Benefits of databases

SIR—I am appalled by the comments of John Maddox (*Nature* **341**, 277; 1989) on databases, which contain a pot-pourri of excuses for inaction. While consistent policies for dealing with databanks are desirable, the variation in quality, subject matter and stages of development makes it inappropriate to consider them all collectively. Each must be judged on its merits. I am concerned with the DNA sequence databases.

The real benefits of the DNA databases will come when they are available promptly, so that a researcher can read a journal and immediately access a sequence whose properties are described. This can be achieved only if the sequence is present in the database before publication. Indeed, as Maddox points out, for a manuscript describing a sequence to be reviewed properly, the original data must be available. Moreover, a sequence in typewritten form is of little value, since even the simplest search requires a computer. It is nonsense to suggest that Indian scientists would be discriminated against if sequence submission to databases were mandatory before publication. No one, whether from East or West, can make proper use of a sequence without an electronic version and a program to assist in its analysis. A computer is every bit as necessary as a centrifuge to today's molecular biologist. Should we relax our requirements for data from scientists with only limited access to centrifuges?

The remarks about the Soviet Union are puzzling. The key issue is not where the data are collected, but whether the data deposited in the databanks are available to everybody, which they are. A scientist in the Soviet Union or India or the United States has essentially equal access to the data, by subscribing to a databank and receiving the data on a regular basis.

The thorny issue of industry versus academy is a red herring. The requirements for publication of sequence papers should be the same for both academic and industrial authors. Methods should be described in sufficient detail so that another researcher can reproduce them. The data underlying the conclusions drawn must similarly be available to the readers. If the manuscript describes a sequence, then it is imperative that the sequence be available. If industry wishes to keep its sequences secret, then it should not publish. Worries about the improper commercial use of nucleotide sequence data seem unfounded. If programs in the public domain are not sufficient to allow analysis of the sequences, then, of course, companies should seize the opportunity and sell software that is capable of that analysis. This is the essence of a free-enterprise system. It must not be a reason for failing to make the databanks as comprehensive as possible.

Scientific journals depend for their existence upon the scientists who perform research just as much as scientists depend upon the journals to promulgate their results. So far, this symbiosis has been a healthy one. However, times are changing. Computers are here to stay and soon journals themselves must be available in electronic form. Equally, the data encapsulated within each scientific article need to be presented in ways that the scientific community can understand and digest. The present system worked well when there were only a few scientists and fewer journals. But electronic databases are now an integral component of the scientific enterprise. Scientists and journal editors have a responsibility to steer the publication process through its next evolutionary stage. A natural and appropriate step in this process will be for the journals and their contributors to join together in ensuring that the sequence databases become the timely resource that they should be. It is a simple matter to require that the author of a sequence paper provides a database accession number for that sequence as a prerequisite to publication. Indeed, this is the only point in the process at which it can be enforced and it is appropriate to do so. To require the granting agencies to police this scheme is unwieldy and impractical. Only when authors cheat by inventing accession numbers should the granting agencies become involved.

RICHARD J. ROBERTS
Cold Spring Harbor Laboratory,
Cold Spring Harbor,
New York 11724, USA

SIR—The staff of the Protein Data Bank (PDB) read with great interest "Making good databanks better".

The PDB, an archival database for three-dimensional structures of biological macromolecules, was established at Brookhaven National Laboratory in 1971. With financial support from the US National Science Foundation, the National Institutes of Health and the Department of Energy, the PDB is compiled and made available to a broad user community worldwide. Currently, the PDB is distributing data for more than 450 structures, with about 140 additional structures in the process of being input. The data input process includes extensive checking by PDB staff, with feedback to depositors allowing for the correction of errors. PDB depositors and users alike find that this quality control constitutes a valuable service.

There is a wide variation in the policy of scientific journals regarding the deposition of crystallographic data, and the PDB relies to a substantial degree upon voluntary contributions. However, as indicated in the article, journals have a right to require that essential supporting data accompany manuscripts submitted for publication and should also arrange to provide access to these data. In the case of crystallographic studies of macromolecules, the only practical means to provide such access is by submission of the data to a database, since the information (atomic coordinates and structure factors) is far too voluminous to be useful in hard-copy form. In recognition of the above considerations, the Commission on Biological Macromolecules of the International Union of Crystallography has recently endorsed a policy that publications should be accompanied by deposition of the appropriate data in the PDB. The policy provides the option for a delay in the release of the deposited data of up to one year from the date of publication for coordinates and up to four years for structure factors, reflecting concern that results from the early stages of analysis will be inaccurate in detail and that investigators should have the opportunity to complete the analysis and interpretation of their data.

Nature should reconsider its policy of not requiring deposition of data in the appropriate databases. In addition to ensuring maximum availability of the information to the scientific community, archiving by the databases removes the very real possibility that essential information will become lost over time.

THOMAS F. KOETZLE
Department of Chemistry,
Brookhaven National Laboratory,
Upton, Long Island, New York 11973, USA

Reversed charges

SIR—One method to slow rising journal costs (*Nature* **341**, 349–350; 1989) is to put more of the publication costs on the authors, because they receive the major benefits of their publications. Many learned societies already have page charges and lower subscription costs than commercially published journals. Page charges should be standard for all journals because publication costs are a legitimate part of a research project. Publication is the final and most important step in a research project as it satisfies the 'publish or perish' directive and places the data in the scientific record.

DAVID R. HERSHEY
Department of Horticulture,
University of Maryland, College Park,
Maryland 20742–5611, USA

■ There are strong objections to page charges in international journals whose contributors work in varied circumstances—Editor, *Nature*. □