STM

The global voice of scholarly publishing

The STM Report

An overview of scientific and scholarly journal publishing

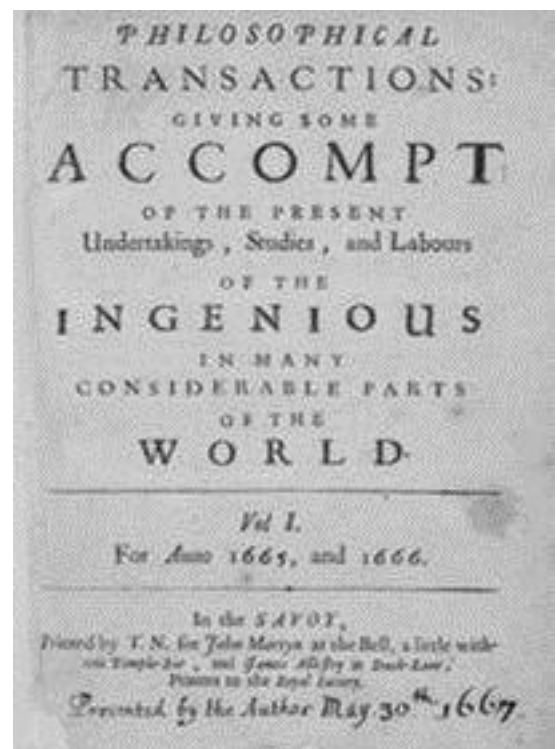Celebrating the 350th anniversary of journal publishing

Mark Ware

Mark Ware Consulting

Michael Mabe

International Association of Scientific, Technical and Medical Publishers

Fourth Edition

March 2015

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.

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 12).

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 16).

The STM market

3. The annual revenues generated from English-language STM journal publishing are estimated at about $10 billion in 2013, (up from $8 billion in 2008, representing a CAGR of about 4.5%), within a broader STM information publishing market worth some $25.2 billion. About 55% of global STM revenues (including non-journal STM products) come from the USA, 28% from Europe/Middle East, 14% from Asia/Pacific and 4% from the rest of the world (page 23).

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 24).

5. Although this report focuses primarily on journals, the STM book market (worth about $5 billion annually) is evolving rapidly in a transition to digital publishing. Ebooks made up about 17% of the market in 2012 but are growing much faster than STM books and than the STM market as a whole (page 24).

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 about 650 publishers producing around 11,550 journals, that is, about 50% of the total journal output by title. Of these, some 480 publishers (73%) and about 2300 journals (20%) are not-for-profit (page 45).

7. There were about 28,100 active scholarly peer-reviewed English-language journals in late 2014 (plus a further 6450 non-English-language journals), collectively publishing about 2.5 million articles a year. The number of articles published each year and the number of journals have both grown steadily for over two centuries, by about 3% and 3.5% per year respectively, though there are some indications that growth has accelerated in recent years. 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 7 and 9 million, depending on definition, although only about 20% of these are repeat authors (pages 27).

8. The USA continues to dominates the global output of research papers with a share of about 23% 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 17% of global output. It is followed by the United Kingdom (7%), Germany (6%), Japan (6%), and France (4%). The rank order changes for citations, however, with the US strongly in the lead with 36% and China at 11th place with 6% (page 38).

Research behaviour and motivation

9. Despite a transformation in the way journals are published, researchers' core motivations for publishing appear largely unchanged, focused on securing funding and furthering the author's career (page 69).

10. Reading patterns are changing, however, with researchers reading more, averaging 270 articles per year, depending on discipline (more in medicine and science, fewer in humanities and social sciences), 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), though new scientific social networks may change this. Researchers spend very little time on average on publisher web sites, "bouncing" in and out and collecting what they need for later reference (page 52).

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 45).

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. Cascade review (transferring articles between journals with reviewer reports) and even journal-independent ("portable") peer review are establishing a small foothold. 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 some interest in its use as a complement to it (for example, the launch of PubMed Commons is notable in lending the credibility of PubMed to post-publication review). There is similar interest in "altmetrics" as a potentially useful complement to review and in other measures of impact. A new technology of potential interest for post-publication review is open annotation, which uses a new web standard to allow citable comments to be layered over any website (page 47).

13. Interest in research and publication ethics continues to be sustained, 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 73).

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 some parallel access for some general journals, including society membership journals, and in some fields (e.g. humanities and some practitioner fields). The number of established research (i.e. non-practitioner) journals dropping their print editions looks likely to accelerate over the coming few years (page 30).

15. Social networks and other social media have yet to make the impact on scholarly communication that they have done on the wider consumer web. The main barriers to greater use have been the lack of clearly compelling benefits to outweigh the real costs (e.g. in time) of adoption. 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 to become more important given the rapid growth in membership of the newer scientific social networks (Academia, Mendeley, ResearchGate), trends in general population, and the integration of social features into publishing platforms and other software (page 72; 134).

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 still account for less than 10% of most STM platform's traffic as of 2014 (though significantly higher in some fields such as clinical medicine) – but this is changing. 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. For the majority of researchers, though, it seems that "real work" still gets done at the laptop or PC (page 24; 30; 139).

17. The explosion of data-intensive research is challenging publishers to create new solutions to link publications to research data (and *vice versa*), to facilitate data mining and to manage the dataset as a potential unit of publication. Change continues to be rapid, with new leadership and coordination from the Research Data Alliance (launched 2013): most 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 cited, visible and accessible. Meanwhile publishers have responded by working closely with many of the community-led projects; by developing data deposit and sharing policies for journals, and introducing data citation policies; by linking or incorporating data; by launching some pioneering data journals and services; by the development of data discovery services such as Thomson Reuters' Data Citation Index (page 138).

18. 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. A number

of initiatives (e.g. CrossRef's TDM tools, PLSclear, and Copyright Clearance Center's and Infotrieve's services to aggregate article content for TDM) have now emerged in terms of the licensing framework (e.g. the STM standard licence clause) and procedures (rights clearance, e.g. PLSclear, CrossRef's TDM tools), content access and aggregation for TDM (e.g. CrossRef's TDM tools, and Copyright Clearance Center's and Infotrieve's aggregation services), and in terms of standardising content formats (e.g. CCC's pilot service), but this area remains in its infancy (page 80; 146).

19. The growing importance to funders and institutions of research assessment and metrics has been reflected in the growth of information services such as research analytics built around the analysis of metadata (usage, citations, etc.), and the growth of a new software services such as CRIS tools (Current Research Information Systems) (page 150).

20. 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 143).

21. While publishers have always provided services such as peer review and copy-editing, increased competition for authors, globalisation of research, and new enabling technologies are driving an expansion of author services and greater focus on improving the author experience. One possibly emerging area is that of online collaborative writing tools: a number of start-ups have developed services and some large publishers are reported to be exploring this area (page 153).

Business models and publishing costs

22. Aggregation on both the supply and demand sides has become the norm, with journals sold in packages to library consortia (see below for open access). Similar models have also emerged for ebook collections (page 19).

23. While the value of the "Big Deal" and similar discounted packages in widening researchers' access to journals and simultaneously reducing average unit costs 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 continue to appear sufficient for the model to retain its importance for some time, though perhaps evolving in scope (e.g. the bundling or offsetting of open access charges) and in new pricing models (page 21; 69).

24. Researchers' access to scholarly content is at an historic high. Bundling of content and the associated consortia licensing model 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). Various surveys have shown that academic researchers rate their access to journals as good or very good, and report that their 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 immediately accessible (page 83).

25. 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 86).

26. 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 84).

27. There is continued 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 have included increased funding for national licences to extend and rationalise cover; walk-in access via public libraries (a national scheme was piloted in the UK in 2014); the development of licences for sectors such as central and local government, the voluntary sector, and businesses (page 84).

28. Average publishing costs per article vary substantially depending on a range of factors including rejection rate (which drives peer review costs), range and type of content, levels of editorial services, and others. The average 2010 cost of publishing an article in a subscription-based journal with print and electronic editions was estimated by CEPA to be around £3095 (excluding non-cash peer review costs). The potential for open access to effect cost savings has been much discussed, but the emergence of pure-play open access journal publishers allows examples of average article costs to be inferred from their financial statements. These range from $290 (Hindawi), through $1088 (PLOS), up to a significantly higher figure for eLife (page 66).

Open access

29. Journal publishing has become more diverse and potentially more competitive with the emergence of new business models. Open access makes original research freely accessible on the web, free of most copyright and licensing restrictions on reuse. There are three approaches: open access publishing ("Gold", including full and hybrid OA journals), delayed free access, and self-archiving ("Green") (page 88).

30. There are around 10,090 (7245 published in English) fully open access journals listed on the Directory of Open Access Journals. OA titles are still somewhat less likely than other titles to appear in selective A&I databases such as Scopus or Web of Science, partly reflecting their more recently establishment, and are (with some notable exceptions) smaller on average than other journals. Consequently the proportion of the 2 million articles published per year that is open access is substantially lower than the proportion of journal titles. Recent estimates place the proportion of articles published in open access journals at about 12% (while OA journals make up about 26-29% of all journals), with 5% more available via delayed access on the publisher's website, and a further 10-12% via self-archived copies (page 31; 98)

31. Gold open access is sometimes taken as synonymous with the article publication charge (APC) business model, but strictly speaking simply refers to journals offering immediate open access on publication. A substantial fraction of the Gold OA articles indexed by Scopus, however, do not involve APCs but use other models (e.g. institutional support or

sponsorship). The APC model itself has become more complicated, with variable APCs (e.g. based on length), discounts, prepayments and institutional membership schemes, offsetting and bundling arrangements for hybrid publications, an individual membership scheme, and so on (page 91; 93).

32. Gold open access based on APCs has a number of potential advantages. It would scale with the growth in research outputs, there are potential system-wide savings, and reuse is simplified. Research funders generally reimburse publication charges, but even with broad funder support the details regarding the funding arrangements within universities it remain to be fully worked out. 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 90; 123).

33. 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. The number of megajournals continues to grow: Table 10 lists about fifty examples (page 99).

34. 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). These policies 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 88).

35. Green OA and the role of repositories remain controversial. This is perhaps less the case for institutional repositories, 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 longer usage half-lives). Green proponents 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; they have little choice about needing to accept submissions from large funders such as NIH, but there has been some tightening of publishers' Green policies (page 102).

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). The scholarly communication supply chain has traditionally been seen as comprising two main players that serve the needs of the scholarly community represented by academics, as authors and readers, and their funders and host institutions; namely, 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 have also played an important and probably growing role. Scholarly communications is evolving, however, and the research funders are increasingly becoming one of the most important parts of the system with the growth of open access and related developments, and other players are playing increasingly important roles (notably data repositories, and software and services providers).

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

---
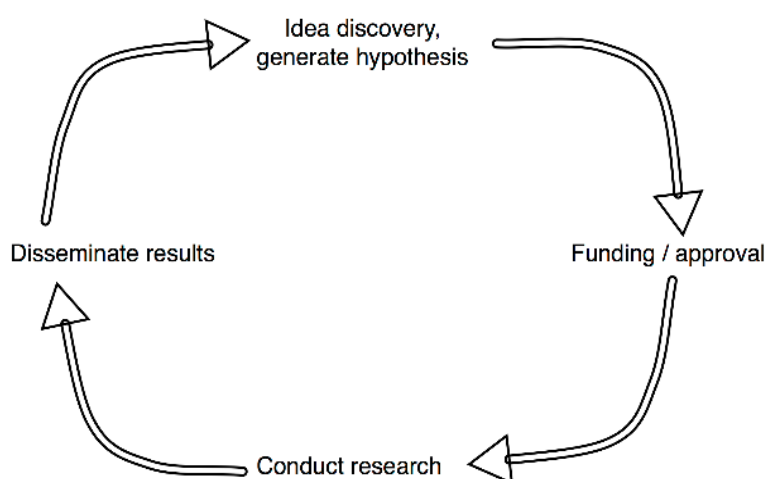
1 STM" is an abbreviation for scientific, technical and medical but has several different meanings. It can be a model of publishing, in which case it includes social sciences and the arts and humanities. It is sometimes used to describe scientific journals. It is also the name of association of publishers ("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

Scholarly communication thus 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 increasingly 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 Inger & Gardner, 2013) 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 exceptions are the wiki and social media. The wiki (and similar online spaces such as forums and dedicated discussion/collaboration platforms that might be thought of as the descendants of the wiki) provides a practical means of facilitating many-to-many written communication which does therefore offer something entirely without parallel in the offline world. Social media, and in particular Twitter, also offers real-time, many-to-many discussion that may also spread simultaneously across multiple web platforms. This perspective may be helpful in balancing some of the techno-centric "solutionist" views that assert that the introduction of digital and web technologies will automatically lead to revolutionary change in scholarly communication (see also *Authors' behaviour, perceptions and attitudes*).

Figure 2: Formal and informal types of scholarly communication
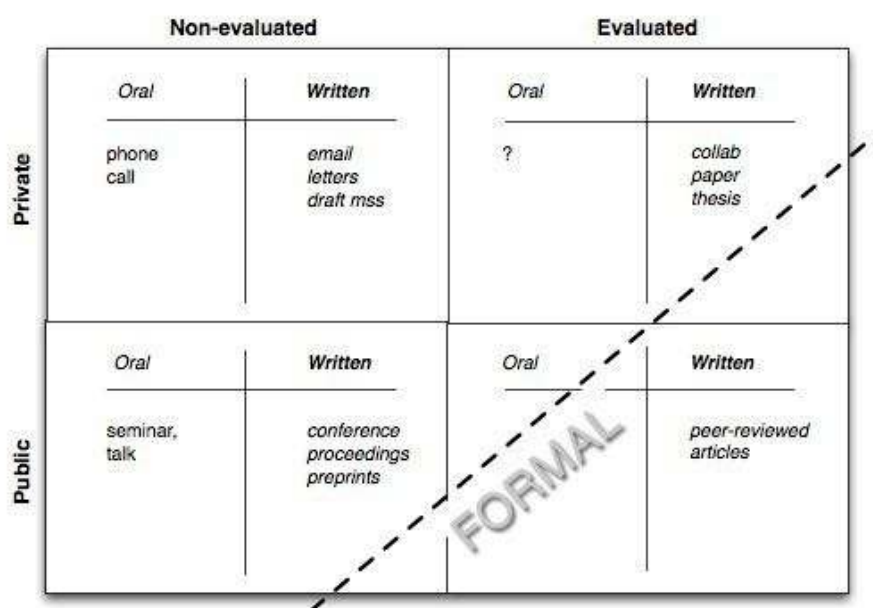
Table 1: Modes of communication

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 and data-sharing 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.

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 (and increasingly to related material, such as datasets). Alternatively this can be seen as part of the dissemination function.

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, tagging and semantic enrichment, etc.).

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, though consortial licensing is reducing this role for the larger publishers – 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, particularly for open access, though access for subscription-based journals is still primarily managed by the library.

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 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 (see *Peer review*).

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, so long as users and the market continue to demand it) printing and binding at least some of the journals on their lists.

- **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, FundRef, 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.

A more elaborate description of the publisher's role was provided in the blog post *82 Things Publishers Do (2014 Edition)* (Anderson 2014a). This is essentially a more granular breakdown of these same functions, but Anderson also emphasises the need for a longterm sustainable model, which in turns requires the generation of a surplus (e.g. for reinvestment in new technology platforms).

Cliff Morgan and coauthors reviewed the role of the publisher in the context of open access developments and suggested a similar set of activities will continue to be required, and estimated 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
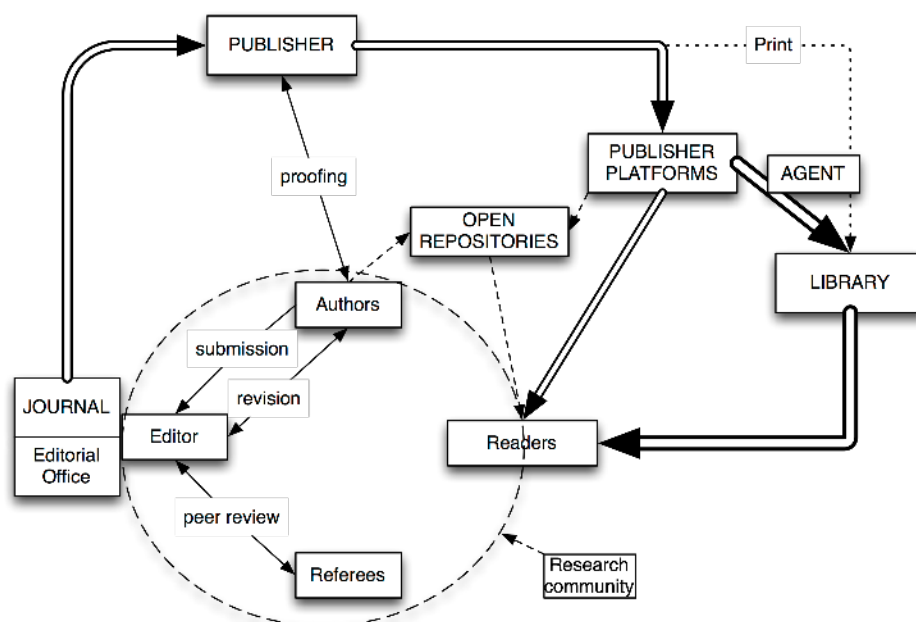
Versions of articles

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), though this has not prevented their widespread continued use.

The CrossRef organisation introduced the CrossMark service in April 2012 to identify (among other things) the version of record (Meyer 2011). There is a visible kitemark that identifies it to the human reader. There is also 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.

Another potential issue is the possible emergence of journals adopting a more fluid notion of the journal article. For instance, the journal *Faculty1000Research* encourages authors to publish (multiple) revised versions of their article, with all versions of an article are linked and independently citable.

2.3.   Sales channels and models

Subscription- or licence-based 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* or *Science* 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 a declining part of the market, especially for larger publishers, as increasingly journals are sold as bundles of titles, either directly to libraries or to library consortia.

While print editions continue (see below), the majority of publishers offer single journal subscriptions in three models: print only, online only, and print and online combined. Most publishers charge less for online-only than print-only, and charge extra for online access to a print subscription.

Individual article sales are growing in popularity (albeit from a very small base), with the proportion of publishers offering increasing from 65% in 2003 to 83% in 2012 (Inger & Gardner, 2013), and related models such as article rental and article packs becoming more common. More important, however, are sales of licences to bundles or collections of journals. Sales of archives (backfiles) are also important, with many libraries keen to acquire the physical files for local storage for a one-off price (with or without a maintenance charge), as well as licensed access models.

Lastly, a key part of the sales model concerns "perpetual access", namely the right of the subscriber to access the previously subscribed-to content after termination of the current subscription. The majority of publishers offer perpetual access, though in some cases there are additional charges. Large publishers are more likely (91%) to offer perpetual access than small publishers (50%) (Inger & Gardner, 2013).

Subscription agents

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, 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 25 agent members but the number of agents has been declining in recent years (the ASA membership was reported at 30 in the 2012 edition of this report, and 40 in the 2009 edition), primarily due to mergers and acquisitions with the industry and the lack of new entrants. (Or more dramatically via business failure, like that of Swets in 2014.) A key reason is the increasing disintermediation of the traditional agent function brought about by move to electronic publishing and in particular the rise of consortia sales. The larger subscription agents are consequently reinventing themselves, for instance as aggregators, publishers, and providers of analytics services. It has been argued that these changes will favour large over small publishers (and thus favour increasing publisher consolidation), because as the former withdraw their high-volume business (replacing it with direct sales to consortia), agents' costs will fall increasingly on the remaining small publishers (Aspesi 2014).

Content bundles

With the rise of electronic publishing, sales of individual journal subscriptions have fallen as a proportion of total sales in favour of **bundles**. 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. 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. Small publishers are more likely to participate in

---

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

multi-publisher bundles such as the ALPSP Learned Journal Collection, BioOne or Project MUSE. A 2012 survey of its library members by the Association of Research Libraries reported that well over 90% of libraries purchased content from the larger publishers as bundles (Strieb & Blixrud, 2013). The nature of bundles had changed compared to 2006, however, with a lower proportion of "all titles" bundles (attributed to pruning required during the 2008/09 recession).

This ARL survey also found that the large majority of licences were still priced on the historic print (sometimes called "prior print") model, similar to the findings of Cox in 2008. In the historic print model, 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. The ARL survey found almost no evidence of uptake of usage-based pricing among its members in 2012; this was echoed in (Inger & Gardner, 2013), which reported that it was still in its infancy and very few (~10%) publishers reported having having this model

- tiered pricing based on a classification of institutions by size; Inger & Gardner (2013) found this was the most popular pricing mechanism after historic print, with size most frequently defined by number of sites. (Classification schemes such as Carnegie or JISC were not popular because they only cover a fraction of most publishers' market.)

- differential pricing based on customer type (e.g. hospital, academic, corporate)

- pricing based on the number of simultaneous users; this model has existed for many years for databases

- an aggregate flat-rate price for all the titles in the bundle.

Despite the apparent stasis in pricing models,[3] industry discussions suggest that there will be more publishers moving away from historic pricing in the coming years.

A key issue for libraries is whether the publisher's licence term for bundles allows cancellations; Cox (2008) 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%). The ARL survey indicates that publishers were in practice more flexible when the economic downturn led to severe constraints in library budgets.

Library consortia

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

---

3 It is worth noting that the historic print model 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.

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[4] lists over 400 consortia in 100+ countries, representing over 26,500 individual libraries[5]; of these, about 350 are responsible for licensing content. The International Coalition of Library Consortia[6] 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 number of individual libraries covered by consortia is of the order of 5000. The ARL survey (Strieb & Blixrud, 2013) reported that the role of the consortia remained central in 2012, with 61–97% (depending on publisher) of reported contracts made via a consortium. According to the last two ALPSP *Scholarly Journals Publishing Practice* reports (Cox & Cox, 2008; Inger & Gardner, 2013), about 90% of larger publishers actively market to consortia, and about half of all 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 offered by only about half of publishers. Cancellation terms are as previously covered for bundles. Newer terms that are starting to become important in bundles include "author-rights" clauses (typically covering self-archiving rights for authors at the licensing institutions) and non-disclosure agreements.

Library system vendors

Library system vendors[7] 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[8] and KBART (Knowledge Bases And Related Tools).[9] Uniquely identifying institutions is important for publishers: the Identify service from Ringgold[10] is the leading commercial service here, with a database of over 400,000 institutions and consortia, while a free dataset OrgRef was launched in DataSalon in 2014.[11]

Vendors have invested substantially in *discovery* tools, including so-called *web-scale discovery*, of which the leading examples are EBSCO Discovery, Proquest Summon, Ex Libris Primo, and OCLC WorldCat Discovery. Collectively these services are installed in approaching 10,000 customer sites. These services provide a simplified search interface (popular with users accustomed to the Google interface), which allows users to discover content from the full range of library holdings (including A&I databases) and web resources in a single

---

4 http://www.ringgold.com/cdo

5 Growth can be indicated by the earlier editions of this report, which recorded 338 active consortia in 2008, up from 164 in 2003, though with relatively small change since 2012

6 http://icolc.net

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

8 http://www.editeur.org/8/ONIX/

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

10 http://www.openrfp.com/pages/identify.html

11 http://www.orgref.org/web/about.htm

search, providing fast results, with relevancy ranking, faceted results browsing, content suggestions, full-text linking, and a variety of social and research-management features. In addition, there are detailed metrics and reporting for institutional use. The popularity of the combination of simplified interface with powerful results is shown by the fact that 81% of libraries adopting discovery services have made them their default search tool. Tests conducted by libraries have shown that use of discovery services increases patron satisfaction and increases use of subscribed-to library content (Somerville & Conrad, 2014; Outsell 2014e; Levine-Clark, McDonald, & Price, 2014).

All these services, like other library-based search tools, find themselves in competition with Google Scholar (and to a lesser extent, Microsoft Academic Search), which offers integration with library holdings, citation links Scholar Metrics, and other features in addition to its signature search capabilities (Inger & Gardner, 2012; Van Noorden 2014b). Indeed, to quote (Somerville & Conrad, 2014), "Google Scholar Library, which enables saving articles directly from the search page in Google Scholar, organizing them by topic, and searching full-text documents within a personal MyLibrary space, is setting heightened expectations for workflow integration solutions". Google Scholar does not disclose the list of journals covered, but is independently estimated to index between 100 and 160 million scholarly documents (Khabsa & Giles, 2014; Orduña-Malea, Ayllón, Martín-Martín, & López-Cózar, 2014).

2.4. *Journal economics and market size*

Journal economics & market size

The total size of the global STM market in 2013 (including journals, books, technical information and standards, databases and tools, and medical communications and some related areas) was estimated by Outsell at $25.2 billion[12] (Outsell 2014c). The market is predicted to grow at about 4% annually through 2017.

Within this overall market for STM information, Outsell have previously estimated the proportion of revenues from journals at about 40%, and from books at 16% (Outsell 2012c). The 2013 market can also be divided into scientific/technical information at $12.2 billion and medical at $13.0 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). The proportion due to advertising has likely fallen since these estimates were made, since advertising revenues have been weak (especially in the immediate aftermath of the 2008/09 recession), while subscriptions and licence income has continued to grow.

By geographical market, Outsell estimates about 55% of global STM revenues (including non-journal STM products) come from the USA, 28% from the EMEA region, 14% from Asia/Pacific and 4% from the rest of the world (principally the Americas excluding USA) (Outsell 2014c). These proportions probably overstate the importance of the USA market for journals alone.

---

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

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). Simba's estimate for the 2013 scientific and technical (i.e. excluding medical) market was $10.7 billion, up just 0.2% on the previous year, though compound annual growth 2010-2012 was 2.3%. It described the S&T books market as being in terminal decline, with scientific and technical book sales falling 4.2% to $2.7 billion in 2012 (Simba 2013b).

The open access segment of the market continues to grow much faster than the market as a whole but remains small in revenue terms. In 2013, Outsell estimated the OA journals market to be worth $128 million and forecast to grow to $336 million by 2015 (Outsell 2013). A more recent estimate from Simba put it at $299 in 2014, and at 2.3% of global STM journal sales in 2013 (i.e. roughly $242 million), up 32% on 2012; Simba forecast OA revenues would triple between 2011 and 2017 (thus reaching $440 million), compared to annual growth of just 1–2% for the market as a whole (Simba 2014).

A rough estimate is that the industry employs an estimated 110,000 people globally, of which about 40% are employed in the EU. An independent survey funded by the AAP's Professional & Scholarly Publishing division in 2014 estimated a total of over 38,000 employed in the USA by over 350 publishers at a payroll cost of $2.3 billion (Czujko & Chu, 2015). 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

Despite slowing growth in recent years, the biggest change in the global economy from an STM perspective has been the rise of China. Although China has become the world's second-largest producer of research papers (see NSF 2014), its share of the global STM market is much smaller than this might suggest. For example, Outsell estimated China comprised less than 5% of the global STM market by revenue in 2011; by contrast, as noted above, the US share was 55% (Outsell 2014c; Outsell 2012e). 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 estimated that the China STM market would grow between 10% and 11% year on year between 2011 and 2013.

Books and ebooks

The market for STM books as a whole has been shrinking in recent years, as declining print revenues have not yet been made up by the relatively fast-growing but still much smaller ebook revenues. Simba estimated that the 2012 global market for medical books fell 2.5% to $2.96 billion, while that for scientific & technical books fell 4.2% to $2.7 billion (Simba 2013b; Simba 2013a).

A 2012 report from Outsell (Outsell 2012d) 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.

The ALPSP *Scholarly Book Publishing Practice* report gives an older but 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 may still be a useful source of background 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 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 e-collections.

The open access market for scholarly books is as yet nascent and tiny, but is potentially significant for open access in some fields – such as the humanities – where the monograph and other scholarly books remain important research outputs. Some initiatives and developments are discussed in the section *Open access books*.

There has been 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 Education'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,[13] followed by 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.[14]

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.

(See also: *Costs of journal publishing*.)

Prospects for the STM market

Although there has been recovery, the effects of the 2008/09 recession are still being felt, with growth in much of Europe stalled. Public sector spending is weak at best in many developed countries as they seek to control budget deficits, leading to inevitable pressure on, and cuts in, institutional and library budgets. At the global level, The International Monetary Fund has cut its global growth forecasts for 2014 and 2015 and warned that the world economy may never return to the pace of expansion seen before the financial crisis (Elliot 2014). For the IMF key risks to the economy remain: financial markets too complacent about the future; tensions between Russia and Ukraine and in the Middle East; and a triple-dip recession in the eurozone could lead to deflation. Other risks include the withdrawal of quantitative easing, and a collapse of the housing market (bubble) in China.

With its limited dependence on advertising, the STM market is far less cyclical than many information markets, but the broadly STM market (as defined by Outsell) has managed compound growth of just 2.7% between 2008 and 2013. Outsell project faster growth in the coming years, with annual growth projected to reach 5%+ by 2016/17 and averaging 4% over the period 2013–2017 (Outsell 2014c), absent any further major economic shocks (e.g. a triple-dip recession the Eurozone, or a new war in Ukraine or 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 much of the developed world – have been flat, with growth

---

13 See MIT OpenCourseWare http://ocw.mit.edu

14 Values from 2008; not inflated to current values

barely keeping pace with inflation, and are likely to remain so. The market growth has come from the emerging regions including China, which continues its goal to move from a manufacturing economy to an "innovation-driven" one by 2020, and similarly these regions will continue to be the focus of revenue growth for most publishers. For example, the Indian market is expected to be on a par with most G8 nations by 2017, according to analysts Simba.

On the plus side, large publishers report strong growth in article submissions and in online usage, reflecting continuing growth in demand for research products. 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. Open access will continue to grow rapidly, although still a small proportion of total revenues. 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.[15]

Although there is limited hard evidence to support this, it seems possible that the recession and continuing 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 new pricing models likely to replace the historic pricing approach in coming years, and perhaps greater linkage of pricing to usage.

2.5. *Journal and articles numbers and trends*

There were about 28,100[16] active scholarly peer-reviewed English-language journals in 2014, collectively publishing approaching 2.5 million articles a year (Plume & van Weijen, 2014). Figure 4 shows the growth in the number of active, peer-reviewed journals recorded in Ulrich's directory between 2002 and 2012; over this period the number grew by about 2.5% a year. At the time of writing, the CrossRef database included over 71 million DOIs, of which 55 million refer to journal articles from a total of over 36,000 journals.More broadly, Google Scholar is estimated to index between 100 and 160 million documents including journal articles, books, and grey literature (Khabsa & Giles, 2014; Orduña-Malea et al., 2014), while the Web of Science database includes about 90 million records.

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 and significant advertising revenues – which the pure research journals do not have (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 10,900 journals from 2550 publishers included in Thomson Reuter's Journal Citation Reports database, of which 8700 are in the Science Edition and

---

15 e.g. Springer for R&D (http://rd.springer.com); Elsevier's acquisition and expansion of Knovel (http://www.elsevier.com/online-tools/knovel)

16 Ulrich's Web Directory listed 28,134 active scholarly peer-reviewed English-language journals on 16 December 2014. The count increases to 34,585 if non-English-language journals are included.

3000 in the Social Sciences Edition: these collectively publish about 1.5 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, Roos, & Lauri, 2009). (NSF 2014), which analyses a subset of the Thomson database, reported that the average number of articles per journal increased from 111 to 168 between 1988 and 2012.) The other major A&I database, Scopus, is intentionally broader in scope, and covers 22,000 peer-reviewed journals from about 5000 publishers. It contains about 53 million records and added just short of 2 million in 2013.

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 5), although the growth did slightly accelerate in the post-war period 1944–78.

Taken over similar timescales, the number of articles has also been growing by an average of about 3% per year. The reason for this growth is simple: the growth in the number of scientific researchers in the world. This is illustrated in Figure 6, 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).

Current article growth may well be higher than this long-term trend; (Plume & van Weijen, 2014) reported the total number of articles growing at an average 6.3% p.a., from 1.3 million to 2003 to 2.4 million in 2013. Similarly, articles indexed in PubMed have increased by 6.7% p.a. over the decade 2003-2013; and articles indexed in both Web of Knowledge and Scopus increased by about 5% on average between 2010 and 2013. (NSF 2014), however, reports article growth between 2001 and 2011 at just 2.8%; this lower figure may be attributable to the smaller dataset employed.)

A recent analysis by (Bornmann & Mutz, 2014) identified three growth phases, which each led to growth rates tripling in comparison with the previous phase: from less than 1% up to the middle of the 18th century, to 2–3% up to the period between the two world wars, and 8–9% to 2012.

