INFORMATION ACCESS:

Building A "GenBank" of the Published Literature

Richard J. Roberts,* Harold E. Varmus, Michael Ashburner, Patrick O. Brown,† Michael B. Eisen, Chaitan Khosla, Marc Kirschner, Roel Nusse,† Matthew Scott,† Barbara Wold

Since the time of the great library of Alexandria, scholars have recognized the value of central repositories of knowledge. As scientists, we are particularly dependent on ready and unimpeded access to our published literature, the only permanent record of our ideas, discoveries, and research results, upon which future scientific activity and progress are based. The growth of the Internet is changing the way we access this literature, as more scientific journals produce online editions to supplement or replace printed versions. We urge journal publishers, their editors, and all working scientists to join together to create public, electronic archives of the scientific literature, containing complete copies of all published scientific papers.

Anyone who has spent time in a library searching for a key paper, result, or method will immediately see one of the benefits of comprehensive repositories. Those gems of information that are often buried within papers, but are not referred to in the abstract or keywords, will become readily retrievable. You will be able to locate descriptions of methods or find the original data that underlie crucial conclusions. You will be able to trace connections between observations originally scattered among many papers in different journals and databases. However, the value of central archives goes well beyond facilitated

Summary of this Article

dEBates: Submit a response to this article

Published dEBates for this article

Related articles in Science

Similar articles found in:

SCIENCE Online
PubMed

PubMed Citation

Search Medline for articles by:

Roberts, R. J. ||
Wold, B.

Search for citing articles in:

ISI Web of Science (4)

Alert me when:

new articles cite this article

Download to Citation Manager

Collections under which this article appears:

Scientific Community

Science and Policy

http://www.sciencemag.org/cgi/content/full/291/5512/2318a 2/25/02
searching and retrieval. Bringing all of the scientific literature together in a common format will encourage the development of new, more sophisticated, and valuable ways of using this information, much as GenBank has done for DNA sequences.

Some have argued that central repositories are of no additional value because many journals already make their online contents freely available after some delay through their own Web sites. However, material that is freely accessible, on a controlled basis, one paper at a time, at a journal’s Web site differs from material that is freely accessible in a single comprehensive collection. The latter can be efficiently indexed, searched, and linked to, whereas the former cannot. Imagine how much less useful DNA sequences would be if instead of GenBank and other global repositories, we had dozens of smaller sequence collections that could only be accessed one at a time through a genome center’s Web site. Only by creating repositories with uniform, explicitly defined, and structured formats, can a dynamic digital archive of life science research literature become possible. Unimpeded access to these archives and open distribution of their contents will enable researchers to take on the challenge of integrating and interconnecting the fantastically rich, but extremely fragmented and chaotic, scientific literature.

To ensure that complete public scientific archives become a fully workable reality, the necessary infrastructure must be constructed. The National Institutes of Health has taken an important step by creating PubMed Central (PMC) (1) with the goal of storing the life sciences literature in digital form and providing free and convenient access, linked to the popular bibliographical database, PubMed. We envision PMC as only the first of many public archives. However, such archives will not realize their potential until they are populated. This requires that journal publishers allow their digital content to be distributed and used through online public archives. Several journals, including the Proceedings of the National Academy of Sciences, the British Medical Journal, Nucleic Acids Research, Molecular Biology of the Cell, and the BioMed Central (2) journals, have already agreed to deposit their content with PMC, following, at most, a short delay after print publication. Publishers now have a wonderful opportunity to reinforce their partnership with the scientific community by supporting extant archives like PMC and by allowing archival material to be freely used and distributed, and we strongly urge them to do so. It would be natural and simple for journals that have already decided to make their back issues freely accessible at their own Web sites to make the same content available in electronic archives. The costs of participating in open archives would be minimal and would be more than offset by the benefits their participation would bring to the scientific community.

