Scientific publishing in transition: an overview of current developments

In this paper we present an overview of the current state of the world of scientific journal publishing. The data used in the paper has been sourced where possible from published reports and articles.

Mark Ware Consulting Ltd

Publishing and Elearning Consultancy

14 Hyland Grove
Westbury-on-Trym
Bristol BS9 3NR
UK

Telephone: +44 117 959 3726
Mobile: +44 7973 824378
Fax: +44 7812 336224

E-mail: mark@markwareconsulting.com

September 2006

Contents

Scientific publishing in transition

Executive summary

In this paper we present an overview of the current state of the world of scientific journal publishing. The data used in the paper have been sourced where possible from published reports and articles (see Notes and references, page 28).

1. Journals form a core part of the process of scholarly communication and are an integral part of scientific research itself. Journals do not just disseminate information, they also provide a mechanism for the registration of the author's precedence; maintain quality through peer review and provide a fixed archival version for future reference. They also provide an important way for scientists to navigate the ever-increasing volume of published material (see page 5).
2. The global market for English-language STM (scientific, technical and medical) journals is about $5 billion. The industry employs 90,000 people globally, of which 40%, or 36,000 are employed in the EU. Another 20–30,000 full time employees are indirectly supported (page 6).
3. There are about 2000 publishers, made up of learned societies, university presses and commercial publishers (though, to blur the picture somewhat, many society journals are published by commercial publishers). Their respective shares of article output are about 30%, 2% and 64% (page 11).
4. There are about 23,000 scholarly journals in the world, collectively publishing 1.4 million articles a year. The number of articles published each year and the number of journals have both grown steadily for over two centuries, by about 3% and 3.5% per year respectively. The reason is the equally persistent growth in the number of researchers, which has also grown at about 3% per year and now stands at around 5.5 million (page 7).
5. The development of online electronic versions of journals has revolutionised scientists' access to the literature. Over 90% of STM journals are now online, and in many cases their publishers have retrospectively digitised earlier hard copy material back to the first volumes. More content is available to more users than at any time in history while the cost of use of each article is falling to well below one euro. The industry has made this possible through the application of sustainable business models and the collective investment of hundreds of millions of euros in electronic developments (page 8).
6. The average total cost of publishing a journal article with a print and electronic edition has been estimated at $3750. This figure can rise substantially for high-quality prestige journals (like *Science*, *Cell* or *Nature*) because the very high rejection rate at these journals means the published articles have to bear the costs of handling the rejected ones (page 11).
7. Independent research by City University (London) in 2004 found that 70% of researchers believed that access to journal literature was better or much better than 5 years ago. Only 10% of authors said that access to the literature was poor or very poor. Another survey found that access to the literature came a long way down a list of possible barriers to research productivity, well behind factors like funding, ability to recruit suitable staff, insufficient autonomy in setting research direction, bureaucracy, lack of job security, etc. (page 13).
8. Journal publishing has become even more competitive over the last 5 years with the emergence of new business models. Open access posits making

original research freely accessible on the web. There are two approaches: open access publishing and self-archiving (page 16).

9. There are some 2000-2400 open access journals in existence, publishing about 2–5% of total articles. They use a variety of funding models, grants, membership subscriptions, sponsorship/advertising, commercial reprints, classified advertising, subscriptions to print editions, volunteer labour, and subsidy or support in kind by the host organisation. The best-known approach, is the "author-side payment" model, where a publication charge (mostly in the range $2–3000) is levied on each accepted article (page 16).
10. It is still too early to say for sure how viable open access publishing will be. Neither of the leading pioneers, Public Library of Science and BioMed Central are even close to profitability. The available data are patchy but, taken together, suggest that achieving widespread sustainability for open access journals will not be particularly quick or easy (page 19).
11. The other route to open access is via self-archiving, whereby the author posts a version of the article (typically the revised manuscript after peer review but prior to copyediting, known as a post-print, rather than the final published article) to an open web-based repository. These repositories can either be central, subject-based collections (e.g. the well-known physics repository, arXiv) or organised to collect the output of a particular institution (page 22).
12. A worrying development for publishers is the emergence of policies by research funders and by authors' employers requiring the deposit of articles in such repositories. The US National Institutes of Health introduced such a policy in 2005, and has subsequently been followed by the Wellcome Trust and some of the Research Councils in the UK, and others in France, Germany and elsewhere (page 25).
13. Publishers fear that widespread systematic self-archiving of this kind will have a serious impact on journal subscriptions, the revenue stream that supports the vast majority of journals. There is evidence from physics and elsewhere that archiving reduces the amount of use of articles get on the publisher's website (readers get the articles from the repository instead). There is also some evidence from a survey of librarians that this is becoming an increasingly important factor in considering journal cancellations. A major study on this subject by Scholarly Information Strategies for the Publishing Research Consortium is due to report on this during October (page 26).

The scientific journal

What is a journal?

There is a spectrum of types of publication that are loosely described as journals, from *Nature* to *Nuclear Physics B* to *New Scientist*, with no clear dividing lines. In this paper, however, we are concerned predominantly with the primary scientific literature: that is, periodicals carrying accounts of original research published after due peer review.

The journal has traditionally been seen to embody four functions:

- *Registration*: establishing the author's precedence
- *Dissemination*: communicating the findings to its intended audience
- *Peer review*: ensuring quality control
- *Archival record*: preserving a fixed version of the paper for future reference and citation.

We take the trouble to restate these fundamentals because it will set the context for a discussion of newer systems – like open archives – that perform some, but not all of these functions.

It is also worth noting that these functions can be seen as much as services for *authors* as for readers. Indeed it has been suggested that when authors transfer rights in their articles to journal publishers for no fee, they are not "giving away" the rights but exchanging them for these services (and others, such as copy editing).

To these might now be added a fifth function, that of *navigation*, that is, providing filters and signposts to relevant work amid the huge volume of published material.

The journals publishing cycle

The movement of information between the different participants in the journal publishing process is usually called "the publishing cycle" and often represented as in Figure 1. Here research information, created by an author from a particular research community, passes through the journal editorial office of the author's chosen journal to its journal publisher, subscribing institutional libraries – often via a subscription agent – before ending up back in the hands of the readers of that research community as a published paper in a journal.

Authors publish to disseminate their results but also to establish their own personal reputations and their priority and ownership of ideas. The third-party date-stamping mechanism of the journal registers their paper as being received and accepted at a certain date, while the reputation of the journal becomes associated with both the article and by extension the author.

The editor of a journal is usually an independent, leading expert in their field (most commonly but not universally a university academic) appointed and financially supported by the publisher. The journal editor is there to receive articles from authors, to judge their relevance to the journal and to refer them to equally expert colleagues for peer review. Peer review is a methodological check on the soundness of the arguments made by the author, the authorities cited in the research and the strength of originality of the conclusions. While it cannot generally determine whether the data presented in the article is correct or not, peer review undoubtedly improves the quality of most papers and is appreciated by authors. The final decision to publish is made by the journal editor on the advice of the reviewers. The review process alone can take from weeks to months, with a similar delay until publication

after the article has been accepted, although electronic publishing has greatly reduced delays in this second stage.

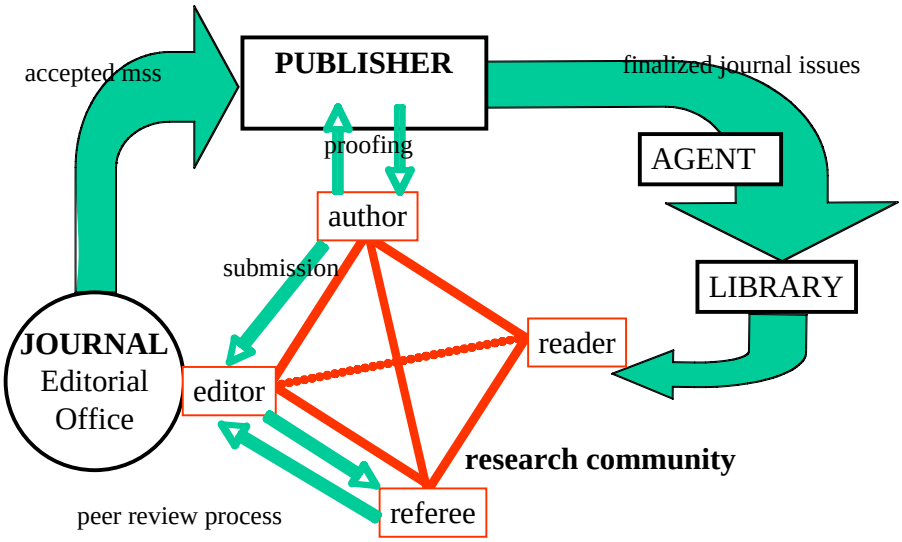

Figure 1: The publishing cycle (courtesy of Michael Mabe)

The role of the publisher is often confused with that of the printer or manufacturer, but it is much wider. Identifying new, niche markets for the launch of new journals, or the expansion (or closure) of existing journals is a key role for the journals publisher. This entrepreneurial aspect seeks both to meet a demand for new journals from within the academic community – and it is noteworthy that journal publishers have been instrumental in the birth of a number of disciplines through their early belief in them and support of new journals for them – but also to generate a satisfactory return on investment. As well as being an entrepreneur, the journals publisher is also required to have the following capabilities:

- Manufacturer – copy editing, typesetting, printing and binding the journals.
- Marketeer – attracting the papers (authors) and new subscribers.
- Distributor – publishers maintain a subscription fulfilment system which guarantees that goods are delivered on time, maintaining relationships with subscription agents, serials librarians and the academic community.
- Electronic host – electronic journals require many additional skill sets more commonly encountered with database vendors, website developers and computer systems more generally.

