stm

The stm report

An overview of scientific and scholarly journal publishing

Mark Ware (Mark Ware Consulting)

Michael Mabe (STM)

September 2009

Published by

International Association of Scientific, Technical and Medical Publishers
Prama House
267 Banbury Road
Oxford OX2 7HT
United Kingdom

About STM

STM is a broad and welcoming organisation which includes large and small companies, not for profit organisations and learned societies, traditional primary and secondary publishers and new players. Members have the opportunity to participate in a body central to the well being of our industry.

The mission of STM is to create a platform for exchanging ideas and information and to represent the interest of the STM publishing community in the fields of copyright, technology developments, and end user/library relations.

STM Aims and Objectives

to assist publishers and their authors in their activities in disseminating the results of research in the fields of science, technology and medicine;

to assist national and international organisations and communications industries in the electronic environment, who are concerned with improving the dissemination, storage and retrieval of scientific, technical and medical information;

to carry out the foregoing work of the Association in conjunction with the International Publishers Association (IPA) and with the national publishers associations and such other governmental and professional bodies, international and national, who may be concerned with these tasks.

STM participates in the development of information identification protocols and electronic copyright management systems. STM members are kept fully up to date (via newsletters, the STM website, and e-mail) about the issues which will ultimately affect their business. STM organises seminars, training courses, and conferences. Its General Assembly is held annually, one day preceding the Frankfurt Book Fair.

Mark Ware Consulting provides publishing consultancy services to the STM and B2B sectors. For more information see www.markwareconsulting.com.

Contents

Executive summary

Scholarly communication and STM publishing

1. STM publishing takes place within the broader system of scholarly communication, which includes both formal and informal elements. Scholarly communication plays different roles at different stages of the research cycle, and (like publishing) is undergoing technology-driven change. Categorising the modes of communication into one-to-one, one-to-many and many-to-many, and then into oral and written, provides a helpful framework for analysing the potential impacts of technology on scholarly communication (see page 8).

2. 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 (page 12).

The STM market

3. The annual revenues generated from English-language STM journal publishing are estimated at about $8 billion in 2008, up by 6-7% compared to 2007, within a broader STM publishing market worth some $16 billion. About 55% of global STM revenues (including non-journal STM products) come from the USA, 30% from Europe, 10% from Asia/Pacific and 5% from the rest of the world (page 16).

4. The industry employs an estimated 110,000 people globally, of which about 40% are employed in the EU. In addition, an estimated 20–30,000 full time employees are indirectly supported by the STM industry globally in addition to employment in the production supply chain (page 16).

5. Although this report focuses primarily on journals, the ebook market is evolving and growing rapidly (page 16).

6. 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 (page 24).

7. There were about 25,400 active scholarly peer-reviewed journals in early 2009, collectively publishing about 1.5 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 between 5.5 and 10 million, depending on definition, although only about 20% of these are repeat authors (pages 18, 23).

8. The USA currently dominates the global output of research papers but the most dramatic growth has been in China and East Asia. China's compound growth rate of 17% per year over the decade to 2005 led to its overtaking all other countries except the USA (page 21).

Research behaviour and motivation

9. Despite a transformation in the way journals are published, researchers' core motivations for publishing appear largely unchanged, focused on funding and furthering the author's career (page 36).

10. Reading patterns are changing, however, with researchers reading more, averaging 270 articles per year, but spending less time per article, with reading times down from 45-50 minutes in the mid-1990s to just over 30 minutes. Access and navigation to articles is increasingly driven by search rather than browsing (page 27).
11. The research community continues to see peer review as fundamental to scholarly communication and appears committed to it despite some perceived shortcomings. The typical reviewer spends 5 hours per review and reviews some 8 articles a year. Peer review is under some pressure, however, notably from the growth in research outputs, including those from emerging economies (page 25).
12. There is growing interest in research and publication ethics, illustrated by the increased importance of organisations like the Committee on Publication Ethics (COPE) and the development of technology solutions to address abuses such as plagiarism (page 38).

Technology

13. The vast majority of STM journals are now available online, with 96% of STM and 87% of arts, humanities and social sciences journals accessible electronically in 2008. In many cases publishers and others have retrospectively digitised early hard copy material back to the first volumes. The proportion of electronic-only journal subscriptions has risen sharply, partly driven by adoption of discounted journal bundles (page 19).
14. Social media and other "Web 2.0" tools have yet to make the impact on scholarly communication that they have done on the wider consumer web. Most researchers do not for instance read blogs regularly or make use of emerging social tools. This may be for a variety of reasons: a reluctance to introduce informal processes into the formal publication process; because the first wave of tools did not take sufficient account of the particular needs of researchers; a lack of incentives for researchers, including the lack of attribution for informal contributions; a lack of critical mass; and simply a lack to time to experiment with new media (page 58ff.).
15. The explosion of data-driven research will challenge publishing to create new solutions to link publications to data, to facilitate data-mining and to manage the dataset as a potential unit of publication (page 59).
16. The much-discussed semantic web, although potentially difficult and expensive to achieve in a formal, comprehensive way, is starting to emerge in pragmatic, domain-bounded approaches such as in chemistry and molecular biology. Semantic web technologies offer significant opportunities to increase research productivity by enhancing journals, improving search and discovery, enriching the user experience, facilitating text- and data-mining and in the longer term supporting the automatic extraction of knowledge from the research literature (page 60).

Business models and publishing costs

17. Aggregation on both the supply and demand sides have increasingly become the norm, with journals sold in packages to library consortia. More than half of journal subscriptions are now sold in bundles of more than 50 titles (page 14).
18. The "Big Deal" and similar discounted packages have been extremely successful in widening researchers' access to journals while simultaneously reducing the average cost per subscription and the average cost per article download. Although the bundle model is under pressure from librarians (e.g. for reasons of inflexibility, lack of control or out-dated pricing models) its benefits appear sufficient for it to remain the dominant business model for some time (page 14).

19. The number of current serials subscriptions per higher education institution in the UK has more than doubled in the 10 years to 2004/05, from 2900 to 7200. Research in 2004 found that 70% of researchers believed that access to journal literature was better or much better than 5 years ago, a finding that has been repeated in later surveys. 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 42).
20. There is growing interest in identifying and addressing specific barriers to access or access gaps, e.g. access by non-members to institution collections or by SMEs (page 43).
21. The average cost of publishing an article in a subscription-based journal with print and electronic editions was estimated by a RIN/CEPA 2008 study to be $3800 (excluding non-cash peer review costs). The study estimated that eliminating print editions would save about £1 billion globally (largely in library costs).
22. Journal publishing has become more competitive with the emergence of new business models. Open access posits making original research freely accessible on the web. There are three approaches: full open access, delayed open access and self-archiving (page 45).
23. There are between 3400 (according to the Open J-Gate directory) and 4300 (DOAJ) open access peer reviewed journals. OA titles appear somewhat less likely than other titles to appear in A&I databases such as Scopus, and are smaller on average than other journals. Consequently the proportion of the 1.5 million articles published per year that are open access is considerably lower than the proportion of journal titles. It is estimated that about 2% of articles are published in full open access journals, another 5% in journals offering delayed open access within 12 months, and under 1% under the optional (hybrid) model (page 20).
24. Gold open access has a number of potential advantages. It would scale with the growth in research outputs and there are potential system-wide savings. But there are major obstacles to widespread uptake: OA publication charges are currently significantly lower than the historical average cost of article publication; about 25% of authors are from developing countries; only about 60% of researchers have separately identifiable research funding; substantial restructuring of funding within universities would be required; and there would be winners and losers among existing journal subscribers, depending on their research intensity (page 52).
25. Research funders are playing an increasingly important role in scholarly communication. Their desire to measure and to improve the returns on their investments emphasises accountability and dissemination. These factors in turn first increase the importance of (and some say the abuse of) metrics such as Impact Factor and secondly lead to the growing number of mandates from funders requiring researchers to self-archive manuscripts in open repositories (page 49).
26. Many publishers remain concerned that Green open access (self-archiving) is essentially parasitical on journal publishing, with no sustainable business model of its own should it (as they fear) undermine journal subscriptions. This potential impact on subscriptions (and other aspects of self-archiving) is the subject of a major EU-funded study Publishing and the Ecology of European Research (PEER) which is due to report in 2011, but in the meantime some publishers and funders have reached bilateral agreements (page 56).

1 Scholarly communication

STM[1] publishing takes place within the broader system of scholarly communication, which includes both formal elements (e.g. journal articles, books) and informal (conference presentations, pre-prints). Apart from the academics (and their funders and host institutions) there are two main players in the scholarly communication supply chain: publishers (responsible for managing the quality control, production and distribution) and librarians (responsible for managing access and navigation to the content, and for its long-term preservation (though this latter role is changing with electronic publishing)).

1.1 The research cycle

The different roles played by scholarly communication can be understood in the context of the research cycle (with the communication role in parentheses) (see Figure 1, from Bargas, cited in Goble 2008):

- Idea discovery, generate hypothesis (awareness, literature review, informal)
- Funding/approval (literature review)
- Conduct research (awareness)
- Disseminate results (formal publication, informal dissemination)

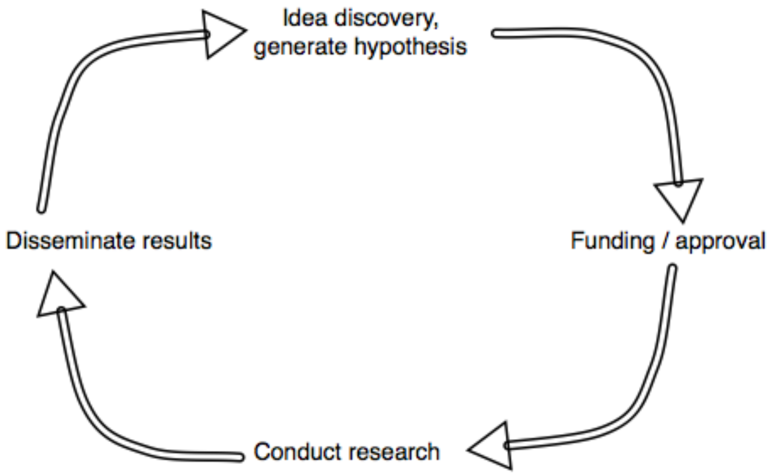

Figure 1: The research cycle

1.2 Types of scholarly communication

As noted above, scholarly communication encompasses a wide range of activities, from conference presentations, informal seminar discussions, face-to-face or telephone conversations, email exchanges, email listservs, formal journal and book publications, preprints, grey literature. One way of categorising scholarly communication is in terms of

[1] "STM" is an abbreviation for scientific, technical and medical but has several different meanings. It can be a model of publishing, in which case it includes social sciences and the arts and humanities. It is sometimes used to describe scientific journals. It is also the name of association of publishers ("***stm***") that is the sponsor of this report. We have employed all usages in this report and trust it is clear from the context which is intended.

whether it is public or private, and whether it is evaluated or non-evaluated. This is illustrated in Figure 2. In this report we are primarily concerned with formal, written communication in the form of journal articles. The boundary between formal and informal communications may be blurring in some areas (for instance, unrefereed author's original manuscripts on the arXiv repository are increasingly cited in formal publications, while journal articles are becoming more informal and blog-like with addition of reader comments) but if anything the central role of the journal article in scholarly communication is stronger than ever.

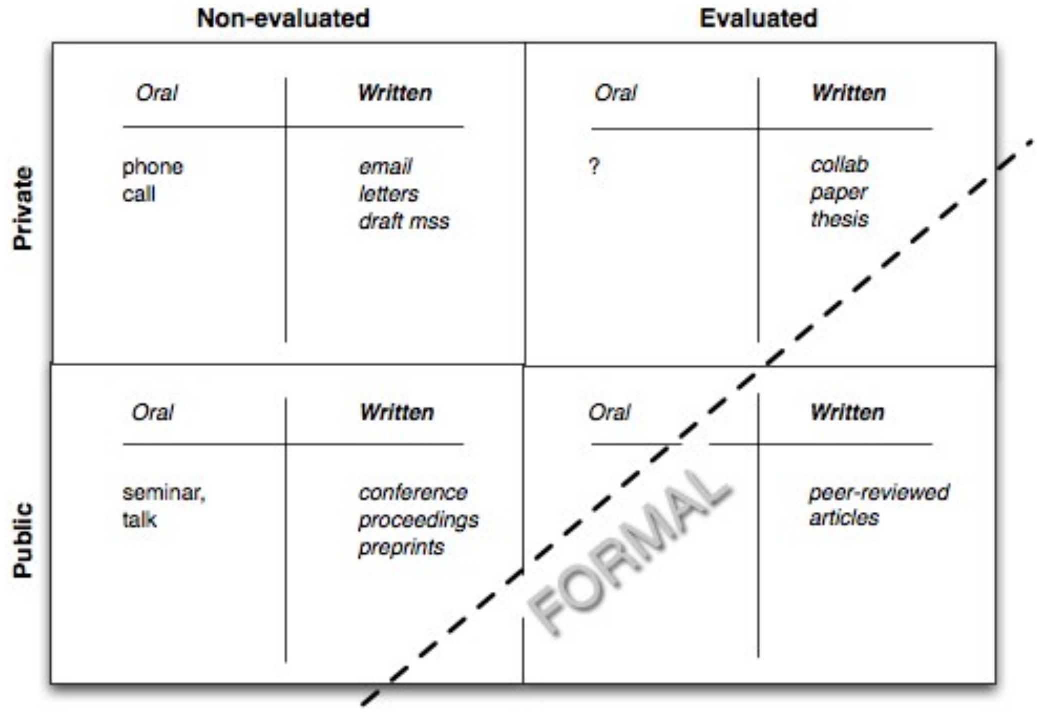

Figure 2: Formal and informal types of scholarly communication

We are also interested, however, in understanding how scholarly communication may be affected by current and future electronic means of communication. We can identify three basic modes for all kinds of human communication: one-to-one, one-to-many, and many-to-many. These can be further categorised into oral and written communications. By considering types of scholarly communication along these dimensions, as illustrated in Table 1, we can see that for the most part, the introduction of electronic and web-based channels has created new ways to conduct old modes of communication (for instance with web-based publications replacing printed publications) but has not offered wholly new modes. The exception is the wiki, which in providing a practical means of facilitating many-to-many written communication does offer something entirely without parallel in the offline world.

Table 1: Modes of communication

Mode	*Connection*	*Old instances*	*New instances*
Oral	One-to-one	Face-to-face conversation Telephone conversation	Instant messaging VOIP telephony Video calls
	One-to-many	Lecture Conference presentation TV/radio broadcast	Instant messaging Web video
	Many-to-many	Telephone conference call?	Web-based conferencing
Written	One-to-one	Letters	Email
	One-to-many	Printed publication	Web-based publications Blogs
	Many-to-many	n/a	Wikis e-whiteboards

1.3 Changes in scholarly communication system

The scholarly communication process is subject to profound transformative pressures, driven principally by technology and economics. At the same time, though, the underlying needs of researchers remain largely unchanged. Changes can be considered under three headings (see also Van Orsdel 2007)

- Changes to the publishing market (e.g. the SPARC programme, new business models like open access; new sales models such as consortia licensing)
- Changes to the way research is conducted (e.g. use of networks; growth of data-intensive and data-driven science; globalisation of research)
- Changes to public policy (e.g. research funder self-archiving mandates; changes to copyright)

The detail and implications of these changes will be discussed further in later sections.

1.4 The RIN Principles for scholarly communication

In 2007 the UK's Research Information Network published a document endorsed by key stakeholders (funders, libraries, research institutions, publishers, etc.) that set out principles for best practice under seven headings (RIN 2007):

1. the pursuit of research aimed at generating new knowledge and understanding
2. assuring the quality of the information outputs generated by researchers
3. ensuring appropriate recognition and reward for all those engaged in the scholarly communications process
4. presenting, publishing and disseminating information outputs digitally, orally, in print and other forms
5. facilitating access to and use of information outputs by researchers and others who have an interest in them
6. assessing and evaluating the usage and impact of information outputs
7. preserving digital, printed and other information outputs, so that those of long-term value are accessible for the indefinite future.

Journals are involved in most of these stages. Although RIN's focus is supporting scholarly communication in the UK, these principles do have universal application.

2 The journal

2.1 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 few clear dividing lines to outsiders. In this report, however, we are concerned predominantly with the scholarly and scientific literature: that is, periodicals carrying accounts of research published after due peer review rather than journalistically based magazines.

The journal has traditionally been seen to embody four functions:

- *Registration*: establishing the author's precedence and ownership of an idea
- *Dissemination*: communicating the findings to its intended audience
- *Certification*: ensuring quality control through peer review and rewarding authors
- *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. Alternatively this can be seen as part of the dissemination function.

2.2 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 3. 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. In the world of electronic publishing, of course, readers also obtain journal articles directly from the publisher in parallel to the library route.

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.

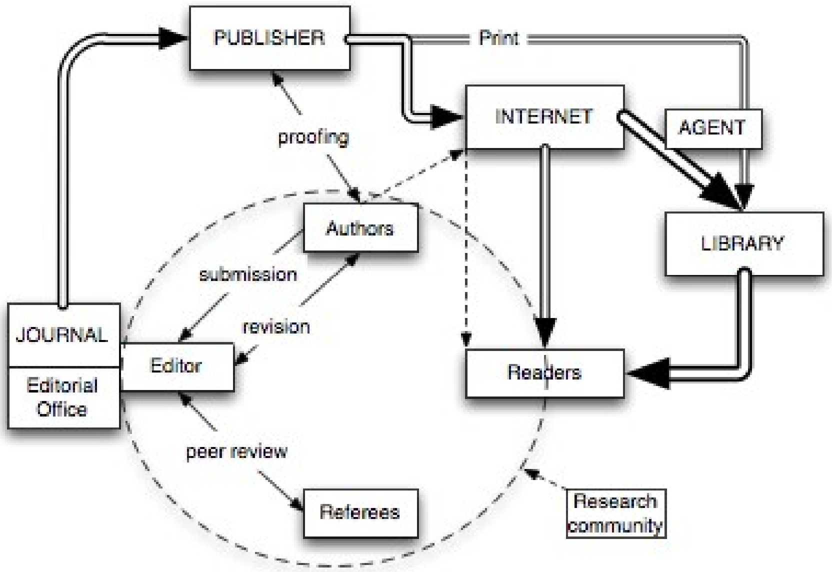

Figure 3: The publishing cycle

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. Peer review is discussed in more depth in a section below.