Figure 4: Growth in the number of active, peer-reviewed journals recorded in Ulrich's directory, 2002–2012
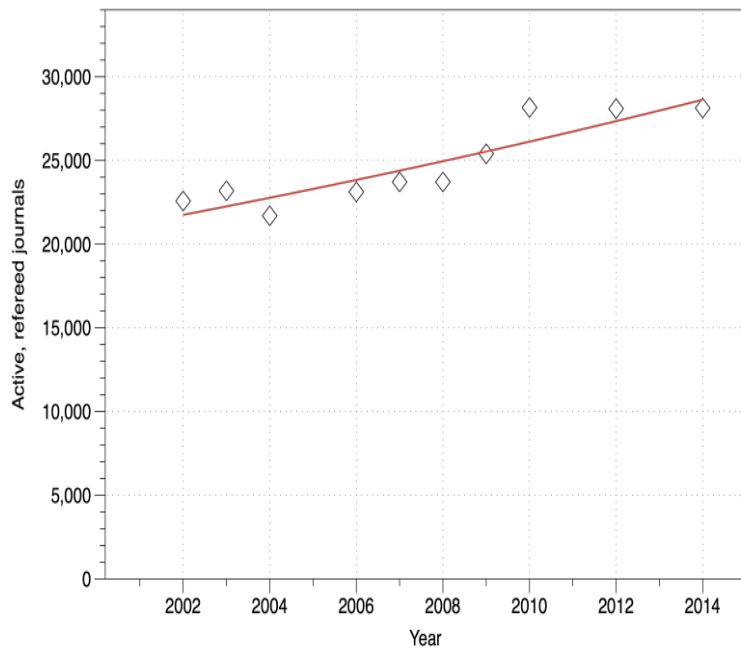

Figure 5: The growth of active, peer reviewed learned journals since 1665 (Mabe 2003)
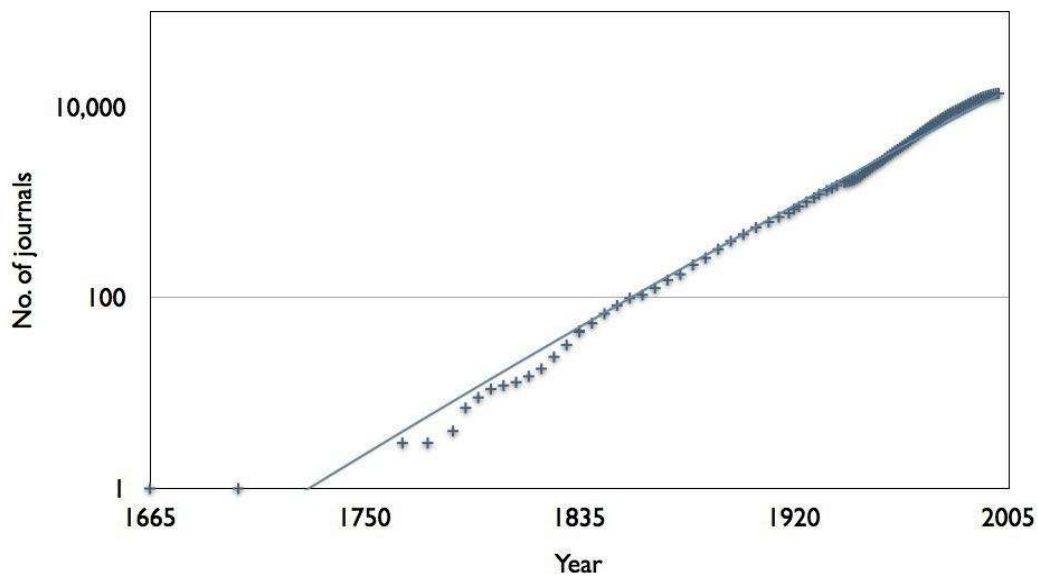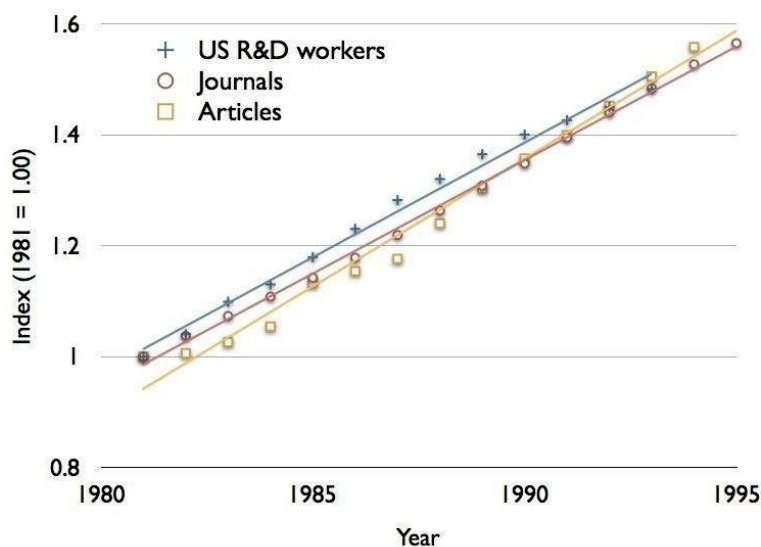
Figure 6: Relationship between numbers of researchers, journals and articles (Mabe 2004), using data from ISI and NSF)

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%).

The latest ALPSP report (Inger & Gardner, 2013) gave similar numbers, suggesting that the market had reached near saturation in terms of online availability, with the large majority of publishers having over 90% of their content available online. Online availability of backfiles is another matter, however, with about 70% of publishers having 90%+ online, and about 20% with less than 50% online. (Bear in mind, however, that this survey reports numbers of publishers, not journals: since the laggards will all be smaller publishers, the proportions for journals and articles will be significantly higher for both current and backfile content.)

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 (some 35% of respondents said they preferred print for viewing content in a 2014 Outsell survey (Outsell 2014b)), 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. Indications from 2015 catalogues and industry discussions are that print editions will, however, finally start to disappear from publisher's lists in significant numbers over the next couple of years.

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 numbers

The number of open access journals listed by the Directory of Open Access Journals[17] was 10,091 as of early December 2014 (of which 7245 were published in English); this represents an increase of 1976 over the 2 years or so since the last STM Report. Not all journals in DOAJ are fully peer-reviewed (though all exercise some form of quality control through an editor, editorial board or peer review)[18]. Ulrich's Directory lists 7111 peer reviewed OA journals, or about 25% 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 over 22,000 peer-reviewed journals, of which 2930 or 13% are open access, while the Web of Science includes some 726 OA journals, or about 9%.

Open access article numbers

Counting the number of open access journals has its challenges (such as filtering out predatory journals), but because journal size varies wildly (e.g. from a small quarterly publishing 20 articles a year up to *PLOS ONE*, which published 31,883 articles in 2014), 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.

Counting open access articles is, however, complicated by issues of definition, and by methodological and measurement challenges. Different researchers use different definitions for categories of OA articles, sometimes for ideological reasons, which makes comparisons of their different estimates hard or impossible. Broadly speaking the four categories of articles counted are (see the later section *Open access*):

- Gold: articles in pure OA journals (whether or not an article publication charge was paid); some studies include hybrid in this category

- Hybrid: articles in subscription journals made openly available immediately on publication, usually as the result of the payment of an Article Publication Charge (APC)

17 http://www.doaj.org/

18 DOAJ moved to a new host, IS4OA (Infrastructure Services for Open Access) in 2012, and the platform was relaunched in 2014. Following the relaunch, new tighter selection criteria were introduced and journals required to complete an application form to demonstrate adherence. (This was partly to eliminate the presence of "predatory" and other low-quality journals.) As of mid 2014, only a very small fraction (a few hundred out of 10,000) of journals listed had been re-certified in this way.

- Delayed: articles in subscription journals made openly available after an embargo period

- Green: copies of article versions available in institutional or other repositories, or on authors' web pages; often embargoed for a period following the publication date; may exist in multiple archived versions and in multiple copies on different repositories

- Other: strictly speaking these are not open access but freely available articles, some legitimately (e.g. promotional availability), some illegitimate (i.e. versions posted in breach of copyright). Automated tools or bots for searching for OA articles may unintentionally (or in a few cases, intentionally) count these

The challenges are greatest for green OA: as well as the challenges of definition and deduplication, the results are not fixed in time because articles can be added retrospectively at any time.

Methodologically, article counts can be made either by querying well curated indexes like Web of Science or Scopus, or by using specialised search engines and bots. The former only works for gold OA articles in journals covered by the indexes, which therefore underestimates the total article numbers (though not necessarily the proportions). The latter is subject to errors in correctly identifying and categorising open available articles discovered; researchers attempt to get round this by manually verifying as large a sample as their time and resources permit in order to estimate the reliability of the automated findings.

With those caveats in mind, Table 2 shows the results of a number of selected studies.The Elsevier study, conducted for the UK's BIS, has slightly lower results than some others but appears to be robustly designed, with lower results due to tighter definitions and more rigorous exclusion of false positives. The Archambault survey, conducted by Science-Metrix for the European Commission, employs unusual definitions that make its results particularly difficult to compare to other studies: unlike most studies, their "Green" category excludes subject repositories (e.g. PubMed Central) and other aggregator sites (e.g. CiteSeerX), and authors' personal web pages; "Other" includes these sites as well as hybrid, delayed, and illegitimately posted copies (which they call "Robin Hood or Rogue OA").

The historical growth in Gold open access is illustrated in Figure 7, from a 2012 article in *Nature* (Van Noorden 2012a), summarising 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. Figure 8 shows a longer-term trend, based on a single set of measurements made in early 2014 (Archambault et al., 2014).

Availability varies substantially by discipline, as illustrated in Figure 9 (from (Björk et al., 2010) and Figure 10 (data from (Archambault et al., 2014). Uptake of the Gold model remains 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 2013 and covered by PubMed that have free full text available (in late 2014) was 34%. The fields most resistant to Gold OA are engineering and physics & astronomy within the STEM subjects, and philosophy & theology and economics & business on the humanities and social science sides. The reasons for this will be varied: physics & astronomy and economics both have thriving preprint cultures, for instance, which may reduce demand for Gold OA.

Table 2: open access articles shares reported by selected studies (see text for details and qualifications)

	(Björk et al., 2010)	(Gargouri, Larivière, Gingras, Carr, & Harnad, 2012)	(Laakso & Björk, 2012)	(Elsevier 2013)	(Archambault et al., 2014)
Gold (total)	5.3%		11.0%	10.2%	12.1%
Gold, with APC				5.5%	
Hybrid	2.0%		0.7%	0.5%	
Gold, no APC ("subsidised")				4.2%	
Delayed free access ("open archives")	1.2%		5.2%	1.0%	
Green (total)	11.9%			11.4%	5.9%
Green, preprints				6.4%	
Green, accepted mss				5.0%	
"Other OA"					30.9
All OA	20.4%	21.0%		~23%	46.9%

Figure 7: Growth in estimates of the fraction of articles published as Gold open access (Van Noorden 2012a)
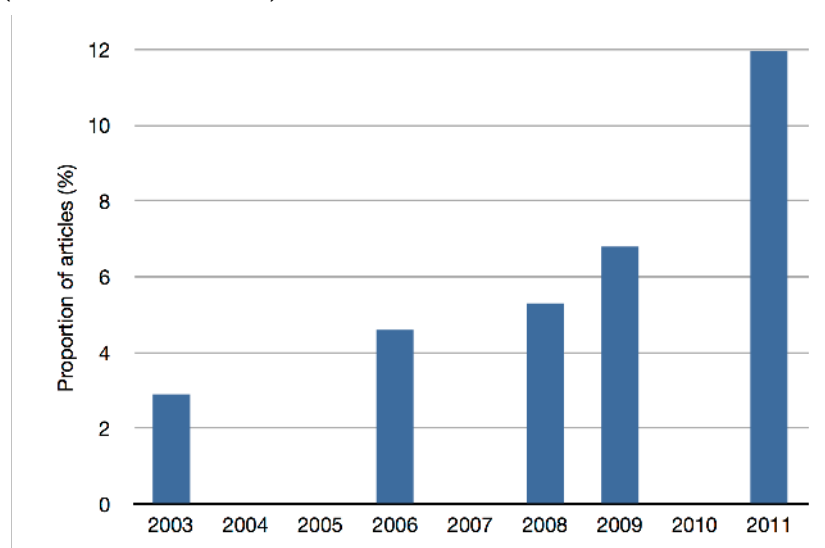
Figure 8: Growth of Gold open access 1996-2013 (from Archambault 2014)
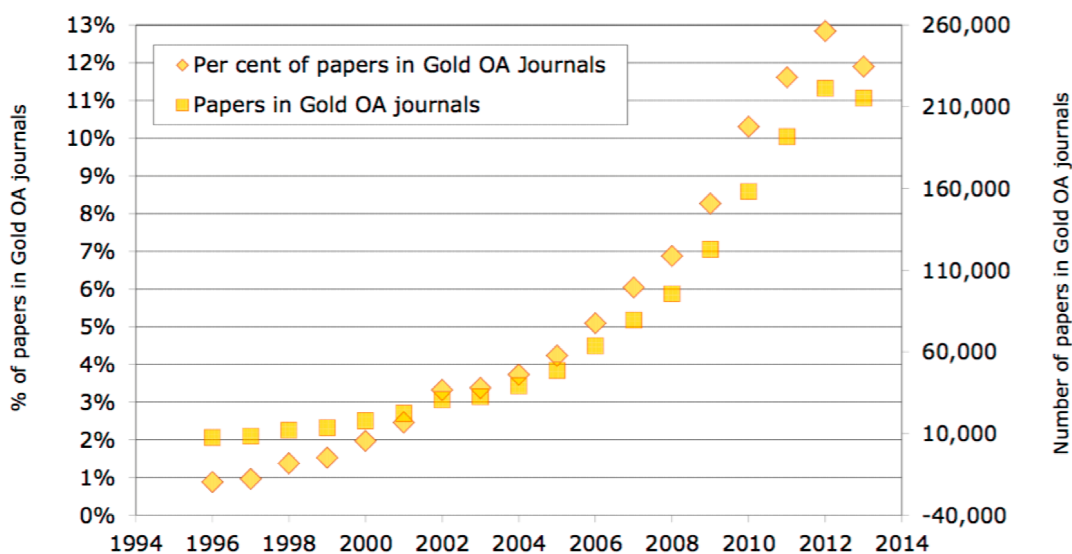
Figure 9: OA availability by discipline. Reproduced from (Björk et al., 2010), doi: 10.1371/journal.pone.0011273.g004
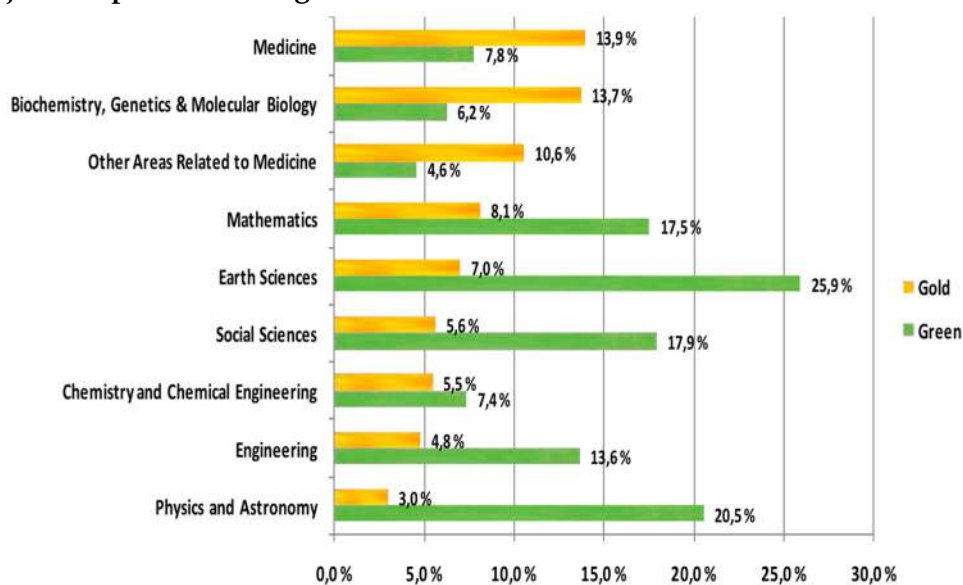
Figure 10: OA availability by discipline, in declining prevalence of Gold OA (data from (Archambault et al., 2014), Table II; Green OA not included because authors' definition is not comparable to other studies)
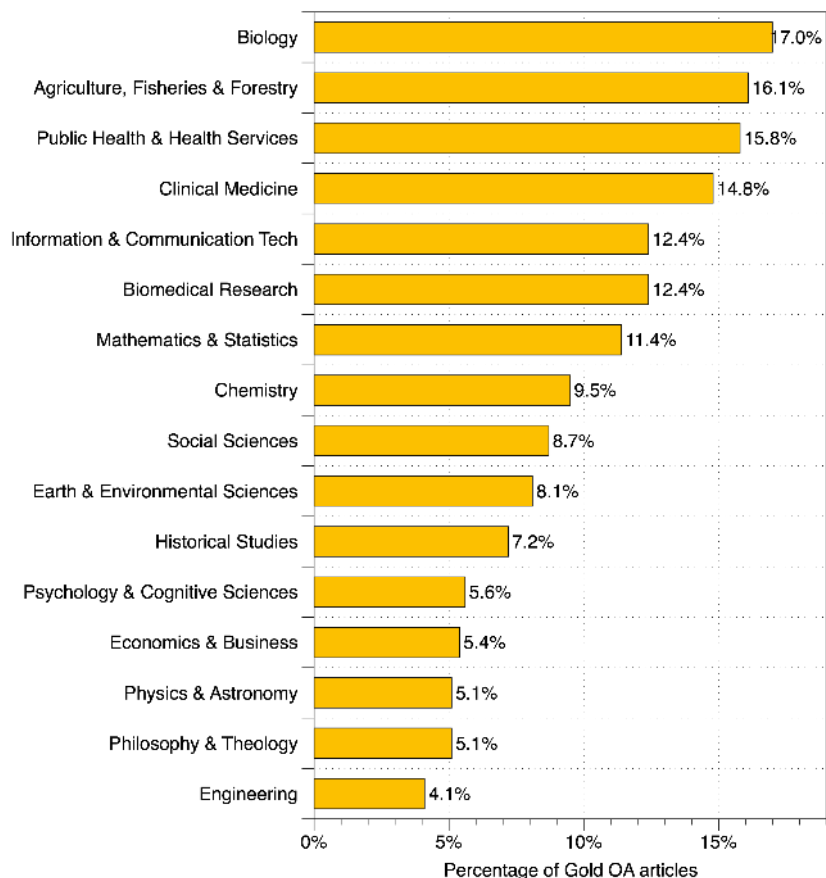
Figure 11: Growth in open access content in PubMed Central
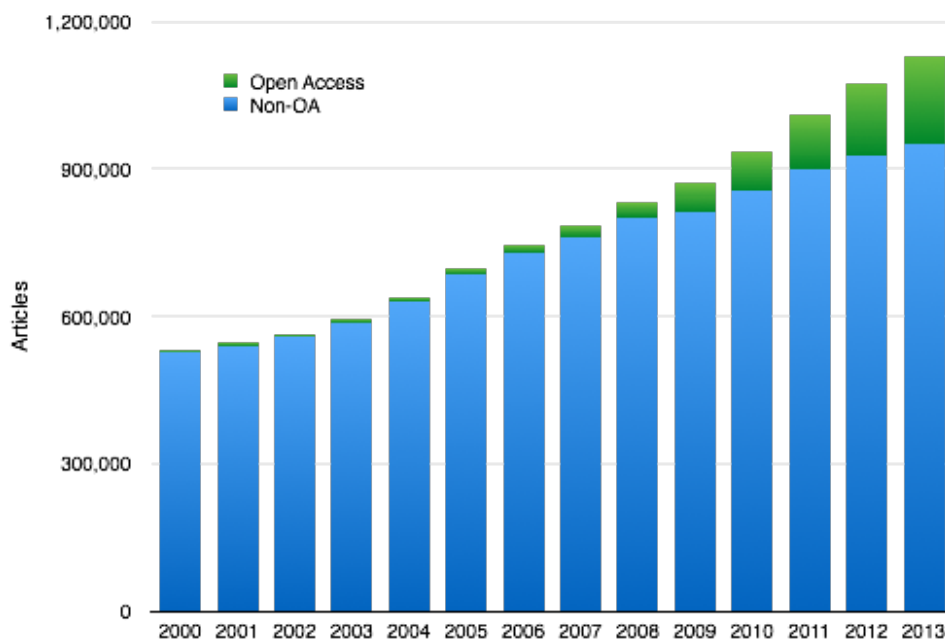
2.6.   Global trends in scientific output

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 over the long term, rising from $522 billion in 1996 to $1.3 trillion in 2009 (NSF 2012) and an estimated $1.6 trillion in 2014 (Battelle 2013). 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 $465 billion (a research intensity of 2.8% of GDP, well above the global average), with a 34% share of all global R&D, though for the first time this position is beginning to be challenged.

Nonetheless, the impact of the 2008/09 recession on global R&D expenditures was marked: the annual growth for OECD countries for 2008–2012 was just half that for 2001–2008 (OECD 2014).

Governments see spending on R&D as critical to innovation, growth and international competitiveness.[19] 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 has been 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. More recent statistics show China's R&D spending doubling between 2008 and 2012 (Rowlands & Nicholas, 2005). 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; Elsevier 2013).

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 (5.2%) compared to 2011 (6.5%), due to continued recessionary effects and the ending of stimulus spending packages, and slowed further to 2.7% in 2013. Battelle forecast some recovery to 3.9% in 2014 (Battelle 2013).

---

19 Expenditure on R&D appears to be a very good investment for governments: while private returns to R&D are estimated to average around 25–30%, social returns are typically 2–3 times larger (Department for Business, Innovation & Skills 2014)

R&D expenditure by subject discipline

The general trend in academic R&D expenditures has been away from the physical and towards the life sciences. In the US, for example, life sciences receives about 60% of funding (and about 55% of this is for medical research), and is the only broad field to have seen a sizeable rise in share of total US academic funding (up by 6 points).

The research priorities of the major regional blocs vary according to their economic needs and internal political pressures. The United States and EU are focused primarily on biological sciences and medical sciences, while Japan's articles are divided among biological sciences, medical sciences, chemistry, and physics. The research priorities of emerging economies have been more focussed toward economic growth and infrastructure development; for example, China's portfolio is currently dominated by chemistry, physics, and engineering, although its 15-year plan for 2006–2020 focuses on energy, water resources, and environmental protection.

Role of industry

The majority of R&D expenditure is funded by industry: about 66% 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; 2013). 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.

Growth of knowledge-intensive industries

The latest NSF Science & Engineering Indicators (NSF 2014) emphasises how the global map of science and technology-related economic activity following the 2008–09 downturn differs from the previous patterns. The overall trend continues to be one of growing importance of knowledge-intensive economies. Knowledge- and technology-intensive (KTI) industries grew their share of the developed economies from 29% to 32% between 1997 and 2012. The growth of KTI industries in the developed world was most evident in China's, whose high-tech manufacturing rose more than fivefold between 2003 and 2012, resulting in its global share climbing from 8% to 24% in 2012.

Numbers of researchers

There is no single comprehensive and widely accepted set of figures for researcher numbers, partly for reasons of difficulty of defining a researcher after leaving academia, and partly because of different approaches to recording these statistics in different countries.

The latest available OECD statistics report a researcher headcount of 8.4 million for 2011 (a full-time equivalent of 6.3 million), covering the OECD plus some key non-OECD countries (e.g. China and Russia) but excluding some other important countries (e.g. India, Brazil). This was an increase of 7.5% on 2010, reflecting a bounce back from the recession when numbers fell. The average annual growth between 2000 and 2011 was 4.2% (headcount) and 3.2% (FTE), suggesting a trend of greater part-time work (OECD n.d.).

The most recently available UNESCO data[20] reports 7.2 million researchers in 2007, up from 5.7 million in 2002 (a CAGR of 4.4%) (UNESCO 2010), while Elsevier's latest report for the UK government gives a lower figure, estimated at 6.73 million for 2011 (Elsevier 2013). (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). The World Bank puts the total somewhat higher, at 8.9 million researchers.

Whichever definition is used the number of global researchers is steadily growing over the longer term, at about 4–5% per year (although with short-term dips during economic recessions, most recently in 2009). 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 2.9% for the G8, and 1% in the US and EU. The most rapid growth has been in South Korea (doubling numbers between 1995 and 2006) and China (tripling between 2005 and 2008). 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 changes

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. For example, the shares of global R&D of both the US and the EU declined substantially between 2001 and 2011, with the US's share falling from 37% to 30%, and the EU share from 26% to 22%. At the same time, the economies of East and Southeast Asia and South Asia saw an increase in their combined share from 25% to 34% of the global total (NSF 2014).

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 (with energy, water resources, and environmental protection as its research priorities). 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 the early 2020s (Battelle 2013; NSF 2012) or even earlier (OECD 2014).

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

---

20 A new edition of the UNESCO Science Report is due to be published in November 2015; see http://www.unesco.org/new/en/natural-sciences/science-technology/prospective-studies/unesco-science-report/

become world class; with the reminder targeted to be the best in China and well known in world (Outsell 2012e).

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 between 11% (NSF 2014) and 17% (Elsevier 2013), and is set to overtake the US well before 2020. (Figures 12 and 13 shows the trends in article outputs for 2008–2012 and 2001–2011 respectively.)

China has not just increased the quantity of its research outputs; the quality is also steadily increasing, although it still remains well behind the US and EU. Its share of total world citations is now close to the world average (Thomson Reuters 2014); this is most marked in computer science, where its relative citation index (the ratio average number of citations to that expected if all global citations were divided equally among countries) of 1.3 is slightly ahead of that of the US. Its proportion of articles among the most highly-cited 1% increased sixfold between 2002 and 2012 (NSF 2014) and is now nearing the world average; and the editors of the Nature Publishing Index 2013 predicted China is "on pace to take over as the top Asia-Pacific contributor to the NPI in the next two or three years" (Nature Editors 2014).

China is not the only country to grow share of world publications. Between 2003 and 2012, all the BRICS countries increased their shares except for Russia, whose share fell to 2.1% from 3.0%. Latin America saw its share of publications over this period rise from 3.8% to 4.9% between 2002 and 2008. The US remains for now in first place, but it share fell the most (in percentage points) from 33% to 27.8% (Thomson Reuters 2014; 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.

Researcher mobility

According to Elsevier, in a global research world, the 1950s idea of a "brain drain" should be replaced by the more nuanced concept of "brain circulation". In this view, the skills and networks built by researchers while abroad accrue benefits to their home country's research base when they eventually return, and often even if they do not return but remain instead as a diaspora (Albanese 2009). Elsevier's work shows that researchers are highly mobile, though mobility varies by country, with UK and Canada being the most mobile with the lowest proportion of "sedentary" researchers (those not publishing outside their home country in the period 1996-2012) at 27%, compared to 60% in Japan and 71% in China. The UK had a particularly mobile researcher population, with almost 72% of active UK researchers having published articles while affiliated with non-UK institutions (Elsevier 2013).

Figure 12: Share of world articles (Source: (Elsevier 2013)
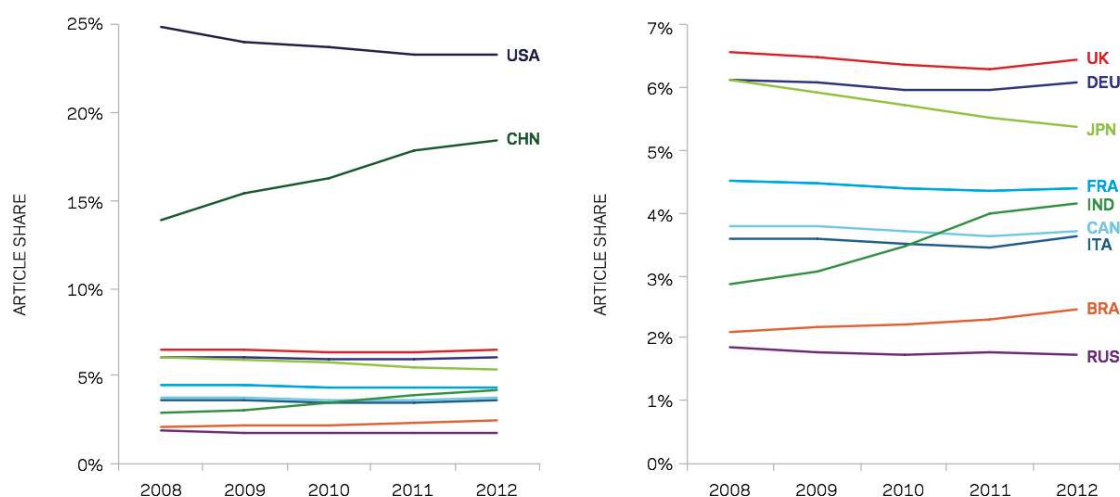
Figure 13: Global shares of article outputs (Source: (NSF 2014)
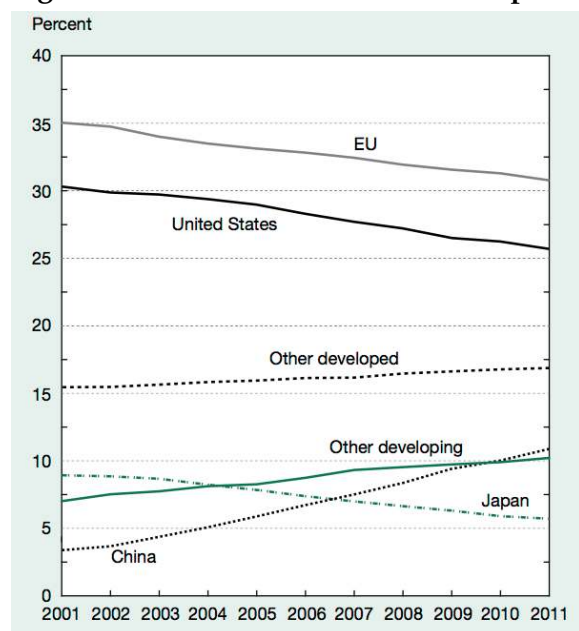
Collaboration and coauthorship

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 15 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 8 in 2012 (NSF 2014); across the Thomson Reuters database as a whole from 3.8 in 2007 to 4.5 in 2011 (ScienceWatch 2012). Figure 4 shows how the average number of authors per paper grew during the second half of the 20th century, while Figure 16 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 (see also *Data Citation*).

International collaboration is also growing: between 1997 and 2012, internationally coauthored articles grew from 16% to 25% of the world's total. 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. The UK, France and Germany all have high international coauthorship at over 50% of papers (Elsevier 2013). China, however, is an exception to the general trend of increasing international collaboration, with rates remaining stable at 27% over the last decade during its very rapid article growth (NSF 2014).

International coauthorship patterns vary by subject discipline. Within STEM subjects, astronomy is the most international field, while psychology, chemistry, social sciences have much lower rates of collaboration (Figure 16; (NSF 2014).

There is a clear benefit to researchers to international 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 14: Coauthorship patterns 1954 to 2000 (from (Mabe & Amin, 2002), using data from Thomson Reuter Science Citation Index)
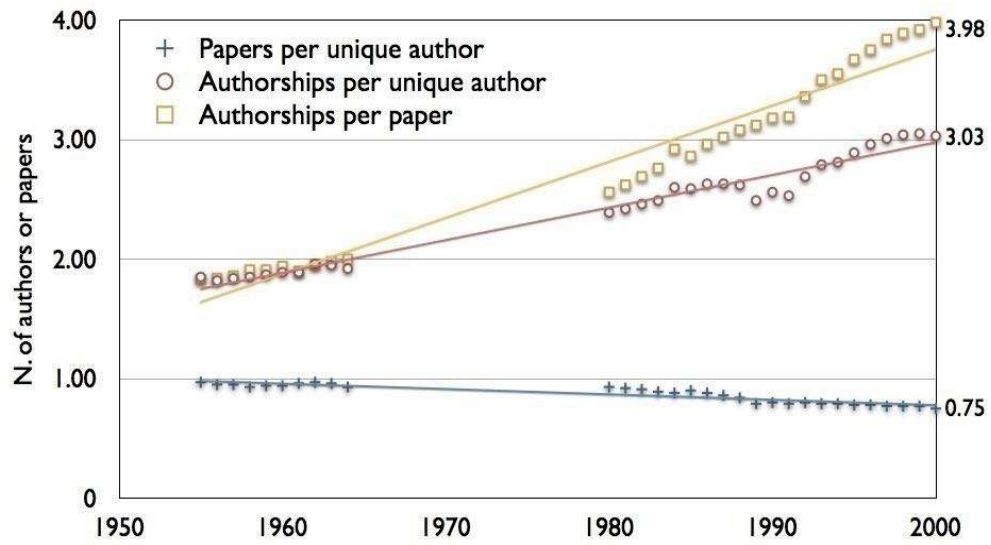
Figure 15: Research articles with international coauthors, by selected region/country/ economy: 1989–2009 (source: (NSF 2012)
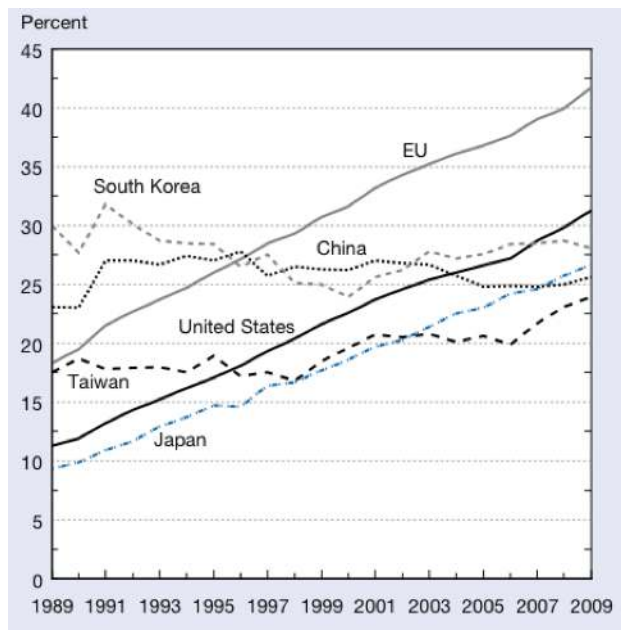
Figure 16: Growth in coauthorship by discipline: global articles in Science Citation Index and Social Sciences Citation Index; 2012 vs 1997 (from (NSF 2014)
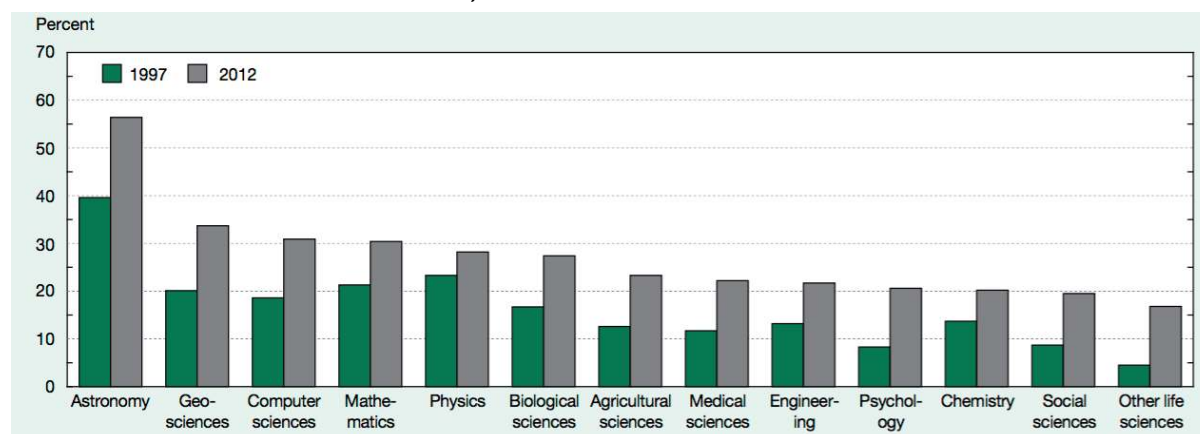
2.7.   Authors and readers

Authors

The global number of active researchers varies by definition used but is estimated to be between 6.7 and 8.9 million (see *Numbers of researchers*). The number of authors differs, however, primarily because by no means all of these will publish an article in a given year. For example, (Plume & van Weijen, 2014) reported that 2.4 million articles were published in 2013 by a total of 4.16 million unique authors. (Total authorships were 10 million because each article had an average of 4.2 authors.) These figures represented steady growth from 2003, when there were about 1.3 million articles published by about 2.1 million unique authors. The Scopus database holds about 15 million author identifiers.

