

## TRENDS IN SCHOLARLY PUBLISHING

*Paul Sheehan*

Paper based on a presentation at a Workshop on Copyright, under the aegis of EBLIDA (European Bureau of Library, Information and Documentation Associations), Dublin, February 1995

*Two forces of change are at work in the academic serials publishing industry. One is the increase in the number of papers and in journal cover prices, to levels where the publication process effectively rations distribution. Secondly, information communications technology has the potential to allow radically different forms of publication. Already some academic groups have begun to publish (unrefereed) papers on pre-print servers operating over the Internet, a publishing model with many inherent problems.*

*Paul Sheehan is Sub-librarian, monographs and serials management, Trinity College Dublin Library, Dublin, Ireland*

It is generally agreed that there is a crisis in serials costs in academic libraries. At the same time, network publishing technologies offer alternative means of publication. However, alternative technology grafted onto the existing pricing structures will not solve the cost crisis for academic libraries. Some cost must be removed from the publication process. A number of separate initiatives, some from within the academic community, may indicate the development of new publishing models.

Copyright in journal papers has traditionally passed smoothly from originator to publisher. The publisher's acquisition of this intellectual property was the quid pro quo for the range of tasks - refereeing, setting-up, printing, distribution - required to bring the paper, after scrutiny, to the attention of a specific readership. Two forces are altering the traditional landscape of property rights in academic publishing.

One is the rate of increase in the volume and price of academic papers. The number of science papers has doubled about every 10-15 years for the last 200 years. But this rate of acceleration actually increased after the second World War, so that in mathematics, for example, the number doubled about every 10 years up to 1990. In other words, half of all mathematics papers were published in the last 10-12 years. Some 850 mathematics papers were published in 1870; about 50,000 are now published each year.<sup>1</sup>

As a result journal titles have proliferated and, in the last decade especially, their cover prices have also increased at a rate usually in excess of the Retail Price Indices (RPI) in either producing or consumer economies in the western world. Publishers have argued that the volume of research is driving the volume of publications, and that price increases reflect increases in the number of pages published. There is data to support the first part of this argument. In the USA between 1976 and 1990 libraries of the Association of Research Libraries Group saw a 50% increase in central funding, while funding for research increased by over 200% in the same period.<sup>2</sup> However, Gillian Page's analysis of the economics of journal publishing indicates that this need not lead to pro rata price increases: "Increases in content do not increase overhead proportionately, so the journal can either offer more pages per dollar or Deutschmark, or be more profitable to the publisher, or

both".<sup>3</sup> Hamaker and Grinell demonstrated a strong correspondence, in 23 Elsevier titles, between increases in the number of pages and price increases.<sup>4</sup>

There is some interesting evidence regarding the pricing policies of commercial and non-profit publishers. John Christensen, analysing chemistry journals at Brigham Young University, concluded that in 1990, while equal numbers of journals were bought from commercial and non-profit publishers, commercial publications cost about four times as much as non-profit publications. Further, when taking into account an impact factor based on ISI citations, "commercial journals were four times as expensive as the non profit journals at providing impact". Between 1990 and 1993, "the cost per million characters for [each of two sets of 9 titles] rose an average 12.4% a year for the commercial journals but only 6.3% annually for the non-profit journals".<sup>5</sup>

Whatever the reasons, the consequence is that there are now few general academic libraries which can support subscriptions to all relevant journals in their main research and teaching areas. In fact few academic libraries have been able to maintain a steady state in journal subscriptions.

The second force of change - developments in information and communications technology - is broadening the options for information exchange in the academic community. Initially, it has allowed information vendors and libraries to investigate alternatives to the print medium. The simplest variation is in the supply of information on CD-ROM disks. This simple physical substitution has, however, been accompanied by a change in terms of contract. In certain circumstances this has meant that, whereas libraries buying printed versions acquired ownership with restricted rights of reproduction (governed by copyright legislation), the contractual terms governing use of the CD-ROM alternative give licensed use for the term of the subscription, not ownership, so that when a subscription ceases, access may be lost.

