STM

The global voice of scholarly publishing

The STM Report

An overview of scientific and scholarly publishing

1968-2018

Celebrating the 50th Anniversary of STM

Rob Johnson
Research Consulting

Anthony Watkinson
CIBER Research

Michael Mabe
The International Association of Scientific, Technical and Medical Publishers

Fifth edition, October 2018

© 2018 STM: International Association of Scientific, Technical and Medical Publishers

Fifth Edition published October 2018

Published by:

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

About STM

STM is the leading global trade association for academic and professional publishers. It has over 150 members in 21 countries who each year collectively publish over 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.

Research Consulting works with publishers, research organisations, funders and policymakers to improve the effectiveness and impact of research and scholarly communication. For more information see www.research-consulting.com.

CIBER Research is an international academic research group founded in 2002 at City University London and subsequently a research arm of the University College London Department of Information Studies. The CIBER team offer both consultancy and research. For further information see www.ciber-research.eu.

Acknowledgements

The STM Report, now in its fifth edition after 12 years, has grown out of all proportion to its first incarnation. This is reflected not just in the growth in the number of pages but also in there now being three main authors, and the breadth of the issues we now cover. This range of content far exceeds the expertise of any one of us. We want to place on record our warm appreciation for the input and advice of the following individuals whose insights have contributed to the finished text:

IJsbrand Jan Aalbersberg, Bev Acreman, Rick Anderson, Eric Archambault, Gaelle Bequet, Peter Berkery, Helen Blanchett, Cherifa Boukacem-Zeghmouri, Sam Burrell, Andrea Chiarelli, Lisa Christian, Barry Clarke, Laura Cox, Tom Cullen , Jeremy Dain, Craig Van Dyck, Michael Evans, Paul Evans, Martin Eve, Victoria Ficarra, Mattia Fosci, Elizabeth Gadd, Richard Gedye, Wouter Haak, Glenn Hampson, James Hardcastle, Eti Herman, Iain Hrynaszkiewicz, Michael Jubb, Rachael Lammy, Carlo Scollo Lavizzari, Adriaan Lemmen, Matt McKay, Jo McShea, Natasha Mellins-Cohen, Ann Michael, Sheila Morrissey, Susan Patton, Dan Penny, Ed Pentz, Theo Pillay, Stephen Pinfield, Andrew Plume, Dan Pollock, Ian Potter, Andrew Preston, Vanessa Proudman, Charlie Rapple, Peter Richardson, Henning Schoenenberger, Eefke Smit, Susan Spilka, Heather Staines, Adrian Stanley, Carol Tenopir, Lennart Velten, Charles Watkinson, Kate Wittenberg, Jie Xu.

As always, any errors or misunderstandings are ours alone.

RJ, AW, MM September 2018

Contents

Executive summary

Scholarly communication and STM publishing

1. STM publishing takes place within the broader system of scholarly communication, which includes both formal and informal elements. Scholarly communication plays different roles at different stages of the research cycle, and (like publishing) is undergoing technology-driven change. Categorising the modes of communication into one-to-one, one-to-many and many-to-many, and then into oral and written, provides a helpful framework for analysing the potential impacts of technology on scholarly communication (see page 11).

2. Journals form a core part of the process of scholarly communication and are an integral part of scientific research itself. Journals do not just disseminate information, they also provide a mechanism for the registration of the author's precedence; maintain quality through peer review and provide a fixed archival version for future reference. They also provide an important way for scientists to navigate the ever-increasing volume of published material (page 14).

The STM market

3. The annual revenues generated from English-language STM journal publishing are estimated at about $10 billion in 2017, within a broader STM information publishing market worth some $25.7 billion. About 41% of global STM revenues (including non-journal STM products) come from the USA, 27% from Europe/Middle East, 26% from Asia/Pacific and 6% from the rest of the world (page 22).

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 46).

5. Although this report focuses primarily on journals, the STM book market (worth about $3.3 billion annually) is evolving rapidly in a transition to digital publishing. Ebooks made up about a third of the market in 2016, having grown much faster than the STM market as a whole in recent years (page 22).

6. There are estimated to be of the order of 10,000 journal publishers globally, of which around 5,000 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 2,300 journals (20%) are not-for-profit (page 40).

7. There were about 33,100 active scholarly peer-reviewed English-language journals in mid-2018 (plus a further 9,400 non-English-language journals), collectively publishing over 3 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. However, growth has accelerated to 4% per year for articles and over 5% for journals in recent years. The reason is the continued real-terms growth in research and development expenditure, and the rising number of researchers, which now stands at between 7 and 8 million, depending on definition, although only about 20% of these are repeat authors (page 25).

8. China has overtaken the US to become the pre-eminent producer of global research papers globally, with a share of about 19%, and on current trends its research spending will also exceed the US's by the early 2020s. The US accounts for 18% of global articles, while India has also seen rapid growth in recent years, and now

produces 5% of global outputs, ahead of Germany, the UK and Japan, each on 4% (page 29).

