The future of electronic data

Will the universities' own electronic repositories affect traditional publishing?

Ann J. Wolpert

ILLUSTRATIONS BY DAVID NEWTON

"Will you manage my digital work?" is a question that twenty-first century university librarians are increasingly being asked. The answer might once have been "Sorry…", but today both university library services and university research and educational output are in the midst of fundamental change. The current response is more likely to be: "Let's talk."

At the risk of stating the obvious, the complex system of relationships and products known as scholarly communication is under considerable pressure. New information technologies (digital formats, the Internet, laptop and desktop computing, data and image capture and manipulation) have created opportunities for communication that were unimaginable in an earlier, print-constrained era.

These information technologies hold great promise for positive change in the ways that scholars, researchers and educators conduct their work. But they have also destabilized the economics of a highly complex communication system. And they challenge some basic assumptions concerning long-held roles in the value chain of traditional scholarly output. All of the stakeholders in the system are currently living in a period of experimentation, and it is not at all clear how the system will look when the dust settles.

All change

Although there is general agreement that the system of scholarly communication must change in response to new information technologies, conversations about the form that any such change might take resemble those of the proverbial blind men describing the elephant. Journal publishers view the situation in the context of revenue and production values. Authors describe their intellectual contributions and usage expectations. Librarians articulate concerns about the scholarly record and the challenge of providing short-term and long-term support to education and research. Scientists have one set of assumptions, humanists another. And the institutions that host authors, and libraries, and sometimes publishers, are trying valiantly to get a grip on the overall costs of the system. Conflicting perceptions mean that change that seems obvious to one group of stakeholders feels counterproductive to another.

Until recently, faculty and their institutions have viewed this stalemate as an annoyance rather than a threat. Networked environments had to be built anyway. Librarians seemed able to manage rising costs through judicious cancellations and collaborations with other institutions. Faculty fussed but generally managed to teach and conduct research much as they always had.

Two events have changed this annoyance to alarm. First, teaching moved onto the Internet, and faculty and their institutions were forced to confront in a meaningful way the operational constraints of the intellectual-property environment as it affects digitally formatted materials. These legal regimes, established primarily to protect digital entertainment products in a global, networked communications environment, map poorly, if at all, to the time-honoured ways that researchers and educators teach and advance knowledge.

Second, bundled, restrictive pricing models for licensed access to formal scholarly publications began to draw the attention of university administrations. These pricing strategies significantly reduce the flexibility of institutionally based libraries to absorb price increases, and adversely affect the ability of libraries to manage the array of inter-institutional educational and research collaborations that have become increasingly important to faculty and students.

These events stirred faculty and universities to begin a series of experiments designed to further test alternatives to traditional publishing. The experiments reflect a concern with the rising cost and declining flexibility represented by conventional publishing, and may even have been initiated for those reasons. More importantly, they illuminate the evolving nature of research and research communication. There is a growing awareness of the fact that, particularly in the high-visibility emerging disciplines, the nature of research and research documentation is changing. Examples include the increased reliance on large data sets or image sets across a variety of disciplines, and the increased commonality of large-scale data analysis. Similarly, extensive supportive documentation is often expected, whether the work is in the sciences or the liberal arts.

Built to serve

Among the more visible experiments are projects to leverage the intellectual output of faculty and institutions through the use of e-print servers and institutional repositories. E-prints are generally understood to be author or institutional archives of electronic versions of scholarly or literary works made accessible through the global Internet. E-print servers seek to address the needs of faculty who work primarily in text (working papers) and need a timely and easy way to share research and ideas across a community of interest.

It is far too early to know whether or how the self-archiving e-print movement will affect the conduct of research and teaching, or of traditional publishing enterprises. E-print servers have so far taken root in only a few disciplines, such as economics, physics, mathematics, computer science and the cognitive sciences. Many of these e-print services, such as CogPrints at the University of Southampton, UK, have been owned and operated in the conventional cottage-industry mode of traditional higher education. In general, these e-print services lack both scalability and a commitment to long-term availability.

A necessary precursor to self-archiving will be a computing environment that is far

© 2002 Nature Publishing Group

more intuitive for the average faculty member than is currently the case. And faculty are not normally eager to function as system administrators. Worth noting is the recent relocation of the seminal 'arXiv.org' e-print server from the Los Alamos National Laboratory to Cornell University. This grandaddy of all archives moved with its creator, Paul Ginsparg, and is now maintained on a server at the Cornell University Library.

Of potentially greater long-term impact to scholarly communication is the body of non-text research and scholarly output that is growing rapidly and inexorably on university campuses. The typical doctoral-level research university currently supports considerable institutional computing capabilities, and its faculty produce and manage data, images, video, sounds, simulations, animations and a host of other media and genres. As the computing costs and support requirements of these resources proliferate in academic departments, academic institutions will be increasingly motivated to find appropriate economies of scale.

