stm

The stm report

An overview of scientific and scholarly journal publishing

Mark Ware
(Mark Ware Consulting & Outsell, Inc.)

Michael Mabe (stm)

Third edition

November 2012

Published by

International Association of Scientific, Technical and Medical Publishers
Prins Willem Alexanderhof 5, The Hague, 2595BE, The Netherlands

About STM

STM is the leading global trade association for academic and professional publishers. It has over 120 members in 21 countries who each year collectively publish nearly 66% of all journal articles and tens of thousands of monographs and reference works. STM members include learned societies, university presses, private companies, new starts and established players.

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.

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

Outsell, Inc. is the world's only research and advisory firm focused solely on media, information and technology. For more information see *www.outsellinc.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 11).

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 14).

The STM market

3. The annual revenues generated from English-language STM journal publishing are estimated at about $9.4 billion in 2011, (up from $8 billion in 2008, representing a 3-year CAGR of about 5.5%), within a broader STM information publishing market worth some $23.5 billion. About 52% of global STM revenues (including non-journal STM products) come from the USA, 32% from Europe/Middle East, 12% from Asia/Pacific and 4% from the rest of the world (page 19).

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 20).

5. Although this report focuses primarily on journals, the STM book market (worth about $4 billion annually) is evolving rapidly in a transition to digital publishing. Ebooks currently make up about 17% of the market but are growing much faster than STM books and than the STM market as a whole. (page 20).

6. There are estimated to be of the order of 5000–10,000 journal publishers globally, of which around 5000 are included in the Scopus database. 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 33).

7. There were about 28,100 active scholarly peer-reviewed journals in mid 2012, collectively publishing about 1.8–1.9 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 6 and 9 million, depending on definition, although only about 20% of these are repeat authors (pages 22).

8. The USA continues to dominates the global output of research papers with a share of about 21% but the most dramatic growth has been in China and East Asia. China's

double-digit compound growth for more than 15 years led to its moving into second position, with 10% of global output. It is followed by the United Kingdom (7%), Japan (6%), Germany (6%) and France (4%). The rank order changes for citations, however, with the US strongly in the lead with 30% and China at 11th place with 4% (page 28).

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 49).

10. Reading patterns are changing, however, with researchers reading more, averaging 270 articles per year, but spending less time per article, with reported 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; at present there is little evidence that social referrals are a major source of access (unlike consumer news sites, for example). Researchers spend very little time on average on publisher web sites, "bouncing" in and out and collecting what they need for later reference (page 37).

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 the rapid growth from emerging economies, which may have temporarily unbalanced the sources of articles and reviewers (page 33).

12. There is a significant amount of innovation in peer review, with the more evolutionary approaches gaining more support than the more radical. For example, some variants of open peer review (e.g. disclosure of reviewer names either before or after publication; publication of reviewer reports alongside the article) are becoming more common. The most notable change in peer review practice, however, has been the spread of the "soundness not significance" peer review criterion adopted by open access "megajournals" like *PLOS ONE* and its imitators. Post-publication review has little support as a replacement for conventional peer review but there is interest in "altmetrics" as a potentially useful complement to review and to other measures of impact (page 35).

13. 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. The number of journal article retractions has grown substantially in the last decade, but the consensus opinion is that this is more likely due to increased awareness rather than to increasing misconduct (page 51).

Technology

14. Virtually all STM journals are now available online, and 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. Consequently the vast majority of journal use takes place electronically, at least for research journals, with print editions providing parallel access for some general journals, including society membership journals, and in some fields (e.g. humanities and some practitioner fields) (page 24).

15. 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. The main barriers to greater use are the lack of clearly compelling benefits of adoption to outweigh the real costs (e.g. in time) of adoption. The rapid development and proliferation of services compound the problem and militate against the development of a critical mass of users for any particular service. Quality and trust issues are also relevant: researchers remain cautious about using means of scholarly communication not subject to peer review and lacking recognised means of attribution. Despite these challenges, social media do seem likely gradually to become more important given the trends in general population, the integration of social features into other software, and the emergence of networks like Mendeley and ResearchGate (page 24; 87).
16. Similarly the rapid general adoption of mobile devices (smartphones and tablets) has yet to change significantly the way most researchers interact with most journal content – accesses from mobile devices account for well under 10% of most STM platform's traffic as of mid-2012 – but this is changing rapidly. Uptake for professional purposes has been fastest among physicians and other healthcare professionals, typically to access synoptic secondary services, reference works or educational materials rather than primary research journals. At the time of writing, use cases, business models and technology choices were still very much in a state of flux (page 20; 24; 88)
17. The explosion of data-intensive research is challenging publishers to create new solutions to link publications to data, to facilitate data mining and to manage the dataset as a potential unit of publication. Change since our last report has been substantial: research funders have introduced or tightened policies requiring deposit and sharing of data; data repositories have grown in number and type (including repositories for "orphan" data); and DataCite was launched to help make research data visible and accessible. Meanwhile publishers have responded by linking or incorporating data; by launching some pioneering data journals and services; by the launch of Thomson Reuters' Data Citation Index; by moving to facilitate text and data mining; and by starting to create new services such as research analytics built around the analysis of metadata (usage, citations, etc.) (page 89).
18. Semantic technologies have become mainstream within STM journals, at least for the larger publishers and platform vendors. Semantic enrichment of content (typically using software tools for automatic extraction of metadata and identification and linking of entities) is now widely used to improve search and discovery; to enhance the user experience; to enable new products and services; and for internal productivity improvements. The full-blown semantic web remains some way off, but publishers are starting to make use of linked data, a semantic web standard for making content more discoverable and re-usable (page 91).
19. Text and data mining are starting to emerge from niche use in the life sciences industry, with the potential to transform the way scientists use the literature. It is expected to grow in importance, driven by greater availability of digital corpuses, increasing computer capabilities and easier-to-use software, and wider access to content. Work remains to be done in terms of standardising content formats (semantic technologies can help here) and the necessary licensing framework (page 54; 93).

Business models and publishing costs

20. Aggregation on both the supply and demand sides have 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. Similar models are also emerging for ebook collections (page 17).

21. While the value of the "Big Deal" and similar discounted packages in widening researchers' access to journals and simultaneously reducing the average cost per subscription and the average cost per article download is recognised, the bundle model remains under pressure from librarians seeking greater flexibility and control, more rational pricing models and indeed lower prices. Nonetheless, its benefits appear sufficient for the model to retain its importance for some time (page 18; 22; 49).
22. Researchers' access to scholarly content is at an historic high. Bundling of content and associated consortia licensing model (especially the Big Deal) has continued to deliver unprecedented levels of access, with annual full-text downloads estimated at 2.5 billion, and cost per download at historically low levels (well under $1 per article for many large customers). 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, while in the US the numbers of serials subscriptions at ARL libraries has grown dramatically to a high point in 2011. Various surveys have repeatedly shown that academic researchers rate their access to journals as good or very good, and that access has improved. The same researchers, however, also identify journal articles as their first choice for improved access. It seems that what would have been exceptional levels of access in the past may no longer meet current needs, and the greater discoverability of content (e.g. through search engines) may also lead to frustration when not everything findable is accessible (page 56).
23. The Research4Life programmes provide free or very low cost access to researchers in developing countries. They have also continued to expand, seeing increases in the volume and range of content and in the number of registered institutions and users (page 59)
24. The most commonly cited barriers to access are cost barriers and pricing, but other barriers cited in surveys include: lack of awareness of available resources; a burdensome purchasing procedure; VAT on digital publications; format and IT problems; lack of library membership; and conflict between the author's or publisher's rights and the desired use of the content (page 58).
25. There is growing interest in expanding access by identifying and addressing these specific barriers to access or access gaps. While open access has received most attention, other ideas explored (for instance by the UK Finch group) have included increased funding for national licences to extend and rationalise cover; walk-in access via public libraries; the development of licences for sectors such as central and local government, the voluntary sector, and businesses (page 59).
26. The average 2010 cost of publishing an article in a subscription-based journal with print and electronic editions has been estimated by CEPA to be around £3095 (excluding non-cash peer review costs). A 2008 RIN/CEPA study estimated that eliminating print editions would save about £1 billion globally (largely in library costs of handling and storing print) (page 3095).
27. Journal publishing has become more diverse and potentially more competitive with the emergence of new business models. Open access posits making original research freely accessible on the web. There are three approaches: open access publishing ("Gold",

including full and hybrid OA journals), delayed free access and self-archiving ("Green") (page 61).

28. There are around 8115 fully open access journals listed on the Directory of Open Access Journals; not all these are fully peer-reviewed although most do, but all have some element of editorial quality control. OA titles are still somewhat less likely than other titles to appear in selective A&I databases such as Scopus or Web of Science, and are (with some notable exceptions) smaller on average than other journals. Consequently the proportion of the 1.8 million articles published per year that are open access is substantially lower than the proportion of journal titles. The most recent estimate places the proportion of 2011 articles published in open access journals at 12% (while OA journals make up about 28% of all journals), with 5% more available via delayed access on the publisher's website, and a further 10-12% via self-archived copies in repositories (page 25; 97)
29. Open access publishing has led to the emergence of a new type of journal, the so-called megajournal. Exemplified by *PLOS ONE*, the megajournal is characterised by three features: full open access with a relatively low publication charge; rapid "non-selective" peer review based on "soundness not significance" (i.e. selecting papers on the basis that science is soundly conducted rather than more subjective criteria of impact, significance or relevance to a particularly community); and a very broad subject scope (in practice limited by authors willingness to submit, the journal's ability to find reviewers, and perhaps in the long run by the usefulness or otherwise to readers to have such broad coverage) (page 73)
30. 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. In the UK, post-Finch government and funder policy has shifted clearly towards Gold rather than Green as a sustainable route to widened access. It is unclear how many other governments will adopt such centralised policy in favour of Gold. Even with broad funder support it remains unclear what a stable financially-sustainable arrangement for Gold open access will look like in detail regarding the funding arrangements within universities. It is unclear where the market will set OA publication charges: they are currently 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; and the more research intensive universities remain concerned about the net impact on their budgets (page 73; 77).
31. 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 have been behind their support of and mandates for open access (and the related, though less contentious policies on data sharing). They have also increased the importance of (and some say the abuse of) metrics such as Impact Factor and more recently are creating part of the market for research assessment services (page 72).
32. Green OA and the role of repositories remain controversial. This is perhaps less the case for institutional repositories (which – despite growth in their numbers and content – remain mainly under-used and disorganised as a source of access to journal content), than for subject repositories, especially PubMed Central. The lack of its own independent sustainable business model means Green OA depends on its not undermining that of (subscription) journals. The evidence remains mixed: the PEER

project found that availability of articles on the PEER open repository did not negatively impact downloads from the publishers's site, but this was contrary to the experience of publishers with more substantial fractions of their journals' content available on the longer-established and better-known arXiv and PubMed Central repositories. The PEER usage data study also provided further confirmation of the long usage half-life of journal articles and its substantial variation between fields (suggesting the importance of longer embargo periods than 6–12 months, especially for those fields with long usage half-lives). Green proponents such as the NIH for their part point to the continuing profitability of STM publishing, the lack of closures of existing journals and the absence of a decline in the rate of launch of new journals since repositories came online as evidence of a lack of impact to date, and hence as evidence of low risk of impact going forward. Many publishers' business instincts tell them otherwise but it is unclear what evidence (short of catastrophic for all concerned) would convince either party of the other's position (page 80).

1. Scholarly communication

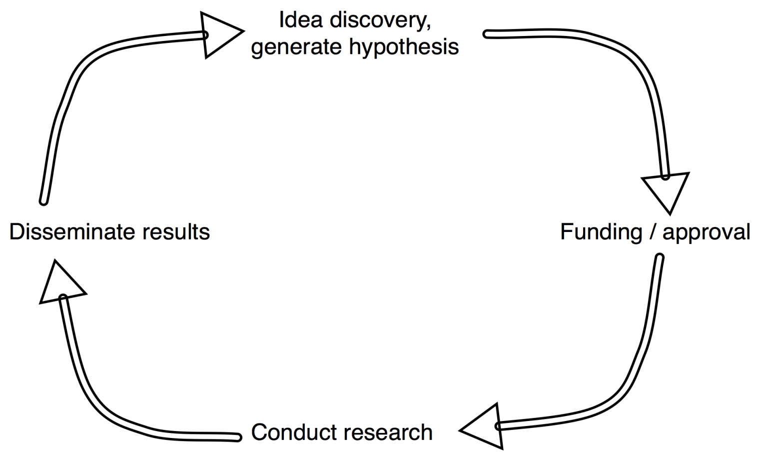
STM[1] publishing takes place within the broader system of scholarly communication, which includes both formal elements (e.g. journal articles, books) and informal (conference presentations, pre-prints). Apart from the academics (and their funders and host institutions) there are two main players in 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)). In some markets (e.g. ebooks, healthcare, industry), aggregators play an important and probably growing role, while for STM generally the research funders are going to be one of the most important parts of the system with the growth of open access.

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] "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.

1.2. Types of scholarly communication

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

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 (see Mabe, 2010 for a more extensive treatment of these arguments). 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. This perspective may be helpful in balancing some of the techno-centric views that assert that the introduction of digital and web technologies will lead to revolutionary change in scholarly communication (see also *Authors' behaviour, perceptions and attitudes*).

Figure 2: Formal and informal types of scholarly communication

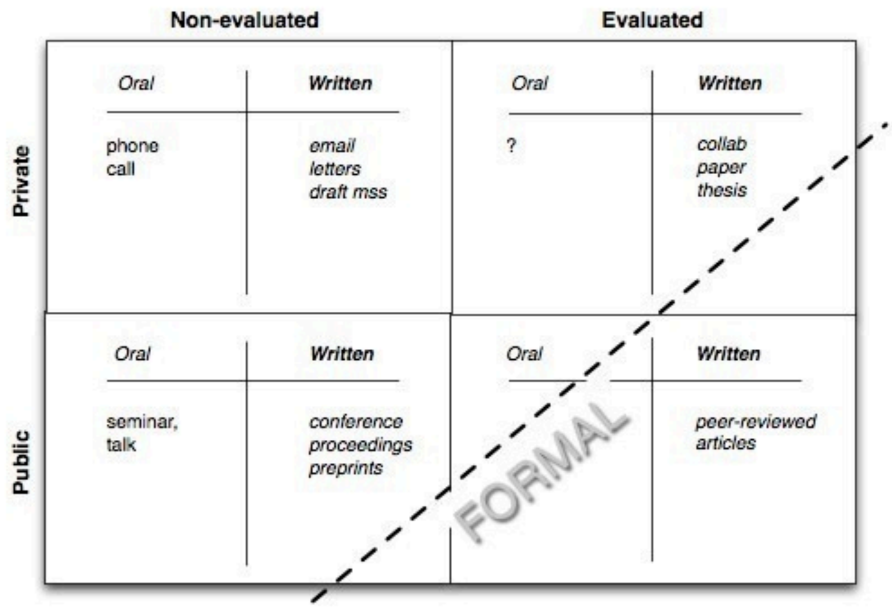


Table 1: Modes of communication

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 (see Authors' behaviour, perceptions and attitudes). Changes can be considered under three headings (see also Van Orsdel, 2008)

- Changes to the publishing market (e.g. new business models like open access; new sales models such as consortia licensing; globalisation and the growth of emerging regions)
- 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.

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 written by the investigators themselves and published after due peer review rather than journalistically based magazines.

The journal has traditionally been seen to embody four functions:

- *Registration*: third-party establishment by date-stamping of the author's precedence and ownership of an idea
- *Dissemination*: communicating the findings to its intended audience usually via the brand identity of the journal
- *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 so much "giving away" the rights as 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.

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

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, while users and the market continue to demand it) printing and binding the journals.
- **Marketeer** – attracting the papers (authors), increasing readership (as important for open access journals as for subscription-based ones) 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.

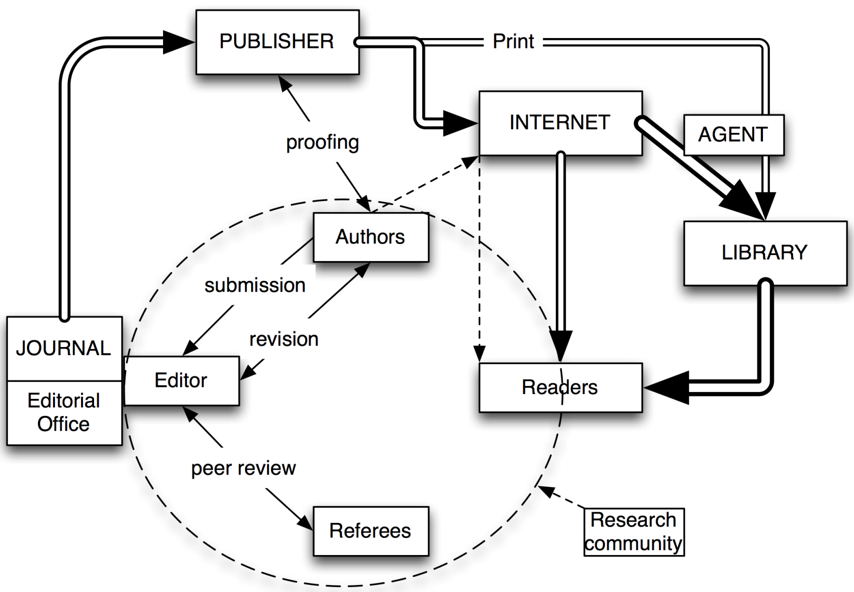
Another way to look at the publisher's role is to consider where they add value. Looking at the STM information arena broadly (i.e. including but not limited to journals), the STM publishers' role can be considered to add value to these processes in the following ways (adapted from Outsell, 2011):

- Sorting and assessment of research outputs: one of the benefits of peer review (Ware, 2008) is the stratification of journals by perceived quality, widely used in assessing research outputs etc.
- Aggregation of content: while other players (e.g. Google, PubMed) are also involved, publishers' aggregation services currently offer widely-used services
- Distillation of evidence: e.g. reference works and meta-reviews
- Creating standards and consensus seeking: a large number of publisher-led initiatives improve the quality, findability and usability of STM content, include CrossCheck, CrossRef, CrossMark, ORCID, etc.
- Granularisation, tagging and semantic enrichment (including development of taxonomies and ontologies), and prioritisation of content, identification, and application of rules: adding value in these ways is likely to become increasingly important
- Systems integration, data structure and exchange standards, content maintenance, and updating procedure: e.g. the SUSHI, KBART standards

- Integration of content from multiple sources: going beyond simple aggregation services, for instance to build sophisticated evidence-based medicine services drawing on multiple content types and sources to support doctors at the point of care
- Creating and monitoring behaviour change: e.g. enforcing standards of disclosure of interest in medical journals; some journals encourage (or require) the parallel deposit of research data
- Development of workflow analytics and best practice benchmarking at the level of the individual, department, institution, and geopolitical entity: e.g. tools to support research assessment.

Cliff Morgan and coauthors review the role of the publisher in the context of open access developments and suggest a similar set of activities will continue to be required, and point out that publishers have collectively invested of the order of $3.5 billion in online publishing technology since 2000 (Morgan, Campbell, & Teleen, 2012).

Figure 3: The publishing cycle



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 potentially confusing (e.g. the post-print definitely does not occur post printing).

The CrossRef organisation has introduced a CrossMark service to identify the version of record (Meyer, 2011). There is a visible kitemark that identify it to the human reader. There would also be defined metadata for search engines etc. The CrossMark does not just identify the article as the version of record but also provides information about the pre-publication process (e.g. peer review) and of post-publication events such as errata, corrections and retractions. The service is still relatively new and not yet widely implemented.

2.3. Sales channels

Journals are marketed to two broad categories of purchaser, namely libraries and individuals (see separate section below for open access journals). Although individual subscriptions (either personal or membership-based subscriptions) can be important for some journals (for example magazine/journal hybrids such as *Nature* and some (especially medical) 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 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 STM 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] currently lists about 30 agent members but the number of agents has been declining in recent years (the ASA membership was reported at 40 in the last edition of this report), 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.

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

The larger subscription agents are consequently reinventing themselves, for instance as aggregators, publishers, and providers of analytics services.

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 were sold in bundles of 50 titles of more (Van Orsdel & Born, 2009), and this fraction is likely to have risen subsequently. According to Cox & 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[3]; 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 consortium) is frequently referred to as the **Big Deal**. The other main pricing models include: usage-based pricing, first tried during the mid-2000s but without gaining much momentum, linkage between licence costs and usage now appear to be becoming increasingly important;[4] tiered pricing based on a classification of institutions by size; 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 ebooks, 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: the Ringgold Consortia Directory Online[5] lists over 400 consortia representing over 26,500 individual libraries;[6] of these, about 350 are responsible for licensing content. The International Coalition of Library Consortia[7] has some 200 members. 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

[3] It is worth noting that this has often been a pragmatic rather than a conservative approach, since the prior print has in many cases been the last point of agreement between the library and the publisher over pricing principles. More advanced database models can have advantages and disadvantages, and neither party wants the disadvantages.

[4] e.g. see the late-2011 deals secured by JISC Collections int he UK and Wiley and Elsevier

[5] http://www.ringgold.com/pages/cdo.html

[6] Growth can be indicated by the earlier editions of this report, which recorded 338 active consortia in 2008, up from 164 in 2003

[7] http://icolc.net

number of individual libraries covered by consortia is of the order of 5000. According to an ALPSP report (Cox & 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 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.

Library system vendors[8] provide the cataloguing, enterprise resource planning and link-resolver and other access systems used by libraries. Although their business relationships 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[9] and KBART (Knowledge Bases And Related Tools).[10]

2.4. Journal economics and market size

Journal economics & market size

The total size of the global STM market in 2011 (including journals, books, technical information and standards, databases and tools, and medical communications and some related areas) was estimated by Outsell at $23.5 billion[11] (Outsell, 2012d). The value of the market as reported in dollars was boosted in 2011 compared to 2010 by a significant appreciation of the euro against the dollar; excluding all currency effects, the Scientific & Technical subsegment of the market grew at 4.3% and Medical at 2.0%.

Within this overall market for STM information, the global 2011 annual revenues from journals were estimated by Outsell at $9.4 billion, and those from books (and ebooks) at $3.8 billion. The market can also be divided into scientific/technical information at $12.8 billion and medical at $10.7 billion.

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, Outsell estimates about 52% of global STM revenues (including non-journal STM products) come from the USA, 32% from the EMEA region, 12% from Asia/Pacific and 4% from the rest of the world (principally the Americas excluding USA, though S Africa and some North African countries are showing strong growth potential) (Outsell, 2012d). These proportions probably overstate the importance of the USA market for journals alone.

Market analysts Simba estimated the STM market in 2011 using a slightly narrower definition than Outsell at $21.1 billion, with journals at about $9 billion (Simba, 2011).

---

[8] See http://www.librarytechnology.org/ for one overview and list of suppliers

[9] http://www.editeur.org/17/ONIX-for-Serials/

[10] http://www.uksg.org/kbart

[11] this and other market size figures are at actual values for cited year, i.e. not updated to current values

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).

China

The biggest change in the global economy since the last STM Report has been the rise of China. Although China has become the world's second-largest producer of research papers, its share of the global STM market is much smaller than this might suggest. For example, Outsell estimate China comprised less than 5% of the global STM market by revenue in 2011; by contrast, as noted above, the US share was 52% (Outsell, 2012d, 2012f). The market is split roughly two-thirds/one-third by value between international and domestic publishers.

Part of the reason for this disparity between research spending and share of the STM market is the early emerging stage of the Chinese research infrastructure. Another reason, however, was the very low pricing that some publishers adopted to enter the market in the early days, a strategy that has continued to depress pricing in the market.

Nonetheless the Chinese market grew at 10% in 2011, significantly outperforming the global market, and Outsell estimates that the China STM market will grow between 10% and 11% year on year between 2011 and 2013.

Books and ebooks

The ALPSP *Scholarly Book Publishing Practice* report gives a detailed picture of the STM book and ebook market in 2009, based on the analysis of 170 publishers' survey responses (Cox & Cox, 2010). Though the market has moved on since 2009, especially in relation to ebooks, this is still a useful source of information. The publishers included (representing a good fraction of the total market) published over 24,000 new titles each year, with a backlist of nearly 350,000 academic and scholarly titles, covering reference, monographs, textbooks, conference reports, professional handbooks and manuals, and research reports. Most of the publishers (over 90%) published for the research and post-graduate market, about two-thirds for under-graduates, and around 40% published general reference titles. While ebook publishing had taken off dramatically compared to an early ALPSP study in 2004, only about two-thirds of publishers were publishing in electronic formats, and for them ebook revenues were under 10% of total book sales.

A more recent picture is given by a 2012 report from Outsell (Outsell, 2012e), which estimated the 2011 global market for ebooks at $670 million, representing about 17% of the STM book market. It was growing much faster than the overall STM market, with 23% growth in 2011 compared to 4.3% for the overall market, and had grown considerably faster than the books sector overall, with a 2008–2011 CAGR of 33.7% compared to 2.1%.

Outsell found that market take-up in the scientific and technical segment was greater than in medical: the latter category comprised 44% of books but only 35% of the ebook market. This seemed paradoxical, given that medical practitioners are among the highest users of digital content and mobile devices in the workplace. The reasons given were that publishers and aggregators were able to sell bundles of content alongside existing channels for scientific and technical, while – although individual medical practitioners were warming to digital books and content – institutional purchasing arrangements were not well structured for bulk purchase of medical ebooks, and there was budget competition for aggregation services and evidence-based medicine and point-of-care products.

Reference content (and to a lesser extent, monographs) were in the vanguard of digital conversion, with publishers reporting digital revenues comprising a substantial majority of reference work sales. By contrast, textbooks were least amenable, with revenues under 10% for digital versions. There were two reasons given. Reference works were easier to digitise (although editorial/production workflows do have to be reengineered for frequent and regular updates), whereas textbooks required more additional functionality to support learning and pedagogy. Second, business models for reference works are more straightforward, while textbook publishers are grappling with the difficulties of adapting the print-based adoption and individual sales models to the digital environment, as well as mixed responses from students to digital textbooks.

A significant difference between books and journals is that academics are far more likely to purchase the books themselves; for example, Tenopir, Volentine, & King, 2012 reported that the single most common source of scholarly readings from books was personal copies (at 39%), well ahead of supply via the library (at 26%), whereas articles were mainly obtained from the library ecollections.