Business models and publishing costs

9. Aggregation on both the supply and demand sides remains the norm, with journals sold in packages to library consortia (see below for open access). Similar models have also become established for ebook collections (page 18).
10. 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. Recent years have seen some high-profile European negotiations stall, and a small minority of institutions elect to return to title-by-title purchasing. Nevertheless, the Big Deal's 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 107).
11. 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 some way over 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 91).
12. 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 92).
13. 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 94).
14. 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 (c. $4,000), excluding non-cash peer review costs. An updated analysis by CEPA in 2018 shows that, in almost all cases, intangible costs such as editorial activities are much higher than tangible ones, such as production, sales and distribution, and are key drivers in per article costs (page 73).
15. The potential for technology and open access to effect cost savings has been much discussed, with open access publishers such as Hindawi and PeerJ having claimed per article costs in the low hundreds of dollars. A recent rise in PLOS's per article costs, to $1,500 (inferred from its financial statements), and costs of over £3,000 ($4,000) per article at the selective OA journal eLife call into question the scope for OA to deliver radical cost savings. Nevertheless, with article volumes rising at 4% per

annum, and journal revenues at only 2%, further downward pressure on per article costs is inevitable (page 74).

Research behaviour and motivation

16. 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 77).
17. 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. This has temporarily unbalanced the sources of articles and reviewers, with a third of all reviews but only a quarter of articles provided by researchers in the USA (page 47).
18. There is a significant amount of innovation in peer review, with the more evolutionary approaches gaining more support than the more radical. 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 gaining support from publishers and funders, although there is evidence they can reduce reviewer acceptance rates. Cascade review (transferring articles between journals with reviewer reports) has gained a foothold, and the "soundness not significance" peer review criterion adopted by open access "megajournals" like *PLOS ONE* is now well-established. Journal-independent ("portable") peer review has not taken hold in earnest, and post-publication review has continued to receive limited support, as evidenced by the recent termination of PubMed Commons.
19. Reading patterns appear to have stabilised following a decades-long increase in the number of articles read per researcher. Researchers are averaging 250 articles per year, depending on discipline (more in medicine and science, fewer in humanities and social sciences), with early indications that the total may even be falling. The decline in time spent per article, down from 45-50 minutes in the mid-1990s to just over 30 minutes in 2012, may also be reversing. Access and navigation to articles is increasingly driven by search rather than browsing, but researchers continue to use a multiplicity of routes to find content. Social media has become significantly more important in all subject areas, but in the sciences usage for this purpose appears to have peaked. Researchers spend very little time on average on publisher web sites, "bouncing" in and out and collecting what they need for later reference (page 57).
20. The deficiencies of the Journal Impact Factor continue to be much discussed, but the growing range of new and alternative metrics have yet to supplant it in the eyes of the research community. There is however growing interest in tracking and demonstrating the broader economic and societal impact of research, underpinned by rising expectations in this regard from funding bodies (page 64).
21. 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 80).

Open access

22. Journal publishing has become more diverse, and potentially more competitive, with a range of new business models now firmly established within the marketplace. Open

access makes original research freely accessible on the web, free of most copyright and licensing restrictions on reuse. There are three main approaches: open access publishing (“Gold”, including full and hybrid OA journals), delayed free access, and self-archiving (“Green”) (page 97).

23. There are around 11,811 (9,172 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 3 million or so articles published per year that is open access is substantially lower than the proportion of journal titles (page 133).

24. Approximately one third of the scholarly literature was available from legal and sustainable open access sources in 2016. Recent estimates place the proportion of articles published in open access journals at 15-20% (while OA journals make up about 26-29% of all journals), with a further 10-15% available via delayed access on the publisher’s website or self-archived copies (page 134).

25. 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, read-and-publish deals, and so on (page 97).

26. Gold open access based on APCs has a number of potential advantages, and has found significant support in some quarters. It would scale with the growth in research outputs, there are potential system-wide savings, and reuse is simplified. Research funders in some jurisdictions will reimburse publication charges, but even with broad funder support the details regarding the funding arrangements within universities and in other regions 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; and charges for full open access articles remain lower than hybrid, though the gap is closing. Calls to redirect subscription expenditures to open access have increased, but the more research-intensive universities and countries remain concerned about the net impact on their budgets (page 101; 139).

27. Open access publishing 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 articles published in megajournals continues to grow, but at a slower rate than previously, while Scientific Reports has usurped PLOS ONE as the leading example in recent years (page 111).

28. Research funders are playing an ever more 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). Recent developments indicate a growing willingness on the part of funders and policymakers to intervene in the STM marketplace, whether by establishing their own publication platforms, strengthening OA mandates or acting to change the incentive structures that drive authors’ publication choices (page 113).

29. Concerns over the impact of Green OA and the role of repositories have receded somewhat, though not disappeared. 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, with indications that Green OA can increase downloads and citations being balanced against evidence of the long usage half-life of journal articles and its substantial variation between fields. In practice, however, attention in many quarters has shifted to the potentially damaging impact of Social Collaboration Networks (SCNs) and pirate websites on subscriptions (pages 114; 174).

Technology

30. Virtually all STM journals are now available online, and consequently the vast majority of journal use takes place electronically, at least for research journals, with print editions providing parallel access for some general journals, including society membership journals, and in some fields (e.g. humanities and some practitioner fields). The number of established research (i.e. non-practitioner) journals dropping their print editions entirely has accelerated in recent years, with others switching to print-on-demand (page 28).

31. Open Science, sometimes called Open Scholarship to bring in the humanities, has supplanted and subsumed Open Access as the overall principle behind the transformation of scholarly communication. There is no satisfactory generally accepted definition because there is little consensus on who the stakeholders are. OSI, funded by UNESCO, includes publishers as stakeholders but some library-based initiatives do not. All definitions emphasise that openness throughout the research cycle is what makes Open Science more than OA (page 148).

