

The best and worst of times

What winners will emerge from the battles over access to scholarly data?

David R. Worlock

The Internet has caused a revolution in research. Scientists now work remotely from one another but in tightly organized groups, with an immediacy of communication that demands instant availability of the corpus of knowledge on which their work is based. In the process, the underpinnings of a system of filtering and selecting results that has been effective for the past 150 years have been questioned and shaken, and the roles of the actors — publishers, librarians, scholarly societies and universities, as well as researchers — are being redefined.

This is not a picture of the manoeuvrings of powerful vested interests confined to a cul-de-sac in the progress of communication, as some have suggested. Rather, it is the proving ground for the future as we evolve from an information-rich to a knowledge-rich society. The patterns of behaviour created by scholarly research publishing will continue to evolve, setting the standards for communication in the business and professional worlds.

Let battle commence

When 28,000 researchers this year signed the Public Library of Science petition (www.publiclibraryofscience.org) demanding that publishers post scholarly articles on freely available, centralized servers six months after publication, a battle — both real and virtual — had begun. The clamour has been strongest in the 'hard' sciences. Andrew Odlyzko, in the current *Nature* web debate (www.nature.com/nature/debates/e-access/Articles/odlyzko.html), reports a story: following an e-print article about a new high-temperature superconductor, "every superconductivity laboratory in the world immediately began to make measurements on this new material and dash into print. Fifty e-prints had been posted on the web by the end of February — before the original paper was even published." In other words, conventional publishing is too slow. There is an inexorable pressure to connect the e-print servers of the world and to create a two-speed scholarly communication economy.

Do these changes undermine traditional, paper-based journals? Scholars still rely heavily on peer-reviewed publication. As authors, of course, they insist on appearing in the most prestigious, branded outlets, even if, as users, their views can be subtly different. Major studies in the United States (see overleaf) indicate that, although print is not under threat, a culture of reading outside core publications is growing; document-

delivery ordering is increasing by as much as 7.7% per annum (Association of Research Libraries research; www.arl.org). This startling growth is occurring as librarians under budgetary constraints reduce subscriptions and use interlibrary loan facilities to the fullest. Indeed, scholars may now be obtaining as many articles from ordering systems as they used to read via their institutions' subscribed publications.

The university and the library are being sharply redefined, in both the terms and style of access by scholars, and in the ownership of tradeable intellectual property in the form of copyrights. Some universities — the Massachusetts Institute of Technology is in the lead — already sell courses on the web. The idea of university consortia trading in the intellectual property created by their researchers is not far behind. And it follows that, if distributed storage but centralized searching is the model, libraries have no role.

Although subscriptions to printed journals still form 35% of average library budgets, document acquisition is 8% and rising. Librarians are evolving into powerful network-resource administrators. Years of library consortia and licensing schemes underline the power of collective bargaining (some publishers report up to 60% of sales to consortia rather than individual library purchasers), and remind libraries that if they have the intellectual property they can be traders, too.

As in every period of rapid change, there are also losers. Outside the consortia, and in the less-developed world, a genuine poverty of access is emerging as never before, with the scholarly rich and poor divided sharply on access and on the ability to stay abreast of the fast-moving research base.

The world of print journals is underpinned by commercial publishing, either via branded journals or by contract publishing for scholarly societies, whose income derives from journal subscriptions. Surprisingly, it is hard to determine the total number of journals: it could exceed 50,000, but there is a profitable core of 12,000–15,000, with publishers maintaining their profit margins by increasing prices when subscription numbers fall.

Rapid consolidation creates economies of scale, prevents players from being cut out by new intermediaries, and accounts for the reinvestment forced on publishers by expen-

sive digital conversion and new formats such as XML. The big are likely to get bigger and the small squeezed.

Publishers' response

How are publishers coping with the changing behaviour and demands of researchers? After a period of denial, a collaboration emerged in the late 1990s that would have been unthinkable a generation earlier. The fruits were the Digital Object Identifier Foundation (www.doi.org) and the CrossRef service (www.crossref.org) — attempts to use metadata (data about electronic data) to improve access to journal articles wherever they were held in distributed publishing systems.

All this begs the ultimate question: who owns the customer? All publishers want to know the answer, so they can match changing habits. Smaller publishers may be able to use CrossRef's search facility to stay on researchers' radar, but librarians and administrators want to make smart deals with major suppliers and intermediaries to secure the best terms.

It may be too late. The Open Archives Initiative promises a lowest-common-denominator universal metadata standard (www.nature.com/nature/debates/e-access/Articles/harnad.html). The Office of Scientific and Technical Information of the US Department of Energy is typical of 'big government' in desiring to provide unifying access standards across thousands of distributed e-print sites in universities. Scholars are looking to the next major change: knowledge representation and the addition of RDF (resource-definition framework) standards to current XML metadata. This will enable ontologies to be used that will allow searching on knowledge structures, not simply on terms and words, in turn creating a new standards debate, as domain-driven ontologies compete at the borders of their disciplines.

Meanwhile, the shape of the article itself as a reporting mechanism is changing. The addition of research files, results databases, software environments for running comparative results, laboratory videos and other materials heralds the day when the article becomes the core knowledge document to which archival and grey literature can be referenced and linked. ■

David R. Worlock is chairman of Electronic Publishing Services Ltd, 26 Rosebery Avenue, London EC1R 4SX, UK.