There is considerable business model innovation in digital textbooks and associated educational market, much of it potentially highly disruptive, including freemium models (e.g. the basic content is free online with charges for additional services such as more functional formats, printing, testing and class-support tools, etc.; examples include FlatWorld Knowledge and Knowmia); Nature Publishing Group's *Principles of Biology* offers students lifetime access to a regularly updated online textbook; advertising-supported (e.g. World Education University); and grant-funded (e.g. Rice University's OpenStax College). There has also been a massive recent growth in open educational resources (OERs): initially consisting of leading universities like MIT making their course materials freely available online,[12] more recently there has been a wave of start-up and spin-off companies such as Coursera, Udacity, Udemy and GoodSemester, and expansion of the university offerings to include certificates (notably MITx and edX). Leading educational publishers, notably Pearson, are moving in the opposite direction by building or acquiring the capabilities to offer an end-to-end service including not just the textbooks and educational content, but also testing, online learning environments, and the creation and delivery of its own courses, and in the near future it will have the power to accredit and grant its own degrees following changes to UK legislation (Pearson, 2012).

Global market 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 £175 billion, made up of £116 billion for the costs of the research itself; £25 billion for publication, distribution and access to the articles; and £34 billion for reading them.[13]

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

---

[12] See MIT OpenCourseWare http://ocw.mit.edu

[13] Values from 2008; not inflated to current values

Prospects for the STM market

The effects of the 2008/09 recession are still being felt, with growth in much of Europe stalled and the US recovery substantially weaker than at this point of the cycle in earlier recessions. Public sector spending is weak in many developed countries as they seek to control budget deficits, leading to inevitable pressure on, and cuts in, institutional and library budgets.

With its limited dependence on advertising, the STM market is far less cyclical than many information markets, and the broadly-defined STM market has managed to sustain growth rates between 4% and 5% since 2007, and is projected to continue growing at around this rate in the short-to-medium term (Outsell, 2012d), absent any further major economic shocks (e.g. a widespread collapse of the Eurozone, or a new war in the Middle East). These overall figures, however, conceal mixed fortunes for the different geographical and product segments within the market. STM publishers' traditional core markets – journal and book sales to institutional libraries in the developed world – have been flat, with growth barely keeping pace with inflation, and are likely to remain so. The market growth has come from the emerging regions, and similarly these will continue to be the focus of revenue growth for most publishers; for example, Simba says the Indian market is expected to be on a par with most G8 nations by 2017.

By product segment, the bright spots have been ebooks and online databases, tools and other services. These are likely to remain sources of growth, with the broader mobile category joining ebooks as a significant growth driver. Another possible source of growth may be the corporate and SME sector: this (apart from pharma and the technical segment to varying degrees) has traditionally been under-served by STM publishers because of the mismatch between journals and the needs of industry. Large publishers such as Springer and Elsevier appear to have recognised an opportunity and restructured accordingly, with dedicated sales and marketing teams focused on the corporate sector, and sector-specific products and portals.[14]

Although there is limited hard evidence to support this, it seems possible that the recession and associated budget pressures may further increase business model innovation. Despite some criticisms, the Big Deal continues to be the major part of the journals (and increasingly the books) market, although it is evolving, for instance with greater linkage of pricing to usage.

2.5. Journal and articles numbers and trends

There were about 28,100 active scholarly peer-reviewed journals in August 2012,[15] collectively publishing about 1.7–1.8 million articles a year (Royal Society, 1995; NSF, 2012; Björk, Roos, & Lauri, 2009). At the time of writing, the CrossRef database included over 56 million DOIs, of which 46 million refer to journal articles.

Journals which published only original research articles 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

[14] e.g. Springer for R&D (http://rd.springer.com); Elsevier's Illumin8 (http://illumin8.com)

[15] Ulrich's web directory listed 28,094 active, peer-reviewed scholarly/academic journals on 3 August 2012. The CrossRef database included 27,000 journals in March 2012. Thomson Reuters' Web of Knowledge database covers some 10,675 journals, while Scopus covers 18,500 peer-reviewed journals

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 5%.

An important subset is the 10,675 journals included in Thomson Reuter's Journal Citation Reports database, of which 8200 are in the Science Edition and 2900 in the Social Sciences Edition: these 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 other major A&I database, Scopus, is intentionally broader in scope, and covers 18,500 peer-reviewed journals.

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).

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

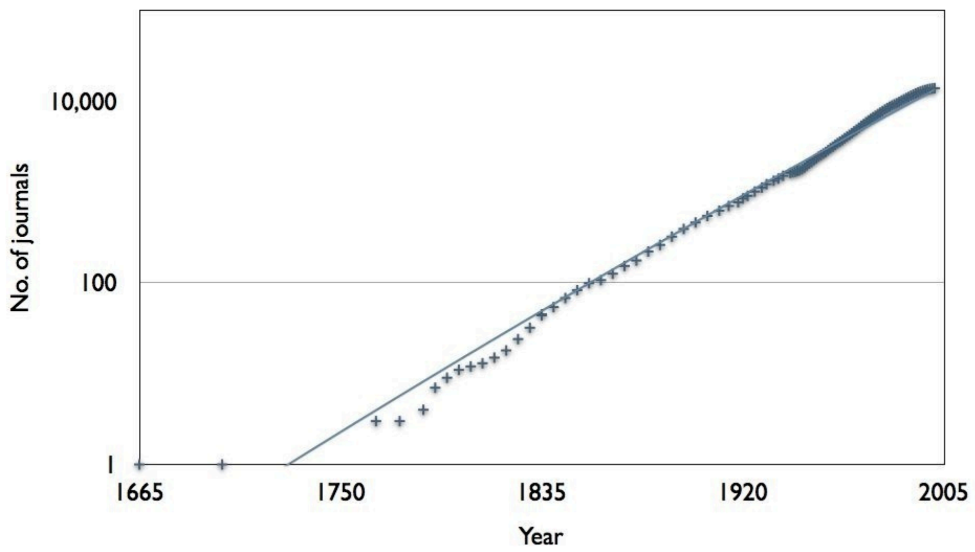
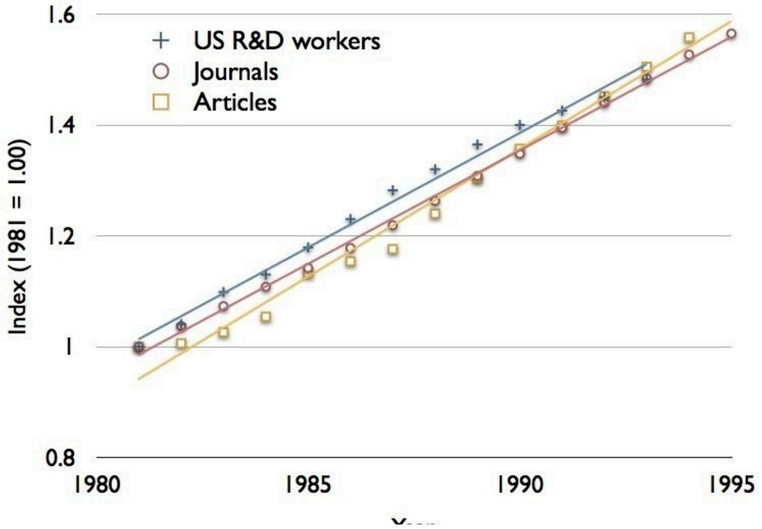


Figure 5: Relationship between numbers of researchers, journals and articles (Mabe, 2004, using data from ISI and NSF)



Online journals

All STM journals are now available online, with just a few exceptions (e.g. very small journals; some journals in the humanities). As far back as 2008, ALPSP's report on scholarly publishing practice (Cox & Cox, 2008) had already found 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%).

Very few journals, however, have yet dropped existing print editions. The main reason is continuing demand from residual parts of the market, including individual and society member copies, and institutional customers in some parts of the world. The factors sustaining this demand for print include its superiority for some uses, concerns about the long-term preservation of digital formats, concerns about access to digital content following subscription cancellation or in the event of publisher demise, caution by some advertisers in switching to digital formats, and tax disincentives in some territories. Print's advantage over digital in terms of portability and readability seem likely to be eroded by the latest tablets, and these mobile formats also appear to be offering some compelling benefits to advertisers. Digital preservation and continuing access issues are addressed in a wide variety of programmes including LOCKSS/CLOCKSS, Portico, national library programmes, etc., while librarians and users are becoming accustomed to online-only journals through Big Deal arrangements and through newly launched titles (including open access journals). While digital printing technologies (including print on demand) make it economically feasible to supply ever-lower levels of print demand

Books are another matter. As noted above (*Books and ebooks*), ebooks made up only about 17% of STM book revenues in 2011. Growth rates are predicted to be high, though, particularly in reference works and monographs in the sciences, while textbooks may take longer to move largely to digital, although there is a lot of innovation in this area.

Open access journal and article numbers

The number of open access journals listed by the Directory of Open Access Journals[16] was 8115 as of early September 2012; this represents an increase of 3750 over the 3 years since the last STM Report, or about 3.5 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). Ulrich's Directory lists 4365 peer reviewed OA journals, or about 13% of the total number of peer reviewed journals included.

The proportion of OA journals included in the major A&I databases is a little lower than the Ulrich's figure, which is not surprising given the higher barrier to inclusion and the lower average age of OA journals. Scopus covers 18,500 peer-reviewed journals, of which 1800 or 9.7% are open access, while the proportion of OA journals covered by Journal Citation Reports is about 8%.

Open access article numbers

Counting the number of open access journals is one thing, but because journal sizes varies wildly (e.g. from a small quarterly publishing 20 articles a year up to *PLOS ONE*, which published 14,000 articles in the first 8 months of 2012), a better measure of the uptake of open access by the research community is the number of articles, in absolute terms and as a proportion of total articles.

The situation is complicated by the four types of open (or free) access that might be counted (see the later section *Open access*): articles in pure OA journals (Gold); OA articles in subscription journals (hybrid); delayed free access to subscription articles following an embargo; copies of articles (which may or may not be the version of record) in open repositories (Green). In addition some subscription journals make articles freely available for limited periods for promotional or other reasons.

Figure 6, from a 2012 article in *Nature* (Van Noorden, 2012a), summarises estimates of the proportion of articles that were Gold open access between 2003 and 2011. The data used draws on different sources including the Web of Science and Scopus databases. The figure of 12% for 2011 is based on estimates of the fraction of OA articles in Scopus made by Mikael Laakso and Bo-Christer Björk (subsequently reported in Laakso & Björk, 2012), and refers to articles immediately available via the publisher's site (i.e. full and hybrid OA), with an additional 5% available via delayed access.

An earlier study by Björk estimated the fraction of the literature published in 2008 that was freely available in 2009 by generating a random sample of articles using a bibliometric database and then searching the web (Björk et al., 2010). The overall findings (for all fields) were that 20.4% of 2008 articles were freely available:

- 8.5% available at publishers' websites (i.e. Gold, hybrid and delayed access)
- 11.9% additionally available on various websites (not just institutional and subject repositories, but also author and departmental websites, etc.)

The methodology of these two studies is not exactly comparable but nonetheless they give an indication of the growth on Gold open access between 2008 and 2011.

The SOAP project (SOAP, 2010) identified about 120,000 articles published in Gold or hybrid OA journals. This would amount to about 8% of articles, in line with Bjork's findings.

[16] http://www.doaj.org/

The Björk paper also gave a breakdown of the 8.5% available at publishers' websites:

- Gold journals: 5.3%
- hybrid articles: 2%
- delayed access: 1.2%

The availability varied substantially by discipline, as shown in Figure 7. Uptake of the Gold model was clearly highest in the biomedical disciplines, where research funding tends to be higher and where research funders. For example, using the PubMed search facility shows that the proportion of articles published in 2011 and covered by PubMed that have free full text available (by any route) was 25.8%, compared to the 20–22% reported by Björk.

Figure 6: Growth in estimates of the fraction of articles published as Gold open access (Van Noorden, 2012a)

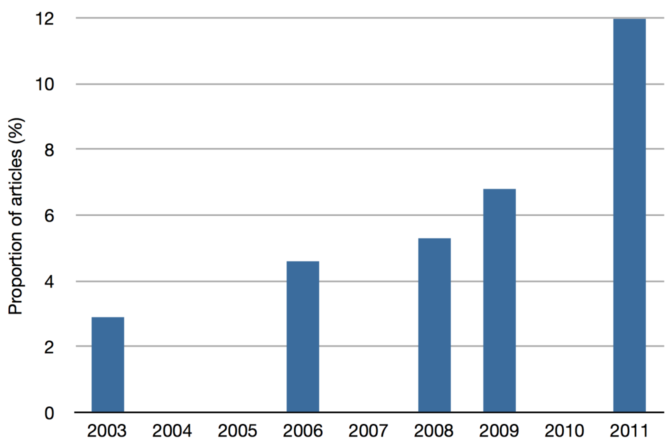
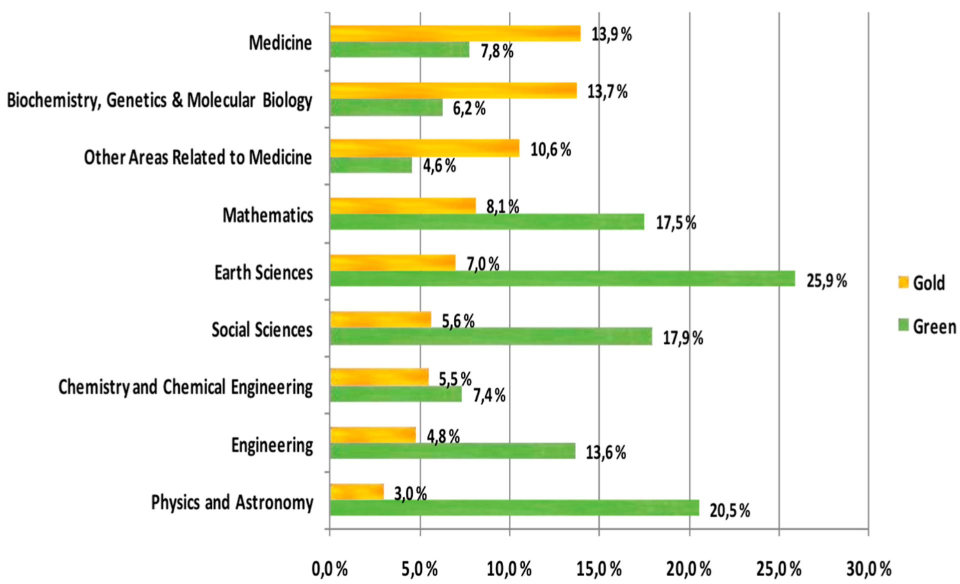


Figure 7: OA availability by discipline. Reproduced from Björk et al., 2010, doi: 10.1371/journal.pone.0011273.g004



2.6. Global trends in scientific output

R&D expenditures

As we have seen, the numbers of research articles are closely correlated to the numbers of researchers, which in turn is closely linked to the amount spent of research and developments.

Global spending on R&D has consistently grown faster than global GDP, rising from $522 billion in 1996 to $1.3 trillion in 2009 (NSF, 2012) and an estimated $1.4 trillion in 2012 (Battelle, 2011). The large majority of this spending (92%) takes place in the three major economic regions of the world, N America, the EU and Asia. The USA spends by far the largest amount compared to other individual countries at $400 billion, though for the first time this position is beginning to be challenged.

Governments see spending on R&D as critical to innovation, growth and international competitiveness. Across the world, the average proportion of national GDP spent on R&D was about 1.7% in 2010, although there is (unsurprisingly) a wide range in this, from oil-rich Saudi Arabia's 0.04%, through India's 0.8%, Canada's 2% to Sweden's 3.7%, with the average for the OECD countries at 2.4%. The trend to increased relative spending on R&D will continue over the long term: although the US set a goal in the 1950s for R&D of 1% of GDP, its expenditure is now pushing 2.9% and many countries (including the EU as a whole) have set targets of 3% of GDP. (OECD, 2011; UNESCO, 2010.)

The growth in R&D spending in China is particularly notable, tripling from 0.6% in 1996 to 1.7% of GDP in 2009, with China's GDP growing by a compound 12% over the same period. China's R&D strategy and some of the impacts are discussed in more detail below. Other emerging countries are also rapidly ramping up their R&D expenditures; Brazil plans to invest 1.5% of its GDP on R&D by 2012 and aim to achieve 2% before 2020 (BIS, 2011).

Although research outputs are driven primarily by the numbers of researchers, there are substantial variations in research productivity, with for example UK researchers generating more articles and more citations per researcher, and more usage per articles compared to all other countries in the top five (US, China, Japan, Germany) (Royal Society, 2011).

The impact of the recession was felt harder in the US and the EU than in China, Brazil and India, allowing these countries to grow their share of global R&D spending faster than otherwise. Global growth in R&D spending in 2012 was estimated to grow a little more slowly in 2012 (5.2%) compared to 2011 (6.5%), due to continued recessionary effects and the ending of stimulus spending packages.

Role of industry

The majority of R&D expenditure is funded by industry: about 62-64% in the US, 54% in the EU (ranging from 45% in the UK to 70% in Germany), and between 60% and 64% in China, Singapore and Taiwan. The fraction of R&D that is performed by industry is even higher, at a little over 70% in the US, for instance (NSF, 2012; Battelle, 2011). This is important for publishing, because the majority of research papers originate from academic authors.

Most of the research included in these expenditure figures is not basic science but more applied R&D. In the US, the fraction of R&D spending on basic sciences is estimated at 18%, and perhaps surprisingly, the share of US R&D devoted to basic science has doubled over the last 50 years. Nearly all of this is performed by academia, though in the past industry and government researchers did substantially more – the days of Bell Labs churning out

Nobel Prizes (13 at the last count) are gone for good. As a consequence, US industry is more dependent on academia for the basic research underpinning innovation than in the past.

Numbers of researchers

There is no single comprehensive and widely accepted set of figures for researcher numbers, principally for reasons of difficulty of defining a researcher once you leave academia. The Royal Society give a figure based on UNESCO data of 7.1 million researchers in 2007, up from 5.7 million in 2002 (Royal Society, 2011), while Elsevier's report for the UK government gives a lower figure, estimated at 5.95 million for 2009 (Elsevier, 2011). (The lower figures are based on the Frascati Manual definition of researcher, which is more tightly defined than UNESCO's "scientist and engineer"; for example, China's total falls from 1.6 million to 1.1 million when re-based on the Frascati definition.) OECD estimates there are about 7.5 researchers and engineers per thousand people in employment (7.5% in the EU, 9% US< 10% Japan). Projecting the UNESCO figures to 2012, assuming the 2002/2007 growth rate was maintained, would give an estimate of 8.7 million.

Whichever definition is used the number of global researchers is steadily growing, at about 4–5% per year. The majority of this growth is driven by emerging countries, with 8–12% annual growth in the leading Asian countries in marked contrast to around 1% in US and EU. One consequence of this is that China will shortly overtake the US and EU in numbers of researchers; similarly, the combined number of researchers from South Korea, Taiwan, China, and Singapore increased from 16% of the global total in 2003 to 31% in 2007 (Royal Society, 2011).

Regional differences

The cumulative effect of sustained above-global-average growth in R&D spending in emerging economies has been a profound shift in the global make-up of research. As the consultants McKinsey described the economic changes, by far the most rapid shift in the world's economic centre of gravity happened between 2000 and 2010, and the same was true for the global research picture. For the first time since WWII, America's leadership is starting to be challenged by China.

China is predicted to overtake the US as the world's largest economy by 2016 or thereabouts. Its R&D spending has trebled since 2005 to £70 billion and spending is planned to increase further to 2.2% of GDP by 2015, and it has set a target of 2.5% of GDP by 2020. Over the last decade its annual growth in R&D expenditure has been over 20%, and on current trends China's research spending will exceed the US's by about 2023 (Battelle, 2011; NSF, 2012).

Underpinning this budgetary growth, China is working hard to strengthen its higher education institutions and research base. "Project 211" aims to strengthen selected HEIs to world standards. The 113 institutions involved educate 80% of PhD students; host 96% of China's key laboratories; and receive 70% of science research funding. Similarly, Project 985 targets selective investment at universities to achieve world status. Of the 39 universities included, two (Peking, Tsinghua) are targeted to be among best in world, a further eight to become world class; with the reminder targeted to be the best in China and well known in world (Outsell, 2012f).

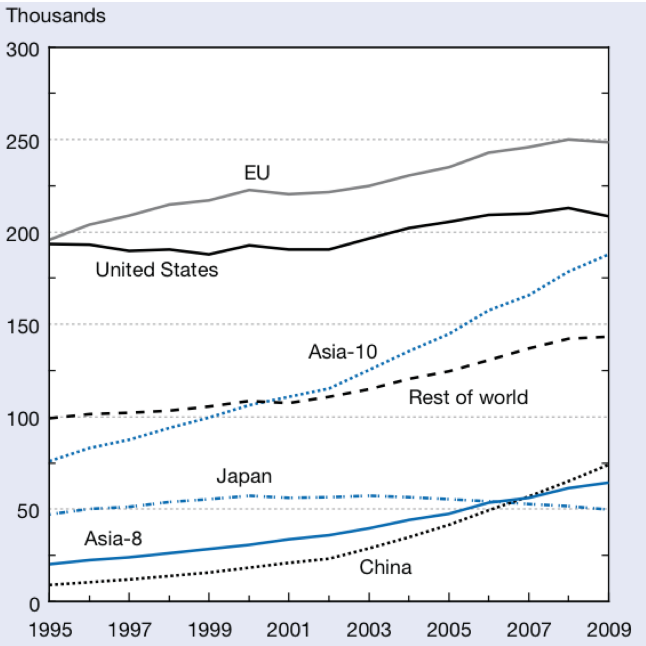
The consequences of the huge growth in research spending on research outputs are predictable. China has overtaken the UK to publish the second largest annual number of research papers, with its share now at over 10%, and is set to overtake the US well before 2020. (Figure 8 shows the trends in article outputs from 1995 to 2009.) At present the impact of these papers is well below its rivals: although ranked 2nd in outputs, it comes 20th in citations per article.

China is not the only country to grow share of world publications. Between 2002 and 2008, all the BRIC countries increased their shares except for Russia, whose share fell to 2.7% from 3.5%. Latin America saw its share of publications over this period rise from 3.8% to 4.9%. The US remains for now in first place, but it share fell the most (in percentage points) from 30.9% to 27.7%. (UNESCO, 2010.) Brazil, India and S Korea's economies are also likely to exceed those of France and Japan in the early 2020s.

The research priorities and emerging strengths of the developing nations are by-and-large different from the historical strengths of developed countries. For example, the UK and US both have comparative strengths in biomedical science and clinical research as well as earth and space sciences, while China (whose research priorities are more tightly focussed than many) is developing strengths in physics, chemistry, maths and engineering.

For UNESCO, these changes amount to a "structural break in the pattern of knowledge contribution to growth at the level of the global economy". In other words, countries no longer need to build their knowledge bases from the ground up via national R&D, but developing countries can (also) build on the world stock of knowledge, make use of under-exploited technology, and do so at less risk. Geographic boundaries are at the same time less relevant for research and innovation and yet more important than ever before.

Figure 8: Science & engineering journal articles produced, by selected region/ country: 1995–2009 (source: NSF, 2012)



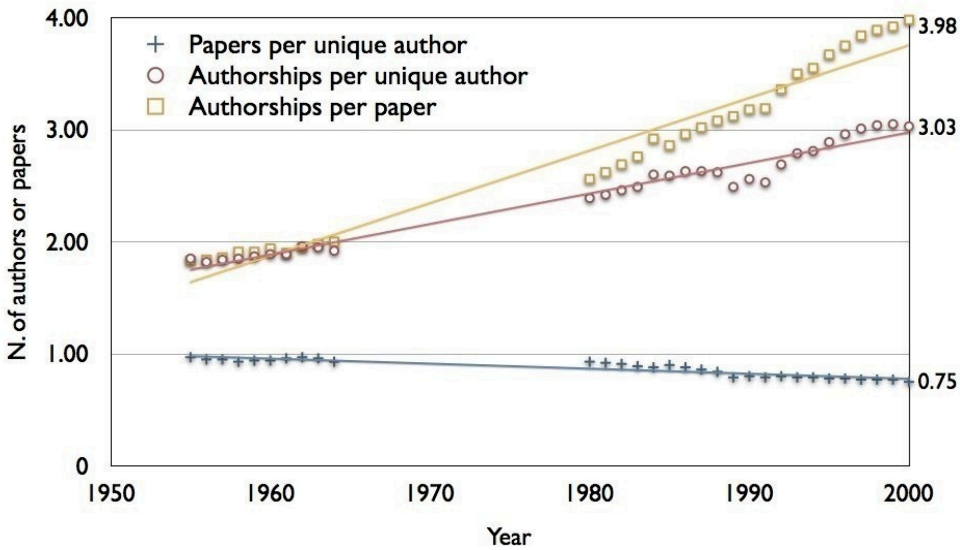
Collaboration and coauthorship

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.

Collaboration is now the norm, reflected in both an increase in the average number of authors and institutions on an article, and in the proportion of international collaboration. In 1988, only 8% of the all articles had international coauthors, but this figure had risen to 23% by 2009, and for the major science and technology regions the proportion ranges from 27% to 42% (NSF, 2012). The Royal Society quote figures based on a different dataset, estimating that today 35% of articles are internationally collaborative, up from 25% 15 years ago (Royal Society, 2011). Figure 10 shows the trends in the proportions of research articles with international coauthors. Interestingly the trend is not upwards for all countries, with the proportion for China and Taiwan staying roughly constant or even declining. International collaboration for Turkey and Iran (not shown in the figure) are also declining slightly. This likely reflects the newness of research institutions in these fast-growing regions.

Overall, the number of author names per US article increased from 3.2 in 1990 to 5.6 in 2010 (NSF, 2012); across the Thomson Reuters database as a whole from 3.8 in 2007 to 4.5 in 2011 (ScienceWatch, 2012). Figure 9 shows how the average number of authors per paper grew during the second half of the 20th century, while Figure 11 shows how the growth in coauthorship has varied by discipline, with the largest numbers of coauthors and largest increases in physics and astronomy, and the smallest coauthorship in mathematics and social sciences. Another reflection of this trend is that coauthored articles grew from 42% to 67% of world output between 1990 and 2010. A more recent trend has been the increase in papers with more than 50 authors, and even with more than 1000 authors ("hyperauthorship"), driven largely by international high-energy physics collaborations. In 1981, the highest number of authors on a paper indexed by ISI was 118, while in 2011 it was 3179. The trend has provoked debate over the nature of authorship, with come calling for the term "contributor" to be distinguished from "author" in such cases.

Figure 9: Coauthorship patterns 1954 to 2000 (from Mabe & Amin 2002, using data from Thomson Reuter Science Citation Index)



The US is a particularly important partner for international collaboration, with 43% of all internationally collaborative papers including at least one US-based coauthor in 2009 (NSF, 2012). In a similar vein, the BRICs' collaboration with each other is minimal, dwarfed by their collaborations with the G7 partners.

There is a clear benefit to researchers to collaborating in terms of increased citations (and to a less marked extent, increased usage). The average number of citations received per article increases with each additional collaborating country (i.e. in addition to the lead author's country); articles with 5 additional countries receive nearly three times as many citations as those with none (Royal Society, 2011). For individual countries the size of the effect varies but tends to be especially strong for developing countries, presumably because they are benefiting from collaborating with better established research teams in developed countries; for China, for example, papers with international collaborators receive 3.1 times as many citations as those with no collaborators beyond the lead institution (Elsevier, 2011).