32. 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 use of Twitter in particular is common, though for scientists in general email remains the main mode of communication. Among Social Communication Networks (SCNs) ResearchGate is the site of choice but researchers use it more for profiling than they do for communication. The publishing community has had some success in persuading ResearchGate not to make available full text when rights of general usage are not available, but much of its content remains illegally posted and hosted (page 174).

33. The rapid general adoption of mobile devices (smartphones and tablets) has yet to change significantly the way most researchers interact with journal content – accesses from mobile devices still accounted 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. Most major publishers now make most, if not all, of their journals optimised for mobile use (page 175).

34. 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 leadership and coordination from the Research Data Alliance: 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 the Scholix initiative has been launched to establish systematically what data underpins literature and what literature references data. Meanwhile publishers have responded by working closely with many of the community-led projects; by developing standardised 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 Clarivate Analytics' Data Citation Index (page 153).

35. Text and data mining are continuing to emerge from niche use in the life sciences industry, with the potential to transform the way scientists use the literature. While recent discussions at European-level have focused on copyright and licensing regimes, uptake remains constrained by a range of challenges, including content aggregation, the costs and technical skills requirements for mining and limited incentives for researchers. A number of initiatives (e.g. CrossRef's TDM tools, PLSclear, and Copyright Clearance Center'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), content access and aggregation for TDM, but this area remains immature (page 83; 178).

36. 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). A range of new web-scale discovery and analytics tools have emerged in recent years, with the potential to enhance our understanding of the relationships between research inputs, activities and outputs (page 173).

37. The growth in the development of Artificial Intelligence and its implementation across industry as a whole has impacted generally on the publishing industry, partly because it can enable cost efficiencies. AI has subsumed much of semantic technology as a guiding principle. Blockchain, although much talked about, has yet to prove its usefulness in practice. 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. Online Collaborative Writing remains a service which has yet to have its day, though a recent emphasis on Annotations is showing promise (page 163).

38. Perhaps the biggest change in scholarly infrastructure has been the development of preprint servers and the growing use of preprints in areas such as biology and chemistry where there had hitherto been little appetite for their take up. Primary journals in general have now accepted that a preprint is not prior publication. There is some concern that preprints (which can be brought up to date) may become a go-to place for the version of record, undermining publisher business models (page 179).

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, as well as data, software and other digital objects). The scholarly communication supply chain has traditionally comprised 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 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 come to play an important and probably growing role. Scholarly communications is evolving, however, and research funders have increasingly assumed a role as one of the most important parts of the system with the growth of open science and related developments. Meanwhile, other players are becoming increasingly active in the distribution and discovery of content (notably data repositories, and a growing array of software and services providers).

1.1 The research lifecycle

The different roles played by scholarly communication can be understood in the context of the research lifecycle (with the communication role in parentheses) (see Figure 1):

- Idea discovery, identification of partners (awareness, literature review, informal)
- Proposal writing/approval (literature review)
- Research process (awareness)
- Publication (formal publication, informal dissemination)