Communications technology has enabled an access versus holdings strategy; or a just too late versus just in case strategy, as it has also been called. CAS-IAS services usually involve a large database of indexed information backed up by the supply of photocopied or faxed documents. The problem here is that libraries and publishers both

see these services as being a solution to their share of the crisis in serials costs. Libraries hope that individual article supply will be a cheaper way of obtaining relevant information than annual subscription. Vendors see it as another revenue stream in a world where even a big publisher, such as Elsevier, loses 3% of its subscription base annually.<sup>6</sup> Pricing policies indicate that the aims of the two main participants are irreconcilable. A document price has, typically, two main components, a processing charge and a copyright clearance charge. There have been some claims of publishers increasing the copyright clearance charge as demand increases.

The area where communications technology threatens to have a really significant impact is not in the management of existing printed material, but in network publishing. Some participants in the debate see this as an opportunity to rectify a situation where a researcher, doing publicly funded work, is involved in a validation and distribution process whereby both the researcher and the parent institution pass property rights in that work, for little financial reward, to a publisher and where limited rights to similar work can only be obtained at very high cost. There is a case for saying that the current economic model of academic journal publishing, especially but not exclusively in the area of Scientific, Technical and Medical (STM) journals, has failed.

The publication process is both demand and supply driven. Publication is an essential part of research, particularly in those fields where research is expensive, and projects are individually funded. Pressure for publication, especially in the STM sector, merely increases the number of already expensive units of cost, the pages published, to the consumer. So the total cost of this literature increases remorselessly, whilst the purchasing capacity of consuming institutions falls consistently. The experience of Christensen at Brigham Young University, where journal costs increased by an average of 10% over the last decade, while budgets rose by 3-6%, is common on both sides of the Atlantic.<sup>7</sup> Instead of widening the distribution of the growing numbers of academic papers, this model increasingly rations them.

The working of these two forces has been described as the push of economic imperatives, and the pull of technical possibilities. Since the latter effect is quickly gaining in strength, the questions raised concern not so much alternative communications media, as alternative publishing models. The model of the printed paper in a learned journal was based on the necessity of recognising specialised functions. Publishing required a commitment of significant capital, and it required specific technical and commercial skills. The publisher, whether learned society or commercial, added the value - achieved through refereeing, sub-editing, setting-up, printing and distribution - required to put papers of a certain standard before a specific readership. This model is now disintegrating. Rudimentary formatting, and sub-editing can be done on a PC. Distribution can be achieved through an Internet connection. The readership defines itself. Therefore, much of the value provided in the printed publication process can now be added more cheaply by other means on the networks.

It is now possible to address almost the whole academic community in the western world, on the Internet. Although much of this traffic is qualitatively different from the publication process which sees a formally refereed paper printed in a journal of repute, there are signs that the printed communication may already be giving way to electronic communication in the marginal areas of formal information exchange.

One of the most interesting of these is the physics pre-print server set up at Los Alamos by Paul Ginsparg. Both Ginsparg<sup>6</sup> and Andrew Odlyzko<sup>9</sup>, in recent articles, state that pre-prints have already become the main method by which experts in some areas of physics and mathematics communicate their latest results to each other. Ginsparg originally set up the server for a high energy physics database, in 1991. This database was receiving 12,000 requests a day by October 1994. A number of other databases have since been formed, so that over 40,000 transactions per day are now being processed on these databases.<sup>10</sup> This process has been described by Ginsparg as a very formal method of communication, with automated submission and indexing; he asserts that in some areas covered by these databases, the conventional journal has been entirely supplanted.

There is substance in Ginsparg's argument that the printed journal is now a clumsy artefact for scholarly communication. The formally refereed article in a reputable journal has limited value as a current awareness tool, particularly in the sciences. Researchers in highly competitive fields rely more on informal methods, principally personal contacts, and to a certain extent pre-prints, to find out where the competition is and how it is faring. There is evidence that the conventional printed form is a strait-jacket. When new printing technologies have become available, this has led to information being presented in a different way in some disciplines, for example, when coloured images became possible in biology journals, the quantity of text diminished substantially.