Figure 10: Research articles with international coauthors, by selected region/country/economy: 1989–2009 (source: NSF, 2012)

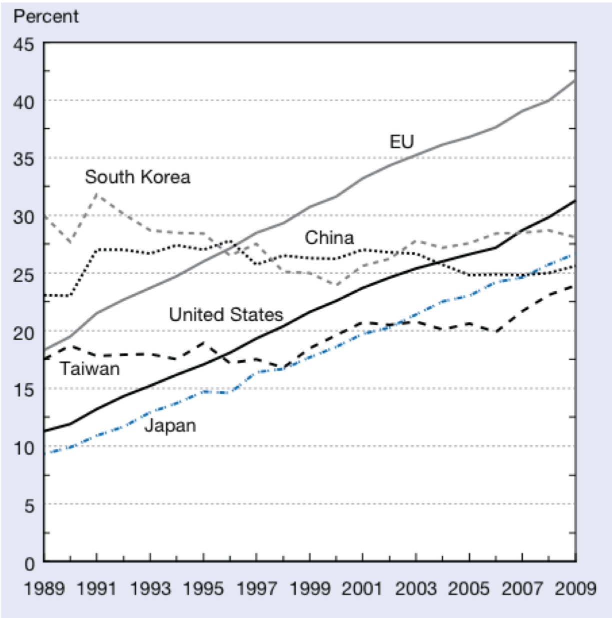
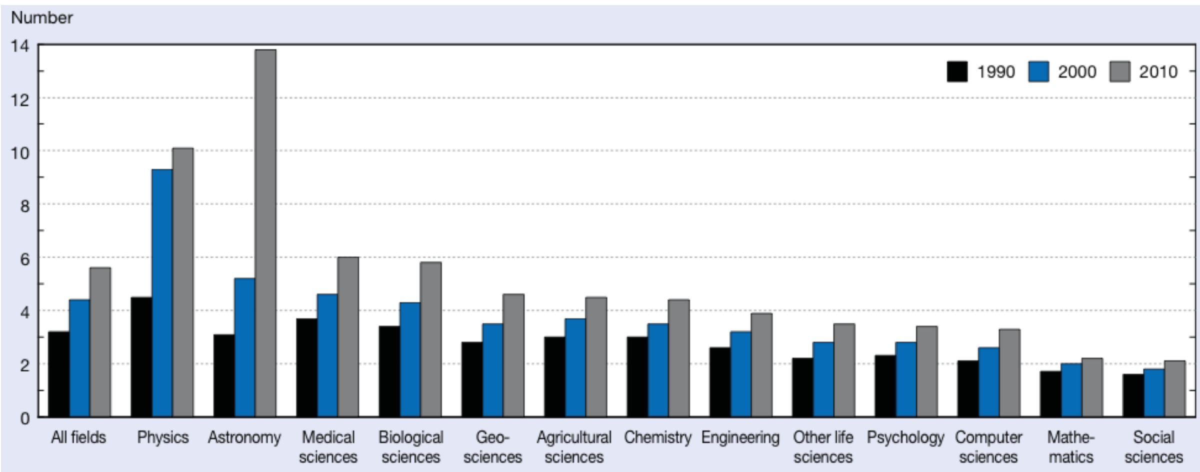


Figure 11: Growth in coauthorship by discipline; US paper 1990–2010 (source: NSF, 2012)



2.7. Authors and readers

The global number of active researchers varies by definition used but is estimated to be between 6.5 and 8.5 million (see *Numbers of researchers*).

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).

Later work from Tenopir & King suggested that about 15 per cent to 20 per cent of scientists in the United States had authored a refereed article. This estimate – and the asymmetry between authors and readers – is corroborated by work from Mabe and Amin who estimated that, of the 5–6 million global researchers then calculated by UNESCO, only around 1 million (circa 18 per cent) were unique repeat authors, while some 2.5 million authors published 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 around 10–15 million.

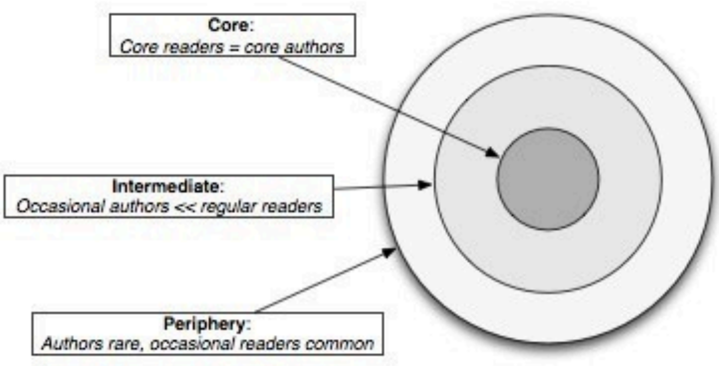
These overlapping author and reader communities can be illustrated as in Figure 12. The degree of overlap between authors and readers will vary considerably between disciplines: in a narrow pure science field like theoretical physics there may be close to 100% overlap, but in a practitioner field such as nursing or medicine the readers will be many times more numerous than the authors.

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 annual downloads of full text articles from publishers' sites are about 2.5 billion (according to an informal STM survey) with perhaps another 400 million downloads from other sites such as repositories. 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, 2009b). A 2005 study showed that articles in the society journal *Pediatrics* were read on average 14,500 times (King et al., 2006).

More recently, the PEER usage study (CIBER Research, 2012a) found that over a six-month period almost every single article (99%) in the study was downloaded at least once from the relevant publisher website, and so was a very large majority, 74%, from a PEER repository. As the authors put it, "the scholarly literature is under heavy scrutiny".

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, 2000; Ware & Monkman, 2008).

Figure 12: 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 8.7 million R&D workers (UNESCO, projected)



2.8. Publishers

There are estimated to be of the order of 5000–10,000 journal publishers globally: the Scopus database covers 18,500 journals from over 5000 publishers, and the long tail making up the remaining 10,000 or so peer-reviewed journals is likely to consist of publishers with just the one journal.

According to Morris (2006), the main English-language trade and professional associations for journal publishers as of 2006 collectively included 657 publishers producing around 11,550 journals, about 50% of the then total journal output by title. Of these, 477 publishers (73%) and 2334 journals (20%) were not-for-profit. 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 5 publish nearly 35% of journals, while three publishers (Elsevier, Springer, and Wiley-Blackwell) have well over 2000 journals each. Among 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).

2.9. 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 it 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. Comparing findings between 1993 and 2005, Mulligan & Mabe, 2011 found little change in researchers' attitudes to peer review: it remained highly valued, and a large proportion continued to be willing to commit to reviewing.

There are a number of arguments in favour of peer review. It could be seen quality assurance process for improving the quality of research studies (as distinct from improving the submitted manuscript prior to publication). Although some see this as one of its purposes, this sets a very high bar for peer review and at present there is little evidence to show its effectiveness in this way (e.g., Jefferson, Rudin, Brodney Folse, & Davidoff, 2007).

On the other hand, 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. Mulligan and Mabe report similar findings, though this belief varied a little by research discipline.

Peer review also acts as a filter, to the benefit of readers. For professional researchers, the most important aspect of this filtering is not just the fact that peer review has taken place, but the basis it provides for the stratification of journals by perceived quality: peer review is the process that routes better articles to better and/or most appropriate journals. While there is an active debate over whether this is the most effective way to filter the literature, it remains for now an important signal for authors (Tenopir, 2010; Ware, 2011).

Peer review can also act as a seal of approval, for instance distinguishing credible peer-reviewed science from non-peer-reviewed materials. This is probably more important for lay readers and journalists than for working researchers.

Critiques of peer review

Peer review is certainly not without its critics. The main criticisms are that it is ineffective; unreliable; poor at detecting errors; offering too much scope for bias, particular in single-blind form; providing scope for reviewer misconduct; and that it is slow, delaying publication unnecessarily (see Ware, 2011). Remedies include open peer review, which it is argued (see below) can both improve the fairness and the quality of review; cascade review, which aims to reduce inefficiency and speed up publication; and post-publication review which, in its most radical form (the "publish then filter" model), could speed up publication by conducting the review after the article has been published.

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, 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. Despite this caution, open review in one form appears to be slowly growing, namely the publication of reviews alongside the published paper; notable examples include the *BMJ*, Biomed Central medical journals, the European Geophysical Union journals; and EMBO, which has made a strong case of the benefits of open review in its journals (Pulverer, 2010).

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. There is, however, some interest in aggregating multiple "signals" of an article's potential impact, including the number of post-publication comments, as a complement to the Impact Factor (see *Article-level metrics and altmetrics*).

A procedural variant on these approaches is cascade peer review. This seeks to avoid the necessity of repeated peer reviewing each time a paper is rejected and resubmitted to another journal, by forwarding (with the author's consent) the article and its accompanying review reports to the new journal. This approach was pioneered by open access publisher BioMedCentral and later became seen as characteristic of the *PLOS ONE*-type megajournal (although it was never a very substantial fraction of *PLOS ONE* submissions). More ambitiously, the journals in the Neuroscience Peer Review Consortium[17] agree to accept manuscript reviews from other members of the consortium, although the journals are with different publishers.

[17] http://nprc.incf.org

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 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 et al., 2009). These figures are full costings, including estimates for the time spent by the academics conducting the review. The publisher's average cost of managing peer review (salaries and fees only, excluding overheads, infrastructure, systems etc.) was reported by the PEER study at $250 per manuscript (Wallace, 2012).

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 appoint and support 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 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%).

A study from Thomson Reuters analysing the data held by its ScholarOne online submission system (covering 4,200 journals from over 365 publishers in 2012) reported 1 million manuscript submissions in 2010 compared to 317,000 in 2005. (These data are not normalised for the increasing numbers of journals using the system; the report does not specify if submissions includes re-submissions but this seems likely.) The average acceptance rate fell slightly over the same period from 41% to 37%. Time from submission to final decision reduced from 65 days on average in 2005 to 59 days in 2010, while time to first decision stayed about constant at 40–41 days (Thomson Reuters, 2012).

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 instead into overhead costs (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

- Production efficiencies: systems can undertake automatic "pre-flight" testing, for instance checking image resolution at the submission stage
- 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 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*).

2.10. Reading patterns

The number of articles that university faculty members report reading per year has steadily increased over time, as illustrated in Figure 13 (Tenopir, 2007; Tenopir et al., 2009). 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*). A more recent UK study by Tenopir reported an average 39 scholarly readings per month, comprising 22 articles, seven books, and ten other publications (Tenopir et al., 2012), amounting to an estimated 448 hours per year spent reading (equivalent to 56 8-hour days).

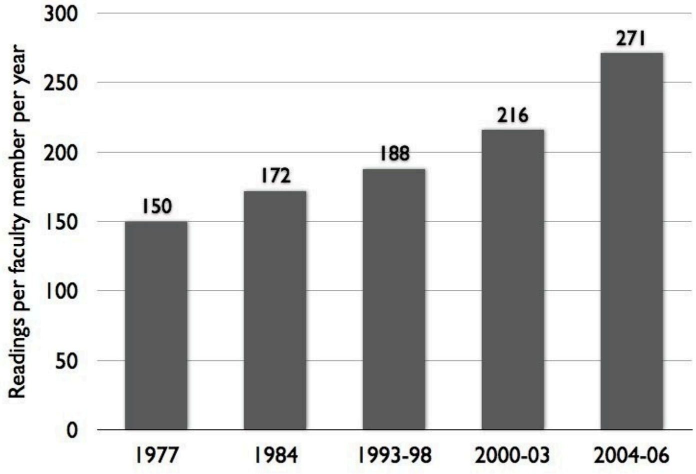
The breadth of reading has also increased over time: in 1977 scientists at Drexel read from an average 13 journals per year, while the figure is now over twice that.

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).

One plausible explanation is given by RIN-funded work done by the CIBER research group (Nicholas & Clark, 2012). Using analysis of publishers' log files, they demonstrate that few users of scholarly websites spend any significant time reading in the digital environment. Session times are short, and only 1–3 pages are viewed, and half of visitors never come back. Researchers reported that only 40% said they had read the whole paper of the last "important" article they had read. Users will download articles for future reading or reference, but in follow-up interviews researchers reported that at least half the articles downloaded were never read (and this is likely to be an optimistic estimate). The CIBER authors argue that researchers in the digital environment have moved from vertical to horizontal information seeking and reading, that is, moving quickly over the surface from article to article ("bouncing, flicking, or skittering") rather than reading deeply. While the authors point to factors in the modern environment that encourage this behaviour (over-supply of articles; lack of discretionary time and more pressured workplaces; multitasking becoming the norm; social media conditioning us to accept fast information), they also suggest that researchers may always have read selectively and in snippets, and that the idea of in-depth scholarly reading as the norm was simply a myth.

Renear and Palmer (2009) discussed the strategies and technology innovations ("strategic reading") that help readers extract information from more papers while spending less time per paper. There is considerable focus on using technology in this way, including semantic web technologies (e.g. taxonomies and ontologies), text and data mining, and the use of new metrics. These are discussed below (see *New developments in scholarly communication*).

Figure 13: Average number of articles that university faculty members reported reading per year (source: Tenopir, 2007)



Access and navigation to articles

Academics use a wide range of methods to locate articles, as illustrated in Figure 14 and in more detail in the more recent data in Figure 15. The growing importance in an online world of searching and parallel reduced importance of browsing is evident in this data (and is reflected in publisher's web logs which typically record around 60% of all article referrals from one search engine, Google). Asking colleagues remained an important strategy albeit ranking behind browsing and searching.

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 2012 study (Inger & Gardner, 2012, updating earlier 2005 and 2008 reports) 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, remaining ahead of web search engines. The academic search engines (Google Scholar, Microsoft Academic Search) appear to have gained ground in 2012 over general search engines.
- Readers strongly valued the content alerting services on journal web sites (journal alerts were the most popular starting point for discovering latest articles) and valued journal homepages as a place to discover latest articles, 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's OPAC and web pages, having suffered initially from the growth of general purpose search engines retain importance as the starting point to navigation, particularly for searching by topic or following up citations. 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 and web scale discovery technologies 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.

The "Generation Y" study investigated the information-seeking behaviours of doctoral students born between 1982 and 1994 (JISC & British Library, 2012). E-journals dominated as the main research resource across all subject disciplines. Although they were described as sophisticated information-seekers and users of complex sources, they were more likely than older researchers to make do with the abstract if they could not retrieve the e-journal article. And while they were active users of new information technologies in general, they were skeptical of the latest web technologies in their research, using only if they could be easily absorbed into existing work practices, with social media lacking legitimacy.

Reading patterns are slowly changing in respect of where it takes place: a significant minority (22%) of respondents to a 2005 survey preferring to conduct their e-browsing from the comfort of home, with medical researchers had the highest response at 29% (Mabe & Mulligan, 2011).

Figure 14: Ways used by university faculty to locate articles (source: Tenopir, 2007)

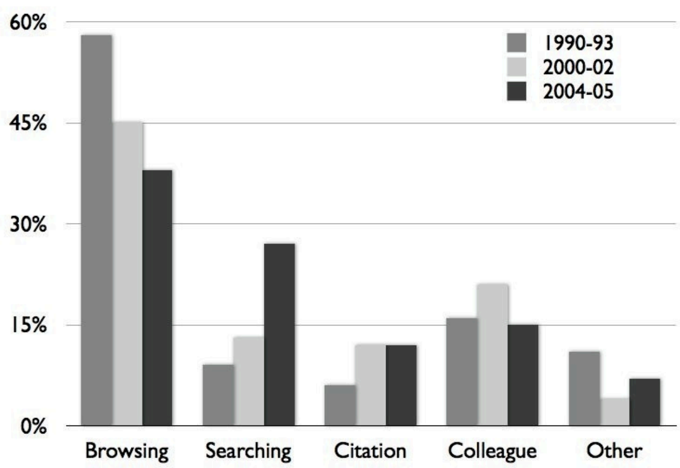
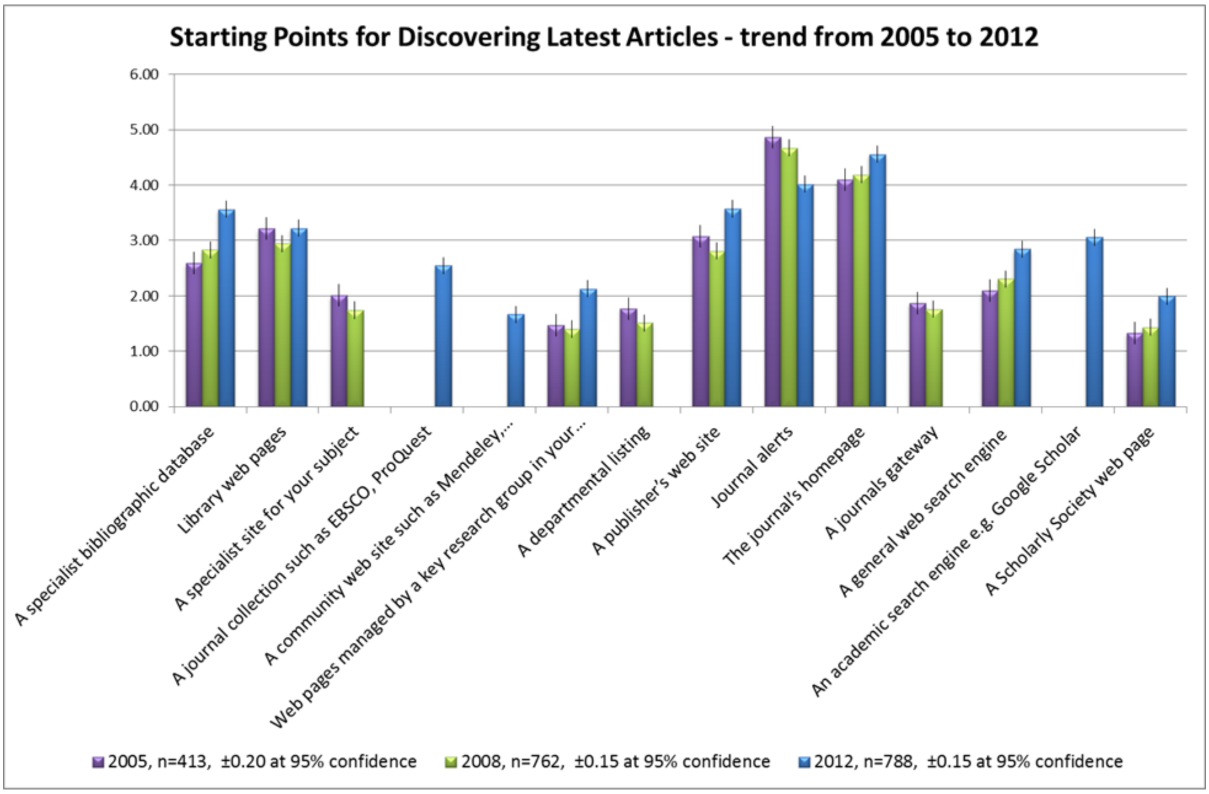


Figure 15: Starting points for discovering latest articles – trend from 2005 to 2012 (source: Inger & Gardner, 2012)



2.11. Disciplinary differences

It is worth noting that the average characteristics described above conceal some important differences between subject disciplines in their patterns of publishing, reading and using scholarly materials.

For example, while the average journal included in the Journal Citation Reports publishes about 120 articles per year,[18] science and technology titles are much larger at about 140 articles and social science and humanities much smaller 45 articles a year. 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 coauthor (with 85% of respondents saying that

[18] Strictly speaking, this refers to the number of "citable items", that is, scholarly works including – but not limited to – articles, reviews and proceedings papers. Data kindly supplied from the Journal Citation Reports® a Thomson Reuters product

75% or more of their output was coauthored), followed by physical sciences and engineering, then the social sciences, with arts and humanities the least likely to coauthor (with 76% saying that 25% or less was coauthored).

- 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 possible decline in the reading (and writing) of books in favour of journal articles, as suggested in the 2005 JISC report, was confirmed in a later RIN study, which found researchers expressing concern about this. It was unclear if it was due to library budget cuts reducing book availability, the greater online availability of journals, or simply the lack of time, but bibliometric analysis confirmed a significant decline in the citation of books as distinct from journal articles and other forms of output (RIN, 2009a).

A fascinating set of case studies in information use, studying in depth how researchers in different disciplines – life sciences, humanities, and physical sciences – discovered, accessed, analysed, managed and disseminated information (RIN, 2009c; RIN, 2011d; RIN, 2012). The various findings are too rich and detailed to be summarised here but the studies repay attention and dispel any notion that there is a single "workflow" adopted by researchers, even within the same disciplines.

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 coauthorship 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 16). 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 17 illustrates differences in the ways readers find articles, with marked variance for instance in the importance of browsing.

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 in the UK between the disciplines in terms of access to resources and to journals in particular. A later RIN study confirmed this for academics (RIN, 2011a), though there were differences between subject areas for industry-based researchers (see *Researchers' access to journals*).
- All authors of whatever discipline claim that career advancement and peer-to- peer communication are the most important reasons for publishing.

Figure 16: Average articles read per university faculty member per year (Source: Tenopir, 2007)

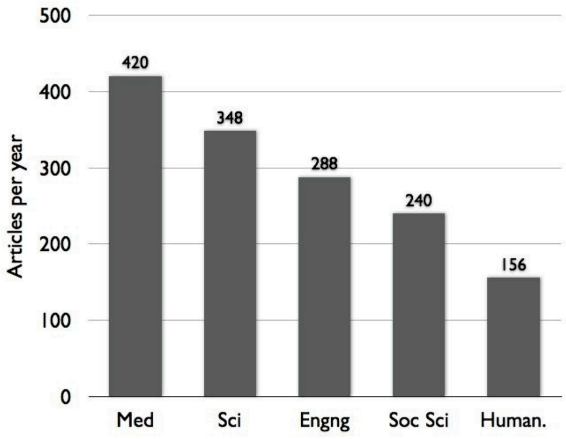
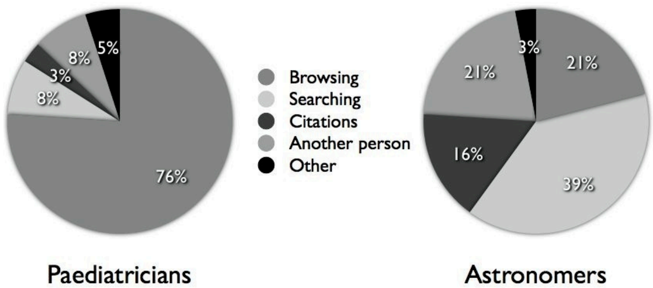


Figure 17: Subject differences in the ways articles are found (Source: Tenopir, 2007)



2.12. 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, 2012).

Electronic journals additionally allow “forward” reference linking, i.e. linking to later work that cites the paper in question, a feature also supported by indexing and discovery services.

The numbers of citations is increasing faster than publications. Comparing the five-year periods 1999/2003 and 2004/2008, the number of publications increased by 33%, while citations increased by 55%. Three factors in this are probably the growth of the literature (i.e, there is simply more to cite), the growth in coauthorship, and a recent trend towards longer reference lists (Elsevier, 2011).

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 2000 to 2010: over this period the United State’s and Japan’s shares declined, while China and other Asian countries’s shares increased.

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 citable items published during that same time period.)

The use of citations data (and in particular the journal-level impact factor) to judge the quality of individual researchers’ and departments’ research outputs, though widespread, is increasingly criticised. The assumption that articles published in the same journal are likely to be of similar quality is not borne out by the data: there is a skewed distribution with 15% of articles accounting for 50% of citations, and 90% of citations generated by 50% of articles (Seglen, 1992). The top half of articles in a journal can thus receive 9 times as many citations as the bottom half. Dissatisfaction with the impact factor is leading to the development of alternative metrics (see below), though for now it retains its primacy.

Average impact factors show considerable variation between subject fields, with the primary reason for variation being the average levels of coauthorship. Hence mathematics with coauthorship of 1.25 has an average Impact Factor of 0.5, while biology has coauthorship and Impact Factor both around 4. The fundamental and pure subject areas have tend to 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. Citation practices thus vary substantially between disciplines; it is possible to correct for this using field-weighted citations when comparing research performance across different fields, though this is frequently not done.

The average number of citations per articles has also been increasing over time; Figure 18 shows that the average for all countries has risen from about 1.7 in 1992 to 2.7 in 2010 (NSF, 2012).

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.)

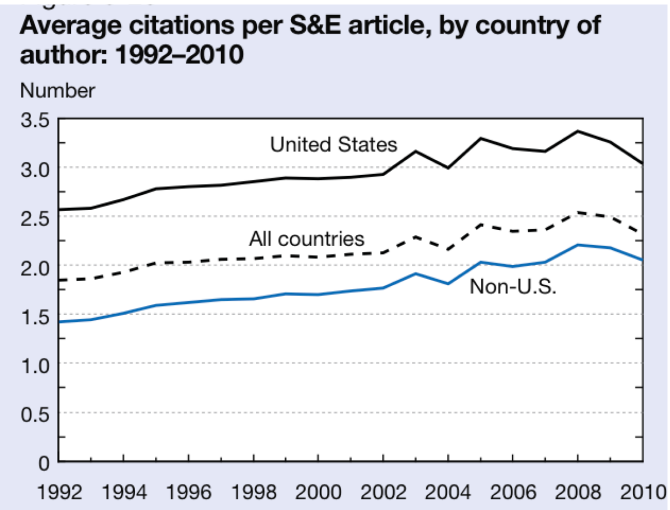
the range is +/-40%. Similarly, an impact factor of 1.50 for a journal publishing 140 articles is not significantly different from another journal of the same size with an impact factor of 1.24. It is thus foolish to penalise authors for publishing in journals with impact factors below a certain value, 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, 2000/2007.

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.

Table 2: Share of world citations of science and engineering articles, by citing year (Source: Science & Engineering Indicators 2012, NSF, 2012)

Region/country	*2000 (%)*	*2010 (%)*
United States	44.8	36.4
European Union	33.3	32.8
China	0.9	6.0
Japan	7.1	5.7
Asia-8	1.8	5.3

Figure 18: Citation inflation: increase in the average citations per article, by country of author (source: NSF, 2012)



Other bibliometric measures

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 compared 39 scientific impact measures (Bollen et al., 2009). Using statistical techniques to categorise the different measures on two dimensions roughly equivalent to prestige and to popularity, they 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". One should note, however, that usage and citation measure different things.

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.

Article-level metrics and altmetrics

This is the approach of the altmetrics movement. It starts from several dissatisfactions with the Impact Factor (or the way it is misused): the journal IF is used as a measure for the quality of an individual article, despite the criticism of this outlined above; second, that citations measure just one narrow aspect of impact; and third, citations (even if measured at the article level) are a slow, lagging indicator. To counter this, the "altmetrics" movement[19] proposes a range of additional metrics to complement metrics provided by citations and downloads to build a more rounded picture of impact (Priem, 2010). The altmetrics draw heavily on social media and tools and include data from Twitter mentions, blog posts, social bookmarking data (e.g. Connotea, Mendeley), as well as news media and article-level comments, annotations and ratings.