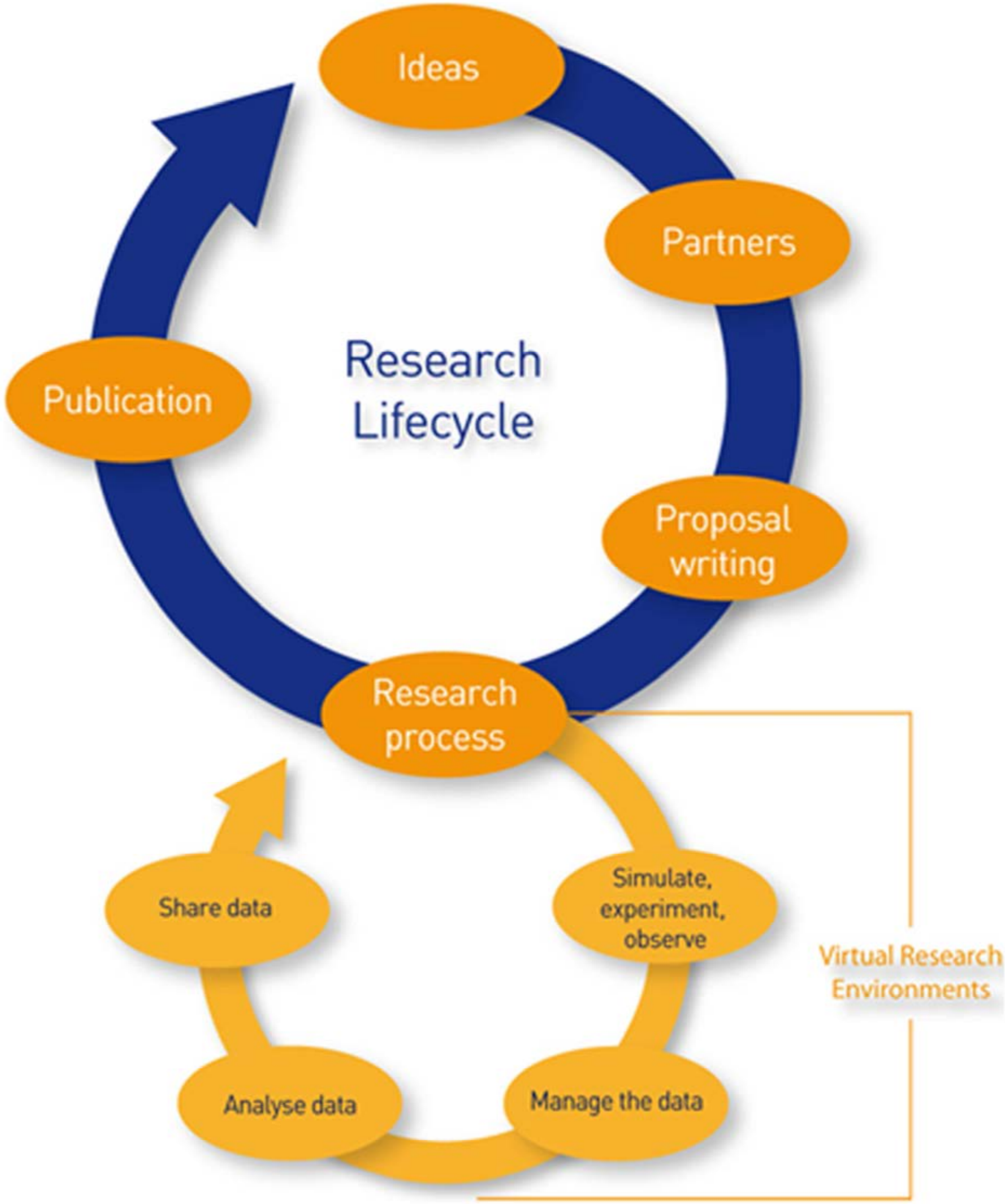
1.2 Types of scholarly communication

Scholarly communication has traditionally encompassed activities including conference presentations, informal seminar discussions, face-to-face or telephone conversations, formal journal and book publications, and grey literature. Over time these forms have been progressively supplemented by new forms of communication, such as email exchanges, email listservs, preprints and, increasingly, social media as well as digital objects.

One way of categorising scholarly communication is in terms of whether it is public or private, and whether it is evaluated or non-evaluated. In this report we are primarily concerned with formal, written communication in the form of journal articles. The boundary between formal and informal communications is blurring in some areas. For instance, unrefereed author's original manuscripts on the arXiv repository are increasingly cited in formal publications and conference proceedings can be of high academic standing in some disciplines, while journal articles are becoming more informal and blog-like with addition of reader comments. Nevertheless, if anything the central role of the journal article in scholarly communication is stronger than ever.

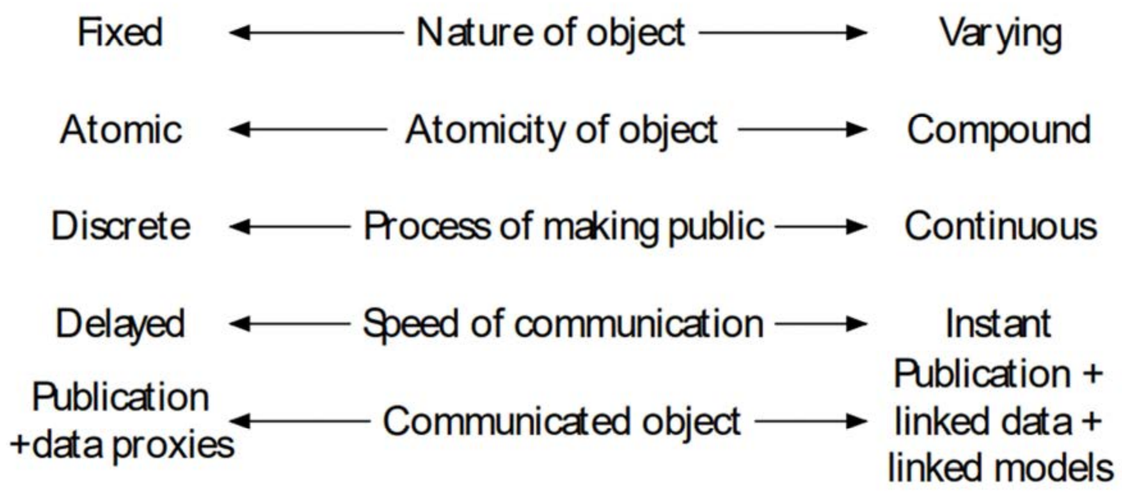
[1] "STM" is an abbreviation for scientific, technical and medical but has several different meanings. It can be a model of publishing, in which case it includes social sciences and the arts and humanities. It is sometimes used to describe scientific journals. It is also the name of the association of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

Figure 1: The research cycle (Tenopir et al. 2011b)



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). The evolving area of information ecology enables these to be further categorised into oral and written communications and other dimensions, such as whether it is live or recorded, immediate or at a distance, or allows for enhancement through technology. By considering types of scholarly communication in this way, 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 yet offered wholly new modes. There are new tools but they are for the same old purposes. Figure 2 shows another aspect of the digital change.

Figure 2: Characterising the future - changing models of research communication (source: Treloar 2014a)



1.3 Changes in the 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.

This type of journal publishing, involving the peer reviewed first reports of phenomena or ideas, is known as "primary publishing" and its outlets "primary journals". Prior to the development of electronic databases and the internet, another type of journal flourished which acted as a navigation device to the primary literature. These so-called abstracting or "secondary journals" arose early in the history of primary publishing and existed right up until the electronic revolution, with about one secondary journal for every 300 primary ones (Price 1963). Following the growth of electronic abstracting and indexing services, the secondary literature has transformed into the myriad indexing services, while they in their turn are facing tough competition from generic internet-based search engines. The third strand of the evolution of the journal publishing world also began quite early with the development of review journals, where experts survey the literature of a field over some years, indicating how current opinion has come to be and the differing strands of research that preceded it or still make it up. This final area of journal publishing is called "tertiary publishing" and has much in common with the monographic book.