Journal economics and market size

The annual revenues generated from English-language STM (Scientific, Technical & Medical) journal publishing are not well documented but are estimated at around $5 billion in 2004[1]. This is a subset of the wider STM publishing market (including books, secondary information services, A&I databases, etc. which was worth $9–12 billion in 2004. STM journals represent a relatively small niche in the overall global publishing and information market, which Outsell estimated to be worth some $263 billion, or even compared to educational publishing ($19.4 billion in 2004).

The industry employs an estimated 90,000 people globally, of which about 40%, or 36,000 are employed in the EU. In addition, an estimated 20–30,000 full time employees are indirectly supported by the STM industry globally (suppliers, freelancers, external editors, etc.)[2].

Journal and articles numbers and trends

There are about 23,000 scholarly peer-reviewed journals[3], collectively publishing about 1.4 million articles a year. An important subset is the 8700 journals included in the ISI Journal Citation database, of which 5900 are in the Science Edition, 1700 in the Social Sciences and 1130 the Arts & Humanities Editions), which collectively publish about 1 million articles annually. This subset is important because it contains the most cited journals, that is, by this measure at least the core literature.

The number of peer reviewed journals published annually has been growing at a very steady rate of about 3.5% per year for over two centuries (see Figure 2, although the growth did slightly accelerate in the post-war period 1944–78). The number of articles has also been growing by about 3% per year over similar timescales. The reason for this growth is simple: the growth in the number of scientific researchers in the world. This is illustrated in Figure 3, which plots the increase in numbers of articles and journals alongside the numbers of US researchers.

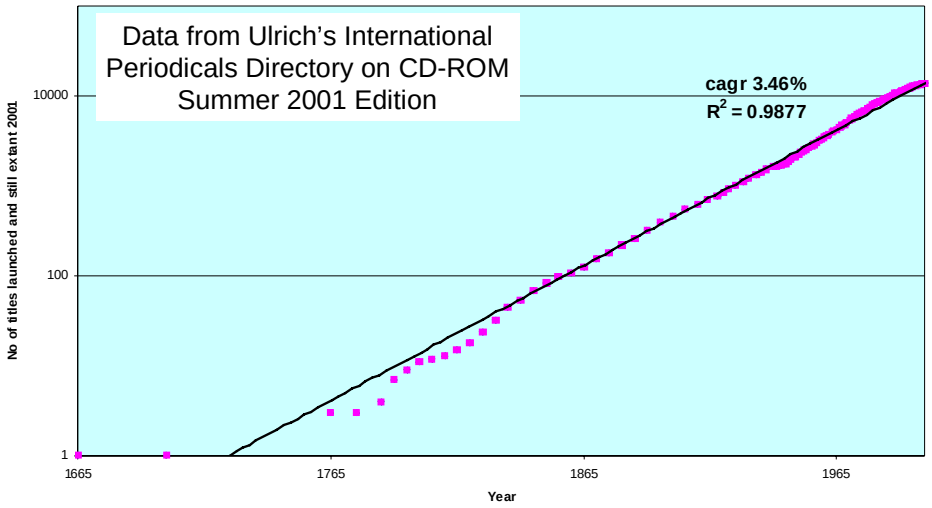

Figure 2: The growth of active, peer reviewed learned journals since 1665[4]

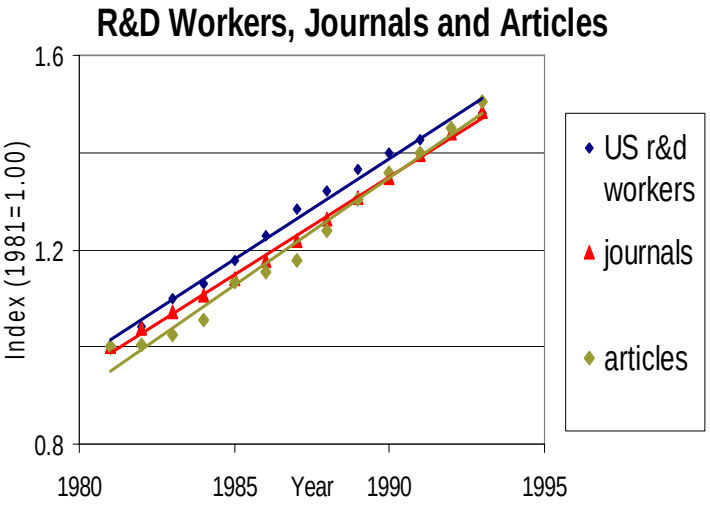

Figure 3: Relationship between numbers of researchers, journals and articles[5]

Online journals

The majority of journals are now available online. A study by Cox in 2005 (based on a publisher survey) found 90% of all journals were online, with 93% of STM and 84% of Arts & Humanities journals. Data in Ulrich's Periodicals Directory suggest a lower proportion of 62% online. The reason for the divergence is likely to be that Ulrich's is more representative of the total global situation, while Cox's sample represents the more advanced development of the US and UK/European publishing industry.

Open Access journals

We shall discuss open access in more detail below but to complete the statistical picture of the journal market we consider the numbers of OA journals. Data from Ulrich's and the Directory of Open Access Journals (DOAJ) suggest that the total number of journals published under some kind of OA model is around 2000–2400. This figure also includes titles that are not fully peer-reviewed and some that are inactive; restricting to just peer-reviewed and active would reduce the total. OA journals therefore represent under 10% of the total journal output, but it is important to recognise that because these journals are smaller than average, these OA journals represent a much smaller proportion (probably 2–5%) of the total articles published.

OA journals also form a numerically small part of the core literature as measured by inclusion in the ISI database. According to a paper published by ISI in 2004[6], there were 239 OA journals covered by the ISI database, or 2.6% of the core ISI set or ~1% of the larger set of journals covered by ISI's *Web of Knowledge*.

Global trends in scientific output

The number of articles catalogued by the ISI's *Science Citation Index (SCI)* and *Social Sciences Citation Index (SSCI)* grew from approximately 466,000 in 1988 to nearly 700,000 in 2003, an increase of 50% (see Figure 4). The growth of publications reflects both an expansion in the number of journals covered by the *SCI* and *SSCI* databases and an increase in the number of articles per journal during this period. Within this overall growth, there are important regional differences, with the EU's output growing faster than the US and overtaking it in the late 1990s (Figure 5). The most dramatic growth, however, is in the output from the East Asia region (China, Singapore, South Korea and Taiwan), which has been around 14–15% over the period (compared to 1–2% for the US, for example).

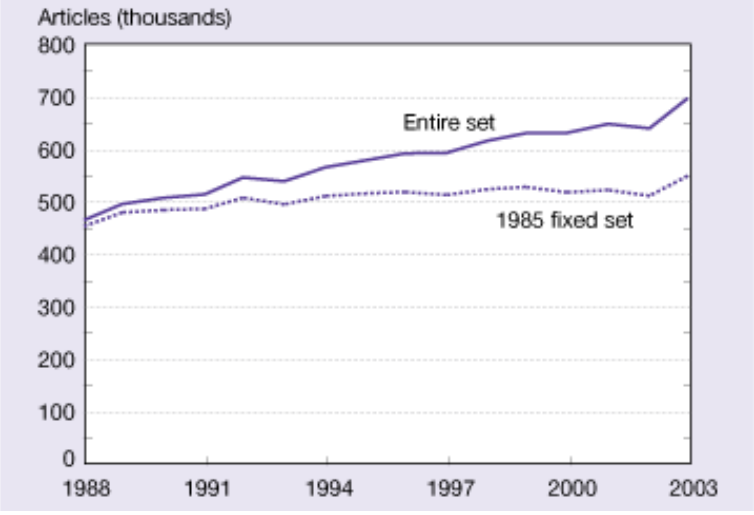

Figure 4: Worldwide scientific article output of selected journal sets: 1988–2003[7]

Research is becoming more international and more collaborative, driven by factors including the scientific advantages of sharing knowledge and know-how beyond a single institution; the lower costs of air travel and telephone calls; increased use of information technology; national policies encouraging international collaboration and the ending of the Cold War; and graduate student study abroad programmes. This growing globalisation of science is reflected in both an increase in the average number of authors and institutions on an article, and in the proportion of foreign addresses. So for articles published in the EU, for example, the average number of co-authors per article increased from 3.33 to 4.81 between 1988 and 2003, while articles with at least one co-author from a non-EU country accounted for 36% of all articles in 2003, up from 17% in 1988[8].

However, at the same time as these co-authorship trends, the annual productivity of each unique author has fallen slightly from one paper per annum per unique author in 1950 to about 0.7 in 2000. As a consequence, although each author is on average getting their name as a collaborator on about four papers each year, they are each responsible for only 0.7 of a paper per annum. Thus the driving force behind the growth in the number of papers in the world remains the number of authors[9].