Further, there is increasing exploitation in scientific papers of data presentations which are beyond the capacity of print. Much of the data relevant to biology and molecular biology papers, for example, cannot be contained within the print medium, and needs to be presented on other media - microfiche or compact disks - which accompany the journal. A more striking example is provided by the Brookhaven National Laboratory in the USA. In its Protein Crystallographic Database, it maintains data files of the co-ordinates of the atomic structure of each of a particular group of proteins.<sup>11</sup> Although a paper describing a new structure will be published in a print journal, the co-ordinate data can only be exchanged over the networks on fast computers, (and may not be governed by the copyright protecting the print version). This data is freely available to any interested researcher on the networks. Furthermore, although it is free, this, and not the printed work, is the primary data, despite the fact that its deposition, while controlled, is not refereed in the usual sense. It is developments such as these which lie behind this quotation from Ginsparg: 'The resources necessary for production and distribution of conventional printed journals allowed publishers to focus on the mechanism, and to avoid any pressure to rethink the intellectual content and quality of the operation'.

Odlyzko gives an example of such a re-think - a model where a pre-print server would be a first stage of publication. In this model papers would acquire clusters of critical comment, which would

form part of the paper. In due course the paper would be refereed, and the referee's comments might also be archived with the paper. The refereed version might also attract comment following subsequent research, and some of this might become part of the paper. His concept is of the paper as a continuum, rather than a discrete bibliographic artefact. He sees this activity being run solely by the scholarly community, without the intervention of publishing intermediaries. The implications for copyright are not clear.

Copyright might vest in parent institutions, or it might be scattered among all intellectual contributors in a unique form of chaos. More fundamentally, copyright would operate not as an instrument to protect commercial publishing interests but rather as the means of protecting the integrity of intellectual works, put into a network environment where corruption could easily occur.

These ideas are not new. Sir Peter Swynnerton - Dyer put forward the idea of a general electronic journal which would be created and owned by the academic community. The Science and Engineering Policy Studies Unit was investigating this as well as other models.

Although a coherent model of network publication has not yet been developed, a great deal of formal information is now available over the network. Much of the information on which this article is based was obtained over the networks at no item cost. Odlyzko's and Ginsparg's articles appeared in conventional journals but copies can also be obtained by e-mail, FTP, or on the WWW. The only condition of use, in Ginsparg's case, was that no particular form of his paper was re-transmitted over the network, as he was making continuous amendments. Most of my references were obtained from the JUGL Bulletin Board for Libraries, again at no direct cost, but which effort is supported by official and commercial grants, and by the significant voluntary efforts of information professionals. Background information came from the *Newsletter on Serials Pricing Issues*<sup>12</sup>, mailed over the Internet free, and created from information supplied and formatted voluntarily by information professionals. In most of these cases, copyright has not been transferred to commercial or learned publisher but has remained the property of individuals.

Another point which arises is that Ginsparg and Odlyzko are academic researchers but they are contemplating new publishing models without other intermediaries, especially publishers, on the horizon. As Odlyzko points out, technological developments will mean that such groups may gain increasing control of their own body of literature. Currently, a PC may have a hard disk capacity of 1Gb. Within ten years that may have increased to over 100Gb. The whole corpus of mathematics papers, in encoded text, can be accommodated on 9Gb of disk capacity.

However, some of the opinions of Ginsparg and Odlyzko need to be treated with care. For example, it seems not to be generally true in physics as a whole that pre-prints are the main method of communicating the latest research, even though the pre-print server may well be a precursor of new self-publishing initiatives, particularly in the sciences. For the time being it remains a fact that academic papers, such as those from which these pre-prints are derived, will ultimately appear in a conventional journal - conventional, that is, in its bibliographic structure, and its economic form - whether in print or electronic format. Also, there are still a range of impediments to the model suggested by Odlyzko.

One concerns values related to context and production. Currently, the status of a printed paper may be closely related to that of the journal in which it appears. A similar value context has yet to be established in the electronic environment. Text encoding, sub-editing and related production values will be important in the electronic environment, and may present technical and financial problems. The maintenance of an archive of papers in their varied states of development may present problems of management.