The primary journal has traditionally been seen to embody four functions (Zuckerman and Merton 1971, Mabe 20012):

- *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 preprint servers and portable peer review services – 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 – sometimes 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 or 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*).

2.2.1 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:

- 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.

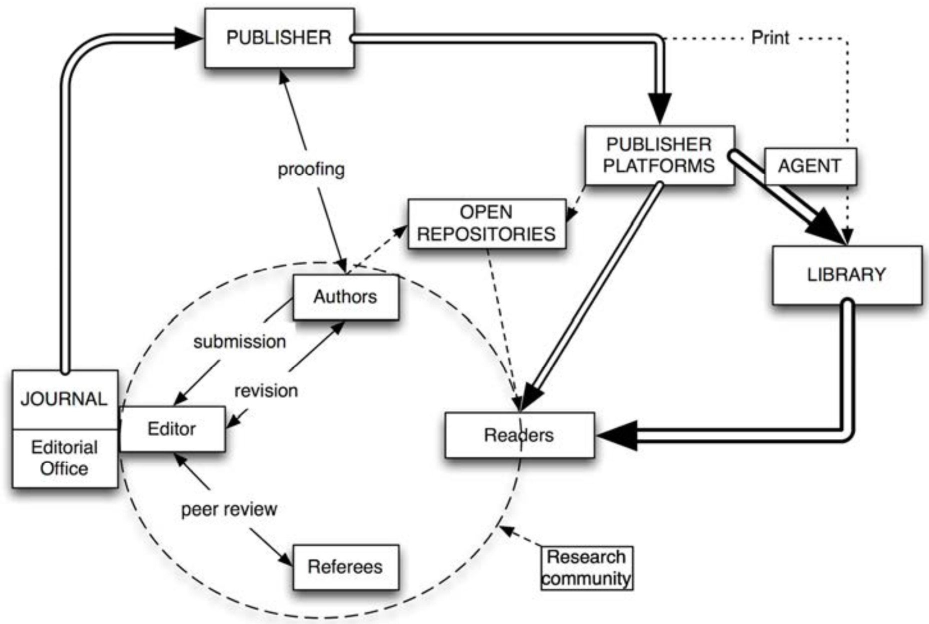
- Aggregation of content: while other players (e.g. Google, PubMed) are also involved, publishers' aggregation services currently offer widely-used services.
- Distillation of evidence: e.g. reference works and meta-reviews.
- Creating standards and consensus seeking: a large number of publisher-led initiatives improve the quality, findability and usability of STM content including ORCID and Crossref services such as Crossref Similarity Check, Crossmark, and the Crossref Funder Registry.
- 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 originally provided in the blog post *82 Things Publishers Do (2014 Edition)* (Anderson 2014a), now updated to *Focusing on Value – 102 Things Publishers Do (2018 Update)* (Anderson 2018). This is essentially a more granular breakdown of these same functions, but Anderson also emphasises the need for a long-term sustainable model, which in turn 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).

The recent moves of larger publishers into information and data analytics is covered by Inchcoombe (2017).

Figure 3: The publishing cycle



2.2.2 Versions of articles

One potential issue with the widespread adoption of self-archiving (not to mention the growth of preprint servers) is that multiple versions of articles are increasingly 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 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 European Commission 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 to the author's original (and sometimes to the accepted manuscript), and post-print to refer to the accepted manuscript. These terms 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.

A notable trend is the emergence of journals adopting a more fluid notion of the journal article. For instance, the platform*Faculty1000Research* encourages authors to publish (multiple) revised versions of their article, with all versions of an article linked and independently citable. This smearing out of the published—unpublished boundary is creating a number of challenges. In 2018 there are calls for a revision of the convention about what constitutes the first published version, with online papers often preceding the official issue dates by months if not years (Keller and Prusiner 2018).

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 has long been a declining part of the market, especially for larger publishers, with the vast majority of journals 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 them 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. Far 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 have tended to be more likely (91%) to offer perpetual access than small publishers (50%) (Inger & Gardner, 2013).

2.3.1 Subscription agents

Subscription agents were an important part of the sales channel: the average library used to place about 80% of its business via agents. Agents acted 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 provided a valuable service to publishers by aggregating library orders and converting them to machine-readable data, handling routine renewals, and so on. The Association of Subscription Agents ceased operation in July 2015. A key reason was the increasing disintermediation of the traditional agent function brought about by the move to electronic publishing and in particular the rise of consortia sales. The larger remaining subscription agents (such as EBSCO and Harrassowitz) 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).

2.3.2 Content bundles and pricing

With the rise of electronic publishing, sales of individual journal subscriptions fell as a proportion of total sales in favour of **bundles**. A decade ago, Cox & Cox (2008) found that nearly all (95%) of large and most (75%) of medium publishers offered bundles of content, though this dropped (for obvious reasons) to 40% of small publishers. Publishers have increasingly offered bundles that include non-journal content, particularly ebooks, reference works and datasets. Small publishers are more likely to participate in multi-publisher bundles such as Aggregagent, 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).

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.

In practice, most large publishers continue to focus on selling some form of bundled e-package content and librarians are still closely evaluating content use, value and quality compared to the price (EBSCO 2018). Despite the apparent stasis in pricing models,[2] industry discussions suggest that there will be more publishers moving away from historic pricing in the coming years. The main drag on this move has been the natural reluctance for anyone to be a loser in a transition.

A key issue for libraries is whether the publisher's licence term for bundles allows cancellations; Cox and 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%).