Historically, publishers have left the job of archiving to the libraries. Library archives have become more accessible as we have moved from indexed abstract books to rapidly updated online abstract searching tools. Public online archives should be viewed as the logical continuation of this tradition and, thus, as a complement to the publisher's
normal activities. For electronic archives to assume this role fully, decades of volumes that currently exist only in printed form will need to be digitized. We do not expect journals to bear the cost of the digital conversion of their printed archives. Indeed, efforts to raise the necessary funds are under way, so that digital conversion of archival volumes can proceed rapidly.

It is important not only that PMC succeed, but also that other institutions be encouraged to provide independent online sites for the distribution and use of the same comprehensive archives. Multiple independent online sites will help ensure ready access for users around the world and will guarantee that no single government or institution can control access to our common scientific heritage. This diversity will also foster innovation in the ways the material in the archives is used.

We feel that if journal editors and publishers were to poll their authors and readers, they would find overwhelming support for such archives. The strength of this support is demonstrated by the growing list of scientists who have signed an open letter (3) advocating free and unrestricted distribution of scientific literature 6 months after publication. We urge our colleagues, especially students and the younger members of the scientific community, to make your views heard. If these efforts are successful, in 10 years, everyone's ability to do science will have been greatly enriched, and we will all wonder how it was possible to work without such archives.

References and Notes

2. [www.biomedcentral.com](http://www.biomedcentral.com)
3. [www.publiclibraryofscience.org](http://www.publiclibraryofscience.org)

R. J. Roberts, New England Biolabs, Beverly, MA 01915, USA. H. E. Varmus, Memorial Sloan–Kettering Cancer Center, New York, NY 10021, USA. M. Ashburner, University of Cambridge, CB2 3EH, UK, and EMBL-European Bioinformatics Institute, Cambridge, CB10 1SD, UK. P. O. Brown, Stanford University School of Medicine, Stanford, CA 94305, USA. M. B. Eisen, Lawrence Berkeley National Laboratory, Berkeley, CA 94720, and University of California, Berkeley, CA 94720, USA. C. Khosla, Stanford University, Stanford, CA 94305, USA. M. Kirschner, Harvard Medical School, Boston, MA 02115, USA. R. Nusse and M. Scott, Stanford University School of Medicine, B. Wold, Biology Division, California Institute of Technology, Pasadena, CA 91125, USA.

*To whom correspondence should be addressed. E-mail: roberts@neb.com

†Howard Hughes Medical Institute investigator.
Summary of this Article

**dEbates:** Submit a response to this article

Published dEbates for this article

Related articles in Science

Similar articles found in:
- SCIENCE Online
- PubMed

PubMed Citation

Search Medline for articles by:
- Roberts, R. J.
- Wold, B.

Search for citing articles in:
- ISI Web of Science (4)

Alert me when:
- new articles cite this article

Download to Citation Manager

Collections under which this article appears:
- Scientific Community
- Science and Policy

dEbate responses to this article:

Read all dEbate responses

**Redundancy of Scientific Literature**
Frank Edgcombe

http://www.sciencemag.org/cgi/content/full/291/5512/2318a
SCIENCE Online, 28 Mar 2001 [Response]

A European Perspective
John Sack

"Public Domain Knowledge," not a Private Property Right
Sterling Stoudenmire

Central Archives at Individual Institutions
Joan Green

No Literature GenBank Now
Alex Bienkowski

Sharing Personal Libraries
Chris Seidel

A Look Libraries
Albert Henderson

Parallel Debate in Nature
Stevan Harnad

Journal Profits
Diane Lester

Fee or Free?
Johannes Velterop

A central repository requires grassroots participation
Michael Mauws

Related articles in Science:

Considerations in Creating Online Archives.
Ira Mellman
Science 2001 292: 51. (in Letters) [Full Text]

SCIENCE'S RESPONSE:
Is a Government Archive the Best Option?.
The Editors
Science 2001 291: 2318-2319. (in Viewpoints) [Summary] [Full Text]