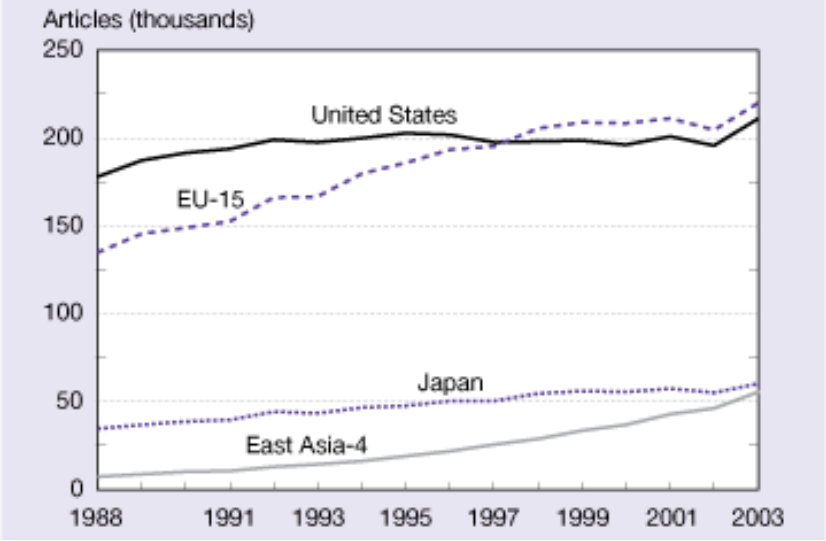

Figure 5: Scientific article output, by major publishing region or country/economy: 1988–2003[10]

Authors and readers

Estimates of the global research community are regularly compiled by UNESCO. Their latest publicly available estimate (source data relating to 2002 or earlier) estimates a base of 5.5 million researchers worldwide[11].

Scientific journal articles are written primarily by academics. For instance, Tenopir and King report that although only 10 to 20% of the scientists in the United States are employed in universities they account for about 75% of articles published[12].

More recent work from Tenopir & King suggests that about 15 per cent to 20 per cent of scientists in the United States have authored a refereed article. This estimate – and the asymmetry between authors and readers – is corroborated by work from Mabe and Amin who estimate that, of the 5–6 million global researchers calculated by UNESCO, only around 1 million (circa 18 per cent) are unique repeat authors[13].

Incidentally, the average scientific paper takes its authors 90–100 hours to prepare[14]. Two to three reviewers will then spend an average of 3–6 hours each on peer review[15].

There is also a distinction to be made between the core active researcher segment and the wider journal-reading community, which is likely to be much larger. Many of these additional readers may be far more peripheral and infrequent readers. This category would also include journal reading by post-graduate and undergraduate students in universities. There appears to be no robust evidence sizing this wider journal reader community but internal research at Elsevier derived from analysing global unique user counts for ScienceDirect suggests the total global journal readership may be between 10–15 million.

Citations and the Impact Factor

Citations are an important part of scientific articles, helping the author build their arguments by reference to earlier work without having to restate that work in detail. They also help readers enormously by pointing them to other related work (surveys show that this is one of the most popular ways authors navigate the literature). Modern electronic journals now also allow "forward" reference linking, i.e. linking to later work that cites the paper in question. The volume of citations worldwide increased from 2.69 million in 1992 to 4.34 million in 2003, an increase of 61%[16]. During this period, the share of cross-national citations grew from 42% to 48%, another sign of the increasing globalisation of science. At the same time, the EU's share of global citations increased strongly, to some extent closing the gap with the US (see Figure 6).

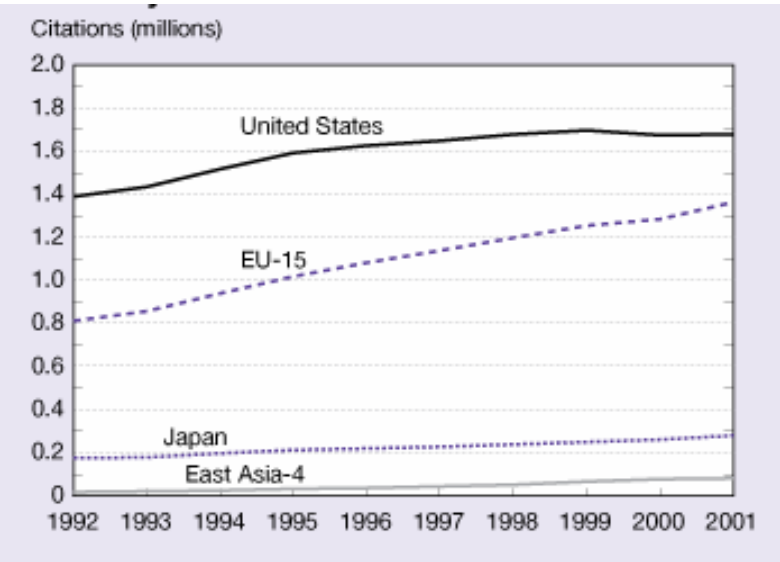

Figure 6: Citations of STM literature, by region/economy: 1998–2003[17]

The number of citations a paper receives is often used as a measure of its impact and by extension, of its quality. The use of citations as a proxy for impact or quality has been extended from articles to journals with the Impact Factor. A journal's Impact Factor is a measure of the frequency with which the "average article" in a journal has been cited in a particular period. (The official definition is that the impact factor is the total number of citations given to a journal in second and third years after publication divided by the total number of citeable items published during that same time period.)

The use of citations data, and in particular the journal-level Impact Factor, to judge the quality of individual researchers' and departments' research outputs, though widespread, is increasingly criticised. The assumption that articles published in the

same journal are likely to be of similar quality is not borne out by the data: there is a skewed distribution with 15% of articles accounting for 50% of citations, and 90% of citations generated by 50% of articles[18]. The top half of articles in a journal can thus receive 9 times as many citations as the bottom half. Some believe that the number of downloads might give a better measure of an article's impact (as noted above, there are many more scientists who are not authors than those who write). The UK Serials Group has commissioned some work to investigate whether it might be feasible to develop a "Usage Factor" based on download statistics[19]. However, all indicators suffer from the "normalisation" problem. That is, how do you correct for the differing effects caused by journal size, type, market size, and discipline to enable one journal to be compared to another. These issues are independent of the choice of indicator, so it is difficult to see how download data will not suffer the same problems as citation[20].

An interesting question is whether articles in open access journals, and articles self-archived by their authors in parallel to traditional publication, receive more citations than they would otherwise have done. This is discussed below in the section on open access.

Publishers

There are estimated to be of the order of 2000 journal publishers globally. The main English-language trade and professional associations for journal publishers collectively include 657 publishers producing around 11,550 journals, that is, about 50% of the total journal output by title. Of these, 477 publishers (73%) and 2334 journals (20%) are not-for-profit[21]. Earlier analysis of Ulrich's directory suggested that about half of all journals came from not-for-profits; the apparent discrepancy may reflect Ulrich's broader coverage. Analysis by Elsevier of the ISI Journal Citation database indicated that the proportions of article output by type of publisher were: commercial publishers (including publishing for societies) – 64%; society publishers – 30%; university presses – 4%; other publishers – 2%.

The distribution of journals by publisher is highly skewed, with two publishers (Elsevier and Springer) having around 2000 journals each. The top 2% (11 publishers) produce more than 70% of the journals in this group, that is, about 35% of all journals. There is a "long tail" of organisations producing a small number of journals, and many of these may not even regard themselves as "publishers" (e.g. academic or government research departments).

Costs of journal publishing

It is helpful to have an understanding of the costs of journal publishing, for instance in relation to the debate over open access publishing (see below).

In publishing, it is common to distinguish between the "first copy costs" and the total cost of publishing. It is also important to clarify whether one is considering just the direct costs (e.g. editing, typesetting, production, marketing and distribution) or whether they include associated indirect costs (e.g. staff costs and other overheads).

The first copy costs are those incurred in bringing a manuscript to the stage where it can be printed or uploaded to a server (or both). These include the costs of managing the peer review (the actual peer review is done by academics *pro bono* but managing the process incurs real costs to publishers – in fact reports of the costs of *managing* peer review vary between $60 and $635 per article, presumably reflecting a variety of practices[22]); substantive editing; copy editing; verifying references and inserting tags to create the online links; preparation of illustrations or special graphics; typesetting and layout; etc. The costs can vary substantially between journals: for example, journals with very high rejection rates incur much higher average costs of peer review management than those with lower rejection rates. First copy costs reported in the

literature vary substantially, with figures ranging from $410 to $10,000 (the latter is for *Science*, which has an exceptionally high rejection rate).

The average total cost of production for a journal article including print is generally thought to be about $3750[23]. Eliminating the print edition could save 10–20% of these costs (this is true for both subscription and open access journals) but many libraries and users are as yet unwilling to give up print.

Journal prices, as well as covering costs, also include in most cases an element of profit (in the case of commercial publishers) or surplus (for not-for-profits). Profits are a major source for reinvestment and innovation. Society publishers frequently use surpluses from journal publishing to support other activities such as conferences and seminars, travel and research grants, public education, etc.)[24].

Journal pricing

Journal pricing has been the source of much debate and controversy, and perceived high prices and high price increases have been one of the factors driving the open access agenda. It is true that journal prices have outpaced inflation, for instance the Association of Research Libraries (ARL) have published statistics which show that the annualised price rise for journals over the period 1984–2004 was 7.6%, compared to the US Consumer Prices Index which rose by an annualised 3.1% over the same period[25]. Similarly Tenopir and King[26] showed that the price inflation ratios between 1975 and 1995 for commercial and society journals were 3.1 and 2.9 respectively (using current dollars).