[2] 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.

Nevertheless there has in recent years been considerable resistance to the big deal, with many libraries deciding to cancel them in favour of a return to title by title purchase. There is no comprehensive data on such cancellations, but an informal survey of North American libraries (Anderson, 2017) and data maintained by SPARC both point to an increase since 2015.[3] An additional concern for universities and libraries has been how best to support open access, particularly in an environment where both licensing and article publication charges for open access have occurred. See *Offsetting.*

This environment has also resulted in an increased desire for national deals that take the open access payments into account. This has not been without some heat, with a standoff on the German deal over the last few years and at time of writing issues arising in France and Sweden, too. [4]

2.3.3 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 increase the purchasing power of libraries in negotiation with publishers, and increasingly to take advantage of bundled electronic content. The numbers of consortia have been growing strongly: the Ringgold Consortia Directory Online[5] lists over 500 consortia in 126 countries, representing over 32,000 individual institutions[6]; of these, about 350 are responsible for licensing content. The International Coalition of Library Consortia[7] has some 200 members. The size and nature of consortia vary considerably, from national consortia to small regional ones, and include academic, medical, public, school and government libraries. The 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 marketed to consortia, and about half of all publishers. Of these, about half used the same pricing model as for their bundles, with the balance negotiating on a case-by-case basis. Consortia deals were 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 were as previously covered for bundles. The early part of the current decade saw a growing focus on "author-rights" clauses (typically covering self-archiving rights for authors at the licensing institutions) and non-disclosure agreements. More recently, there has been a push for the inclusion of "text and data-mining" clauses in model agreements, and an emphasis (particularly in Europe) on linking licensing negotiations with open access and promoting transparency.[8]

2.3.4 Library system vendors

Library system vendors[9] 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 the chain

[3] See https://sparcopen.org/our-work/big-deal-cancellation-tracking/

[4] See https://www.timeshighereducation.com/news/sweden-cancels-elsevier-contract-open-access-dispute-spreads for situation as of May 2018

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

[6] 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

[7] http://icolc.net/

[8] https://libereurope.eu/blog/2017/09/07/open-access-five-principles-for-negotiations-with-publishers/

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

that links readers to publishers' content. Publishers work with systems vendors on supply-chain standards such as ONIX for Serials[10] and KBART (Knowledge Bases And Related Tools).[11] Uniquely identifying institutions is important for publishers: the Identify service from Ringgold[12] is the leading commercial service here, with a database of some 500,000 institutions and consortia. A free dataset, OrgRef, was launched by DataSalon in 2014, but has been superseded by other developments, notably the Organization ID Registry Initiative,[13] and is no longer actively maintained. Formal plans for a Research Organization Registry (ROR) were due to be announced in late 2018. Other services include Digital Science's GRID,[14] while ISNI has recently updated its Organizations Registry to enable organizations to change and correct their own records and allow the research community to identify author affiliations persistently and authoritatively.[15]

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 (now owned by ProQuest), 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 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. Tests conducted by libraries have typically 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). Three-quarters of US library directors continue to see an index-based discovery service as a highly important priority for their library, according to a survey by Ithaka S + R. Nevertheless, there are indications that libraries are reluctant to invest additional resources in these services, and are increasingly comfortable with scholars beginning their research process outside of the library (Ithaka S+R, 2017).

Discovery also remains problematic, with an estimated five to 15 clicks required for most researchers to find and access an article.[16] Recent years have seen the emergence of a new breed of open access discovery tools, including Unpaywall, OA Button, Lazy Scholar and Kopernio, all of which offer browser extensions, and free-to-use aggregators such as Bielefeld Academic Search Engine (BASE) and Core. Increasingly, OA discovery services are being integrated into library discovery tools and link resolvers, as well as traditional A&I databases. Kopernio was acquired by Clarivate Analytics, owner of Web of Science, in April 2018,[17] and recently announced a collaboration with the California Institute of Technology (Caltech), with a view to offering one-click access to both subscription and open-access content.[18] Meanwhile, 1Science's 1findr service aims to combine web-scale discovery and analytics functions in a single product, and Digital Science launched Anywhere Access in mid-2018 as a managed, cloud-based and COUNTER-compliant discovery solution.

[10] http://www.editeur.org/8/ONIX/
[11] http://www.uksg.org/kbart
[12] http://www.openrfp.com/pages/identify.html
[13] https://orcid.org/content/organization-identifier-working-group
[14] https://www.grid.ac/
[15] http://www.isni.org/content/isni-organizations-registry-identifying-organizations-scholarly-supply-chain
[16] See https://www.thebookseller.com/news/new-venture-mendeley-co-founder-jan-reichelt-680621
[17] See https://www.nature.com/articles/d41586-018-04414-8
[18] https://clarivate.com/blog/news/clarivate-analytics-caltech-collaborate-offer-researchers-one-click-access-published-research/

All these services, find themselves in competition with Google Scholar (and, to a much lesser extent, Microsoft Academic Search), which offers integration with library holdings, citation links via Scholar Metrics, and other features in addition to its signature search capabilities (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".

2.4 Journal economics and market size

2.4.1 Journal economics & market size

The total size of the global STM market in 2017 (including journals, books, technical information and standards, databases and tools, and medical communications and some related areas) was estimated by Outsell (2018a) at $25.7 billion.[19] The 2017 market can also be divided into scientific/technical information and solutions at $13.8 billion and medical at $11.9 billion. The medical information market is predicted to grow at 4.6% annually through 2021, with scientific and technical growing at a slower rate of 2.9%.