A number of tools have emerged to support the tracking, reporting and visualisation of altmetrics (see *http://altmetrics.org/tools/* for a current list).

There are some preliminary indications that social media activity may predict citations, though the evidence is not strong (e.g. Eysenbach, 2011). The main criticism of using social media mentions, as well as of article-level comments and ratings, as a measure of impact is that it is unclear what they are measuring beyond immediacy and popularity. Articles with eye-catching and unusual titles (particularly if they contain sexual terms) seem likely to be as strong candidates for high-volume bouncing around the internet echo chamber as work with genuine long-term impact.

[19] not to be confused with the Altmetrics project and app (http://altmetric.com), which is a tool developed by Digital Science to collect and present altmetric data on an article's webpage

Usage and the Journal Usage Factor

Total global downloads of articles from publishers' sites have been estimated at between 1.1 billion in 2010 (as shown in Table 3) and 2.5 billion (according to an informal STM survey), with perhaps another 400 million from other sites such as repositories.

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). This would be particularly be the case for clinical medical journals, or other journals with a large practitioner readership.

The UK Serials Group commissioned work to investigate whether it might be feasible to develop a "Usage Factor" based on download statistics. 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. UKSG and COUNTER then commissioned CIBER to conduct more detailed investigations which were published in 2011 (CIBER Research Ltd, 2011), following which UKSG/COUNTER issued in March 2012 "The COUNTER Code of Practice for Usage Factors: Draft Release 1" (COUNTER, 2012).

The draft Code defines the publication and usage period as two concurrent years: that is, the usage factor for 2009/2010 will be based on 2009/2010 usage data for articles published in 2009/2010. The Usage Factor: Journals (UFJ) will be the "median value of a set of ordered full-text article usage data"; the median is proposed rather than the mean because the data is highly skewed, with most items having low use, and a few used many times. It will be reported annually as an integer (greater precision is deprecated because the level of variation means there is a lot of statistical noise). It will integrate articles-in-press from the accepted manuscript stage, and will incorporate usage from multiple platforms, reflecting the heterogenous sources of article usage.

Patterns of usage were found by CIBER to vary considerably between different document types and versions. Consequently it is proposed there should be two versions of the UFJ: one based on usage to all paper types except editorial board lists, subscription information, and permission details, and a second based on scholarly content only (short communications, full research articles, review articles).

CIBER found that there was little correlation between the proposed UFJ and citation-based measures such as the Impact Factor. This was not surprising as they measure different things (reflecting reader and author choices respectively). Highly cited papers do tend to be highly downloaded, but the reverse is not necessarily true, particularly in fields with high proportions of practitioners. Citations and downloads have different profiles over time: most downloads occur in a peak a few months wide immediately following publication, while citations build over a longer period of 2–3 years.

The consensus view emerging seems to be that downloads (as a proxy for readings) is a potentially useful complement to citation data but that it should not be seen to replace it, because they reflect different aspects of "using" a research paper. Download and reading papers is more important during the early stages of research design and of article writing, while citing tends to occur more towards the end of the process.

Table 3: Article downloads by country, 2010 (source: Elsevier, 2011)

Country	Article downloads (millions)	Proportion of global total (%)
Global total	1065	100.0
USA	327	30.7
China	105	9.9
UK	100	9.4
Germany	70	6.6
Japan	62	5.8

2.13. 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 provided one of the more reliable estimates of journal costs. CEPA have subsequently updated their estimates for a later report (RIN, 2011c), giving the average 2010 journal article cost of production (print + electronic) at £3095. This was 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): £1261
- variable costs (printing, paper, distribution): £581
- indirect costs (staff and overheads): £666
- surplus: £586

Note that RIN included 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 (Campbell & Wates, 2009).

The PEER project reported the average cost of managing peer review at $250 per submitted manuscript and the average production cost at $170–400 per accepted manuscript (in each case the figures refer to salary and fees only, excluding overheads, infrastructure, systems etc.) (Wallace, 2012).

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 2008 RIN report quoted a narrower range. For low rejection rate journals the RIN authors gave a figure of £1670, with high rejection rate journals at £4091. RIN's figure for popular hybrid journals (*Science*, *Nature*, etc.) was £4116, though other estimates have placed it at $10,000 or even higher.

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; Thorn, Morris, & Fraser, 2009). RIN estimated the average profit/surplus in 2008 at 18% of revenues, equivalent to £517 per paper (these figures were not updated for the 2011 report), 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 system-wide cost savings that would arise overall if 90% of journals were to switch to e-only publishing at £1.08 billion, 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 figure 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.

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 increase in serials expenditures between 1986 and 2011 was 6.7%, while the US Consumer Prices Index rose by an annualised 2.9% over the same period (ARL, 2011).

The reasons for historic journal price increases have been varied and include (adapted from King & Alvarado-Albertorio, 2008): growth in article output leading to increased numbers of articles per journal, which with a parallel increase in average article length led to larger journals; reduction in page and colour charges; the "new journal" effect (growth of scholarship leads to the burgeoning of new fields, which in turn leads to 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); increased special requirements and features; conversion of back issues to electronic format; publishers increasing prices to compensate for falling subscription numbers and currency effects; and, of course, cost inflation (especially salary and paper costs), which has annualised at about 3% per annum for the last twenty or more years.

In summary then, the observed annual average journal price inflation during the 1990s and 2000s has a number of components, of which organic growth in the literature (3%) and cost inflation (3%) were the most important, followed by electronic delivery and conversion costs, new journal specialisation and attrition (price spiral) and currency fluctuation effects (~1%).

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. For instance, between 2004 and 2008, total UK university spending rose in real terms by 22% while library spending on "information content" rose by 15% (RIN, 2011b). In the US, the proportion of university funds devoted to libraries fell in 2009 for the 14th year in succession, dropping below 2% for the first time (ARL, 2011). This is partly attributable to efficiency gains (e.g. bundled and consortium-based purchasing, other shared services, outsourcing of cataloging and reference services, and staff reductions) but also reflects the failure of libraries to make their case for sustaining their share of a growing total budget.

Effect of bundling and consortia licensing on prices

Statistics using publishing subscription prices have become increasingly misleading, however, because these figures do not represent what libraries have actually paid, due to 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 partly for this reason.)

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. 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, 2006, p.133). This fall has continued, with an average price per download in UK academic institutions falling in real terms from £1.19 in 2004 to £0.70 in 2008, a reduction of 41% (RIN, 2011b).

2.14. Authors' behaviour, perceptions and attitudes

There have been numerous studies of author behaviour, perception and attitudes but two pioneering 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, & Huntingdon, 2004; Rowlands & Nicholas, 2005), and a survey commissioned by Elsevier in collaboration with CIBER and NOP in 2005 (Mabe, 2006; Mabe & Mulligan, 2011). More recently studies by RIN and Harley have extended and amplified these findings (RIN, 2009a; Harley et al., 2010).

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. Their findings in respect of open access have to some extent now been overtaken by events (for instance, a majority of authors believed that mass migration to open access would undermine scholarly publishing, yet this is now government policy in the UK at least – see *Open access*), but some points remain current:

- 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.
- CIBER found that authors had little knowledge of institutional repositories and there was also evidence that a significant minority (38%) were unwilling to use IRs. With the exception of a few special cases, this remains true today (e.g. see Wallace, 2012).

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 that remain current 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.
- The importance of peer review was again underlined. [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%).
- There was high demand for articles published more than 10 years earlier [that is, prior to the introduction of electronic journals].

A similar picture in relation to what authors want from journals was given by an analysis of 10 years' worth of data from Elsevier's Author Feedback Programme (Mabe & Mulligan, 2011). Comparing data for 2002 and 2009 (incorporating responses from nearly 100,000 researchers) showed very stable attitudes, with factors reflecting the quality, the relevance and speed of publication remaining the most important, and ranked in identical order. The data also showed support for peer review and similar levels of reviewing across the decade.

This sense of continuity and preference for existing approaches and tools was illustrated in a RIN study into researchers' use of and attitudes towards Web 2.0 and social media (RIN, 2010).

A major UC Berkeley study (Harley et al., 2010) similarly found researchers remaining focussed on conventional formal publication, and very cautious about new models of web-based scholarly communication. Researchers used a range of communication methods at different stages of the research cycle, and these varied from discipline to discipline with biology standing out as having the narrowest range of types of outlet (i.e. primarily research journals). They found "no evidence to suggest that "tech-savvy" young graduate students, postdoctoral scholars, or assistant professors are bucking traditional publishing practices" and that "once initiated into the profession, newer scholars—be they graduate students, postdoctoral scholars, or assistant professors—adopt the behaviors, norms, and recommendations of their mentors in order to advance their careers". In fact it was established researchers that could afford to be more experimental. (An earlier Californian study reported similar findings, with senior faculty more open to innovation than younger, more willing and experiment and to participate in new initiatives, and also found more appetite for change in arts and humanities than in other disciplines (University of California, 2007).) The Harley study did though identify topics where attention was required, including: re-examination of the methods and timing of peer review; new models of publication able to accommodate varied lengths, rich media and embedded data links; and support for managing and preserving new digital research methods and outputs (e.g. components of natural language processing, visualisation, complex distributed databases, and GIS, etc.).

2.15. 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.

There has been concern in recent years at the fast-growing number of retractions, which have increased from about 30 a year in the early 2000s to more than 400 in 2011, despite a rise of only 44% in papers over the period (Van Noorden, 2011). Even so, it only represents perhaps 0.02% of papers, though in surveys, around 1–2% of scientists admit to having fabricated, falsified or modified data or results at least once. It seems probable that the increase in published retractions is positive, coming from an increased awareness of the issues and better means of detection rather than an increase in misconduct itself. One problem with retractions is the tendency for authors to continue citing the withdrawn paper; adoption of the CrossMark initiative should help curb this, or at any rate alert readers who follow the citations.

Committee on Publication Ethics

The Committee on Publication Ethics (COPE)[20] 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 7000 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. All COPE members are expected to follow its Code of Conduct and Best Practice Guidelines for Journal Editors, of which the most recent revision was published in 2011 (COPE, 2011).

[20] http://publicationethics.org/

Other organisations with an interest in publishing ethics

The International Committee of Medical Journal Editors (ICMJE)[21] provides detailed guidance on ethical matters relating to medical publishing (many of which are equally applicable to other areas), including authorship and contributorship, editorship, peer review, conflicts of interest, privacy and confidentiality, and protection of human subjects and animals in research.

The World Association of Medical Editors (WAME)[22] also addresses ethical issues, and has published a policy statement on conflict of interest in peer-reviewed medical journals (WAME, 2009).

The Retraction Watch blog writes regularly on article retractions and the issues raised. Its authors have proposed journals adopt a Transparency Index which would specify things like the journals peer review policy, whether it used plagiarism detection software, its mechanism for dealing with allegations of errors or misconduct, and whether its corrections and retractions conformed to ICMJE and COPE guidelines (Marcus & Oransky, 2012).

CrossCheck

CrossCheck[23] 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 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.

Other tools for detecting misconduct include screening with image-editing software and data review (digit preference analysis can detect fabricated data).

2.16. 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.

[21] http://www.icmje.org

[22] http://www.wame.org

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

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.

The most recent review of copyright in the UK and the EU (the Hargreaves report and subsequent government consultations[24], and Copyright in the Information Society[25] programme respectively), covered the topics raised by the digital environment that are relevant under any regime:

- *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 under review 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
- *Orphan work*s 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 key recommendations made by Hargreaves relevant to publishers are (Hargreaves, 2011):

- the proposed establishment of a digital copyright exchange to make it easier to get copyright clearance. It was seen as part of the solution to the problem of "orphan works" (ones where the original copyright holders cannot be traced); the UK BIS Select Committee recommended in addition a separate registry for orphan works
- an exception (to copyright protection) for format-shifting; publishers have some concerns about the potential for sharing of ebooks, though it is favoured by academic libraries and the British Library to facilitate long-term preservation and for the fair dealing/research copying exception
- an exception to allow text and data mining (TDM). This is an active area of development in STM; the Select Committee preferred to see publishers developing usable and affordable licensing schemes (see below, *Text mining and data mining*)
- that copyright exceptions could not be overridden by contract.

[24] http://www.ipo.gov.uk/ipreview.htm

[25] http://ec.europa.eu/internal_market/copyright/copyright-infso/index_en.htm

In the US too there is an active debate on the need for copyright reform. The Register of Copyrights intends to bring forward legislation aimed at dealing with orphan works, to update the fair use rules in relation to library uses, and to enable mass digitisation of commercially unavailable works (Samuelson, 2012). The last point arises following the failure of the Google Book settlement, and relates to a new rights category: commercially unavailable or "out-of-commerce" works are in copyright but not available for sale (roughly the digital equivalent of out of print). France has passed legislation allowing its national library to digitise such out-of-commerce works.

Perceptions and understanding of copyright

It is worth noting that much of the debate about copyright in STM sector takes place within a context of widespread ignorance and misunderstanding of copyright and the rights available under the current regime. For example, 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-archive the published version, which very few publishers permit. The study concludes that publishers had failed to communicate their copyright policies effectively.

This picture, of copyright and associated use and re-use rights being little- or mis-understood, recurs in other studies of academics, and even with librarians. For example, a RIN study on access gaps identified confusion about licensing and particularly walk-in rights, especially for e-resources (RIN, 2011a), and lack of knowledge about copyright has been cited as one of the reasons for author hesitancy in depositing in archives.

Text and data mining rights

Text and data mining (TDM) has been identified as an important and growing way of using STM content. It is discussed in more detail under *New developments in scholarly communication* but deserves an entry within this Copyright section because the rights issues are under active debate and in flux.

At the time of writing, it was uncommon for STM journal licences to permit TDM without further consent of the publisher, and most publishers (other than open access publishers) did not have publicly available policies, but dealt with each request on a case-by-case basis (Smit & van der Graaf, 2011). The requirement to contact each publisher individually would create an onerous burden for a researcher that wanted to mine a substantial fraction of the literature.[26] In response to these issues, the Hargreaves report proposed a copyright exception for TDM; even if enacted in the UK, however, this would not solve most researchers' problems because the issue is global and because it would deal only with the right to mine content already licensed, whereas a more general problem is mining both licensed and unlicensed content. A better way forward could be a comprehensive licensing regime. A small but important step was the model licence terms to cover TDM recently agreed by STM, ALPSP and P-D-R (Pharmaceutical Documentation Ring).[27]

[26] There are, for example, 587 publishers with more than 1000 papers published in PubMed since 2000, clearly an infeasible number for most people to negotiate with

[27] http:// is.gd/UXnRMl

The issue is also bound up with open access. For example, the UK Research Councils 2012 access policy requires authors' copies of articles deposited in archives to permit TDM, and for open access articles published in journals to similarly permit TDM in order to be compliant with the policy. Some publishers (e.g. Wiley-Blackwell, Springer) have changed the licensing of their OA articles from the CC-BY-NC to CC-BY in response to these and other pressures.

Machine readable and embedded licences

One potential solution to the problems of orphan works and of misunderstandings over what rights were available to users of digital content could be to embed the licence in a machine readable format within the resource itself. This already occurs to some extent with certain types of media file, notably music and videos for online sale. In these arenas it is often associated with digital rights management (DRM) arrangements, but this is not necessary: the licences can simply assert ownership and specify allowed downstream uses and licensing requirements.

The Linked Content Coalition[28] is working to develop a "framework for a fully interoperable and fully connected standards-based communications infrastructure". This would include the whole supply chain as well as the end user, and potentially involves all media types, not just STM.

2.17. 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 fully 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 journals are 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 perceived 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.[29] Its digital archiving services are available to publishers worldwide and are used by many major publishers including Elsevier, Springer, Wiley

[28] http://www.linkedcontentcoalition.org/

[29] http://www.kb.nl/hrd/dd/index-en.html

Blackwell, Taylor & Francis, OUP, and Sage. The e-Depot also offers archiving services to repositories in the Netherlands.

- LOCKSS (Lots of Copies Keeps Stuff Safe).[30] 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, based at Stanford, allows libraries to collect and store local copies of subscribed content under a special licence (more than 500 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 collaborative organisation of scholarly publishers and research libraries using the LOCKSS technology, covering about 7400 journals from 35 publishers.
- Portico is a not-for-profit preservation service for scholarly content[31], initially as a JSTOR project before spinning out as an independent organisation. It offers a permanent managed archive of ejournal and ebook (and other digital) collections, with libraries benefiting from protection against loss of access caused by defined trigger events (e.g. the titles being no longer available from the publisher or other source) It also offers a facility for post-cancellation access. As of mid-2012 it had about 750 participating libraries, 150 publishers covering over 14,000 journals and 156,000 ebooks, representing a total of 338 million files.
- The Alliance for Permanent Access[32] (APA) aims to develop a shared vision and framework for a sustainable organisational infrastructure for permanent access to scientific information, pursued through information exchange, collaborations and specific projects. A related organisation, APARSEN (APA Records of Science in Europe) aims to build a long-lived Virtual Centre of Digital Preservation Excellence.

2.18. TRANSFER code

The *UKSG Transfer Code of Practice*[33] 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 the Code was at Version 2.0, and was endorsed by some 36 publishers, including all the large journal publishers.

In addition to maintaining the Code, the Transfer working group also maintains an alerting service (including a notifications database, forms and list), and provides informal advice.

2.19. 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

[30] http://www.lockss.org

[31] http://www.portico.org

[32] http://www.alliancepermanentaccess.eu

[33] http://www.uksg.org/transfer

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. For example, the ARL statistics (ARL, 2011) 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 68,375 in 2011 (not all these will be peer-reviewed journals), while at the same time the unit cost of serials fell steadily from a peak in in 2000. Similarly, the number of current serials subscriptions per higher education institution in the UK more than doubled in the 10 years to 2004/05, from 2900 to 7200 (Creaser et al., 2006). SCONUL figures show a similar growth in UK access and statistics for Australia show a similar pattern.

The two *E-journals: their use, value and impact* reports from the Research Information Network (RIN, 2009b, 2011b) 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, and continued to rise at over 20% annually to 2008, while the cost of access fell to about £0.70 per article by 2008 (£0.65 at the most research-intensive institutions). The studies 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.

Current levels of access

Assessing the current level of access to scholarly journals is a key question for governments and other policy makers, and yet the studies on this made to date all suffer from methodological weaknesses to a greater or lesser extent (Meadows, Campbell, & Webster, 2012). This was particularly the case for the results of the recent consultations made by government bodies (OSTP 2012; European Commission, 2012a, 2012b); to be fair, these were explicitly consultation exercises rather than market research studies, but the dangers arise if the results are taken as being representative or generalisable.

These methodological differences and weaknesses thus make different surveys difficult to compare and interpret. The largest recent survey was conducted by CIBER in late 2011 on behalf of RIN (RIN, 2011a), based on 2645 responses to 20,000 invitations (13.2%). The survey confirmed again the central importance of journal articles (and to a lesser extent, conference papers). In universities and colleges, 93% said research papers were easy or fairly easy to access, and 72% said that access had improved over the last five years. This finding was in line with earlier surveys using similar methodology and appears to suggests on the face of it little problem in the way of access.

Similarly, a survey conducted by Outsell for the Australian Go8 Library group (Group of Eight & Outsell, 2010) analysed 1,175 responses (8.5%) from a population of 13,807 Australian researchers. It found 91% of respondents said that access to information resources met their needs very well or adequately.

And yet when respondents in the CIBER survey were asked for which of a range of resources they would most like to see access improved, a large majority (39% in the case of universities and colleges) identified journal articles as their first choice.

And in the European Commission survey, where the majority of respondents were librarians, almost 84% disagreed or disagreed strongly with the statement, "There is no access problem to scientific publications in Europe". Respondents to the OSTP consultation also argued for stronger government mandates and centralised repositories to improve access.

How to reconcile these positions? To start with, the RIN authors observe that "easy" access to most of the literature is not enough for many researchers. Although levels of access in universities were typically good overall, there were areas where access was less easy, notably in industry and for other groups such as independent professionals without access to academic libraries (Ware, 2009).

More generally, what would have been exceptional in the past may no longer meet current needs. Meadows speculates that because researchers know that almost all journal articles are digitally available, they are frustrated and express dissatisfaction when they are unable to access particular resources. Another factor may be the increased visibility and ease of finding of research articles through search engines, and the increased use of these to find scholarly content.

As the Finch Report noted (Finch Working Group, 2012), most researchers in academia and in large research-intensive companies have access to a larger number of journals than ever before, but they want more:

> *"online access free at the point of use to all the nearly two million articles that are produced each year, as well as the publications produced in the past; and the ability to use the latest tools and services to analyse, organise and manipulate the content they find, so that they can work more effectively in their search for new knowledge."*

Barriers to access

Barriers to access are an important issue: the RIN survey findings suggested "that information barriers can lead to significant non-productive activity and lost opportunities on the part of researchers and knowledge workers". Similarly the Finch Report saw improved access as promoting enhanced transparency, openness and accountability, and public engagement; closer linkages between research and innovation; economic growth; improved efficiency in the research process; and increased returns on the investments made in research.

The most commonly cited barriers to access in all the surveys and consultations discussed above were cost barriers and pricing: the high price of journal subscriptions and shrinking library budgets were cited by 85% or more of respondents in both the EC and OSTP consultations. The RIN survey also found that the most common barrier was when researchers had to pay to access content: the majority of respondents for whom access to journals was important felt they did not have enough access through existing arrangements. As well as high subscription prices, the RIN respondents also felt that prices charged for individual articles were too high.

While cost barriers were the most important, they were not the only one identified in these (and earlier) surveys. Other barriers cited include: lack of awareness of available resources; a burdensome purchasing procedure; VAT on digital publications; format and IT problems (including digital rights management issues); lack of membership of a library with access to content; and conflict between the author's or publisher's rights and the desired use of the content.

In this context the Finch Group was set up in the UK with a brief to examine ways to expand access. Its recommendations are primarily focussed on moving to open access in the longer term (see *Open access*) but its recommendations included several measures intended to broaden access in the short term during the transition to open access: increased funding for national licences to extend and rationalise cover; walk-in access to the majority of journals to be provided in public libraries; the development of licences for sectors such as central and local government, the voluntary sector, and businesses.

SMEs

Public policy interest in access to the scientific literature by small and medium-sized enterprises (SMEs) has grown. SMEs have been seen as a 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. The report found, however, that 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. Some of these approaches were pursued by the Finch Group, although it also noted that the fraction of SMEs that undertake R&D is very small.

Access in developing countries

In various 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 – perhaps unsurprisingly – worse in the rest of the world.

There are a number of schemes providing free or heavily discounted access to the scientific literature to researchers in developing countries.

The Research4Life programmes[34] are collaborations between UN agencies, STM publishers, universities and university libraries, philanthropic foundations and technology partners. There are currently three programmes:

- HINARI, launched in January 2002 in conjunction with the World Health Organisation, offers free or 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 8100 journals from 145 publishers, 4700 public institutions in 105 eligible countries registered for access.

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

- 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 over 3000 journals from 70 publishers in the fields of food, agriculture, environmental science and related social sciences. Over 2400 institutions have registered for access to AGORA.
- 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 3900 journals. Subjects include environmental chemistry, economics, law and policy, and other environmental subjects such as botany, conservation biology, ecology and zoology. Over 2200 institutions have registered for access to OARE.
- ARDI (Access to Research for Development and Innovation)[35] was launched in partnership with the World Intellectual Property Organization in 2009 and joined Research4Life in 2011, aimed at promoting the integration of developing and least developed countries into the global knowledge economy

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[36] of nearly 700 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[37] 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)[38] 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. 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.[39]

There are also some concerns that providing free access to Western journals (or equivalently, offering waivers of open access fees) may have unintended consequences in undermining nascent indigenous publishing (e.g. Dickson, 2012). Many of these programmes monitor this effect carefully.

---

35 http://www.wipo.int/ardi/en/

36 http://highwire.stanford.edu/lists/devecon.dtl

37 http://www.inasp.info/file/b9ba352ea3b1516f931be58b66f912d9/perii-eipm.html

38 http://www.eifl.net

39 http://www.authoraid.info/

3. Open access

Open access refers to the making available of content (especially, though not exclusively, journal research articles) in online digital copies, free of charge, and increasingly free of most copyright and licensing restrictions, and free of technical or other barriers to access (such as digital rights management or requirements to register to access).[40]

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" using the NISO Versions preferred term (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").

The question of what reuse rights are included has also assumed greater importance since the last edition of this report, partly as a result of the growing importance of text and data mining, although this may be redundant since publishers are also working on enabling this in the absence of open access (see *Text and data mining*).

In terms of timing (the "*when*") there are three options: prior to (formal) 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" (supply-side) business model or sponsored model
- **Delayed open access**: Stage 3, but delayed; existing business model
- **Self-archiving** (the "Green" route): Stage 2, either immediate or delayed; no independent 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.

3.1. *Full open access ("Gold" OA)*

In 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.

[40] e.g. see http://www.earlham.edu/~peters/fos/overview.htm

- Hybrid (or optional) 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.

3.2. 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 hybrid journals both use this approach. Many full and hybrid OA journals also offer paid-for "institutional memberships", whereby members of the paying institution can pay reduced (or sometimes no) publication charges (Björk & Solomon, 2012b has a more detailed account of pricing approaches used in OA journals).

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 *Transition and sustainability issues*).

The hybrid 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, now offer hybrid schemes. Uptake by the market has, however, been small (~2% or so) and the model continues to be regarded with a degree of suspicion by some librarians and funders, with concerns over whether the hybrid open access fees will lead to lower subscription prices (the so-called "double-dipping" issue). In fairness, all of the publishers using this model have said they will take the effect of OA fees into account when setting subscription prices going forward; a diminution in pricing benefits all subscribers, though, not just the institution where an author has paid for their article to be made open access.

Not all open access journals use publication charges: about half the journals listed on the Directory of Open Access Journals do not list author fees. Instead these journals use 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. The fact that a numerical majority of DOAJ *journals* may not make publication charges is potentially misleading, however, but the majority of *articles* published in OA journals do make charges (e.g. Dallmeier-Tiessen et al., 2010). An example of a different approach is the high-profile initiative to establish open access across the field of high energy physics without author charges is discussed below (see *SCOAP3*).

Table 4 lists a selection of 2012 publication charges from major societies and commercial publishers. Fees for full and hybrid open access journals fall in the range $1300–5000, except for Hindawi with a median $600 charge, and with lower fees of $500–1000 typically charged for case reports, short communications and some areas without much research funding. 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 (e.g. based on length, or for rapid peer review).