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).

A more recent study looked at the most productive authors, defined as those who had published at least once every year over the 16-year period under study (1996–2011). It found a total of 15.2 million publishing scientists of which just 150,608 (or less than 1%) managed to publish a paper every year. This active core, however, was responsible for 42% of papers and 87% of the very highly cited papers. Many of these prolific scientists are likely the heads of laboratories or research groups whose names are attached to the outputs of their teams (Ioannidis, Boyack, & Klavans, 2014).

Readers

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. More recently the scientific social network Academia.edu has reported have more than 16 million registered users, which suggests that this figure may be an underestimate (see *Scientific social networks*).

These overlapping author and reader communities can be illustrated as in Figure 17. 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).

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 17: overlapping author and reader communities
About 4 million authors publish each year (Plume & van Weijen, 2014), out of a global population of approximately 8.5 million R&D workers (based on UNESCO figures)
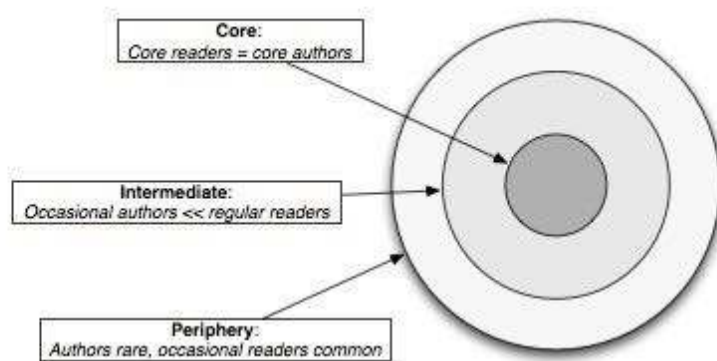
2.8. Publishers

There are estimated to be of the order of 5000–10,000 journal publishers globally: the Scopus database covers 22,000 journals from over 5000 publishers, and the long tail making up the remaining 10,000 or so peer-reviewed journals not covered by Scopus is likely to consist of publishers with just the one journal.

The membership of main English-language trade and professional associations for journal publishers (ALPSP, SSP and STM) include most of the larger publishers but of course only a small fraction of the wider global total of publishers. According to (Morris 2006), as of 2006 these 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 four publishers (Elsevier, Springer, Wiley-Blackwell, and Taylor & Francis) have well over 2000 journals each (Table 3). 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).

Table 3: The 10 largest publishers, by number of journals

Publisher	Number of journals	Cumulative % of all journals
Springer (exc. NPG)	2987	10.6%
Elsevier	2500	19.5%
Wiley	2388	28.0%
Taylor & Francis	2105	35.5%
SAGE	750	38.1%
Wolters Kluwer (inc. Medknow)	672	40.5%
Hindawi	438	42.1%
CUP	350	43.3%
OUP	362	44.6%
NPG (inc. Frontiers)	178	45.2%

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 typically consists of 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 – see *Critiques of peer review*); for example in a Publishing Research Consortium survey 93% disagreed that peer review was unnecessary (Ware & Monkman, 2008); see also (Sense About Science 2009). Despite this overall commitment, however, there appears to be growing 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, however,(Mulligan & Mabe, 2011) found little change in researchers' core attitudes to peer review: it remained highly valued, and a large proportion continued to be willing to commit to reviewing.

Strong support for this continuity comes from a more recent large-scale survey conducted by CIBER (Nicholas et al., 2015). This found that the key benefit of peer review was seen to be its role in providing the central pillar of trust, although improvements to the article were also valued. The biggest criticisms were slowness; hands-off editors that did not intervene sufficiently; light-touch peer review – researchers valued detailed, robust review; and variable quality of reviewing. Researchers (especially younger ones) were willing to use non-peer reviewed materials but far less likely to cite them: this was seen to be a formal activity where peer-reviewed content was required. Peer review also remained important to choice of journal, and the Impact Factor remained important. Social media and open access were not seen to be important agents for changing attitudes towards peer review: researchers had moved from a print-based system to a digital system, but it had not significantly changed the way they decided what to trust.

Benefits of peer review

There are a number of arguments in favour of peer review. It could be seen as a 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 and administrators (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. The growth of "predatory journals" and similar, however, may have increased the need for distinguishing journals and articles

that are properly peer reviewed from those masquerading as such (see *Peer review validation* below).

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 main 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 has historically been more common in the humanities and social sciences than in the "hard" sciences, with clinical journals falling between the two. This may starting to change, however, with some leading scientific journals like Conservation Biology and some of the Nature research journals adopting it (Cressey 2014).

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. Alternatively, journals using double-blind review may need to incur significant editorial costs to remove clues to authors' identities from manuscripts before review.

Open peer review

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.

Historically, 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, support for open review appears to be growing: the publication of reviews alongside the published paper, either signed or unsigned, is becoming more

widespread. 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).

Post-publication peer review

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 and transparent process. A well-known example is the journal *PLOS ONE*, though fewer than 10% of articles in PLOS journals have received comments. As with pre-publication open peer review, academics have been reluctant to participate in most of the trials and initiatives in this area to date. 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.

Some recent developments, however, indicate that this may turn out to be a concept whose best is yet to come:

- The NIH has launched PubMed Commons, a pilot service allowing commenting on the 22 million articles in the PubMed database. Commenting is restricted to authors in PubMed (a pretty broad category in this field) and users can both comment and rate the usefulness of other comments. PubMed's central position in biomedicine gives the approach considerable credibility

- The open access publisher Frontiers (part owned by Macmillan Science & Education) have developed an innovative evaluation system that using algorithmic methods to provide a post-publication evaluation of published research, allowing readers to indirectly contribute to an article ranking

- PLOS has trialled Open Evaluation, allowing users (in a private beta test) to rate articles on four dimensions: interest level, the article's significance, the quality of the research, and the clarity of the writing. It is presumably no coincidence these are typically the same questions posed to reviewers in conventional pre-publication review

- Academia.edu, which currently claims more than 16 million registered users, announced that it had acquired the startup Plasmyd, with a view to integrate its technology specifically to support post-publication peer review.

In a closely related area, there is also growing interest in aggregating multiple "signals" of an article's potential impact, including the number of post-publication comments (both on the journal website and elsewhere on the web), as a complement to the Impact Factor (see *Article-level metrics and altmetrics*).

Cascade peer review

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 agree to accept

manuscript reviews from other members of the consortium, although the journals are with different publishers. In practice, this consortium has had little impact, with only a tiny fraction of papers being transferred.

Despite this, a new peer review consortium was announced in 2013 by eLife, BioMed Central, the Public Library of Science, and the EMBO (the European Molecular Biology Organisation) (Clarke 2013).

Portable peer review

In cascade review, peer review is carried out by a journal in the usual way, and if the paper is rejected the review may accompany the paper to a new journal submission.

A more radical idea is for authors to commission their own peer reviews *prior* to journal submission. This might be as a pre-submission process intended to improve the paper before submission in the conventional way, or even a fully "portable" review that participating journals could agree to accept (if not necessarily to be bound by).

There are two organisations offering portable peer review services, each with quite different business models. Rubriq[21] provides authors with peer review in return for a fee of \$500–650. Authors can absorb this as a cost of improving the papers or, Rubriq hopes, open access journals might discount their APCs to reflect the value of the submitted review. Peerage of Science[22] offers a platform for journal-independent review which publishers can similarly scan for potential submissions. It does not charge authors but seeks to cover its costs by charging publishers a fee of about \$400 per published paper.

Peer review abuse and misconduct

A few researchers have exploited loopholes in the peer review system to rig peer review in their favour. A 2014 Nature article detailed how some authors had set up fraudulent accounts on online peer systems for both fictitious and actual researchers using multiple generic email addresses (e.g. Gmail) generated for the purpose (eLife 2014). They were then able to propose themselves as reviewers for papers they submitted. In the past 2 years, journals have been forced to retract more than 110 papers in at least 6 instances of peer-review rigging. In one major case, a 14-month investigation by SAGE uncovered some 130 suspicious reviewer accounts and 60 papers articles were retracted. (See also *Publishing ethics*.)

Rewarding reviewers

Peer review of journal articles has traditionally been seen as part of the professional obligations of the researcher. Fees are almost never paid (one rare exception is *Collabra*, a newly announced megajournal from CDL; there are also examples of payment in kind, such as waiving submission fees, waiving or discounting APCs, providing time-limited access to subscription-based resources, etc.).

Surveys of researchers and publishers' day-to-day experience suggest that there is very little demand for such fees (although anti-corporate sentiment may contrast large publishers' profits with the fact that peer review is unpaid). There does appear, however, to be demand for greater formal recognition for the work of reviewers. At present, where blinded peer review is employed, such recognition typically takes the form of an annual statement from

---

21 http://www.rubriq.com

22 https://www.peerageofscience.org

the journal listing and thanking its reviewers. Researchers can and do list reviewing activities on their *curricula vitae*.

More direct ways of rewarding review via recognition are emerging:

- Publons offers a service whereby reviewers can post their peer review history online, and then showcase this as they choose (for instance in the CVs). At the time of writing, Publons had recorded nearly 30,000 reviews from 59,000 reviewers, covering 2800 journals. (Van Noorden 2014d) carries interviews with some of the more prolific users

- Publishers are experimenting with more direct approaches. For example, Elsevier held an open competition for ways to improve peer review; the winning entry proposed "reviewer badges and rewards scheme", which is now being implemented. Reviewers can display badges (generated via Mozilla OpenBadges) on social media pages. In a second phase a "reviewer recognition" platform has been developed for approximately 40 journals. Upon completion of a review for one of these titles, reviewers are provided with a link to a personal page on the platform that displays their reviewer activity (van Rossum 2014)

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 some 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 submitted 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: (Inger & Gardner, 2013) reports that only 5% of publishers were without peer review systems (sharply down from 35% in the previous 2008 survey). The majority of publishers opted for one of the three market leaders, Editorial Manager (Aries), eJournal Press and Manuscript Central (Thomson Reuters).

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 handling 1 million manuscript submissions in 2010 compared to 317,000 in 2005.[23] (These data are not normalised for the increasing numbers of journals using the system.) 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 (Morris 2009).

The scale of the peer review operation managed by publishers is illustrated by the throughputs of the leading online submission systems. By 2013 the ScholarOne system was handling a total of 1.6 million original submissions per year (or 2.2 million including resubmissions). Its rival Editorial Manager processes a total of 2 million manuscripts a year on behalf of 5,800 journals from over 250 publishers.

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*).

- Integrated e-commerce: OA article processing charges, or page or colour charges can be managed using publishers' own systems or third-party plug-ins such as RightsLink (Copyright Clearance Center)

- Metadata collected at submission or acceptance can be used to create integrations with other services; for instance, see *CHORUS* which depends on collection of FundRef data.

Peer review certification

The reader of an article has no way of knowing whether an individual article has been peer reviewed, and if so, to what standard, without a good working knowledge of the journals in a field. Lay readers will typically not have this knowledge, but even expert researchers will encounter articles from outside their domain, and journals with otherwise good peer review do not always clearly label (at the article level) which articles have been per reviewed and which not. To counter this problem, some initiatives propose certification of the peer review process at the article and/or journal level.

---

23 By 2013, the number of original submissions had risen to 1.6 million, or 2.2 million for all submissions (source: Thomson Reuters)

One such is PRE (Peer Review Evaluation),[24] "a suite of services designed to support and strengthen the peer-review process". Its first service, PRE-val, verifies for the end user that content has gone through the peer review process and provides information relevant to assessing the quality of the process. An article-level service, preSCORE was also originally proposed.

In a similar vein, the medical editor of BioMed Central has called for a "kitemark" to identify research papers that have been peer reviewed by people with the necessary skills (Patel 2014).

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 18 (Tenopir 2007; Tenopir, King, Edwards, & Wu, 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).

A 2008 international survey (Tenopir, Mays, & Wu, 2011) found that researchers in the sciences reported spending time reading scholarly content of between 12.3 hours/week (health sciences) and 15.3 hours/week (life sciences); while social science researchers said they spent a (somewhat implausible?) 25.9 hours/week (while not reading any more articles in total).

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). Researchers in health sciences spend the least time per article at around 23 mins compared to 30–31 mins in the other sciences and social sciences (Tenopir et al., 2011). 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

---

24 http://pre-val.org

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 & 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 18: Average number of articles that university faculty members reported reading per year (source: (Tenopir 2007)
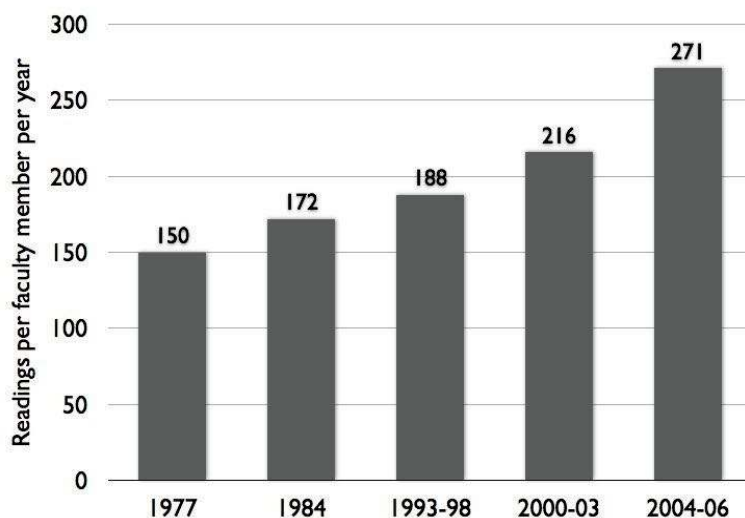
Access and navigation to articles

Academics use a wide range of methods to locate articles, as illustrated in Figure 19 and in more detail in the more recent data in Figure 20. 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 (Mabe & Mulligan, 2011), 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.

- There were some notable difference between disciplines: for example, researchers in humanities and education research were much more likely to use the library web pages for article searching than those in physics and astronomy

- Regional differences may also be important: for example, Asian researchers were more likely to come to an article from direct searching, rather than from journal alerts or article citations

- 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.

An interview study by Newman & Sack (2013) provides some useful qualitative background to the quantitative data on reading and information discovery behaviours. The following, for example, suggests there is still a lot to be gained by providing easy-to-use tools for helping researchers stay current:

> *"Most interviewees do not have a systematic strategy for keeping up to date. […] Interviewees rely heavily on cited references in known items, recommendations received from colleagues, or contents of a small number of familiar journals. Only a few get alerts from abstracting and indexing databases supplemented by alerts from important journals. Several expressed frustration at their lack of skill in finding current information.*

> *"An interviewee in the Computer Science department stated, 'I have constant guilt feelings about not doing enough to keep current.'"*

The Ithaka S+R/JISC 2012 survey of UK academics confirms this picture, with the leading way to keep up with current research reported to be attending conferences of workshops, followed materials suggested by other academics, skimming new issues of key journals, and skimming tables of contents. Electronic discovery tools (recommender systems, saved keyword alerts) were of much lesser importance, with social media being the least important method (Ithaka S+R, JISC, & RLUK, 2013).

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).

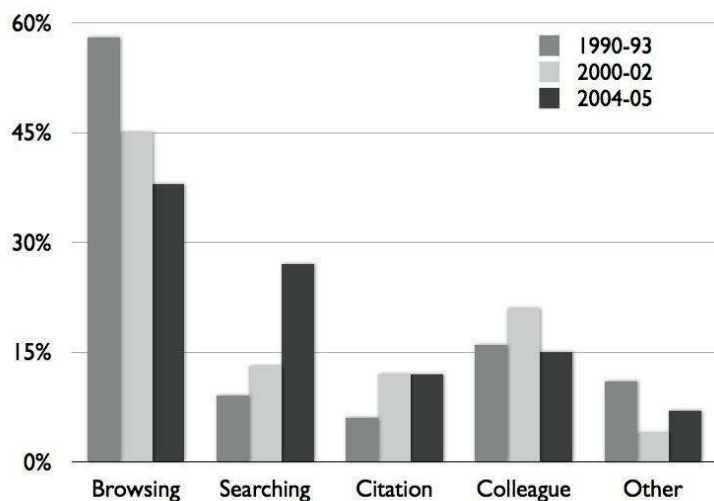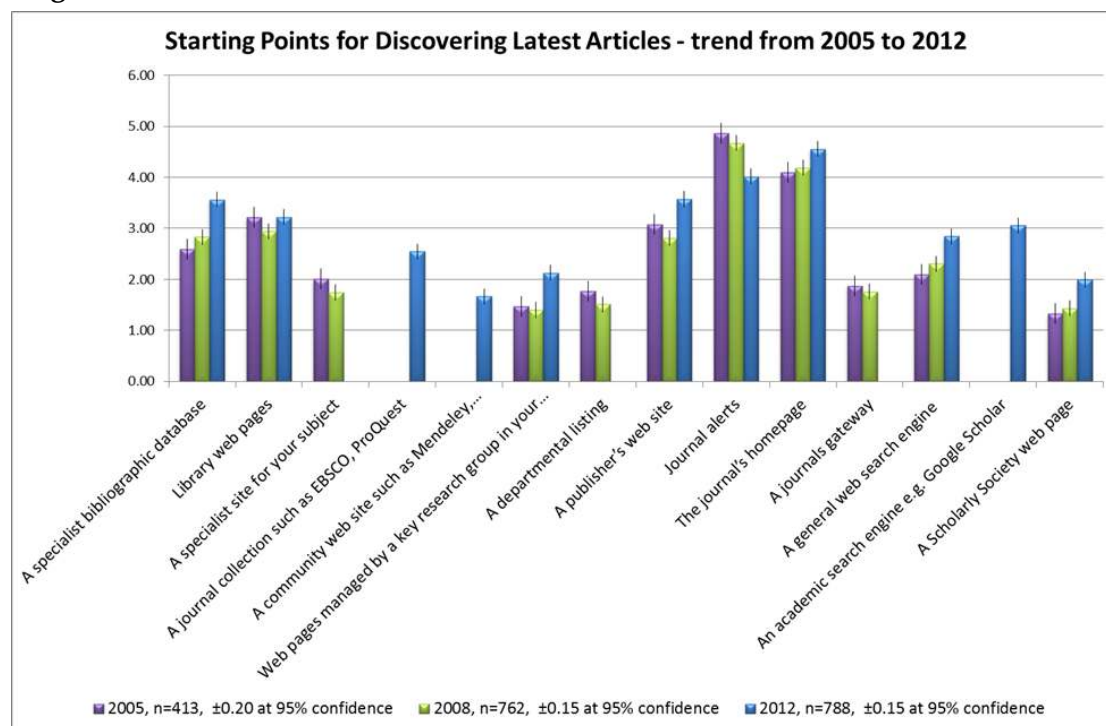**Figure 19: Ways used by university faculty to locate articles (source: (Tenopir 2007)**

Figure 20: 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,[25] 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 (Mabe & Mulligan, 2011) 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 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). Another study identified pressures created by assessment exercises as a factor in in this change (Adams & Gurney, 2014); see *Effects of research assessment on researcher behaviour*).

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

---

25 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

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 21). 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 22 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 also marked differences between disciplines in the attitudes of researchers towards open access. Some of these reflect funding structures (e.g. the lack of external research funding in the humanities and mathematics), while others reflect long-standing norms in the research communities (e.g. a preprint culture predating open access). See *Open access* for more details.

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 21: Average articles read per university faculty member per year (Source: (Tenopir 2007)
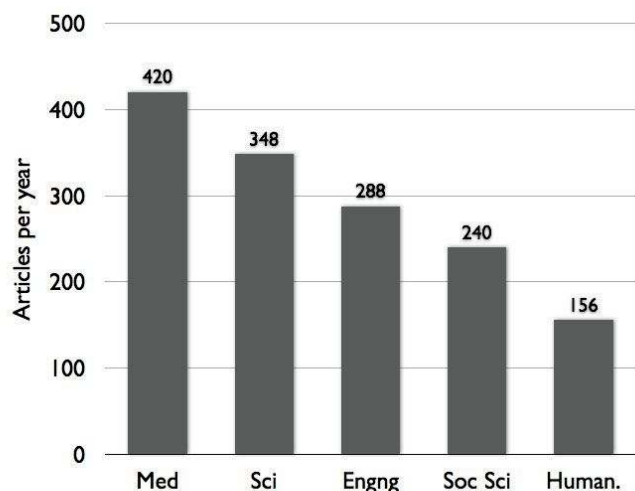
Figure 22: Subject differences in the ways articles are found (Source: (Tenopir 2007)
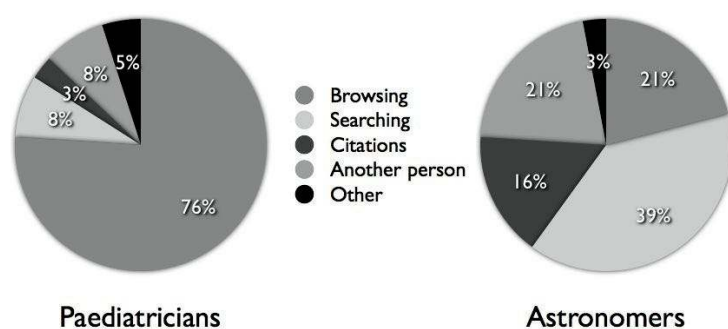
2.12. Citations and the Impact Factor

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.

Citation inflation

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%. Figure 23 shows the trend in average citations per article for the period 1992–2012; the average for all countries has risen from about 1.7 in 1992 to 2.5 in 2012

(NSF 2014).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).

Figure 23: Citation inflation: increase in the average citations per article, by country of author (source: (NSF 2014)
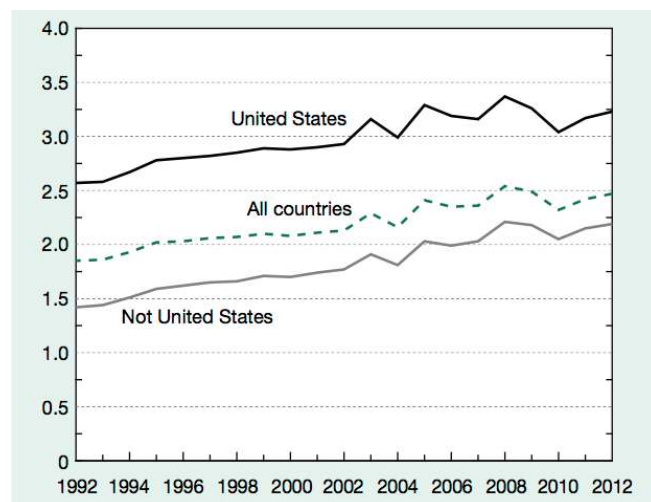
International trends in citation

As with article publication patterns, the regional shares of citations are changing as a result of these globalisation pressures. Table 4 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 more recent data in Figure 24 confirms this trend continues (Elsevier 2013)[26].

The growing internationalisation of research is reflected in an increasing proportion of citations from outside the country of authorship. Like international coauthorship (see *Collaboration and coauthorship*), international citation has grown steadily over the last two decades for all major scientific countries with the exception of China. In 1992, 69% of citations to Chinese scientific articles came from outside China; by 2012, the proportion had dropped to 49%, suggesting China's expanding article output is being used mostly within China (NSF 2014).

Table 4: 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

---

26 the apparent differences between these figures and the NSF data used in Table 4 are most likely due to the different datasets used: Elsevier used the full Scopus database, with around 22,000 journals; NSF used a subset of Web of Science containing 5087 journals

Figure 24: Share of world citations 2008-2012 (Source: (Elsevier 2013)
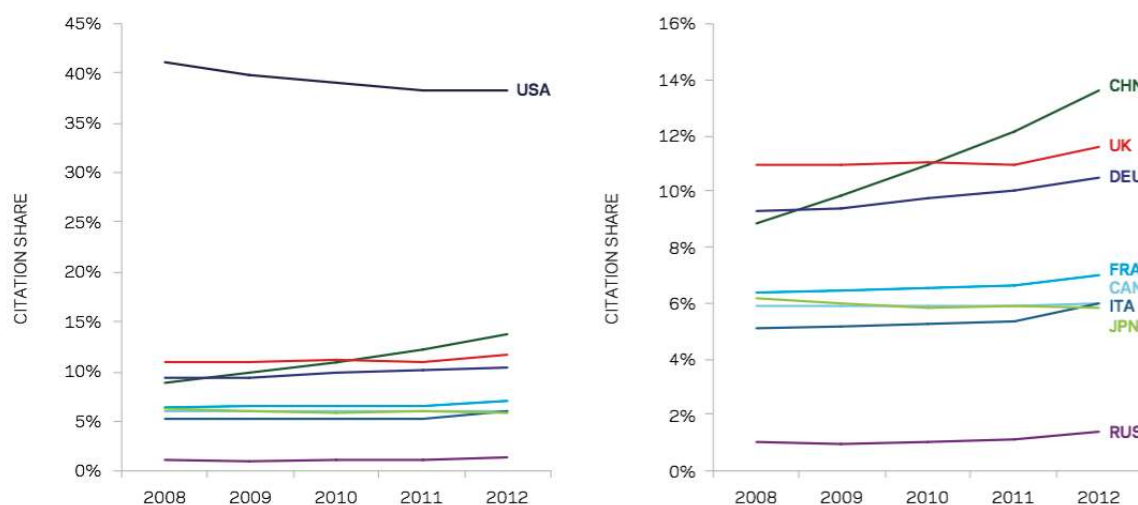
Citations and impact metrics

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.

The distribution of citations follows the widely-found Pareto pattern, with about 80% of citations coming from about 20% of articles. For example, Scopus data for citations to 2008 articles made in 2008–2012 showed almost exactly this result, while 32% of papers remained uncited (Elsevier 2013).

At the other end of the scale, the proportion of papers in the most-cited 1% is used as an impact measure by countries and institutions. Figure 25 shows that while the US and EU have remained constant, China's share has steadily increased between 2002 and 2012. The Nature Publishing Index shows a similar trend (Nature Editors 2014).

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 (the Source Normalised Impact per Paper (SNIP) is one way to account for this; see *Other bibliometric measures*). 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.

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 Collins & Tabak (2014).

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 (see *Open access citation advantage*).

Figure 25: Share of the world's top 1% of cited articles from US, EU and China: 2002–12 (from: (NSF 2014)
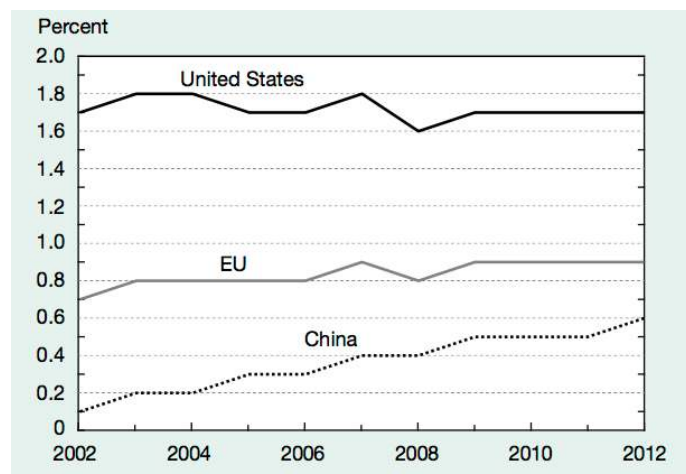
Effects of research assessment on researcher behaviour

Goodhart's Law[27] says that when a measure becomes a target, it ceases to be a good measure. In other word, it stops truly reflecting the original variable, but increasingly measures the effectiveness of the organisation or individual at maximising the measure, and in doing so may also change behaviour in undesirable ways.

---

27 http://en.wikipedia.org/wiki/Goodhart's_law

There is clear evidence that research assessment exercises such at the REF (UK's Research Excellence Framework) or ERA (Excellence in Research for Australia) have changed researcher behaviour. For instance, Adams & Gurney (2014) analysed UK data to show that researchers submit journal articles in preference to the outputs that elsewhere they say are central to their field, they skew their selection to high-impact journals, and they submit pieces [for assessment] from such journals even when they are not well cited and, sometimes, not even research papers. The authors suggest that this is because they believe that the brand of a journal known to have high average impact is a better proxy "signal" in place of real evidence of excellence. Submission behaviour was observed to change over successive RAE cycles leading to a progressive concentration on journal articles. To enable this relative growth, there was a shift out of conference proceedings in engineering and out of scholarly monographs in the social sciences.

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 Source Normalised Impact per Paper (SNIP) uses the Scopus database to measure contextual citation impact by weighting citations based on the total number of citations in a subject field.

- SCImago Journal Rank (SJR)[28] is a freely available journal-level metric. It is a prestige measure based on the idea that not all citations are the same, that is, citations are weighted according to the prestige of the citing journal.

- Google provides Scholar Metrics,[29] a free journal-level citation impact metric based on the h-index

- **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, de Sompel, Hagberg, & Chute, 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.

---

28 http://www.scimagojr.com

29 http://googlescholar.blogspot.co.uk/2014/06/2014-scholar-metrics-released.html

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[30] 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. CiteULike, Mendeley), as well as news media and article-level comments, annotations and ratings.

A number of tools and services have emerged to support the tracking, reporting and visualisation of altmetrics, including Altmetric, PlumX (acquired by EBSCO in early 2014), PLOS Impact Explorer (based on Altmetric), PageCritic, and others. (See *http://altmetrics.org/tools/* for a current if incomplete list.)

As interest grows in tracking impact at the article or author level, authors may increasingly seek to maximise their impact (or at least, their altmetric scores). Two initiatives aimed at supporting this are:

- Kudos[31] aims to help authors expand readership of their research publications and increase citations, via a structured process that includes writing a lay summary and using social media effectively. It is free to researchers but charges publishers for related services

- ImpactStory[32] lets researchers create online profiles which profile their research outputs (papers, datasets, presentations, software), and track and show the altmetric impacts of the same. It charges researchers a subscription of $60/year

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.

It is also worth bearing in mind that citations and usage at the article level are usually characterised as having low levels for the majority of individual articles. The numbers are so low that trying to turn them into a meaningful discriminatory metric will be bedevilled by

---

30 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

31 https://www.growkudos.com

32 https://impactstory.org

the counting error: most articles will have data at the level of statistical noise and be indistinguishable from each other.

Typed citations and contributor roles

At present a citation is a blunt instrument: it is not apparent from the fact of the citation what the author's intent was: agreement, disagreement, etc. To improve the value of citations there have been proposals to "type" citations in a structured way. The benefits would be primarily in text and data mining applications and for visualisation of research networks. One initiative is CiTO, the Citation Typing Ontology (Shotton 2010); also (Shotton 2009). The prospects for authors adopting such a structured process do, however, seem remote at present. (See also *Open annotation*.)

In a similar vein, it is the norm for papers in most fields to have multiple authors and yet the roles of the various authors may vary significantly. Contributor roles might include study conception, methodology, investigation, data analysis and statistics, writing, etc. The roles may be described in the acknowledgement sections of papers (particular in medical journals) but the data is unstructured and inconsistently applied. To address this, a group of editors, journals and publishers are working on the development of a standard taxonomy for describing contributor roles that could be used in STM journals (Allen, Scott, Brand, Hlava, & Altman, 2014; Meadows 2014).