The figures do not, however, represent what libraries have actually paid, because of the efficiencies of electronic delivery and the growth of multi-journal licences. LISU (Loughborough University's Library and Information Statistics Unit) note in their 2005 annual report that such deals were partly responsible for *lowering* the average cost per title of current UK serial subscriptions by 23% over the 5-year period to 2003/04[27].

The reasons for journal price increases are varied and include[28]:

1. Increased numbers of articles produced by researchers, as described above (at around 3% per annum). This is a fundamental driver for journal costs. This leads to:
2. Increased numbers of articles per journal: from 1975 to 2001 a journal tracking study[29] showed that the average number of articles per year published in science journals increased from 83 to 154 articles per title[1].
3. Increased average length of articles: the same study indicated an increase from 7.4 to 12.4 pages per article. This, combined with 1 above, leads to:
4. Increased size of journals: similarly, the size of science journals (including non-archival content) increased from 820 to 2216 pages per year.
5. Increased special requirements or features such as specialized language, special graphics, mathematical equations, chemical compounds, citations, linkages, moving graphics and images and links to numeric databases.

[1] It is also the case that science journals are on average more expensive than those in social sciences or the arts. The reason for this appears to be primarily that science journals are bigger: STM journals publish between 4 and 10 times as many articles as social science journals (see Figure 7).

6. Conversion of back issues to electronic format, provision of search options, and other value added attributes associated with electronic publishing.
7. Publishers increasing prices to compensate for falling subscription numbers (which may of course lead to a spiral of further demand reduction, etc.) and currency effects (journals may be produced in one currency area and sold in another, leading to potential exchange losses).
8. And, of course, inflation (especially salary and paper costs), which have annualised at about 3.3% per annum for the last twenty or more years.

An additional factor is the relative economic inefficiency of new journals when they are started (journals start with 4, 6, 8, even occasionally 12 issues rather than a slow increment of articles). When new journals being introduced at 3.5% by title per annum are factored into overall subscription inflation, this can contribute up to 1% of the average 7.6% rise experienced by libraries.

In summary then, the observed 7.6% annual average journal price inflation for the last twenty years has a number of components: organic growth in the literature (3%), cost inflation (3.3%), electronic delivery and conversion costs, new journal specialisation (up to 1%) and attrition and currency fluctuation effects (ca. 0.5–1%).

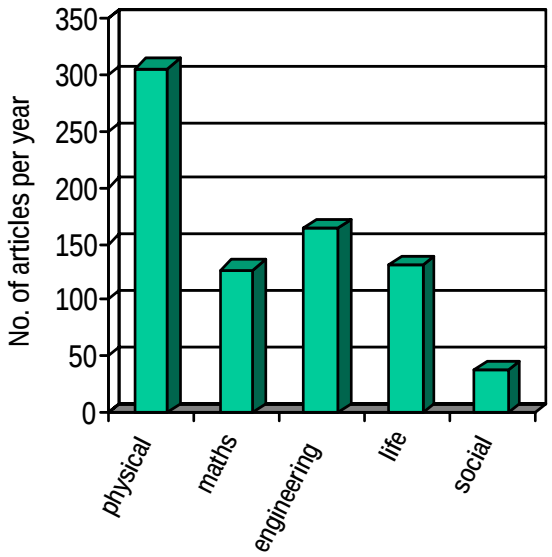

Figure 7: Articles per journal per year, by discipline[30]

Authors' behaviour, perceptions and attitudes

There have been numerous studies of author behaviour, perception and attitudes but two recent pieces of work stand out for their large scale (4000–6000+ respondents) and rigorous methodology and design: the two surveys conducted by CIBER (part of University College London) and published in 2004 and 2005[31], and a survey commissioned by Elsevier in collaboration with CIBER and NOP in 2005[32]. These surveys represent the best and most current data on these topics.

In "New journal publishing models: an international survey of senior researchers" Rowlands & Nicholas report on the second CIBER survey, which received responses from 5513 senior journal authors. Among their findings were:

- In choosing where to publish, being able to retain copyright or to be able to place a copy of the pre- or post-print on the Web or in a repository were not of importance to most authors.

- The crucial importance of peer review was re-emphasised.
- Senior authors and researchers believe downloads to be a more credible measure of the usefulness of research than traditional citations.
- Two significant shifts were reported to have occurred since the previous CIBER survey (2004). First, the research community is now much more aware of the open access issue. There has been a large rise in authors knowing quite a lot about open access and a big fall in authors knowing nothing at all about it. Secondly, the proportion of authors who thought they had published in an open access journal grew from 11% (2004) to 29%[33]. [This perception is probably erroneous because as we saw above only ~2–5% of papers are actually being published in open access journals. It would seem that authors are confusing barrier-free access through their institutional desktops to journals purchased by their libraries with genuinely free-to-read publications.]
- A clear majority of authors believes that mass migration to open access would undermine scholarly publishing. (A good proportion of these, however, thought this would be a good thing, reflecting dissatisfaction with the status quo.)
- There is little enthusiasm for author- or reader-facing charges [e.g. submission/publication charges or pay-per-view charges, respectively].
- Authors had very little knowledge of institutional repositories: less than 10% declared that they know "a little" or "a lot" about this development. There was also evidence that a significant minority (38%) were unwilling to use IRs.

The Elsevier/CIBER/NOP 2005 survey used a similar methodology to the CIBER surveys – online questionnaires with 6344 responses – but supplemented this with 70 follow-up depth telephone interviews. Among its key findings were:

- Although the superficially most important reason given for publishing was to disseminate the results, the underlying drivers were funding and furthering the author's career. This pattern was similar to an earlier study conducted in 1993 except that "establishing precedence" and "recognition" had increased in importance.
- Researchers are ambivalent towards funding bodies: 63% think they have too much power over what research is conducted. But despite concerns about the pressure to publish in high impact journals, funding bodies do not dictate the choice of journal. [It should be noted that this survey was conducted before funding body mandates about article deposit were introduced; attitudes may have significantly altered as a result.]
- Authors are divided when it comes to deciding whether to publish in a prestigious or niche journal: 43% agree while 39% disagree that it is important to publish in a prestigious general journal rather than a more appropriate specialised journal.
- The importance of peer review is underlined. There was near universal belief that refereed journals were required. The large majority believe that peer review improves an article. Respondents were committed to peer review: 85% were willing to act as reviewers.
- A majority – 60% – believed that the publisher adds value – but 17% did not, with more thinking so in Computer Science (26%) and Mathematics (22%).

- Reading patterns are slowly changing: a significant minority (22%) of respondents preferred to conduct their e-browsing from the comfort of home. (Medical researchers had the highest response at 29%.)
- Electronic versions have not yet completely taken over: the majority disagree that an article will only be read if available electronically.
- There was high demand for articles published more than 10 years ago. [This date is important because few journals launched online versions before then, so electronic access to this literature depends on someone – usually the publisher – retrospectively digitising this material.]
- Knowledge of both institutional and subject-based repositories was fairly low, with 33% knowing a little or a lot about institutional repositories, and 38% about subject-based repositories. [These figures were, however, considerably higher than the 10% reported by CIBER, see above.]

Researchers' access to journals

The development of online versions of scientific journals has led to greatly increased access to the scientific literature at greatly reduced cost per use. This has been largely because the very low marginal costs of electronic distribution have allowed publishers to offer access to sets of journals (up to and including the complete output of the publisher) for relatively small additional licence fees compared to the previous total print subscriptions at the institution. On the demand side, libraries have formed consortia to enhance their buying power in negotiating electronic licences with publishers, also resulting in access to more journals for their readers.

Statistics show that the number of journals acquired per library has increased dramatically since the advent of electronic journals in the late 1990s, and the cost paid per journal has fallen[34]. For example, the number of current serials subscriptions per higher education institution in the UK has more than doubled in the last 10 years to 6900[35]. Usage of previously unsubscribed journals in such licences is remarkably high[36], and cost per use is falling to low levels[37].

Examples of evidence of this widened access include:

- From the 2004 CIBER survey: "A surprising finding of the survey is the very high level of reported satisfaction with access to the journals literature: 61% of authors said that this was currently 'good' or 'excellent', meaning that they have access to all or at least most of the materials they need. Only 10% of authors said that matters were 'poor' or 'very poor'."
- Nearly 70% of researchers in the same survey believed access to journal content was better or much better now than 5 years ago.
- A survey of immunologists and microbiologists by CIBER for the Publishing Research Consortium (PRC) found that they were "generally satisfied with their level of access to the journals system and a large majority (83.7%) agree that major improvements in journal accessibility have been made over the past five years."[38]
- The PRC survey also found that access to the literature came a long way down a list of possible barriers to research productivity, well behind factors like funding, ability to recruit suitable staff, insufficient autonomy in setting research direction, bureaucracy, lack of job security, etc.
- Elsevier reported that EU libraries with relatively large collections of Elsevier print journals in 1999 (334 journals on average) had access in 2005 to 3.7 times as many Elsevier titles via ScienceDirect (1221 titles on average).