The role of the publisher has often been 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/electronic service provider** – copy editing, typesetting & tagging, and (for the time being) printing and binding the journals.
- **Marketeer** – attracting the papers (authors), increasing readership 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.

Versions of articles

One potential issue with the widespread adoption of self-archiving is that multiple versions of articles will be available to readers (and others, such as repository managers). In order to help create a consistent nomenclature for journal articles at various stages of the publishing cycle, NISO (National Information Standards Organization) and ALPSP have collaborated on a recommended usage (NISO 2008). The NISO recommended terms are

- AO = Author's Original
- SMUR = Submitted Manuscript Under Review
- AM = Accepted Manuscript
- P = Proof
- VoR = Version of Record
- CVoR = Corrected Version of Record
- EVoR = Enhanced Version of Record

For many purposes (such as much of this report) this represents a finer-grained structure than is necessary for discussing journal publishing. ***stm*** in its discussions with the EU and others refers instead to Stage 1 (the author's original manuscript), Stage 2 (the accepted manuscript) and Stage 3 (the final paper - any of the versions of record).

The term pre-print is also used to refer the author's original (and sometimes to the accepted manuscript), and post-print to refer to the accepted manuscript. These terms are deprecated because they are ambiguous and confusing (e.g. the post-print definitely does not occur post printing).

The CrossRef organisation has proposed the introduction of a CrossMark to identify the version of record. This would be a visible kitemark (logo) that would identify it to the human reader. There would also be defined metadata for search engines etc. The CrossMark would not just identify the article as the version of record but would also provide information about the pre-publication process (e.g. peer review) and of post-publication events such as errata, corrections and retractions.

2.3 Sales channels

Journals are marketed to two broad categories of purchaser, namely libraries and individuals. Although individual subscriptions (either personal or membership-based subscriptions) can be important for some journals (for example magazine/journal hybrids such as *Nature* and society journals), purchase and use of individual subscriptions has been falling for many years and as they are in any case typically priced at very high discounts, the large bulk of the journals market by revenue is made up of sales to libraries.

Traditionally library sales were in the form of subscriptions to individual journals. This is still an important part (currently around half) of the market but increasingly journals are sold as bundles of titles, either directly to libraries or to library consortia.

Subscription agents are an important part of the sales channel: the average library is estimated to place about 80% of its business via agents. Agents act on behalf of libraries, allowing the library to deal with one or two agents rather than having to manage relationship with large numbers of journal publishers, each with different order processes, terms & conditions, etc. Agents also provide a valuable service to publishers by aggregating library orders and converting them to machine-readable data, handling routine renewals, and so on. Discounts offered to agents by publishers have traditionally been lower than in many other industries and are falling (and not-for-profit publishers have traditionally not offered discounts at all) so that agents make their revenue by charging fees to libraries. Agents have a venerable history, with the first (Everett & Son) established in 1793. The Association of Subscription Agents[2] current lists about 40 agent members but the number of agents has been declining in recent years, primarily due to mergers and acquisitions with the industry and the lack of new entrants. One reason is the increasing disintermediation of their function brought about by move to electronic publishing and in particular the rise of consortia sales.

With the rise of electronic publishing, sales of individual journal subscriptions have fallen as a proportion of total sales in favour of **bundles**. One 2008 survey estimated that over half of all journals are now sold in bundles of 50 titles of more (Van Orsdel & Born 2009). According to Cox (2008), nearly all (95%) of large and most (75%) of medium publishers offer bundles of content, though this drops (for obvious reasons) to 40% of small publishers. Small publishers are more likely to participate in multi-publisher bundles such as the ALPSP Learned Journal Collection, BioOne or Project MUSE. Cox found that most publishers still priced bundles on the basis of the "prior print" model; that is, the library is offered electronic access to all the titles in the bundle at a price reflecting the library's existing print subscriptions (which are typically retained) plus a top-up fee for electronic-only access to the non-subscribed titles. This top-up model (especially when the bundle includes all of the publisher's output and the sale is to a consortia) is frequently referred to as the **Big Deal**. The other main pricing models include: usage-based pricing, which was tried during the mid-2000s but appears to have largely dropped from favour; pricing based on a classification of institutions by size, which also seems to be reducing in importance; and pricing based on the number of simultaneous users, which has been growing. A key issue for libraries is whether the publisher's licence term for bundles allows cancellations; Cox found that only 40% of publishers allowed cancellations, with commercial publishers interestingly being much more likely to permit cancellations than not-for-profits (46% vs 24%). Publishers are increasingly offering bundles that include non-journal content, particularly e-books, reference works and datasets. This is a trend that is likely to continue.

The growth of sales of titles in bundles has been paralleled by the increasing importance of sales of such bundles to **library consortia** (though it is important to recognise the two different concepts – some publishers deal with consortia but do not offer bundled content). Consortia arose in order to provide efficiencies by centralising services (e.g. shared library management systems, catalogues, ILL, resources etc.) and centralising purchasing, to increase the purchasing power of libraries in negotiation with publishers, and increasingly to take advantage of bundled electronic content. The numbers of consortia have been growing strongly: one industry directory (Cox 2009) recorded 338 active consortia in 2008, up from 164 in 2003, while the International Coalition of Library Consortia[3] has some 150 members. The true total depends somewhat on definitions (for instance not all library consortia purchase content, for instance about 40 of the 150 ICOLC members) but is probably somewhat larger than the Cox figure. The size and nature of consortia vary considerably, from national consortia to small regional ones, and include academic, medical, public, school and government libraries. The total number of individual libraries covered by consortia is of the order of 5000. According to Cox (2008), about half of publishers actively market to consortia (90% of larger publishers). Of these, about half use the same pricing model as for their bundles, with the balance negotiating on a case-by-case basis. Consortia deals are now typically (60%) for a 3-year period, with 30% on a 1-year and 10% on a 2-year basis, with price caps now more widespread. Cancellation terms are as previously covered for bundles.

---

[2] http://www.subscription-agents.org/

[3] http://www.library.yale.edu/consortia/

Library system vendors[4] provide the cataloguing, enterprise resource planning and link-resolver and other access systems used by libraries. Although their business relationship are thus primarily with libraries rather than publishers, they are an important part of chain that links readers to publishers' content. Publishers work with systems vendors on supply-chain standards such as ONIX for Serials[5] and KBART (Knowledge Bases And Related Tools)[6].

2.4 Journal economics and market size

The annual revenues generated from English-language STM journal publishing are estimated at about $8 billion in 2008, up by 6-7% compared to 2007.

It may be helpful to consider scholarly journal publishing into the context of publishing and the wider information industry. According to Outsell (2009), the overall information industry (which includes news, directories, and a wide variety of commercial information sources was worth some $400 billion in 2008. More relevantly, within this industry Outsell quantifies the "scientific, medical and technical information" sector at $23.7 bn. This figure includes not just publishers but geophysical data providers ; just including the traditional STM publishers gives a market size of about $16 bn. This figure includes all types of STM publishing such as books and reference works, databases, A&I services as well as journals. Journals at about $8 billion therefore make up about half of the overall STM market by value.

Journals publishing revenues are generated primarily from academic library subscriptions (68-75% of the total revenue), followed by corporate subscriptions (15-17%), advertising (4%), membership fees and personal subscriptions (3%), and various author-side payments (3%) (RIN 2008).

By geographical market, about 55% of global STM revenues (including non-journal STM products) come from the USA, 30% from Europe, 10% from Asia/Pacific and 5% from the rest of the world (adapted from EPS 2006). These proportions probably overstate the importance of the USA market for journals alone.

The industry employs an estimated 110,000 people globally, of which about 40% are employed in the EU. In addition, an estimated 20–30,000 full time employees are indirectly supported by the STM industry globally (freelancers, external editors, etc.) in addition to employment in the production supply chain (source: Elsevier estimates).

Books & e-books

This report focuses on journal publishing. It is worth noting, however, that electronic books are increasingly offered by STM publishers on the same electronic platforms as host their journals. (Unlike in trade publishing, the development of acceptable stand-alone e-book readers and the use of technical DRM solutions have not been particularly important in the STM e-book market.) Two-way electronic linking between book and journal content will become increasingly common (not just citation linking but linking of related material). Business models are still evolving but journal-type models such as subscription and "big deal"-type access will likely sit alongside outright purchase, purchase with periodic updating, and pay-per-view. Publishers are also likely to expand the bundling of journal and e-book content.

---

4 See http://www.librarytechnology.org/ for one overview and list of suppliers

5 http://www.editeur.org/onixserials.html

6 http://www.uksg.org/kbart

Academic libraries, particularly at institutions with teaching as well as research interests appear keen to develop e-book collections. According to one survey (Sharp & Thompson 2009) the top 5 reasons for this were user convenience (e.g. 24/7 and off-campus access); it was seen as part of a strategic move to electronic access; student demand for multiple copies (particularly for use around key deadlines such as exams); their searchability; and the reduced pressure on shelf space. Value for money was also important. The preferred purchasing models were (in declining order) individual titles purchased from aggregator; individual titles purchased from publishers; aggregator packages; publisher packages; and reader-driven acquisition (purchase triggered by use). Obstacles to uptake reported were difficulties in finding out what is available in electronic format (no equivalent of Books in Print); the variety and complexity of business models and licence terms; library staffing implications, with boundaries blurring between books and journals; too few textbooks available, although demand is perhaps strongest here. Other issues that have held back the development of an e-book market have included a variety of formats; diversity of software and hardware products and platforms; lack of agreement on standards; digital curation issues; and discoverability and access.

Despite these issues academic libraries are starting to build substantial e-book collections. The library catalogue for the Max Planck Society (MPG), for example contained nearly 41,000 e-books in late 2008 (substantially more than the number of journal titles), of which the largest collections were: IEEE Conference Proceedings (10,600), SPIE Conference Proceedings (6500), Safari TechBooks (6000), Springer Lecture Notes (4,800), NetLibrary (inc. free titles, 4500), and SourceOECD (3200).

Large journal publishers have been active at moving their book lists onto the existing electronic platforms. For example, Springer launched its e-book programme in 2006 with 10,000 titles; there are now around 30,000 with 5000 being added annually. Springer reported that 25% of the 130 million combined journal and e-book full-text downloads it delivered in 2008 were from e-books, and that 10% of its existing customers were subscribing to e-books (McClure 2009).

There does appear also to be unmet demand from students and their teachers for electronic textbooks. Publishers, however have been cautious about providing course texts online as there is a lack of evidence about demand and concerns over impacts on print sales, and all parties have been held back by uncertainty as the best pricing and licensing models. In the UK, the JISC national e-books observatory[7] is addressing these issues with a detailed study of student use of pilot list of online course texts at 127 participating universities. The study continues but initial findings seem to suggest that e-books supplement rather than replace print, in particular easing pressure on short loans collections, and indeed many of the textbooks in the study actually increased their print sales against what was expected.

ALPSP are currently conducting a survey on book and e-book publishing practice. Results are expected to be published in late 2009[8].

Possible impact of economic crisis

At the time of writing, the likely impact of the global economic crisis on STM publishing was not clear. In the USA, library and consortia associations issued statements cautioning that many if not most of their members would be facing budget cuts in the coming 2–3 years (ARL 2009; ICOLC 2009). Van Orsdel & Born (2009) also reported anecdotal accounts suggesting possible cuts of the order 5–15% for the next two years.

---

7 http://www.jiscebooksproject.org/

8 See the ALPSP website: http://is.gd/N6AU

It is possible budget pressures may accelerate switching to electronic-only subscriptions, assuming there is a price advantage (that is not swallowed by VAT charges) in doing so.

Another interesting speculation is that the recession may indirectly favour "author-pays" open access, insofar as the substantial funds proposed under stimulus packages in the USA and elsewhere may be more easily made available to pay publication charges (via research grants) than transferred to the library budget.

There is also much talk at time of writing of a retreat from the Big Deal in response to economic pressures on libraries in favour of title-by-title electronic access and pay-per-view options. While this may help ease financial pressure it will certainly reduce the extremely wide access that the Big Deal arrangements have allowed.

Advertising only makes up a small proportion of the STM market but is likely to be disproportionately hit, as in previous downturns. The pharma market was 10% down in 2008 according to Elsevier; this does not just relate to the economic crisis (the weakness of the new drug pipeline is a major factor) but this important area for many medical journals is likely to be weak.

2.5 Journal and articles numbers and trends

There were about 25,400 active scholarly peer-reviewed journals in early 2009[9], collectively publishing about 1.5 million articles a year (Björk, Roos & Lauri 2009).

Journals which published only research, comprise about 95% of journals, with the balance consisting of the so-called hybrids, academic journals with extensive journalistic content that effectively weld magazine and research journal characteristics together. These hybrids are sold to both individuals and institutions, have high circulation and significant advertising revenues – which the pure research journals do not have (from Mabe 2008). The largest single subject area is biomedical, representing some 30% of journals, with arts & humanities a minority at under 5%.

An important subset is the 9360 journals included in the Journal Citation database, of which 6400 are in the Science Edition, 1800 in the Social Sciences and 1160 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. Journals included in the Thomson citation database are also on average substantially larger than those not included (publishing 111 articles per year compared to 26, according to Björk *et al.* 2009)

The number of peer reviewed journals published annually has been growing at a very steady rate of about 3.5% per year for over three centuries (see Figure 4), 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 5, which plots the increase in numbers of articles and journals alongside the numbers of US researchers. Similar data is available for other OECD countries confirming this effect (source: Elsevier).

---

9 Ulrich's web directory listed 25,378 active, peer-reviewed scholarly/academic journals on 16 March 2009. Thomson Reuters' Web of Knowledge database covers some 23,000 journals, while Scopus covers "over 16,000 peer-reviewed journals"

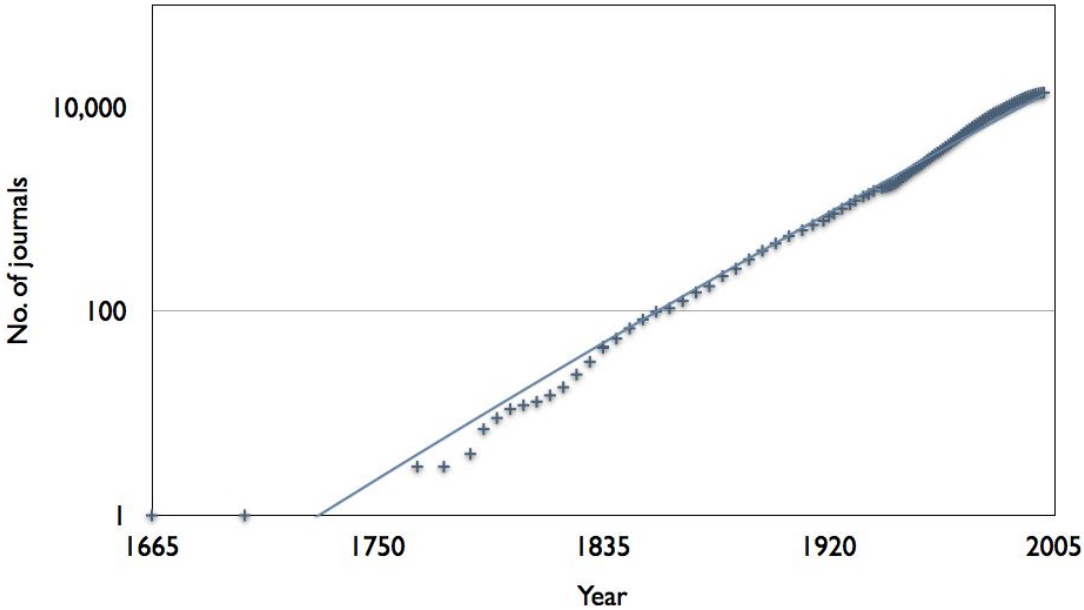

Figure 4: The growth of active, peer reviewed learned journals since 1665 (Mabe 2003)

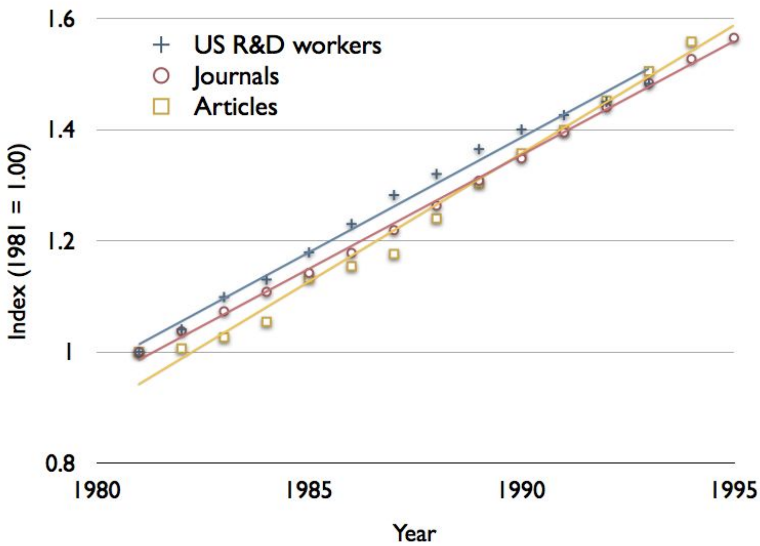

Figure 5: Relationship between numbers of researchers, journals and articles (Mabe 2004, using data from ISI and NSF)

Online journals

The vast majority of STM journals are now available online. According to the ALPSP's 2008 report on scholarly publishing practice (Cox & Cox 2008), 96% of STM and 87% of arts, humanities and social sciences journals were accessible electronically in 2008. This represented a steady increase compared to comparable surveys conducted in 2003 (STM 83%, AHSS 72%) and 2005 (STM 93%, AHSS 84%). Only 10 of the publishers surveyed had yet to put their journals online; all were small publishers and six were university presses.

The rate at which libraries are switching from print to electronic subscriptions is illustrated for UK universities in Figure 6.