San Francisco Declaration on Research Assessment (DORA)

Dissatisfaction among researchers as well as some journals and publishers with the way research assessment is conducted was made evident in the San Francisco Declaration on Research Assessment (DORA) in late 2012 (American Society for Cell Biology & et al, 2012). The Declaration points out that research outputs are many and varied, and rehearses the arguments against use of the Impact Factor for research assessment. Its key recommendation is to not use journal metrics as a surrogate for article quality for research assessment purposes, but it also makes a number of recommendations for publishers and metrics providers:

- greatly reduce the emphasis on journal impact factor in promotion
- make article-level metrics available to encourage a shift away from journal-level metrics
- remove all reuse limitations on reference lists in research articles and make them freely available
- remove or reduce the constraints on the number of references in research articles
- be open and transparent by providing data and methods used to calculate all metrics
- provide the data under a licence that allows unrestricted reuse, and provide computational access to data, where possible

The declaration had been signed by over 12,000 individuals and about 550 organisations at the time of writing.

Changes in citation behaviours

In addition to the trends outlined above, two recent papers from the Google Scholar team have provided evidence that shows authors are citing a higher proportion of older papers than in the past, and that highly-cited papers are more likely to be found in non-elite journals (Acharya et al., 2014; Verstak et al., 2014). In both cases the authors speculate that

online availability and growing ease of discovery (e.g. via search engines or other discovery tools) of older and more obscure journal content has played a role.

Citations by patents

Citations to STM articles made within patents are sometimes used as another measure of wider impact beyond academe. Citations are typically much older than in the scientific literature, mainly because of the delay in granting patents; for example, the NSF analysis looks at an 11-year window after a 5-year lag. In the US, the proportion of patents citing academic literature increased from 12% to 15% between 2003 and 2012 (NSF 2014). The majority of cited articles fall into three fields: biological sciences (48%), medical sciences (23%), and chemistry (11%).

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 5) 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 wider impact than do citations (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). The COUNTER Code of Practice for Usage Factors is now available in Release 1 (COUNTER 2014).

The 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 (UFJ1) as defined as "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. Two UFJ's may be calculated: the publisher usage factor (based on fulltext usage on the publisher's COUNTER-compliant platform), and the consolidated usage factor (derived from the total usage on a group of COUNTER-compliant platforms).

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. Journal-level usage factors will have application in library acquisition settings and perhaps for authors selecting journals to submit to, but in many cases article-level metrics will be more relevant for the same reasons as discussed above.

Table 5: 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 estimated 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.) 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 RIN (RIN 2008). The largest part of this saving comes not from publisher costs but 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

The potential for open access to effect cost savings has been much discussed (e.g. see *Open access*). However the emergence of pure-play open access journal publishers allows some evidence of average article costs to be inferred from their financial statements:

- PLOS's annual report for 2013/14 shows total costs (including overheads) of $29.6 million for about 34,000 articles published, giving an average of $1088 per article ($871 excluding overheads). This combines the low-cost *PLOS ONE* with the higher-cost

selective journals, suggesting that the average for *PLOS ONE* would have been lower (PLOS 2014)

- eLife's financial statement for its first full year of operation showed total costs of £2644k, equating to an average cost per article of £8370 (or about $14,000 at £/$=1.67; 316 articles were published in 2013). While eLife is a highly selective journal with a costly editorial staff, this figure should be treated with caution, as it is not revealed what fraction of the costs were one-offs or deliberately larger than necessary to build future scale (eLife 2014). Output more than doubled to 636 articles in 2014, which seems likely to have reduced the average significantly

- at the other end of the cost scale, Hindawi was reported in *Nature* as publishing 22,000 article in 2012 at an average cost of $290 per article (Van Noorden 2013). Hindawi uses a low-cost publishing model and is situated in a relatively low-wage part of the world (Egypt)

The same *Nature* article quoted other publishers' stated costs, albeit not supported by published accounts:

- the *Proceedings of the National Academy of Sciences* estimated their average cost at $3700 per published article

- *Nature*'s Editor-in-Chief was quoted as estimating its internal costs as £20–30,000 per paper

- PeerJ said their average costs were in the "low hundreds of dollars" per article

On the other hand, substantial savings at large existing publishers may not be that easy to find: the financial analyst firm Bernstein Research estimated that a full transition to open access would save a subscription publisher around 10–12% of its cost base (Aspesi 2012).

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) noted 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).

This was also illustrated in a 2012 report (Gantz 2012); see also (Gantz 2013) which challenged the common interpretation of the ARL statistics cited above. The report argued that while library serial expenditures had indeed increased three-fold between 1990 and 2010, the ARL libraries' collections had tripled in size through new acquisitions and through expanded content in existing holdings. Average cost per journal was therefore the same as in 1990. The apparent 6-fold increase in journal prices reported by ARL was not incorrect as such, but was based on the list price for print, whereas libraries were now purchasing bundles of electronic content. This was illustrated by the increase in average cost per journal acquired between 1990 and 2000, followed by its decline to 1990 levels by 2010.

2.14. Authors' behaviour, perceptions and attitudes

There have now been numerous studies of author behaviour, perception and attitudes. Two pioneering pieces of work stand out for their large (at the time) 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). Later studies by RIN and Harley have largely extended and amplified the CIBER findings (RIN 2009a; Harley & et al, 2010); while more recent work has documented authors'

evolving attitudes toward open access (Taylor & Francis 2014; Nature Publishing Group 2014).

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% thought 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].

Motivations for publishing

The fundamental needs of researchers with regard to scholarly communication have been studied over the last 20 years or so, and vary depending on their role, that is whether acting as an author or a reader. The core needs of authors are to be seen to report an idea first; to feel secure in communicating that idea; [for empirical subjects] to persuade readers that their results are general and arise from enactment of a canonical (scientific) method; to have their claim accepted by peers; to report their idea to the right audience; to get recognition for their idea; and to have a permanent public record of their work (Mabe 2012).

Looking at the specific motivations for publishing, the most important motivation reported in a 2005 survey was "dissemination" (73%), with "furthering my career" and "future funding" the key secondary motivations. Comparing these results to a similar study in 1993 showed little change in these three motivations or their rank order, but the secondary motivations, "recognition" and "establishing precedent" had clearly increased, especially the latter (Mulligan & Mabe, 2011).

Choice of journal

Multiple surveys has shown that the main factors affecting author choice of journal are the journal's quality, its relevance, and speed of publication (in that order). These attitudes have remained very stable over time. For example, an analysis of 10 years' worth of data from Elsevier's Author Feedback Programme (Mabe & Mulligan, 2011) allowed comparison of data for 2002 and 2009 (incorporating responses from nearly 100,000 researchers) and showed that quality, the relevance and speed of publication remained the most important factors, and ranked in identical order. This overall picture was confirmed in a 2012 survey of UK academics (Ithaka S+R et al., 2013), and was confirmed again in a 2014 Nature Publishing Group survey, which reported the top five factors to be journal reputation, relevance, quality of peer review, Impact Factor, and speed to first decision (Nature Publishing Group 2014).

Author perceptions of, and attitudes towards open access

There is interest in whether the open access status affects authors' choice of journals. Recent surveys suggest that the three main factors remain pre-eminent for most authors, but that OA status is emerging as important secondary factor. For example, a recent NPG survey found that a minority (37%) of science researchers cited immediate open access as a very or quite important factor in journal selection compared to 90–96% citing relevance or quality factors (Nature Publishing Group 2014). For those that chose OA journals, the most frequently given reason was that the journal only offered open access; that is, they had chosen the journal for other reasons. The second most frequently given reason for selecting an OA journal, however, was that they believed research should be open available immediately after publication. Interestingly, funder and institutional mandates were unimportant reasons for choosing OA publication, with the most important stated reasons being the belief that research should be freely available, followed by the belief that OA publications were more widely read.

Open access status may also be a negative factor for journal choice, at least insofar as it involves publication charges: the Ithaka 2012 survey of UK academics found the fourth most important factor in journal choice to be "The journal permits academics to publish articles for free, without paying page or article charges", in this case ahead of speed of publication (Ithaka S+R et al., 2013).

Several large-scale surveys have explored the attitudes of authors towards open access, including NPG's Author Insights surveys, and Taylor & Francis's Open Access Surveys. These both ran annually in 2013 and 2014, allowing some estimation of changes in attitudes (Nature Publishing Group 2014; Taylor & Francis 2014). These surveys are complementary, with the NPG being stronger on the sciences and the T&F on the humanities and social sciences. They contain a wealth of detail but some highlights include the following:

- that there are benefits of open access seem generally accepted: only 11% said OA had no fundamental benefits

- a good majority of researchers believe that open access offers wider circulation (81%) and higher visibility (75%) for their work, and these beliefs strengthened between 2013 and 2014

- about half researchers think OA publication is faster than in subscription journals [although it is unclear whether this is actually the case]

- researchers are divided on whether OA journals are more heavily cited: 29% agreed but 31% disagreed, while 39% were neutral. However, more agreed and fewer disagreed in 2014 compared to 2013

- CC-BY licences are unattractive to a significant fraction of authors: 65% of T&F respondents did not find it acceptable for their work to be used without their prior consent for commercial gain. When asked to state preferences for different open access licences, the most popular choice was CC-BY-NC-ND, ahead of CC-BY-NC and CC-BY-ND, and well ahead of CC-BY which was easily the least preferred option. CC-BY was the most or second most preferred open licence for only 11% of respondents compared to 53% for CC-BY-NC-ND. (T&F respondents were, however, biassed towards the humanities and social sciences.)

- support for *PLOS ONE*-style "soundness not significance" peer review may be ebbing, with levels of support dropping between 2013 and 2014

- the main reasons for depositing articles in repositories were a personal responsibility to make work freely available, and requests for the article from other researchers

- conversely, the main reasons for not depositing articles in repositories were lack of understanding about publisher policy, and lack of available time

- rigorous (but rapid) peer review was the most important of the services author expect in return for a publication charge, closely followed by rapid publication

- looking forward, authors believe journals will remain as the principal publication outlet, demarcating quality research, but a significant proportion of research papers will be published only in subject or institutional repositories that will coexist with journals

Attitudes to peer review

Researchers consistently express support for peer review in the surveys listed above, as well as in surveys dedicated to exploring peer review (see *Peer review* for more detail). Mabe's longitudinal data showed that attitudes towards peer review did not significantly vary during the period 2002–2009 (Mabe & Mulligan, 2011).

Attitudes towards social media and "science 2.0"

This same 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 were bucking traditional publishing practices" and that "once initiated into the profession, newer scholars—be they graduate students,

postdoctoral scholars, or assistant professors—adopt[ed] 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.).

In the past few years, however, numbers of registered users of scientific social networks including Academia.edu, ResearchGate and Mendeley have rapidly grown, suggesting researchers may be becoming more willing to use some kinds of social media or networks for professional purposes. (See *Scientific Social Networks* for more details.)

See also *Social media*.

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, citation rings, peer review rigging, 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)[33] 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 9000 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

---

33 http://publicationethics.org/

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).

Other organisations with an interest in publishing ethics

The International Committee of Medical Journal Editors (ICMJE)[34] provides detailed guidance on ethical matters relating to medical publishing (many of which are equally applicable to other areas), including authorship and contributorship, sharing of research data (including clinical trials data), editorship, peer review, conflicts of interest, privacy and confidentiality, and protection of human subjects and animals in research. The ICMJE Recommendations (previously known as the Uniform Requirements for Manuscripts Submitted to Biomedical Journals) amount to an *ad hoc* standard that is widely adhered to (ICMJE 2013).

The World Association of Medical Editors (WAME)[35] 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 and other automated detection tools

CrossCheck[36] 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 for photo or image manipulation, and data review (digit preference analysis can detect fabricated data).

The arXiv repository has its own dedicated software for screening submission for potential plagiarism. A 2014 study looked at patterns of potential plagiarism within arXiv across the whole corpus of 757,000 articles from mid-1991 to mid-2012. Text reuse was fairly common: after filtering out review articles and legitimate quoting, about one in 16 arXiv authors was found to have copied long phrases and sentences from their own previously published work. About one out of every 1000 of the submitting authors copied the equivalent of a

---

34 http://www.icmje.org

35 http://www.wame.org

36 http://www.crossref.org/crosscheck.html

paragraph's worth of text from other people's papers without citing them. Perhaps the most interesting finding was that the more a paper reuses already published work, the less frequently that paper tends to be cited (Citron & Ginsparg, 2014).

2.16. Copyright and licensing

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 (about two-thirds of large publishers now prefer this grant of licence option (Inger & Gardner, 2013)); 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 typically retain copyright and release the work under a Creative Commons licence or similar (see below) which allows use and reuse but imposes conditions, such as attribution of the author, which depend on copyright. However, OA under a traditional copyright regime is also possible in principle.

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.

Copyright reforms

The most recent reviews of copyright in the UK and the EU (the Hargreaves report and subsequent government consultations[37], and Copyright in the Information Society[38] and subsequent Licences for Europe[39] programmes 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

---

37 http://is.gd/3il5QW

38 http://ec.europa.eu/internal_market/copyright/copyright-infso/index_en.htm

39 http://ec.europa.eu/licences-for-europe-dialogue/

- *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

- *"Out-of-commerce"* works are works that are still protected by copyright but are no longer available through normal channels of commerce (for example, out-of-print books where the publisher does not intend to reprint or issue an ebook edition). Various governments have proposed or enacted the right of public or national libraries or cultural institutions to make these available via digitisation

UK copyright changes

In the UK, a number of the key recommendations made by Hargreaves that were relevant to publishers have now been at least partly implemented (Hargreaves 2011; Intellectual Property Office 2014):

- a copyright exception to allow text and data mining (TDM) has now been implemented, despite this being an active area of development in STM (the Select Committee preferred to see publishers developing usable and affordable licensing schemes). This permits users "make copies of works 'for text and data analysis'", provided this is for non-commercial research, and that copies are accompanied by "sufficient acknowledgement" (where practicable). (See also below, *Text and data mining*)

- the UK government announced[40] funding in March 2013 for the Copyright Hub and Digital Rights Exchange, which had been proposed by Hargreaves to make it easier to get copyright clearance, including for the case of "orphan works" (ones where the original copyright holders cannot be traced). It is intended to provide a resource of information about copyright; a focal point for registries of different rights and their owners; and a platform for licensing transactions

- a copyright exception for format-shifting: this came into effect on 1 October 2014 despite objections from a variety of rights-holders. The exception only covers personal copying, however, so falls short of the rights that academic libraries and the British Library were hoping for to facilitate long-term preservation

- Hargreaves also recommended that copyright exceptions could not be overridden by contract. This was implemented with respect to the new copyright exceptions (quotation, parody and private copying/format shifting)

EU copyright consultations and developments

As noted above, the EU has conducted consultations around possible changes to copyright within the Copyright in the Information Society and subsequent Licences for Europe programmes, which concluded in November 2013. This was followed by a further public consultation on review of EU copyright rules, with a report and proposals for next steps issued in July 2014.[41]

---

40 See: Government gives £150,000 funding to kick-start copyright hub, https://www.gov.uk/government/news/government-gives-150-000-funding-to-kick-start-copyright-hub

41 http://ec.europa.eu/internal_market/consultations/2013/copyright-rules/index_en.htm

The agenda has been similar to those of the UK Hargreaves report, including important STM issues such as text and data mining, but also included matters such as cross-border access and portability of services, and user-generated content and micro-licensing. The Commission is working a white paper due for release in late 2014, "A copyright policy for Creativity and Innovation in the European Union", setting out its proposed changes; an internal draft was leaked in July 2014 (Baker & McKenzie 2014).

The Commission's so-called ongoing initiatives are of less relevance to STM publishers, include out-of-commerce works, private copying levies, access to copyright works for people with print disabilities, and the online distribution of audiovisual works.

US and other territories

In the US too there is an active debate on the need for copyright reform. The Register of Copyrights already 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 "out-of-commerce" works.

The reform agenda is, however, substantially wider. Maria Pallante, the US Register of Copyrights testified before Congress in March 2013 to the need for reform in a wide list of areas, including: "clarifying the scope of exclusive rights, revising exceptions and limitations for libraries and archives, addressing orphan works, accommodating persons who have print disabilities, providing guidance to educational institutions, exempting incidental copies in appropriate instances, updating enforcement provisions, providing guidance on statutory damages, reviewing the efficacy of the DMCA, assisting with small copyright claims, reforming the music marketplace, updating the framework for cable and satellite transmissions, encouraging new licensing regimes, and improving the systems of copyright registration and recordation."

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 reuse 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. Recent surveys of authors confirm that confusion about copyright and their retained rights persist (e.g. Taylor & Francis 2014).

Model licences

Model and sample licences have been developed by a number of organisations including publisher organisations, intermediaries, and purchasing bodies. Use of such licences is desirable for two main reasons: it simplifies transactions and the operation of the market, and because the licences typically represent "best practice" following substantial consultation and negotiation among interested parties. Examples include:

- IFLA Licensing Principles: these are in fact not a model licence but a set of principles governing contracts between libraries and publishers. Originally drafted in 2001, most recently updated May 2014. http://www.ifla.org/publications/ifla-licensing-principles-2001

- LicensingModels: a set of licences for electronic resources originally developed in 1999 in collaboration with the major subscription agents and subsequently extended by John Cox Associates (the site is now maintained by Ringgold). Licences included academic libraries, academic consortia, corporate library, public library, ebooks, and 30/60-day free trials. http://www.licensingmodels.org

- P-D-R Model Licence was developed by ALPSP, STM and the Pharma Documentation Ring covering licence terms between publisher and pharmaceutical companies. The 2012 update includes a new clause with guidance on rights for text and data mining. http://www.p-d-r.com/content/publications/

- Text and data mining (TDM): STM has developed sample licences covering TDM of subscribed content, and for TDM of previously un-subscribed content. http://www.stm-assoc.org/text-and-data-mining-stm-statement-sample-licence/

- STM open access licences: see *Open access licences* below

- JISC Model Licences apply of course only to JISC agreements, including NESLi2, with sublicences for archives, databases, and SHEDL. http://www.jisc-collections.ac.uk/model_licence

- in the US there is no equivalent national procurement, but model licences include the LIBLICENSE model licences, as well as those created by various large consortia. http://liblicense.crl.edu/licensing-information/model-license/

SERU

SERU (Shared Electronic Resource Understanding) Recommended Practice is a NISO Best Practice. It provides an alternative to a licence agreement where library and publisher agree, primarily designed for (and utilised in) the North American market. The SERU statement expresses commonly shared understandings of the content provider, the subscribing institution and authorised users; the nature of the content; use of materials and inappropriate uses; privacy and confidentiality; online performance and service provision; and archiving and perpetual access. The benefit is to simplify procurement of electronic resources by avoiding the need for a bilateral licence.

Originally adopted in 2008 for e-journals, it was updated to its current version in 2012 which covers a wider range of content including ebooks. Publishers, libraries and consortia that are willing to use SERU join the registry (available at its website), though this does not commit them to using it for future orders (NISO SERU Standing Committee 2012).

Open access licences

For open access journals, the article is released under a licence that allows users to access, copy and reuse the content under specified circumstances. From the authors' perspective, the typical arrangement is for them to retain copyright but to sign a licence agreement with the publisher allowing the latter to issue the work under the specified open access licence, although other arrangements are possible.

The licences most frequently used for open access journals are those developed by Creative Commons. The latest versions (v4.0) were launched in November 2013; the main area of development compared to v3.0 was further internationalisation; improved interoperability with other licences; anticipation of future developments to make them longer lasting; and specific requirements for data, science and education.

Creative Commons are sometimes described as "some rights reserved" (in contrast to the "all rights reserved" copyright statement); the principle is quite different from placing material in the public domain (i.e. waiving rights). The licences come in multiple flavours:

- CC-BY allows users maximum freedom in re-using content: essentially all copying and reuse is permitted provided the author (copyright holder) is acknowledged, including the creative of derivative works, and reuse for commercial purposes. This is the licence preferred by most open access advocates including the Open Access Scholarly Publishers Association (OASPA)

- CC-BY-NC is the same as CC-BY except the reuse for commercial purposes is not allowed (without first obtaining permission, as with standard copyright). Many open access advocates prefer the CC-BY licence, arguing that commercial use is a fuzzy term, and that allowing commercial exploitation of publicly funded research is in the public interest

- CC-BY-NC-ND additionally exclude the creation of derivative works. OASPA does not permit its use by its members because it sees derived use as fundamental to the way in which scholarly research builds on what has gone before

- CC-BY-SA: the "share-alike" rider requires those creating derivative works to attach the same share-alike licence. This is (perhaps surprisingly) deprecated by most open access advocates; for example, OASPA does not permit its use by its members because material distributed within a share-alike article could only be combined and redistributed with other share-alike content

New open access model licences were released by STM in August 2014.[42] These were intended to be complementary to Creative Commons licences. They were designed to cover recent developments such as multi-language access, text mining, and also the specific instance of commercial use to cover paid advertising being associated with open access content.

The licences were not well received by open access advocates and campaigners; a coalition (including funders, institutions, publishers, curators and the users of public resources) issued a statement calling on STM to withdraw them (*Global Coalition of Access to Research, Science and Education Organizations Calls on STM to Withdraw New Model Licenses* 2014). STM's statement in response (STM 2014) noted that there were multiple views on the issue, including among its own members: some preferred the efficiency of standardising on a single licence, while others preferred to offer choices and options to authors that may reflect

---

42 http://www.stm-assoc.org/open-access-licensing/

particular concerns. In the end this seems likely to be a matter that will be decided by the market.

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 remain under active debate and in flux.

At the time of writing, it was still relatively 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; Inger & Gardner, 2013). 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.[43]

In a parallel public response to these issues, the Hargreaves report proposed a copyright exception for TDM; this has now been enacted in the UK, but fails to 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 for this more general case could be a comprehensive licensing process, covering multiple publishers. A small but important step was taken in 2012 with the model licence terms to cover TDM agreed by STM, ALPSP and P-D-R (Pharmaceutical Documentation Ring).[44]

Publishers have issued statements of commitment to facilitating TDM for non-commercial use (STM 2013a). In this regard STM has developed model licence terms that could be added to existing publisher-library licences to support TDM under defined terms (STM 2012).

A number of more ambitious cross-industry collaborative initiatives have emerged, notably those led by CrossRef, CCC and PLS, which are discussed below in the section on *Text and data mining*.

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 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

---

43 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

44 http://is.gd/UXnRMl

necessary: the licences can simply assert ownership and specify allowed downstream uses and licensing requirements.

The Linked Content Coalition[45] worked between 2011 and 2013 to develop a Rights Data Network to provide "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. The Rights Data Network would be "a network of authoritative linked data in which all key entities in the rights data network had standard, resolvable identifiers; these identifiers were linked in standard ways; and the management of the identifiers and links was under registry procedures which ensure that they are under appropriate authority, and that parties with a legitimate interest in an entity can make sure that interest is correctly and publicly recognised."

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 elsewhere. At this stage, however, the challenges appear to be at least as much organisational as technical.

The main solutions currently in use are as follows:

- National library services: the earliest and best known of these is the e-Depot at the Koninklijke Bibliotheek.[46] Its digital archiving services are available to publishers worldwide and are used by many major publishers including Elsevier, Springer, Wiley 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).[47] As the name suggests it works on the principle of redundancy, similar to the way that multiple print journal holdings provide

---

45 http://www.linkedcontentcoalition.org/

46 http://www.kb.nl/en/organisation/research-expertise/long-term-usability-of-digital-resources/information-for-international-publishers

47 http://www.lockss.org

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)[48] is a sustainable collaborative organisation of some 200 scholarly publishers and 750 research libraries based on the LOCKSS technology.

- Portico is a not-for-profit preservation service for scholarly content[49], 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 late-2014 it had about 920 participating libraries, 274 publishers covering more than 20,600 journals and 340,000 ebooks, and other content, representing a total of 705 million files.

- The Alliance for Permanent Access[50] (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) is a network of excellence supporting research into barriers to the long-term accessibility and usability of digital information and data.

According to an ALPSP report (Inger & Gardner, 2013), Portico, followed by LOCKSS/CLOCKSS, was the most popular option for both large and medium publishers. All of the large publishers in the survey had some kind of archival arrangements, but nearly a fifth (18%) of small publishers did not.

2.18. TRANSFER code

The *UKSG Transfer Code of Practice*[51] 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 3.0, and was endorsed by some 50 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. Phillpotts and colleagues provide more background on the evolution of the Code (Phillpotts, Devenport, & Mitchell, 2015).

---

48 http://www.clockss.org/clockss/Home

49 http://www.portico.org

50 http://www.alliancepermanentaccess.eu

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

2.19. Researchers' access to journals

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

Statistics show that the number of journals acquired per library has increased dramatically since the advent of electronic journals in the late 1990s, and the cost paid per journal has fallen. 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); RIN 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 consultations made by government bodies (OSTP 2012; European Commission 2012b; European Commission 2012a); 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. A survey conducted by CIBER in late 2011 on behalf of RIN (RIN 2011a) analysed 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 proportion (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.

The Ithaka survey of UK academics found that when a wanted item was not held in the library collection, the highest share of respondents reported that they look for a freely available version online, while the second highest share just gave up, both of which outranked using the library's interlending or document supply service (Ithaka S+R et al., 2013).

This was the context in which 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 (see *RCUK policy*): increased funding for national licences to extend and rationalise coverage; walk-in access to the majority of journals to be provided in public libraries (see *Public access*); 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.

There has been relatively little further research on this issue since the 2009 survey mentioned above. (Houghton, Swan, & Brown, 2011) investigated access by SMEs in Denmark, looking at levels of access and use, whether there were any barriers to access, access difficulties or gaps, and the costs and benefits involved in accessing research findings. Access to academic research was found to bring substantial benefits. Twenty-seven per cent of the products and 19% of the processes developed or introduced during the last three years would have been delayed or abandoned without access to academic research, with these new products contributing an average 46% of annual sales. About half of respondents rated research articles as very or extremely important, and a similar proportion (55%) reported difficulties accessing research articles. The most widely used means of access to non-open access

materials were personal subscriptions and in-house library or information services. Public libraries, inter-library loans and pay-per-view (PPV) were little used.

Public access schemes may help SMEs, though walk-in access (such as the UK *Access to Research* pilot programme) were not highly rated in the PRC research (see *Public access*).

In the past few years new services have launched offering access to journal articles at prices lower than the full "pay per view" price on the publisher. Providers include DeepDyve, Proquest Udini, ReadCube Access, and Infotrieve (Copyright Clearance Center). The access is limited either by time (article rental) or by features (e.g. disabling printing and local saving), with business models including one-off charges, or monthly or annual plans, and plans for groups or companies. In a similar vein, Reprints Desk's Article Galaxy Widget allows users to search for articles and then find the lowest cost access option available.

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[52] are collaborations between UN agencies, STM publishers, universities and university libraries, philanthropic foundations and technology partners. The partnership's goal is to help attain six of the UN's eight Millennium Development Goals by 2015, reducing the scientific knowledge gap between industrialised countries and the developing world. There are currently four programmes that collectively provide some 7,700 institutions in 109 developing world countries with free or low cost access to over 30,000 journals from 180 countries and other fulltext resources:

- 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

- 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

- 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

- ARDI (Access to Research for Development and Innovation) 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).

---

52 http://www.research4life.org

Other schemes include:

- HighWire Press offers free access for developing countries to a list[53] of about 500 high-quality journals, based simply on software that recognises from where the user is accessing the site

- Some publishers offer similar schemes independently, e.g. the Royal Society of Chemistry, the National Academies Press

- INASP's PERI scheme ended in 2013 but INASP continues to support access to research through its Strengthening Research and Knowledge Systems[54] programme

- eIFL (Electronic Information for Libraries)[55] partners with libraries and library consortia to build capacity, advocate for access to knowledge, encourage knowledge sharing and initiate pilot schemes for innovative library services

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.[56]

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.

---

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

54 http://www.inasp.info/en/work/what-we-do/programmes/srks/

55 http://www.eifl.net

56 http://www.authoraid.info/

3. Open access

Open access refers to the making available of content (especially journal research articles, though there is growing interest in open access for other research outputs including monographs and conference proceedings) in online digital copies, free of charge, and free of many or 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).[57]

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 is a matter of debate and has assumed greater importance since the last edition of this report, with some funders requiring CC-BY licensing while substantial proportions of authors show reluctance to waive some rights such as commercial reuse rights (see *Open access licences*). Another factor is the growing importance of text and data mining, although this is more complex since TDM solutions need to work across corpuses of both OA and non-OA content (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.

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; subscription-based 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. 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, and entrepreneurial activity which has increased the supply of credible open access journals to authors.

---

57 e.g. see http://www.earlham.edu/~peters/fos/overview.htm

Research funder policies have likely been the most important factor in creating an environment for open access. Notable milestones 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[58] (June 2012), in particular introducing sanctions for non-compliance and a move to CC-BY licences

- the UK Research Councils new unified policy (announced July 2012 and introduced in April 2013), which largely develops the Finch recommendations (RCUK 2012); see also *RCUK Policy* below)

- the European Union's new 7-year research programme, Horizon 2020, which came into effect in 2014 and covers the EU's €80 billion funding. Requirements on authors were tightened, with a target for 60% of funded openly available by 2016 (de Vrieze 2012). The policy supports Gold OA via reimbursement of APCs; in addition, a version of all articles (including Gold) must be deposited in an open archive no later than publication, and made openly available within 6 months (or 12 months for HSS)

- in the US, the debate has been around NIH mandate and its possible extension. Competing legislation (Research Works Act (RWA) and the Federal Research Public Access Act (FRPAA) were abandoned but a third piece of legislation, the America COMPETES Bill was passed in 2011, and required the Office of Science and Technology Policy (OSTP) to coordinate access policies across the federal funding agencies. This led to the OSTP issuing a memo in 2013 that required the larger agencies to develop plans to provide public access – see *Office of Science and Technology Policy* below

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 (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 such as Nature Publishing Group and the AAAS, and 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 some policy-making and funders towards the Gold model took place around 2012. This was 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

---

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

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.

The momentum towards to policy preference for Gold appears now to have slowed. While it is the case that, as pointed out by HSBC in an analysis of open access policy the academic publishing sector (Graham 2013), essentially all important research funders support Gold open access to the extent that APCs are reimbursed, very few express a preference for Gold, and many actively promoted Green over Gold.

3.2. Open access business models

As open access has grown in scale and matured, and has expanded to disciplines in which research funding is not as important as in the experimental sciences, it has become clear that a single simple business model will not suit all situations. While the basic APC model remains important, a substantial number of variations on this theme have emerged, as well as growing interest in non-APC models (see Table 6).

Table 6: Open access business models. (See also Björk & Solomon (2012b) for a more information on pricing approaches used in OA journals)

Model / strategy	Description	Examples
Article publication charge (APC)	Fee levied on acceptance to cover costs of publication and related services. Various discounts and waivers are common	Widespread
Page & other publication charges	Additional charges levied on top of basic APC, e.g. for mss longer than specified limits, inclusion of colour / rich media, etc	Science Advances (AAAS); PhysRevX; Some hybrid journals where colour charges are standard
Submission fees	Non-refundable fee payable on submission regardless of outcome of peer review A possible viable model for high-rejection rate journals	rare: e.g. Hereditas; JMIR (submission fees are surprisingly more common in subscription-based journals)
Prepayments	Block purchase of APCs in return for discounts	Taylor & Francis
Institutional memberships	A package of other relevant models such as institutional-based discounts, prepayment, bundling, offsetting, etc.	BMC; PLOS
Offsetting	Capping of total subscription + OA charges to a particular institutional or consortial customer	IOP; RSC; JISC "Total Cost of Ownership"; Austria FWF

Model / strategy	Description	Examples
Bundling	Combination charge covering subscription/licence fees plus OA publication charges or institutional membership	Springer
Institutional-based discounts linked to subscriptions	Discounted APCs for authors at institutions subscribing to other journals from same publisher	OUP
Individual membership	Individuals purchase memberships for one-off fees (tiered); all coauthors must be members (up to maximum number); members required to participate (e.g. via peer review) to remain in good standing	PeerJ
Freemium	Open access to a basic online version plus charges for additional or added-value services (print, enhanced electronic formats, etc.)	JMIR; OECD (data/books); Knowledge Unlatched (books)
APCs supported by third party	Often intended as transitional support r.t. permanent model Discounted (or zero) APCs Supported by societies, institutions, foundations, etc.	Some BMC transfers-in; MedKnow; Versita (De Gruyter Open); eLife (at present)
Sponsorship (non-APC models)	Sponsors cover costs with no intention to adopt APCs Sponsors include: societies; institutions, research organisations; foundations; research funders. Volunteerism may also be involved	Eurosurveillance (ECDC) Clinical Phytoscience (Springer) Asia & the Pacific Policy Studies (Wiley) Journal of Law and the Biosciences (OUP/Duke/Harvard/Stanford)
Library Partnership Subsidy	Creation of new library consortia for the collective funding of open access publishing	Open Library of the Humanities SCOAP3 Knowledge Unlatched (monographs)

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 costs are paid for in a way that does not require payment for access. There are two main variants:

- Immediate full OA: the entire contents of the journal are made freely available immediately on publication.

- 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.

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 (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*).

Hybrid journals

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 (~1–2% or so; see *Open access article numbers*).

Bird reported hybrid uptake rates for OUP journals (where she works) and for some other publishers (Bird 2010). 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 quoted 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 arose subsequent to Bird's article is *Nature Communications*, which had an opt-in rate (at $5000 per article) of over 40%, though it has now converted to full open access.

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). 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. (See also *Offsetting*.)

Offsetting

In response to concerns about double-dipping and pressures from some large consortia purchasers, a report commissioned by Wellcome and other bodies recommended three possible approaches to making the hybrid market more transparent and competitive (Björk & Solomon, 2014). These included refunding APCs at list price; tiered APCs with price caps; and funders only reimbursing part of the APC with universities making up the difference.