The range of fees shown in Table 4 is reflected in the reported average APC paid by the research funder Wellcome Trust of $2365. The average APC of all open access articles appears to be lower, however, according to a study of 1,370 OA journals that published a total of 100,697 articles in 2010 (Björk & Solomon, 2012c). Björk and Solomon found an overall average fee $906, ranging by discipline from about $1100 in biomedicine, through $530 for technology and engineering, to $240 for arts and humanities. The distribution of APCs in this study was bimodal, with a higher peak at $1600–1800 and a lower one around $600–800, corresponding to different kinds of publishers. This suggests the market average APC seen here is likely to rise as open access expands, since the "professional" publishers have far greater capacity to expand than do the small-scale and part-time publishers typically responsible for the very low fees.

The move to very permissive re-use licences as a condition of APC payment by funders will also have an effect. Traditionally journal income has come from a variety of sources not just subscriptions. A significant element has been copyright fees for secondary uses (paid through the Collective Management Organisations like CCC and CLA) and commercial reuses or reprinting, especially in the case of pharmaceutical industry reprints that are used in drug promotion to physicians. The recent requirement of both the Wellcome Trust and RCUK that Gold fees will only be paid to publishers who use the extremely permissive Creative Commons CC-BY licences, essentially remove revenue from the Pharma Industry reprints and from copying elsewhere. One possibility is that APCs will vary according to the licence conditions, with those that remove revenue possibilities from publishers being set higher.

Nonetheless the fees reported here are mostly lower than the existing industry average full cost per article (based on e.g. RIN, 2008, 2011c).

In order not to exclude authors from low-income countries or those who lack the funds, most if not all full open access journals will waive charges for such authors. An allowance for the proportion of waived or absent author fees therefore needs to be made when setting APCs or in calculating market size from listed APCs.

Table 4: Publication charges for a selection of full and hybrid OA journals. Various discounts (society members, subscribing/"member" institutions, low-income countries, etc.) not shown. (Source: publisher websites, October 2012; £/$=1.6, €/$=1.3)

Journal/publisher	***Full/Hybrid OA***	***Charge (US$)***
American Institute of Physics	Full/Hybrid	1350–1800
American Physical Society	Full/Hybrid	1500–2700
BioMed Central	Full	630–2635 (median 2025)
Hindawi	Full	300–1750 (median 600)
BMJ Group	Full/Hybrid	1920–4000
Cambridge University Press (147 journals)	Hybrid (Full planned)	2700 (STM) 1350 (HSS)
Elsevier - *Cell Reports* - case reports, comms	Full/Hybrid	Mostly ~3000 5000 500–600
New Journal of Physics/IOP-DPG	Full	1440

Journal/publisher	Full/Hybrid OA	Charge (US$)
Oxford University Press	Hybrid Full	1615–3000 475–2770
PLOS - *PLOS ONE*	Full	2250–2900 1350
Royal Society (London)	Hybrid Full	2380 1932
Springer (see also BMC above)	Hybrid	3000
Wiley-Blackwell	Hybrid Full	3000 1450–3900

Hybrid content 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.

3.3. Delayed open access

Under this model, the journal makes its contents freely available after a period, typically 6–12, or in some cases 24 months. A growing number of journals (particularly in the life science and biomedical areas) have adopted delayed open access policies, for example, many of the (society-owned) life science journals on the HighWire platform.

The business model depends on the embargo period being long enough not to compromise subscription sales; this is discussed in more detail below (see *Transition and sustainability issues*).

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

3.4. Open access via self-archiving ("Green" OA)

The "green" route to open access is by self-archiving, which makes available a Stage 2 version of the article (the 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 *Transition and sustainability issues*).

The author (or more likely in practice, 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 PubMed Central in biomedicine).

The OpenDOAR website[41] divides repositories into the following categories (data as of October 2012):

- Institutional: 1824 (83%)
- Disciplinary: 235 (11%)
- Aggregating: 96 (5%)
- Governmental: 52 (3%)

The Registry of Open Access Repositories (ROAR) reported a total of 2955 repositories, of which 465 were in the US and 243 in the UK (as of October 2012).

The broad categorisation into institutional and subject repositories potentially conceals wide variations in scope, function and cost. For example treating arXiv as equivalent to PubMed Central is misleading. The latter has been described as "a proper electronic library": its functions including conversion of multiple input formats into structured XML, correction of the structural, content, and consistency errors that occur when converting text for digital preservation, and provision of the conversion process to print a "clear" PDF version of downloaded articles as required (Terry, 2005).

At present, the largest collections, highest visibility and most use seem to be squarely with the subject repositories (e.g.Björk et al., 2010; and see also The Web Ranking of Repositories, http://repositories.webometrics.info). Romary & Armbruster (2009) argue for the superiority of central (not necessarily subject) repositories: first, funder mandates are more effective than institutional in driving deposit and they are best served by single infrastructures and large repositories which enhance the value of the collection, and second, their analysis shows institutional repositories to be more cumbersome and less likely to achieve a high level of service than central repositories.

The numbers of repositories has increased substantially in recent years, with growth primarily from the institutional repository category. The PEER Baseline report gave the following reasons for the growth in the numbers of institutional repositories (Fry et al., 2009):

- opening up access to scholarly publications
- increased visibility (and possibly usage and citations)
- showcasing institutional research outputs
- the increasing availability of public funds (in the UK, via JISC; in Europe, via DRIVER project funding)
- an increasingly competitive educational sector.

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.

[41] www.opendoar.org

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;[42] it currently lists a total of 2955 archives of which 1885 are identified as institution or department level research repositories (up from 1300 in the 2009 edition of this report). The alternative OpenDOAR service[43] lists 2207 repositories of which 1824 are categorised as institutional. The growth rate is impressive but there would need to be around 7000–9000 IRs to give comprehensive coverage.

The PEER Baseline report gave the following reasons for the growth in the numbers of institutional repositories (Fry et al., 2009), and in addition to the two main objectives shown above identified the following factors:

- a need to showcase institutional research outputs
- the increasing availability of public funds to support development of IRs (in the UK, via JISC; in Europe, via DRIVER project funding)
- an increasingly competitive educational sector.

The numbers of articles deposited by authors in their IRs has grown much more slowly, and many IRs (except perhaps in the Netherlands) remain forlornly underused by depositing authors (e.g. see Salo, 2008; Albanese, 2009). (The total number of articles included in the 1885 repositories listed by Eprints is about 9.8 million, or a mean of 5220, but these totals include all types of record, including bibliographic records imported from other sources, and the distribution is skewed with a small number of large successful 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 (see Wallace, 2012, and *PEER project* below). Stevan Harnad estimates that there is an upper limit on what advocacy and persuasion can achieve in terms of the rate of voluntary deposit of e-prints of about 15% of eligible articles; 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 grey literature, research data and other institutional content, and those (such as

[42] roar.eprints.org/

[43] www.opendoar.org

the University of California's eScholarship repository) who see the role primarily in terms of scholarly communication and publishing (Albanese, 2009). The UK Finch Group saw the role of IRs as being more in the former category (Finch Working Group, 2012).

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[44] (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, albeit with less comprehensive coverage. It currently holds over 790,000 e-prints. (See below, *Recent developments in open access*, for a discussion of arXiv's funding model.)

RePEc (Research Papers in Economics)[45] 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 1.2 million research pieces from 1,500 journals and 3,300 working paper series. It differs from arXiv in several ways: first, it is decentralised (and volunteer-based) bibliographic database rather than a centralised repository, integrating content from some 1400 archives; second, 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, and facilitates a variety of specialised services built using its data.

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 has been supported by many publishers who have voluntarily deposited on behalf of their authors either the author's manuscript version (stage 2) or in some cases the full text (stage 3), 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 there were 2.5 million research articles hosted on PMC, of which 504,000 were in the open access subset (the others are freely available but not open access in the specialised sense used by PMC, i.e. to adopt a Creative Commons licence including permitting redistribution and reuse).

---

[44] http://www.arxiv.org

[45] http://repec.org

A UK version of PMC is hosted and managed by the British Library. From November 2012 this will become Europe PubMed Central.[46]

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 billion). Similar policies are now becoming widespread; the SHERPA Juliet website[47] listed (as of October 2009) 133 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[48] recorded 136 full institutional and 34 sub-institutional mandates in October 2012. High profile institutions adopting mandates include Harvard, MIT, UCL, ETH Zurich, Fraunhofer-Gesellschaft, and the University of California.

The early impact of mandates was muted: authors are generally not motivated to self-archive (e.g. see the discussion of the *PEER project* findings), and in the absence of monitoring and enforcement this activity tends to get given a low priority. This was particularly true for institutional mandates, but even the high-profile funder mandates have seen less than comprehensive compliance to date: for NIH it was about 75% and for Wellcome about 55% in mid-2012 (in both cases with significant assistance from the publishers themselves). The situation is changing, however, led by funders making compliance a higher priority: for example, the Wellcome Trust announced in June 2012 that it was tightening its open access policy, including sanctions on researchers that failed to comply.[49]

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;[50] of the 563 publishers included:

- 27% allow archiving of both author's original and accepted manuscript
- 32% allow archiving of accepted manuscript

[46] http://europepmc.org

[47] http://www.sherpa.ac.uk/juliet/

[48] http://www.eprints.org/openaccess/policysignup/

[49] http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm

[50] http://www.sherpa.ac.uk/romeo/

- 8% allow archiving of the author's original manuscript
- 33% do not formally support archiving.

Some 67% of publishers therefore permit archiving in some form. The proportion of journals will 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 the publisher's final online version. Publishers have also introduced 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. One of the original institutional repositories, 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 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) (Houghton et al., 2009). 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).

A 2007 survey of US institutional repositories (Rieh et al., 2007) found that the funding in almost all cases came from the library and that there was no additional budget provided (i.e. funds were taken from the routine library operating costs). Budget amounts were not given but breakdowns by type of expenditure were provided: ~37% for staff; vendor fees ~38%; hardware ~10%; software ~2.5%; software/hardware maintenance and backup ~12.5%.

More recently, the PEER project found it very difficult to obtain data on the set-up and running cost of institutional repositories, with investments in platform set-up, and costs in software upgrade and repository maintenance treated as sunk costs and not accounted for separately, and costs spread across multiple departments. The project was able to obtain estimates of the cost of technical staff support; it reported a cost per reference in the range €2–50, and cost per full-text article of €2.5–53.2. The wide range reflected the efficient scaling with size of holdings, i.e. the lower costs per item refer to the larger repositories, and *vice versa*.

The UK Repository Support Network (http://www.rsp.ac.uk) has used illustrative hardware costs of £2000–150,000 and suggests a £20k set-up should handle 50–100,000 papers. The site lists other areas of start-up and ongoing cost (primarily staff time) but gives no indication of the likely levels for these.

Large disciplinary repositories are naturally more expensive. Cornell University Library estimates the annual running costs for the arXiv at $826k. The National Institutes of Health has estimated that the cost of administering its self-archiving policy would be $4 million.[51] This is, however, a small fraction of the total cost of PMC, reflecting just the cost of collecting, processing and converting NIH-funded manuscripts to the PMC archival format.

[51] http://publicaccess.nih.gov/Collins_reply_to_Pitts121611.pdf

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 (fewer publishers permit archiving of the version of record). Most funder and institutional mandates require deposit of at least 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).

The larger institutional repositories are working with publishers to provide links from the archived version to the version of record. (The CrossMark service will be valuable here in distinguishing the version of record from other versions; see *Versions of articles* above.)

3.5. 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, and SCOAP3 could arguably fit this category).

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.[52]

A final "variant" might be mentioned, which is false open access. A number of surveys (e.g. 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.

3.6. System-wide perspectives

As policy-makers' interest in open access has grown there have been a number of attempts to study the economic impacts of open access, including the system-wide effects for scholarly communication, and (more controversially) the wider economic impacts.

As noted previously, a 2008 report (RIN, 2008) 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

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

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.

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.

The estimates were contested, primarily by publishers who argued that the analysis underestimated the efficiencies of the current subscription system and the levels of access enjoyed by UK researchers, and that many of the savings hypothesised would depend on the rest of the world adopting author-pays or self-archiving models. Many of the figures used in the Houghton model were inaccurate estimates rather than industry derived data.

In addition to the system savings, Houghton estimated increased economic returns to UK public-sector R&D arising from increased access might be worth around £170m. This figure is clearly more speculative, resting on hard-to-test assumptions about the levels of current access and the marginal rate of return to any increased access.

A 2011 UK study, *Heading for the Open Road* (RIN, 2011c), attempted to address the limitations of these two studies, and in particular looked at the issues arising from a dynamic transition from the current regime to various scenarios for increased access via open access (Gold and Green were modelled separately) and other routes (e.g. increase licensing, transactional access), rather than consider a static hypothetical economy in which close to 100% conversion to open access had already occurred. Its key conclusions were that the open access routes offered the greatest potential to policy-makers interested in promoting access. Although Green was capable of increasing access it came with risks of damage to the publishing system it terms of subscription cancellations and concerns that it was not self-sustaining. Gold open access was the preferred route in the long run, for its underlying sustainability, the potential for greater transparency and lower barriers to entry, and the potential for higher benefit/cost ratios and savings to UK public purse and to UK universities *provided* average APCs were not too high (the study used an APC of £1450 for its lower-APC scenario, and estimated the threshold average APC above which Gold would not be cost-effective at about £2000).

There are methodological constraints to all such studies which limit the confidence that can be placed in their findings:

- non-cash items: for example, including estimates of researchers' time saved by improved access is problematic because the saving is not realised in cash terms but assumed to be translated into greater efficiency. While the estimates may be plausible, an analogous problem of identifying increased economic productivity due to adoption of information technologies has proved surprisingly hard
- there are large uncertainties associated with several of the key variables

- in particular, the economic multiplier effects included by both the Houghton and the 2011 RIN report[53] result in large numbers that can swamp the other effects, and yet rest on untested assumptions
- none of the approaches have a good way of realistically modelling the likely heterogeneous take-up of open access, i.e. to reflect the likely situation where policy and implementation varies not just from country to country, but between institutions within the same country.

3.7. Recent developments in open access

Drivers of open access

The main drivers of uptake have the interventions and policies of research funders and policy-makers, and the growth and maturity of the open access publishing sector, an entrepreneurial activity which has increased the supply of credible open access journals to authors.

Research funder policies have likely been the most important factor in creating an environment for open access. Notable recent developments have included:

- publication of the UK Finch Group report (June 2012; Finch Working Group, 2012). Its recommendations were subsequently accepted by the UK government, marking a clear shift in policy in favour of open access for research articles, concluding that the "principle that the results of research that has been publicly funded should be freely accessible in the public domain is a compelling one, and fundamentally unanswerable"
- the tightening of the Wellcome Trust policy[54] (June 2012), in particular introducing sanctions for non-compliance and a move to CC-BY licenses
- the UK Research Councils new unified policy (announced July 2012, which largely develops the Finch recommendations as accepted by UK Science Minister David Willetts; RCUK, 2012)
- the European Union, is developing a new 7-year research programme, Horizon 2020, which comes into effect in 2014 and covers the EU's €80 billion funding. Requirements on authors will certainly be tightened, with speculation that there will be a target for 100% of funded openly available (the target on the previous programme, FP7, was 20%), a possible maximum for APCs, and requirements on reuse rights and embargo lengths (de Vrieze, 2012)
- in the US, the debate has been around NIH mandate and its possible extension. The Research Works Act (RWA), which would have reversed the NIH mandate was withdrawn and is effectively dead. Conversely, the Federal Research Public Access Act (FRPAA) which would extend NIH-type mandates to other federal funding bodies has been reintroduced but is unlikely to be passed. A third piece of legislation, the America COMPETES Bill was passed in 2011, required the Office of Science and Technology Policy (OSTP) to coordinate access policies across the federal funding agencies; consultation around this is ongoing.

---

[53] Both studies used the Solow-Swan growth model

[54] http://www.wellcome.ac.uk/News/Media-office/Press-releases/2012/WTVM055745.htm

The growth and maturing of the open access publishing industry is reflected in the growth of number of OA journals (see *Open access journal and article numbers*). Björk (2011) described the development of the sector from a volunteer model (often led by an individual scholar) in the 1990s, through a wave in which long-established journals, in particular society journals and journals from regions such as Latin America, made their articles OA when they started publishing parallel electronic versions, followed by adoption of OA as a business model from 2002, initially by new entrants and then by incumbent publishers both commercial and non-commercial. The adoption of the model by prestigious publishers, particularly non-commercial ones such as OUP, the Royal Society, and many leading societies, helps build credibility for the model for authors. Similarly while many authors would be reluctant to publish in new journals without impact factors, many OA journals have now existed long enough to establish credible impact factors (e.g. Björk & Solomon, 2012a).

Shift of policy focus towards Gold

A shift in thinking among policy-making and funders towards the Gold model appears to have taken place in 2012. This is particularly the case in the UK, where the Finch report not only recommended that outputs from research funded by the taxpayer should be made open access, but that the preferred option should be to do this via the Gold model, with funding made available to cover publication charges. The Group (and the UK government) appears to have accepted the advantages of Gold over Green primarily in terms of providing a sustainable business model for OA, but also in terms of avoiding risk of damage to a successful UK industry. The 2012 RCUK policy also accepts the benefits of the Gold model and has proposed a method for funding APCs (block grants to universities). Research intensive universities (e.g. the UK's Russell Group), however, remain concerned about the potential cost impacts.

Megajournals

The publishing model pioneered by *PLOS ONE* has proved highly successful and arguably represents one of few innovations to the scholarly journal model to have had significant widespread impact. The model consists of three key parts: full open access with a relatively low APC; rapid "non-selective" peer review based on "soundness not significance" (i.e. selecting papers on the basis that science is soundly conducted rather than more subjective criteria of impact, significance or relevance to a particularly community), plus a policy keeping review straightforward (e.g. avoiding where possible requests to conduct additional experiments and resubmit); and a very broad subject scope (essentially limited by authors willingness to submit and the journal's ability to find reviewers). In addition, the model was associated with the cascade peer review model (although this was in practice never very important for *PLOS ONE* in terms of numbers of submissions), and the journal promoted rapid publication, partly as a consequence of simpler peer review (although as it has grown it has struggled to keep publication times any faster than other leading journals in the field).

The success of *PLOS ONE* the megajournal model has led to widespread emulation by other publishers; examples include *AIP Advances*, *BMJ Open*, *SAGE Open*; *Scientific Reports* (Nature Publishing Group); *G3: Genes, Genomes, Genetics* (Genetics Society of America); *Biology Open* (Company of Biologists); and more planned (e.g. both the IEEE and the IET plan engineering megajournals). Other publishers have adopted elements of the model (broad scope, rapid publication, low-cost open access) but retained a more traditional selection peer review process: *Physical Review X* (APS); *Open Biology* (Royal Society); *Cell Reports* (Elsevier).

It remains unclear whether the megajournal model is something entirely new or the latest incarnation of brand extension or subject field journals of last resort. As *PLOS ONE* attracts articles from outside its core biomedical community this may have an effect on its actual average impact factor potentially reducing its attractiveness. It is also unclear whether an "all subjects" journal will really be stable in the long-term, considering its size. These factors go to the heart of the fundamental forces that have shaped journal publishing.

Other developments in open access

SCOAP3

SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics)[55] 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 was 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. SCOAP3 would then use this funding to allow publishers to publish the same journals but under the new open access model with centralised funding eliminating the need for author charges.

Having received sufficient pledges (technically, non-binding expressions of interest) to cover the projected budget, SCOAP3 conducted a tendering process during 2012. The tender process concluded in September 2012,[56] identifying 12 journals from 7 publishers for participation. These journals published 6600 articles during 2011, a large majority of the high-quality peer-reviewed HEP literature. There were some omissions, notably the American Physical Society's *Physical Review Letters*, the bid for which was rejected on price.

Articles funded by SCOAP3 will be available open access in perpetuity, under a CC-BY license, while publishers will reduce their subscription fees accordingly.

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 and nuclear physics share 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.

arXiv

The arXiv repository was launched in 1991 but as it has grown its host organisation (now Cornell) has struggled to justify the funding requirements. In August 2012 arXiv announced a new funding model[57] covering the period 2013–17 consisting of three sources of revenue: cash and in-kind support by Cornell LIbrary; grant funding from the Simons Foundation; and collective funding from the member institutions, i.e. institutions in high energy physics

[55] http://scoap3.org/

[56] http://scoap3.org/news/news94.html

[57] See http://arxiv.org/help/support

that have voluntarily agreed to make contributions toward the costs. Cornell hopes to raise $330k per year (36% of the total running costs) from the member contributions from some 126 institutions each paying $1500–3000 annually (a tiered rate depending on size of institution).

eLife, PeerJ

The growth of the size and mainstream acceptance of open access publishing appears to be encouraging a wave of experimentation and innovation (Van Noorden, 2012b). A good example is PeerJ, planned for launch in late 2012: founded by ex-PLOS and Mendeley staff, and backed by Tim O'Reilly's venture fund, PeerJ proposes a model in which authors take out a membership entitling them to publish articles. Lifetime memberships are priced between $99 and $349, with the highest band allowing authors to publish unlimited papers. Each author on a multi-paper (up to a maximum of 12) has to be a paid-up member; the average paper in PubMed has around 5–6 authors, so the effective rate may end up in practice nearer the megajournal rate ($1350) than $99. The model also has an element of of viral marketing built in, given that researchers coauthor papers with a changing cast of collaborators. The significance of PeerJ at this point lies not in its impact on the market but in representing the willingness of credible publishing professionals and risk capital to experiment with radical innovation in academic publishing.

A quite different approach is represented by the launch (also in late 2012) of the journal *eLife*[58] by three research funders, the Howard Hughes Medical Institute, the Max Planck Society and the Wellcome Trust. Explicitly setting out to create an open access competitor to the leading general science journals (*Cell*, *Nature*, *Science*), the *eLife* journal is described by its founders as the first step in a programme to catalyse innovation in research communication. *eLife* may or may not adopt a conventional Gold model (it will be free to authors to publish for a currently unspecified initial period); its significance lies in the unprecedented direct participation by research funders in the primarily publishing.

Open access for scholarly books

The initial focus of the open access movement was on access to research articles in journals. There has been growing interest in open access to other kinds of content, including educational resources and scholarly books, particularly monographs. The Finch report (Finch Working Group, 2012) recommended that interested parties should work together to promote further experimentation in open access publishing for scholarly monographs. The OAPEN (Open Access Publishing in European Networks)[59] project is a collaborative initiative to develop and implement a sustainable Open Access publication model for academic books in the Humanities and Social Sciences, originally EU co-funded as part of the eContentplus project. OAPEN launched the Directory of Open Access Books[60] in April 2012; as of September 2012 there were 33 publishers and over 1200 OA books listed.

Most approaches to finding a viable model for providing open access to monographs in the humanities have been based on either the delayed model or on providing online access to a basic electronic version in parallel with charging for higher-value versions such as print, ereader editions, enhanced ebook editions, and so on. In the sciences the Gold model of

[58] http://www.elifesciences.org

[59] http://www.oapen.org

[60] http://www.doabooks.org

author publication charges has been adopted by a few publishers[61] for multi-author monographs, where the individual chapter is equivalent to the journal article.

Reuse rights

Since the last edition of this report, there has been a clear trend towards placing greater emphasis on the licensing and reuse rights attached to open access articles. A number of funder and institutional mandates now require not just that some version of funded research articles are made freely available, but that they are licensed using the Creative Commons CC-BY licence to facilitate redistribution and reuse with the fewest restrictions (e.g. RCUK, 2012). The growing interest in text and data mining is one reason (see *Text and data mining*). A number of publishers have switched from the CC-BY-NC to the CC-BY licence for open access articles; dropping the "non-commercial" restriction will entail publishers foregoing any commercial reuse revenues such as reprints for pharmaceutical companies and other rights income (an important source of income for medical journals).

"Predatory publishers"

The reputation of open access publishing has been tarnished in some commentators' eyes by the emergence of so-called "predatory publishers" (Beall, 2012). These are alleged to take advantage of the low barriers to entry in OA publishing to launch large numbers of journals, and then use large-scale indiscriminate email to market to authors, sometimes not disclosing the (full) cost of publication until after acceptance, and listing editorial members who had not agreed to serve, and otherwise preying on researchers' need to publish or perish.

3.8. Transition and sustainability issues

The previous edition of this report included a somewhat theoretical discussion of whether open access would be sustainable. In the intervening three years, while not everyone has been persuaded, the actions of the scholarly community and the market suggest that the debate has moved more to *what* is necessary to make it sustainable (accepting it as a goal), and to the problems of how a transition could be managed. (These categories overlap, of course.) Gold open access is growing fast, but at present it remains only a small part of the market and there are valid questions about how a scaling-up would be achieved.

The key issues are:

- Is there a economically self-sustaining model of open access, and if so, what is it likely to look like?
- What will be the impact on economics of publishing: will economic returns be sufficient to continue to attract current publishers, or alternatively what might be the impacts of restructuring
- Will the same models for open access work in all fields, or for all types of journal (the "one size fits all" problem)?
- How will funding be managed during a transition?
- How will funding mechanisms be arranged (in detail), and what impacts will these have on scholars as well as publishers?

[61] e.g. Springer (http://www.springeropen.com/books); InTech (http://www.intechopen.com)

- What will be the impact of heterogeneous uptake, with different governments, funding bodies and institutions adoption different policies, and different cultural norms across disciplines?
- What will be the geopolitical impacts: how will the changes affect researchers in emerging economies and those in less developed economies?

Gold: a sustainable model for open access

There is general agreement that under appropriate circumstances Gold open access offers a viable business model that can be both economically self-sustaining and provide wider economic and access advantages over the subscription model. This has been reflected in policy-oriented studies (e.g. RIN, 2011c, Finch Working Group, 2012), in the profitability of new open access publishers (Ithaka S+R, 2011, PLOS, 2012), and in behaviour of existing commercial publishers in launching open access journals, and is also recognised in statement signed by STM members supporting sustainable open access.[62]

Ignoring for the moment differences between disciplines and other complications, it is axiomatic that for the model to be sustainable, the prices that authors and their funders are willing to pay need to be greater than full costs (including sufficient surplus for ongoing investments and to cover the cost of capital). The average current cost of producing an article has been estimated at £2364 (say $3800) (RIN, 2011c). As we saw above, one study estimated the average APC paid in 2010 across OA articles published in the journals in the Directory of Open Access journals at just $906 (Björk & Solomon, 2012c). Table 3 showed that most leading publishers charge APCs in the range $1300–5000 (though Hindawi's median APC is $600), while the average APC paid by the Wellcome Trust in late 2010 was $2365. Whichever figure is chosen, however, it is lower than the current reported average cost and revenue per article.

To be viable, therefore, prices need to be higher, or costs lower, or both.