Within this overall market for STM information, Outsell estimate 2017 revenues from journals at $9.9 billion (38.5% of the total), and from books at $3.2 billion (12%) (Outsell 2018a).[20]

By geographical market, Outsell estimate about 41% of global STM revenues (including non-journal STM products) come from the USA, 27% from the EMEA region, 26% from Asia/Pacific and 6% from the rest of the world (principally the Americas excluding USA) (Outsell 2018a). These proportions probably overstate the importance of the US market for journals alone. Revenues from the EMEA region have remained relatively stable as a proportion of the total in recent years, while the share of revenues from Asia/Pacific have increased substantially, largely at the expense of the US. This primarily reflects China's rapid development, overtaking the US to become the world's largest producer of research papers (NSB 2018). Nevertheless, the very low pricing that some publishers adopted to enter the Chinese market in the early days continues to depress revenues from that part of the world.

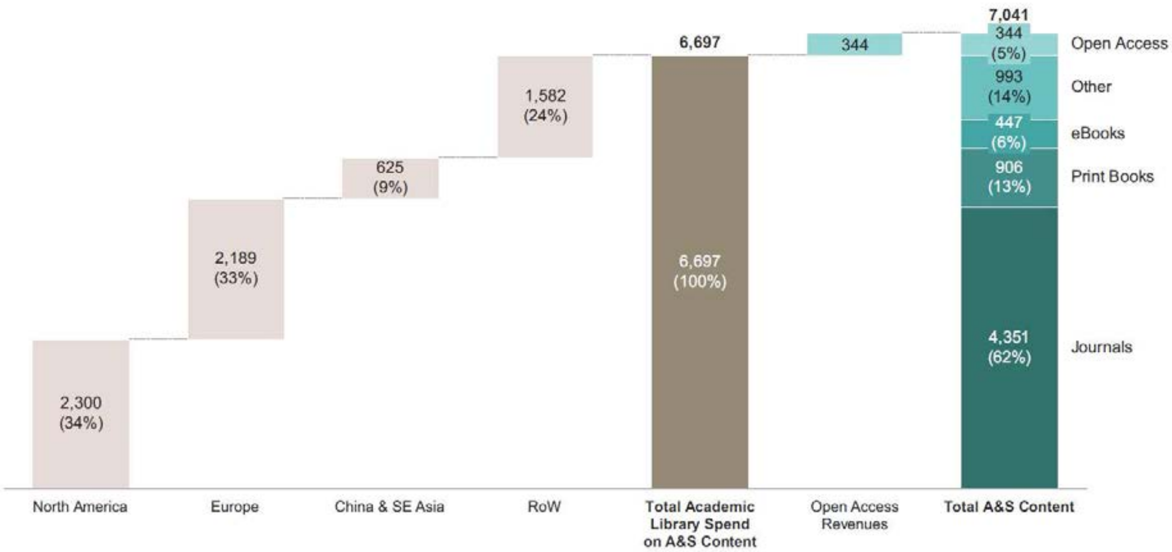
Market analysts Simba have estimated the scientific and technical publishing and medical publishing markets at $9.9 billion each, equating to a total STM market of $19.8 billion. Their definition of the market is slightly narrower than Outsell's and is based on five content delivery channels: journals, books, online content, abstracting and indexing and other activities (Simba 2017a and 2017b).

Meanwhile, OC&C have estimated global spending on academic and scientific content by academic libraries alone at just over €7 billion (c.$8.1 billion) in 2016 (cited in Springer Nature 2018). Academic libraries have traditionally been the primary source of journal revenues, estimated at 68-75% of the total. Other revenue sources include 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 fallen steadily since these estimates were made, with membership fees and personal subscriptions also likely to be in decline.

Figure 4: Global spending on academic and scientific content in 2016 by region and product (in € million, of total) (Source: OC&C, cited in SpringerNature 2018)

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

[20] Our previous report quoted a figure of $10 billion for the journals market in 2013, which now appears to be a slight over-estimate. Outsell's version of record for 2013 sizes the journals market at $9.1 billion, indicating that the market has grown at an average of 2.1% per year since that time.



The open access segment of the market continues to grow much faster than the market as a whole, but remains small in revenue terms. OC&C estimated global open access revenues at $344m in 2016 (Figure 4), up from an estimated $128 million in 2013 (Outsell 2013). Delta Think (Carlin 2017) put a higher figure of $470m on the OA market for 2016, estimating that OA represented 20-22% of market volume and 5-9% of market value. Most analysts expect OA's double-digit growth rates in recent years to continue into the future, outstripping the more modest growth rates expected for the journals market as a whole. In practice, the growth of offsetting arrangements and read-and-publish deals are likely to make it increasingly difficult to separate OA and subscription revenues in the coming years (see *Section 3.2.3* Gold - Hybrid).

2.4.2 Books and ebooks

Outsell (2018a) estimated the market for scientific and technical books at $719m in 2017, and for medical books at $2,486m. OC&C report that combined spending on books and eBooks fell by 1.2% in the period from 2011 to 2016 (cited in SpringerNature 2018), though Outsell's figures suggest there has been a return to growth since 2015. OC&C also report a shift in academic library spending from print books (which have seen a decline at -4.1% from 2013 to 2016) to eBooks (CAGR of 6% from 2013 to 2016). OC&C projects this trend to continue, with eBooks expenditure expected to grow at a CAGR of 4.1-4.6% from 2016 to 2020.