Separately, Jisc Collections developed an approach it called "Total cost of ownership" of scholarly communication: managing subscription and APC payments together (Lawson 2015). Pilot schemes have been agreed with a number of publishers. Options under

consideration include publishers offering credits against future APCs when subscriptions are taken out; or conversely to offer credits against future subscription payments when APCs are paid; or bundling of subscriptions with future APCs for modest additional payments.

Similar offsetting models have also emerged in the marketplace in which open access charges are offset against subscriptions or otherwise managed as a single procurement. Examples include a pioneering "Gold for Gold" model by RSC,[59] offsetting schemes from IOP Publishing (Jump 2014a).

Article publication charges

Table 7 lists a selection of 2014/15 publication charges from major societies and commercial publishers. Fees for full open access journals mostly fall in the range $1000–5000, except for Hindawi with a median $600 charge, and with lower fees of $400–1000 typically charged for case reports, short communications and some areas without much research funding. Hybrid APCs tends to be more expensive and in a narrower range, typically around $3000. The situation is also complicated by a wide range of discounts, bundling and offset models (see Table 6). Some journals impose additional charges (e.g. based on length, or for more rapid peer review, or for different licences).

The range of fees shown in Table 7 is reflected in the reported average APC paid by the research funder Wellcome Trust of $2365, though this included a substantial proportion of hybrid APCs. The average APC for all open access articles appears to be lower, however: a 2014 study based on OA journals indexed in Scopus (Björk & Solomon, 2014) found the average APC for full OA journals between $1418 (pure OA publishers) and $2097 (subscription publishers), while that for hybrid OA was $2727. This report also showed more variation in APCs between journals for full OA than for hybrid, and that APCs correlated with Impact Factor for full OA, but not for hybrid.

These averages are higher than in an earlier study by the same authors; one explanation is that the former used Scopus rather than DOAJ as the sampling source, and hence better established (at least two years old) and likely more professionally published journals. The earlier study looked at 1,370 OA journals that published a total of 100,697 articles in 2010 (Björk & Solomon, 2012c). The authors 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 suggested the market average APC seen here was 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 ALPSP Scholarly Journals Publishing Practice report also provides information on average APCs and their distribution (Inger & Gardner, 2013): the median APC by publisher was $1350 for full OA journals and $2500 for hybrid OA. The authors comment that compared to earlier editions, average APCs for full OA had fallen while those for hybrid largely remained around $3000.

Another study using the DOAJ dataset in May 2014 reported a mean APC of $1221 (median $1145) for a sample of non-zero-APC journals (Morrison & et al, 2014), p.9). This study also reported a long list of APC discounts and surcharges (p.11).

---

59 http://www.rsc.org/Publishing/librarians/GoldforGold.asp

The move to the most permissive reuse 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 requirement of funders like the Wellcome Trust and RCUK that Gold fees will only be paid to publishers who use the Creative Commons CC-BY licences, thus essentially removes 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 (e.g. until recently Nature Publishing Group used this model, with a roughly 10% premium for CC-BY over CC-BY-NC-ND, though it subsequently adopted CC-BY as its default licence).

Nonetheless the fees shown here are mostly lower than the previously reported existing industry average full cost per article (based on e.g. (RIN 2008). This may be one reason why hybrid APCs are higher than for pure OA journals.

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. For example, PLOS's annual report showed that it provided waivers amounting to about 8% of APC income (PLOS 2014).

Table 7: Publication charges for a selection of full and hybrid OA journals. Various discounts (society members, subscribing/"member" institutions, low-income countries, etc.) not shown. Zero APCs are subsidised by third parties, not promotional offers (Source: publisher websites, Dec 2014; £/$=1.6, €/$=1.3)

Journal/publisher	*Full/Hybrid OA*	*APC, full (US$)*	*APC, hybrid (US$)*
American Institute of Physics	Full/Hybrid	1350–2200	1500–2500
American Physical Society	Full/Hybrid	1700	1700-2700
BioMed Central	Full	1015–2650 (median 2075) (exc. zeros)	–
BMJ Group	Full/Hybrid	2160–2720	3120–4800
Cambridge University Press	Full/Hybrid	600–1600 (STM)	2700 (STM) 1500–2700 (HSS)
Elsevier	Full/Hybrid	500–5000	3000 (most)

Journal/publisher	Full/Hybrid OA	APC, full (US$)	APC, hybrid (US$)
F1000 Research	Full	1000	–
Hindawi	Full	300–2250 (median 600)	–
Nature Publishing Group	Full/Hybrid	1350–5200	3975 (CCBYNCND) 4400 (CCBY)
Oxford University Press	Full/Hybrid	0–2270	1600–4000
PLOS	Full	1350 (PLOS ONE) 2250–2900 (Others)	–
Royal Society (London)	Full/Hybrid	2160	2880
Springer (see also BMC above)	Full/Hybrid	1070 (SpringerPlus) 0–1700	3000
Wiley-Blackwell	Full/Hybrid	0; 800–4500	3000 (most)

"Flipped" journals

Converting a journal from subscriptions to open access is known as "flipping" the business model. The Open Access Directory[60] lists some 196 examples (unfortunately not all accurate), while Table 8 provides a more selective list of examples.

The factors publishers will take into account in identifying candidates for flipping include: a modest subscription revenue; expected longer term growth in authorship than in subscriptions; higher rejection rates; attractive to authors; available (and used) funding for OA in the discipline; the volume of existing hybrid articles; and the ratio of current revenues to published articles (Jones 2014a).

---

60 http://oad.simmons.edu/oadwiki/Journals_that_converted_from_TA_to_OA

Table 8: Examples of flipped journals

Journal	Publisher	Launched	Flip	IF	APC
Nucleic Acids Research	OUP	1974	2005	8.8	$1,450
Acta Veterinaria Scandinavica	BMC	1959	2006	1.38	$2,350
Genetics Selection Evolution	BMC	1960	2009	3.75	$1,745
Evolutionary Applications	Wiley	2008	2012	4.57	$1,950
Aging Cell	Wiley	2002	2014	5.94	$2,800
Cancer Science	Wiley	1907	2014	3.53	$2,250
Influenza & Other Resp. Viruses	Wiley	2007	2014	1.89	$2,500
Journal of Diabetes Investigation	Wiley	2010	2014	1.50	$3,000
Developmental Cognitive Neuroscience	Elsevier	2011	2014	3.71	$1,500
Stem Cell Research	Elsevier	2007	2014	3.91	$1,800
Int J Infectious Diseases	Elsevier	1996	2014	2.33	$1,750
Epidemics	Elsevier	2009	2014	2.38	$1,800
EJC Supplements	Elsevier	2003	2014	-	$3,000
"Central European Journal of" series (x8)	De Gruyter	~2009	2015	0.43	tba
Nature Communications	NPG	2010	2015	10.74	$5,200
Chemical Science	RSC	2011	2015	8.60	tba

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 *The BMJ*. The open access publisher BioMed Central also uses this model for a few journals.

Non-APC models

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, as a majority of *articles* published in OA journals probably do make charges (e.g. Dallmeier-Tiessen, S et al 2010). Nonetheless, the scale of non-APC Gold open access appears substantial (see Table 2; Elsevier, 2013).

Sponsored or subsidised OA journals that do not charge authors or readers are becoming a little more common at larger and commercial publishers, as well as in their more traditional homes (research organisation, societies, etc.). In some cases these may be transitional arrangements intended to attract authors as the journal becomes established (e.g. *eLife*) but in other cases there is no intention to introduce APCs.

Non-APC models are of particular interest in the humanities and social sciences where research funding is much lower than in the experimental sciences (Edwards 2014). Historians have been notably vocal in rejecting the APC model (e.g. Mandler 2014).

Library partnership subsidy

Library partnership subsidy (also called consortia open access) is a non-APC model involving the creation of *ad hoc* library consortia for the collective funding of open access publishing. It is a new model with relatively few examples, which while sharing this core concept, differ significantly in their details:

- The Open Library of the Humanities is using the library partnership subsidy model to fund its new, PLOS-inspired humanities megajournal platform. Indicative annual charges are around $925 per library, assuming 200 libraries participate and 250 articles per year[61]

- Knowledge Unlatched (monographs): this rather complex model creates a market platform on which members of a participating libraries can opt in to tranches of monographs proposed by participating publishers. The fee charged to these libraries allows the publisher to upload a basic online version (e.g. a simple HTML layout), which is then freely available to all. The price paid per library declines as the number of libraries join. Publishers retain the right to sell print and higher added-value versions such as ebook formats and digitally enhanced editions. Despite its complexity, however, KU appears to be gaining traction and released its first pilot tranche of titles in 2014[62]

- The high-energy physics consortium led by CERN, SCOAP3, might also be classified as a library partnership subsidy model. In this case, the conversion of almost all the leading titles in the field from subscriptions to open access was achieved, following a process of soliciting library pledges and a subsequent tender (see *SCOAP3*)

---

61 https://www.openlibhums.org/about/library-partnership-subsidy-lps-flyer/

62 KU sought support from a minimum of 200 libraries to unlatch its Pilot Collection. Over 250 had signed up by the deadline of February 2014. By October some 13,000 downloads had been recorded, averaging 40 per book per week.

- The arXiv preprint server has tripartite funding: roughly one third from the Simons Foundation, one third (in kind) from Cornell library, and the final third from collective subsidy from (most of) the institutions making most use of the arXiv.

One of the unresolved issues with this model is the "freeloader" problem: a library (and its patrons) get access whether or not they participate in the funding. Knowledge Unlatched has thought hard about this, and attempted to build in incentives to avoid it (e.g. discounts on the print version). However much libraries may wish to support open access, without some restraining features (e.g. multi-year contracts) it will surely always be easier to cut a subsidy budget (where content remains freely available) than the subscription to other, equally valuable, content.

A similar model has been proposed in a white paper by K | N Consultants. This envisages a central administrative organisation and an independent review panel overseeing the funding of publishing via grants, with funding raised from universities and research institutions, funders and donors (K | N Consultants 2014).

3.3. Types of open access journal

As well as categorising by their business models, it may also be helpful to describe the main types of OA journal, as shown in Table 9.

Table 9: Types of open access journals

Journal model / strategy	Business models	Description	Examples
Megajournals	APCs Institutional memberships Individual memberships	Broad scope "Objective" peer review Low cost / high volume	PLOS ONE; Optics Express; BMJ Open; Scientific Reports (NPG); SAGE Open
Broad-scope selective journals	APCs Library Partnership Subsidy	Broad scope, as for megajournals, but with traditional selective peer review	Open Biology (Royal Society); Nature Communications (NPG); Open Library of the Humanities
Topical OA journals	APCs Page / publication charges Submission fees APCs supported by third party Sponsorship (non-APC models)	"Standard" research journal but OA rather than subscriptions Increasingly the norm for new launches (e.g. all Springer launches since 2011)	Most OA journals not in the above categories

Journal model / strategy	Business models	Description	Examples
Cascade / second tier journal	APCs APC may be shared with referring journal	Often a positioning strategy to test OA while protecting flagship journals Common society strategy Case Reports a growing sub-category	J Nutrition Science (CUP); Physiological Reports (Wiley / Physoc / APS)
Publisher & multiple society collaboration	APCs APC is typically shared with referring partner where relevant	Multiple society partner journals refer rejected papers (cascade) Partners share APCs	Wiley Open Access journals – e.g. Ecology and Evolution; Immunity, Inflammation and Disease; Energy Science & Engineering; etc.
Flipped journals	As for topical OA journals	Journal converted from subscription to OA model	see Table 8 Open Access Directory lists 196 examples (not all accurate) http:// is.gd / 456h5v
Hybrid journals	APCs	Subscription journal of any type but with optional paid-for OA at article level	Most subscription journals in biomedicine published by large publishers
OA Conference series	OA with charge to organisers, based on volume and services included (equivalent to APC)	Proceedings published as a serial May include services to organisers, e.g. use of online tracking system Low cost, semi-automated processes	IOP publishing – J Phys Conference Series

Megajournals

The fastest growing part of the open access market is the "megajournal" sector. This publishing model, pioneered by *PLOS ONE* has proved highly successful and arguably represents one of few innovations to the scholarly journal model yet 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 has been 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; see Table 10 for examples. 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: for example, *Physical Review X* (APS); *Open Biology* (Royal Society); *Cell Reports* (Elsevier); *Nature Communications* (NPG); *Science Advances* (AAAS).

During 2014, however, *PLOS ONE*'s previously consistent year-on-year growth was interrupted: by the end of the year monthly output had fallen by 25% from its December 2013 peak, prompting speculation as to the causes (Davis 2014), although the total for the year was in fact less than 5% down on 2013.

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 still unclear whether an "all subjects" journal will really be stable in the long-term, considering its size, and if not, whether it revert to something more like a collection of traditional subject-based journals, or evolve towards a newer article- (and data-) based model of scholarly communication. These factors go to the heart of the fundamental forces that have shaped journal publishing.

Table 10: Open access megajournals, with date of launch and output (note, broad-scope selective journals like Science Advances, Nature Communications, Palgrave Communications, eLife, etc. are not included)

Name	1st Publication	Total Output (as of 2014)
Optics Express (OSA)	1997	27,290
Zootaxa	2001	15,713
PLOS ONE	2006	114,211
Ecosphere (Ecological Society of America)	2010	526
mBio (American Society of Microbiology)	2010	1,018
FEBS Open Bio	2011	255
AIP Advances	2011	1,390
BMJ Open	2011	2,987
SAGE Open	2011	822
QScience Connect	2011	91
G3 (the Genetics Society of America)	2011	719

Name	1st Publication	Total Output (as of 2014)
Scientific Reports (Nature)	2011	7,325
Springer Plus	2012	1,392
Cureus	2012	232
The Scientific World Journal (Hindawi)	2012	4,914
F1000 Research	2012	707
Biology Open (Company of Biologists)	2012	451
PeerJ	2013	736
SAGE Open Medicine	2013	93
CMAJ Open (Canadian Medical Association)	2013	66
BMJ Open Respiratory Research	2013	36
Open Heart (BMJ)	2013	74
IEEE Access	2013	186
Journal of Engineering (IET)	2013	122
Royal Society Open Science	2014	52
BMJ Open Diabetes Research & Care	2014	32
Elementa (BioONE)	2014	29
Frontiers in Medicine (Frontiers/NPG)	2014	32
Frontiers in Surgery (Frontiers/NPG)	2014	46
Open Library of the Humanities	2014	-
Cogent Economics & Finance (Cogent/T&F)	2014	56
Others in Cogent Series (T&F) – 15 planned in total	2015	-
Collabra (U California Press)	2015	-
[Unnamed at time of writing] (Elsevier)	2015	-

3.4. 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. The best known were the DC Principles Group of society publishers using the HighWire system, primarily in the life sciences. Although the group's free access articles are no longer shown separately, the HighWire platform currently hosts a total of over 2.5 million freely available articles, of which the majority are from delayed access journals.[63]

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.

---

63 http://highwire.stanford.edu/lists/freeart.dtl

A study in 2012 by Laakso and Björk identified 492 journals using this model, publishing a combined total of 111,312 articles in 2011. About 78% of these articles were made open access within 12 months from publication, with 85% becoming available within 24 months. Delayed OA journals have on average twice as high average citation rates compared to closed subscription journals, reflecting the fact that many were leading society journals in their fields. The authors concluded that delayed OA journals constituted an important segment of the openly available scholarly journal literature, both by their sheer article volume as well as by including a substantial proportion of high impact journals (Laakso & Björk, 2013).

3.5. 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 someone acting on their behalf) deposits the article in an open repository or other open web space. The 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). Deposit by authors on their so-called home pages (typically unstructured space on the institutional web server) is also important.

The proportion of Green articles made available by these three main routes is difficult to estimate accurately. For example, (Björk, Laakso, Welling, & Paetau, 2014) reports three studies that estimated substantially different proportions, as follows:

- homepages etc.: 27-49% (of Green articles)

- institutional repositories: 19-44%

- subject repositories: 29–43%

One issue for differing estimates is the choice of different time windows for the studies. For example, it is estimated that over half of Green articles are not uploaded until at least a year after publication (a fact that also reduces the benefit of Green OA to readers) (Björk 2014).

Taylor & Francis's survey in 2014 reported that 23% of authors said they had deposited their last published paper in an institutional repository, 23% on personal or departmental website, 12% in a subject repository, and 52% did not deposit. Excluding those who did not deposit, this would equate to roughly 40% each for institutional repositories and personal/departmental websites, and 20% in subject repositories (Taylor & Francis 2014). This data is unlikely to accurately reflect the proportion of papers available, however, because many authors may be unaware of direct deposit of papers by publishers.

A good overview of Green open access is given in (Björk et al., 2014) for those seeking more detail.

The OpenDOAR website[64] divides repositories into the following categories (data as of December 2014):

- Institutional: 2257 (83%)

---

64 http://www.opendoar.org

- Disciplinary: 296 (11%)

- Aggregating: 98 (4%)

- Governmental: 77 (3%)

The Registry of Open Access Repositories (ROAR) reported a total of 3914 repositories, of which 723 were in the US and 249 in the UK (as of December 2014); 2603 were institutional, and 259 were "research cross-institutional" (similar to OpenDOAR's disciplinary category).

The broad categorisation into institutional and subject repositories potentially conceals wide variations in scope, function and cost. For example treating arXiv, RePEc and PubMed Central as equivalent is misleading. The arXiv contains mainly authors' accepted manuscripts; RePEc is essentially an indexing service over some 1600 repositories; while PMC is a highly centralised database. Indeed PMC 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).

Repositories contain a wide variety of content types, not just journal articles: theses, book chapters, working papers, conference papers, and others are also common (see Figure 26).

At present, although Green articles are spread across institutional and subject repositories and homepages, the largest collections, highest visibility and most use seem to be still with the subject repositories (e.g. Björk et al., 2014), (Björk et al., 2010); and see also The Web Ranking of Repositories, http://repositories.webometrics.info). Romary & Armbruster (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 infrastructure of institutional repositories has strengthened, however, and they are becoming one of the services expected from the university library, and most institutions now have a repository in use (Björk et al., 2014).

Figure 26: repository content by type (source: OpenDOAR)
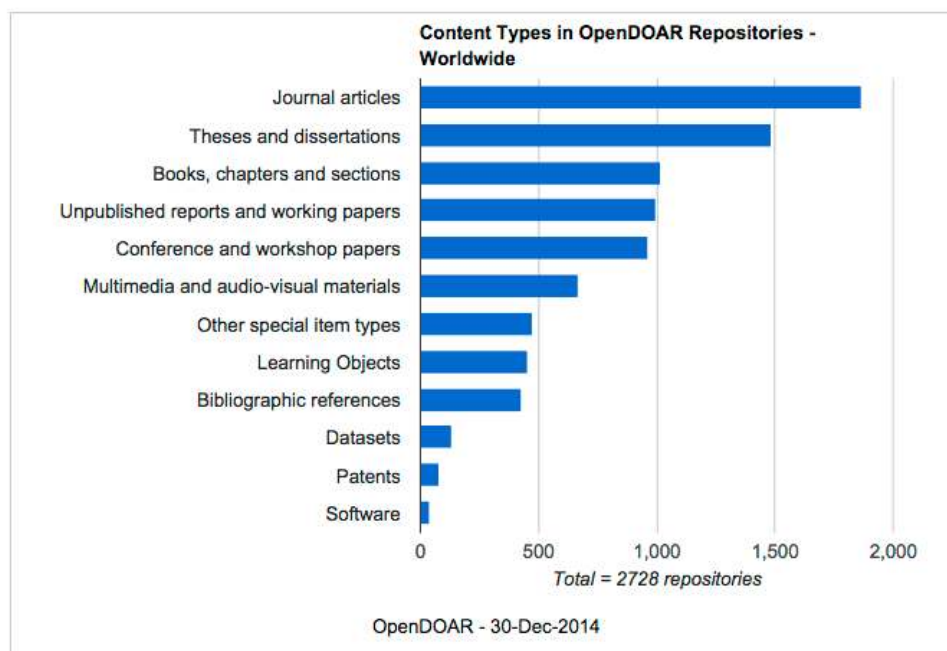
The numbers of repositories has increased substantially in recent years, with growth primarily from the institutional repository category, with OpenDOAR recording about 800 in 2006 rising to 2260 by 2014. Their combined records total over 14.5 million, growing annually at around 35%. Not all these records are full text, of course, but the proportion is rising.

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.

Perhaps unsurprisingly, new subject repositories are more rarely launched (though see *New preprint repositories* below). Pinfield has described the growth of repositories (Pinfield et al., 2014).

Institutional repositories

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

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

The two main objectives for having an institutional repository are:

- to provide open access to institutional research output;

- 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 powerful 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 (see above), although the complexity of services that they offer varies significantly.

The numbers of articles deposited by authors in their IRs has also grown, initially slowly but total deposits are now growing at about 35% annually. Many IRs (except perhaps in the Netherlands) do remain underused by depositing authors (e.g. see Björk et al., 2014; Salo 2008; Albanese 2009), but there are some clear exceptions (e.g. Queensland's University of Technology's repository has served more than 10 million downloads). (The total number of articles included in the 1885 repositories listed by Eprints in 2012 was about 9.8 million, or a mean of 5220, but these totals included 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 majority of authors remain 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 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). (See also: *Library publishing*; *Data-intensive science*)

Subject-based repositories

Central subject-based repositories have been around for much longer than institutional repositories. Björk reviewed the status of subject repositories in 2013. He concluded that they catered to a strong market demand when they first emerged, but the later development of Internet search engines, the rapid growth of institutional repositories and the tightening up of journal publisher OA policies seems to be slowing their growth. (Björk 2014). The leading subject repositories do appear to be in rude health, however, as the following examples show.

arXiv

One of the earliest was arXiv, established in 1991 at Los Alamos by Paul Ginsparg and now hosted by the Cornell library. arXiv[65] (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 1 million preprints. (See below, *Recent developments in open access*, for a discussion of arXiv's funding model.)

As the arXiv has grown (Figure 27) its host organisation (now Cornell, originally LANL) has struggled to justify the funding requirements. In August 2012 arXiv announced a new funding model[66] 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 that have voluntarily agreed to make contributions toward the costs. Cornell hoped 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). In 2014 the projected costs were $886k, which works out at less less than $10 per paper added (Van Noorden 2014a).

Figure 27: Growth in arXiv; physics and maths remain the most important subjects (Source: Nature News, (Van Noorden 2014a)

---

65 http://www.arxiv.org

66 See http://arxiv.org/help/support
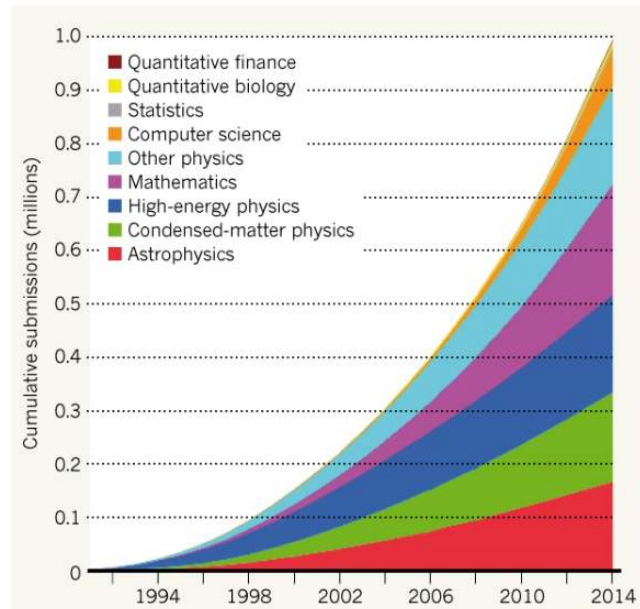
RePEc

RePEc (Research Papers in Economics)[67] 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 1.4 million research pieces from 1,800 journals and 3,800 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 1600 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.

PubMed Central

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 3.3 million research articles hosted on PMC, of which 945,000 were in the open access subset (the others are freely available but not open access in the sense used by PMC, released under Creative Commons licence permitting redistribution and reuse).

Europe PubMed Central[68] is based on PubMed Central with some additional services and functionality (McEntyre et al., 2011). It is part of the PMC International collaboration that also includes PMC Canada.

New preprint repositories

Two new preprint archives were launched in the life sciences in 2013, somewhat against conventional wisdom for the field which (unlike physics or economics) has no prior preprint culture (and for example, NPG had previously launched Nature Precedings in 2007, only to close it in 2012).

PeerJ Preprints is part of the PeerJ ecosystem, supporting the author's workflow from early draft, through peer review, to publication and beyond. Pricing follows PeerJ's freemium model: to deposit an author must be a PeerJ member, with a free membership allowing just one preprint per year, and more extensive usage requiring paid-for membership levels. After a slightly slow start in mid-2013, submission rates picked up in 2014 with 729 preprints deposited over the year.

---

67 http://repec.org

68 http://europepmc.org

BioRxiv launched later than PeerJ towards the end of 2013 but has already outpaced it in terms of numbers of submissions, with 918 preprints deposited in 2014. It was closely modelled on the physics arXiv, as its name suggests. It offers some additional features such as public commenting, supplementary information and links to external databases.

Other comparable services in life sciences include the journal *F1000Research*, which publishes papers immediately on submission, combining elements of preprint server with open peer review. The figshare data archive has a liberal deposit policy and a scan of its archive shows some users treating it as a place to share preprints.

Other types of repository

SSRN

SSRN, or Social Science Research Network,[69] can be classified as a subject repository but with enough differences to make it an interesting case study. It was founded in 1992; this early start reflects the long-standing preprint culture in economics (rather similar to that in physics). Organisationally, it is now a corporation with a budget in excess of $1 million, though it remains dependent on the work of volunteers (e.g. about 1000 act as Advisory Editors, Editors and Network Directors). It currently hosts abstracts of some 580,000 papers and fulltext versions of 480,000. It has delivered over 80 million fulltext PDF downloads over its lifetime, with current usage around 1 million downloads per month.

Its partnerships with a large fraction of the publishers in its field enable it to provide an indexing service in addition to its repository capabilities (somewhat similar to the respective roles of PubMed and PubMed Central).

SciELO

SciELO (Scientific Electronic Library Online) is not a conventional repository but a bibliographic database and a digital library of open access journals. The SciELO's model is used for cooperative electronic publishing in developing countries. Launched in 1997 originally in Brazil, it currently operates in 12 countries with three more "in development". As of mid-2014, it hosted 1161 open access journals containing nearly 0.5 million articles. SciELO announced in 2013 an agreement with Thomson Reuters for the integration of the SciELO Citation Index into Web of Science.

Redalyc

Redalyc[70] (Red de Revistas Científicas de América Latina y El Caribe, España y Portugal) is a bibliographic database and collection of open access journals, specialising in the scientific outputs and interests of Latin America. Launched in 2002, it now covers some 930 journals and 365,000 articles. Its services include bibliometric indicators, socio-scientific networks, journal collections, and usage metrics.

Other OA aggregators and indexing services

A number of services have been developed to improve discovery and use of specifically open access content. Table 11 gives an overview of some examples.

---

69 http://www.ssrn.com/

70 http://www.redalyc.org/

Table 11: Open access aggregators and indexes

Service	Launched	Description	Records	URL
BASE (Bielefeld Academic Search Engine)	2009	Search engine for academic open access web resources using OAI-PMH. Fulltext and metadata searching	> 65 million documents from more than 3,200 sources	http://www.base-search.net/about/en/
DOAJ	2002; 2013	Article-level search of a subset of the journals registered in DOAJ	1.8 million articles from 6032 journals (out of total of 10,135 journals)	http://doaj.org
OAIster	2002	Catalogue of open access records harvested using OAI-PMH, searchable via OCLC's WorldCat	> 30 million records from more than 1,500 contributors	http://www.oclc.org/oaister.en.html
OpenDOAR	2006	Allows search of the repositories registered in the OpenDOAR directory via a Google Custom Search	~10–15 million?	http://www.opendoar.org/search.php
Paperity	2014	Startup aiming to be "a multi-disciplinary aggregator of peer-reviewed Open Access journals and papers, both gold and hybrid". Uses proprietary web crawler/harvester based on redex	> 390,000 open articles from 2,200 journals	http://paperity.org/
ScienceOpen	2014	Multi-function platform combining preprint repository, open peer review, publishing and aggregation	> 1.4 million articles	https://www.scienceopen.com
Scilit	2013	Experimental OA search engine from open access publisher MDPI	~1.85 million articles	http://www.scilit.net/

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[71] listed (as of December 2014) 140 research funders, of which 109 had deposit policies (varying from requiring to just encouraging open access archiving),

---

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

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. (This likely underestimates the global number of policies, since SHERPA/Juliet has a UK emphasis.) 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[72] recorded 415 full institutional and 65 sub-institutional mandates in December 2014. Another directory of institutional and funder mandates, MELIBEA,[73] listed 349 institutional and 150 funder policies. 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[74]. Nonetheless it will take time for the policy to be effectively communicated: for instance, only 30% of respondents to the 2014 Taylor & Francis author survey said they understood the RCUK policy and many "appeared to be unsure whether the policy applies to them, since over half [55%] were unable to say whether or not their future articles would need to be published in accordance with the policy or not." (Taylor & Francis 2014)

Recent influential policies requiring author deposit have included those from RCUK (see below), the Bill and Melinda Gates Foundation[75] (notable for its strong requirements: immediate open access with no embargoes, CC-BY or equivalent licensing).

In the US, the OSTP memorandum requiring federal agencies to ensure public access to the outputs of publicly funded research seems likely to extend the impact of the NIH policy to other areas, though this is still working through.

Office of Science and Technology Policy memorandum

A key development in open access in the United States was the White House Office of Science and Technology Policy memorandum of February 2013 regarding public access to federally funded research (OSTP 2013). This specified that all agencies with research budgets greater than $100 million were required to make research outputs – specifically, "any results published in peer-reviewed scholarly publications that are based on research that directly arises from Federal funds" – freely available with a maximum delay of 12 months following publication.

---

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

73 http://www.accesoabierto.net/politicas/default.php. As well as listing policies, MELIBEA also calculates a metric, the "estimated open-access percentage"

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

75 http://www.gatesfoundation.org/How-We-Work/General-Information/Open-Access-Policy

The memo did not specify how this was to be achieved but required agencies to develop plans to meet these requirements. Four categories of solution emerged: agencies could build their own new dedicated repositories; the NIH's PubMed Central repository could have been extended or cloned as a paid-for inter-agency service; SHARE, a library-led system based linking higher-education infrastructure (including institutional repositories); and CHORUS, a publisher-led system leveraging existing publishing platforms and technologies.

CHORUS

CHORUS – Clearinghouse for the Open Research of the United States – was formed by a group of publishers and service providers, as a not-for-profit public-private partnership to develop a service that would enable funding agencies to meet the OSTP requirements.[76] (It is now managed by CHOR, Inc., a not-for-profit organisation.)

CHORUS acts as an information bridge, linking to freely accessible journal articles resulting from federally funded research directly on publisher platforms. CHORUS provides five core functions: identification, discovery, access, preservation, and compliance with policy requirements. It depends on *FundRef* (qv), a standard way to report funding sources for published scholarly research managed by CrossRef; when adopted by publishers it allows papers funded by federal agencies to be identified and made available via CHORUS. The system thus requires no new major infrastructure but provides an information and access layer on top of existing publisher platforms, and with development and hosting costs falling largely on publishers. In addition to this access layer, CHORUS also allows agencies to create discovery portals to their content, and offers "dashboards" to enable all stakeholders to monitor public-access compliance. It will provide open APIs to allow anyone to create overviews and discovery tools for federally-funded research. CHORUS proposes to support text and data mining through CrossRef's TDM tools and LicenseRef projects (see *Text and data mining*).

CHORUS also includes a preservation function via a partnership with Portico (see *Long term preservation*) so that long-term public access to federally-funded papers is not dependent on the continued existence of CHORUS or indeed the publishers.

Publishers considering participating in CHORUS will have to decide whether to make available the final version of record or the accepted manuscript (Cochran 2014).

At the time of writing one agency, the Department of Energy had selected CHORUS as its means of complying with the OSTP requirements. The largest agency, NIH is committed to its own platform, PubMed Central, while the remaining agencies' proposals were not yet public.

Open access critics of CHORUS point to the greater functionality of PubMed Central (including full text search and sophisticated interface and discovery tools), and see it as a way of preserving the value and primacy of the publisher's platform (e.g. Eisen 2013).

SHARE

An alternative vision is offered by SHARE[77] (SHared Access Research Ecosystem), a collaborative initiative of the Association of Research Libraries (ARL), the Association of American Universities (AAU), and the Association of Public and Land-grant Universities

---

76 http://www.chorusaccess.org/

77 http://www.arl.org/focus-areas/shared-access-research-ecosystem-share

(APLU), with initial funding from the Sloan Foundation. SHARE is not primarily focussed on meeting OSTP requirements but more generally addresses the need for preservation of, access to, and reuse of research outputs.

The service architecture is planned to consist of four layers: a notification service; registry; discovery; and mining and reuse. A prototype of the notification service is under development by COAS (Centre for Open Science): it will use both "push" protocols and a harvesting service drawing on a wide range of sources (including institutional and subject repositories, publishers, CrossRef, etc.) to collect information about "release events" (e.g. article publication or the dissemination of research data). These notifications will be distributed as a set of metadata to stakeholders such as funding agencies, research offices, and institutional and disciplinary repositories. Like CHORUS, SHARE will not store copies of research outputs but will maintain a registry of content that will subsequently support a discovery layer.

SHARE will not only provide access to journal articles but also intends to include research data within its remit (Lynch 2014).