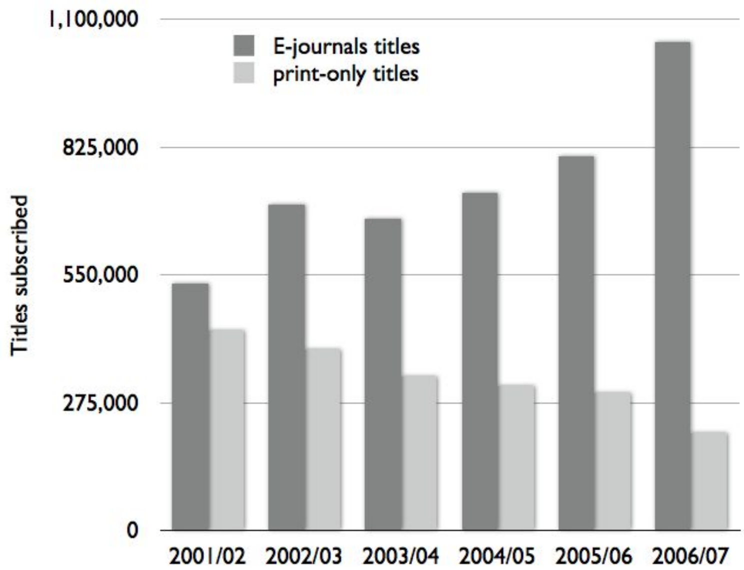

Figure 6: The total numbers of electronic and print-only journal titles subscribed to by the 115 UK universities (Source: SCONUL, from RIN 2009)

Open Access journal and article numbers

The number of open access journals listed by the Directory of Open Access Journals was just about 4360 as of September 2009[10], with the number increasing by about 2 per day. Not all journals in DOAJ are peer-reviewed (though all exercise some form of quality control through an editor, editorial board or peer review). A study by Morris also showed that the journals covered by the DOAJ were not all active, with about half not having published an article in the year of the study (Morris 2006a). An alternative directory, Open J-Gate[11] listed 5943 open access journals, of which 3416 were peer reviewed, while Ulrich's lists 2245 peer reviewed OA journals.

Using the Open J-Gate figure for peer-reviewed journals would indicate that open access journals make up about 10% of all peer-reviewed journals, measured by number of journal titles. OA journals appear less likely than others to be included in A&I databases, however, with only 7.5% of the 16,000 journals covered by Scopus being OA.

Open access journals are also smaller on average than other journals. The proportion of articles freely available on publication in 2006 was estimated to be 4.6% (Björk et al. 2009); this will over-state the numbers published in open access journals per se, because some non-OA journals make some of their articles freely available (e.g. for promotional or other reasons). Other authors have put the proportion of articles published in OA journals lower at about 2% (Campbell & Wates 2009).

A study published in 2009 (Matsubayashi *et al.* 2009) reported that 26% of biomedical research articles published in 2005 could be found in a freely-available free-text version on

---

10 http://www.doaj.org/

11 http://openj-gate.com

the web (in a search done in early 2006). The majority of these were found on journal websites (rather than in institutional or subject repositories) and of these, it is likely that a substantial proportion came from the HighWire free online text articles, i.e. subscription journals with delayed free access.

2.6 Global costs of the scholarly communication system

A 2008 RIN report by Cambridge Economic Policy Associates estimated the total system costs of conducting and communicating the research published in journals at £175bn, made up of £116bn for the costs of the research itself; £25bn for publication, distribution and access to the articles; and £34bn for reading them.

The £25bn for publication includes publishing and library costs; the publishing costs total £6.4 bn: of this, £3.7 bn is fixed first copy costs, including £1.9 bn in non-cash costs for peer review and £2.7 bn is variable and indirect costs, including publishers' surplus. Excluding the non-cash peer review costs, publishing and distribution therefore costs £4.9bn, or about 3% of the total costs.

The annual revenues from journal publication are about $8bn (see below). The most recent report on the journals market by Simba puts the annual revenues from journal publication at US$7.7 billion.

2.7 Global trends in scientific output

According to the NSF analysis (NSF 2008), the number of articles catalogued by Thomson Reuter's Science Citation Index (SCI) and Social Sciences Citation Index (SSCI) grew by 2.3% p.a. between 1995 and 2005 (from 564,645 to 709,541). The growth of articles reflects both an expansion in the number of journals covered by the SCI and SSCI databases (from 4,093 in 1988 to 4,906 in 2005) and an increase in the number of articles per journal (from 117 in 1988 to 139 in 2005) 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 7). The most dramatic growth, however, is in the output from the East Asia region (China, Singapore, South Korea and Taiwan); between 1995 and 2005, China's output grew at 17% and Taiwan's at 16% per year, compared to 0.6% for the USA and 1.8% for the EU, while the UK's output was flat and Russia's change in output was negative.

Despite this differential growth, the USA continues to dominate world output of scientific papers, with 29% of the total, followed by Japan with 8% and the United Kingdom, Germany, and China with 6% each. This means that just these five countries are responsible for 55% of the world's journal articles, while 23 countries accounted for 90%.

China's high growth rate means that it moved from 14th to 5th ranked in world output over the 10 years, to 2nd place in engineering and chemistry, and to 3rd place in physics and mathematics. Assuming the growth rates were sustained, China's output will already have overtaken all other countries except the USA, and would overtake the USA in 2017. Of course, it is not feasible that China could sustain its historical growth rates but this projection gives a vivid picture of how important China will become.

India with compound annual growth of 4.5%, however, failed to grow as fast as other Asian countries and lost rank in some fields.

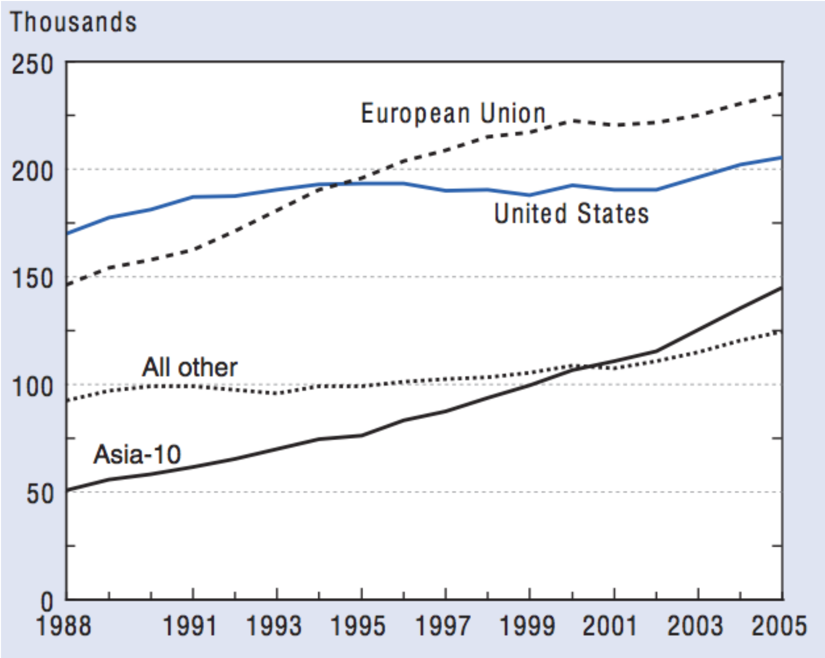

Figure 7: Scientific article output, by major publishing region 1988–2005 (Source: Science and Engineering Indicators, NSF 2008)

Research continues to become ever 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 international collaboration. 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 (NSF 2006).

Overall, the number of author names per article increased from 3.1 in 1988 to 4.5 in 2005 (NSF 2008). Another reflection of this trend is that co-authored articles grew from 40% to 61% of world output between 1988 and 2005.

Interestingly for proponents of the special relationship, UK-US collaborations are significantly less common than would be predicted from the overall proportion of international collaboration in each country’s output.

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 1954 to about 0.7 in 2000. As a consequence, although each author is on average getting their name as a collaborator on about three 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 (Mabe & Amin 2002).

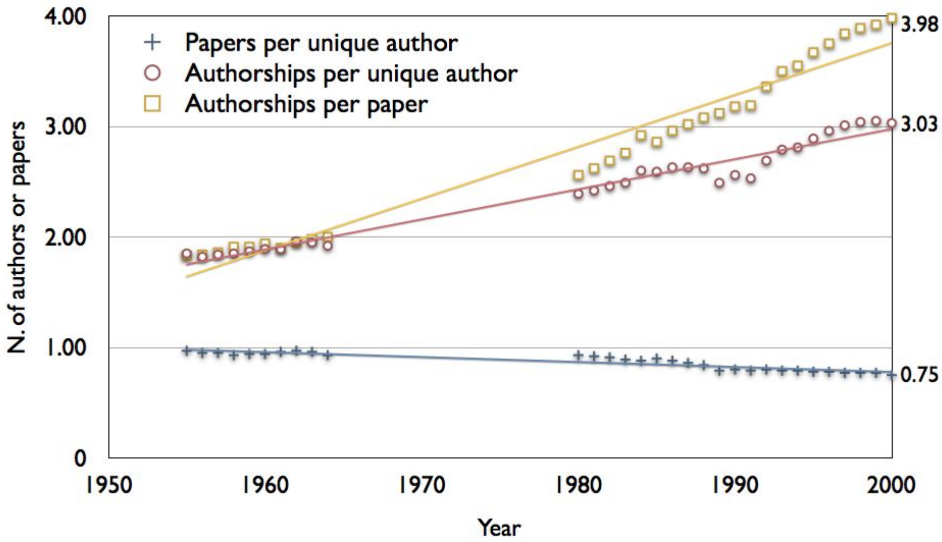

Figure 8: Co-authorship patterns 1954 to 2000 (from Mabe & Amin 2002, using data from Thomson Reuter Science Citation Index)

2.8 Authors and readers

Estimates of the global research community are compiled by UNESCO. Their latest publicly available estimate (source data relating to 2002 or earlier) estimates a base of 5.5 million researchers worldwide (UNESCO 2005). Tenopir & King quote a figure of about 10 million scientists (Tenopir & King 2000).

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 (King & Tenopir 2004).

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, while some 2.5 million authors publish at least once over a 5 year period (Mabe & Amin 2002).

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.

These overlapping author and reader communities can be illustrated as in Figure 9. The degree of overlap between authors and readers will vary considerably between disciplines: in a narrow pure science field like theoretical physics there may be close to 100% overlap, but in a practitioner field such as nursing or medicine the readers will be many times more numerous than the authors.

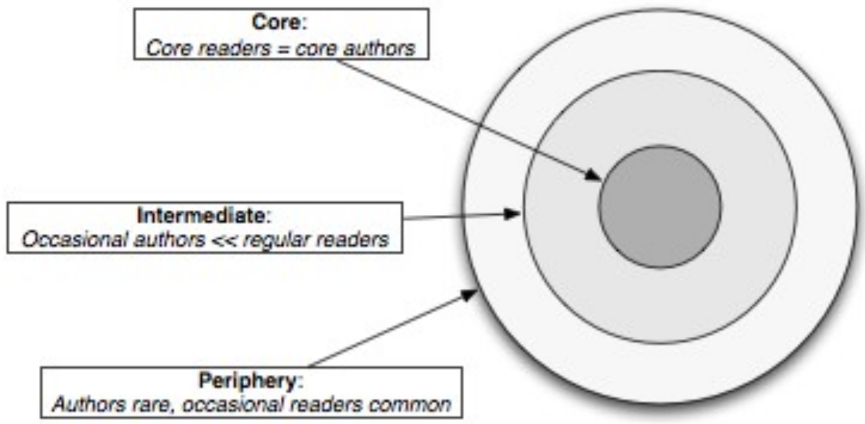

Figure 9: overlapping author and reader communities. About 1 million authors publish each year, or 2.5 million who publishing in a 5-year period (ISI data), out of a global population of about 6 million R&D workers (UNESCO) or 10 million "scientists" (Tenopir & King)

It used to be believed that the average scientific paper was very little read. This misunderstanding arose from the flawed rescaling of pioneering work done by Garvey and Griffith on reading of journals (King, Tenopir & Clarke 2006). Electronic publishing has allowed one aspect of article use to be measured precisely, namely article downloads. Although not every download will translate into a full reading, it is estimated that around 1.8 billion full text articles are downloaded every year. In the UK universities, 102 million full text articles were downloaded in 2006/07, an average of 47 for every registered library user, with an annual rate of growth of about 30% (RIN 2009). A 2005 study showed that articles in the society journal *Pediatrics* were read on average 14,500 times (King, Tenopir & Clarke 2006).

Incidentally, the average scientific paper takes its authors 90–100 hours to prepare (King & Tenopir 2004). Two to three reviewers will then spend an average of 3–6 hours each on peer review (Tenopir & King 2000; Ware & Monkman 2008).

2.9 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 (Morris 2006b). 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 Thomson-Reuters 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. At one end of the scale, 95% or more publish only one or two journals, while at the other end, the top 100 publish 67% of all journals. The top 10 publish about 35% of journals, while four publishers (Elsevier, Springer, Taylor & Francis and Wiley-Blackwell) have well over 1000 journals each. Amongst the "long tail" of organisations producing just one or two journals, many of these may not even regard themselves as "publishers" (e.g. academic or government research departments) (Morris 2007).

2.10 Peer review

Peer review is fundamental to scholarly communication and specifically to journals. It is the process of subjecting an author's manuscript to the scrutiny of others who are experts in the same field, prior to publication in a journal. (It is also used for the evaluation of research proposals.) This review process varies from journal to journal but is typically two or three reviewers reporting back to a journal editor who takes the final decision. The average acceptance rate across all STM journals is about 50%.

Academics remain strongly committed to peer review despite some shortcomings (for instance, the potential for bias); for example in the PRC survey 93% disagreed that peer review was unnecessary (Ware & Monkman 2008; see also Sense About Science 2009). Despite this overall commitment, however, there is support among authors for improvements to the system, notably in relation to the time taken and in the potential for bias on the part of reviewers.

One reason researchers support peer review is that they believe it improves the quality of published papers. In the PRC survey, researchers overwhelmingly (90%) said the main area of effectiveness of peer review was in improving the quality of the published paper, and a similar percentage said it had improved their own last published paper.

Types of peer review

There are two types of peer review in broad use, single-blind review (in which the reviewer is aware of the author's identity but not vice versa) and double-blind review (in which reviewer and author are not aware of the other's identity). Single-blind review is substantially the more common (e.g. 84% of authors in the PRC survey had experience of single-blind compared to 44% for double-blind review) but there is considerable support expressed by academics for the idea of double-blind review, presumably in response to the perceived potential for bias in single-blind review. Double-blind review is currently more common in the humanities and social sciences than in the "hard" sciences, with clinical journals falling between the two.

A fundamental flaw of double blind review is the difficulty of actually masking the identity of the author from the reviewers. Most authors usually cite their own previous work, often more so than other sources; their subject matter and style may also give away their identity to knowledgeable peers.

A newer approach to dealing with the criticisms of single-blind review is open peer review: in this model, the author's and reviewers' identities are known to each other, and the reviewers' names and (optionally) their reports are published alongside the paper. Advocates of open review see it as fairer because, they argue, somebody making an important judgement on the work of others should not do so in secret. It is also argued that reviewers will produce better work and avoid offhand, careless or rude comments when their identity is known. Open peer review is much less common than the two standard types (22% of authors said they had some experience of it in the PRC survey). Authors express limited support for it in surveys and seem reluctant to participate in practice (for instance in *Nature*'s open peer review trial (Campbell *et al.* 2006)). The most important reason is probably that reviewers are concerned about the possible consequences of being identified as the source of a negative review.

More recently, electronic publishing technology has allowed a variant of open review to be developed, in which all readers, not just the reviewers selected by the editor, are able to review and comment on the paper and even to rate it on a numerical scale following publication. This post-publication review could occur with or without conventional pre-publication peer review. The benefits are seen to be that it takes account of comments from a wider range of people ("the wisdom of crowds") and makes the review a more living process. A well-known example is the journal PLoS ONE. As with pre-publication open peer review, academics seem reluctant to participate. In addition to the same concerns as attach to pre-publication open review, academics also cite their lack of time for writing substantial comments on published papers.

Time spent on peer review

Peer review inevitably takes time. Practice varies between disciplines, with review times measured in weeks (or less) for rapid-publication journals in fast-moving life science disciplines but can be much longer (months, or more) in mathematics and in the humanities and social sciences. In the PRC survey authors reported average review times of about 3 months. On average, authors regarded review times of 30 days or less as satisfactory, but satisfaction levels dropped sharply beyond 3 months, and fewer than 10% were satisfied with review times longer than 6 months.

The commitment of the scholarly community to peer review is illustrated by the time spent. In the PRC survey, reviewers reported spending a median 5 hours (mean 9 hours) on each review, and on average reviewed about 8 papers a year. The majority of reviews were, however, completed by a more productive subset of reviewers who managed nearly twice as many reviews as the average. The global cost of peer review is substantial, albeit a largely unpaid, non-cash cost: a RIN report estimated this at £1.9 billion annually, equivalent to about £1200 per paper (RIN 2008). The Houghton report used a slightly higher figure, at £1400 per paper (Houghton 2009).

Publisher's role in peer review

The publisher's role in peer review, at its most fundamental, is to create and support the journal and its editor and editorial office. Operationally the publisher's role has been to organise and manage the process, and more recently to develop or provide online tools to support the process. Online submission systems are now increasingly the norm: while a survey of publishers for ALPSP found the overall market penetration by publisher was 65% (Cox & Cox 2008), this relatively low figure disguises the fact all large publishers use online systems. An international survey of journal editors conducted in late 2007 reported that 76% of journal editors were using online submission systems on their journals, with their use more common in life sciences (85%) and markedly less common in humanities and social sciences (51%).

The use of online submission systems has reduced the overall time required for peer review and reduced some of the associated direct costs (e.g. in paper handling and postage) but often these have been transformed into overhead costs instead (software, hardware and training). By enabling a fully-electronic workflow it has also permitted some additional benefits, including the following:

- Faster publication times: the systems can create a fully linked version of the author's final peer reviewed manuscript that can be published online immediately on acceptance
- Support for reviewers and editors: automatic linking of references in the author's manuscript can help editors identify reviewers and help reviewers assess the manuscript. Some publishers also provide editors with access to A&I databases to help with assessment and selection of reviewers. Newer artificial intelligence systems based

on text mining[12] can also integrate with online submission systems and aid in the identification of reviewers

- Plagiarism detection: the CrossCheck system allows submitted articles to be compared to published articles and to articles on the web (see *Publishing ethics*).