Another concerns transmission. The Super Journal Project<sup>13</sup> which looked at technical models of electronic journal publishing on SuperJANET, established that network speed is critical. A fully functional electronic journal in a host-based model would need to display not just characters, but also graphics and formulae in colour. One conclusion of the project was that although the SuperJANET backbone running at 34 mbs was sufficiently fast to support quick

transmission of this data, local Ethernet links running at 10mbps would slow browsing to a tedious pace. 'The implication is that the network speed needs to be maintained right through the network at a minimum of 34mbps in order to make browsing of electronic journals effective.'

A further impediment concerns the reason for publication, particularly in the sciences. Research projects are now so expensive that the publication of results is not just a validation process. It is a justification of the funding. Research investment per [scientific] paper was reported as £33,000 in the USA and £8-15,000 in the UK in 1993.<sup>14</sup> Formally reported research is now a performance indicator which will influence future funding. With competition for research funding increasing, researchers are not likely to risk publication in a medium which does not protect their work. A critical event in the current publication cycle is the receipt dating of a paper with a journal, which effectively gives that work protection against plagiarism, and primacy with respect to the later published work of competitors. Ginsparg directly addresses these points, and says that, in specific areas of physics, the mere number of refereed research papers is too coarse an instrument for the assessment of job applications or research grant proposals. Anyway, the research communities in such fields are small and the members well known to each other. So a pre-print can be of great value. Or to put the converse case, (according to Ginsparg), 'certain physics journals play no role whatsoever for physicists'.

What of the role of the commercial or institutional publisher? David Brown, in his paper 'Electronic publishing within the UK publishing industry', which forms part of the *Working papers of the Information Technology Subcommittee of the HEFC's Libraries Review*, refers to a study of the role commercial and learned publishers could have in network publishing. This concluded that of the existing ten functions performed by publishers, the set which involved describing the content of articles, the coding and tagging, would best be left to the professional publishers, whereas the remainder could be performed by others on the network, or the network itself.

With the future shape of the electronic product so unclear, both in terms of format and economic

character, the growing trend of the producers of the intellectual product, the researchers, taking their place in the debate is the most noteworthy. The universities of North Carolina and South Carolina, and Duke University, are investigating proposals to end the practice of faculty staff routinely transferring copyright to publishers. Instead, a model where universities and their faculties would develop partnerships with publishers, like the not-for-profit university and association presses, is under review.<sup>15</sup> The surrender of copyright in research publications to publishers has left the academic research community at a disadvantage, when this information has had to be bought back to seed further research. The pricing mechanism has meant that it is increasingly being rationed. So, if the value added by conventional publishing can be achieved more cheaply by other means in an electronic model and if the value added by some traditional library functions can be substituted more cheaply by electronic means, there are powerful reasons for researchers in the academic community to investigate them.

#### References:

1. Odlyzko, A. Tragic loss or good riddance? The impending demise of traditional scholarly journals. *Notices of the American Mathematical Society*. Vol 42, Jan 1995, pp 49-53.
2. Brown, D. Electronic publishing within the UK publishing industry in *Libraries and IT. Working papers of the Information Technology Subcommittee of the HEFC's Libraries Review*. Bath: UKOLN, 1993.
3. Page, Gillian. *Journal publishing: principles and practices*. London, Butterworth, 1988.
4. Hamaker, C. and Grinell, S. Cost Analysis of monographs and serials in Lee, Sul (Ed) *Library material costs and access to information*. New York, Haworth Press, 1990
5. Christensen, J. The cost of chemistry journals. *Chemistry and Industry*, 20 March 1995, p236.
6. Brown, D. Ibid.
7. Christensen, J. Ibid
8. Ginsparg, P. Various articles and statistics. Available from: <http://XXX.lanl.gov/blurb>
9. Odlyzko, A. Ibid

10. Ginsparg, P. Ibid
  11. Personal communication
  12. *Newsletter on Serials Pricing Issues*. Ed. Marcia Tuttle. Available free by e-mail from listserv@unc.edu. Type in body of message: subscribe prices [name]
  13. Pullinger, David J. *The Superjournal project. Electronic journals on Superjanet*. Bristol; Institute of Physics Publishing, 1994
  14. Donovan, Bernard. Why journals are published: the author, the editor and the review process. *Serials*. Vol 6, No 3, November 1993, pp69-73
  15. *Serials*. Vol 6, No 3, November 1993, p19.
-