RCUK Policy

The Research Councils UK policy[78] was developed during 2012 and came into effect in April 2013. It built on earlier policies dating from 2005 and was clearly influenced by the Finch Report. Unlike most other funder policies, it recognised the benefits of Gold over Green open access, all else being equal and funds available. The policy is summarised in the decision tree shown in Figure 28.

In order to cover the cost of article processing charges a block grant was made to universities and eligible research organisations. RCUK is in the process of reviewing the operation of its policy in the light of experience; the compliance rate reported by universities in the first year of operation (2013/14) vary between 35% (Hull) and 97–89% (LSE; Huddersfield).[79]

The costs incurred by UK institutions in 2013/14 for compliance with the RCUK and (slightly more stringent) HEFCE policies[80] was estimated by Research Consulting at £9.2 million, of which Gold APCs were less than 10% (see Figure 29 for the breakdown). It is clear from these high administrative costs that system-wide efficient systems for payment and tracking of APCs do not yet exist.[81] Jisc made a similar point in its evidence to the review of the RCUK OA policy: "The workflows and information flows imply new joins between stakeholders and systems, many of which are not used to / set up to join up efficiently. Legacy systems, lack of technical standards, tensions between different needs, etc. mean that

---

78 http://www.rcuk.ac.uk/research/openaccess/policy/

79 http://e2eoa.org/2014/10/01/open-access-reports-to-rcuk-collected-here/

80 HEFCE requires all materials to be deposited on acceptance in an open repository, in order to be eligible for submission to future REF assessments – http://www.hefce.ac.uk/whatwedo/rsrch/rinfrastruct/oa/policy/

81 Publishers' internal systems are becoming more efficient, e.g. through adoption of outsourced services such as CCC's RightsLink (www.rightslink.com/), but this does not necessarily affect inefficiencies at the payer end. Another intermediary is Open Access Key (OAK), whose systems are intended to improve efficiency at the institutional as well as publisher end (https://www.openaccesskey.com)

the data that would underpin a scalable APC environment are often absent, incomplete, wrong, ill-defined, and/or stuck somewhere" (Jisc 2014).

Figure 28: RCUK open access policy summarised in a decision tree

Figure 29: Compliance cost of the RCUK open access policy (CC-BY-SA Research Consulting)
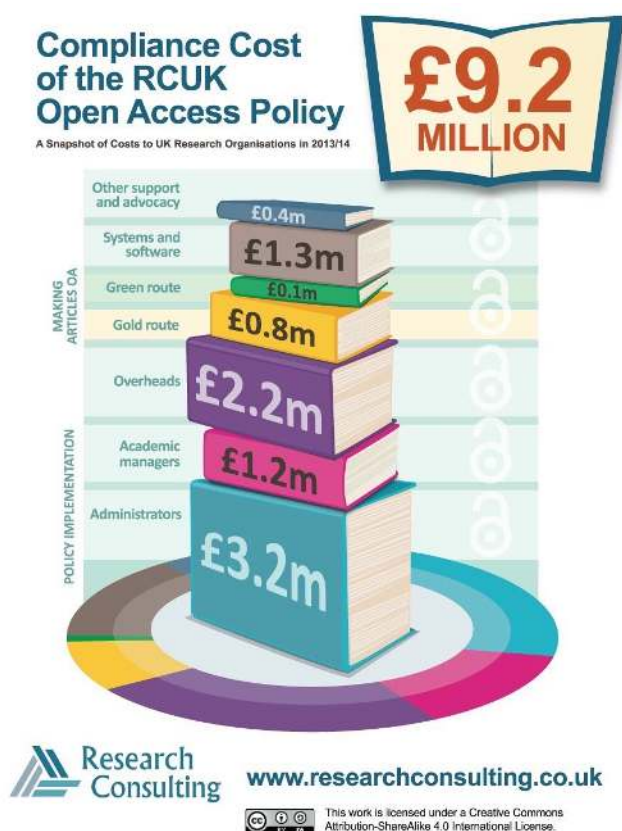
Publishers' policies on self-archiving

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. In response to more systematic deposit and discovery tools, policies are increasingly distinguishing between archiving on personal websites (with more liberal policies), institutional repositories, and subject repositories (with tighter requirements, reflecting the perceived greater threat to subscriptions). Policies will also vary to reflect specific funder's requirements; for instance very few publishers will not allow deposit to PubMed Central. (See also *Scientific social networks*.)

A database of publisher policies is maintained by the SHERPA/RoMEO project;[82] of the 1766 publishers included (as of December 2014):

- 34% allow archiving of both author's original and accepted manuscript
- 34% allow archiving of accepted manuscript
- 7% allow archiving of the author's original manuscript
- 25% do not formally support archiving.

Some 75% 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.

---

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

Some publishers also allow authors to archive the final publisher version, though this is rarer (around a third of publishers allow such posting to institutional or subject repositories, according to (Inger & Gardner, 2013), Table 7), with this much more common among small publishers, and large publishers significantly less likely to allow than in 2005). Some publishers add riders, such as requiring a link from the archived manuscript to the publisher's final online version. Publishers also commonly have 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). This gave an average total cost of about £20 per article.

A 2007 survey of US institutional repositories (Rieh, Markey, St Jean, Yakel, & Kim, 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[83] (2006–2013) used illustrative hardware costs of £2000–150,000 and suggested a £20k set-up should handle 50–100,000 papers. The site listed other areas of start-up and ongoing cost (primarily staff time) but gave no indication of the likely levels for these.

Large disciplinary repositories are naturally more expensive overall, but economies of scale may permit lower per-article costs. For example, Cornell University Library estimates the 2014 annual running costs for the (highly automated) arXiv at $886k, less than $10 per new article deposited. The National Institutes of Health has estimated that the cost of

---

83 http://www.rsp.ac.uk

administering its self-archiving policy would be $4 million.[84] At around 90,000 articles per year, this works out at about $44 per article. 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.

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 repositories (both institutional and subject) 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).

3.6. *Other open access variants*

Willinsky (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 – this is very similar to the "library partnership model" shown in Table 6).

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.[85]

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 asking for reported levels of use of, or authorship in, open access journals may suffer from this confusion.

3.7. *SCOAP3*

SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics)[86] is an ambitious project 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

---

84 http://publicaccess.nih.gov/Collins_reply_to_Pitts121611.pdf

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

86 http://scoap3.org/

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 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 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.

SCOAP3 identifying 12 journals from 7 publishers for participation in the first wave following a tender process concluding in September 2012,[87]. 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.

The SCOAP Repository[88] went live in in early 2014; as of the end of the year it hosted some 4300 articles.

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.

3.8. *Open access to 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.

OAPEN Library is an online library and publication platform for freely accessible academic books, mainly in the area of Humanities and Social Sciences. It evolved from the earlier Open Access Publishing in European Networks[89] project, 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[90] in April 2012; as of December 2014 there were 79 publishers and over 2482 OA books listed, with numbers of publishers and books both growing by over 40% annually. Many of the DOAB publishers are university presses but commercial publishers are also represented, including Bloomsbury

---

87 http://scoap3.org/news/news94.html

88 http://repo.scoap3.org

89 http://www.oapen.org

90 http://www.doabooks.org

Academic, Brill, De Gruyter, Palgrave Macmillan, Springer, Taylor & Francis, and Ubiquity Press.

OAPEN-UK[91] is a separate research project running 2010–2015. It is a Jisc/AHRC-funded collaborative research project gathering evidence to help stakeholders make informed decisions on the future of open access scholarly monograph publishing in the humanities and social sciences.

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 (e.g. Milloy 2013; Ferwerda 2014). The APC model is rare because of the higher costs of book publication and lack of available funding, but some examples do exist (e.g. Palgrave Open, Ubiquity Press). In the sciences the Gold model of author publication charges has been adopted by a few publishers[92] for multi-author monographs, where the individual chapter is equivalent to the journal article.

More recently two innovative models have been explored: Knowledge Unlatched's library partnership subsidy (see *Library partnership subsidy* in the section on open access business models above for details); and crowdfunding models whereby the publisher sets a target price at which point the title is released in electronic format (e.g. Open Book Publishers and De Gruyter, partnering with Unglue.it)

University of California Press unveiled a new open access books platform, Luminos, in early 2015. Its business model[93] combines APCs, library partnership subsidy and print sales. There is a base Title Publication Fee of $15,000: authors are required to meet $7500 of this (with waivers available), with the balance coming from voluntary library memberships ($1000 per library per year) and revenue from print sales, plus UCP subsidy.

3.9. Public access

The political advantages, if nothing more, of providing public access to (publicly-funded) research was recognised by the Finch report as well as being one of the key drivers of US government policy (see *Office of Science and Technology Policy memorandum*). In the UK, a pilot programme to provide walk-in access via public libraries was launched in 2013 as "Access to Research", supported by a couple of dozen leading publishers providing access to over 10 million articles via about 80% of UK library services.[94] Some publishers, notably the American Physical Society in the US, also provide similar public library access on an independent basis.

3.10. System-wide and economic 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.

---

91 http://oapen-uk.jiscebooks.org/

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

93 http://www.luminosoa.org/why-oa.php

94 http://www.accesstoresearch.org.uk

RIN/CEPA study

As noted above, 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 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.

The Houghton report

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.

Heading for the Open Road

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).

Other economic studies of open access

Other studies that have attempted to assess the economic impact of open access include the following:

- Houghton extended his original report (covering the UK) to Denmark and the Netherlands, and subsequently prepared a summary comparing the analyses for each country (Houghton 2009)

- Houghton also used the same methodology to model the impact of the then-proposed Federal Research Public Access Act (FRPAA) in the United States (Houghton 2010)

- A 2011 study for Jisc looked at the potential benefits to the private sector of open access to higher education and scholarly research (Parsons, Willis, & Holland, 2011)

- The potential for open access to increase (economic) innovation has been explored by Dagmara Weckowska, based on ESRC-funded interviews (Weckowska 2014)

Limitations of economic studies

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[95] 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.

A possible further issue is that rational debate about the merits of open access is too often drowned out by advocacy, defence of entrenched positions, and wishful thinking, as librarian Rick Anderson has described (Anderson 2014c).

3.11. Other developments in open access

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). An example is *PeerJ*, launched in late 2012: founded by ex-PLOS and Mendeley staff, and backed by Tim

---

95 Both studies used the Solow-Swan growth model

O'Reilly's venture fund (and a subsequent equity investment by SAGE), PeerJ proposes a model in which authors take out a membership entitling them to publish articles. Lifetime memberships are priced between $99 and $299, 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 price may end up in practice nearer the typical 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 (it has published a total of only 730 articles to date, and no other publishers have yet to adopt its model), 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*[96] 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. And if nothing else, eLife gives the lie to those who believe open access publishing costs next to nothing: its annual report showed its average cost per published article in 2013 (its first full year of publication) was $14,000, although this looks likely to have fallen sharply by 2014 as output doubled (eLife 2014).

Reuse rights

The clear trend towards funders placing greater emphasis on the licensing and reuse rights attached to open access articles continues, while at the same time the issue has become more contentious among researchers.

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*). The *How Open Is It?* guide illustrates this perspective, presenting a spectrum of openness under categories such as reader rights, reuse rights, machine readability, etc. (SPARC & PLOS, 2013).

In response to these mandates, a number of publishers switched from the CC-BY-NC to the CC-BY licence as their default 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); a few publishers have responded by charging more for CC-BY licensing than for CC-BY-NC.

On the other hand, surveys of authors as well as their behaviour when offered a choice (some publishers allow authors to select their preferred licence), shows that a substantial fraction of authors prefer CC-BY-NC to CC-BY licensing, and indeed may even prefer

---

96 http://www.elifesciences.org

traditional copyright assignment to the publisher over CC-BY (e.g. Taylor & Francis 2014). This is particularly the case in the humanities and some of the social sciences.

"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. Another issue that has received less coverage than predatory publishers is that of "highjacked" journals, where a website is fraudulently created to mimic a legitimate journal's site in order to attract submissions and APC fees (Jalalian & Mahboobi, 2014).

Surveys of researchers indicate that perceived quality remains a reason for a substantial minority for not choosing open access journals to submit to (NPG 2014; Frass, Cross, & Gardner, 2014). Legitimate open access publishers have responded by establishing the Open Access Scholarly Publishers Association[97] (OASPA) which requires members to adhere to a code of conduct and provides a complaints process. The Directory of Open Access Journals has also responded by cleaning its database of journals and publishers that did not meet criteria similar to those of OASPA, after discovering that at least 900 suspect journals were included (Anderson 2014b). There have been concerns, however, that some journals and organisations have been mislabelled as predatory as a result of applying simplistic or subjective criteria.

Open Access Button

The Open Access Button[98] is a browser-based bookmarklet that users are encouraged to use when they are unable to access content by reason of a paywall. It performs three functions: first, it collects information from users to build a picture of access gaps. Second, it gives users an opportunity to users to broadcast their experience (e.g. via Twitter). Lastly, it tries to suggest alternate versions of the same article either a version in an open repository, or if not found, alternate articles on similar topics that are openly accessible. It also contacts authors of flagged articles and suggests they make a version available, and if so, the original requesters are notified. The Button is therefore primarily an advocacy campaign for open access rather than a significant channel for accessing content. (See also *Scientific Social Networks*.)

3.12. Transition and sustainability issues

The actions of the scholarly community and the publishing market make it clear that the open access debate has now moved on to *what* is necessary to make it open access sustainable (rather than *whether* this is possible), and to the problems of *how* a transition should 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 (about 10% of articles, but only ~2.5% of revenues) and there are valid questions about how a scaling-up would be achieved.

The key issues are:

---

97 http://oaspa.org

98 https://www.openaccessbutton.org

- 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 as open access scales up, and what impacts will these have on scholars and institutions, as well as on publishers?

- 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 based on article publication charges

There is general agreement that under appropriate circumstances APC-based 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; 2014), and in behaviour of existing commercial publishers in launching open access journals, and is also recognised in the statement signed by STM members supporting sustainable open access.[99]

Ignoring for the moment differences between disciplines and other complications, for the Gold 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, studies have variously estimated the average APCs charged (excluding hybrid) at between $906 and $2097, depending on the sample used (Björk & Solomon, 2012c; 2014; Morrison & et al, 2014). Table 7 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 (strongly weighted to hybrid APCs). 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 (Björk & Solomon, 2014)), 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

---

99 http://www.stm-assoc.org/publishers-support-sustainable-open-access/

have been about half its previous size (hypothetical because it assumed a wholesale conversion to OA).[100]

It is also clear that open access publishing can be profitable at an APC of $1350 or lower, given sufficient scale and a low-cost approach, as demonstrated by PLOS's results (which show average costs of $1088 per article, across all PLOS journals). Journals with higher editorial costs than *PLOS ONE*, whether due to high rejection rates or a greater level of editorial services, would need to charge higher APCs: the *Proceedings of the National Academy of Science* estimated it would need an APC of $3700 (Van Noorden 2013); EMBO has said it needs to charge "much more than $2000" to cover costs on its journals (Pulverer 2014); *Nature*'s internal costs have been estimated by its publisher at well over $10,000 (Jha 2012); and *eLife*'s average cost per article was $14,000 in 2013 (its first full year), though this seems likely to fall sharply in 2014 as output doubled.

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 additional, non-research article content, nor the increasing demands for novel tools to become standard.

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; Archambault et al., 2014). 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.

---

100 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.

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.

As it seems unlikely that these kinds of transition funding will be forthcoming from governments and research funders at a global level, funders and publishers are exploring market-based solutions. One promising option is the offsetting and other bundling approaches (see *Offsetting*).

Another transition issue is that the economic case for Gold rests on non-cash savings (e.g. researchers' time) and uncertain economic multiplier effects. In an austerity 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 out 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's annual reports show waivers running at 7–8%. 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).

The sustainability of non-APC Gold models is harder to model. This is not a trivial or purely low-quality part of the scholarly communication landscape: one credible estimate based on analysis of the Scopus database put the prevalence of Gold open access articles without APCs (4.2%) at comparable levels to those with APCs (5.5%) (Elsevier 2013), and nearly all the larger journal publishers have examples in their portfolios.

Whether the existing approaches (largely based on subsidy or sponsorship) can be scaled is unclear, while the emerging models (e.g. library partnership subsidy or crowdsourcing) are unproven or have unresolved weaknesses (such as free-riding).

APC-based Gold OA does not appear to be a good model for the very prestigious, top-tier journals like *Nature* or *Science* that depend on expensive editorial quality control, because of the very high APCs that would have to be charged. The launch of *eLife* is intended to

challenge this assumption but, although it has successfully attracted high-quality articles, this remains inconclusive until it declares its longterm business model.

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 approach being adopted by EMBO and some other higher rejection rate journals is to establish manuscript transfer arrangements (see *Cascade peer review*) with other journals that can rebate some of the APC back. Another possibility for high rejection rate journals may be to introduce submission fees (Pulverer 2014), though it also comes with significant disadvantages: publishers are loathe to risk deterring submissions in a competitive market, and funders do not favour it (because it primarily funds non-publication).

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 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 global uptake has remained very low (around 1–2%) but there are exceptions with some journals and publishers reporting higher rates (see *Hybrid journals*).

Funder and institutional concern about double-dipping remain. New business models that bundle subscription and APC costs or allow offsetting may allow hybrid to play a more important role in the future (see *Offsetting*).

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. Laakso and colleagues described delayed open access as an overlooked high-impact category of openly available scientific literature, providing access to up to 5% of (non-current-year) journal articles (Laakso & Björk, 2013).

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, except that embargoes here are fully under the control of the publisher, and are dealt with in the following section.

3.13. Effect of self-archiving on journals

Publishers continue to have concerns 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.

PEER project

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.[101] 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 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.

---

101 http://www.peerproject.eu/

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 30 (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*).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 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 each instance 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 30: Article downloads from publisher sites by age and subject area (CIBER Research 2012a)
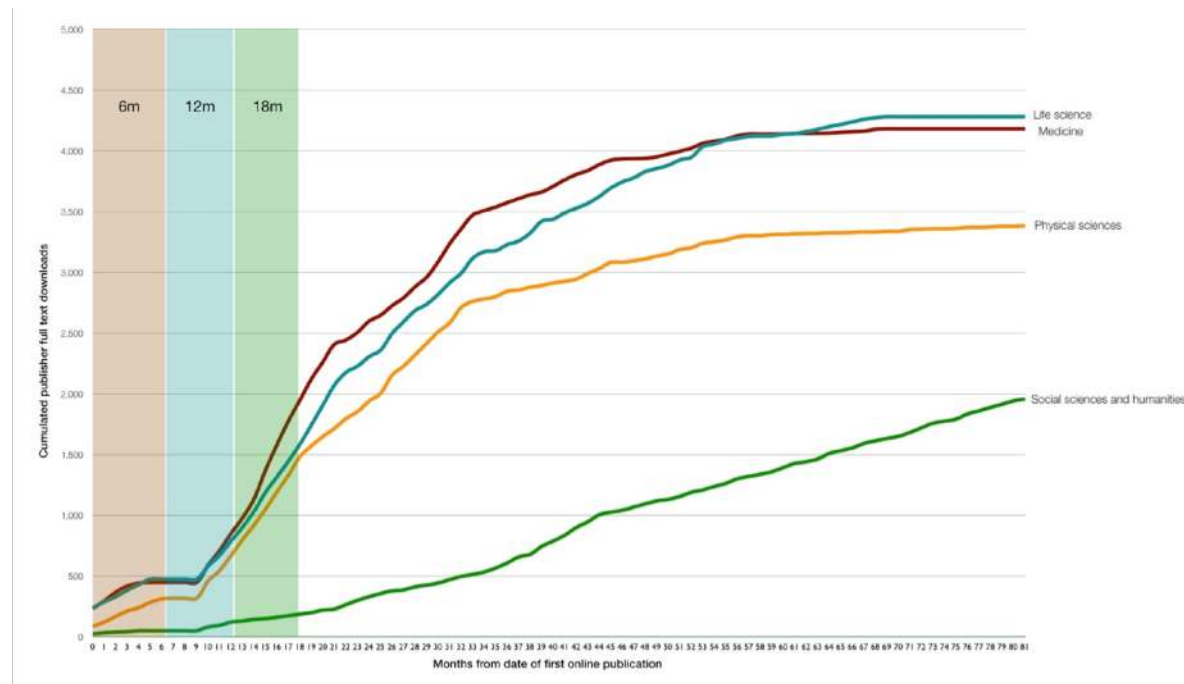
3.14. Open access impacts on use

Impact on usage (downloads)

There is now a substantial body of studies of the impact of open (or free) access on usage behaviour, including downloads and citations, going back at least 15 years – indeed this is probably one of the largest areas of the open access literature.[102] (A similar question with respect to the impact of making research data open is now also being explored – see *Data citation*.)

While many of these are flawed through the lack of proper methodology to control for other factors affecting usage (including early view effects and selection bias, see below), 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).

---

102 For example, the OpCit Project's online bibliography is a large source of papers in this area (OpCit n.d.)

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). The numbers and locations of coauthors are also known to affect citation rates but are also frequently not controlled.

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, Plume, McVeigh, Pringle, & Amin, 2007). Citation patterns differ between subject disciplines, however, so this still leaves it open that there may be an effect in other fields.

Davis (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 (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". And 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".

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 publication 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.

A more recent study on economics and business journals was clear: "the enormous effects found in previous studies were an artefact of their failure to control for article quality, disappearing once we add fixed effects as controls" (McCabe & Snyder, 2014). The authors did find, however, that the lack of an overall aggregate effect masked differences between online sources: in this field, availability on JSTOR boosted citations by 10%.

Conversely another study by the same authors (McCabe & Snyder, 2013) found that "moving from paid to open access increases cites by 8% on average in our sample. The benefit is concentrated among top-ranked journals. In fact, open access causes a statistically significant reduction in cites to the bottom-ranked journals in our sample, leading us to conjecture that open access may intensify competition among articles for readers' attention, generating losers as well as winners."

Nature Communications was a hybrid journal (now converted to full open access) with a very high rate of uptake (40%+) of the open option. This allowed a comparison of the usage and citation of OA and non-OA in the same journal. The NPG-commissioned report by Research Information Network (RIN) found a large and statistically significant increase in downloads (confirming many earlier studies). It reported that there "appeared to be" a small increase in citations while conceding that the study had not controlled for selection bias, parallel Green availability, or number and location of authors (RIN 2014).

The effect is therefore still unclear, but best available evidence at this point tends to suggest that open access articles in the aggregate probably do not receive more lifetime citations, but they do get them sooner due to early view and selection bias effects. There may also be particular circumstances in which open access does increase citations: for example, availability on some platforms, for some fields, or for high-ranked journals.

Whether or not the citation effect exists and how large it might be should increasingly be a matter of academic interest, as the proportion of literature that is open access steadily increases. The effect is, however, one of the widely claimed benefits of open access and appears to influence authors (e.g. 29% of authors in the 2014 Taylor & Francis survey said that OA journals were more highly cited than subscription journals, up from 25% in 2013 (Taylor & Francis 2014)).

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. We believe that there is a reasonable consensus that the trends discussed here are important to scholarly publishing, although there is certainly scope for debate as to their relative importance.

For a more regularly updated view, we recommend following the reports of the STM Future Lab Committee[103]. The most recent report of this committee (STM Future Lab Committee 2014) identified three key themes for 2014–16 to be "The machine is the new reader" (enriched content, text and data mining, smart articles, etc.); "The return to the author" (better author/researcher experience, augmented peer review, new authoring tools, production automation, etc.); and "New players changing the game" (social media, new start-ups, self-publishing, open authoring tools, etc.).

It's also worth noting that the key trends identified by the STM Future Lab Committee in preceding years have not becoming obsolete overnight, but build towards and feed into the current scenarios:

- in 2013: the key themes were "Hybrid reader experience" (the future of the PDF, etc.), "Open access implementation" (open courseware, open ID, open reference data, etc.), and "From validation to augmentation" (curation and analytics, altmetrics, new metrics)

- in 2012: "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). Three key strands – API platforms, research data, and identity management – were central to making content more "actionable".

A 2012 review of the development 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;[104] ecommerce and monetisation; convergence and integration; functional (active) content; and analytics. While some of the detail of this report is inevitably dated, many of the themes remain important.

Outsell's evaluation of the STM platform provider market (Outsell 2014d) saw some more recent trends – need for better discoverability and access to data; a focus on the researcher rather than the research "container"; – information rather than format – emerging alongside the continuing drivers – user behaviour and workflow needs; business model customisation; and a need for standards.

---

103 http://www.stm-assoc.org/standards-technology/committee/

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

These various topics and themes are explored in the sections following.

4.1.   "Science 2.0" or "Open Science"

Although the term "science 2.0" has been in use for at least a decade as a loose way to refer to new digital (often web-based) approaches to research (e.g. Waldrop 2008), it has more recently been used in a more formal way by European policy-makers. The European Commission conducted a public consultation entitled *"Science 2.0": Science in Transition* (European Commission 2014) which posited Science 2.0 as a systemic change in scientific research and its organisation, enabled by digital technologies and driven by the globalisation of the scientific community and by the need to address the Grand Challenges.[105]

This consultation paper saw Science 2.0 as encompassing three broad trends: a significant increase in research outputs; data-intensive science (see below); and an increase both in the number of scientists and in the audience for science (including both professional and lay groups ("citizen science") in both cases).

The policy interest in Science 2.0 lies partly in the mooted implications for society and the economy, including science becoming more responsive to societal needs; greater openness that could improve trust in science; economic benefits accruing from increased rates of innovation (e.g. due to SMEs getting open access to research data); new services and productivity gains arising from use of "big data" techniques and text and data mining; etc. (Although the evidence for most of these is, as the paper concedes, currently thin.)

From a publishing perspective, open access is seen by the EC paper as an important part of Science 2.0, together with the need for other changes including speeding up publication and reforming the existing peer review system. Related areas where it sees the need for change are in the challenge of reproducing research results due to the lack of available data; and the need for reform of the ways research is assessed and incentivised, for instance rewarding open data approaches on the same scale as journal articles.

Most recently Science 2.0 has been rebranded as Open Science as the claims of Science 2.0 were increasingly seen as too broad and not supported by the research community. Open science also has the advantage of not appearing to look back to the now dated Web 2.0 era.

4.2.   FORCE11 and "Science in Transition"

A number of initiatives have emerged primarily from the research community (i.e. rather than being publishing-led), with examples including:

- FORCE 11 is a network describing itself as a community working together in support of the goal of advancing scholarly communication. Apart from organising a well-regarded biennial conference[106] (originally called *Beyond the PDF*), its 2011 manifesto offers a realistic analysis of the new opportunities and challenges. It focuses its attention on the need for tools for researchers as producers of STM output, enhanced products for researchers as consumers, and tools and services for reputation management (FORCE 11 2011). FORCE 11 has been closely involved in a number of the specific issues discussed below, e.g. see *Data citation*.

---

105 http://grandchallenges.org

106 e.g. https://www.force11.org/meetings/force2015

- The "Science in Transition" project shares many of the same themes as the the EC "Science 2.0" consultation. It has been influential in the Netherlands, though its impact elsewhere has been less important[107]

4.3. 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? In practice, publishers support both behaviours, 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 is the open API (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.

4.4. Social media

Social media and networks (sometimes referred to as Web 2.0) offer the potential to enhance informal and formal scholarly communication. Their impact is growing, though it remains limited compared to conventional channels of scholarly communication.

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 social media 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 lacked 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.

---

107 http://www.scienceintransition.nl/english

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 a crop of Web tools for biologists and why they were 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; Crotty 2010).

In addition to this relatively low *active* use of social media (i.e. posting content), researchers also make little *passive* use of social media as a a source of information and awareness. For example, Ithaka found that very few respondents saw blogs and social media as important to their research, and specifically, following other researchers through their blogs or social media was by far the least important way for researchers to keep up with their fields (Ithaka S+R 2013). Interestingly, a significantly higher percentage of scientists (albeit still low, at around 15%) said they shared the findings of their research via social media than used social media to keep up with new research (around 5%); perhaps they saw social media as a way to extend impact beyond the academy rather than a way to communicate with peers.

Web 2.0 ideas could be used supplement peer review, for instance by allowing readers to add comments and ratings to the article after publication. Where tried (e.g. rating and commenting for journal articles), uptake has been very low to date, and there are serious questions as to what is measured through such techniques (Harley & et al, 2010). Nonetheless new initiatives continue to emerge in this area and it may yet have a role to play (see *Post-publication peer review*).

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. Scientific social networks have grown very rapidly (see below). 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 the least play a part in content discovery and sharing.

Wikis

Wikipedia is not just the best known general-purpose user-generated encyclopaedia but also (despite continuing scepticism in some quarters about the quality of its content) one that 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, therefore, 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.

A related service, Wikidata acts as the central storage for the structured data for Wikimedia sister projects including Wikipedia. A proposal to establish Wikidata as the central hub for linked open research data provisionally titled "Wikidata for research" is being developed, coordinated by the Museum für Naturkunde Berlin, in collaboration with Wikimedia Germany[108]. It remains to be seen, however, whether this will be successful.

Academic publishers have been slow to adopt wikis as a publishing platform, most likely because the wiki model relies on open, editable and reusable content which is not easy to monetise.[109] 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.

Scientific social networks

Scientific social networks have grown significantly since the last edition of this report. The three main networks, which all launched around 2008, are Academia.edu[110] (which has reported over 16 million registered users), Mendeley (around 3.5 million users, and acquired by Elsevier in April 2013), and ResearchGate (over 5 million users). A fourth network, Colwiz, launched in 2011 and currently has about 260,000 users. Awareness of the networks also appears to be correspondingly high, especially for ResearchGate in STM fields and Academia in social sciences and humanities (Van Noorden 2014c).

The numbers of documents by users are also substantial: Mendeley reported that its users had uploaded over 470 million documents; ResearchGate reported that 14 million documents were "accessible" via its platform.

Researchers use these platforms for a variety of purposes, with the most popular motivation being to maintain an online profile to make themselves and their work more discoverable. Other popular uses include posting content, finding related researchers, tracking metrics, and discovering new and recommended research papers (Van Noorden 2014c). At present, one activity that appears relatively unimportant is direct interaction and discussion; these seem more to be tools for researchers to raise their profiles and become more discoverable, and to access workflow tools and services, rather than community tools of social interaction.

The potential use of these networks for sharing of journal articles in breach of copyright has caused concern for publishers with "take down" notices being issued in some cases (Economist 2014). At the time of writing industry initiatives to address the issue included development of a set of principles to be voluntarily used by stakeholders as a guide for sharing articles via social networks, and possible licensing arrangements.

Bibliography management software (such as Endnote (Thomson Reuters), Flow (Proquest), Pages (Springer), Zotero, etc.) also allows users to share their research libraries with other

---

108 http://blog.wikimedia.de/2014/12/05/wikidata-for-research-a-grant-proposal-that-anyone-can-edit/

109 Two initiative mentioned in the last edition of this report (Elsevier's WiserWiki and SciTopics) have now been discontinued. A newly launched example is IWA Publishing's WaterWiki, which combines formally published reference information with community generated content http://www.iwawaterwiki.org/xwiki/bin/view/Main/WebHome

110 A commercial start-up, backed by venture capital, despite the "edu" domain

users but typically the sharing is inherently one-to-one or one-to-few, or restrictions on the numbers of users with whom content may be shared are explicitly enforced.

4.5. 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 computing devices in the general population 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 tablet sales exceeded PC sales in 2013. The numbers of mobile devices are starting to dwarf PCs – in 2013, there were over 2 billion smartphones and tablets in use, compared to 1.5 billion PCs (Meeker 2014; Blodget & Cocotas, 2012). Unsurprisingly, global mobile web traffic is growing rapidly, standing at 25% in May 2014 compared to 14% a year earlier.

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 (Wolters Kluwer Health 2013). A similar level of use is found in US nurses, with 65% of nurses using mobile devices for professional purposes at work at least 30 minutes per day, while 20% use them for two hours or more. Wolters Kluwer's *Nursing Drug Handbook* apps were reported in 2014 to have been downloaded 450,000 times (Wolters Kluwer Health 2014).

The cost/benefit equation is clearer for busy professionals than for most academic researchers, but mobile device use is rising in this group too, with growth mostly coming from increased tablet uptake. Mobile traffic at the leading STM platforms was still only around 10% in 2014, albeit growing rapidly year-on-year. (Inger & Gardner, 2012)'s survey showed an overwhelming preference for accessing online articles on a desktop or laptop PC over tablet or phone; mobile device use was higher in medical compared to academic sectors, but still very much a minority activity.

Use cases for mobile are still emerging and developing (Outsell 2012b). 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 growth of tablet uptake – e.g. point-of-care drug information is ideally delivered through a device that is always in the 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 may have 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, most larger STM journals and platforms offered a mobile-optimised interface (e.g. using responsive or adaptive design[111]). 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 still 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.

4.6.   Research data

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 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.

---

111 http://en.wikipedia.org/wiki/Responsive_web_design

Impact on journal publishing

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, OpenAIREplus) in addition to the initiatives listed below (see *Data repositories*).

- 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, a membership model providing services to users (e.g. as at PANGAEA). See also *Data journals* below.

It is widely accepted that research data should be openly available to other researchers (subject to specific constraints such as protecting patient confidentiality). STM publishers have included their recognition of this principle in the 2007 STM Brussels Declaration[112] and the subsequent STM/DataCite statement[113] in 2012.