2.11 Reading patterns

The number of articles read per year by university faculty members has steadily increased over time, as illustrated in Figure 10. Other sources give similar estimates of around 250-270 articles per year for university academics, while non-university scientists read only about half as many (King & Tenopir 2004). There are substantial differences between disciplines (see *Disciplinary differences*).

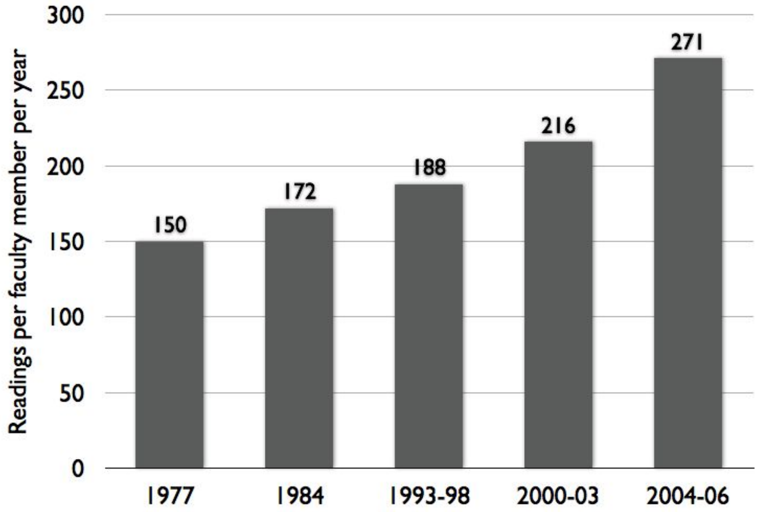

Figure 10: Average number of articles read per university faculty members per year (Source: Tenopir 2007)

The breadth of reading has also increased over time: in 1977 scientists at Drexel read from an average 13 journals per year, while the figure is now over twice that.

The average time spent reading a journal article remained at around 45–50 minutes between 1977 and the mid-1990s, but has since fallen to just over 30 mins (Renear & Palmer 2009). This was despite the average length of journal articles increasing substantially (from 7.4 to 12.4 pages between 1975 and 2001). Renear and Palmer (2009) discussed the strategies and technology innovations that help readers extract information from more papers while spending less time per paper.

Access and navigation to articles

Academics use a wide range of methods to locate articles, as illustrated in Figure 11. The growing importance of searching and parallel reduced importance of browsing is evident.

---

12 For example, as offered by suppliers such as Collexis

However, asking colleagues remains an important strategy albeit ranking behind browsing and searching.

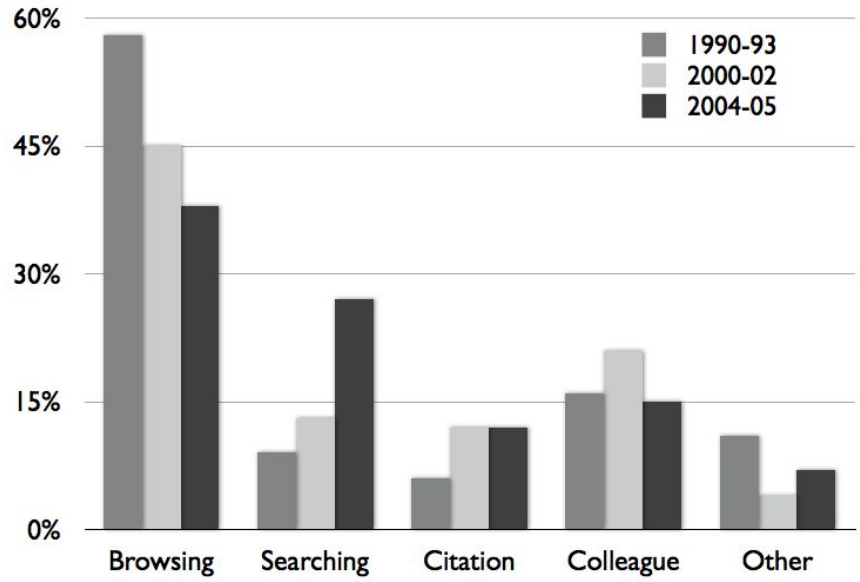

Figure 11: Ways used by university faculty to locate articles (Source: Tenopir 2007)

The source of reading of articles shifted substantially away from personal subscriptions towards library-provided access between the 1970s and the 1990s.

The ways readers access and navigate to journal content on the web have consequences for publishers and librarians. Inger & Gardner's 2008 study (updating a 2005 report by SIS) focussed on citation searching, core journal browsing, and subject searching, and presented these findings:

- Readers are more likely than ever before to arrive within a journal web site directly at the article or abstract level, rather than navigating from the journal homepage (let alone the publisher's homepage). This is of course partly driven by the growing use of search engines, particularly Google and Google Scholar, to locate scholarly content but what was notable in the survey was the multiplicity of routes used by readers. Specialist bibliographic databases were still the single most popular option for readers searching for articles on a specific topic, just ahead of web search engines.
- Readers valued the content alerting services on journal web sites but placed much less value on personalisation and search functions (presumably because they prefer to search across multiple journal/publisher sites using external search tools). RSS alerts were still a minority tool but had grown enormously in popularity between 2005 and 2008.
- The Library OPAC and the library's own web pages, having suffered initially from the growth of general purpose search engines were once more of growing importance as the starting point to navigation. Library controlled web space had the advantage of linking only to content that had been paid for by the library and met library selection criteria. The library's deployment of link resolver technology had further strengthened their importance.

- Inger reported that publishers know that personalisation features are little used by readers but remained under pressure from editorial board and society members to include this level of functionality.

2.12 Disciplinary differences

It is worth noting that the average characteristics described above conceal some important differences between subject disciplines.

While the average journal included in the Thomson Reuters citation database publishes about 115 articles per year, science and technology titles are much larger and social science and humanities much smaller. This is part of the explanation for why journal prices are substantially higher in the former compared to the latter disciplines.

The UK's JISC 2005 report on disciplinary differences (Sparks 2005) was based on a survey of UK academics but there is little reason to think that its findings would not have wider application. Its findings included:

- Article output is significantly different in the different disciplinary groups, with the "hard" sciences (physical and biomedical sciences and engineering) publishing the most with about 7.5 articles per three-year period, the social sciences next (5 articles) and the arts/humanities the least (under 3).
- The degree of joint authorship is also significantly different and follows similar patterns, with biomedical authors most likely to co-author (with 85% of respondents saying that 75% or more of their output was co-authored), followed by physical sciences and engineering, then the social sciences, with arts and humanities the least likely to co-author (with 76% saying that 25% or less was co-authored).
- As is well known, the role played by journal articles is much more important to scholarly communication in STM areas than in the arts & humanities (where books and monographs play a more significant role). The report suggested, however, that this difference might be closing, with journal articles playing a more important role in A&H. A possible reason suggested was the emphasis research assessment places on (high impact factor) journal publication.
- The peak age of needed articles varied substantially by discipline, with the peak age in humanities being about 20 years ago, in chemistry, engineering and medicine 10 years ago, and computer science, life sciences and information science 5 years ago.

The "certification" function of the journal is much less important in some disciplines than others, as shown by the willingness in some disciplines to accept a preprint (unrefereed author's original manuscript) as a substitute for the final published version of record. Certification appears less important in theoretical and large-scale experimental disciplines (high energy and theoretical physics, maths, computer science), where co-authorship is high and/or the small size of the field means the quality of each researcher's work is known personally to peers, but more important in small-to-medium experimental fields (life sciences, chemistry, geology, etc.). It should be noted that in terms of sheer numbers of researchers these latter fields provide the vast bulk of all researchers in the world.

There are considerable difference in the reading and article-seeking behaviours between disciplines. For instance the number of articles read by faculty members in medicine is nearly three times that in the humanities (see Figure 12). These numbers will reflect both the relative importance of the journal article in the fields and the nature of what constitutes a "reading" and the complications of interpreting fields like medicine with a predominating practitioner component. Figure 13 illustrates differences in the ways readers find articles, with marked variance for instance in the importance of browsing.

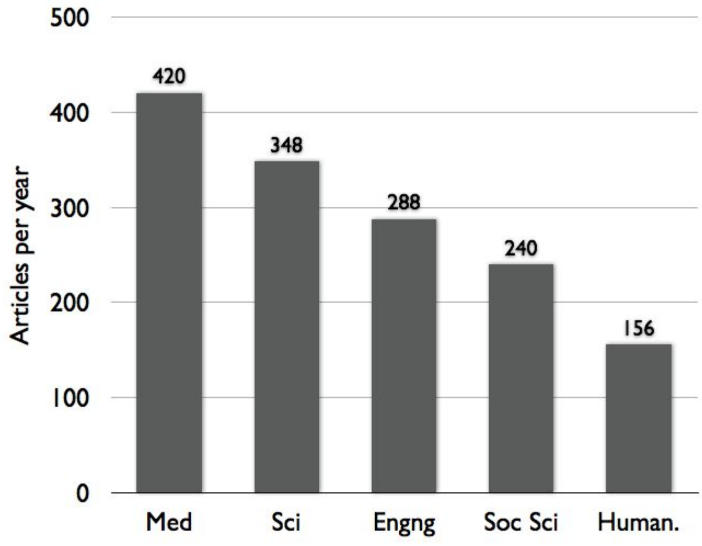

Figure 12: Average articles read per university faculty member per year (Source: Tenopir 2007)

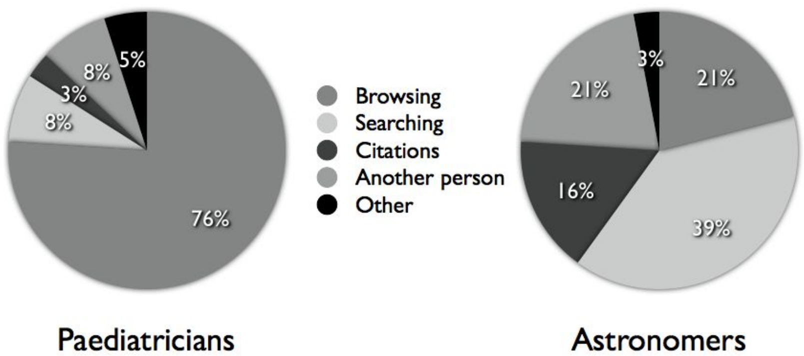

Figure 13: Subject differences in the ways articles are found (Source: Tenopir 2007)

There are marked differences between the disciplines in authors' attitudes towards peer review. Broadly speaking, the PRC survey showed authors in the physical sciences & engineering thought peer review was more effective, and were more satisfied with its current operation than authors in the humanities and social sciences. Double-blind peer review was much more common in HSS (94% of authors had experience of it) compared to the physical sciences & engineering (31%), and HSS authors expressed a much stronger preference for double-blind over single-blind review than did other authors.

There are, however, areas where there appear to be no (or only small) differences between disciplines:

- The JISC study found there was little difference between the disciplines in terms of access to resources and to journals in particular.

- All authors of whatever discipline claim that career advancement and peer-to- peer communication are the most important reasons for publishing.

2.13 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, e.g. see Inger & Gardner, 2008). 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 (as recorded in the Science and Social Science Citation Indexes[13]) worldwide increased from 2.69 million in 1992 to 4.34 million in 2003, an increase of 61%. During this period, the share of cross-national citations grew from 42% to 48%, another sign of the increasing globalisation of science.

As with article publication patterns, the regional shares of citations are changing as a result of these globalisation pressures. Table 2 shows the changes from 1995 to 2005: over this period the United State's share declined while the EU and Asia-10 countries increased.

Table 2: Share of world citations of science and engineering articles (Source: Science & Engineering Indicators 2008, NSF 2008)

Region/country	*1995*	*2000*	*2005*
United States	49.6	44.8	40.8
European Union	30.6	33.3	33.7
Other Western Europe	2.3	2.5	2.5
Asia-10	8.2	9.8	12.9
Other Asia	0.0	0.0	0.1
Other former USSR	1.0	1.0	0.8
Near East/North Africa	1.0	1.1	1.2
Central/South America	0.7	1.0	1.5
Sub-Saharan Africa	0.3	0.3	0.3

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

[13] These data (from NSF 2008) are based on journals in the Science Citation and Social Sciences Citation Indexes published by Thomson Reuters

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 (Seglen 1992). The top half of articles in a journal can thus receive 9 times as many citations as the bottom half.

Average impact factors show considerable variation between subject fields, with in general the fundamental and pure subject areas have higher average impact factors than specialised or applied ones. The variation is so significant that the top journal in one field may have an impact factor lower than the bottom journal in another area. Related to subject variation is the question of multiple authorship. The average number of authors varies by subject (see *Disciplinary differences*). Given the tendency of authors to refer to their own work, this variation is reflected in varying citation levels.

Another problem with the use of impact factors as a quality measure is that the figure is a statistical average, which will show statistical fluctuations. These are particularly important for smaller journals (because smaller samples mean larger statistical fluctuation). For a journal of average size (about 115 articles per year), a year-to-year change in the impact factor of less than +/-22% is not significant, while for a small title (less than 35 articles p.a.) the range is +/-40%. Similarly, an impact factor of 1.50 for a journal publishing 140 articles is not significantly different from another journal of the same size with an impact factor of 1.24. It is thus foolish to penalise authors for publishing in journals with impact factors below a certain value, say 2.0, given that for an average-sized journal, this could vary between 1.5 and 2.25 without being significant. For a fuller discussion of these issues, see Amin & Mabe (2000).

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.

Alternative metrics

Given the shortcomings of the impact factor, other metrics have been proposed, either as complements or as alternatives. Some of the better known are as follows:

- **the immediacy index**, which measures how soon after publication articles in a journal are cited
- **the cited half-life** is a measure of how long articles in a journal continue to be cited after publication
- **the h-index** is defined as: an author has an index *h* if *h* of their *Np* papers have at least *h* citations each, and the other (*Np* - *h*) papers have at most *h* citations each. This is intended to give a measure of quality and sustainability of scientific output of individual academics rather than for journals
- **the eigenfactor** uses network theory algorithms similar to the Pagerank method used by Google to measure the influence of journals by looking at how often they are cited by other influential journals.

In fact there are many more possible measures. The MESUR team based at Los Alamos recently published a paper comparing 39 scientific impact measures (Bollen *et al.* 2009). They used statistical techniques to categorise the different measures on two dimensions roughly equivalent to prestige and to popularity, two rather different aspects of scientific impact. The authors concluded that the impact factor measured a particular aspect that "may not be at the core of the notion of 'scientific impact'. Usage-based metrics such as Usage Closeness centrality may in fact be better consensus measures".

In practice, use of the impact factor is so widespread that it looks unlikely to be dropped even if there are technically better measures, particularly if those metrics are complex, though it would be wiser to consider a range of measures rather than relying on any single metric.

Usage Factor

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 commissioned work to investigate whether it might be feasible to develop a "Usage Factor" based on download statistics[14]. The report, issued in mid-2007, concluded that it would be feasible to develop a meaningful journal Usage Factor and that there was support in the library and publisher communities to do this. Librarians said that that a UF, if available, would be an important factor in acquisition and cancellation decisions. The report identified some technical and structural issues that would need to be resolved, including the perception that online usage data is more easily manipulated than citation data. The UKSG currently plans to issue an RFP by mid-2009 for further modelling and analysis work, so it will be some time yet before any agreed Usage Factor emerges.

It will be interesting to see how they will deal with the "normalisation" problem. That is, the need to 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 citations. The UF team is aware of these issues, however, and is considering, for instance, using different definitions for different fields. Another issue to be resolved is whether different users' downloads have the same weight: will student downloads be counted the same as faculty ones?

2.14 Costs of journal publishing

An understanding of the costs of journal publishing has become important not just for publishers but also for the wider scholarly community because of the debate over the serials crisis and open access.

A 2008 RIN report conducted by Cambridge Economic Policy Associates looked in detail at the costs involved in the journals publishing process (RIN 2008), including library access provision costs and non-cash cost incurred by scholars in conducting peer review and in searching for and then reading articles. This report provides one of the most reliable estimates of journal costs. It estimates the total cash cost (i.e. excluding the non-cash peer review costs) of producing the average journal article at £2863. This is made up as follows:

- first copy costs (the costs incurred regardless of the number of copies distributed, e.g. peer review management, copy-editing, typesetting & origination): £1136
- variable costs (printing, paper, distribution): £608
- indirect costs (staff and overheads): £608
- surplus: £517

Note that RIN include surplus in this figure, so that the cost is that seen by the purchaser rather than producer. Taking this into account the relative proportions are broadly similar to the averages for Wiley-Blackwell journals given in Campbell & Wates (2009).

[14] http://www.uksg.org/usagefactors

It is important to remember these figures are averages. First copy costs in particular show considerable variation depending on the type of journal. The earlier RIN/EPS Baseline report (EPS 2006) quoted figures from the literature ranging from $350 to $2000, but the later report quotes a narrower range. For low rejection rate journals the RIN authors give a figure of £1023, with high rejection rate journals at £1427. Their figure for popular hybrid journal (*Science*, *Nature*, etc.) of £1452 is, however, considerably lower than some figures that have been quoted in the past (up to $10,000), and the additional cost compared to a low rejection rate journal does seem low when considering the inhouse staff costs incurred by such journals in their peer review. (For example, for a 96% rejection rate journal 24 papers have to be peer reviewed by at least two peer reviewers for one to be published, leading to an article cost perhaps as much as three times that at journals with the average rejection rate of around 50%.)

RIN also estimate variations in indirect cost by publisher type at £705 per article for commercial publishers against £428 for society publishers. We are not aware of any other systematic data which would validate this.

Journal prices, as well as covering the publisher's costs, also include in most cases an element for profit (in the case of commercial publishers) or surplus (for not-for-profits). Profits are a major source for reinvestment and innovation. For their part, societies frequently use surpluses from journal publishing to support other activities such as conferences and seminars, travel and research grants, public education, etc. (Baldwin 2004). RIN estimate the average profit/surplus at 18% of revenues, equivalent to £517 per paper, as shown above, with variations between commercial publishers (£642) and society publishers (£315) that at least partly reflect their differing tax status as much as actual profitability (not for profits do not pay corporation tax so the fairest comparisons would be between post-tax profits and surpluses rather than pre tax).

