectiveness of that process over the long term:

**Universal adherence, without exception, to a principle of full disclosure and unrestricted access to data and materials that are central or integral to published findings will promote cooperation and prevent divisiveness in the scientific community, maintain the value and prestige of publication, and promote the progress of science.**

In the committee's view, there should be a single scientific community that operates under a single set of principles regarding the pursuit of knowledge. This includes a common ethic with regard to the integrity of the scientific process and a long-held commitment to the validation of concepts by experimentation and later verification or falsification of published observations.

The focus of this report is on the life sciences, but the principles and standards considered in the committee's deliberations are of a fundamental nature. Although different fields have different accepted norms and practices, the committee hopes that its recommendations will be of interest to scientists in general and that they will prompt additional thoughtful discourse and debate in the scientific community at large.