Research Data Alliance

The Research Data Alliance, created in 2013, has been an important development (Treloar 2014). The RDA seeks to improve data interoperability across boundaries (e.g. national, disciplinary, producer/consumer) working at the infrastructure level (where infrastructure includes not just hardware but also software, content and format standards, and human actors). It operates through a series of Working Groups, each addressing tasks with a relatively short (18 month) timeframe to produce specific deliverables such as policies, technical specifications, etc. RDA Interest Groups have broader perspectives and longer (indefinite) timeframes, and can spawn new Working Groups, coordinate activities of multiple Working Groups, and so on. The RDA/WDS Publishing Data Interest Group[114] is one of the most important for STM publishers, with related Working Groups including those on Data Citation, Publishing Data Bibliometrics, Publishing Data Services, Publishing Data Workflows, and the BioSharing Registry.

Data repositories

The number of data repositories has grown substantially: there are two main directories, Databib.org, which lists 995; and re3data.org, which lists over 1000 and reports an average of 10 repositories added to its register per week.

---

112 http://www.stm-assoc.org/public-affairs/resources/brussels-declaration/

113 https://www.datacite.org/joint-statement-stm-and-datacite

114 https://rd-alliance.org/groups/rdawds-publishing-data-ig.html

Data repositories have been developed to host "orphan data", that is, datasets for which there (currently) exists no recognised disciplinary repository. Examples include DRYAD[115], which specifically hosts the data underlying peer-reviewed articles in the basic and applied biosciences; Zenodo[116], a European project linked to OpenAIRE; and figshare, a commercial service from Digital Science.

In addition to the central orphan data repositories, it is increasingly the norm for institutional repositories (often managed by the university library) to offer data deposit services.

Data citation

Since the last edition of this report there has been substantial progress on data citation and linking (a process that has been a good example of multiple stakeholder collaboration[117]), leading to the publication of the Joint Declaration of Data Citation Principles[118]. (Callaghan 2014) gives a good account of these developments; see also (Murphy 2014).) This Declaration sets out eight principles for the purpose, function and attributes of citations: importance; credit and attribution; evidence; unique identification; access; persistence; specificity and verifiability; and interoperability and flexibility.

The advantages of citing data are to make data more discoverable; to increase research transparency, encourage researchers to share data and thus help with reproducibility; to allow creation of data citation metrics or otherwise give credit.

Citing datasets in journal articles is relatively straightforward, since it simply extends existing citation practices. It would also be valuable for the reverse citation to be created at the same time, i.e. for datasets to be linked back to the articles that cited them. At present this is achieved manually, with journals simply notifying data repositories by email but this is clearly not scalable even if some of this workflow can be automated. To address this issue, a registry has been proposed to act as an intermediary between journals and data repositories.

Some additional challenges include the following:

- dynamic data: some datasets are not static but are updated on a continual basis. While desirable that data should be up-to-date, this creates a problem for the fixity of the scientific record. The RDA Data Citation Working Group is investigating possible technical solutions, and DataCite has issued some guidance

- micro-attribution and micro-citation: this refers to the situation where a contributor is responsible for a small part of a much larger database.

It does appear that researchers are increasingly citing datasets. A 2014 study cross-comparing Scopus reference lists with data repositories listed in Databib found that such citations have grown by 19% annually (CAGR) between 1996 and 2013, reaching 30,000 papers in 2013 (Huggett 2014).

---

115 http://datadryad.org

116 http://zenodo.org

117 e.g. contributors included the CODATA Task Group on Data Citation, the FORCE11 data citation synthesis group, DataCite, the Digital Curation Centre, and others

118 https://www.force11.org/datacitation

Data sharing and journal policies

As with open access to journal articles, research funders are playing an important part in mandating the open sharing of research data. 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[119]. Other major mandates include the OSTP memorandum in the US and the Horizon 2020 funding in the EU. The Digital Curation Centre's website provides an overview of the main UK research funders' data policies.[120]

While the benefits of making research data available are widely accepted, in practice there are a number of obstacles. Researchers are not always strongly motivated to share their own data. A 2014 survey found that about half had made research data available (typically as journal supplementary materials), but reported a long list of reasons why they were reluctant to share. The most frequently given were intellectual property or confidentiality (42%), funder / institution did not require sharing (36%), concerns about the research being scooped or misunderstood / misinterpreted (26%), and a further eight reasons (Ferguson 2014).

On the other hand, important reasons for having shared data included that it was standard practice in the researcher's community, to increase impact and visibility of research, and because the journal required it. As the third point suggests, the adoption by journals of policies either encouraging or requiring the sharing of research data associated with articles has been growing and is increasingly becoming the norm (Sturges et al., 2014). Publishers can help shape community norms, as well as reflect them in their policies, but strong data policies may need careful introduction and communication, as illustrated by some of the reaction to PLOS's data policy introduction in early 2014 (Bloom 2014). One approach may be to offer incentives as well as requirements: for example, PLOS, CDL and DataONE are collaborating on development of data-level metrics (analogous to article-level metrics) to provide feedback on data usage, views, and impact (Lin 2014).

Data journals

The growing interest in improving discovery and reuse of research data, and in providing opportunities for researchers and data producers to publish and gain acknowledgement for their research data outputs, has led to a growing number of data journals to be launched.

While conventional journals may link to (or embed) research data within the familiar rhetorical structure of the scientific article, data journals offer a platform (normally open access) for publication of "data articles" or "dataset papers" that are typically short articles providing a technical description of a dataset. Some data journals also publish (i.e. host) the dataset themselves, but others link to datasets hosted on dedicated data repositories. Where data journals link to external datasets there are often minimum requirements for the third-party hosting (e.g. *Geoscience Data Journal* specifies the data centre must be able to mint a DOI).

Table 12 list some example data journals. It is unclear whether data journals will remain as distinct entities, as their functions might instead be met by existing journals (by expanding

---

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

120 http://www.dcc.ac.uk/resources/policy-and-legal/overview-funders-data-policies

their articles types to include data papers) or through other ways layering description, extended metadata and attribution/reward mechanisms over data repositories.

Table 12: Data journals (examples)

Journal	*Publisher*	*URL*
Biodiversity Data Journal	Pensoft	http://biodiversitydatajournal.com
Dataset Papers in […] series (11 topics)	Hindawi	http://www.datasets.com/
Earth System Science Data	Copernicus	http://earth-system-science-data.net/
Ecological Archives – Data Papers	Ecological Society of America	http://esapubs.org/archive/instruct_d.htm
F1000 Research	Science Navigation Group	http://f1000research.com
Genomics Data – Data in Brief papers	Elsevier	http://www.journals.elsevier.com/genomics-data/
Geoscience Data Journal	Wiley	http://www.geosciencedata.com
GigaScience	BGI / Biomed Central	http://www.gigasciencejournal.com
International Journal of Robotics Research	SAGE	http://ijr.sagepub.com/
Journal of Open […] Data (4 journals)	Ubiquity Press	e.g. http://openarchaeologydata.metajnl.com/

Other developments and initiatives

Other developments and initiatives in the field of research data include:

- DataCite[121] (launched in December 2009) addresses 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

- new discovery tools have been created: Thomson Reuters' Data Citation Index[122] (launched in October 2012), supports data discovery, reuse and interpretation; OpenAIREplus harvests and indexes the metadata from open access scientific datasets across multiple repositories; the DataCite Metadata Search tool allows discovery of datasets registered with DataCite

- ODE (Opportunities for Data Exchange, an Alliance for Permanent Access project)[123] has produced a set of recommendations for journals, including the introduction of stricter

---

121 http://datacite.org

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

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

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.

Some other notable initiatives include: BioSharing, which works to build stable links between journals, funders with data-sharing policies, and standardisation efforts in biosciences; BioDBCore, a community-defined, uniform, generic description of the core attributes of biological databases; and ISA Commons, which produces an open-source framework for data sharing centred around the general-purpose ISA-Tab file format.

4.7. Semantic web and semantic enrichment

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.

 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);

- 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 13. 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*).

Table 13: 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

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 reuse (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).

Carol Goble and others have argued that for scientific purposes linked data is insufficient. They proposed further layers sitting over linked data publishing, to support provenance, quality, credit, attribution and methods to provide the reproducibility that enables validation of results, thus making the dataset a "first class research object" (Bechhofer et al., 2013).

4.8.   New article formats and features

Publishers and others continue to innovate and investigate potential new ways to explore, present, format and share research articles and related content on the web. Some recent developments include the following:

- Enhanced HTML-based formats: example include Elsevier's Article of the Future, Wiley's Anywhere Article, and eLife's Lens formats, as well as similar initiatives from other publishers. These have been based on research into how researchers use online articles, and primarily aim at improving and streamlining the user experience, for instance dividing the screen into regions so that the text can be viewed alongside images or references. Another advantage of redesigning the online layout using HTML5 is that it can natively support mobile-friendly views

- Enhanced PDFs: recognising that researchers will often prefer to use the PDF (particularly for local storage, annotation, etc.), new more feature-rich and web-connected versions of the PDF format have been developed, of which the best known are ReadCube and Utopia Docs. Publishers in fields heavily reliant on 3D information – earth sciences, geophysical, geospatial, engineering, medical scanning, etc. – are starting to adopt the 3D PDF format

- Article versioning: platform developments that allow articles to be updated or expanded, while rigorously preserving the original version(s) and its publication record. PLOS and F1000Research have actively explored this area, though many other publisher platforms also support it

- eLife's Research Advance article type performs a related function by allowing researchers to publish significant "additions" to original research papers, so that they can report (substantial) progress in their research programmes rapidly and efficiently without need to write a full new paper

- Dynamic ("live") figures: rather than publishing figures as flat images, which makes reuse of the underlying data either difficult or impossible, figures could be presented as dynamically generated images from data stored with the article. F1000Research has taken this idea a stage further by allowing the user to interact with the code that

generates the figure, so that, for example, parameters could be varied and the different results explored.[124]

- Data visualisation: there is a very large number of file formats used to store experimental research data. The usefulness of including such datasets in the article supplementary data can be much enhanced by providing visualisation tools. Some publishers (e.g. Taylor & Francis, NPG, PLOS) use a service offered by Figshare for storage and visualisation of such dataset.

- Article viewing and sharing: ReadCube's Content Sharing Initiative allows users to share subscribed content with non-subscribers via a special link. At present the service has only been adopted by ReadCube's sister company, Nature Publishing Group

- Microarticles: this is Elsevier's name for a new short article format designed to let authors publish useful data, method descriptions or other valuable research results (including intermediate and null/negative results), that might otherwise remain unpublished

- Geotagging: much research in a wide variety of fields from archaeology and epidemiology to environmental and earth sciences includes location-specific information. Until recently the only way to locate research relevant to a particular location or region was to use keyword searching, which is imprecise and haphazard. Search based on geotagging allows precise searching, map-based interfaces (as in Google Maps) and other advantages. Examples include JournalMap, a scientific literature search engine that finds research based on location and biophysical attributes combined with traditional keyword searches; and Elsevier's Geofacets, which provides peer-reviewed maps including context from their source publications aimed at geoscientists.

4.9. 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 *Text Mining and Scholarly Publishing* (Clark 2013) gives a good introduction to TDM (see also Clark, Jensen, & Campbell, 2014; Smit & van der Graaf, 2011; and 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.

The main challenges for more widespread adoption are legal uncertainties as to what is permitted, and the lack of an efficient licensing regime (see *Text and data mining rights*); technical issues such as standard content formats including basic common ontologies; the need for content aggregation to permit mining cross-publisher corpuses; the costs and technical skills requirements for mining; and a lack of understanding on the part of publishers. This last point was illustrated in an ALPSP report: "a large number of the publishers surveyed have little or no understanding of text mining, and many suggest in

---

124 An example article using this format is at http://f1000research.com/articles/3-176/v1

their comments that they have never been approached by a client about text mining" (Inger & Gardner, 2013).

These challenges are being addressed in a number of initiatives:

- STM publishers issued a statement in November 2013 committing its signatories to implementing the STM sample licence clause, or otherwise to permit non-commercial TDM of subscribed-to content at no additional cost; to develop the mine-ability of content; and to develop platforms to allow integration of holdings across institutions for TDM purposes (STM 2013b)

- CrossRef's text and data mining tools (originally Prospect):[125] this offers a metadata API and services that can provide automated linking for TDM tools to the publisher full text, plus a mechanism for storing licence information in the metadata, and optionally, a rate-limiting mechanism to prevent TDM tools overwhelming publisher websites

- Copyright Clearance Center (CCC) is piloting a new service for use by life science companies.[126] This will aggregate article content from multiple rightsholders in a single service with normalised metadata, and authorises access to and downloading of subscribed-to content by researchers within commercial life science organisations. The system thus has the potential reduce the necessity for one-off licensing negotiations, along with the associated administration costs, while providing royalties to rightsholders when their content is used for TDM

- Infotrieve has developed a service that allows researchers to search across sets of full text content to build corpuses of content for text mining by aggregating content from multiple publishers.[127] Infotrieve was subsequently acquired by CCC, so this service may become integrated with CCC's own

- PLSclear is a web service to simplify the process of making and managing requests to access publisher content for TDM, working as an online clearing-house for research requests[128].

4.10. Reproducibility

The lack of reproducibility of scientific research published in journals is increasingly perceived as a serious problem (sometimes called the "reproducibility crisis"). The issue came to widespread public attention in 2013 with the publication of a cover story in *The Economist* (Anon 2013) but had been on the radar for research funders for much longer. The NIH has taken a lead to develop policies to address the issue (Collins & Tabak, 2014), including better training for investigators; more systematic evaluation of grant applications; greater transparency of research data including a proposed new Data Discovery Index as well as more rigorous enforcement of its data sharing requirements; and the launch of Pubmed Commons to support open discussion on published articles (see *Post-publication peer review*).

---

125 http://tdmsupport.crossref.org

126 http://www.stm-assoc.org/
2013_05_20_FACT2_Billington_CCCs_Text_and_Data_Mining_Pilot_Service.pdf

127 http://www.businesswire.com/news/home/20140707005558/en/Infotrieve-Spearheads-Text-Mining-Initiative

128 http://www.plsclear.com/Pages/ClearWizard.aspx

Reproducibility is a complex, multi-dimensional problem with roots deep in research process, organisation and culture, but is also affected by some aspects of publishing. These include incentives and pressures for early publication; selective publication of positive findings; and weak challenge of statistical analysis in peer review.

Publishers and journals have responded in a number of ways, including the introduction or enforcement of policies on the registration of trials; introduction of policies on data deposit and sharing (see *Data sharing and journal policies*); encouraging or requiring the sharing of computer code as well as research data; strengthening peer review, for instance by adoption of reviewer checklist and by making greater use of statistical experts during review.

Other publisher approaches might include publication of negative findings; extension of the prior registration model from clinical trials to other types of study;[129] and semantic markup of entities like reagents and antibodies to ensure unique identification.

4.11. Big data & analytics

"Big data" refers to collections of data too large to be handled and analysed in conventional databases systems.[130] 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 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* 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

---

129 e.g. the Elsevier journal Cortex has introduced a new article format called the Registered Report: http://www.journals.elsevier.com/cortex/news/registered-reports-a-new-article-format-from-cortex/

130 this is something of a moving target but currently measured in petabytes and exabytes

- 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, with data covering the research performance of 4,600 research institutions and 220 countries

- 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 above (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 (~470 million records) and users (~3.5 million). This allows it to create statistics and article recommendations, and to create services for institutions, for instance to help librarians understand how their collections are used by their patrons.

4.12. Identity and disambiguation

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; 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).[131] ORCID (the organisation) is a non-profit collaboration involving participants from across the research and scholarly communication worlds (around 150 organisations, 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. The ORCID registry was launched in late 2012 and had at the time of writing over 1.1 million live IDs.[132]

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. For example, new services that ORCID include functionality to support grouping of works by identifier, better management of duplicate works, several social features, and the ability to import Bibtex bibliography files (Cochrane 2014).

---

131 http://about.orcid.org

132 ORCID statistics: https://orcid.org/statistics

4.13. Research management and 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, 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; also called Research Information Management, or RIM) within institutions.[133] 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 three 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 its PURE service; Thomson Reuters, whose CRIS Converis (previously AVEDAS, acquired by Thomson Reuters and integrated with its Research in View service), and InCites analytics suite are supported by the Web of Knowledge database; and Digital Science (the sister company to Nature Publishing Group), which has a presence in this nascent market through its ownership of Symplectic. There are also some non-commercial national-level initiatives such as METIS (Netherlands) and CRIStin (Norway).

The main services provided are subscription-based tools and services (e.g. to analyse relative competitive strengths of research programmes, identify collaborators, measure individual/team research performance, etc.); custom research and analytics[134]; and data licensing for internal analysis.

A related kind of service is that provided by UberResearch (another Digital Science company), which supports research funders' decisions on which grants to fund by allowing comparison with existing publicly funded (but not necessarily published) research.

There is a separate market for corporate research analytics services, for example in the pharmaceutical and high-tech engineering sectors but these services are outside the scope of this report.

For metrics such as institutional comparisons to be used to support management and policy decisions they have to be reliable and comparable to metrics used by other institutions. The Snowball Metrics project addresses this issue by aiming to create and share universally agreed research-related metrics, complete with standardised "recipes" for how they should be calculated, including the data sources available for doing so (Jump 2014b).

4.14. FundRef

FundRef is a collaborative pilot project of scholarly publishers and funding agencies, facilitated by CrossRef, to provide a standard way of reporting funding sources for

---

133 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

134 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)

published scholarly research. Essentially the system consists of a standard set of codes to represent funding agencies, with some 4000 funders covered initially, and backend systems to tie articles and funders together via DOIs. The intended benefits are helping funders report on the research and development outcomes they supported, more systematic reporting of research funding in publications, and text mining applications.

Like any metadata, it will become more valuable only when widely adopted. Until recently few journals have required authors to use FundRef data to capture funding during the article submission. Nonetheless at the time of writing CrossRef statistics showed some 386,000 DOIs had associated FundRef metadata. The use of FundRef seems likely to grow rapidly because it is valued by research funders and, in particular, its adoption will be necessary to support CHORUS (qv).

4.15. Library publishing

There has been an expansion of interest among academic libraries in providing publishing services over the last 4 years or so (Jones 2014b). A 2011 ARL report highlighted the potential but described the field as evolutionary with many of the programmes being exploratory (Ivins & Luther, 2011). The *Library Publishing Directory 2015*, however, presents a picture of vibrant activity, reporting some 124 case studies, mostly in the USA and Canada (Lippincott 2014). Most library publishers (90%) work with local academic departments, but more than half provide publishing services to third-party organisations such as learned societies and research institutes, thus publishing some 194 journals.

Libraries surveyed published a total of 432 faculty-driven (as opposed to student-driven, of which there were 214), campus-based journal titles, nearly all of which were open access. The OA journals rarely (10%) charged APCs, instead covering the publishing costs from the library budget. The total number of library-published journals is likely to larger than reported in the Directory, however; for example, Open Journal Systems hosts many thousand journals and bepress around 700 journals, many of which may be library-published.

California Digital Library's recently announced open access megajournal *Collabra* is an example of library publishing comparable to publishing industry equivalents. Most library publishing, however, combines lightweight publishing services with lightweight technical solutions such as Open Journal Systems, bepress, DSpace, and WordPress, with Ubiquity Press also now competing in this space. The services are becoming more sophisticated, though, including metadata assignment (80% of library publishers), peer-review management (25%) and marketing (41%). Importantly, discoverability is not being neglected, for instance through provision of metadata to web-scale discovery services (qv) like Primo and Summon.

4.16. Open Annotation

Open annotation, a new open specification for web-based annotation[135], offers the potential for richer types of commentary and and discourse to be supported in a layer sitting over journal (and other academic) content (Carpenter 2013).

Open annotation shares some features with simpler forms of annotation (e.g social bookmarking services) but supports multiple annotation types, including bookmarking, highlighting, tagging and commenting. Annotation does not require either the permission of

---

135 http://openannotation.org

the annotated website or that it installs any new software. Publishers may, however, choose to run their own open annotation services which could allow for instance richer features to be offered to subscribers or registered users.

Annotations can be linked not just to web documents but to specific location within pages, right down to the sentence level, permitting more meaningful and interactive commentary. Additionally, annotation and linking is not limited to text: the standard supports annotation of non-textual materials such as images, maps and videos. Open annotations are also citable and can be preserved as part of the scholarly record.

A leading provider of open annotation services is the not-for-profit Hypothes.is, which also organises an annual conference (iAnnotate). Other organisations developing tools and services within the scholarly sphere include Annotator (Open Knowledge Foundation), Domeo (Mass. General Hospital), and PubPeer. General-purpose web annotation tools that might be co-opted for scholarly purposes include Genius and Diigo.

Enhanced PDF readers aimed at STM audiences offer alternative (non-standards based) ways of sharing annotation, for example ReadCube, Utopia Docs, Colwiz, Mendeley, etc.

4.17. Learned societies

Learned societies and associations serve a variety of missions and purposes, typically embracing the advancement of field or discipline, or the advancement of the interests of the members, or both. Publishing journals has long been a central part of their roles (along with conferences and meetings, education, and so on), but while the journal remains important for advancing the discipline, its value as a membership benefit is being cast into doubt by changes in the publishing landscape, notably the move away from print, the wide availability of journals through the consortia licensing model, and to a lesser extent (to date) by the open access journal.

Society members continue to support the idea of their societies publishing journals. For instance the Ithaka survey found the two most valued roles were perceived to be "Organises conferences and other in-person meetings" and "Publishes peer-reviewed academic journals" (Ithaka S+R et al., 2013). On the other hand, many surveys have shown that an association of a journal with an established society is of only minor importance at best to authors in choosing a journal to publish in (e.g. Nature Publishing Group 2014).

New roles and member benefits are not easy to find. Online community and networking services are often advanced as a natural digital analogue of the society's real-life meetings and networks, but few societies have yet been successful in this. Members express little interest in their societies developing such services (e.g. Ithaka S+R et al., 2013), perhaps because wider networks offer greater value than the niche focus of most societies, leaving opportunities in the space open to entrepreneurs like Academia, Mendeley, and ResearchGate.

These issues (and other factors such as the often antiquated governance structures (Outsell 2014a) lead many societies to adopt a low risk, low innovation stance in journal publishing. This is particularly evident in relation to open access, where societies (with of course some notable exceptions) have not been among the leaders and innovators, often fearing financial or reputational risk, or damage to their flagship titles.

Pressures such as these are likely to continue the longstanding trend for society journals to move from self-publishing to partnership with publishers. The publishers that specialise in society partnering are starting to explore new ways to help societies improve the

membership offer (beyond the member journal subscription: see for example Wiley's partnership with Knode). Given the competitive market for society journal contracts, we might expect to see more innovation in this area in coming years.

4.18. Author services and tools

Publishers have of course always provided services such as peer review and copy-editing to authors, but increased competition for authors, globalisation of research (hence a greater proportion of authors with weaker English language skills), and new enabling technologies are driving an expansion of author services. These could be grouped in presubmission services; production; information and alerts; marketing and promotion; and discounts and other services.

Presubmission services. These include journal selection tools (e.g. Research Square's JournalGuide, CoFactor, and Edanz Journal Selector; a customised version of the latter is available as part of the Springer Author Academy site); language and translation services (most publishers outsource, though some (e.g. OUP) do it themselves); presubmission enquiries and screening; journal information pages (these are becoming increasingly open about sharing current data on their author-related performance such as peer review times, production times, etc.)[136].

Production. Manual services such as redrawing or relabelling figures are now rare, but have been replaced by automated services such as reformatting of reference lists (and removal of unnecessary styling requirements for submitted manuscripts generally), and e-proofing tools.

Information and alerts. Tracking and status reporting during production; citation alerts following publication.

Marketing and promotion. Given authors' growing need to maximise the visibility and impact of their work, there is plenty of scope here: article-level metrics and usage statistics; advice and tools/services for authors to promote their own papers, and integration of services like Kudos, Publiscize or ImpactStory; toll-free shareable links for subscription content or shareable versions (e.g. ReadCube/NPG).

Discounts and other. Discounts on books etc. are longstanding offerings; new digital services include access to bibliographic databases and tools for editors and reviewers.

In addition to improving the range and quality of services to authors, publishers are also seeking to improve the user experience (simplifying processes, eliminating unnecessary stages, generally improving ease of use, etc.). The online submission and tracking system is therefore an important part of this equation.

4.19. Collaborative writing and sharing tools

Although there has been discussion for some time of the potential benefits of offering collaborative writing tools aimed specifically at scientific authors, the dominance of Microsoft Word has limited the demand (Perkel 2014). Google Docs is freely available and has created awareness of the benefits of online writing tools, but lacks many features required in scientific writing.

---

136 see for instance at BMJ (http://heart.bmj.com/site/about/) or Elsevier (e.g. http://journalinsights.elsevier.com/journals/0377-0257

The startup Overleaf (originally WriteLaTeX, in which Macmillan's Digital Science recently invested) offers such a service, which is claimed to have over 150,000 users at more than 1000 institutions and over 2 million documents created.

Other academic online writing tools include Authorea, Fidus Writer and shareLaTeX. The Authorea platform offers publishing services in addition to writing.[137] Another service, Annotum, offers a writing, peer review and publishing platform based on WordPress with extensions to support scholarly content. The Plot.ly website allows the collaborative creation of graphs on a cloud-based platform; graphs can be shared either on the platform or by using code to embed, allowing users access to the underlying data.

At present all these services are used by a tiny minority of scientists. This may change with publisher endorsement and integration. For instance, WriteLaTeX offers publishers a web service to provide "one-click" submission from Overleaf to the publisher's system. The typesetter River Valley has developed a somewhat similar service, RVPublisher, marketed primarily at publishers. Some publishers are also actively exploring this area: Elsevier has reported working on the creation of authoring tools to support semantic mark-up, and Wiley is similarly researching options for capturing more structured information from authors.

4.20. 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.

Open notebook science has been adopted by a tiny (close to vanishingly small) minority of researchers. We were skeptical in the last two editions 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 (including the two live examples cited previously). 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.

---

137 Authorea have created an interesting interactive paper describing "the paper of the future" using features of the Authorea to demonstrate its features: https://www.authorea.com/users/23/articles/8762/_show_article

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 clear trends.

The web 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. If anything, the need for certification has become stronger, driven by increased competition, globalisation, research assessment and funders' growing emphasis on "impact".

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. The newer scientific social networks (Academia, Mendeley, ResearchGate) have rapidly expanded in recent years but at present much of the motivation for creating accounts seems to be about increasing visibility and impact, rather for active online discussion or collaboration, or a more convenient form of reference management.

Widespread adoption of smartphones has had limited impact, and even though tablets are starting to change the work practices of some physicians and other healthcare professionals, among others, they also appear to had little impact on most researchers' working practices.

Researchers' access to scholarly content is at a historic high. Bundling of content and associated consortia licensing model 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 continues apace, associated with accelerating growth in research outputs. 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 (and are in marked contrast to the retrenchment in Russia).

The Research4Life programmes (HINARI, AGORA, OARE, and ARDI) have again continued to expand, seeing further 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 these 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. Despite regular criticism, the model seems more likely to evolve (e.g. bundling open access charges; new pricing structures not linked to historic print holdings) than to be superseded.

Despite all this, improving access via open access has clearly dominated publishing industry and policy developments. The role of funders and governments has remained central, ranging from the Green-focused OSTP memorandum in the United States to the Gold-oriented RCUK policy in the UK. In the last report we discerned a clear policy shift towards Gold. It is now clear that this view was over-influenced by then-recent events in the UK:

outside the UK, Austria and the Netherlands, the dominant policy stance is very much more Green-tinted. One the one hand, virtually all funders will reimburse Gold APC payments, but on the other, their open access policies tend to lean more towards Green.

Nonetheless, open access journals continue to grow rapidly in number and in output. Estimates vary, but DOAJ lists over 10,000 OA journals, and Gold (including without APC) now represents 11-12% of articles, Green at least another 12%, and delayed access perhaps another 5%. Megajournals remain the fastest growing segment, with recent announcements from publishers as diverse as Elsevier and California Digital Library.

Turning to Green open access, the numbers of institutional repositories has continued to grow, with ROAR-listed repositories up from nearly 3000 in 2012 to over 3900 now. 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). Indeed, by one study, self-archiving in unsystematic, unstructured ways such as author or departmental home pages may currently be more important than institutional repositories. Subject repositories (PubMed Central, arXiv (recently passing its 1 millionth article milestone), SSRN, RePEc etc.) remain 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.

Budgetary implications no doubt make it easier for policy-makers to opt for Green, but the lowering of trust in publishers and suspicion of the profit motive that we have commented on in the last two editions have also been a factor. This mistrust has if anything grown in importance, though much of the debate around the future of scholarly communication (especially open access) remains characterised by lack of hard evidence and rhetorical argument on both sides, as was discussed by librarian Rick Anderson (Anderson 2014c). There is a danger in over-weighting the noisy minority, but high-profile criticism in national newspapers (with global audiences), author boycotts, and increasingly difficult publisher-government licensing negotiations all help to create a climate in which (for example) some learned societies now shy away from working with the largest commercial publishers. Away from the frontline of open access activism – an activity of interest to a tiny fraction of active researchers – publishers and researchers continue to work productively and fruitfully together, though it can be easy to lose sight of this in the heat of the debate.

The management and sharing of research data, and their linking to and integration with the research literature, have moved decisively centre-stage. The creation of the Research Data Alliance was a visible reflection of the importance the research community is placing on this topic, while ensuring research data is made available is a priority for research funders. The benefits are potentially huge, as set out in the influential Royal Society report *Science as an Open Enterprise.* There were tangible developments within both the research and publishing communities, including in terms of policy, an expansion of data repositories, and publishing innovation including data journals. We remain, however, at the very start of this revolution. In a related area, there was also progress in developing solutions to enable text and data mining, though demand (away from the pharma/biotech sector) remains nascent.

Within the publishing industry, mergers and acquisitions activity appears buoyant. Partly this reflects long-standing structural trends such as economies of scale and scope (though these have been accentuated by the web), such as the planned merger of Springer and Macmillan Science & Education, but also reflects the growing importance of technological

innovation and the shift from content to content plus services (e.g the acquisitions by Elsevier of Knovel and Mendeley, and by Thomson Reuters of AVEDAS).

~ ~ § ~ § ~ ~

Looking to the future, it would be reassuring to be able to hold onto some constants in a fast-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 open access megajournals may seem to run counter to these fundamentals in respect of the undifferentiated branding and lightweight peer review. However, it is early days for these titles and we may yet see them as much driven by other factors, such as being an option for papers by the more prestigious titles (the cascade model) and new author entrants or simply examples of brand extension (e.g. *Nature Communications*, *Royal Society Open Science*).

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 over the longer term, though it is much more cyclical than public funding. Traditional academic acquisition budgets will grow slowly (if at all) in real terms, despite some belated bounce-back from the 2008/09 recession, particularly in Europe where recovery appears stalled at best (and the IMF recently warned that global economic growth may never return to the pre-crisis rates). Buyers of all kinds will seek demonstration of the value of their purchases (through usage and perhaps in more sophisticated, metrics-driven ways). Emerging markets will continue to provide the best growth opportunities, though perhaps slowing compared to the recent past. And faced with hyperabundance of content, readers will value relevance, usability, insights and answers over raw access.

Open access will continue to 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 few years and may yet stall at a mixed market position. But the momentum in uptake among authors, publishers and funders is clearly there and if anything will accelerate in the short term at least.

The research and publishing communities are still working out 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? What will become the market rate for publication charges, currently anywhere from $250 to $5000 (and indeed zero)? Will market forces push down APCs overall, and if so what will be the consequences, or will prices stratify, as in other competitive markets?

The demand for open reuse rights (e.g. CC-BY licensing for articles, CC0 for data) for OA content will also continue, for the same main reason – demand from research funders (who rightly or wrongly see CC-BY and its ilk as important for reuse and in particular for new approaches to reuse such as text and data mining) – and for the same reason seems likely to be conceded, despite being unpopular with a substantial fraction of researchers.

Green OA and the role of repositories will remain controversial with publishers. This is less the case for institutional repositories (which – despite growth in their numbers and content – remain under-used by authors, and a relatively unimportant channel for reader discovery and use), than for subject repositories, especially PubMed/PMC. The latter is acting increasingly like a publisher, and investing in its platform and related features (e.g. data integration, advanced article formats, a commenting platform). Embargo lengths will also continue to be the subject of debate, though we do not anticipate much movement.

Open access will also not just be for journal articles: OA monograph models are starting to emerge and are expected to expand. Other kinds of open books and open educational resources will also become more important.

Not all publishers are equal: the ALPSP survey (Inger & Gardner, 2013) showed that publishing policies and practices at smaller publishers were notably less forward looking and innovative than at larger and medium-sized publishers. Society publishers (except for the largest), will continue to find it difficult to adapt to industry and research community changes, including open access, while their parent societies struggle with maintaining relevance to their memberships.

The role of the library continues to evolve. While the core function of providing and supporting access to scholarly content will remain, libraries will potentially play important roles in expanding the use of institutional repositories, including supporting the local management of research data; in managing open access funding and payment schemes; and in the newly emergent area of library publishing.

In the consumer world, successful new media brands are increasingly transcending content "containers" (i.e. categorisation by types such as magazine, news, television, book, audio, etc.). There will be a similar shift (albeit not a new one) for STM users and purchasers away from the container (journal, monograph) to the research content itself, and services built around this. This may lead to a further expansion of comprehensive collection licensing and perhaps to growth of the "all you can eat" subscription services aimed at small corporates or individuals.

The digital transition away from print will of course continue. Although essentially all journals have electronic editions, legacy print continues, especially in books. For research journals, virtually all meaningful use has now migrated online (even if the downloaded PDF is later printed off), and the next few years will see a growing fraction of these journals dropping their print editions. For society journals, especially in the clinical and professional fields, and for the general-purpose journals, at least some of the residual print use seems likely to move to new mobile devices.

