UPSIDE: Uniform Principle for Sharing Integral Data and Materials Expeditiously

In October 2001, a National Academies committee was convened to examine the responsibilities of authorship in the biological sciences. This blue ribbon group was drawn broadly from academia, the commercial sector, and scientific publishing and was chaired by Tom Cech, president of the Howard Hughes Medical Institute.

The committee evaluated the responsibilities of authors to share data and materials referenced in their publications, the role of journals to impose requirements for data and material sharing, and whether a common set of requirements for sharing does or should exist. For example, many felt that it was inappropriate and a bad precedent to report portions of the sequence of the human genome without full disclosure (1), and the committee considered these issues broadly. In 2003 the Cech committee published the results of the study (2). The report puts forth recommendations for effective common practices and provides specific real-life scenarios to test how these practices could be applied. The committee concludes that core scientific standards accepted by the great majority of the community do exist and should be more widely adopted.

PNAS polices have long followed the core values stated in the Cech report, and we are vocal proponents of broad and transparent data sharing. We think the report should be required reading for all life scientists and journal editors. There are a few areas where we have not explicitly spelled out, as the report recommends, our policy and procedures on ensuring proper data sharing. Accordingly, we have now made these explicit in our Information for Authors, published in this issue and posted on the PNAS web site (www.pnas.org/misc/iforc.shtml).

The heart of the Cech committee report is what they termed "UPSIDE," or the "uniform principle for sharing integral data and materials expeditiously." An abbreviated form of the five core principles that make up UPSIDE are shown in Table 1. The central tenet of UPSIDE was well stated:

> 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. . . . [T]he 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 . . . but also to provide them in a form on which other scientists can build with further research. All members of the scientific community . . . have equal responsibility for upholding community standards as participants in the publication system, and all should be equally able to derive benefits from it.

Table 1. The core principles of UPSIDE

Principle 1. 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.
Principle 2. If central or integral information cannot be included in the publication for practical reasons (for example, because a data set 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, online). 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.
Principle 3. 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 the repositories by the time of publication.
Principle 4. 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.
Principle 5. 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 authors provide publication-related material they may not demand an exclusive license to commercialize a new substance that the recipient generates. Nor is it appropriate for the providers to require coauthorship simply by providing the raw materials. The merits of adopting an efficient and equitable material transfer agreement (MTA) are many, but too often the terms of an MTA are idiosyncratic and an impediment to sharing publication-related materials. As the report recommends, "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." The report contains a helpful list of dos and don'ts for MTA policy.

The Cech committee also recommends that journals clearly state the consequences for authors who are noncompliant and do not share their materials in a timely fashion. It has long been our policy that PNAS reserves the right to ban all authors of a paper from future publication in PNAS if they fail to provide requested materials in a timely fashion. Now this statement is included in the Information for Authors. It is reasonable in some cases (e.g., mouse lines) for authors to charge a modest amount to cover the cost of preparing and shipping the requested materials. This too is outlined in the Information for Authors.

The report also calls for mandatory data deposition in publicly accessible repositories. PNAS has long required protein coordinate deposition in the Protein Data Bank at the Research Collaboratory for Structural Bioinformatics or its equivalent. In addition, we were among the first journals to require nucleic acid deposition in the GenBank database. Structural information about nucleic acids must be deposited in the Nucleic Acid Database Project (NDB). Currently, PNAS strongly recommends that fMRI data be deposited at the fMRI Data Center, and we encourage the standardization of microarray data along the existing MIAME guidelines. Authors are not permitted to house large datasets on their own web site or that of their institution if the data are critical to the conclusions of the paper.

We differ from the committee recommendations in only one respect. The report proposes that an MTA term include ''a requirement that the material be used for research purposes; that is where the primary intention of the research is the fundamental increase in knowledge.'' PNAS does not endorse this restriction because we believe that it would limit the sharing of materials with industry. Requiring a company to not make commercial use of scientific results seems wrong in principle and hardly enforceable. The overriding principle is that once something is published it should be freely shared by everyone.

Scientists' responsibilities to share their data and reagents are self-governed largely by community standards where, in exchange for acknowledgement, authors provide the information essential to their findings. This, of course, is how science moves forward. Tom Cech put it well: ''The upside of UPSIDE is twofold: it keeps science honest, and it fosters the progress of science. Both are worth nurturing and protecting.''

Nicholas R. Cozzarelli, *Editor-in-Chief*

1. Cozzarelli, R. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 8159.

2. Cech, T.R. (2003) *Sharing Publication-Related Data and Materials: Responsibilities of Authorship in the Life Sciences*. Available at www.nap.edu/books/0309088593/html.