One of the mooted advantages of the open access model (from an economic perspective) is its greater price transparency and hence price competition; if true (and market developments do suggest that price is being used by new entrants as a competitive element), the prospects for higher prices would be remote. As a 2009 Outsell report illustrated, substantial substitution of open access publishing for subscription-based journals under these circumstances would lead to a shrinking in the size of STM publishing market by revenue (Outsell, 2009); on their assumptions the market would hypothetically have been about half its previous size (hypothetical because it assumed a wholesale conversion to OA). The report did go on to model growth scenarios for different levels of uptake and pricing, assuming increased article output due to differing levels of R&D spending growth, and showed that the market could under some circumstances recover to pre-OA levels after a period, but only at APCs higher than those currently seen in the market.

On the assumption that the market dynamic will be to lower prices, the concern of some industry commentators is then that there could be undesirable unintended consequences. First, pressure to lower costs could lead to corners being cut and quality reduced ("the race to the bottom", e.g. Anderson, 2012). For some types of publishing, a low-cost no-frills option appears to be what the market wants – witness the growth of *PLOS ONE* – but the approach does not fit the more highly selective journals carrying significant amounts of

[62] http://www.stm-assoc.org/publishers-support-sustainable-open-access/

additional, non-research article content, nor the increasing demands for novel tools to become standard (e.g. plagiarism detection).

Second, pressures on revenues and thin margins could increase pressures on editors or publishers to reduce scientific standards to accept more articles. Arguably the same or similar pressures exist under a subscription model (since publishers have been able to pass on higher prices as journals expanded), and the answer is surely the same in both cases: journals with poor standards will increasingly be unable to attract good authors or editorial board members, and will languish accordingly.

To date the uptake of the Gold model has varied substantially by discipline, with greatest uptake in biomedicine and the lowest in the humanities, maths and (perhaps more surprising) chemistry (e.g. see Björk et al., 2010). Consequently most (but by no means all) entrepreneurial publishing activity has also concentrated on this area.

Factors favouring the uptake in biomedicine include the high level of research funding and research funders that have set the agenda. With government policies moving towards open access for *all* scholarly outputs, some question how this will be managed in disciplines where external funding is not the norm (e.g. maths, humanities), or just more generally where authors are unfunded (only 60% of authors overall are grant supported). The answer appears to be that universities (or other employers) would fund APCs from central resources (e.g. via the block grants proposed in the new RCUK policy), though this raises other issues such as who decides how to ration a finite publication fund (e.g. see Crotty, 2012).

A key obstacle to wider adoption is funding the transition. For individual institutions, adopting a (national or local) policy in favour of Gold open access would increase their costs via APCs while they were still paying for the continuing subscription-based journals (e.g. Swan & Houghton, 2012). The same is true at a national level, if a country adopts a policy favouring Gold open access significantly in advance of the rest of the world. These issues were modelled in the Heading for the Open Road report (RIN, 2011c) and discussed in the Finch report, which recommended that the UK government provide an additional £38 million per year during the transition, plus one-off costs of £5 million, to cover these effects during the transition. Unsurprisingly these were the recommendations not adopted by the government (though it did provide an additional one-off £10 million). As some have pointed out, scaling these transition costs up to a global scale would lead to very large costs.

One transition issue is that the economic case for Gold rests on non-cash savings (e.g. researchers' time) and uncertain economic multiplier effects. In a recessionary environment such benefits carry less weight when set against the cash costs of implementation.

Another way in which open access could shrink the market could be through the impact on rights income (which would presumably be lost if a CC-BY licence were adopted), and on corporate subscriptions (these currently represent approximately 15–17% of journal income, but corporations contribute only a small fraction (around 5%) of papers). This need not be an issue for journal finances if it is priced into the APCs, though some point that this would in effect represent an undesirable transfer of payments from corporations to universities.

Similarly APC pricing needs to factor in an allowance for waivers if these are offered (e.g. to authors from developing countries). This also represents a transfer or subsidy, though few would object. OUP has reported that for about 70 journals its waiver rates have been stable at 6–7%, while PLOS cites a level under 10%. Researchers from middle-income countries may feel uncomfortable about requesting waivers, though; one survey suggested that the fraction of authors paying APCs from their personal funds was substantially greater in such countries.

There are other concerns about the impacts of open access on emerging and less developed economies. One is the impact on local publishers: if open access becomes the norm, authors might desert local journals because they would be unable to waive APCs (because the majority of their authors would qualify), in favour of Western journals offering waivers. A waiver system is therefore not desirable in the long run, except perhaps for the poorest countries (Dickson, 2012).

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.

Loss of print-based advertising would be an issue for some journals, including the wide-circulation general journals, although advertising overall represents only 4% of journal revenues overall. This issue is, however, more to do with the digital transition generally than an issue for open access; there are for instance some indications that tablet editions are proving attractive to advertisers.

One issue not addressed by this discussion so far is whether the political zeitgeist, and in particular public attitudes towards the internet, could make paid-for publishing unsustainable. The potential threat comes from attitudes that online content should be free; that sharing of content is the default option on the web; that the notion of intellectual property is outmoded; and that public funding automatically equates to public access. Michael Mabe has discussed these issues in a recent book chapter, and concluded that the battle is not yet lost so long as a copyright framework can be maintained and politicians understand the risks involved (Mabe, 2012).

Hybrid (optional) open access

Hybrid was originally proposed as a lower-risk route for subscription journals to move towards open access without risking all in a one-off transition. Of course, some publishers may have introduced it less in expectation of a near-term transition to OA than to take advantage of available funding, and to offer authors a route for compliance with funder policies.

In practice average uptake has remained very low (around 2%) but there are exceptions. Bird (2010) reported hybrid uptake rates for OUP journals (where she works) and for some other publishers. OUP has adopted the Oxford Open for over 80 journals. Overall uptake was 6% but varied substantially by discipline, from 2% in the humanities and social sciences, through 4% in medicine, 6% in maths, to 10% in the life sciences. Some life science titles had much higher uptake, e.g. *Human Molecular Genetics* at 18% and *Bioinformatics* at 30%. OUP's figures appear higher than those reported by other publishers. Bird speculated this might be because OUP offers a 50% discount for authors at subscribing institutions (usually the majority of authors.) She quotes uptake at other publishers as follows: Nature Publishing Group: 5% across their specialist STM titles, with some titles showing higher uptake (e.g. *EMBO Journal* at 11%); Wiley-Blackwell surveyed other publishers' sites, finding 1-2% overall, but with some titles up to 20%; and Wiley-Blackwell's own uptake was "very low" overall, but with two biomedical journals at 10-20% in 2008. One stand-out example that has arisen subsequent to Bird's article is the launch of *Nature Communications*, which has an opt-in rate (at $5000 per article) of over 40%.

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 of society publishers using the HighWire system, who collectively make available over 600,000 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 (historic) 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.

3.9. Effect of self-archiving on journals

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

Some support for this hypothesis was given by a in a now-dated 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 disposed 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 number of studies, including the PEER project, have demonstrated the preference of researchers for the version of record, at least for some stages of the research publishing cycle.

One issue has been whether self-archiving can lead to reduced article downloads from the publisher's website, given the importance of usage in librarians' selection and cancellation decisions, and for trends in the market for usage to be a factor in pricing journals. PEER (Publishing and the Ecology of European Research), an EU-funded project that ran for a total of 45 months between 2008 and 2012, involving 12 participating publishers and six repositories from across the EU, has provided the most comprehensive and detailed study yet of the impact of archiving on open repositories.[63] The project findings covered a broad range of topics, including publisher/repository economics and behavioural research (some of which are reported elsewhere in this report), but the usage studies provided the most pertinent data on the effect of repositories on journals.

The usage study (CIBER Research, 2012b) was designed as a randomised controlled trial to compare downloads of articles from publisher websites with and without parallel availability of the article in the PEER network of repositories. The key finding of the study was that, far from reducing publisher downloads, exposure of articles in PEER repositories

[63] http://www.peerproject.eu/

was correlated with a modest increase in downloads from the publisher site. The overall increase was 11.4% (for which the 95% confidence interval was 7.5% to 15.5%). The researchers suggest the likely explanation is that higher digital visibility to search engines that PEER deposit created by virtue of high quality metadata (publisher metadata was enhanced and extended on ingestion in many cases) and a liberal policy on indexing by search engine robots. The authors concluded that there is no experimental evidence to support the assertion that PEER repositories negatively impact publisher downloads, and argue that a binary "repository versus publisher" opposition is a false dichotomy, and that "that are players in a complex scholarly communications ecosystem where visibility is king and the key players are increasingly the general search engines". Another factor that needs to be taken into account was that each journal in the PEER project was able to select an embargo period suitable for it. This means that a one-size fits all approach which might be considered the likely norm in the real world was not reflected in the project.

This is not necessarily the whole story, though: first, the PEER researchers take pains to stress that their findings apply only to the PEER repositories, which were untypical for a variety of reasons. For instance, PEER found the publisher uplift to be statistically significant in the physical sciences, but this finding is contrary to the (uncontrolled) experience of some physical science publishers in relation to coverage by the arXiv repository, which is both much better known in the field and also – crucially – contains essentially the complete contents of journals in some sub-fields of physics, although most of it is Stage 1 or preprint content. Second, even if the finding were general, the increased usage generated by repositories would not necessarily mean that libraries facing budget cuts might not preferentially select journals whose contents were available on repositories over ones that were not. For example, a 2012 survey conducted for ALPSP and the Publishers Association asked librarians whether they would continue to subscribe to journals if the majority of content was freely available online after a 6 months' embargo (Bennett, 2012); 34% of respondents said they would cancel some STM journals, and 10% that they would cancel all such journals, and for AHSS journals, 42% would cancel some and 23% all affected journals.

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. The PEER usage study also provided data on the lifetime usage profile of articles. Figure 19 (taken from CIBER Research, 2012a) shows the cumulative publisher downloads for different subject areas following publication. Usage only starts to plateau for the life sciences, medicine and physical sciences around 56 months, while social sciences and humanities continued to rise steadily at 80 months. The key issue for subscribers, though, is less the overall length of age profile than its shape, i.e the importance (and hence value) placed by researchers on access to the version of record during the first 6, 12 or 24 months. In any case, regardless of the evidence, policy-makers including the UK Research Councils and the EU are increasingly opting for shorter embargo periods (see *Self-archiving policies and mandates*).

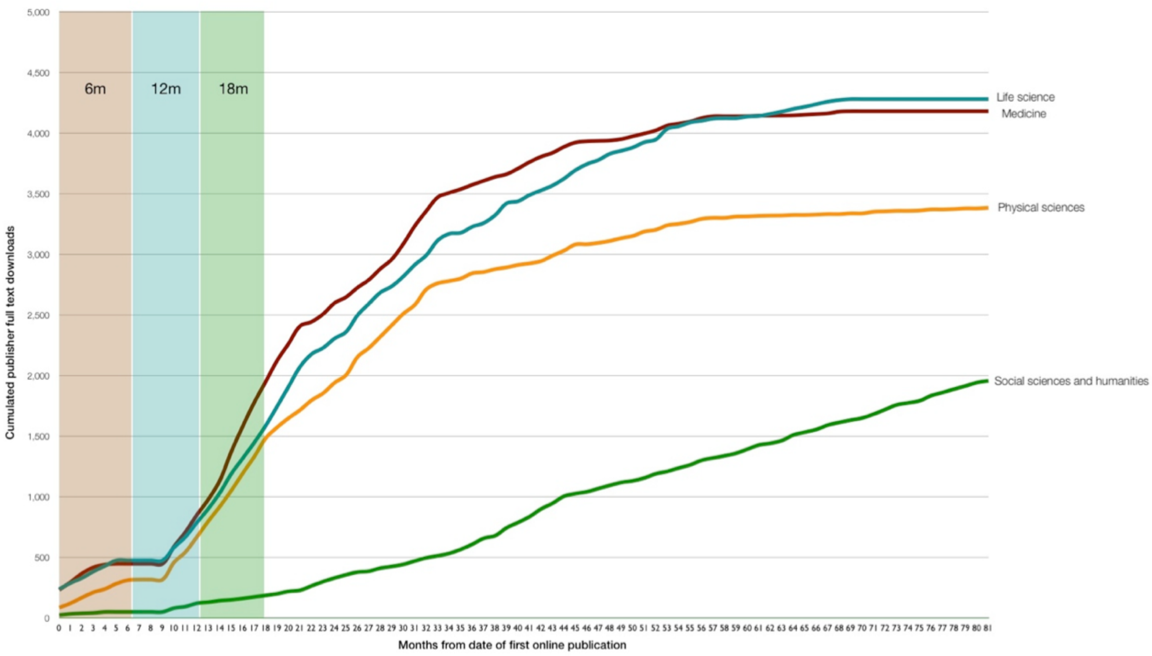
PEER project

In addition to the usage data study, the PEER project had two other main research topics, looking at behavioural and economic aspects (Wallace, 2012).

The key behavioural finding is probably not part of the behavioural project *per se*, but simply the complete lack of interest by authors in depositing articles under the scheme. The initial plan was to populate the archive half with articles deposited direct by publishers, and half by authors. Despite sending nearly 12,000 invitations only 170 papers were deposited by authors. This may have had something to do with the experimental nature of the project, and that PEER would have been previously unknown to them, but there was also anecdotal evidence that some researchers considered making journal articles accessible via Open Access to be beyond their remit. Indeed, authors who associated open access with self-archiving were in the minority. Overall, the PEER behavioural project concluded that "Academic researchers have a conservative set of attitudes, perceptions and behaviours towards the scholarly communication system do not desire fundamental changes in the way research is currently disseminated and published." They were not necessarily negative about repositories but were certainly very guarded, and unpersuaded that the benefits justified changing their own behaviour.

The PEER economics study uncovered a little new information. It confirmed that peer review had real costs and had few economies of scale, and estimated the publisher average cost of peer review per submitted manuscript (salary and fees only, excluding overheads, infrastructure, systems etc.) at $250. Excluding peer review, average production costs were estimated to be in the range $170–400 per paper published (again excluding overheads etc.). Publisher platforms had annual maintenance costs of $170–400,000, on top of set up and development costs typically costing hundreds of thousands of dollars.

Figure 19: Article downloads from publisher sites by age and subject area (CIBER Research, 2012a)



3.10. Open access impacts on use

Impact on usage (downloads)

There appears to be widespread agreement that freely available articles are downloaded significantly more than comparable articles. For example, the *Journal of Experimental Biology* compared the optional OA and the non-OA articles in 2007, and found the full-text versions of the OA articles were downloaded approximately 40% more than the non-OA articles (Bird, 2010).

More robust evidence comes from a 2011 randomised controlled trial found OA articles were downloaded significantly more often, with HTML downloads roughly doubling and PDF downloads increasing 62% (Davis, 2011; Davis & Walters, 2011).

Part of the increase in downloads may not be due to increased (human) use of the content: When OUP converted *Nucleic Acids Research* to open access, article downloads more than doubled, but most of the increased use was attributed to search engines with only an additional 7–8% use beyond this (Bird, 2008). *NAR* was, however, a leading mature journal (and hence likely to be subscribed widely), and it publishes in an area unlikely to be of interest to outside the professional research community.

Easy ubiquitous availability may also change what a "use" may be, as researchers' behaviour changes. For instance, CIBER, 2008 and others have shown that users "power browse" through an initial hit list of articles (typically found from a search), skimming and discarding many while retaining a few for later study and use. Outsell, 2009 pointed out that this means that while articles are seen as being of uniform value by publishers, for the researcher the value may vary from zero (instantly discarded) to significant.

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 (e.g. Lawrence, 2001). 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 (Kurtz et al. 2005).

A bibliometric study in 2010 (Gargouri et al., 2010) attempted to control for selection bias by comparing self-archived papers in institutions with strong (and reasonably well observed) mandates requiring deposit with those from other institutions (with low, author-selected deposit). They found a strong statistical correlation between open access status and citations, with the effect strongest for the most cited articles.

Craig and colleagues 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 (Craig et al., 2007). Citation patterns differ between subject disciplines, however, so this still leaves it open that there may be an effect in other fields.

Davis (2011) reported a randomised controlled trial that found the open access articles received significantly more downloads and reached a broader audience within the first year, yet were cited no more frequently, nor earlier, than subscription-access control articles. In a later review, Davis & Walters (2011) assessed the available evidence and concluded that the impact on citations was "not clear. Recent studies indicate that large citation advantages are simply artefacts of the failure to adequately control for confounding variables". But he found that "the conclusions of Craig and colleagues were well supported by subsequent work. After controlling statistically or methodologically for confounding effects, there is little evidence that open access status has an independent effect on citation counts".

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

Björk and Solomon also conducted a different kind of study, comparing citation performance of open access and subscription journals (Björk & Solomon, 2012a). They found that overall citation rates were about 30% higher for subscription journals, but after controlling for journal age, discipline, and the location of the publisher, the differences largely disappeared. Open access journals with article processing charges were more highly cited than OA journals without APCs. In medicine, OA journals launched in the last 10 years receive about as many citations as the subscription journals launched in the same period.

Whether or not the citation effect exists and how large it might be is increasingly a matter of academic interest, however, as the proportion of literature that is open access steadily increases.

4. New developments in scholarly communication

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

Given the accelerating rate of change, covering trends in technology in a report like this that is updated only every three years presents some challenges. There was at the time of writing nonetheless fairly good consensus that the trends discussed here are important to scholarly publishing, although there may be debate as to their relative importance.

For a more regularly updated view, we recommend following the reports of the STM Future Lab Committee[64]. The most recent report of this committee (STM Future Lab Committee, 2011) identified the key theme for 2012–15 to be "From Discoverability to Actionability of Content", or in other words a shift of focus from technologies aimed at supporting search and discovery (e.g. platform architecture, semantic enrichment, SEO, etc.) to ones that make the content more useful, more interactive, more usable and more reusable (e.g. APIs, data integration, data and text mining, semantic web technologies, productivity and workflow tools). The report identified three key topics: API platforms, research data, and identity management which were central to this view of making content more "actionable".

A more detailed look at the STM publishing platform (Outsell, 2012a) identified the following key themes: identifying the role of the publisher platform in an increasingly open information ecosystem; the growing importance of the user experience; discoverability; moving beyond simple personalisation; and social media and networks. These were discussed in the context of the following technology trends: mobile; semantic enrichment; search tools; APIs;[65] ecommerce and monetisation; convergence and integration; functional (active) content; and analytics.

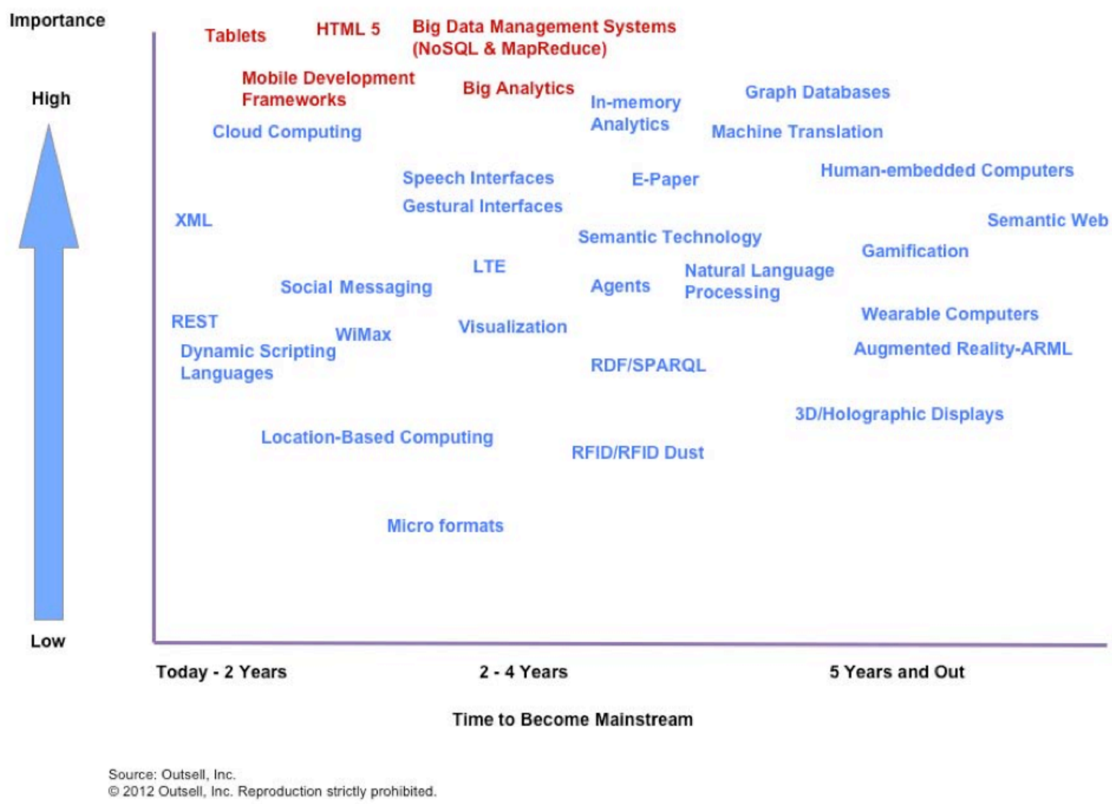
Looking at the information industry more broadly, Outsell's technology watch report (Outsell, 2012b) identified five key technologies driving change (see Figure 20): tablets, mobile development frameworks, HTML5, "big data" management systems, and "big analytics" (i.e. tools to mine and analyse big data).

Each of these various topics and themes is picked up in the sections following.

[64] http://www.stm-assoc.org/future-lab-committee/

[65] Application Programming Interface, e.g. see http://en.wikipedia.org/wiki/Application_Programming_Interface

Figure 20: Five Disruptive Technologies Plus Others That Matter (source: Outsell)



4.1. Publishing platforms and APIs

It is well known that the large majority of searches do not start on the publisher's site (e.g. up to 60% of web referrals come from search engines). Given this, what it the role of the publisher platform in the researcher's workflow? If researchers are journal- and publisher-agnostic, and want to get in and out of the publisher's site as quickly as possible having found and downloaded the PDF (CIBER, 2008), should publishers design sites to be (smart) repositories of (smart) content with maximum open web discoverability and open APIs, fine-tuned for fastest possible delivery of content through whatever service the end-user chooses to access? Alternatively, should publishers invest in semantic enrichment, increased engagement, adding or integrating workflow tools to create a rich, productive environment? Publishers appear to conclude that both behaviours need to be supported, whether a power browser bouncing in and out of the site, or a researcher in a more exploratory phase seeking a more immersive or interactive experience.

A key technology feature for the STM platform will be open APIs (here "open" means that the specification is freely available, not the content). The strategic reason is that much of the value of the platform will increasingly lie in its interoperability (e.g. ability to integrate content from multiple sources, to integrate and share data, to add functionality, and to allow users to access their content from within their chosen starting point or workflow tool). More tactically, deployment of modern APIs will allow publishers to develop new products and services faster, to develop internal workflow process and manage them more easily, and to support multiple devices more easily.

4.2. Social media

Social media and networks (sometimes referred to as Web 2.0) offer the potential to enhance informal and formal scholarly communication, although their impact to date has been limited.

A number of studies have looked at researchers' use of social media. RIN's report *If you build it, will they come?* found low take-up, with under 15% using regularly (RIN, 2010). Only a small group, around 5%, used blogs and other tools to publish the outputs and work in progress. The main barrier to greater use that RIN identified was the lack of clarity over potential benefits: the costs of adoption were not trivial, and without clear and quick benefits researchers preferred to stick with the services they already knew and trusted. The rapid development and proliferation of services meant it was hard to keep track of them, or assess their potential benefits, and their proliferation tended to mean that each lacked the critical mass users needed. There were also a second set of barriers around quality and trust: researchers were discouraged from using new forms of scholarly communication that were not subject to peer review or lack recognised means of attribution. And contrary to the stereotype, there were only small differences in use by demographic factors including age. RIN's overall conclusions was that there was little evidence to suggest that web 2.0 would prompt in the short or medium term the kinds of radical changes in scholarly communications advocated by the open research community.

Other studies showed have found similar results (Ithaka S+R, 2010; Procter et al., 2010; RIN, 2009a). More anecdotally, David Crotty has written thoughtful accounts of the current crop of Web tools for biologists and why they are not more successful, seeing 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 (Crotty, 2008, 2010).

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*). Where tried (e.g. rating and commenting for journal articles), uptake has been very low, and there are serious questions as to what is measured through such techniques (Harley et al., 2010).

On the other hand, trends in social media use in the general population are so strong that many believe that they will become a more substantial part of scholarly communication over time, particularly as they become more tightly integrated into PC and mobile operating systems. For example, Mendeley, ResearchGate, and CiteULike have all developed substantial user bases. There is some indications that Twitter may be able to play a role in predicting highly cited papers (Eysenbach, 2011). The growing adoption of article-level metrics may also create more awareness of the use of Twitter or blogs to discuss or promote journal articles, and hence perhaps a positive feedback effect. And closer integration of social features into services (as with Mendeley), rather than trivial inclusion of a "Like" button can build social behaviours more naturally. Overall, therefore, there a case for believing social media will at least play a part in content discovery.

Looking specifically at wikis, 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. Examples include WikiPathways, which uses standard wiki software to create a site "dedicated to the curation of biological pathways by and for the scientific community", and OpenWetWare, which promotes sharing of information among the biology and biological engineering research 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.[66] The journal *RNA Biology* has required since 2008 authors of articles on RNA families also to submit a draft article on the RNA family for publication in Wikipedia, with the hope that the Wikipedia page would become the hub to collect later information about the RNA family. This policy has not been widely emulated by other journals.

4.3. Mobile access and apps

Professionals of all types are under increasing pressures to perform more complex tasks at an accelerating pace in an environment greater regulation and accountability and overloaded by ever-increasing amounts of data. It is not surprising that in these circumstances that mobile access to information, tools and services has the potential to create huge benefit.

The adoption of mobile devices has been, and continues to be extremely rapid, even by the standards of the internet age. The number of smartphones sold in 2011 exceeded the number of PCs, and in just a few years the numbers of mobile devices will dwarf PCs – by 2015, the number of internet users will have doubled compared to early 2012, and most of these users will be accessing the internet via mobile devices (Meeker, 2012; Blodget & Cocotas, 2012).

Uptake is even more rapid among some professional groups than in the public at large; for example, over 80% of US physicians own smartphones, 62% a tablet, and nearly 90% regularly use a smartphone or tablet to access clinical information between patient sessions (Outsell, 2012c).

The cost/benefit equation is clearer for busy professionals than for most academic researchers, but mobile device use is rising fast in this group too, with growth mostly coming from increased tablet uptake. As of mid-2012, the leading STM platforms were seeing under 10% of their traffic coming from mobile devices, but mobile traffic was doubling year-on-year.

Use cases for mobile are still emerging and developing. The first generation of apps tended to simply provide access to information (that is, they show something), rather than allowing the user to achieve something within their workflow (i.e. do something). So STM publishers initially addressed the core needs of "looking up and keeping up", i.e. searching for facts and small pieces of information, and keeping abreast of developments via RSS or eToC feeds or similar. Clinical calculators are a little more interactive but play a similar role.