Electronic-only publishing cost savings

The potential cost savings from moving to online-only publishing have typically been given by publishers at 10-20% of costs. RIN (2008) estimated the global cost savings that would arise overall if 90% of journals were to switch to e-only publishing at £1.08 bn, offset by a rise of £93m in user printing costs. The largest part of this saving comes from library savings (from not having to handle, bind, preserve print copies etc.), with reductions in publication and distribution costs equal to 7% of the total publishing costs. Eliminating the profit/surplus elements, this is equivalent to 9% of the publisher's costs, slightly under the publisher estimates.

Open access and possible cost savings

These are discussed below in the section on open access publishing.

2.15 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 1986–2004 was 7.6%, compared to the US Consumer Prices Index which rose by an annualised 3.3% over the same period (ARL 2004). Similarly Tenopir and King (2000) showed that the price inflation ratios between 1975 and 1995 for commercial and society journals were 3.1 and 2.9 (using current dollars).

Later statistics using publishing subscription prices are increasingly misleading, however, because these figures do not represent what libraries have actually paid, because of the efficiencies of electronic delivery and the growth of multi-journal licences. (ARL and LISU have both stopped recording the number of subscriptions in their annual statistics for this reason.) 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 (Creaser, Maynard & White 2005, p.133). This fall has continued, with an average price per download of £0.80 in the UK in 2006/07 (RIN 2009).

The reasons for journal price increases have been varied and include (adapted from King & Alvarado-Albertorio 2008):

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 (King & Tenopir 2004) showed that the average number of articles per year published in science journals increased from 83 to 154 articles per title15
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. The reduction in author charges – at one time page charges were common, especially in society journals but were often dropped in the face of competition from commercial journals. More recently, publishers have been dropping colour charges for competitive reasons (though colour reproduction costs have dropped, too).
6. The new journal effect: the growth of scholarship leads to an expansion of articles (as noted above) and also to the burgeoning of new fields, which in turn leads to the launching of new journals. On average new journals will tend to be in niche areas with low circulations (at least initially) and will tend to be relatively inefficient economically, and hence will tend to have higher subscription prices. 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.
7. 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.
8. Conversion of back issues to electronic format, provision of search options, and other value added attributes associated with electronic publishing.
9. 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).

---

[15] 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 14).

10. And, of course, cost inflation (especially salary and paper costs), which has annualised at about 3.3% per annum for the last twenty or more years.

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 (price spiral) and currency fluctuation effects (ca. 0.5–1%). It could be argued that in the absence of other factors, while inflation and article growth remain at 3.3% and 3% respectively, increases of less than 6.3% per annum would reflect significant cost saving or margin loss by the publisher.

The serials crisis arose not just because of these pressures on prices, but also because growth in research budgets (which translates into increased article output) has consistently outpaced growth in library budgets. Similarly, spending on libraries has fallen from 4% to 3% of average UK Higher Education Institutional spending since 1980; similar declines can also be observed in the US as well.

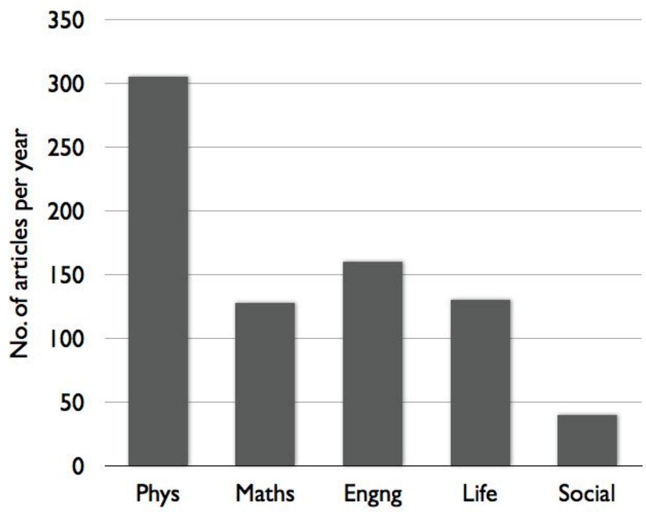

Figure 14: Articles per journal per year, by discipline (Figure from Mabe 2004, based on data from Tenopir & King 2000)

Effect of consortia licensing on prices

One increasingly used measure of journal pricing is the cost per download. Partly because scholars are becoming more used to using electronic content and partly because the "Big Deal" and similar consortia licences provide access to a lot of additional content at relatively low additional cost, the average price paid per downloaded article has fallen substantially. Recent Elsevier estimates put their average figure at $2.75 per download and as low as $0.60 at one large American university, while a RIN report estimated the average cost for UK academics at £0.80 in 2007/07 (RIN 2009).

2.16 Authors' behaviour, perceptions and attitudes

There have been numerous proliferating studies of author behaviour, perception and attitudes but two pieces of work stand out for their large international 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 (Rowlands & Nicholas

2004; Rowlands *et al.* 2005), and a survey commissioned by Elsevier in collaboration with CIBER and NOP in 2005 (Mabe 2006; Mabe & Mulligan In Press).

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 author's original or accepted manuscript 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 believed downloads to be a more credible measure of the usefulness of research than traditional citations.
- A clear majority of authors believed 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 was little enthusiasm for author- or reader-facing charges [e.g. submission/publication charges or pay-per-download 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 (Coles 1993) conducted in 1993 except that "establishing precedence" and "recognition" had increased in importance. The transition to electronic publishing between 1993 and 2005 had thus created hardly any differences in author motivations.
- Researchers were ambivalent towards funding bodies: 63% think they had too much power over what research is conducted. But despite concerns about the pressure to publish in high impact journals, funding bodies did not dictate the choice of journal. [This survey was conducted before funding body mandates about article deposit were introduced and hence was unable to explore researchers' views on this topic.]
- Authors were divided when it comes to deciding whether to publish in a prestigious or niche journal: 43% agreed while 39% disagreed that it was important to publish in a prestigious general journal rather than a more appropriate specialised journal.
- The importance of peer review was underlined. There was near universal belief that refereed journals were required. The large majority believed that peer review improved an article. Respondents were committed to peer review: 85% were willing to act as reviewers. [See also *Peer review*.]
- A majority – 60% – believed that the publisher added value – but 17% did not, with more thinking so in Computer Science (26%) and Mathematics (22%).
- Reading patterns were 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 had 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 was 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.]

A PRC paper published in 2009 looked at authors' perceptions of the rights they retained in their articles following publication and compared this to what publishers actually permit (Morris 2009). The study found that authors underestimate what they could do with pre-publication versions (e.g. self-archiving, use in course packs, provide copies to colleagues) while overestimating what publishers' policies allowed them to do with the published version. In particular, many authors believed they could self-archived the published version, which very few publishers permit. The study concludes that publishers have failed to communicate their copyright policies effectively. Part of the blame may be due to the use of the counter-intuitive term "postprint" to designate not the published version but the accepted manuscript (i.e. "post" peer review, but not post printing). Use of this term is now deprecated (see *Versions of articles*) but is still used for example by the RoMEO database and others.

2.17 Publishing ethics

There has been a growing awareness of the need for higher (or at least more transparent) ethical standards in journal publishing to deal with issues such as conflict of interest, ghost-writing, guest authorship, authorship disputes, falsification and fabrication of data, scientific fraud, unethical experimentation and plagiarism. Much of the criticism has been addressed at the intersection of the biomedical journals and pharmaceutical industry but the issues are by no means unique to this sector.

The adoption of online submission systems has made it easier for journals systematically to collect information such as declarations on competing interests, ethical consents, etc. It is increasingly the norm for journals in relevant fields to publish such declarations alongside the paper.

Committee on Publication Ethics

The Committee on Publication Ethics (COPE[16]) was established in 1997 and provides a forum for publishers and editors of scientific journals to discuss issues relating to the integrity of the work submitted to or published in their journals. It has over 5200 members, mostly editors of scientific journals. It holds quarterly meetings and provides its members with an auditing tool for their journals to measure compliance with its best practice guidelines.

CrossCheck

CrossCheck[17] is a plagiarism detection tool set up by the CrossRef organisation specifically for the scholarly journal sector. Although software is widely available that can compare a text to documents on the web, such services are not useful for checking a scientific manuscript because the scientific literature databases are not accessible to such services. CrossCheck remedies this by creating a collaborative database of STM content (contributed

---

[16] http://publicationethics.org/

[17] http://www.crossref.org/crosscheck.html

by participating publishers) allied to commercial plagiarism detection software (currently iThenticate). Users of the service can compare submitted manuscripts to the published literature. The software provides an automated report on the degree of matching between documents but the final decision on whether this represents plagiarism, repeat publication or some other more benign cause remains a matter for human judgement.

2.18 Copyright

A robust copyright (or more generally, intellectual property) regime that is perceived to be equitable by the large majority of players in the system is a precondition for commercial content and media industries, and journal publishing (open access included) is no exception. In the case of subscription-access journals, authors either transfer copyright to the publisher (while retaining certain defined rights) or grant the publisher a licence to exploit another set of defined rights; in either case the outcome is much the same, to allow the publisher to exploit commercially the rights in return for services provided to the author (peer review, copy-editing, kudos etc.). In the case of open access journals, authors may often retain copyright and release the work typically under a Creative Commons licence which allows use and re-use but imposes conditions, such as attribution of the author, which depend on copyright. However, OA under a traditional copyright regime is also possible and common.

Copyright and other IP law (such as patent law) seeks to establish a balance between granting monopoly rights to the creator (in order to encourage creativity and innovation) and the interests of wider society in having unrestricted access to content. This balance may need to be kept under review, for example to stay abreast of developments in technology. The digital transition has presented many challenges to the traditional copyright regime based on control of copies and integrity of documents – a single digital document can serve the world and it is essentially never entirely unalterable. Much of activity in this area (e.g. in the UK Gowers review) has been in the area of patents or aspects of copyright not of immediate consequence to journal publishing but the following topics currently under review are relevant:

- Digital copyright exceptions. Copyright exceptions are provided where it is judged in the public interest to allow special cases that are exempt from some normal copyright limitations. They are governed under international treaty by the Berne 3-step test: exemptions must be confined to a special case; that does not interfere with the normal exploitation of the work; and does not unreasonably prejudice the legitimate interests of the rights-holder. Exceptions being reviewed include: the archiving needs of libraries (e.g. to replace damaged originals from an archival copy or to convert to content to a new format as old formats become obsolete); support for the blind and visually impaired; inter-library lending; access within libraries to digitised content acquired in print formats; teaching course-packs; orphan works (see European Commission 2008; ***stm*** 2008; ALPSP 2008).
- Orphan works are copyright works for which the user is unable to identify and/or contact the rights holder. Such works risk exclusion from legitimate exploitation because copyright-compliant users may prefer non-use over risk of infringement. In order to avoid this, an orphan works exception allows exploitation where the user has made a "diligent search" to identify the rights holder. The issue is currently topical because of its potential application to mass digitisation (such as that conducted by Google); some have suggested the need for diligent search could be obviated in such circumstances (because the benefits to society from the results of mass digitisation outweigh the damage to rights holders), while STM publishers generally hold that the requirement for diligent search should remain (STM 2008). A Memorandum of Understanding on diligent search guidelines was signed in the EU by representatives of rightsholders and cultural organisations in 2008 (Anon 2008).

In the US, the Fair Copyright in Research Works Act (the "Conyers bill") seeks at the time of writing to amend US law to prevent federal government agencies from imposing terms or conditions regarding licenses or rights on research funding grants. The Act is specifically aimed at over-turning the NIH archiving mandate, arguing that it is in breach of existing copyright law (a point disputed by its proponents). It currently appears unlikely to make law but its existence reflects the deep unease many publishers and others have about the NIH and similar mandates (see *Open Access*).

It is worth noting that much of the debate about copyright in STM sector takes place within a context of widespread ignorance of copyright and the rights available under the current regime. For example, see the recent PRC paper discussed above (in *Authors' behaviour, perceptions and attitudes*) comparing authors' perception of the rights retained in their articles following publication and compared this to what publishers actually permit (Morris 2009).

2.19 Long term preservation

In the print world, long term preservation was the clear responsibility of the library community (rather than publishers). Preservation was ensured by the proven durability of (acid-free) paper, the multiple dispersed collections and the enduring nature of the host institutions.

With electronic journals, matters are not so straightforward. The fundamental issue is that the problems of long term digital preservation are not yet resolved: although storing the binary data seems feasible (by regularly transferring to new storage media as the old ones become obsolete), the problem is that the data may not be interpretable in the future, for example if the relevant hardware and/or operating systems are not available. A less fundamental, but still important practical issue is the fact that most electronic journal is accessed from the publisher's server; the library itself does not possess a copy to preserve but cannot rely on the publisher necessarily to be in existence at an arbitrary date in the future. This lack of a proven solution for long term preservation has been one of the factors holding back librarians from converting to electronic-only subscriptions.

The technical issues are being addressed by research programmes, for instance at the Koninklijke Bibliotheek (National Library in the Netherlands), at the Digital Curation Centre and British Library in the UK and the PADI project at the National Library of Australia.

The main current solutions are as follows:

- National library services: the earliest and best known of these is the e-Depot at the Koninklijke Bibliotheek[18]. Its digital archiving services are available to publishers worldwide and are used by many major publishers including Elsevier, Springer, Blackwell, OUP, and Sage. The KB recently announced a project with the Directory of Open Access Journals (University of Lund) to develop preservation services for open access journals.
- LOCKSS (Lots of Copies Keeps Stuff Safe)[19]. As the name suggests it works on the principle of redundancy, similar to the way that multiple print journal holdings provide security. The LOCKSS system allows libraries to collect and store local copies of subscribed content under a special licence (more than 300 publishers have given permission for their content to be preserved in the LOCKSS system). The software allows each library server continually to compare its content with others and thus identify and repair any damage. CLOCKSS (Controlled LOCKSS) is a sustainable

---

18 http://www.kb.nl/dnp/e-depot/e-depot-en.html

19 http://www.lockss.org/lockss/Home

collaborative organisation of scholarly publishers and research libraries using the LOCKSS technology.

- Portico is a not-for-profit preservation service for scholarly content[20]. It began as an initiative within JSTOR before spinning out as an independent organisation. It currently has just under 500 participating libraries and 70 publishers.
- The Alliance for Permanent Access[21] aims to develop a shared vision and framework for a sustainable organisational infrastructure for permanent access to scientific information. It organises an annual conference and is responsible for the PARSE project (Permanent Access to the Records of Science in Europe)[22]. PARSE.Insight is a two-year project co-funded by the European Union, concerned with the preservation of digital information in science, from primary data through analysis to the final publications resulting from the research. Its output is intended to guide the European Commission's strategy about research infrastructure.

2.20 TRANSFER code

The TRANSFER Code of Practice[23] is a voluntary statement of best practice for the transfer of journals between publishers. It is designed to minimise the potential disruption to librarians and end-users. It specifies roles and responsibilities for the transferring and receiving publishers and covers matters like perpetual access to previously subscribed content, transfer of the digital content and subscription lists, communication with interested parties, and transfer of the journal URL and DOIs. At the time of writing some 24 publishers had endorsed the Code, including all the large journal publishers.

---

20 http://www.portico.org/index.html

21 http://www.alliancepermanentaccess.eu

22 http://www.parse-insight.eu

23 http://www.uksg.org/transfer

3 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 (ARL 2004). For example, the ARL statistics show that the number of serials purchased per ARL library declined during the 1990s, reaching a low point of 13,682 in 2001, but has subsequently dramatically increased to 23,849 in 2006 (not all these will be peer-reviewed journals). Similarly, the number of current serials subscriptions per higher education institution in the UK has more than doubled in the 10 years to 2004/05, from 2900 to 7200 (Creaser, Maynard & White 2006). SCONUL figures show a similar growth in UK access and statistics for Australia show a similar pattern.

A recent report from the Research Information Network (RIN 2009) illustrated the dramatic impact of consortia licensing on access within higher education institutions in the UK. For example, full text article downloads more than doubled between 2003/04 and 2006/07 to around 102 million, while the cost of access fell to about £0.80 per article (falling to £0.60 at the most research-intensive institutions). The study found that there was a positive correlation between universities' expenditure on electronic journals and volume of downloads. It also found that journals use and expenditure was strongly positively correlated with research outcomes, independent of institutional size.

Usage of previously unsubscribed journals in such licences is remarkably high (Sanville 2001), and cost per use is falling to historically low levels (JISC 2005).

Illustrations 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." (Rowlands & Olivieri 2006)
- This 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).

3.1 Access gaps

This is not the whole story, of course, and unmet needs remain. For example, the Rightscom survey conducted in mid-2005 found that about half of UK researchers in all disciplines said they had experienced problems in gaining access to the resources needed to carry out their research. (The survey did not quantify either the frequency or the severity of the problems experienced.) The main problems cited were that the library did not subscribe to the journal needed (or buy the books needed, in the case of arts & humanities), followed by lack of access to databases and (in the case of physical sciences) conference proceedings.

Various surveys have asked academics to rate their access to the literature. The CIBER survey in 2004 mentioned above found 61% of academics said their access was excellent or good. A GfK/NOP survey for the PRC in 2008 found almost exactly the same figure (60%), with 29% saying "it varies" and 12% "poor" or "very poor". Similar results were found by yet another PRC-sponsored survey at the end of 2007 (Ware & Monkman 2008).

In a survey in 2006 of immunologists by Rowlands and Olivieri (2006) also for the PRC, 35% of respondents said they experienced some difficulty in getting access to all the articles they needed.

In these surveys reported access was best in the wealthy Anglophone countries (US, Canada, UK, Australia), less good in smaller European countries and the middle East, followed by Asia and worse in the rest of the world.

An area of apparently growing interest in public policy terms is access to the scientific literature by small and medium-sized enterprises (SMEs). SMEs are known to be a major source of innovation and job creation and hence of particular importance in the global downturn. SMEs have not been part of the core market for journal publishers as they do not generally purchase subscriptions, but have typically accessed the literature through library, database and document supply services. A survey for the Publishing Research Consortium (Ware 2009) found that people in UK high-tech SMEs valued information more highly, and read more journal articles, than those in larger companies. Of those that considered information important, 71% felt they had good access, and 60% that it was better than 5 years ago. However, the report found more than half sometimes had difficulty accessing an article, and outline a number of possible steps that could be taken to improve access: pay-per view access could be made simpler, with a more appropriate payment mechanism for companies, and lower prices; higher education journal licences could include online as well as walk-in access for local businesses; and a comprehensive, centrally administered national licence could be explored.