Digital data

This makes institutional repositories a more likely long-term outcome than the individually sponsored e-print server. Indeed, the current proliferation of institutionally sponsored digital repository services suggests that a variety of experiments are already well underway. The California Digital Library is developing a suite of services that should empower the entire University of California system by providing tools to house and distribute the intellectual output of university faculty.

The DSpace initiative at the Massachusetts Institute of Technology Libraries (an open-source initiative that will be implemented in a federation of between five and seven North American university libraries within a year) is designed to host and preserve a variety of text and non-text formats (see *Nature* **419,** 869; 2002). In Britain the Cambridge University Library is particularly interested in developing the capability for managing digital teaching objects on behalf of faculty and the university. And the digital collections at the California Institute of Technology (Caltech Digital Collections), Indiana University (Digital Library of the Commons) and the University of Michigan (OAIster) are yet more examples of content-management services designed to support the work of faculty.

Protocols and standards for these repositories are being established through the efforts of such organizations as the Scholarly Publishing and Academic Resources Coalition, the Open Archives Initiative, the World Wide Web Consortium, and Creative Commons. Funding sources thus far include national governments, foundations and individual institutions. Almost universally, institutions are turning to their research libraries

for assistance in designing and implementing these new information services, as the work of the Digital Library Federation and the Coalition for Networked Information will attest. After all, academic research libraries have a long tradition of leveraging institutional investments, supporting the faculty enterprise, and developing sustainable preservation strategies for a variety of content formats.

A companion set of services is emerging as options for the creation of publications from large institutional repositories. Berkeley Electronic Press, for example, offers the University of California and other institutions the possibility of producing formal peer-reviewed publications through their university presses, as well as the alternative of simpler online publication strategies for those who do not need or want peer-review capabilities.

A slightly different model is Project Euclid at Cornell University. Euclid is a set of tools designed to advance scholarly communication by addressing the unique needs of low-cost independent and society journals. An alternative to traditional publishing, Euclid provides robust functionality through full-text searching, reference linking, interoperability through the Open Archives Initiative, and long-term retention of data.

Peering into the future

For the foreseeable future, it is likely that traditional peer-reviewed journals will persist in their historical niche of documenting the record of advances in disciplines. Well-regarded and heavily-used e-print services in a variety of subject areas have yet to eliminate peer-reviewed journals as the tool of choice for the permanent record of a discipline. Likewise, faculty will not lightly abandon an evaluation system that has served them reasonably well for centuries. The importance of peer recognition cannot be overstated, whether it be in the form of acceptance for publication or in the honour associated with participation in the editorial process. As peer-reviewed journals themselves move online, however, and their print volumes disappear from library shelves under preferred licensing arrangements, it will be interesting to see whether their current market distinction becomes less obvious and less relevant.

The most interesting aspect of institutional repositories will be in the alternative forms of communication that emerge as faculty begin to exploit the capabilities of an open, interoperable system. Disciplines will have the unprecedented luxury of designing new scholarly communication systems that reflect the ways they need and want to teach and conduct research. As the volume of material in institutional repositories grows, and harvesting protocols enable the revelation and repackaging of works in new ways, the ideas of the next generation of faculty will no longer be bound by the constraints of traditional print. They may even reinvent the mechanics of peer review to take advantage of repository tools and functionality.

Meanwhile, this is a time of experimentation, and no one can be confident that their predictions will hold true. Only two outcomes look certain: first, that the status quo will not remain unchanged; and second, that research institutions have some fairly compelling reasons to encourage the availability of scholarly information and documentation on the Internet. The experiments currently underway will help to determine whether there are equally compelling reasons to provide those online information resources with the benefits of management within persistent institutional repositories. ■

Ann J. Wolpert is at Massachusetts Institute of Technology, 77 Massachusetts Avenue, 14S-216, Cambridge, Massachusetts 02139, USA.

The Case for Institutional Repositories
- **www.arl.org/sparc/IR/ir.html**

DSpace
- **http://dspace.org/index.html**

EPrints
- **www.eprints.org**

The Berkeley Electronic Press
- **www.bepress.com**

University of Michigan OAIster
- **http://oaister.umdl.umich.edu**

MIT's ArchNet
- **http://archnet.org/lobby.tcl**

California Digital Library
- **http://escholarship.cdlib.org**

Open Archives Initiative
- **www.openarchives.org**

Research Papers in Economics (RePEc)
- **http://econpapers.hhs.se**

ArXiv.org e-Print archive
- **http://arXiv.org**

Caltech Collection of Open Digital Archives (CODA)
- **http://library.caltech.edu/digital/default.htm**

Indiana University Digital Library of the Commons
- **http://dlc.dlib.indiana.edu**

University of Southampton CogPrints
- **http://cogprints.ecs.soton.ac.uk**

© 2002 Nature Publishing Group

Copyright © 2002 EBSCO Publishing

Copyright © 2003 EBSCO Publishing

Copyright © 2003 EBSCO Publishing

Copyright © 2003 EBSCO Publishing

Copyright of Nature is the property of Nature Publishing Group and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.