Access in developing countries

There are a number of schemes providing free or heavily discounted access to the scientific literature to researchers in developing countries. Some of the more notable ones include:

- HINARI, a collaboration between publishers, WHO and Yale University Library, offers free access to over 3300 biomedical journals to countries with the lowest per capita incomes, and access for a nominal fee ($1000 for the full collection) for the next band of countries, 113 countries in total. Downloads by developing country researchers are running at an annual rate of well over 4 million articles.
- HINARI's sister programme, AGORA, provides access to the journal literature in food and agriculture, and a third programme, OARE, was launched in 2006 to provide access in environmental sciences.
- HighWire Press offers free access for developing countries to a list of 320 high-quality journals, based simply on software that recognises from where the user is accessing the site. Bepress (Berkeley Electronic Press) has a similar arrangement.
- Some publishers offer similar schemes independently, e.g. the Royal Society of Chemistry, the National Academies Press.
- INASP's PERI scheme negotiates affordable, sustainable country-wide licences that provide access free at the point of use for researchers.
- eIFL (Electronic Information for Libraries) provides country-wide access to thousands of titles in social sciences, humanities, business and management by libraries in nearly 40 countries of the Soros Foundations' network.

Open access

Open access is the idea of making original research freely accessible on the web, ideally immediately on publication[39]. It is therefore strictly speaking a property of an article, rather than a journal. There are two approaches (or "routes") to open access:

- Open access publishing – the "gold" route, whereby the journal makes the article freely accessible.
- Open access self-archiving – the "green" route, where the author (or someone acting on their behalf) deposits a version of the published article, typically a pre- or post-print, in an open repository.

There are numerous variants on each of these approaches. We shall discuss these briefly in the next sections and look at the current state of play.

Open access publishing

There are three main variants of OA publishing:

- Immediate full OA: the entire contents of the journal are made freely available immediately on publication. A well-known example is *PLoS Biology*.
- Hybrid and optional OA: here only part of the journal content is made immediately available. There are two distinct models:
    - The journal makes its research articles immediately available but requires a subscription to access other "value added" content such as commissioned review articles, journalism, etc. An example is *BMJ*.

- The journal offers authors the option to make their article OA in an otherwise subscription-access journal in return for payment of a fee (e.g. Springer's Open Choice or OUP's Oxford Open schemes).
- Delayed OA: the journal makes its contents freely available after a relatively short period, typically 6–12 months (e.g. the majority of journals on the HighWire platform). A growing number of journals (particularly in the life science and biomedical areas) have adopted delayed open access policies.

The optional model potentially provides a relatively low risk way for established subscription journals to experiment with open access, in effect allowing the market (i.e. authors, or their funders) to decide what value they place on open access. Nearly all the major journal publishers, both commercial and not-for-profit, are now offering optional schemes (Sage and Taylor & Francis are notable exceptions), with four major publishers announcing schemes in August 2006 alone. Some publishers have said publicly that they will reduce the subscription price in proportion to the revenues raised from OA publication charges, while others have remained silent on this point.

Willinsky has identified nine different sub-species of open access[40]. Apart from those listed above and the self-archiving route, he includes "dual mode" (print subscription plus OA online version); "per capita" (OA made available to countries based on per capita income – see discussion of developing country access above); "abstract" (open access to journal table of contents and abstracts – most publishers offer this); and "co-op" (institutional members support OA journals – an example is the German Academic Publishers).

A new variant of hybrid open access emerged very recently, whereby the articles submitted by members of American Society of Plant Biology will be published in the society's journal *Plant Physiology* with full immediate open access[41].

Open access citation advantage

A number of studies have addressed the question of what the effect of open access self-archiving might be on the citations an article receives. The common-sense hypothesis is that an archived article will receive more use, and hence be cited more often (and earlier), than one only available in a subscription journal. However, since other academics are the source of virtually all citations an article gets, an overall increase in citation numbers would only be possible if a significant proportion of the active researchers in the field of the journal did not already have access. There is as yet no clear cut answer to this. Most studies have shown that this does appear to be the case but have failed to deal with the issue of causation: did the articles receive more citations because they were OA, or did their authors select their better papers to archive or are better authors more likely to self-archive? In response to this, Kurtz[42] (2005) reported that in astronomy, self-archiving was correlated with increased citations but that the increase was explained by self-selection bias, with no evidence of any OA effect. The leading bibliometrician Moed[43], studying articles archived on arXiv, also found a strong self-selection and "early view" effects but limited evidence for an open access advantage.

A related question is whether articles published in an open access journal receive more citations than they would have done had they been published in a subscription journal. Eysenbach[44] looked at articles published in the *Proceedings of the National Academy of Sciences* (PNAS) and concluded that "OA articles are more immediately recognised and cited by peers than non-OA articles published in the same journal", even after allowing for self-selection.

The initial results of studies commissioned by OUP into their open access journals also show an unclear outcome. LISU could find no citation effect so far. CIBER

looking at the download data concluded that the "OA effect" here may simply be due to search engines: the journals were opened up to crawling by Google shortly before going OA[45].

OA publishing business models

The best-known OA publishing model is the "author-side payment" model, where the author (or usually his/her research funder or institution) pays a publication charge. Full immediate OA journals and optional OA journals both use this approach. Many full and optional OA journals also offer paid-for "institutional memberships", whereby members of the paying institution can pay reduced (or sometimes no) publication charges.

Research by the Kaufman-Wills Group for ALPSP published in 2005, however, showed that the (small) majority of OA journals did not make author charges (in fact, author charges are more common (in the form of page charges, colour charges, reprint charges, etc.) among subscription journals). Instead these journals used a variety of funding models, including grants, membership subscriptions, sponsorship/advertising, commercial reprints, classified advertising, subscriptions to print editions, volunteer labour, and subsidy or support in kind (witting or unwitting) by the host organisation.

Table 1: Publication charges for a selection of full and optional OA journals[46]

Journal/publisher	*Full/Optional OA*	*Charge (US$)*
American Physical Society	Optional	975–1300
BioMed Central	Full	1100–1700
Blackwell	Optional	2500
BMJ Publishing Group (exc. BMJ)	Optional	2200–3145
Cambridge University Press (some)	Optional	2700
Elsevier (some journals only)	Optional	3000
New Journal of Physics/IOP-DPG	Full	1080
Oxford University Press	Optional (49)/Full (1)	1900
PLoS	Full	2000–2500
Royal Society (London)	Optional	370–550 per page
Springer	Optional	3000
Wiley (45 biomedical journals only)	Optional	3000

Publication charges were set at low levels by the OA publishing pioneers PLoS and BioMed Central. Both have had to raise their fees substantially: PLoS raised its fee of $1500 to $2000 or $2500 (depending on journal), while BMC has raised its fee from its original $500 to between $1140 and $1710 for the majority of its journals. Both publishers have yet to break even financially. Fees for full and optional open access journals now mostly fall in the range $2–3000, with some notable exceptions (see Table 1). The situation is complicated a little by the fact that some OA publishers offer reduced fees to authors at institutions that agree to pay an institutional subscription fee, while some impose additional charges. For example, the OUP charge for Nucleic Acids Research is reduced from $1900 to $950 for "member institutions" who pay an annual subscription of $2700, but it is increased by a charge of $170 per page over 9 pages. For simplicity we have ignored these complications in the table.

In order not to exclude authors from low-income countries or those who lack the funds, most if not all full open access journals offer to waive charges for such authors.

This will potentially affect the financial sustainability of this model unless an allowance is made for the proportion of waived or absent author fees in the normal charge (see Waivers of publication charges, page 21).

How viable is open access publishing?

The short answer, as Kaufman-Wills said in their report in 2005, is that it is still too early to tell for sure. Author awareness and knowledge of open access are growing but until recently have been low. Experimentation is the order of the day at most journal publishers, large and small, commercial and not-for-profit. The possible effects of research funder mandates (which do not specify OA publication but encourage and validate the open access concept – see section on self-archiving policies and mandates, page 25) are unclear.

Despite this fast-moving and hazy background, there are some experiences and studies that are worth reporting:

- *Nucleic Acids Research*, one of OUP's flagship journals (Impact Factor 7.55), converted to full OA from January 2005. (Previously it had been a delayed OA journal, with content freely available after 6 months.) OUP revealed in June 2006 that NAR's income per article had dropped from $4647 in 2004 to $3622 in 2005[47]. Income from subscriptions (which include a print copy) declined steadily. It was also mentioned that dealing with payments from individual authors had generated an increased administrative burden. It is unclear how sustainable this will be: revenue loss combined with increased costs does not suggest a rosy future.
- Figures published in *Nature*[48] in June 2006 showed that the early break-even hoped for by PLoS is some way off. It lost $1 million 2005, and more importantly its author fees and advertising revenues covered only 35 per cent of total costs. PLoS has received some $13 million in grant funding. In response to the *Nature* article, PLoS has pointed out that its journals are at an early stage of their lives and but also said that it envisaged relying on grant support for the foreseeable future, despite the increase in author charges mentioned above. While grant support may be viable for one or two journals, funding all 23,000 titles through this mechanism is clearly unsupportable.
- BioMed Central, the largest full OA journal publisher, launched in 1999 and now has a total of 162 journals. It reported a loss of £2.9 million (excluding exceptional inter-group income) on a turnover of £1.8 million for 2005[49]. The loss in 2004 was similar, at £3.0 million. The number of OA articles published in 2005 and 2006 increased while the author charges nearly doubled, and they also changed their institutional membership model so that they pay in proportion to the number of articles published (rather than a flat fee). BMC pointed out that they are engaged in several other activities, including R&D activities relating to databases and database related services, outside of its core open access business, which account for a substantial proportion of the costs and hence reported losses.
- Cairo-based Hindawi Publishing, a commercial STM publisher with some 40-odd OA journals, has reported that its OA publishing is already profitable, even though it charges typically only $450 per article. As well as author charges it also offers subscriptions to the print editions, albeit at fairly low prices. It benefits from operating in a cheap labour market compared to its main competitors but it also attributes its success with OA to acquiring well-established journals that it converts to OA rather than starting from scratch, and is careful only to acquire journals in well-funded fields and those familiar with page charges.