3.2 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.

The Research4Life programmes[24] are collaborations between UN agencies, STM publishers, universities and university libraries, philanthropic foundations and technology partners. There are currently three programmes:

[24] http://www.research4life.org

- HINARI, launched in January 2002 in conjunction with the World Health Organisation, offers free or very low cost online access to major journals, full-text databases and other resources in biomedical and related social sciences to local, not-for-profit institutions in developing countries. On launch it offered access to some 1500 journals from 6 major publishers; this has now expanded to a list of over 6200 journals from 150 publishers.
- AGORA, set up in October 2003 by the Food and Agriculture Organization of the UN and major publishers, enables access to a digital library collection of about 1300 journals in the fields of food, agriculture, environmental science and related social sciences.
- OARE (Online Access to Research in the Environment), launched in late 2006 in partnership with United Nations Environment Programme, offers access to the environmental literature with over 1300 journals from over 340 publishers. Subjects include environmental chemistry, economics, law and policy, and other environmental subjects such as botany, conservation biology, ecology and zoology.

The programmes offer free access to the poorest countries (by GNP per capita) and very low cost access (typically about $1000 per institution for the complete package).

Other schemes include:

- 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 PERii scheme negotiates affordable, sustainable country-wide licences that provide access free at the point of use for researchers and supports global research communication.
- 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.

The problems of accessing and using literature in developing countries are not limited to affordability. The Research4Life, INASP and eIFL all recognise the broader issues and variously provide training, outreach and support, advocacy, bandwidth improvement. Support is also provided for authors, for instance through INASP's AuthorAid programme[25].

[25] http://www.authoraid.info/

4 Open access

Open access is the idea of making original research articles freely accessible on the web,[26]. It is therefore strictly speaking a property of an article, rather than a journal. The different approaches to open access can be considered in terms of *what* is made open, *when* it is made open, and *how* it is made open.

Three "*what*" stages may be distinguished:

- Stage 1 — author's un-refereed draft manuscript for consideration by a journal, often called (especially in physics) a preprint ("author's original" or "submitted manuscript under review" using the NISO Versions preferred terms (see *Versions of articles*))
- Stage 2 — author's final refereed manuscript accepted for publication by a journal and containing all changes required as a result of peer review ("Accepted manuscript")
- Stage 3 — final published citable article available from the journal's website ("Version of record")

In terms of timing (the "*when*") there are three options: prior to publication, immediately on publication, and at some period after publication (an "embargo" period). The question of "*how*" is largely one of the business model (if any).

Using this framework allows us to distinguish the main types of open access in current use:

- **Full open access** (the "Gold" route): whereby the journal makes the Stage 3 version available immediately on publication, using a "flipped" business model
- **Delayed open access**: Stage 3, but delayed; existing business model
- **Self-archiving** (the "Green" route): Stage 2 (or sometimes Stage 1), either immediate or delayed; no business model.

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

4.1 Full open access

The final published paper is made available online immediately on publication using a business model in which publication is paid for rather than access. There are two main variants:

- Immediate full OA: the entire contents of the journal are made freely available immediately on publication. A well-known (though not typical) example is *PLoS Biology*.
- Optional (or hybrid) OA: here only part of the journal content is made immediately available. 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

---

[26] 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 text/data-mining, teaching or writing, subject only to minimal constraints such as attribution and non-commercial re-use. For example, the Berlin and Bethesda statements require the author to consent in advance to let users "copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship...". In this paper we shall ignore this distinction and focus on the key debate around access.

or OUP's Oxford Open schemes). This can be seen as a route to convert subscription journals to full open access; many publishers offering this have said that the subscription price will be reduced in line with uptake of the OA option.

4.2 Full OA 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.

This approach has advantages, not least that it scales with increases in research output. It provides universal access to scholarly content and offers a business model for publishers. There are clearly obstacles to wider adoption, though, which are discussed below (see *How sustainable is open access publishing?*)

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.

Not all open access journals use publication charges: research by the Kaufman-Wills Group for ALPSP published in 2005, showed that the (small) majority of OA journals did not make author charges (in fact, author charges were 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. An example of a high-profile initiative to establish open access across the field of high energy physics without author charges is discussed below (see *SCOAP3*).

Table 3 lists a selection of 2009 publication charges. Some OA publication charges appear to have been originally set at levels that were designed for impact rather than being related to actual costs. For instance, the OA publishing pioneers PLoS and BioMed Central have both had to raise their fees substantially: PLoS raised its fee of $1500 to $2200–2850 (depending on journal), while BMC has raised its fee from its original $500 to between $1050 and $1995 for the majority of its journals. Fees for full and optional open access journals now mostly fall in the rather wide range $1000–3000. The lowest charges are levied by Hindawi, an open access publisher of some 150+ journals based in Cairo. The situation is also 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. Most fees, however, still appear lower than the industry average cost per article (e.g. RIN 2008).

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.

Table 3: Publication charges for a selection of full and optional OA journals (Source: publisher websites, April 2009)

Journal/publisher	***Full/Optional OA***	***Charge (US$)***
American Institute of Physics	Full/Optional	1500–1800
American Physical Society	Full/Optional	975–1300
BioMed Central - case notes, research notes	Full	1050-1995 280-820
Hindawi	Full	275–975, plus 2 at 1400
BMJ Publishing Group (exc. *BMJ*)	Optional	2220–3145
Cambridge University Press (some)	Optional	2700
Elsevier - *Cell*	Optional	Mostly ~3000 5000
New Journal of Physics/IOP-DPG	Full	1200
Oxford University Press - subscribing institutions - non-subscribing institutions	Optional/Full	 1800 3000
PLoS - *PLoS ONE*	Full	2200–2850 1300
Royal Society (London)	Optional	360–540 per page
Springer	Optional	3000
Wiley-Blackwell	Optional	3000

SCOAP3

SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics)[27] is an ambitious plan originating from CERN to convert all journal publishing in high energy physics (HEP) to a sustainable form of open access. Within HEP, some 5000–7000 articles a year are published, 80% of them in a small core of 6 journals from 4 publishers. Virtually all these articles appear author's original and/or final manuscripts on arXiv prior to publication, and so the journals are losing (or have already lost) their dissemination function. The key remaining functions are seen to be high quality peer review and acting as "the keeper of records". SCOAP3 has estimated the global cost of journal publishing in HEP at around $13 million (based on 5000–7000 articles at $2000 per article). The idea is to form a coalition of national HEP funding bodies, libraries and consortia that will agree to contribute up to this level (by redirecting subscriptions), with national contributions based on the fraction of HEP articles per country. Once pledges have been received from sufficient countries, SCOAP3 will issue tenders for publishers to publish the same journals but under the new open access model with centralised funding via SCOAP3 eliminating the need for author charges. At the time of writing, despite nearly two years of discussion, only about

[27] http://scoap3.org/

60% of the funding pledges had been received and SCOAP3 had yet to present a timetable for its next steps (creating an organisation for international governance and issuing tenders). For more information see Salvatore Mele's presentation to the SCOAP3 Forum (Mele 2009).

SCOAP3 suggest that their project could act as a pilot with lessons for other fields. HEP is relatively unusual, however, with a high proportion of articles concentrated in a few journals and a very high proportion already open access via self-archiving. Astrophysics (a related field) shares these characteristics, as do some other parts of theoretical physics, but it is difficult to see how the model could be applied to fields with much more diverse publications ecology such as the biomedical sciences.

Hybrid journals

In another hybrid business model, 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 open access publisher BioMed Central also uses this model.

4.3 Delayed open access

Under this model, the journal makes its contents freely available after a relatively short period, typically 6–12, or in some cases 24 months (e.g. many of the life science 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 business model depends on the embargo period being long enough not to compromise subscription sales; this is discussed in more detail below (see *How sustainable is open access?*)

Publishers have typically selected journals for this model in areas where they expect access not to damage sales, for instance very newsworthy journals or those in rapidly developing fields.

4.4 Other open access variants

Willinsky (2003) identified nine different sub-species of open access. 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 less common variant of hybrid open access is whereby the articles submitted by members of a learned society will be published in the society's journal with full immediate open access[28].

A final "variant" might be mentioned, which is false open access. A number of surveys, most recently one by the UK Biosciences Federation (2008), have demonstrated that academics confuse open access with free-at-the-point-of-use online access provided by their institutions. Responses to surveys on authors indicating high levels of use of, or authorship in, open access journals may suffer from this confusion.

---

[28] an example is American Society of Plant Biology's journal *Plant Physiology*, see http://www.plantphysiol.org/cgi/content/full/142/1/5

4.5 Open access via self-archiving

The "green" route to open access is by self-archiving, which makes available a stage two version of the article (either an author's original or accepted manuscript), either immediately or delayed. Self-archiving has no independent business model, in that it relies on the assumption that making Stage 2 versions freely available will not compromise the sales of Stage 3 versions (i.e. journal subscriptions). This assumption is discussed below (see *Sustainability of open access*, page 52).

The author (or someone acting on their behalf) deposits the article in an open repository. This repository might be an *institutional* repository run by the author's institution (typically a university) or a central subject-based repository (such as arXiv in physics and PubMed Central in biomedicine).

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. original author's and accepted manuscripts), 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 relatively easily to 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, although the complexity of services that they offer varies significantly. The Eprints project maintains an information database of repositories[29]; it currently lists a total of 1300 archives of which 713 are identified as institution or department level research repositories. The alternative OpenDOAR service lists 1367 repositories of which 1099 are categorised as institutional.

The numbers of articles deposited by authors in their IRs has grown much more slowly, and most of these IRs (except perhaps in the Netherlands) remain forlornly underused (e.g. see Salo 2007; Albanese 2009). (The total number of articles included in the 713 repositories listed by Eprints is about 3.2 million, or a mean of 4500, but the distribution is 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 still 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%

29 http://roar.eprints.org/

of eligible articles; the adoption of institutional mandates is intended to achieve higher deposit rates.

The future of IRs is unclear, with a continuing debate between those who see them primarily as part of the digital infrastructure of the university, perhaps playing an important role in managing research data, and those (such as the University of California's eScholarship repository) who see the role primarily in terms of scholarly communication and publishing (Albanese 2009).

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 and now hosted by the Cornell library. arXiv[30] (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 as at least the author's original manuscript. 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 over 530,000 e-prints.

RePEc (Research Papers in Economics)[31] 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 285,000 working papers, 435,000 journal articles and other content. It differs from arXiv in that it does not contain full-text articles, that is the journal article records are for abstracts and bibliographic information only, although many have links to full text versions including to the publisher's site for the full version. It is also different in that publishers collaborate with RePEc to deposit bibliographic records of their journal articles. In many ways RePEc is thus more like a free bibliographic database than a repository.

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 is the designated repository for researchers funded by the NIH and other biomedical research funders. 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. Since 2004, PMC has taken accepted manuscripts from authors for archiving in support of the NIH funding policy discussed above. At the time of writing (April 2009) there were 1.3 million research articles hosted on PMC, of which 136,000 were open access (the others are freely available but not open access).

A UK version of PMC is hosted and managed by the British Library and other international versions are expected at some stage.

---

30 http://www.arxiv.org

31 http://repec.org

Self-archiving policies and mandates

In 2004, The US National Institutes of Health introduced a policy encouraging researchers that it funded to deposit a copy of their accepted manuscripts in the repository PubMed Central. Compliance with this voluntary policy was low (<5%) and NIH consequently changed its policy to require researchers to deposit, with effect from April 2008. The NIH mandate allows authors to defer deposit for up to 12 months after publication.

Although not the first, the NIH policy received much attention because of the size of its research budget (ca. $30 bn). Similar policies are now becoming widespread; the SHERPA Juliet website[32] listed (as of May 2009) nearly 50 research funders with deposit policies, including all the UK Research Councils, the Wellcome Trust, the Howard Hughes Medical Institute, the European Research Council, the DFG and the Fraunhofer in Germany, and Australian Research Council. Embargo periods vary from 6 to 12 months, or in some cases "at the earliest opportunity" while respecting publishers policies.

In addition to research funders, some host institutions have also adopted similar policies. The Eprints/ROARMAP website[33] recorded 37 such institutional mandates in March 2009. High profile institutions adopting mandates include Harvard, MIT and the University of California.

It remains to be seen what impact mandates will have. A survey of UK academics for RCUK (LISU & SQW Consulting 2008) found that mandates were little regarded, with 66% of academics regarding funder mandates as "not at all important" in influencing their publishing policy.

The six months embargo period appears to owe its popularity to it being chosen by Varmus and the original Public Library of Science boycott letter of 1999 where journals that did not make their contents available for free after six months would not be submitted to, refereed or edited by the signatories. Six months was chosen as an embargo because the high frequency weekly journals favoured by the life science community were not thought to be affected by it, although no evidence was gathered.

Publishers' policies on self-archiving

Most publishers have fairly liberal policies on allowing authors to archive versions of their articles on the web, although generally these policies were originally introduced on the understanding that the archiving would not be systematic. A database of publisher policies is maintained by the SHERPA/RoMEO project[34]; of the 563 publishers included:

- 30% allow archiving of both author's original and accepted manuscript
- 21% allow archiving of accepted manuscript
- 11% allow archiving of the author's original manuscript
- 38% do not formally support archiving.

Some 62% of publishers therefore permit archiving in some form. The proportion of journals will likely be higher still, since the largest publishers generally do allow some form of 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 archived manuscript to

32 http://www.sherpa.ac.uk/juliet/

33 http://www.eprints.org/openaccess/policysignup/

34 http://www.sherpa.ac.uk/romeo/

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.

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) (MIT 2003). A survey for ARL (Bailey 2006) 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.

Houghton (2009) used an estimate of £100,000 for the annual costs of higher education institutional repositories (including an element for senior management's time in policy and advocacy activities). On top this the cost of the time taken by academics in depositing their articles was estimated at about £10 per deposit, or about £1.6 million for the UK as a whole (or £15 million globally).

Large disciplinary repositories can be much more expensive. The National Institutes of Health has estimated that the cost of administering its self-archiving policy for 100% compliance would be $3.5 million.

Multiple versions of articles

One potential issue with the widespread adoption of self-archiving is that multiple versions of articles will be available to readers (and others, such as repository managers).

Authors will self-archive either the author's original or the accepted manuscript, or in some cases both (few publishers permit archiving of the version of record). Most funder and institutional mandates require deposit of the accepted manuscript. It is possible that an author may self-archive different versions in more than one repository (e.g. an institutional and a central repository).

4.6 Sustainability of open access

Full open access ("Gold")

There is no reason why in principle gold open access should not be a sustainable business model in some disciplines, although it is unlikely to be successful for all disciplines (for instance where research grants are rare – only 60% of authors overall are grant supported). The pioneer OA publishers BMC (now owned by Springer) and Hindawi have, for example, been reported anecdotally to be profitable (although the levels of profitability were not disclosed). In order for a business model to be sustainable, it must generate sufficient surplus to provide for ongoing investment and to reward providers of capital. For gold OA, this can occur if prices that authors are willing or able to pay are on average (taking into account the fraction of authors whose fees are waived – e.g. 25% of authors are based at poor institutions in developing countries) higher than the costs of producing the article. The market (such as it is) to date has thrown up a wide range of gold OA charges, from well under $1000 (Hindawi) to $2850 (PLoS), with a median probably somewhere between $1000 and $2000. This is clearly substantially less than the current industry average cost of producing a journal article (print + electronic), which as we saw above is about $4000 (RIN 2008).

For gold OA to be viable, therefore publication charges would have to be higher or alternative cost savings made (or both). Publication charges have certainly risen from the early levels, as noted above, but at present it is not clear that there is widespread appetite for substantial higher prices. Turning to possible cost savings, the first area for saving is to eliminate the print edition (an option that is of course equally open to subscription journals). In itself, this would not be sufficient to close the gap, offering savings of perhaps £200-300 per article. Open access proponents argue that additional savings might come from reduced sales & marketing costs from not having to promote subscriptions, sell and negotiate consortia license deals, etc. (publishers would have to market to authors but they do this anyway); from reduced back-office costs (e.g. no need for access control systems, although this is likely to be offset by increased infrastructure and processing costs of handling larger numbers of smaller payments); elimination of the subscription agent function (though something similar might have to be invented to aggregate author payment charges); elimination of rights-management functions, etc. The size of these potential savings is disputed.

Other challenges include:

- Corporate subscriptions represent approximately 15–17% of journal income, but corporations contribute only a small fraction (around 5%) of papers. This income would therefore be lost as it would not be converted to author payments.
- Rights income would be lost.
- Loss of print-based advertising: although advertising overall represents only 4% of journal revenues, for some journals it is an important part of their income. At present advertising revenues are still linked to print publication – online advertising rates are so much lower than print (by 1–2 orders of magnitude) that any increased online traffic would not compensate. This issue is, however, more to do with the transition to electronic publishing than OA journals per se.

In addition pricing has to allow for waivers to authors unable to pay, for instance in developing countries, where about 25% of authors are based. OUP's experience with about 70 journals is that waiver rates have been stable at 6–7%.

It seems likely that Gold OA is not a good model for the very prestigious, top-tier journals like *Nature* or *Science* that depend on expensive editorial quality control. This appears to be the case at PLoS, for instance: Butler (2008) reported in *Nature* that PLoS's high-end journals were struggling to achieve profitability and had been sustained by charitable donations, whereas the less prestigious PLoS ONE journal, which employs a system of light peer review, and the PLoS "community journals" were making a positive financial contribution.

Optional (hybrid) open access

As noted above, optional open access can be seen as a route to full open access. As such its viability would be subject to the same constraints as full open access discussed above. While uptake remains low, of course, the journals are essentially operating on a conventional subscriptions-based model.

Optional OA schemes in subscription journals mostly began around 2004-05. The overall proportion of authors that have chosen to take up the OA option is currently very low, perhaps 1–2% on average across all publishers offering it, though significantly higher in some specialised areas (e.g. bioinformatics, computational biology) where take-up can reach 20-30%. Take-up is likely to be higher in areas where awareness of the open access is high and where research is funded by institutions willing to allow OA publication charges to be recouped from grants. The total number of papers published via this route is still very small overall, about 8000 papers in 2008, roughly 0.5% of all papers.