Although most of the current interest is generated by the rapidly expanding tablet market, there seems likely to be applications that remain well-suited to smartphones despite the

[66] Two initiative mentioned in the last edition of this report (Elsevier's WiserWiki and SciTopics) have now been discontinued

growth of tablet uptake – e.g. point-of-care drug information is ideally delivered through a device that is always in your pocket.

On the tablet, additional uses include long-form reading, more immersive self-study and other education applications, and active engagement with research content (still in its infancy, but could include annotation and highlighting, adding papers to bibliographic systems, and tagging, though to perhaps creating presentations or other new content). In the future there will be increasing integration of mobile apps with workflow and enterprise systems (e.g. medical records and e-prescribing systems, and similar).

There is one more important difference between mobile app-based access and PC web-based access to journals. Mobile devices are personal, rarely shared, thus tying usage data to the individual rather than the institution as happens with web access (where access control is typically by IP range). The app environment allows much richer data to be collected (with appropriate consents) about the user's interaction with the app/content. And the app ecosystem (i.e. device plus cloud plus App Store etc.) encourages purchases via a single click (including from within the app itself), tied to the individual's credit card (via the App Store) rather than the library budget.

Business models are, like use cases, still developing. For research journal publishers, the default option has simply been to provide mobile access as a (necessary) additional service. Mobile subscriptions are increasing, however, offering a new opportunity for individual and member subscriptions. Reports suggest much higher engagement with advertising in tablet versions of medical journals than with web version, and hence higher prices and advertising renewals (Edgecliffe-Johnson, 2012), suggesting that tablet editions may offer a route into fully digital versions for journals with advertising content (and a potential route for societies to drop their membership print editions). In the general public mobile app market, in-app purchases dwarf revenues from app purchases or subscriptions, and this model clearly has potential in STM (e.g. for individual issues, additional chapters of text or reference works, etc.).

There are important technology choices to be made for publishers in addressing the overlapping issues of mobile access and apps, that go beyond the scope of this report. At the time of writing, by no means all STM journals and platforms offered a mobile-optimised interface (e.g. using responsive design[67]). For app development, publishers have to choose between native apps (written in the development language for each individual device), webapps (written using open standards especially HTML5), or hybrid apps (combining native code with web content). Native apps currently offer the best user experience (e.g. greatest speed and responsiveness, and tightest integration with the device features), whereas HTML5 offered the promise of a standards-based, write once for all browsers approach with lower development and maintenance costs.

4.4. Data-intensive 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-intensive science, there is an abundance of data which

[67] http://en.wikipedia.org/wiki/Responsive_web_design

can be explored to create and test new hypotheses. The late Microsoft researcher Jim Gray argued that this enabled a fundamentally different, new way to conduct science, the "fourth paradigm" (e.g. see Lynch, 2009), joining the earlier three paradigms of theory, experimentation and computer simulation.

Since the last edition of this report, there have been substantial developments in this field:

- research funders have introduced policies (or tightened existing policies) requiring the deposit and sharing of research data. The NIH data sharing policy, for example, now expects data to be shared "no later than the acceptance for publication of the main findings from the final dataset", and requires researchers to include a data management plan with all new grant applications[68]
- the number of data repositories is growing: the list maintained by DataCite currently includes over 150 repositories[69]
- data repositories have been developed to host "orphan data", that is, datasets for which there (currently) exists no recognised disciplinary repository. Examples include and OpenAIRE[70] and DRYAD[71], which specifically hosts the data underlying peer-reviewed articles in the basic and applied biosciences
- DataCite[72] was launched in December 2009 to address the challenges of making research data visible and accessible. It assigns persistent identifiers for datasets (currently based on DOIs), and provides registration and search services
- Thomson Reuters launched the Data Citation Index[73] in October 2012, supporting data discovery, reuse and interpretation. At launch the DCI initially indexed some 70 data repositories
- data journals have started to emerge. Examples include *F1000 Research* (Science Navigation Group); *Biodiversity Data Journal* (Pensoft); *Earth System Science Data* (Copernicus); and *GigaScience*, a combined database/journal collaboration between BGI (formerly the Beijing Genomics Institute, and the world's largest sequencing centre) and BioMed Central
- ODE (Opportunities for Data Exchange, an Alliance for Permanent Access project)[74] has produced a set of recommendations for journals, including the introduction of stricter editorial policies on the availability of underlying data, recommending data archives, providing citation guidelines for data using persistent identifiers, and launching or sponsoring data journals.
- Other notable initiatives include: BioSharing, which works to build stable links between journals, funders with data-sharing policies, and standardisation efforts in biosciences;

---

68 http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#time

69 http://datacite.org/repolist

70 http://openaire.cern.ch

71 http://datadryad.org

72 http://datacite.org

73 http://wokinfo.com/products_tools/multidisciplinary/dci/

74 http://www.alliancepermanentaccess.org/index.php/community/current-projects/ode

BioDBCore, a community-defined, uniform, generic description of the core attributes of biological databases; ISA Commons, which produces an open-source framework for data sharing centred around the general-purpose ISA-Tab file format; and DCXL (Digital Curation for Excel), a California Digital Library project supported by Microsoft Research aiming to support individual scientists with data management, sharing, and archiving by building an open-source add-in for Excel.

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

- researchers will increasingly 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. Publishers, data repositories and the various individual research communities will need to agree on the respective roles for data hosted by journals (e.g. in supplementary materials files) and in repositories. In most cases it seems likely that it will be preferable for the data to be hosted in properly managed repositories.
- there will need to be two-way linking between journal articles and research databases. There are research projects working in this area (e.g. SURFshare, DRYAD, OpenAIRE) in addition to the initiatives listed above.
- the dataset will start to become a (mainstream) unit of publication, with quality control 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.). There are (at least) two possible business models: one is simply to base the quality control on the peer review of the linked journal article; and second, membership model providing services to users (e.g. as at PANGAEA).

4.5. Semantic web

It is convenient to distinguish between the *semantic web*, and the use of semantic technologies and semantic enrichment of content.

These concepts involve 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, as originally articulated by Web creator Sir Tim Berners-Lee, remains complex, expensive and difficult to achieve, but more pragmatic, domain-bounded approaches are already adding significant value in STM publishing as well as across the Web in general.

Semantic enrichment is the tagging of content to add value by identifying and linking terms and concepts of interest to a particular domain, organised into structured taxonomies . While this can be done manually (indexing is an example), in practice its large-scale deployment in STM had to wait for the development of automated tools.[75] This can be thought of as a multi-stage process:

- Automatic extraction of metadata (which will be specific to each domain): the identification of terms and subsequent mapping these to defined entities (again domain specific);
- Defining the relationships between entities (e.g., Condition X is a symptom of disease Y and a side-effect of drug Z);

[75] e.g. from TEMIS, Access Innovations, Silverchair, and others

- Creation of links between the entities within and across documents to build a structured knowledge base;
- Use of analytics to derive new knowledge .

The benefits of semantic enrichment are shown in Table 5. They fall into three broad areas:

- *Smarter content*: a key benefit of semantics is to improve search and discovery, providing powerful new ways to find related material, explore new areas, put research into a broader context, and so on. Use of taxonomies allows users to find content even when their search terms do not exist in the article, and to discover related content. Matching content to user interests can be used to deliver personalised content recommendations. User interests can be self-declared but the technique becomes more powerful when automated techniques are used, which could be semantic or statistical analysis of the article content viewed, or behavioural data or derived from collaborative filtering.
- *Enabling new products and services*: for example, closer matching of advertisements to user profiles and/or displayed content has already been shown to dramatically improve click-through rates and hence achievable yields. Grouping content by semantically defined areas can allow new subscription products to be created, with content dynamically updated to match.
- *Internal productivity*: semantic enrichment can also be used by publishers to automate their own internal editorial and production workflows, for example through (semi-)automated editorial mark-up or providing recommendations for peer reviewers.

Adoption of these semantic technologies will also facilitate text and data mining techniques. This in effective 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 (see *Text and data mining*).

4.6. Linked data

Linked (open) data is a way of publishing data on the web in a structured way that facilitates interlinking of the data to other resources, and thus makes the data more useful. Built using standard web technologies, linked data allows relationship between things to be expressed, which greatly facilitates navigation between, and integration of, multiple information sources. Linked open data is the same thing except that an "open" licence is used, permitting sharing, extension and re-use (Miller, 2010).

It is potentially a way for publishers to make their content more discoverable and increase usage within new services. It can thus be seen as equivalent to supplying metadata through automated feeds to A&I and library systems suppliers. The technology is much more powerful, not least because it enables third parties (e.g. libraries or application developers) to create new services integrating multiple sources.

At present the major search engines appear to be holding back from fully committing to linked open data (though Microsoft's Academic Search makes use of it for some services, e.g. the graphical visualisation of search results) in favour of an alternative, simpler approach to structured data called microdata using the Schema.Org specification.

Some therefore see linked data as lacking a "killer app" which would drive more rapid adoption. Nonetheless it is backed by (and under active development at) major libraries including the Library of Congress and British LIbrary, and by the OCLC Online Computer Library Center.

Publishers are starting to explore the potential: for example, Nature Publishing Group has released its article metadata via linked open data. Elsevier is also exploring its use, while Thomson Reuters (publishers of the Web of Science) also support linked data in some areas (including the OpenCalais service).

Table 5: Publisher Benefits from Deployment of Semantic Content Enrichment (source: TEMIS/Outsell; Outsell, 2012a)

Area	*Example services*	*Benefits*
Smarter content	SEO Faceted search Linking Recommendations Personalisation	Increased usage Lower marketing costs Improved renewal rates Increased transaction revenues Author perceptions
Semantically derived products	Semantic (targeted) advertising Knowledge bases Topic pages Collections and 'slices'	New revenue streams Increased yields Re-use of existing assets
Workflow productivity	Automated content processing (e.g. tagging, linking) Content discovery Peer reviewer recommendations	Lower costs Improved consistency Scalability Reduced time to market

4.7. Text and data mining

Text and data mining (TDM) has the potential to transform the way scientists use the literature (*Nature*, 2012). It is expected to grow in importance, driven by greater availability of digital corpuses, increasing computer capabilities and easier-to-use software, and wider access to content (the Publishing research Consortium report *Journal Article Mining* gives a good introduction to TDM (Smit & van der Graaf, 2011; see also JISC, 2012).

TDM draws on natural language processing and information extraction to identify patterns and find new knowledge from collections of textual content. Semantic enrichment and tagging of content are likely to enhance TDM capabilities. At present TDM is most common in life sciences research, in particular within pharmaceutical companies, but relatively little used elsewhere. One barrier to more widespread adoption is the lack of an efficient licensing regime (see *Text and data mining rights*) but there are also technical issues such as standard content formats including basic common ontologies.

4.8. Big data and analytics

"Big data" refers to collections of data too large to be handled and analysed in conventional databases systems.[76] Tools for handling big data were developed at Yahoo (Hadoop), Google (MapReduce), and Amazon, driven by the need for search engines and large consumer web sites to handle enormous amounts of user data in real time. Large datasets on this scale arise

[76] this is something of a moving target but currently measured in petabytes, and exabytes

from the web itself, customer and user data (e.g. Walmart, Facebook), in healthcare data, location data, device data ("the internet of things"), and of course scientific research (e.g. CERN's Large Hadron collider processes 40 million images per second).

Consultants such as McKinsey have predicted large economic benefits to firms and to society from adopting big data techniques – for example, they estimate annual benefits to US healthcare at $300 billion, the annual consumer surplus from using personal location data at $600 billion, and so on (McKinsey, 2011).

The scale of these challenges may seem to put big data beyond the reach of STM publishing and information suppliers, but this is not necessarily true. A special issue of Elsevier's *Research Trends* has recently discussed examples from the world of research but also including the use of big data in science policy, research investments, and bibliometric analysis (Halevi, 2012).

Various types of STM data may be amenable to big data techniques, including research data, full text collections (i.e. text mining), metadata including citations, and usage and behavioural data. Some specific examples of big data in STM include:

- PlantWise is a CABI initiative to improve food security and the lives of the rural poor by reducing crop losses. CABI collates data from plant clinics in the field, including information on pests, diseases, and other intelligence, and has this uploaded to central repositories via scanners. CABI is then able to blend this data with information from its own publications and third-party sources. By utilising its own CAB Thesaurus, it can extract information and store it as semantically structured data. Combining this with other datasets allows the use of advanced analytics to create predictive pest maps and pest risk indexes
- Elsevier's SciVal can analyse huge volumes of citation and other data to create maps of the relative competitive strengths of the research base at a national level
- The journal/database GigaScience is a collaboration between BGI (formerly the Beijing Genomics Institute, and the world's largest sequencing centre) and BioMed Central. It combines journal articles with a huge dataset, and provides data analysis tools and cloud-based computing resources
- Data mining is discussed in more detail below (see *Text and data mining*). One example is Ariadne Genomics (purchased by Elsevier in 2012), which provides services for life science researchers (especially in pharmacos) to mine information from the literature
- Mendeley uses big data technology (Hadoop and MapReduce) to process the volumes of data arising from the interaction between its database of articles (~50 million articles) and users (~1.5 million). This allows it to create statistics and article recommendations, and (with Swets) to create the Mendeley Institutional Edition, which helps librarians understand how their collections are used by their patrons.

4.9. Identity and trust

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).

Although there are number of initiatives to address this issue, the most important of these for STM publishing is ORCID (Open Researcher and Contributor ID).[77] ORCID (the organisation) is a non-profit collaboration involving participants from across the research and scholarly communication worlds (including universities, funders, research organisations, data repositories and professional societies as well as publishers). It provides two services: a registry to create and maintain the ID and associated data for each individual researcher; and an API platform to support system-to-system communication and authentication.

Individuals can obtain their own IDs and manage their record free of charge, and organisations may join to link their records to ORCID, receive updates, and to register their employees and students.

The importance of ORCID goes beyond simple disambiguation of researcher names: a robust method of uniquely identifying individual contributions and networks between researchers will facilitate or improve a host of services, including research analytics (see next section), social media and networking services, and others.

4.10. Research analytics

An emerging market for services built on STM publishing information is that of research analytics: research information management systems linked to analytics tools. The idea is to provide insight for academic institutions and their research managers (e.g. university pro vice-chancellors of research), research funders, and governments into the quality and impact of research programmes. The analytic tools use bibliographic data including citations, building on previous cruder approaches (such as using the Journal Impact Factor), to assess quality of output with more sophisticated data analysis, and integration with current research information systems (CRIS) within institutions.[78] CRIS systems integrate information on the institutions researchers' and research groups' activities and outputs, pulling in information from internal systems, including HR, finance, grant tracking systems, and research project progress reports, as well as external data, in particular bibliographic datasets, and other external proprietary and public datasets (e.g. patents or funding).

The two main companies active in this market are Elsevier, whose SciVal suite of analytic tools (supported by the Scopus database) were complemented by the 2012 acquisition of the Danish CRIS vendor Atira; and Thomson Reuters, whose CRIS Research in View, and InCites analytics suite are supported by the Web of Knowledge database.

Digital Science, the sister company to Nature Publishing Group, also has a presence in this nascent market through its ownership of Symplectic, a UK CRIS vendor.

The main services provided are subscription-based tools and services (e.g. to analyse relative competitive strengths of research programmes, identify collaborator, measure individual/ team research performance, etc.); custom research and analytics[79]; and data licensing for internal analysis.

---

[77] http://about.orcid.org

[78] euroCRIS, the European Organisation for International Research Information, hosts an annual conference and manages the CERIF (Common European Research Information Format) standard: http://www.eurocris.org

[79] An interesting example is the report Elsevier did for the UK Department of Business, Innovation and Skills on the international competitiveness of the UK research base (Elsevier, 2011)

4.11. 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, not just the final boiled-down journal article. Two examples are Useful Chemistry[80] and Cameron Neylon's open laboratory notebook.[81]

Open notebook science has been adopted by a tiny (close to vanishingly small) minority of researchers. We were skeptical in the last edition of this report that this would change quickly; correctly, as it turned out, with little progress in this direction and many of the experiments now mothballed. Most researchers are too concerned about confidentiality and intellectual property rights, about being scooped, and that it would limit their publication options, and more fundamentally whether there is value in sharing at this stage of the research process. One core idea, however, that of greater sharing and reuse of research data has become mainstream, as discussed above.

[80] http://usefulchem.wikispaces.com/

[81] http://biolab.isis.rl.ac.uk/camerons_labblog

5. Conclusions

It is our intention to continue 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 become the dominant means 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 continue to remain largely unchanged.

In particular, the social media and networks whose memberships and use are growing so fast in the general population have yet to make much impact on researchers' professional activities. More recent trends (such as the emergence of Mendeley and ResearchGate, and the growing integration of social features into software and web platforms) do suggest this may start to change, though the impacts seem more likely to be incremental than radical.

Similarly, widespread adoption of smartphones has had limited impact, but the more recent tablet devices are already starting to change the work practices of some physicians and other healthcare professionals, among others.

Researchers' access to scholarly content is at a historic high. Bundling of content and associated consortia licensing model (especially the Big Deal) has continued to deliver unprecedented levels of access, with annual full-text downloads estimated at 2.5 billion, and cost per download at historically low levels (well under $1 per article for many large customers).

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 (and has overtaken the US in some subject disciplines). The expansions of the research bases in India and Brazil are also striking.

The Research4Life programmes (HINARI, AGORA, OARE, and ARDI) have also continued to expand over the last three years, seeing increases in the volume and range of content and in the number of registered institutions and users. The Third World still lags the West in digital infrastructure (and research capacity more generally) but the success of the programmes means that researchers in the poorest countries need not be restricted from accessing the scholarly literature by reason of unaffordable subscriptions.

While the increases in access and associated value delivered by the Big Deal are recognised, it has come under increasing pressure in tight economic circumstances, with libraries seeking greater flexibility and control, more rational pricing models and indeed lower prices. Notwithstanding some noisy renegotiations conducted in public, the model seems likely to retain its importance in the near future at least.

Not everyone is in a position to benefit from access via a Big Deal, of course, and there has been growing interest from both policy-makers and publishers in expanding access to other groups such as small and medium-sized businesses, healthcare professionals not affiliated to research (teaching) hospitals, independent professionals, and other interested parties. For example, the Finch report made recommendations including licensing solutions and walk-in access in public libraries, while the industry has been exploring new models such as article rental and small individual bundles.

The principal focus for improving access has of course been on open access. The number of OA journals has grown substantially over the last three years, (from 4360 to 8115, according to the DOAJ), and more importantly the proportion of OA articles has also grown (to 17% of Scopus content published in 2011, according to one recent estimate, counting both Gold (12%) and delayed (5%) OA articles).

The principal driver of open access has the role of research funders, which have moved on from adopting the now-standard mandates requiring research outputs (i.e. articles) to be made openly available, to growing activism and direct intervention in publishing (e.g. the *eLife* journal).

The OA debate has if anything become more politicised, with competing legislative initiatives in the US (RWA, FRPAA, etc.) and various public consultations in the UK, US and Europe. In the UK at least, the debate post-Finch appears to have shifted decisively towards the Gold model, mainly because of its sustainable financial model; while some parts of Europe may be minded to follow suit, it is not clear that the US is presently inclined to adopt such a centralised, top-down approach to Gold OA.

Turning to Green open access, the numbers of institutional repositories have grown substantially over the last three years (ROAR lists nearly 3000 repositories) and many of these have been able to expand their collections. Despite this expansion and the now widespread funder/institutional mandates, self-archiving as an individual activity in institutional repositories remains of limited interest to much of the scholarly community (outside a few fields where sharing preprints or working papers was already the norm). Subject repositories (PubMed Central, arXiv, SSRN, RePEc etc.) are more attractive to researchers, however, both as authors and (perhaps more so) as readers, and this continues to worry publishers concerned about the impact on subscriptions.

In the last report, we pointed to a lowering of trust in publishers and a debate around the future of scholarly communication (especially open access) characterised by lack of hard evidence and rhetorical argument on both sides. The scorecard here is mixed. Collaborative processes like the Finch report in the UK, the OSTP Roundtable and subsequent America COMPETES initiatives in the US, and the European PEER project gave cause for optimism, while the Research Works Act and continued politicisation of the "public access" agenda did not. Away from the frontline of open access activism – an activity of interest to a vanishingly small fraction of active researchers – publishers and researchers continue to work productively and fruitfully together in time-tested and in innovative ways, though it can be easy to lose sight of this in the heat of the debate.

~

As we move forward, we hope to hold onto some constants in a changing world, notably the core functions of the journal (registration, dissemination, certification and providing an archival record). The core motivations of authors do indeed appear to remain remarkably fixed, in terms need for attribution and recognition, for quality control including peer review, for visibility and the widest reach for their ideas. The success of *PLOS ONE* and other megajournals currently seems to run counter to these fundamentals. However, it is early days for these titles and we may yet see them as driven by other factors, such as being places for rejected papers and new author entrants or simply examples of brand extension.

Other trends have existed long enough to feel like part of the landscape: the relentless growth in volume and complexity of research outputs, and their increasingly data-centric nature. The growth in outputs from emerging markets, especially China, India and Brazil, will continue. The proportion of R&D (though not basic research) funded by industry will continue to rise. Traditional academic acquisition budgets will grow slowly (if at all) in real terms, and buyers of all kinds will seek demonstration of the value of their purchases (through usage and perhaps in more sophisticated ways). There will not be a return to a pre-recession budget environment: this is the "new normal". Emerging markets will continue to provide the best growth opportunities, though perhaps slowing compared to the past three years. And faced with hyperabundance of content, readers will value insights and answers over raw access.

Nonetheless open access will be one of the defining features of the next stage of STM publishing. It will be a complex transition, and will certainly not be completed over the next three years. But the momentum in uptake among authors, publishers and funders is clearly there and will likely increase for a period.

It is still unclear what a stable financially-sustainable arrangement for Gold open access will look like in detail: how precisely will author funding be arranged between research grants, institutional block grants, library and departmental budgets and other sources? Will there really be the substantial additional transitional funding envisaged in the Finch report, particularly for the research-intensive universities? What will become the market rate for publication charges, currently anywhere from $500 to $5000? Will market forces push down APCs, and if so what will be the consequences?

The present demand for open re-use rights (e.g. CC-BY licensing) for OA content will also continue, for the same main reason – demand from research funders – and for the same reason seems likely to be conceded.

Green OA and the role of repositories will remain controversial. This is less the case for institutional repositories (which – despite growth in their numbers and content – remain mainly under-used and disorganised), than for subject repositories, especially PubMed/PMC. The latter is acting more like a publisher, and investing in a platform and related features (e.g. data integration). Embargo lengths will also continue to be the subject of debate.

Open access will also not just be for journal articles: OA books and open educational resources will also become more important in STM areas.

The digital transition will continue. Although all journals have electronic editions legacy print continues, especially in books, and the digital share of the combined STM market revenues is only about 60%.

For research journals, this is largely an accounting and cost issue, as virtually all meaningful use has migrated online, and at least some of the residual print use in clinical, professional and general-purpose journals seems likely to move to new mobile devices.

Accelerating mobile adoption will be a second defining feature of the next three years, particularly for clinical and practitioner areas. It remains early days: mobile devices represent well under 10% of STM platform accesses, probably nearer 5%, though increasing very rapidly.

Use cases are still emerging and developing, with "looking up and keeping up" still dominant, but long-form reading and educational use are growing fast, while interaction with research content is still in its infancy. From the limited evidence to date, tablet use appears more likely to displace print consumption than use of other electronic devices, occupying a new niche alongside desktop and smartphone screens.

Business strategies are equally rudimentary. Mobile is different in ways not yet fully appreciated. The devices are personal, (usually) tied to an individual's credit card or other billing system, and capable of providing highly detailed, richer user metrics. The most prevalent business case has been to add value or convenience to existing subscriptions, perhaps hoping also for increased use and engagement. There has certainly been some experimentation with pricing and with in-app purchase and freemium models, and for clinical journals there are encouraging signs that tablet editions may at last provide a locus for digital advertising. In general, though, publishers lack a coherent near-term strategy for a return on their investments in mobile.

Sticking with the nexus of technology and business models, we would expect publishers with sufficient relevant capabilities to seek to add value to core journal content, including active content, visualisation and analytics, and moving towards workflow tools and systems. These kinds of developments will also favour aggregation (recognising that single publisher outputs are frequently insufficient) and convergence of content types (or at least their greater cohabitation – books, journal articles, conference papers will not lose their separate roles and identities); partnerships and collaborations will thus increase in importance. Developing platforms of this kind will lead to publishers increasingly thinking in terms of services rather than (existing) products, and will also tend to shift the needle a little further from "content" to "software", with (larger) publishers becoming more like technology companies. The ability to add value will also benefit from better, more detailed and more fine-grained understanding of user needs and behaviours.

The defining features of the STM technology strategy will a combination of an open, interoperable platform with open APIs, and widespread deployment of semantic technologies. Semantic enrichment will make content smarter, improving discoverability and use, and will be one way of making content more interactive. It will also enable new products and services, and support internal productivity. Linked data also has the potential to improve discoverability, provide context and support re-use, though uptake is limited to date.

More platforms will feature interactive content using approaches similar to those explored in the Elsevier "article of the future" and similar pilots. Finding the balance between features that genuinely improve the reader experience or enhance researcher productivity, and those that add to the complexity or unfamiliarity of the interface, however, is not easy. It will also require continuing publisher investment in platforms simply to stay current.

Data will play a larger role in STM publishers' lives for two reasons. First, data will become an increasingly central part of research outputs; journals will need to cite and provide access to the underlying data (typically hosted on a data repository), but also to directly incorporate some kinds of data. Data publication is also a possibility.

Second, STM publishers will have access to more data than ever about their users and the usage of their content. Having the capability to make use of this data for analytic and development purposes will provide an advantage.

To conclude, the final defining feature of the coming years will be the accelerating pace of market and technology innovation, even as the core values remain constant. STM publishing can rightly pride itself on its history of innovation, but the game is changing and future revenue growth will more innovation-led, and potentially disruptive innovation more common. In the digital world, user expectations are increasingly set by the leading consumer brands. Publishers will have to come to terms with a faster rate of change, more frequent development and release cycles, and more external innovation.