- *Geometry & Topology* and *Algebraic & Geometric Topology* announced in July 2006 that they were to convert from open access to subscription journals after 10 and 5 years respectively. The reason given was the failure of their business model, which relied on volunteer labour (which now wishes to move on to other endeavours) and declining library "subscriptions" (these subscriptions were not required to access the journals but to allow libraries to link to their sites).
- *Journal of High Energy Physics* is a very successful journal (high volume of submissions, high impact factor), originally launched in 1997 by the SISSA research institution in Trieste as an open access journal. SISSA tried to run the journal on a combination of institutional support (i.e. by SISSA), sponsorship by high-energy physics institutions and a (very) limited number of print subscriptions. It recognised that the journal could not survive on this model because the sponsors were unwilling to meet the bills – ironically this was partly because of the success of the journal led to rapid growth in its size and cost. SISSA brought in the UK Institute of Physics as a partner and converted the journal to a subscription model. (It is still freely available to researchers in low income and developing countries.)

Clearly we need to be cautious in interpreting these data insofar as they apply to a relatively small number of journals. Taken together, though, the data suggest at the least that achieving widespread sustainability for open access journals will not be particularly quick or easy.

One thing that does appear clear is that there are likely to be different experiences in different disciplines: one size does not fit all. An author-pays model is much less likely to succeed in areas where direct research funding is low, such as mathematics.

Impact on research-intensive institutions

One concern with open access publication charges is that they may fall disproportionately on research-intensive institutions, where the aggregate bill for publication charges higher than their current subscriptions. (Conversely heavy users but low contributors to journals, such as industry, would gain.) Davis at Cornell analysed the holdings of a sample of ARL (Association of Research Libraries) institutions and compared the subscription costs to the potential costs under an author-pays model. He found that that the institutional cost of funding the scholarly journals system for the vast majority of ARL institutions would be likely to be higher under an author-pays model than current subscription fees[50].

Corporate subscribers

About 20% of journal subscription revenue comes from subscriptions by corporations who while heavy readers of journals, write proportionately fewer articles (about 5% of total articles). Switching to an author-side payment system would thus disproportionately benefit industry at the expense of academia. There would in effect be a transfer of costs within the total journal system from corporations to universities and research institutions, and it seems unlikely that this is an extra burden that stretched academic budgets would be in a position to shoulder.

CERN plan

A task force at CERN (the European high-energy physics laboratory) announced in June 2006 a plan to convert particle physics journals to open access. As reported by Peter Suber in the September issue of SPARC Open Access News:

> *CERN is putting together a coalition of funding agencies, laboratories, libraries, and scientists to convert willing journals in the field of particle*

> *physics to OA. The plan is to raise the money to pay reasonable processing fees for every article in participating journals. The journals could drop subscriptions, go full OA, and charge neither readers nor authors. This is the first time that any organization has tried to convert all the TA [toll access] journals in a field to OA.*

It is far too early to know how successful the CERN plan will be.

Waivers of publication charges

For the author-pays open access model to be viable, the fees charged must cover at least the direct and indirect costs of processing and disseminating the article, plus some margin for central administration costs, reinvestment (e.g. in electronic systems), and profit (or "surplus" in the case of not-for-profits). (In some special cases, such as medical journals, it is possible to raise revenues from advertising and corporate reprints, but this is very much the exception for scholarly journals.)

Most if not all full open access journals offer to waive the publication charge in the case of financial hardship or for authors from low-income countries. (By contrast, few if any optional open access schemes offer waivers.) For these journals, therefore, the actual amount received per article published depends on the proportion of authors who claim the waiver. At present we are unaware of any systematic data on this but we note the following:

- Numerous author surveys have reported authors' unwillingness to pay publication charges at economic levels: for instance, see Figure 8 taken from the 2004 CIBER survey. This may be changing, as authors become more familiar with the concept of OA journals, and also, if more funding bodies encourage OA publishing and provide funding (see page 25).

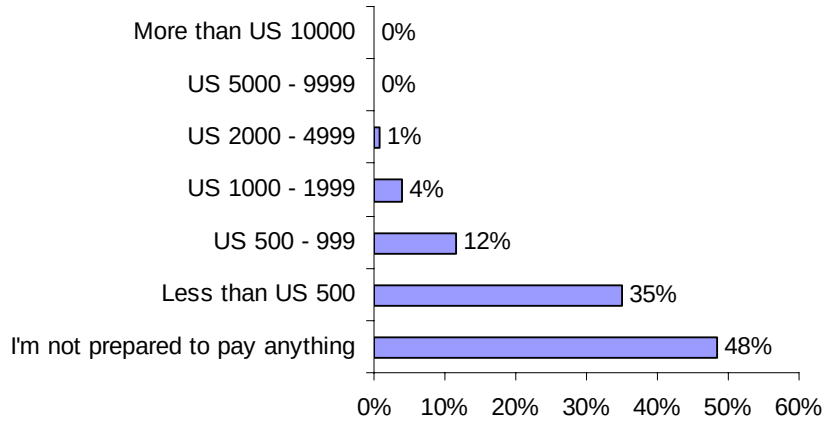

Figure 8: Amount authors said they were prepared to pay to publish in open access journals (from CIBER 2004 survey)

- *Nucleic Acids Research* (OUP) reported >90% of authors paid the publication charge ($950–1900) during its first year of full OA.
- The *New Journal of Physics*, an editorially successful open access journal launched in 1998, reports that waivers (on an article charge of $1080) also currently run at about 10% of authors[51].
- To ensure sustainability, an OA publisher would have to increase its article charges to account for both the level of waivers and non-payment. Developing world authors, usually but not exclusively the world's poorest, represent 25%

of all authors in ISI. If 25% can't or won't pay then OA article fees would need to be increased by 33% more than the average estimated article cost of $3750 to remain financially neutral.

Uptake of optional open access

When CIBER surveyed researchers in 2004 on their willingness to pay charges to publish in open access journals they found:

> "*There is little evidence here of much stomach on the part of authors to pay author charges at anything remotely near the rates that some commercial publishers claim are necessary for a long-term sustainable business model to develop. In fact, there is outright hostility*."

Things are clearly moving, however, and we saw earlier how authors' awareness and knowledge of OA increased between the 2004 and 2005 CIBER surveys. For optional OA journals, the interesting figure is the proportion of authors that chose to take up the OA option. OUP have reported some data from the first full year of its Oxford Open scheme, shown in Table 2 below[52]. A total of 360 OA papers were published, or 7.9% of the total of 4575. As expected, the data show greatest uptake in the life sciences, followed by medicine, with limited uptake in other disciplines. Three life sciences titles in the areas of molecular and computational biology have seen over 20% uptake. The highest of these was for *Bioinformatics*, which has published over 50 open access papers in 2006. Online subscription prices for 2007 have been adjusted for these journals to reflect this uptake.

Table 2: Uptake of Oxford Open optional open access scheme (data is for first full year, July 2005–June 2006)

Subject area	*No. of journals*	*Papers published*	*OA papers*	*% Uptake*
Medicine	21	2019	104	5.2
Life sciences	16	2203	246	11.2
Social sciences & humanities	9	293	8	2.7
Maths	2	60	2	3.3
Total	48	4575	360	7.9

Blackwell's Online Open optional open access scheme was launched in late 2005, initially on a trial basis. They reported to us that in 2006 up to 1 August, a total of 73 articles had been published under the scheme, or less than 0.14% of Blackwell's article output. This understates the uptake, however, because OO is only available for a small fraction of Blackwell's journals at present (98 journals out of 665).

Open access via self-archiving

The other ("green") route to open access is by self-archiving, that is, where the author deposits a version of the published article, typically a pre- or post-print, in an open repository. This repository might be an *institutional repository* run by the institution (typically a university) or a central *subject-based repository* (such as arXiv in physics and PubMed Central in biomedicine).

Versions of articles

We distinguish three main versions of the article that the author might archive:

- The pre-print: this is the author's final draft manuscript of the article prior to submission to a journal and the peer review process.
- The post-print: the author's manuscript after it has been peer reviewed and the comments of the reviewers and the journal editor added, but prior to copy editing and other additions (such as reference links).
- The publisher's version: the final version following copyediting, typesetting and layout, tagging for reference linking and links to other services. If provided to the author, this would typically be a PDF file.

The term "e-prints" is sometimes used to refer to both pre-prints and post-prints.