Though mobile access may not (yet) have had the impact on STM that some might have anticipated, given its transformative effects in some part of consumer media, expanding mobile adoption will be an important trend for the next three years, particularly for clinical and practitioner areas. It remains early days: mobile devices represent under 10% of STM platform accesses (though higher for some clinical platforms) and increasing 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 (say annotation, reference management, or writing) 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, potentially occupying a new niche alongside desktop and smartphone screens.

Business strategies are also still evolving. 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 been some experimentation with pricing and with in-app purchase and freemium models, though with limited success, 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; for some, providing a mobile accessible site (e.g. through responsive design) may simply become a standard platform cost.

The new scientific social networks have such large registered user bases it seems impossible that they will not play some role in the evolving landscape. At the very least they are likely to become another channel for discovery and sharing of new content, and for the discovery of potential collaborators.

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 makes content smarter, improving discoverability and use, and will be one way of making content more interactive. It also enables new products and services, and supports internal productivity.

Publishing platforms will be increasingly convergent with respect to content type (i.e. hosting journals, books and data equally) and neutral as to business model. Mishaps such as hybrid open access articles slipping behind paywalls, or platforms requesting rights payments for reuse of CC-BY material are a consequence of retro-fitting OA features onto platforms built with the subscription model in mind: understandable, but in need of a fresh approach.

STM publishing platforms are now starting to catch up with user interface/user experience developments common in the consumer web, though this has been relatively neglected to date. It reflects a growing focus on the researcher (as opposed to the library), driven partly by the redefining of the customer in the OA model, but also by a focus on research assessment and metrics. For similar reasons attribution, citation and credit will become more fine-grained, both in terms of more granular definitions of authorship (or contributorship), and in terms of what can be cited (e.g. individual data elements within a larger, and indeed dynamic, dataset).

More platforms will feature interactive content using approaches similar to those explored in the eLife Lens, Wiley Anywhere Article, and similar initiatives. 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, as data becomes an increasingly central part of research outputs, journals will need not only 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 will become increasingly common, initially supported by the "data paper" model.

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.

Whether all this amounts to the "Science 2.0" or "open science" transformation envisaged by some policy-makers and other advocates is unclear. The driving factors are undeniable: growth in research outputs, increase in research numbers and the globalisation of research, and the move to data-intensive science. But even if marrying these to expansion of open access to articles and data does not lead to the increase in economic innovation (driven by faster, more socially-responsive science) that policy-makers seek, it nonetheless represents a clear challenge for the publishing sector.

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. With innovation a key success factor, recruitment of talent from outside the traditional STM publishing sector, partnering, and acquisition of technology start-ups will all become more common.

6. Information sources

In addition to the details references, the following sources of information may be helpful.

6.1. Publisher organisations

- The International STM Association: broad coverage, including copyright, public affairs, and standards and technology www.stm-assoc.org

- ALPSP (Association of Learned, Professional and Society Publishers): a wide range of information resources are available from its website, including the periodic *Scholarly Journals Publishing Practices* survey (currently in its 4th edition) www.alpsp.org

- OASPA (Open Access Scholarly Publishers Association) oaspa.org/

6.2. Global statistics and trends

- Battelle Global R&D Funding Forecast. Produced annually www.battelle.org/media/publications/global-r-d-funding-forecast

- NSF Science & Engineering Indicators 2014. Produced every two years (the previous editions are well worth reviewing as well as the current edition) http://www.nsf.gov/statistics/seind14/

- OECD Stat.Extracts, Main Science and Technology Indicators http://stats.oecd.org/Index.aspx?DataSetCode=MSTI_PUB#

- UNESCO Science Report – the most recent edition (2010) is somewhat dated; the 2015 edition is due to be published in November 2015 http://www.unesco.org/new/en/natural-sciences/science-technology/prospective-studies/unesco-science-report/

6.3. Open access

- Open Access News and Resources (Copyright Clearance Center/ALPSP) http://www.copyright.com/content/cc3/en/open_access.html

- The CREATe 2014 working paper *Open Access Publishing: A Literature Review* covers some 750 literature references for those looking for in-depth coverage (Frosio 2014) http://www.create.ac.uk/publications/000011

- OpenDOAR (Directory of Open Access Repositories): includes statistics and charts on growth of repositories, types of content, etc. www.opendoar.org/

- ROAR (Registry of Open Access Repositories): similar to OpenDOAR but also includes ROARMAP, a registry of OA mandates and policies roar.eprints.org/

- SHERPA/RoMEO: a database of Publisher copyright policies with respect to self-archiving and reuse http://www.sherpa.ac.uk/romeo/

- SHERPA/JULIET: Research funders' open access policies http://www.sherpa.ac.uk/juliet/index.php?la=en

6.4. Publishing industry research and analysis

- Outsell: research and advisory service focusing on information content strategy and use , serving a wide range of vendors, buyers, and users of information www.outsellinc.com/

- Simba Information: market intelligence and forecasts in the media and publishing industry www.simbainformation.com/

- Publishing Research Consortium: a number of useful reports are freely available, including on peer review, text and data mining, and aspects of open access www.publishingresearch.org.uk

- Research Information Network: studies and reports on developments in scholarly communications. Current projects at http://www.researchinfonet.org; the influential reports commissioned by RIN in its earlier guise between 2006 and 2011 are listed here (may require some Googling) http://www.researchinfonet.org/links/

7. References

Acharya, A., Verstak, A., Suzuki, H., Henderson, S., Iakhiaev, M., Lin, C. C. Y., & Shetty, N. (2014). Rise of the rest: The growing impact of non-elite journals. *Eprint arxiv:1410.2217*. Retrieved from arXiv.

Adams, J., & Gurney, K. (2014). *Evidence for excellence: Has the signal overtaken the substance?*. Retrieved from http://www.uberresearch.com/first-digital-research-report-launched-by-digital-science-looking-wether-there-is-evidence-for-excellence/

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

Allen, L., Scott, J., Brand, A., Hlava, M., & Altman, M. (2014). Publishing: Credit where credit is due. *Nature News*, *508*(7496), 312. doi:10.1038/508312a

American Society for Cell Biology, & et al. (2012). *San Francisco declaration on research assessment*. Retrieved from http://am.ascb.org/dora/

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/

Anderson, K. (2014a). 82 things publishers do (2014 edition). *Scholarly Kitchen Blog.* Retrieved from http://scholarlykitchen.sspnet.org/2014/10/21/updated-80-things-publishers-do-2014-edition/

Anderson, R. (2014b). *Housecleaning at the directory of open access journals*. Retrieved from http://scholarlykitchen.sspnet.org/2014/08/14/housecleaning-at-the-directory-of-open-access-journals/

Anderson, R. (2014c). Is rational discussion of open access possible? *Insights: The UKSG Journal*, *27*(2), 171-180. doi:10.1629/2048-7754.162

Anon. (2013, October 19). How science goes wrong. *Economist*. Retrieved from http://www.economist.com/news/leaders/21588069-scientific-research-has-changed-world-now-it-needs-change-itself-how-science-goes-wrong

Archambault, E., Amyot, D., Deschamps, P., Nicol, A., Provencher, F., Rebout, R., & Roberge, G. (2014). *Proportion of open access papers published in peer-reviewed journals at the european and world levels 1996–2013*. Retrieved from http://science-metrix.com/files/science-metrix/publications/d_1.8_sm_ec_dg-rtd_proportion_oa_1996-2013_v11p.pdf

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

Aspesi, C. (2012). *Reed Elsevier: Transitioning to open access - are the cost savings sufficient to protect margins?*. Retrieved from http://www.richardpoynder.co.uk/OAcosts.pdf

Aspesi, C. (2014). *Reed Elsevier - goodbye to Berlin - the fading threat of open access*. Retrieved from http://www.richardpoynder.co.uk/Aspesi.pdf

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

Baker & McKenzie. (2014). *Internal draft of European Commission's copyright policy paper revealed*. Retrieved from http://www.lexology.com/library/detail.aspx?g=ad633f55-0f79-48f1-8ede-93b5606aa818

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

Battelle. (2013). *Global R&D funding forecast*. Retrieved from http://www.rdmag.com/articles/2013/12/2014-r-d-magazine-global-funding-forecast

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

Bechhofer, S., Buchan, I., De Roure, D., Missier, P., Ainsworth, J., Bhagat, J.,... Goble, C. (2013). Why linked data is not enough for scientists. *Future Generation Computer Systems*, *29*(2), 599-611. doi:10.1016/j.future.2011.08.004

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. *J Med Internet Res*, *13*(4), e115. doi:10.2196/jmir.1802

Björk, B. -C. (2014). Open access subject repositories: An overview. *Journal of the Association for Information Science and Technology*, *65*(4), 698-706. doi:10.1002/asi.23021

Björk, B. -C., & Solomon, D. (2012a). Open access versus subscription journals: A comparison of scientific impact. *BMC Med*, *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., & Solomon, D. (2014). Developing an effective market for open access article processing charges. Wellcome Trust. Retrieved from http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtp055910.pdf

Björk, B. C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology*, *62*(2), 237-250. doi:10.1002/asi.22963

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

Bloom, T. (2014). PLOS' new data policy: Part two. *PLOS EveryONE Blog.* Retrieved from http://blogs.plos.org/everyone/2014/03/08/plos-new-data-policy-public-access-data/

Bollen, J., de Sompel, H. V., 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

Bornmann, L., & Mutz, R. (2014). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Eprint arxiv:1402.4578*. Retrieved from arXiv.

Callaghan, S. (2014). Preserving the integrity of the scientific record: Data citation and linking. *Learned Publishing*, *27*(5), 15-24. doi:10.1087/20140504

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.

Carpenter, T. (2013). IAnnotate — whatever happened to the web as an annotation system. *Scholarly Kitchen Blog.* Retrieved from http://scholarlykitchen.sspnet.org/2013/04/30/iannotate-whatever-happened-to-the-web-as-an-annotation-system/

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

Citron, D. T., & Ginsparg, P. (2014). Patterns of text reuse in a scientific corpus. *Proceedings of the National Academy of Sciences*, 201415135. doi:10.1073/pnas.1415135111

Clark, A., Jensen, L., & Campbell, D. (2014). *Member briefing: Text and data mining*. ALPSP. Retrieved from http://www.alpsp.org/Ebusiness/ResearchPublications/ AdviceNoteGuidelines.aspx

Clark, J. (2013). *Text mining and scholarly publishing*. Retrieved from http:// www.publishingresearch.net/documents/ PRCTextMiningandScholarlyPublishinFeb2013.pdf

Clarke, M. (2013). Game of papers: ELife, BMC, PLOS and EMBO announce new peer review consortium. *Scholarly Kitchen.* Retrieved from http://scholarlykitchen.sspnet.org/ 2013/07/15/game-of-papers-elife-bmc-plos-and-embo-announce-new-peer-review-consortium/

Cochran, A. (2014). *Implementing CHORUS: Big decisions loom for publishers*. Retrieved from http://scholarlykitchen.sspnet.org/2014/07/09/implementing-chorus-big-decisions-loom-for-publishers/

Cochrane, A. (2014). Interview with Laurel Haak of ORCID: Supporting the efforts with membership and integration. *Scholarly Kitchen Blog.* Retrieved from http:// scholarlykitchen.sspnet.org/2014/09/18/interview-with-laurel-haak-of-orcid-supporting-the-efforts-with-membership-and-integration/

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

Collins, F. S., & Tabak, L. A. (2014). Policy: NIH plans to enhance reproducibility. *Nature News*, *505*(7485), 612. doi:10.1038/505612a

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. (2014). *Release 1 of the COUNTER code of practice for usage factors*. Retrieved from http://www.projectcounter.org/usage_factor.html

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

Cressey, D. (2014). Journals weigh up double-blind peer review. *Nature News.* doi:10.1038/ nature.2014.15564

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/

Czujko, R., & Chu, R. (2015). *Scholarly and scientific publishing workforce in the U.S.* AAP/AIP. Retrieved from http://publishers.org/press/158/

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

Davis, P. (2014). Peak PLOS: Planning for a future of declining revenue. *Scholarly Kitchen.* Retrieved from http://scholarlykitchen.sspnet.org/2014/12/02/peak-plos-planning-for-a-future-of-declining-revenue/

Davis, P. M. (2011). Open access, readership, citations: A randomized controlled trial of scientific journal publishing. *FASEB J, 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. *J Med Libr Assoc, 99*(3), 208-17. doi:10.3163/1536-5050.99.3.008

Department for Business, Innovation & Skills. (2014). *Science and innovation: Rates of return to investment.* Retrieved from https://www.gov.uk/government/publications/science-and-innovation-rates-of-return-to-investment

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

Economist. (2014). No peeking…. *The Economist.* Retrieved from http://www.economist.com/news/science-and-technology/21593408-publishing-giant-goes-after-authors-its-journals-papers-no-peeking

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

Edwards, D. C. (2014). How can existing open access models work for humanities and social science research? *Insights: The UKSG Journal, 27*(1), 17-24. doi:10.1629/2048-7754.135

Eisen, M. (2013). *A CHORUS of boos: Publishers offer their "solution" to public access.* Retrieved from www.michaeleisen.org: http://www.michaeleisen.org/blog/?p=1382

eLife. (2014). *A foundation for change: The elife Sciences 2013 annual report.* eLife. Retrieved from http://2013.elifesciences.org/#toc/all

Elliot, L. (2014, October 7). IMF says economic growth may never return to pre-crisis levels. *The Guardian.* Retrieved from www.theguardian.com: http://www.theguardian.com/business/2014/oct/07/imf-economic-growth-forecasts-downgraded-crisis

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

Elsevier. (2013). *International comparative performance of the UK research base – 2013*. Retrieved from https://www.gov.uk/government/publications/performance-of-the-uk-research-base-international-comparison-2013

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

European Commission. (2014). *Public consultation – "Science 2.0": Science in transition*. Retrieved from http://ec.europa.eu/research/consultations/science-2.0/background.pdf

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

Ferguson, C., Marcus, A., & Oransky, I. (2014). Publishing: The peer-review scam. *Nature News*, *515*(7528), 480. doi:10.1038/515480a

Ferguson, L. (2014). How and why researchers share data (and why they don't). *Wiley Exchanges Blog.* Retrieved from http://exchanges.wiley.com/blog/2014/11/03/how-and-why-researchers-share-data-and-why-they-dont/

Ferwerda, E. (2014). Open access monograph business models. *Insights: The UKSG Journal*, *27*, 35-38. doi:10.1629/2048-7754.46

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

FORCE 11. (2011). *FORCE 11 manifesto: Improving future research communication and e-scholarship*.

Frass, W., Cross, J., & Gardner, V. (2014). *Taylor & Francis open access survey June 2014*. Retrieved from http://www.tandf.co.uk/journals/explore/open-access-survey-june2014.pdf

Frosio, G. (2014). *Open access publishing: A literature review*. CREATe. Retrieved from http://www.create.ac.uk/publications/000011

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

Gantz, P. (2012). *Digital licenses replace print prices as accurate reflection of real journal costs*. Retrieved from http://publishers.org/_attachments/docs/library/2012%20bulletin.pdf

Gantz, P. (2013). Journal print subscription price increases no longer reflect actual costs. *Learned Publishing*, *26*(3), 206-210. doi:10.1087/20130309

Gargouri, Y., Larivière, V., Gingras, Y., Carr, L., & Harnad, S. (2012). Green and gold open access percentages and growth, by discipline. *ArXiv Preprint http://arxiv.org/abs/1206.3664*

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

Global Coalition of Access to Research, Science and Education Organizations Calls on STM to Withdraw New Model Licenses. (2014). [Web page]. Retrieved from http://www.plos.org/global-coalition-of-access-to-research-science-and-education-organizations-calls-on-stm-to-withdraw-new-model-licenses/

Goble, C. (2008, September 23). 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

Graham, D. (2013). *Academic publishing: Survey of funders supports the benign open access outcome priced into shares*. HSBC Global Research.

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. (2009). Open access: What are the economic benefits? A comparison of the United Kingdom, Netherlands and Denmark. doi:10.2139/ssrn.1492578

Houghton, J. (2010). *Economic and social returns on investment in open archiving publicly funded research outputs*. SPARC. Retrieved from http://sparc.arl.org/sites/default/files/vufrpaa.pdf

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

Houghton, J., Swan, A., & Brown, S. (2011). Access to research and technical information in Denmark. Retrieved from http://eprints.soton.ac.uk/272603/

Huggett, S. (2014). A quick look at references to research data repositories. *Research Trends (Elsevier).* Retrieved from http://www.researchtrends.com/issue-38-september-2014/a-quick-look-at-references-to-research-data-repositories/

ICMJE. (2013). *Recommendations for the conduct, reporting, editing and publication of scholarly work in medical journals*. Retrieved from http://www.icmje.org/recommendations/

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

Inger, S., & Gardner, T. (2013). *Scholarly journals publishing practice: Academic journal publishers' policies and practices in online publishing.* ALPSP. Retrieved from http://www.alpsp.org/Ebusiness/ProductCatalog/Product.aspx?ID=359

Intellectual Property Office. (2014). *Intellectual property – guidance: Exceptions to copyright.* Retrieved from https://www.gov.uk/exceptions-to-copyright

Ioannidis, J. P. A., Boyack, K. W., & Klavans, R. (2014). Estimates of the continuously publishing core in the scientific workforce. *PLOS ONE, 9*(7), e101698. doi:10.1371/journal.pone.0101698

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

Ithaka S+R. (2013). *US faculty survey 2012.* Ithaka S+R. Retrieved from www.sr.ithaka.org: http://www.sr.ithaka.org/research-publications/us-faculty-survey-2012

Ithaka S+R, JISC, & RLUK. (2013). *UK survey of academics 2012.* Retrieved from http://www.sr.ithaka.org/research-publications/ithaka-sr-jisc-rluk-uk-survey-academics-2012

Ivins, O., & Luther, J. (2011). *Publishing support for small print-based publishers: Options for ARL libraries.* ARL. Retrieved from http://www.arl.org/component/content/article/6/1170

Jalalian, M., & Mahboobi, H. (2014). Hijacked journals and predatory publishers: Is there a need to re-think how to assess the quality of academic research? *Walailak Journal of Science and Technology, 11*(5), 389-394. doi:10.14456/WJST.2014.16

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

Jha, A. (2012). Open access to research is inevitable, says *Nature* editor-in-chief. *The Guardian.* Retrieved from http://www.theguardian.com/science/2012/jun/08/open-access-research-inevitable-nature-editor

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. (2014). *Jisc's evidence to the review of the RCUK OA policy.* Retrieved from http://scholarlycommunications.jiscinvolve.org/wp/2014/09/18/jiscs-evidence-to-the-review-of-the-rcuk-oa-policy/

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

Jones, J. (2014a). To flip or not to flip. In *ALPSP international conference.* Retrieved from http://blog.alpsp.org/2014/09/open-access-daily-challenge-new.html

Jones, P. (2014b). What's going on in the library? Part 1: Librarian publishers may be more important than you think. *Scholarly Kitchen Blog.* Retrieved from http://

scholarlykitchen.sspnet.org/2014/12/01/whats-going-on-in-the-library-part-1-librarian-publishers-may-be-more-important-than-you-think

Jump, P. (2014a). IOP launches 'offsetting' scheme to cut cost of open access. *Times Higher Education.* Retrieved from http://www.timeshighereducation.co.uk/news/iop-launches-offsetting-scheme-to-cut-cost-of-open-access/2013546.article

Jump, P. (2014b). Snowball metrics: No pinch of salt needed. *Times Higher Education.* Retrieved from http://www.timeshighereducation.co.uk/news/snowball-metrics-no-pinch-of-salt-needed/2014496.article

Khabsa, M., & Giles, C. L. (2014). The number of scholarly documents on the public web. *PLoS ONE, 9*(5), e93949. doi:10.1371/journal.pone.0093949

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

K | N Consultants. (2014). *A scalable and sustainable approach to open access publishing and archiving for humanities and social sciences.* Retrieved from http://knconsultants.org/toward-a-sustainable-approach-to-open-access-publishing-and-archiving/

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

Laakso, M., & Björk, B. -C. (2013). Delayed open access: An overlooked high-impact category of openly available scientific literature. *Journal of the American Society for Information Science and Technology, 64*(7), 1323-1329. doi:10.1002/asi.22856

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

Lawson, S. (2015). 'Total cost of ownership' of scholarly communication: Managing subscription and APC payments together. *Learned Publishing, 28*(1), 9-13. doi:10.1087/20140103

Levine-Clark, M., McDonald, J., & Price, J. (2014). Examining the impact of discovery systems on usage. *Wiley Exchanges Blog.* Retrieved from http://exchanges.wiley.com/blog/2014/07/21/examining-the-impact-of-discovery-systems-on-usage/

Lin, J. (2014). Making data count: PLOS, CDL, and DataONE join forces to build incentives for data sharing. *PLOS Tech Blog.* Retrieved from http://blogs.plos.org/tech/making-data-count-plos-cdl-and-dataone-join-forces-to-build-incentives-for-data-sharing/

Lippincott, S. (2014). *Library publishing directory 2015.* Library Publishing Coalition. Retrieved from http://www.librarypublishing.org/resources/directory/lpd2015

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.

Lynch, C. (2014). The need for research data inventories and the vision for SHARE. *Information Standards Quarterly*, *26*(2). Retrieved from www.niso.org: http://www.niso.org/ publications/isq/2014/v26no2/lynch/

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

Mandler, P. (2014). Open access: A perspective from the humanities. *Insights: The UKSG Journal*, *27*(2), 166-170. doi:10.1629/2048-7754.89

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

McCabe, M., & Snyder, C. (2013). Identifying the effect of open access on citations using a panel of science journals. *[preprint].* Retrieved from http://papers.ssrn.com/sol3/ papers.cfm?abstract_id=2269040

McCabe, M. J., & Snyder, C. M. (2014). Does online availability increase citations? Theory and evidence from a panel of economics and business journals. *Review of Economics and Statistics.* doi:10.1162/REST_a_00437

McEntyre, J. R., Ananiadou, S., Andrews, S., Black, W. J., Boulderstone, R., Buttery, P.,... Vaughan, P. (2011). UKPMC: A full text article resource for the life sciences. *Nucleic Acids Research*, *39*(Database issue), D58-65. doi:10.1093/nar/gkq1063

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. (2014). An interview with Amy Brand on a proposed new contributor taxonomy initiative. *Scholarly Kitchen.* Retrieved from http://scholarlykitchen.sspnet.org/

2014/08/20/an-interview-with-amy-brand-on-a-proposed-new-contributor-taxonomy-initiative/

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

Meeker, M. (2014). *Internet trends 2014*. Retrieved from http://kpcbweb2.s3.amazonaws.com/files/85/Internet_Trends_2014_vFINAL_-_05_28_14-_PDF.pdf?1401286773

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

Milloy, C. (2013). Innovative approaches to publishing open access monographs – it's not business as usual. *Jisc Inform*, (37). Retrieved from http://www.jisc.ac.uk/inform/inform37/OpenAccessMonographs.html

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

Morrison, H., & et al. (2014). Open access article processing charges: DOAJ survey may 2014. *[preprint]*. Retrieved from http://sustainingknowledgecommons.files.wordpress.com/2014/10/oa-apcs-article-2014-october-171.pdf

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

Murphy, F. (2014). Data and scholarly publishing: The transforming landscape. *Learned Publishing*, *27*(5), 3-7. doi:10.1087/20140502

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

Nature Editors. (2014). *Nature Publishing Index 2013 Asia-Pacific*. Retrieved from http://www.natureasia.com/en/publishing-index/asia-pacific/supplement2013

Nature Publishing Group. (2014). 2014 Author Insights survey. *figshare*. doi:10.6084/m9.figshare.1204999

Newman, M. L., & Sack, J. (2013). Information workflow of academic researchers in the evolving information environment: An interview study. *Learned Publishing*, 26(2), 123-131. doi:10.1087/20130208

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

Nicholas, D., Watkinson, A., Jamali, H. R., Herman, E., Tenopir, C., Volentine, R.,... Levine, K. (2015). Peer review: Still king in the digital age. *Learned Publishing*, 28(1), 15-21. doi: 10.1087/20150104

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

NISO SERU Standing Committee. (2012). *SERU recommended practice (RP-7-2012)*. NISO. Retrieved from http://www.niso.org/workrooms/seru/

NPG. (2014). *Author Insights 2014*. figshare. Retrieved from http://dx.doi.org/10.6084/m9.figshare.1204999

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

NSF. (2014). *Science and Engineering Indicators 2014*. Arlington, VA: NSB 14-01. Retrieved from http://www.nsf.gov/statistics/seind14/

OECD. (n.d.). *OECD.StatExtracts: Main Science and Technology Indicators.* Retrieved from stats.oecd.org: http://stats.oecd.org/Index.aspx?DataSetCode=MSTI_PUB#

OECD. (2011). *Main Science and Technology Indicators*. OECD. Retrieved from www.oecd.org/sti/msti

OECD. (2014). OECD science, technology and industry outlook 2014. doi:10.1787/sti_outlook-2014-en

OpCit. (n.d.). *The effect of open access and downloads ('hits') on citation impact: A bibliography of studies*. Retrieved from http://opcit.eprints.org/oacitation-biblio.html#most-recent

Orduña-Malea, E., Ayllón, J. M., Martín-Martín, A., & López-Cózar, E. D. (2014). About the size of Google Scholar: Playing the numbers. *arXiv eprint* http://arxiv.org/abs/1407.6239

OSTP. (2012). *US office of science and technology policy requests for information on public access to scholarly journals and data*. Retrieved from http://www.whitehouse.gov/administration/eop/ostp/library/publicaccess

OSTP. (2013). Increasing access to the results of federally funded scientific research. Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

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

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

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

Outsell. (2013). *Open access: Market size, share, forecast, and trends*. Retrieved from http://www.outsellinc.com/store/products/1135

Outsell. (2014a). *Professional associations and learned societies: Current issues and outlook*. Retrieved from http://www.outsellinc.com/store/products/1276

Outsell. (2014b). *Sensored world, sensible choices: Information industry outlook 2015*. Retrieved from http://www.outsellinc.com/store/products/1272

Outsell. (2014c). STM 2014 market size, share, forecast, and trend report. Retrieved from http://www.outsellinc.com/1280/products/1268-scientific-technical-medical-information-2014-market-size-share-forecast-and-trend-report

Outsell. (2014d). *STM platform providers: Market update*. Retrieved from http://www.outsellinc.com/store/products/1248

Outsell. (2014e). *Web-Scale discovery services: The evolution continues*. Outsell. Retrieved from http://www.outsellinc.com/1178/products/1264-im-report-web-scale-discovery-services-the-evolution-continues

Parsons, D., Willis, D., & Holland, J. (2011). *Benefits to the private sector of open access to higher education and scholarly research*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.368.1820&rep=rep1&type=pdf

Patel, J. (2014). Why training and specialization is needed for peer review: A case study of peer review for randomized controlled trials. *BMC Medicine, 12*(1), 128. doi:10.1186/s12916-014-0128-z

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

Perkel, J. M. (2014). Scientific writing: The online cooperative. *Nature, 514*(7520), 127-8. doi:10.1038/514127a

Phillpotts, J., Devenport, T., & Mitchell, A. (2015). Evolution of the Transfer Code of Practice. *Learned Publishing, 28*(1), 75-79. doi:10.1087/20150112

Pinfield, S., Salter, J., Bath, P. A., Hubbard, B., Millington, P., Anders, J. H., & Hussain, A. (2014). Open-access repositories worldwide, 2005-2012: Past growth, current characteristics, and future possibilities. *Journal of the Association for Information Science and Technology, 65*(12), 2404-2421. doi:10.1002/asi.23131

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

PLOS. (2014). *PLOS progress update 2013/2014*. Retrieved from http://www.plos.org/about/plos/progress-update/

Plume, A., & van Weijen, D. (2014, September). Publish or perish? The rise of the fractional author…. *Research Trends*, (38). Retrieved from www.researchtrends.com: http://www.researchtrends.com/issue-38-september-2014/publish-or-perish-the-rise-of-the-fractional-author/

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

Pulverer, B. (2014). The cost of open access. *Wiley Exchanges.* Retrieved from http://exchanges.wiley.com/blog/2014/10/23/the-cost-of-open-access/

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 U.S. *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-

RIN. (2014). *Nature Communications: Citation analysis*. Nature Publishing Group. Retrieved from http://www.nature.com/press_releases/ncomms-report2014.pdf

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. (2011). *Knowledge, networks and nations: Global scientific collaboration in the 21st century*. The Royal Society. http://sro.sussex.ac.uk/id/eprint/45410

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*, 9 July 2012.

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

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

Shotton, D. (2010). CiTO, the citation typing ontology. *Journal of Biomedical Semantics*, *1*(Suppl 1), S6. doi:10.1186/2041-1480-1-S1-S6

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

Simba. (2013a). *Global medical publishing 2013-2014*. Retrieved from http://www.simbainformation.com/Global-Medical-Publishing-7681197/

Simba. (2013b). *Global scientific and technical publishing 2013-2014.* Retrieved from http://www.simbainformation.com/Global-Scientific-Technical-7681199/

Simba. (2014). *New Simba report sizes the market for open access journal publishing.* Retrieved from http://www.simbainformation.com/about/release.asp?id=3618

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

Somerville, M., & Conrad, L. (2014). *Collaborative improvements in the discoverability of scholarly content. A SAGE white paper.* SAGE. Retrieved from http://www.sagepub.com/repository/binaries/pdf/improvementsindiscoverability.pdf

SPARC, & PLOS. (2013). *How open is it?* Retrieved from http://www.plos.org/wp-content/uploads/2013/12/hoii-guide.pdf

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

STM. (2012). Text and data mining: STM statement & sample licence. Retrieved from http://www.stm-assoc.org/text-and-data-mining-stm-statement-sample-licence/

STM. (2013a). A statement of commitment by STM publishers to a roadmap to enable text and data mining (TDM) for non-commercial scientific research in the european union. Retrieved from http://www.stm-assoc.org/2013_11_11_Text_and_Data_Mining_Declaration.pdf

STM. (2013b). *Text and data mining for non-commercial scientific research.* Retrieved from http://www.stm-assoc.org/text-and-data-mining-stm-statement-sample-licence/

STM. (2014). STM response to 'Global coalition of access to research, science and education organisations calls on STM to withdraw new model licenses'. Retrieved from http://www.stm-assoc.org/2014_08_07_STM_response_to_call_for_license_withdrawal.pdf

STM Future Lab Committee. (2014). *STM Future Lab trend watch 2014.* Retrieved from http://www.stm-assoc.org/future-lab-trend-watch/

Strieb, K., & Blixrud, J. (2013). *The state of large-publisher bundles in 2012.* ARL. Retrieved from http://publications.arl.org/rli282/13

Sturges, P., Bamkin, M., Anders, J. H., Hubbard, B., Hussain, A., & Heeley, M. (2014). Research data sharing: Developing a stakeholder-driven model for journal policies. *Journal of the Association for Information Science and Technology.* http://eprints.nottingham.ac.uk/id/eprint/3185

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/

Taylor & Francis. (2014). *Open access survey June 2014.* Retrieved from http://www.tandfonline.com/page/openaccess/opensurvey/2014

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., Mays, R., & Wu, L. (2011). Journal article growth and reading patterns. *New Review of Information Networking*, *16*(1), 4-22. doi:10.1080/13614576.2011.566796

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 Biol*, *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

Thomson Reuters. (2014). *The research & innovation performance of the G20*. Retrieved from http://sciencewatch.com/grr/the-g20-nations

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

Treloar, A. (2014). The Research Data Alliance: Globally co-ordinated action against barriers to data publishing and sharing. *Learned Publishing*, *27*(5), 9-13. doi:10.1087/20140503

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: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 Noorden, R. (2013). Open access: The true cost of science publishing. *Nature News*, *495*(7442), 426. doi:10.1038/495426a

Van Noorden, R. (2014a). The arXiv preprint server hits 1 million articles. *Nature News*. doi:10.1038/nature.2014.16643

Van Noorden, R. (2014b). Google Scholar pioneer on search engine's future. *Nature News*. doi:10.1038/nature.2014.16269

Van Noorden, R. (2014c). Online collaboration: Scientists and the social network. *Nature News, 512*(7513), 126. doi:10.1038/512126a

Van Noorden, R. (2014d). The scientists who get credit for peer review. *Nature News.* doi: 10.1038/nature.2014.16102

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

van Rossum, J. (2014). How we can better support and recognize reviewers. *Elsevier Editor's Update.* Retrieved from http://editorsupdate.elsevier.com/issue-42-march-2014/can-better-support-recognize-reviewers/

Verstak, A., Acharya, A., Suzuki, H., Henderson, S., Iakhiaev, M., Lin, C. C. Y., & Shetty, N. (2014). On the shoulders of giants: The growing impact of older articles. *ArXiv Preprint ArXiv:1411.0275.* doi:arXiv:1411.0275

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

Waldrop, M. M. (2008). Science 2.0. *Scientific American, 298*(5), 68-73. doi:10.1038/scientificamerican0508-68

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

Weckowska, D. (2014). *Open access publishing and innovation*

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

Wolters Kluwer Health. (2013). *2013 physician outlook survey.* Retrieved from http://www.wolterskluwerhealth.com/News/Documents/White%20Papers/Wolters%20Kluwer%20Health%20Physician%20Study%20Executive%20Summary.pdf

Wolters Kluwer Health. (2014). *Mobile device internet and social media use and habits survey report.* Retrieved from http://www.wolterskluwerhealth.com/News/Pages/Wolters-Kluwer-Health-Survey-Finds-Nurses-and-Healthcare-Institutions-Accepting-Professional-Use-of-Online-Reference--Mobi.aspx