6. References

Albanese, A (2009). Institutional repositories: Thinking beyond the box. *Library Journal*, *23*(2), 107-116. Retrieved from Ingenta: http://www.libraryjournal.com/article/CA6639327.html

Amin, M, & Mabe, M (2007). Impact factors: Use and abuse. *Perspectives in publishing*. Elsevier Science. (Original work published October, 2000) Retrieved from http://is.gd/Tlp9f6

Anderson, K (2012). The race to the bottom — data, pertussis, roads, fires, and scholarly publishing. *The Scholarly Kitchen Blog.* Retrieved from http://scholarlykitchen.sspnet.org/2012/05/16/the-race-to-the-bottom-data-pertussis-roads-fires-and-scholarly-publishing/

ARL (2011). *ARL annual statistics 2010-11*. Association of Research Libraries. Retrieved from http://www.arl.org/stats/annualsurveys/arlstats/arlstats11.shtml

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

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

Battelle (2011). *Global R&D funding forecast*. Retrieved from http://battelle.org/docs/default-document-library/2012_global_forecast.pdf?sfvrsn=2

Beall, J (2012). Predatory publishers are corrupting open access. *Nature News*, *489*(7415), 179. doi:10.1038/489179a

Beckett, C, & Inger, S (2006). *Self-Archiving and journal subscriptions: Co-existence or competition?*. Publishing Research Consortium. Retrieved from http://www.publishingresearch.net/self_archiving2.htm

Bennett, L (2012). *ALPSP PA potential results of six month embargo*. Retrieved from http://www.publishingresearch.org.uk/documents/ALPSPPApotentialresultsofsixmonthembargofv.pdf

Biosciences Federation (2008). *Surveys of the financial contribution of bioscience societies to UK universities and of researchers' attitudes to open access and self-archiving*. Retrieved from http://www.publishingresearch.net/links.htm

Bird, C (2008). Oxford Journals' adventures in open access. *Learned Publishing*, *21*(3), 200-208. doi:10.1087/095315108X288910

Bird, C (2010). Continued adventures in open access: 2009 perspective. *Learned Publishing*, *23*(2), 107-116. doi:10.1087/20100205

BIS (2011). *Innovation and research: Strategy for growth*. UK Department for Business, Innovation and Skills. Retrieved from http://www.bis.gov.uk/assets/biscore/innovation/docs/i/11-1387-innovation-and-research-strategy-for-growth.pdf

Björk, B-C (2011). A study of innovative features in scholarly open access journals. *Journal of Medical Internet Research*, *13*(4), e115. doi:10.2196/jmir.1802

Björk, B-C, & Solomon, D (2012a). Open access versus subscription journals: A comparison of scientific impact. *BMC Medicine*, *10*, 73. doi:10.1186/1741-7015-10-73

Björk, B-C, & Solomon, D (2012b). Pricing principles used by scholarly open access publishers. *Learned Publishing*, *25*(2), 132-137. doi:10.1087/20120207

Björk, B-C, & Solomon, D (2012c). A study of open access journals using article processing charges. *Journal of the American Society for Information Science and Technology*, *63*(8), 1485-1495. doi:10.1002/asi.22673

Björk, B-C, Roos, A & Lauri, M (2009). Scientific journal publishing: Yearly volume and open access availability. *Information Research*, *14*(1). Retrieved from http://informationr.net/ir/14-1/paper391.html

Björk, B-C, Welling, P, Laakso, M, Majlender, P, Hedlund, T, & Gudnason, G (2010). Open access to the scientific journal literature: Situation 2009. *PloS One*, *5*(6), e11273. doi:10.1371/journal.pone.0011273

Blodget, H., & Cocotas, A (2012). *The future of mobile*. Retrieved from www.businessinsider.com: http://www.businessinsider.com/the-future-of-mobile-deck-2012-3?op=1

Bollen, J, de Sompel, H, Hagberg, A, & Chute, R (2009). A principal component analysis of 39 scientific impact measures. *PloS One*, *4*(6 e6022). doi:10.1371/journal.pone.0006022

Campbell, P (2006). Report of nature's peer review trial. *Peer-to-Peer*. Retrieved from http://blogs.nature.com/peer-to-peer/2006/12/report_of_natures_peer_review_trial.html

Campbell, R, & Wates, E (2009). Journal publishing: Implications for a digital library policy. In D. Baker & W. Evans (Eds.), *Digital library economics.* Chandos.

CIBER (2008). *Information behaviour of the researcher of the future*. Retrieved from http://www.jisc.ac.uk/media/documents/programmes/reppres/gg_final_keynote_11012008.pdf

CIBER Research (2012a). *PEER usage study: Descriptive statistics for the period March to August 2011*. PEER. Retrieved from http://www.peerproject.eu/reports/

CIBER Research (2012b). *PEER usage study: Randomised controlled trial results*. PEER. Retrieved from http://www.peerproject.eu/reports/

CIBER Research Ltd (2011). *The journal usage factor: Exploratory data analysis*. Retrieved from http://ciber-research.eu/CIBER_news-201103.html

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

COPE (2011). *Code of conduct and best practice guidelines for journal editors*. Retrieved from http://publicationethics.org/files/Code_of_conduct_for_journal_editors_Mar11.pdf

COUNTER (2012). *The COUNTER code of practice for usage factors: Draft release 1*. Retrieved from http://www.projectcounter.org/documents/Draft_UF_R1.pdf

Cox, J, & Cox, L (2008). *Scholarly publishing practice: Academic journal publishers' policies and practices in online publishing. Third survey*. ALPSP. Retrieved from http://www.alpsp.org/ngen_public/article.asp?id=0&did=0&aid=2446&st=scholarly%20publishing%20practice&oaid=0

Cox, J, & Cox, L (2010). *Scholarly book publishing practice*. ALPSP. Retrieved from http://www.alpsp.org/Ebusiness/ProductCatalog/Product.aspx?ID=41

Craig, I D, Plume, A M, McVeigh, M E, Pringle, J, & Amin, M (2007). Do open access articles have greater citation impact?:: A critical review of the literature. *Journal of Informetrics*, *1*(3), 239-248. doi:10.1016/j.joi.2007.04.001

Creaser, C, Maynard, S, & White, S (2006). *LISU annual library statistics 2005*. Retrieved from http://www.lboro.ac.uk/microsites/infosci/lisu/pages/publications/als06.html

Crotty, D (2008). Web 2.0 for biologists–are any of the current tools worth using? *Benchmarks blog*. Retrieved from http://cshbenchmarks.wordpress.com/2008/04/03/web-20-for-biologists-are-any-of-the-current-tools-worth-using/

Crotty, D (2010). Rules of thumb for social media in science. *The Scholarly Kitchen blog.* Retrieved from http://scholarlykitchen.sspnet.org/2010/04/05/rules-of-thumb-for-social-media-in-science/

Crotty, D (2012). Are university block grants the right way to fund open access mandates? *The Scholarly Kitchen blog.* Retrieved from http://scholarlykitchen.sspnet.org/2012/09/13/are-university-block-grants-the-right-way-to-fund-open-access-mandates/

Dallmeier-Tiessen, S et al. (2010). *Open access publishing – models and attributes (SOAP)*. Retrieved from http://edoc.mpg.de/478647

Davis, P M (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, 25(7), 2129-34. doi:10.1096/fj.11-183988

Davis, P M, & Walters, W H (2011). The impact of free access to the scientific literature: A review of recent research. *Journal of the Medical Library Association : JMLA*, *99*(3), 208-17. doi: 10.3163/1536-5050.99.3.008

Davis, P M, Lewenstein, B V, Simon, D H, Booth, J G, & Connolly, M J L (2008). Open access publishing, article downloads, and citations: Randomised controlled trial. *BMJ (Clinical Research Ed.)*, *337*, a568. doi:10.1136/bmj.a568

Dickson, D (2012). Developing world gains open access to science research, but hurdles remain. *The Guardian*. Retrieved from http://www.guardian.co.uk/global-development/2012/sep/03/developing-world-open-access-research-hurdles?CMP=twt_gu

Edgecliffe-Johnson, A (2012, July 24). Tablets lift journals' ad revenues. *Financial Times*. Retrieved from http://on.ft.com/MFMldR

Elsevier (2011). *International comparative performance of the UK research base, report for UK Department of Business, Innovation and Skills*. Retrieved from http://www.bis.gov.uk/assets/biscore/science/docs/i/11-p123-international-comparative-performance-uk-research-base-2011.pdf

EPS (2006). *UK scholarly journals: 2006 baseline report*. Research Information Network. Retrieved from http://www.rin.ac.uk/our-work/communicating-and-disseminating-research/uk-scholarly-journals-2006-baseline-report

European Commission (2012a). *Areas of untapped potential for the development of the European Research Area: Preliminary summary and analysis of the response to the ERA framework public consultation*. Retrieved from http://ec.europa.eu/research/era/pdf/era-summary-report-2012_en.pdf

European Commission (2012b). *Online survey on scientific information in the digital age*. Retrieved from http://ec.europa.eu/research/science-society/document_library/pdf_06/survey-on-scientific-information-digital-age_en.pdf

Eysenbach, G (2011). Can tweets predict citations? Metrics of social impact based on Twitter and correlation with traditional metrics of scientific impact. *Journal of Medical Internet Research*, *13*(4), e123. doi:10.2196/jmir.2012

Finch Working Group (2012). *Accessibility, sustainability, excellence: How to expand access to research publication*. Retrieved from http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf

Fry et al. (2009). *PEER behavioural research: Baseline report*. Retrieved from http://www.peerproject.eu/fileadmin/media/reports/Final_revision_-_behavioural_baseline_report_-_20_01_10.pdf

Gargouri, Y A, Hajjem, C, Larivière, V, Gingras, Y, Carr, L, Brody, T, & Harnad, S (2010). Self-Selected or mandated, open access increases citation impact for higher quality research. *PloS One*, *5*(10), e13636. doi:10.1371/journal.pone.0013636

Goble, C (2008). *The future of research (science & technology); presentation to British Library board*. Retrieved from http://www.slideshare.net/dullhunk/the-future-of-research-science-and-technology-presentation

Group of Eight & Outsell (2010). *Building value frameworks for academic libraries – cost–benefit study*. Retrieved from http://go8.edu.au/__documents/go8-committees/go8-librarians/anu_outsell_8_april_2010.pps

Halevi, G (2012). *Research trends: Special issue on big data*. Elsevier. Retrieved from http://www.researchtrends.com/wp-content/uploads/2012/09/Research_Trends_Issue30.pdf

Hargreaves, I (2011). *Digital opportunity: A review of intellectual property and growth*. Retrieved from http://www.ipo.gov.uk/ipreview-finalreport.pdf

Harley, D et al. (2010). *Assessing the future landscape of scholarly communication: An exploration of faculty values and needs in seven disciplines*. Retrieved from escholarship.org: http://escholarship.org/uc/item/15x7385g

Houghton, J, Rasmussen, B, Sheehan, P, Oppenheimer, C, Morris, A, Creaser, C, . . . Gourlay, A (2009). *Economic implications of alternative scholarly publishing models*. JISC. Retrieved from http://www.jisc.ac.uk/publications/reports/2009/economicpublishingmodelsfinalreport.aspx

Inger, S, & Gardner, T (2012). *How readers discover content in scholarly journals.* Renew Training. Retrieved from http://www.renewtraining.com/How-Readers-Discover-Content-in-Scholarly-Journals-summary-edition.pdf

Ithaka S+R (2010). *Faculty survey 2009: Key strategic insights for libraries, publishers, and societies*. Retrieved from http://www.sr.ithaka.org/research-publications/faculty-survey-2009

Ithaka S+R (2011). *Hindawi publishing corporation: Growing an open-access contributor-pays business model*. Retrieved from www.sr.ithaka.org: http://www.ithaka.org/ithaka-s-r/research/case-studies-in-sustainability-2011/SCA_IthakaSR_CaseStudies_Hindawi_2011.pdf

Jefferson, T, Rudin, M, Brodney Folse, S, & Davidoff, F (2007). Editorial peer review for improving the quality of reports of biomedical studies. *Cochrane Database of Systematic Reviews (Online)*, (2), MR000016. doi:10.1002/14651858.MR000016.pub3

JISC (2012). *The value and benefits of text mining*. Retrieved from http://www.jisc.ac.uk/publications/reports/2012/value-and-benefits-of-text-mining.aspx

JISC & British Library (2012). *Researchers of tomorrow: The research behaviour of Generation Y doctoral students*. Retrieved from: http://www.jisc.ac.uk/publications/reports/2012/researchers-of-tomorrow.aspx

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

King, D W, & Tenopir, C (2004). An evidence based assessment of author pays. *Nature Web Focus on access to the literature*. Retrieved from http://www.nature.com/nature/focus/accessdebate/26.html

King, D W, Tenopir, C, & Clarke, M (2006). Measuring total reading of journal articles. *D-Lib Magazine*, *12*(10). doi:10.1045/october2006-king

Kurtz, M. J., Eichhorn, G., Accomazzi, A., Grant, C., Demleitner, M., Henneken, E., & Murray, S. S. (2005). The effect of use and access on citations. Information Processing & Management, 41(6), 1395 - 1402. doi:10.1016/j.ipm.2005.03.010

Laakso, M, & Björk, B-C (2012). Anatomy of open access publishing: A study of longitudinal development and internal structure. *BMC Medicine*, *10*(1), 124. doi:10.1186/1741-7015-10-124

Lawrence, S. (2001). Free online availability substantially increases a paper's impact. *Nature*, 411(6837), 521. doi:10.1038/35079151

Lynch, C (2009). Jim Gray's fourth paradigm and the construction of the scientific record. In T. Hey, S. Tansley, & K. Tolle (Eds.), *The fourth paradigm: Data-Intensive scientific discovery.* Richmond: Microsoft.

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

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

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.

Mabe, M (2010). Scholarly communication: A long view. *New Review of Academic Librarianship*, *16*, 132-144. doi:10.1080/13614533.2010.512242

Mabe, M (2012). Does journal publishing have a future? In R. Campbell, E. Pentz, & I. Borthwick (Eds.), *Academic and professional publishing.* Oxford: Chandos.

Mabe, M, & 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, & Mulligan, A (2011). What journal authors want: Ten years of results from Elsevier's author feedback programme. *New Review of Information Networking*, *16*, 71-89. doi:10.1080/13614576.2011.574495

Marcus, A, & Oransky, I (2012). Bring on the transparency index. *The Scientist*, (12 August 2012). Retrieved from http://the-scientist.com/2012/08/01/bring-on-the-transparency-index/

McKinsey (2011). *Big data: The next frontier for innovation, competition, and productivity*. Retrieved from http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation

Meadows, A, Campbell, R, & Webster, K (2012). The access question. *Learned Publishing, 25*(3), 189-193. doi:10.1087/20120305

Meeker, M (2012). *Internet trends*. Kleiner Perkins Caufield Byers. Retrieved from http://www.kpcb.com/insights/2012-internet-trends

Meyer, C. A (2011). Distinguishing published scholarly content with crossmark. *Learned Publishing*, *24*(2), 87-93. doi:10.1087/20110202

Miller, P (2010). *Linked data horizon scan*. JISC. Retrieved from http://cloudofdata.s3.amazonaws.com/FINAL-201001-LinkedDataHorizonScan.pdf

MIT (2003). *MIT's DSpace experience: A case study*. Retrieved from 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

Morgan, C, Campbell, B, & Teleen, T (2012). The role of the academic journal publisher and open access publishing models. *International Studies Perspectives*, *13*(3), 228-234. doi:10.1111/j.1528-3585.2012.00495.x

Morris, S (2006). Data about publishing. *ALPSP Alert*, (112), 8. Retrieved from www.alpsp.org: http://www.alpsp.org/Ebusiness/ResearchPublications/ALPSPAlert/ALERTARCHIVE.aspx

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

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

Mulligan, A, & Mabe, M (2011). The effect of the internet on researcher motivations, behaviour and attitudes. *Journal of Documentation*, *67*, 290-311. doi:10.1108/00220411111109485

Nature (2012). Gold in the text? (Editorial). *Nature*, *483*(7388), 124. doi:10.1038/483124a

Nicholas, D, & Clark, D (2012). 'Reading' in the digital environment. *Learned Publishing, 25*(2), 93-98. doi:10.1087/20120203

NISO (2008). *Journal article versions (JAV): Recommendations of the NISO/ALPSP JAV technical working group (April 2008), NISO-RP-8-2008*. Retrieved from http://www.niso.org/publications/rp/RP-8-2008.pdf

NSF (2012). *Science & engineering indicators 2012*. Retrieved from http://www.nsf.gov/statistics/seind12/

OECD (2011). *Main science and technology indicators*. OECD. Retrieved from www.oecd.org/sti/msti

Outsell (2009). *Open access primer (public version)*. Retrieved from http://www.outsellinc.com/images/user_assets/Outsell_Mkt_Report_14dec09_Open_Access_Primer_Public.pdf

Outsell (2011). *Scientific, technical & medical information: 2011 market forecast and trends report*. Retrieved from www.outsellinc.com: http://www.outsellinc.com/store/products/1040

Outsell (2012a). *Evolution of the STM publishing platform – an industry overview and roadmap*. Outsell. Retrieved from http://www.outsellinc.com/store/products/1083

Outsell (2012b). *Five technologies to watch 2012-2013*. Retrieved from http://www.outsellinc.com/store/products/1050

Outsell (2012c). *Mobile in STM: Case studies of accelerating change*. Retrieved from http://www.outsellinc.com/store/products/1096

Outsell (2012d). *STM 2012 market size, share, forecast & trend report*. Retrieved from www.outsellinc.com: http://www.outsellinc.com/products/index/1107-scientific-technical-medical-information-2012-market-size-share-forecast-and-trend-report

Outsell (2012e). *STM e-books: 2012 market size, share, and forecast*. Retrieved from www.outsellinc.com: http://www.outsellinc.com/store/products/1100

Outsell (2012f). *STM in China: 2012 market size, share & forecast*. Retrieved from www.outsellinc.com: http://www.outsellinc.com/store/products/1091

Pearson (2012). Pearson launches higher education college. Retrieved from www.pearson.com: http://www.pearson.com/news/2012/august/pearson-launches-higher-education-college.html?article=true

PLOS (2012). *PLOS progress update 2011-2012*. Retrieved from http://www.plos.org/wp-content/uploads/2012/09/2011_PLOS_Progress_Update_hi.pdf

Priem, J (2010). Altmetrics: a manifesto. Retrieved from altmetrics.org: http://altmetrics.org/manifesto/

Procter, R, Williams, R, Stewart, J, Poschen, M, Snee, H, Voss, A, & Asgari-Targhi, M (2010). Adoption and use of Web 2.0 in scholarly communications. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *368*(1926), 4039-4056. doi:10.1098/rsta.2010.0155

Pulverer, B (2010). Transparency showcases strength of peer review. *Nature*, *468*(7320), 29-31. doi:10.1038/468029a

RCUK (2012). *Research Councils UK policy on access to research outputs*. Retrieved from http://www.rcuk.ac.uk/documents/documents/RCUK%20_Policy_on_Access_to_Research_Outputs.pdf

Renear, A H, & Palmer, C L (2009). Strategic reading, ontologies, and the future of scientific publishing. *Science*, *325*(5942), 828. doi:10.1126/science.1157784

Rieh, S, Markey, K, St Jean, B, Yakel, E, & Kim, J (2007). Census of institutional repositories in the US. *D-Lib Magazine*, *13*(11/12). Retrieved from http://www.dlib.org/dlib/november07/rieh/11rieh.html

RIN (2008). *Activities, costs and funding flows in the scholarly communications system in the UK*. Research Information Network. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/activities-costs-and-funding-flows-scholarly-commu

RIN (2009a). *Communicating knowledge: How and why UK researchers publish and disseminate their findings*. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/communicating-knowledge-how-and-why-researchers-pu

RIN (2009b). *E-journals: Their use, value and impact*. Research Information Network. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/e-journals-their-use-value-and-impact

RIN (2009c). *Patterns of information use and exchange: Case studies of researchers in the life sciences*. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/using-and-accessing-information-resources/patterns-information-use-and-exchange-case-studie

RIN (2010). *If you build it, will they come? How researchers perceive and use Web 2.0*. London: Research Information Network. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/use-and-relevance-web-20-researchers

RIN (2011a). *Access to scholarly content: Gaps and barriers*. Retrieved from http://rinarchive.jisc-collections.ac.uk/node/1172

RIN (2011b). *E-journals: Their use, value and impact – final report*. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/communicating-and-disseminating-research/e-journals-their-use-value-and-impact

RIN (2011c). *Heading for the open road: Costs and benefits of transitions in scholarly communications*. Retrieved from http://www.rin.ac.uk/system/files/attachments/Dynamics_of_transition_report_for_screen.pdf

RIN (2011d). *Reinventing research? Information practices in the humanities*. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/using-and-accessing-information-resources/information-use-case-studies-humanities

RIN (2012). *Physical sciences case studies: Information use and discovery*. Retrieved from http://rinarchive.jisc-collections.ac.uk/our-work/using-and-accessing-information-resources/physical-sciences-case-studies-use-and-discovery-

Romary, L, & Armbruster, C (2009). *Beyond institutional repositories*. Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1425692

Rowlands, I, & Nicholas, D (2005). *New journal publishing models*. CIBER. Retrieved from http://www.ucl.ac.uk/ciber/ciber_2005_survey_final.pdf

Rowlands, I, Nicholas, D, & Huntingdon, P (2004). *Scholarly communication in the digital environment: What do authors want,*. CIBER. Retrieved from http://www.ucl.ac.uk/ciber/ciber-pa-report.pdf

Royal Society (1995). *Peer review – an assessment of recent developments*. London: The Royal Society.

Royal Society (2011). *Knowledge, networks and nations: Global scientific collaboration in the 21st century*. The Royal Society.

Salo, D (2008). Innkeeper at the Roach Motel. *Library Trends*, *57*(2), 98-123. doi:10.1353/lib.0.0031

Samuelson, P (2012). Reforming copyright is possible. *The Chronicle of Higher Education.*

ScienceWatch (2012). *Multiauthor papers: Onward and upward*. Thomson Reuters. Retrieved from http://sciencewatch.com/multiauthor-papers-onward-and-upward

Seglen, P O (1992). The skewness of science. *Journal of the American Society for Information Science*, *43*(9), 628-638. doi:10.1002/(SICI)1097-4571(199210)43:9%3C628::AID-ASI5%3E3.0.CO;2-0

Sense About Science (2009). *Peer review survey 2009: Preliminary findings*. Sense About Science. Retrieved from http://www.senseaboutscience.org/pages/peer-review.html

Simba (2011). *Global STM publishing 2010-2011*. Retrieved from https://www.simbainformation.com/Global-STM-Publishing-6059305/

Smit, E & van der Graaf, M (2011). *Journal article mining: A research study into practices, policies, plans ... and promises*. Publishing Research Consortium. Retrieved from http://www.publishingresearch.org.uk/documents/PRCSmitJAMreport2.30June13.pdf

SOAP (2010). *Open access publishing – models and attributes*. Study on Open Access Publishing. Retrieved from edoc.mpg.de: http://edoc.mpg.de/478647

Sparks, S (2005). *JISC disciplinary differences report*. JISC/Rightscom. Retrieved from http://www.jisc.ac.uk/media/documents/themes/infoenvironment/disciplinarydifferencesneeds.pdf

STM Future Lab Committee (2011). *STM publishing industry: 2012 technology trend watch*. Retrieved from http://www.stm-assoc.org/future-lab-trend-watch-2012/

Swan, A, & Houghton, J (2012). *Going for gold? The costs and benefits of gold open access for UK research institutions: Further economic modelling*. UK Open Access Implementation Group. Retrieved from http://repository.jisc.ac.uk/610/

Tenopir, C (2000). *Towards electronic journals: Realities for scientists, librarians, and publishers*. Washington, DC: Special Libraries Association. Retrieved from http://works.bepress.com/carol_tenopir/7/

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

Tenopir, C (2010). *Research publication characteristics and their relative values*. Publishing Research Consortium. Retrieved from www.publishingresearch.net/projects.htm

Tenopir, C, King, D W, Edwards, S, & Wu, L (2009). Electronic journals and changes in scholarly article seeking and reading patterns. *Aslib Proceedings*, 61(1), 5-32. doi:10.1108/00012530910932267

Tenopir, C, Volentine, R, & King, D W (2012). Scholarly reading and the value of academic library collections: Results of a study in six UK universities. *Insights: The UKSG Journal*, *25*(2), 130-149. doi:10.1629/2048-7754.25.2.130

Terry, R (2005). Funding the way to open access. *PLoS Biology*, *3*(3), e97. doi:10.1371/journal.pbio.0030097

Thomson Reuters (2012). *Global publishing – changes in submission trends and the impact on scholarly publishers*. Retrieved from http://scholarone.com/media/pdf/GlobalPublishing_WP.pdf

Thorn, S, Morris, S, & Fraser, R (2009). Learned societies and open access: Key results from surveys of bioscience societies and researchers. *Serials: The Journal for the Serials Community*, *22*(1), 39-48. doi:http://dx.doi.org/10.1629/2239

UNESCO (2010). *UNESCO science report 2010*. Retrieved from http://www.unesco.org/new/en/natural-sciences/science-technology/prospective-studies/unesco-science-report/unesco-science-report-2010/

University of California (2007). *Faculty attitudes and behaviors regarding scholarly communication: Survey findings from the University of California*. Retrieved from http://osc.universityofcalifornia.edu/responses/materials/OSC-survey-full-20070828.pdf

Van Noorden, R (2011). Science publishing: The trouble with retractions. *Nature*, *478*, 26-28. doi:10.1038/478026a

Van Noorden, R (2012a). Britain aims for broad open access. *Nature*, *486*(7403), 302. doi:doi:10.1038/486302a

Van Noorden, R (2012b). Journal offers flat fee for 'all you can publish'. *Nature News*, *486*(7402), 166. doi:10.1038/486166a

Van Orsdel, L C (2008). The state of scholarly communications. *The Serials Librarian*, *52*(1-2), 191–209. doi:10.1300/J123v52n01_16

Van Orsdel, L C., & Born, K (2009, April 15). Reality bites: Periodicals price survey 2009. *Library Journal.* Retrieved from http://www.libraryjournal.com/article/CA6651248.html?q=reality%20bites%20van%20orsdel

de Vrieze, J (2012). Horizon 2020: A €80 billion battlefield for open access. *Science Insider.* Retrieved from http://news.sciencemag.org/scienceinsider/2012/05/horizon-2020-a-80-billion-battle.html

Wallace, J (2012). *PEER project: Final report*. PEER. Retrieved from www.peerproject.eu: http://www.peerproject.eu/reports/

WAME (2009). *Conflict of interest in peer-reviewed medical journals*. World Association of Medical Editors. Retrieved from http://www.wame.org/conflict-of-interest-in-peer-reviewed-medical-journals

Ware, M (2008). *Peer review: Benefits, perceptions and alternatives*. Publishing Research Consortium. Retrieved from http://www.publishingresearch.net/documents/PRCsummary4Warefinal.pdf

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

Ware, M (2011). Peer review: Recent experience and future directions. *New Review of Information Networking*, *16*(1), 23–53. doi:10.1080/13614576.2011.566812

Ware, M, & Monkman, M (2008). *Peer review in scholarly journals: Perspective of the scholarly community—an international study*. Publishing Research Consortium. Retrieved from http://www.publishingresearch.net/PeerReview.htm

Willinsky, J (2003). The nine flavours of open access scholarly publishing. *Journal of Postgraduate Medicine*, *49*(3), 263-267