It is worth noting that the stand-alone PDF file may lack some capabilities compared to the version on the publisher's system (e.g. internal navigation, multi-resolution images, the ability to download references to software such as EndNote). As scientific communication increasingly incorporates "born-digital" electronic content (e.g. interactive molecular structures, built-in links to source data), the "flat" PDF version will increasingly diverge from the definitive online version. (Although, at present, authors still tend to print out articles to read[53].)

The proliferation of different versions of an article on the web potentially creates problems for authors in determining which is the definitive version. An ALPSP/NISO Working Group has started to look at this issue[54].

Publishers' policies on self-archiving

Perhaps surprisingly, most publishers have fairly liberal policies on allowing authors to archive copies of their articles on the web. A database of publisher policies is maintained by the SHERPA/RoMEO project[55]; of the 169 publishers included:

- 45% allow archiving of both pre- and post-print
- 25% allow archiving of post-print
- 15% allow archiving of the pre-print
- 21% do not formally support archiving.

Some publishers also allow authors to archive the final publisher version, though this is rarer. Some publishers add riders, such as requiring a link from the pre- or post-print to the publisher's final online version. Publishers are, however, beginning to introduce embargo periods (i.e. not allowing self-archiving for a set period after publication) with a view to protecting subscriptions.

Subject-based repositories

Central subject-based repositories have been around for much longer than institutional repositories. One of the first is arXiv, established in 1991 at Los Alamos by Paul Ginsparg. arXiv[56] (which pre-dates the world wide web) was designed to make efficient and effective the existing practice of sharing article pre-prints in high-energy physics. Perhaps because it built on this existing "pre-print culture" and because high-energy physicists were early adopters of electronic networks, it was enthusiastically adopted by this community, so much so that virtually all articles in the field are self-archived in at least a pre-print form. arXiv has now expanded its coverage to some (but by no means all) other areas of physics, mathematics, computer science and quantitative biology. It currently holds 383,443 e-prints.

RePEc (Research Papers in Economics)[57] was another early repository, again building on the pre-existing culture in economics of sharing pre-publication articles known as working papers. RePEc now holds records for 185,000 working papers, 213,000 journal articles and other content. It differs from arXiv in that many of the journal article records are for abstracts and bibliographic information only, with links to the publisher's site for the full version.

A subject-based repository of great current interest to publishers is PubMed Central (PMC). Rather than originating in volunteer efforts from the community itself, PMC is a project of the US National Institutes of Health (NIH). It builds on PubMed, the earlier bibliographic database that includes Medline, by adding full text. PMC works with publishers who voluntarily deposit the full text, which can be made available immediately (for full open access journals) or after an embargo period (for delayed open access journals). PMC has also worked with publishers to digitise back content, which must then be made freely available. More recently (in support of the NIH funding policy discussed below) PMC has accepted post-prints from authors for archiving. At the time of writing (September 2006) there were 3719 such post-prints available on PMC. Recently it was announced that the British Library would host and manage a UK version of PMC.

Institutional repositories

An institutional repository is an online database for collecting and preserving – in digital form – the intellectual output of an institution, particularly a research institution.

For a university, this would include materials such as research journal articles (i.e. pre-prints and post-prints), and digital versions of theses and dissertations, but it might also include other digital assets generated by normal academic life, such as administrative documents, course notes, or learning objects.

The two main objectives for having an institutional repository are:

- to provide open access to institutional research output by self-archiving it;
- to store and preserve other institutional digital assets, including unpublished or otherwise easily lost ("grey") literature (e.g., theses or technical reports).

Universities can also benefit from showcasing their research outputs.

The IR movement dates from the early 2000s with the launch of DSpace at MIT in 2002 and the slightly earlier development of Eprints software at Southampton.

IR software uses a technical standard (OAI-MHP) that enables the article metadata to be harvested by special search engines such as OAIster or Google Scholar. This allows users to relatively find articles of interest regardless of which institutional repository hosts them, though this distributed search is less effective than a centralised database such as PubMed, which uses a controlled vocabulary (or taxonomy) of keywords.

The number of IRs has grown (and is growing) rapidly. The Eprints project maintains an information database of repositories; it currently lists a total of 735 archives of which 369 are identified as institution or department level research repositories.

The numbers of articles deposited by authors in their IRs has grown much more slowly, and most of these IRs (except in the Netherlands) are nearly empty with just "a few hundred"[58] articles. (The total number of articles included in the 369 repositories listed by Eprints is about 1.2 million, or a mean of 3200, but the distribution is very skewed with a small number of large repositories and a long tail of small ones.) At present it appears that the large majority of authors are either ignorant of or indifferent to the potential benefits of self-archiving. Stevan Harnad

estimates that there is an upper limit on what advocacy and persuasion can achieve in terms of the rate of voluntary deposit of e-prints of about 15% of eligible articles.

Self-archiving policies and mandates

A newer and fast-moving development in self-archiving has been the introduction by research funders and by institutions of policies to request or mandate (i.e. require) authors to self-archive their research papers in open repositories. The motivation has been partly the slow voluntary uptake by authors of self-archiving and also a feeling that publicly funded research should be made publicly available.

Research funders with such mandates now include:

- US National Institutes of Health: as the largest funder of biomedical research in the world by far (its budget is about $28 billion), this is a particularly important mandate. At present, NIH policies "requests and strongly encourages" (rather than requires) authors to deposit their final post-print in PubMed Central at the date of publication, with the option to delay open access release for up to 12 months.
- Wellcome Trust (UK): another large biomedical funder. Unlike NIH, Wellcome *requires* authors to deposit articles in PMC (or UK PMC once this is established), no later than 6 months after publication.
- The UK Research Councils have decided to have separate policies reflecting the different research communities' attitudes to self-archiving. The Medical Research Council has followed the Wellcome Trust's lead, requiring deposit within 6 months of publication. The Economic and Social, the Biotechnology and Biological Sciences and the National Environmental Research Councils also require deposit, in this case "at the earliest opportunity".
- CNRS (France) is establishing its own e-print archive and says its "highly incentivised institutional self-archiving policy for our researchers will ensure that the majority of CNRS publications [are] deposited in the archive" (time period not specified) – i.e. strong encouragement but not a formal requirement.
- Deutsche Forschungsgemeinschaft (DFG, Germany) expects deposit in an archive, where possible, with discipline-specific delay periods of 6-12 months – i.e. strong encouragement but not a formal requirement.

Universities or university departments with deposit policies include:

- Queensland University of Technology (Australia): requires all research outputs (not just journal articles) to be deposited, subject to some exclusions. 2400 research articles deposited to date.
- University of Southampton Department of Electronics and Computer Science (UK): departmental policy requires deposit. 1480 research articles deposited to date.
- University of Lund (Sweden): deposit of research articles is strongly recommended but not formally required. 6240 research articles deposited to date.

Costs of repositories

There is a wide range of reports of the costs of introducing and managing an institutional repository. DSpace at MIT estimated its annual running costs at $285k (staff $225; operating costs $25k; $35k)[59]. A survey for ARL in 2006[60] found start-up costs ranged from $8,000 to $1,800,000, with a mean of $182,550 and a median of

$45,000. The range for ongoing operations budgets for implementers is $8,600 to $500,000, with a mean of $113,543 and median of $41,750.

NIH has reported that the cost of handling submissions and administering the policy was $1 million for fiscal 2005. If the compliance rate grew to 50%, the cost would grow to $2 million/year. If the compliance rate were 100% (65,000 articles/year), the cost would be $3.5 million/year.

Effect of self-archiving on journals

Perhaps not surprisingly, publishers are concerned about the possible impact of widespread self-archiving of journal articles. The common-sense hypothesis is that if compulsory mandates lead to very high levels of deposit, libraries (whose budgets are likely to remain under pressure indefinitely) will increasingly choose to rely on the self-archived version rather than subscribe to the publisher's version.

Some support for this hypothesis was given by a recent survey of librarians undertaken by Ware[61]. Availability of articles in repositories was cited as either a "very important" or an "important" possible factor in journal cancellation by 54% of respondents, even though ranking fourth after (i) decline of faculty need, (ii) reduced usage, and (iii) price. When respondents were invited to think forward five years, availability in a repository was still the fourth-ranking factor, but the relevant percentage had risen to 81%. Whilst this is not evidence of actual or even intended cancellation as a consequence of the growth of repositories, it suggests that repositories are an important new factor in the decision process, and one which is growing in significance.

There is certainly evidence that self-archiving can lead to reduced article downloads from the publisher's website. The Institute of Physics, American Physical Society and London Mathematical Society have reported[62] that journals covered by arXiv experience significantly fewer downloads from their sites than other comparable journals. In the case of IOP, the figures quoted are dramatic, with core high energy physics titles experiencing 100 times fewer downloads from the IOP site.

A reduction in usage or downloads is important because low or declining journal usage is one of the key factors used by librarians to select journals for cancellation[63]. There is also some trend in the market for usage to be a factor in pricing journals.

At the time of writing, however, hard evidence for a causal link between self-archiving and subscriptions is thin. For instance, the Institute of Physics and American Physical Society have reported that there is no detectable impact on subscriptions of the journals covered by arXiv (i.e. their rate of subscriptions decline was no faster than other comparable physics journals). But publishers, understandably enough, are keen not to wait until there is incontrovertible evidence of damage to subscriptions. A major study on this issue by Scholarly Information Strategies for the Publishing Research Consortium is due to report during October.