For example, OUP published in 2008 the results of their open access initiative (Bird 2008). The average uptake was 7%, rising to 11% in the life sciences (see Table 4). Uptake was greatest in the journals *Human Molecular Genetics* (at 17%) and *Bioinformatics* (24%).

Table 4: Uptake of Oxford Open optional open access scheme for 2007 (reproduced with permission from Bird 2008)

Subject area	*No. of journals*	*Articles published*	*Open access articles*	*% uptake*
Medicine	30	5,799	289	5
Life sciences	19	3,609	388	11
Social sciences and humanities	13	598	14	2
Mathematics	3	614	29	5
Total	65	10,620	720	7

It is possible the growth of gold and optional open access may be slowed by the publicity given to the green archiving by the introduction of funder and institutional mandates (see above); authors may feel they can achieve the same result (open access) and comply with the mandates without incurring charges.

4.7 Delayed open access

Delayed access journals provide free access (though not usually open access) to their content after an embargo period set by the journals. The best known are the DC Principles Group and other publishers using the HighWire system, who collectively make available some 2 million articles in this way.

The viability of the delayed open access business model rests on the willingness of libraries to continue to subscribe to journals even though the bulk of their content is freely available. There are two (related) key factors to be taken into account, the length of the embargo period and the subject area. The arguments on these points are essentially the same as applied to self-archiving, and are dealt with in the following section.

4.8 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.

Support for this hypothesis was given by a in a 2006 report by SIS for the Publishing Research Consortium (Beckett & Inger 2006). This study surveyed the purchasing preferences of librarians and concluded that librarians were likely to substitute OA for subscribed materials, provided the materials were peer reviewed (as is the case with all funder/institutional mandates) and provided the materials were not embargoed for too long. The last point was critical: librarians were far less likely to favour OA versions over subscriptions where the OA version was embargoed for 12 or 24 months, but an embargo of 6 months or less had little impact on their preference. This was, however, a survey of librarians. A more recent survey by the Biosciences Federation indicated that researchers might have a stronger preference for the version of record.

A previous survey of librarians undertaken by Ware (2006) also explored this issue. 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[35] that journals covered by arXiv experience significantly fewer downloads from their sites than other comparable journals. In the case of IOP, the figures quoted were dramatic, with core high energy physics titles experiencing 100 times fewer downloads from the IOP site.

A reduction in usage or downloads on the publisher's site matters because low or declining journal usage is one of the key factors used by librarians to select journals for cancellation (Ware 2006). There is also some trend in the market for usage to be a factor in pricing journals. (See also *Usage Factor*.)

A key issue in this debate 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, while OA proponents argue there is little evidence for this. Publishers also argue that there should not be a single embargo period for every discipline, as the patterns of journal use are quite different across field. For example, Figure 15 (using data from ScienceDirect) shows how lifetime article downloads accumulate over time from the date of publication. While 60% of lifetime use of an article in the rapid-usage life sciences occurs in the first 12 months after publication, this drops to 44% for chemistry and to 36% for social sciences. The social sciences and mathematics show less than half their lifetime downloads after 2 years, and indeed all subjects (except rapid-usage life sciences) show 60% or less of lifetime usage at this stage.

At the time of writing, hard evidence for or against a causal link between self-archiving and subscriptions remained thin. A major EU-funded study, Publishing and the Ecology of European Research (PEER), will build a body of evidence, however, by systematically observing journal use aimed at answering this question, as well as related questions such as whether archiving increases access, how it affects the broader ecology of European research, what factors affect researchers' willingness to deposit, and it will explore models that illustrate how traditional publishing systems could coexist with self-archiving.

---

35 E.g. see http://www.alpsp.org/events/2005/PPR/default.htm

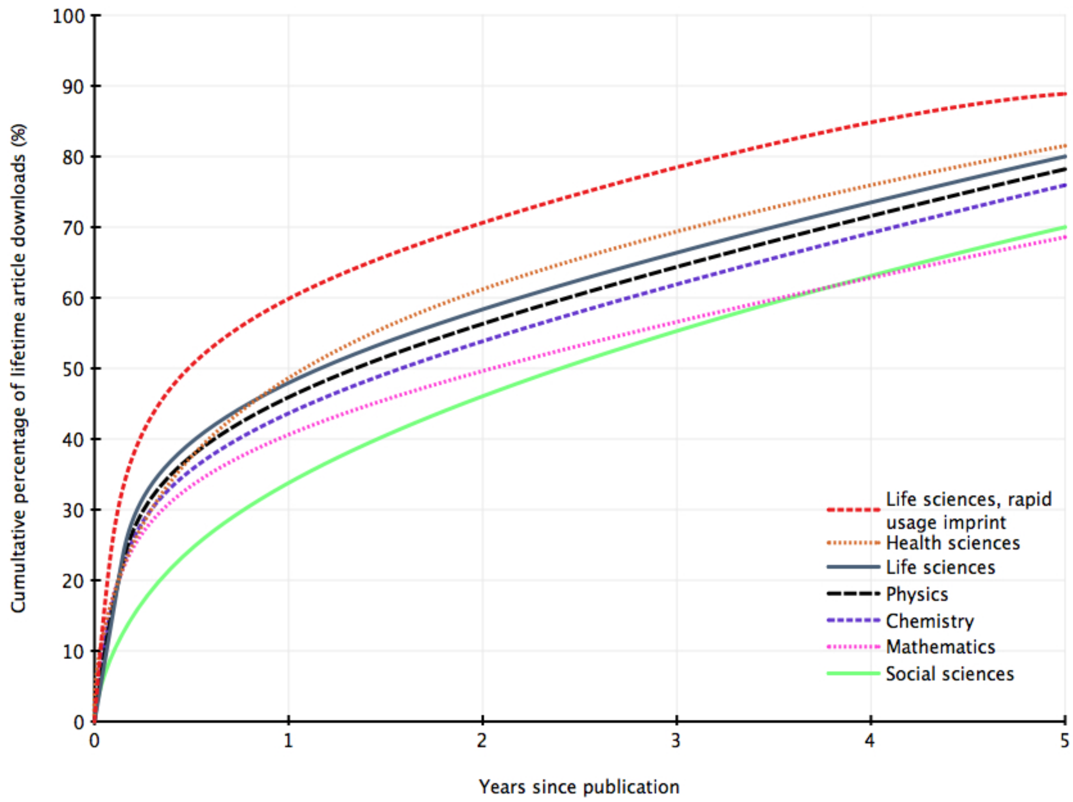

Figure 15: Article usage over time (Source: ScienceDirect)

4.9 System-wide perspective

As noted previously, the RIN/CEPA 2008 report estimated the total costs of journal publishing and distribution at £4.9bn (excluding non-cash peer review costs), out of a total £25bn for publishing and library costs. The authors then modelled the impact of converting to a system in which 90% of articles were published under an author-side fee. They estimated that there would be cost savings across the system of about £560m, split almost equally between publishers and librarians. (These savings were on top of global savings of about £1bn from switching to electronic-only publishing.) Libraries would save some £2.9bn in subscriptions, but this would be offset by author side charges of virtually the same amount. The costs and benefits would fall unequally across institutions: research-intensive institutions would tend to pay more in publication fees than they currently do for library subscriptions, while the reverse would be true in other institutions. The savings also exclude any additional administrative costs required to manage author-side payments at publishers, funders and institutions. For a large publisher, the number of articles published per year may be as much as 30-40 times as great as the number of subscription/licence accounts, so this last factor could add substantial costs. (It is possible a system analogous to subscription agents might evolve to aggregate publications charges; this might be a possible role for reproductive rights organisations under an open access regime.)

A JISC report (Houghton *et al.* 2009) published the following year by the economist John Houghton estimated system-wide savings accruing to open access publishing in the UK alone at £212m, less the author-side fees of £172m, giving a net saving of £41m. (This appears roughly comparable in scale to the £560m global savings estimated in the RIN report.) The largest single part of the savings (£106m) came from research performance savings, including reduced time spent by researchers on search and discovery, seeking and obtaining permissions, faster peer review through greater access, and less time spent writing due to greater ease of access e.g. for reference checking. Funders should, according to Houghton, therefore be comfortable with diverting research funds to pay for open access charges because the savings in research performance etc. would outweigh the cost.

In response, publisher organisations (PA, ALPSP & ***stm*** 2009) have argued that the analysis was deeply flawed. It underestimated the efficiencies of the current subscription system and the levels of access enjoyed by UK researchers. Many of the savings hypothesized would depend on the rest of the world adopting author-pays or self-archiving models. The calculated savings would remain hypothetical unless translated into job losses; for example some 200 library job losses would be required to realize the estimated £11m savings in library costs. Critics also argue that Houghton *et al.* underestimated the costs of switching to an author-pays model because they underestimated the true costs of publishing an article only, and because additional costs such as the infrastructure required to manage the many small publication charges were not included.

In addition to the system savings, Houghton suggested increased economic returns to UK public-sector R&D arising from increased access might be worth around £170m. This appears speculative, resting on flawed and untested assumptions about the levels of current access and the marginal rate of return to any increased access.

4.10 Open access citation advantage

A number of studies have addressed the question of what the effect of open access might be on the citations an article receives. The common-sense hypothesis is that an openly available 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.

Most studies have shown that it does appear to be the case that self-archived articles receive more citations than non-archived articles, with figures for the advantage ranging from 200% to 700%, but it is important to separate three separate effects: the *early view* effect posits that archived articles may have received more citations at a given point because they had been available for longer; *selection bias* occurs if authors are more likely to archive their better work, or if better authors are more likely to self-archive; the *open access* effect is the component due purely to the fact that the article was open access.

Craig *et al.* (2007) in a review of the literature concluded that the most rigorous study then available (i.e. Moed 2007, covering condensed matter physics) demonstrated a clear early view effect with the remaining difference in citation due to selection bias but no evidence to support an open access effect. Citation patterns differ between subject disciplines, however, so this still leaves it open that there may be an effect in other fields.

In 2008 Davis et al. published the results of the first (and so far only) randomised control trial investigating the issue (Davis 2008). The study randomly assigned articles on publication in 11 American Physiological Society journals as open access or subscription access (i.e. within the same journals). The study found that open access articles were no more likely to be cited than subscription articles in the first year after publication. This ignores citations occurring after the first 12 months but the authors believed their time frame was adequate to detect an open access effect if it existed. The study has nonetheless been criticised for this limitation but the authors plan further analysis to test whether the conclusions hold over a longer period.

The best available evidence at this point therefore suggests that open access articles do not receive more lifetime citations but they do get them sooner due to early view and selection bias effects.

5 New developments in scholarly communication

Technology is driving (or creating the opportunity for) profound changes in the ways research is conducted and communication, both of which are likely to have impacts on journal publishing.

5.1 Web 2.0

So-called Web 2.0 tools offer the potential to enhance informal and formal scholarly communication, although their impact to date has been limited.

Blogs and social networking allow informal discussion and information sharing. There are estimated to be some 1000-1500 scientific blogs, although at present the numbers of researchers who regular read blogs (~15%) and the average time spent reading blogs are both low. Social networks (e.g. Nature Network) have also been somewhat slow to become established (at least in comparison to the uptake of social networking in the general population) – recent (2007–08) surveys put the proportion of scientists using social networking sites for professional purposes at about 10-15%. One inhibiting factor may be the perception of such sites as being for personal rather than professional use. If so, this may change with the developments of networking sites such as BioMedExperts, pubScholar and SciLink which allow users to explore and expand the social network created by the web of literature citations.

Social bookmarking (or shared reference management) systems allow users to store internet bookmarks and categorise them (with "tags") so that as well as being available for the user's own future use, they can be shared, for example with colleagues or with anyone interested in the field represented by the categories used. There are at least three competing services aimed at academics, CiteULike, Connotea (Nature Publishing Group) and (more recently launched) 2collab (Elsevier). Most of the leading electronic journals platforms offer clickable icons for at least one of these services, most allowing easy use and promoting the services. These services potentially offer a number of benefits to academics. One obvious use is to allow a research group to share literature discoveries with each other and to maintain a single shared bibliography. Perhaps more interestingly, it would be possible to use the combined metadata of the user population to identify articles related to a particular article in ways that were not necessarily obvious from the content or keywords. As with social networking, however, take-up of these services has been relatively slow to date.

Web 2.0 ideas could be used supplement peer review, by allowing readers to add comments and ratings to the article after publication (see *Peer review*).

Wikipedia is not just the best known general-purpose user-generated encyclopaedia but despite initial and continuing scepticism in some quarters about the quality of its content, it is increasingly used by researchers and academics, albeit not for critical information. There are a number of coordinated projects ("WikiProjects") aimed at improving the number and quality of articles within specific disciplines.

Although of some interest, Wikipedia itself is unlikely to have much impact on core areas of scholarly communication. More relevant are specific projects that utilise the core functionality of the wiki platform for research or other scholarly purposes. Perhaps the most exciting are wiki-based projects that allow the research community to create and maintain shared databases and resources. One example is WikiPathways, which uses standard wiki software to create a site "dedicated to the curation of biological pathways by and for the scientific community".

Academic publishers have been slow to adopt wikis, most likely because the wiki model relies on open, editable and reusable content which is not easy to monetise. Three examples may be worth mentioning, although none is conclusive:

- WiserWiki (Elsevier) was launched in early 2008 with content seeded from an existing (out of print) textbook. Only qualified doctors are allowed to edit or create pages.
- Elsevier's SciTopics is a wiki-like service that allows invited experts to maintain pages on topics of their choice, with summaries of the topics, further reading and weblinks supplied by the expert, supported by automatically generated links to related articles in Scopus and search results from Scirus.

The journal *RNA Biology* recently changed its policy to require authors of articles on RNA families also to submit a draft article on the RNA family for publication in Wikipedia. The hope is that the Wikipedia page will become the hub to collect later information about the RNA family.

As noted at the start of this section, the research community has to date been surprisingly slow to adopt Web 2.0 solutions to scholarly communication needs (particularly in contrast to the enormous impact that Web 2.0 has had on the web generally). The reasons for this may be several, including some innate conservatism in the system. One problem may have been the proliferation of competing services: at least four academic social bookmarking services and over a dozen scientific social networking sites exist. Until a clear leader emerges, would-be users (apart from early-adopting enthusiasts) may be actively discouraged from participating by this situation. More recently, services have emerged such as FriendFeed that allow users to aggregate content from various different services which may be one way forward.

Crotty (2008) has written a thoughtful account of the current crop of Web tools for biologists and why they are not more successful. He argues that there has been too much "Web 2.0 for the sake of Web 2.0", copying without thought from the consumer sector, and too much emphasis on the social rather than the timesaving aspects. He sees the main reasons for lack of adoption as being lack of time; lack of incentive; lack of attribution; lack of critical mass; inertia; and inappropriate tools that do not fit the culture of science.

According to Cox & Cox (2008), only about 25% of journal publishers had provided any Web 2.0 technologies (defined as wikis, forums, podcast, blogs, tagging, or "other"). This may not be so surprising, given that uptake appears slow and publisher investments in Web 2.0 remain difficult to monetise.

5.2 Data-driven science

Computers, networks and a variety of automatic sensors and research instrumentation have created an explosion in data produced in research. This does not just create a data management problem (which is as great in lab-bench science such as chemistry as in "big science" projects) but also has the potential to change the way science is done. In the traditional scientific model, the researcher first develops a hypothesis which is tested by gathering the necessary data. In data-driven science, there is an abundance of data which can be explored to create and test new hypotheses.

The ramifications are very diverse but potential impacts on publishing are:

- researchers will want (machine-readable) access to the data underlying the results presented in journal articles both for personal exploration of the data and to permit large-scale data-mining. It is not clear whether hosting this data (e.g. as supporting information files) is best done by journals or whether journals should support deposit in community repositories (e.g. by requiring deposit as a condition of publication, as happens for instance for molecular genetics articles and GenBank). It is noteworthy, however, that as readers academics will say they want the data of their colleagues but as authors they are much less willing to supply it (Mabe & Amin 2002).

- there will need to be two-way linking between journal articles and research databases. There are some research projects working in this area (e.g. SURFshare, eCrystals Federation).
- the dataset may become a (mainstream) unit of publication, with peer review and attribution. As this happens, databases may become more like journals (and vice versa), thus requiring the apparatus of peer review (editor and editorial board, reviewers, etc.). One problem with this vision is there is as yet no known sustainable business model for such data journals. The European Commission is considering issuing a call for proposals in this area.

5.3 Semantic web

The concept of the semantic web involves tagging information published on the web (both articles and data) in a structured way that encodes semantic meaning that can be extracted automatically by computers. The formal concept of a universal semantic web would be very difficult (not to say expensive) to achieve, and may even by impossible in principle, but pragmatic, domain-bounded approaches should be able to add significant value (e.g. see Shotton 2009 or the RSC's Project Prospect[36]).

In combination with the integration with the research data discussed in the previous section, semantic web technologies offer opportunities to enhance journals including

- Improved quality control and support for peer reviewers by automatic validation and consistency checking of manuscripts.
- Improved searching and discovery tools. An early example is Elsevier's Illumin8, a research tool aimed at corporate researchers, which uses semantic indexing to provide e.g. summarised answers and interrelationships that are semantically related to the context of the search query. On the A&I side, the Astronomy Data Service is planning to use semantic technology (i.e. RDF) to interlink astronomy observations and datasets (Kurtz *et al.* 2009).
- The development of semantic recommendation engines will provide powerful new ways to find related material, explore new areas, put research into a broader context, and so on.
- An enriched reader experience with not just a more interactive form of the research paper but with intelligent (i.e. semantically aware) linking.
- In the slightly longer term, adoption of semantic web technologies will facilitate text- and data-mining techniques that hold great promise for accelerating the productivity of researchers. This effectively turns the published literature into a structured database. As well as the technical challenges and licensing issues, new business models to support this may also be required.

5.4 Open notebook science

Open notebook science (also sometimes called open source research) is based on the belief that sharing and collaborating will achieve more than secrecy and competition. It draws its inspiration explicitly from the open source movement in computer software. The idea is to share all research outputs, including work-in-progress and detailed experimental results,

---

36 http://www.rsc.org/Publishing/Journals/ProjectProspect/index.asp

not just the final boiled-down journal article. Two examples are Useful Chemistry[37] and Sortase Cloning[38].

Open notebook science has been adopted by a tiny (close to vanishingly small) minority of researchers and although there will probably be some growth, it seems likely it will stay that way. Most researchers are too concerned about confidentiality and intellectual property rights, about being scooped, and that it would limit their publication options.

5.5 Identity

Unambiguously identifying researchers and their work across the heterogeneous systems that make up the electronic scholarly communication environment is bedevilled by several problems: researchers with identical names (e.g. John Smith); different arrangements or transliterations of the same name; and researchers changing names (e.g. on marriage).

A number of initiatives exist to (partially) address this issue, including Thomson Reuter's ResearcherID, the SciLink social networking site, Author Resolver and the automatic and manual processes in A&I databases like Scopus. The CrossRef organisation has suggested a more fundamental approach (CrossRef Contributor ID) that would combine an author's OpenID with the DOI to link the author to the article.

---

37 http://usefulchem.wikispaces.com/

38 http://chemtools.chem.soton.ac.uk/projects/blog/blogs.php/blog_id/10

6 Conclusions

It is our intention to update this report every 3 years or so. If we take this opportunity to look back over the last 3-5 years, we can see a number of important trends:

- The internet has now reached early maturity as the vehicle through which scholarly communication takes place. Despite this radical change of medium, authors' motivations for publishing in research journals and their views on the importance of peer review remain largely unchanged.
- Building on electronic delivery, the bundling of content and associated consortia licensing model (especially the Big Deal) has delivered an unprecedented increase in access to scholarly content, with annual full-text downloads estimated at 1.8 billion, and cost per download at historically low levels (under $1 per article for many large customers). More than half of journal titles are now bought in bundles of 50 titles or more.
- The Third World still lags the West in digital infrastructure but the success of the Research4Life programmes (HINARI, AGORA, OARE) means that researchers in the poorest countries are not restricted from accessing the scholarly literature by reason of unaffordable subscriptions.
- Globalisation of the scholarly communication system proceeds apace. Perhaps most notable has been the growth of article output from East Asia and particularly China, which is now the second largest producer of research articles in the world.
- The value for money that the Big Deal and similar licences have brought, has largely contributed to the ending of the serials crisis, though that is not to say that the issue of journals cancellations has gone away.
- The debate has instead moved on from serials pricing to open access. Open access journals have continued to grow in number but although the DOAJ lists some 4360 journals, the proportion of the annual scholarly literature that is published in full immediate OA journals remains small at about 2%. About 5% of articles are published under delayed or hybrid OA models, while the uptake of optional OA is still small (~0.1% of all articles). These overall averages do conceal much higher proportions in some fields and the influence of the OA movement, for instance in political circles is much higher than these figures might warrant.
- Self-archiving, the Green route to open access, has been slow to capture the imagination of the scholarly community outside a few fields where sharing preprints or working papers was already the norm.
- The most notable trend in OA in recent years has therefore been the emergence of self-archiving mandates from research funders (lead by the NIH, whose earlier voluntary request for deposit convincingly demonstrated that researchers en masse have very little interest in self-archiving unless compelled) and by institutions (notably Harvard etc.).
- Much of the debate around open access (and scholarly communication generally) has been characterised by a lack of hard evidence and by rhetorical argument on both sides of the debate. Perhaps as a consequence of the serials crisis, trust in publishers is lower than one would like and the scholarly communications community – publishers, researchers, institutions, libraries and funders – is in danger of fragmenting. This may be changing, however: one could point to the establishment in the UK of the Research Information Network with a commitment to supporting evidence-based policy or to project such as PEER.

Looking forward to the next 3-5 years, what might we expect?

- The global recession will certainly take a toll over the next 1-2 years, with impact on library budgets and journal cancellations. It is possible that budget pressure may accelerate a switch to electronic-only subscriptions. Those (relatively few) journals that do enjoy substantial advertising revenues are likely to see these fall sharply. More speculatively, the recession might give a boost to full open access journals relative to subscription journals, insofar as the substantial parts of government stimulus packages devoted to research funding could more easily in the short term find its way into publication charges than into library budgets.
- Article output will continue to grow, and in particular, the growth of East Asia and especially China will continue to substantially outpace growth elsewhere. On present trends Chinese output would overtake the US within the next decade or so and the quality of Chinese research is increasing rapidly. The journals system will need to accommodate this growth and regional shift in terms of business models and in terms of the peer review system.
- Despite the increased access and value for money argument, the Big Deal is under pressure by reason of its perceived lack of flexibility in collection development and inappropriate pricing models (e.g. prior print). We would, however, expect the licensing of bundled content to remain the dominant model over the next few years, albeit with increased flexibility.
- Journal publishing platforms and associated technology have by no means reached maturity and publishers are likely to continue to invest substantially in their development. Areas for development will include Web 2.0; workflow and services; integration of dataset; semantic enhancement of content; and enhanced discoverability and usability of content generally.
- As yet the scholarly community has been slow to embrace Web 2.0 tools but this is likely to change as the tools improve, adapt more closely to the real needs and existing practices of researchers and provide genuine benefits.
- Peer review will remain central to scholarly communication; despite criticisms of aspects of the present system, academics are cautious about change in the area and any innovation is likely to be slow and limited. The importance of journal brands to authors and readers is thus likely to be maintained.
- Open access (or more broadly, the evolution of journals business models) will continue to be the hot topic. Political, societal and cultural pressure for increased access to all kinds of content will not go away and publishers have not yet convinced the whole community that evolution of the existing business models provides the most effective and sustainable way to achieve this. It is possible that some of the momentum behind full open access publishing will fall away if uptake continues at the current overall low levels and as the focus switches to mandated self-archiving. In this area, it seems unlikely that the NIH mandate will be overturned but over the next few years we should have more evidence about the impact of self-archiving on journal sales, from the report of the PEER project and elsewhere. One thing does seem clear, that one size does not fit all: what may work in one field will not necessarily be appropriate for another.

7 References

Albanese, R. (2009) Institutional Repositories: Thinking Beyond the Box, *Library Journal* 1 March 2009 http://www.libraryjournal.com/article/CA6639327.html

ALPSP (2008) Response to European Commission Green Paper 'Copyright in the Knowledge Economy', available from http://www.alpsp.org/

Amin, M. and Mabe, M. (2000). Impact Factors: Use and Abuse. Elsevier White Paper, http://tinyurl.com/f4akm

Anon (2008) Memorandum of Understanding on Diligent Search Guidelines for Orphan Works, available from the website of the European Commission via http://is.gd/GxxV

ARL (2004) ARL Annual Statistics 2004, http://www.arl.org/stats/arlstat

ARL (2009) Statement to Scholarly Publishers on the Global Economic Crisis, February 2009. http://www.arl.org/bm~doc/economic-statement-2009.pdf

Bailey, C.W. (2006). SPEC Kit 292: Institutional repositories, Association of Research Libraries http://www.arl.org/bm~doc/spec292web.pdf

Baldwin, C. (2004). What do societies do with their publishing surpluses? ALPSP/Blackwell http://www.alpsp.org/ForceDownload.asp?id=78

Beckett, C. and Inger, S. (2006) Self-Archiving and Journal Subscriptions: Co-existence or Competition? Publishing Research Consortium http://www.publishingresearch.net/self_archiving2.htm

Biosciences Federation (2008) - Surveys of the financial contribution of bioscience societies to UK universities and of researchers' attitudes to Open Access and self-archiving http://www.bsf.ac.uk/journals/BSF_survey_report_July_2008_FINAL.pdf

Bird C. (2008) Oxford Journals' adventures in open access *Learned Publishing* 21, 200-208 doi: 10.1087/095315108X288910

Björk, B., Roos, A. and Lauri, M. (2009) Scientific journal publishing: yearly volume and open access availability, *Information Research*, 14(1) paper 391 http://InformationR.net/ir/14-1/paper391.html

Bollen, J. et al. (2009) A principal component analysis of 39 scientific impact measures arXiv:0902.2183v1

Butler D (2008) PLoS stays afloat with bulk publishing *Nature*, 454 (3),11 doi:10.1038/454011a

Campbell, P. et al. (2006) Report of Nature's peer review trial, Nature Peer-to-Peer blog, http://blogs.nature.com/peer-to-peer/2006/12/report_of_natures_peer_review_trial.html

Campbell, R. and Wates, E. (2009) Journal publishing: implications for a digital library policy, chapter in *Digital Library Economics*, edited by David Baker and Wendy Evans (Chandos)

Coles, B.R. (1993) The STM Information System in the UK. London: British Library R&D Report No. 6123

Cox, J. and Cox, L. (2008) Scholarly publishing practice: Academic journal publishers' policies and practices in online publishing. Third survey (ALPSP) http://is.gd/3Q932

Cox, L. (2009) Consortium Purchasing Directory 4th Edition (Frontline Global Marketing Services) http://www.frontlinegms.com/42532.html

Craig, I. et al. (2007) Do Open Access Articles Have Greater Citation Impact? (PRC & Journal of Informatics) http://www.publishingresearch.net/Citations.htm

Creaser, C., Maynard, S. and White, S. (2005) LISU Annual Library Statistics 2005 http://www.lboro.ac.uk/departments/ls/lisu/pages/publications/als05.html

Creaser, C., Maynard, S. and White, S. (2006). LISU Annual Library Statistics 2006 http://www.lboro.ac.uk/departments/ls/lisu/pages/publications/als06.html

Crotty, D. (2008) Web 2.0 for Biologists–Are any of the current tools worth using? posting on Bench Marks blog http://www.cshblogs.org/cshprotocols/2008/04/03/web-20-for-biologists-are-any-of-the-current-tools-worth-using/

Davis (2008) Open access publishing, article downloads, and citations: randomised controlled trial *BMJ* 2008;337:a568 doi:10.1136/bmj.a568

EPS (2006) UK scholarly journals: 2006 baseline report. An evidence-based analysis of data concerning scholarly journal publishing (RIN) http://is.gd/3Q5Dw

European Commission (2008) Green Paper 'Copyright in the Knowledge Economy' Brussels, COM(2008) 466/3 http://is.gd/3Q5HH

Goble (2008) *The Future of Research*, British Library Awayday presentation, http://www.semanticgrid.org/presentations/BritishLibrary2008GOBLEpublished.ppt

Houghton *et al.* (2009) Economic Implications of Alternative Scholarly Publishing Models, JISC http://is.gd/3Q5KF

ICOLC (2009) Statement on the Global Economic Crisis and Its Impact on Consortial Licenses, January 19, 2009. http://www.library.yale.edu/consortia/icolc-econcrisis-0109.htm

Inger, S. & Gardner, T. (2008) How Readers Navigate to Scholarly Content, available from http://www.sic.ox14.com/publications.htm

JISC (2005) Nesli2 Usage Statistics Study, JISC, http://www.jisc.ac.uk/uploaded_documents/nesli2_usstudy.pdf

Kaufman-Wills Group (2005), The Facts about Open Access: A Study of the Financial and Non-financial Effects of Alternative Business Models for Scholarly Journals (ALPSP) http://is.gd/3Q5UB

King, D. & Alvarado-Albertorio, F.M. (2008). Pricing and other means of charging for scholarly journals, *Learned Publishing* 21(4) 248-272 doi:10.1087/095315108X356680

King, D. and Tenopir, C. (2004). An evidence based assessment of author pays. Nature Web Focus on Access to the Literature http://www.nature.com/nature/focus/accessdebate/26.html

King, D., Tenopir, C. and Clarke, M. (2006) Measuring Total Reading of Journal Articles *D-Lib Magazine* 12(10) doi:10.1045/october2006-king

Kurtz et al. (2009) The Smithsonian/NASA Astrophysics Data System (ADS) Decennial Report, arXiv:0903.3228

LISU and SQW Consulting (2008) Open Access to Research Outputs: Final report to RCUK, http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/news/oareport.pdf

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 doi:10.1108/00012530210441692

Mabe, M (2004) presentation at 6th Nordic Interlending Conference, Trondheim

Mabe, M & Mulligan, A. (in press) Researcher motivations, behaviour and attitudes in the Internet Age

Mabe, M. (2003). The growth and number of journals, *Serials* 16(2) 191-197 http://uksg.metapress.com/link.asp?id=f195g8ak0eu21muh

Mabe, M. (2006) Journal futures: How will researchers communicate as the Internet Matures?, presentation at Council of Science Editors Annual conference, Tampa 2006

Mabe, M. (2008) STM Publishing: the known Knowns, the known Unknowns ... and all points in between. Presentation at Academic Publishing in Europe conference, Berlin

Matsubayashi, M. et al. (2009) Status of open access in the biomedical field in 2005, *J Med Libr Assoc* 97(1): 4–11. doi: 10.3163/1536-5050.97.1.002

McClure, M. (2009) Turning a New Page in Ebooks, *Information Today* 26(4) April http://www.infotoday.com/IT/apr09/McClure.shtml

Mele, S. (2009) Presentation to the SCOAP3 Forum, Seattle, http://scoap3.org/files/Seattle-140309-Mele.pdf

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

Moed, H.F. (2007) The effect of "open access" on citation impact: An analysis of ArXiv's condensed matter section, *Journal of the American Society for Information Science and Technology* 58(13) 2047-2054 doi 10.1002/asi.20663

Morris (2007) Mapping the journal publishing landscape: how much do we know? *Learned Publishing* 20(4) 299-310 doi: 10.1087/095315107X239654

Morris, S. (2006a) When is a journal not a journal? A closer look at the DOAJ. *Learned Publishing* 19(1) 73-76 doi: 10.1087/095315106775122565

Morris, S. (2006b) ALPSP Alert No. 112 August 2006, page 8 www.alpsp.org/ForceDownload.asp?id=74

Morris, S. (2009) Journal authors' rights: perception and reality (Publishing Research Consortium) http://www.publishingresearch.net/author_rights.htm

NISO (2008) Journal Article Versions (JAV): Recommendations of the NISO/ALPSP JAV Technical Working Group (April 2008), NISO-RP-8-2008 http://www.niso.org/publications/rp/RP-8-2008.pdf

NSF (2006) Science and Engineering Indicators 2006, National Science Foundation http://www.nsf.gov/statistics/seind06/

NSF (2008) Science and Engineering Indicators 2008, National Science Foundation http://www.nsf.gov/statistics/seind08/

Outsell (2009) Information Industry Market Size and Share Rankings: Preliminary 2008 Results http://www.outsellinc.com/store/products/795

PA, ALPSP and ***stm*** (2009) Joint comments by the Publishers Association, the Association of Learned and Society Professional Publishers and the International Association of STM Publishers on the Houghton report. http://is.gd/3Q74S

Renear, A. and Palmer, C. (2009) Strategic Reading, Ontologies, and the Future of Scientific Publishing, *Science* **325**, 828 doi 10.1126/science.1157784

RIN (2007) Research & the Scholarly Communications Process: Towards Strategic Goals for Public Policy http://is.gd/3Q7ap

RIN (2008) Activities, costs and funding flows in the scholarly communications system in the UK http://is.gd/3Q7cm

RIN (2009) E-journals: their use, value and impact. A Research Information Network report http://is.gd/3Q7gF

Rowlands, I. and Nicholas, D. (2005). New Journal Publishing Models, CIBER report. http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf

Rowlands, I. and Olivieri, R. (2006). Journals and scientific productivity: a case study in immunology and microbiology. Publishing Research Consortium http://www.publishingresearch.net/journals_scientific.htm

Rowlands, I., Nicholas, D., and Huntingdon, P. (2004). Scholarly communication in the digital environment: what do authors want, CIBER report

Salo, D. (2008) Innkeeper at the Roach Motel, *Library Trends* **57**:2 (Fall 2008), http://minds.wisconsin.edu/handle/1793/22088

Sanville, T. (2001). A Method out of the Madness: OhioLINK's collaborative response to the serials crisis, *Serials* 14(2) doi: 10.1300/J123v40n01_12

Seglen, P. (1992) The Skewness of Science. *Journal of the American Society for Information Science* 43, 628–638 [quoted In Craig, The Effect of Self-Archiving and Open Access on Citations] doi: 10.1002/(SICI)1097-4571(199210)43:9<628::AID-ASI5>3.0.CO;2-0

Sense About Science (2009) Peer Review Survey 2009: Preliminary Findings http://www.senseaboutscience.org.uk/index.php/site/project/29/

Sharp, S. & Thompson, S (2009) E-books in academic libraries: new challenges, best practice and lessons learned, presentation at UKSG Annual Conference 2009, http://www.uksg.org/event/conference09/breakout_sessions/

Shotton, D. (2009) Semantic publishing: the coming revolution in scientific journal publishing, *Learned Publishing* 22 (2) 85-94 doi: 10.1087/2009202

Sparks, S. (2005) JISC Disciplinary Differences Report. (JISC/Rightscom) www.jisc.ac.uk/.../Disciplinary%20Differences%20and%20Needs.doc

stm (2008) Digital copyright exceptions and limitations for scholarly publications in the education and research communities http://is.gd/3Q7WI

Tenopir, C. (2007) What does usage data tell us about our users? Presentation at Online Information, London

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

UNESCO (2005). UNESCO Science Report 2005. www.unesco.org/science/psd/publications/science_report2005.pdf

Van Orsdel, L.C. (2007) The State of Scholarly Communications: An Environmental Scan of Emerging Issues, Pitfalls, and Possibilities, *The Serials Librarian*, 52, 191-209 doi: 10.1300/J123v52n01_16

Van Orsdel, L.C. and Born, K. (2009) Reality Bites: Periodicals Price Survey 2009, *Library Journal* 15 April 2009 http://is.gd/3Q81A

Ware, M. (2006) Open archives and their impact on journal cancellations, *Learned Publishing* 19(3) doi: 10.1087/095315106777877511

Ware, M. (2009) Access by UK small and medium-sized enterprises to professional and academic information. A Publishing Research Consortium report. http://www.publishingresearch.net/SMEaccess.htm

Ware, M. & Monkman, M. (2008) Peer review in scholarly journals: Perspective of the scholarly community – an international survey. A Publishing Research Consortium report. http://www.publishingresearch.net/PeerReview.htm

Wates, E. and Campbell, R. (2007) Author's version vs publisher's version: an analysis of the copy-editing function, *Learned Publishing* 20: 121–129 doi: 10.1087/174148507X185090

Willinsky, J. (2003). The Nine Flavours of Open Access Scholarly Publishing. *J postgrad med*, 49, 263–267 http://is.gd/3Q8DB