A key issue in this regard is the existence and length of any permitted embargo periods. Publishers argue that reducing or eliminating embargoes, as has been proposed in relation to funder mandates, for instance, would put journal subscriptions at greater risk. Some evidence to support this can also be found in Ware's survey: librarians said that delayed open access was not a factor in cancellations – the large majority (92%) said the embargo would have to be less than 6 months for them to consider cancelling the subscription. (The critical embargo period varied by discipline and was shortest for STM journals.)

New developments in scholarly communication

The system of scholarly communication, of which journals form one part, is developing in response to opportunities created by new network technologies. Scientists are using wikis, blogs, social bookmarking tools (e.g. Connotea), collaborative news sites (Dissect Medicine) and other new tools. One interesting and innovative example of a possible new online-only model is *Signaling Gateway*, a collaboration between the Nature Publishing Group and the Alliance for Cellular Signaling. It combines raw data with reports on experimental procedures and protocols, and reference datasets (managed by AfCS); structured data on 3500 proteins (edited and collated by NPG); and news and comment (written and commissioned by NPG). The site is partly funded by advertising.

Some potential areas for further research

Some areas that are not well understood or present current challenges include:

- The impact of open access self-archiving on journal subscriptions is an important issue that clearly needs further study. As noted above, a major study on this issue by Scholarly Information Strategies for the Publishing Research Consortium is due to report during October.
- Usage statistics: there are no publicly available statistics on article usage comparable to the ISI citation data. The issue will become more complex as self-archiving grows.
- Improved measures of article impact to supplement citations and impact factor, perhaps based on downloads.
- Multiple versions: we discussed three different versions (pre-print, post-print, final publisher's version) but we could have identified more and the problem of multiple versions could get more complex if new types of peer review (such post-publication review and commentary) become common. ALPSP has initiated a project with NISO to address this problem.
- Research looking at how scientists make use of journals and related resources in their research, how access to journals adds value and whether the overall system could be developed to enhance research productivity.
- The true extent of unmet demand for journals.

Postscript

It is important to base policy on good data that accurately describe the industry in meaningful ways. In the past, for instance, some have compared journal prices (e.g. between types of publisher) without taking into account the size of the journal or the number of articles included, which is clearly meaningless.

We also need to recognise that the publishing industry and the wider environment are both changing rapidly and not devise solutions to yesterday's problems. To take another example from journal pricing, journals are increasingly purchased in discounted multi-journal licences, which makes discussion based on individual print journal prices increasingly misleading.

We hope this briefing paper will help foster an informed debate on the issues facing scientific journal publishing.

Notes and references

[1] This $5bn figure averages estimates from EPS ($4.7bn) and Simba ($5.4bn)

[2] Elsevier estimates

[3] Ulrich's web directory listed 22,808 active, peer-reviewed scholarly/academic journals on 21 August 2006

[4] Mabe, M. (2003). The growth and number of journal, *Serials* 16(2) 191-197

[5] From a presentation by M. Mabe, 6th Nordic Interlending Conference, Trondheim, 2004

[6] McVeigh, M.E. (2004) Open Access Journals in the ISI Citation Databases, Thomson Corporation

[7] Figure taken from NSF Science and Engineering Indicators 2006

[8] NSF Science & Engineering Indicators 2006

[9] Mabe M.A. and Amin M. (2002). Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing, *Aslib Proceedings: new information perspectives*, 54(3) 149-157

[10] Figure taken from NSF Science and Engineering Indicators 2006

[11] UNESCO (2005). UNESCO Science Report

[12] King, D. and Tenopir, C. (2004). An evidence based assessment of author pays. Nature Web Focus on Access to the Literature

[13] Mabe M.A. and Amin M. (2002). Dr Jekyll and Dr Hyde: author-reader asymmetries in scholarly publishing, *Aslib Proceedings: new information perspectives*, 54(3) 149-157

[14] King, D. and Tenopir, C. (2004). An evidence based assessment of author pays. Nature Web Focus on Access to the Literature

[15] Tenopir, C. and King, D. (2000). Towards electronic journals: realities for scientists, librarians, and publishers. Special Libraries Association, Washington

[16] These data, from NSF Science & Technology Indicators 2006) are based on journals in the Science Citation and Social Sciences Citation Indexes published by ISI.

[17] Figure taken from NSF Science and Engineering Indicators 2006

[18] Seglen, P. 1992. The Skewness of Science. *Journal of the American Society for Information Science* 43, 628–638

[19] See http://www.uksg.org

[20] For a fuller discussion, see Amin, M. and Mabe, M. (2000). Impact Factors: Use and Abuse. Elsevier White Paper, http://tinyurl.com/f4akm

[21] ALPSP Alert No. 112 August 2006

[22] King, D. (2006). Scholarly journal pricing: a literature review & commentary (pre-print)

[23] E.g. the Open Society Institute's (OSI) "Guide to Business Planning for Launching an Open Access Journal" uses guidelines of US$3750 per article when identifying minimum revenues required per article. Publishing industry consultants John Cox Associates also estimated mean costs per article to be between US$3500 and US$4000 in 2004.

[24] Baldwin, C. (2004). What do societies do with their publishing surpluses? ALPSP/Blackwell

[25] http://www.arl.org/stats/arlstat/

[26] Tenopir, C. and King, D. (2000). *Towards Electronic Journals: Realities for Scientists, Librarians, and Publishers*. Washington, D.C.: Special Libraries Association

[27] Creaser, C., Maynard, S. and White, S. (2005). LISU Annual Library Statistics 2005, p. 133

[28] This list is adapted from King, D. (2006). The economic cost of journal publishing: a literature review and commentary (pre-print)

[29] King et al. (1981); Tenopir & King (2004)

[30] Figure from presentation by Mabe, M. (2004), using data from Tenopir and King 2000

[31] Rowlands, I., Nicholas, D., and Huntingdon, P. (2004). Scholarly communication in the digital environment: what do authors want, CIBER report; and Rowlands, I. and Nicholas, D. (2005). New Journal Publishing Models, CIBER report.

[32] E.g. presentation by Mabe, M. at Council of Science Editors Annual conference, Tampa, 2006: Journal futures: How will researchers communicate as the Internet Matures?

[33] Internal research by Elsevier reports that 65 per cent of authors who claimed to have last published in an open access journal had in fact published in a "traditional" journal, but the trend is still notable.

[34] Monograph and Serial Expenditures in ARL libraries 1986–2004, in ARL Statistics (2004), http://www.arl.org/stats/arlstat/graphs/2004/monser04.pdf

[35] Creaser, C., Maynard, S. and White, S. (2005). LISU Annual Library Statistics 2005

[36] Sanville, T. (2001). A Method out of the Madness: OhioLINK's collaborative response to the serials crisis, *Serials* 14(2)

[37] http://www.jisc.ac.uk/uploaded_documents/nesli2_usstudy.pdf

[38] Rowlands, I. and Olivieri, R. (2006). Journals and scientific productivity: a case study in immunology and microbiology. Publishing Research Consortium

[39] Some would also argue that to be open access the user must also be permitted to make free use of the article, e.g. in teaching or writing, subject only to minimal constraints such as attribution and non-commercial re-use. In this paper we shall ignore this distinction since the key debate is focussed on access.

[40] Willinsky, J. (2003). The Nine Flavours of Open Access Scholarly Publishing. *J postgrad med*, 49, 263–267

[41] http://www.plantphysiol.org/cgi/content/full/142/1/5

[42] Kurtz, M. et al. (2004). The effect of use and access on citations. *Information Processing and Management 41(6), 1395–1402*

[43] Moed, H.F. (2006). New developments in citation analysis and research evaluation, *Information Services and Use* 26 (2), 135–137.

[44] Eysenbach, G. (2006). Citation advantage of open access articles. *PLoS Biology* 4(5)

[45] http://www.oxfordjournals.org/news/oa_report.pdf

[46] Source: publisher websites, September 2006

[47] Presentation by Martin Richardson in June 2006

[48] Butler, D. Open-access journal hits rocky times, in *Nature*, 20 June 2006

[49] BioMed Central Ltd's annual accounts for 2005, from Companies House, UK.

[50] Davis, Philip and the Cornell University Library Task Force on Open Access Publishing. Dec 22, 2004. Calculating the Cost per Article in the Current Subscription Model.

[51] John Haynes (IOP), personal communication, September 2006

[52] Mithu Mukerjee (OUP), personal communication, September 2006

[53] E-Journal User Study (E-Just) – see http://ejust.stanford.edu/

[54] See http://www.niso.org/committees/Journal_versioning/JournalVer_comm.html

[55] http://www.sherpa.ac.uk/romeo.php

[56] www.arxiv.org

[57] http://repec.org/

[58] Van Westrienen, G. & Lynch, C. Academic Institutional Repositories. *D-Lib Magazine*, 11(9)

[59] MIT (2003) MIT DSpace—A Case Study. http://dspace.org/implement/case-study.pdf

[60] Bailey, C.W. (2006). SPEC Kit 292: Institutional repositories, Association of Research Libraries

[61] For a summary of the survey, see: Ware, M. (2006.) Open archives and their impact on journal cancellations, *Learned Publishing* 19(3)

[62] E.g. see http://www.alpsp.org/events/2005/PPR/default.htm

[63] Ware, M., ibid.