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. However for publishers the library market remains crucial, as it is for journals. Academic Library Book Purchasing Trends (ProQuest 2016) provides a bottom line that is not unexpected:

- Most library budgets have been flat or have had small changes only.
- Monograph budgets are eroding as funds shift to finance electronic journals due to price increases.
- Ebooks account for a growing proportion of book budgets as libraries move from print to electronic collections.

The business models adopted by publishers and by aggregators licensed to sell publisher content in digital form have long been a source of frustration among the library community, especially as publishers usually sell without digital rights management (DRM) and aggregators whose bundles cut across brands, and are thus more useful, are only allowed to sell with DRM. The Charlotte Initiative in its deliberations and documentation provides the arguments involved.[21] One basic assumption is that ebooks are likely to become gradually more important both for libraries and for end users. Patron Driven Acquisition (PDA), also called Demand Driven Acquisition (DDA), is increasingly favoured, with librarians concerned to avoid another "Big Deal", where most of the usage goes to only a small number of the titles licensed (Unsworth 2017). However, book publishers confronted with the budgetary situation as described above are keen to move ebook packages into the library's "continuation" budget out of their "one off" book budgets.

There are also some movements to bring digital books and journals together. Book chapters are now usually given a DOI and can be bought separately, though to date there is not much evidence of significant interest. Many major publishers now routinely make their STM books available online in subject collections, while others go further. Wiley Online covers both books and journals, for example, and has subject-based offerings. Cambridge University launched Cambridge Core in 2018, bringing book and journal content together for the first time in what has been described as a sophisticated, high-performance replacement for Cambridge Journals Online and Cambridge Books Online. Its Cambridge Elements product type goes still further, offering a book-journal hybrid intended to serve as a dynamic reference source.[22]

The open access market for scholarly books has been slow to grow, but is potentially significant in some fields – such as the humanities – where the monograph and other scholarly books remain important research outputs. Key initiatives and developments are discussed in the Section 3.2.6 *Open Access Books.*

2.4.3 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.[23]

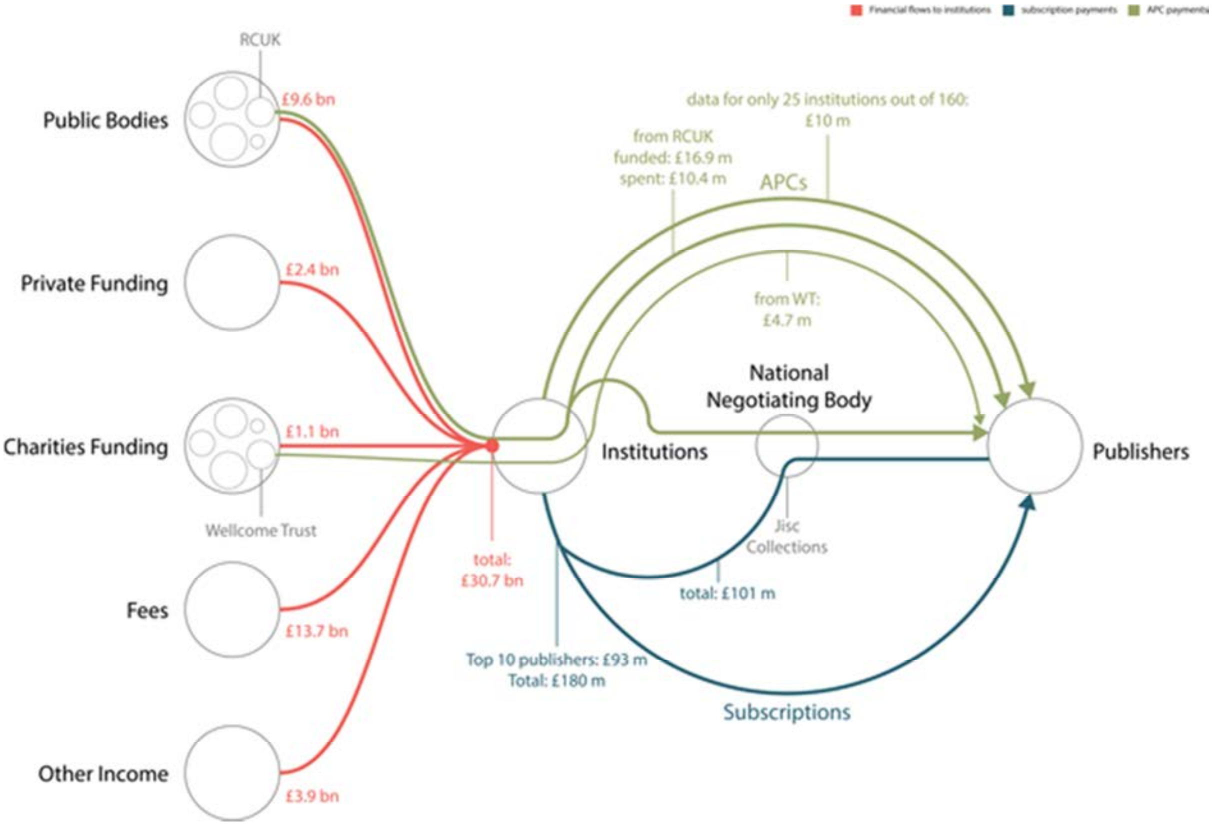
The £25 billion for publication included publishing and library costs; the publishing costs total £6.4 billion: of this, £3.7 billion is fixed first copy costs, including £1.9 billion in non-cash costs for peer review and £2.7 billion is variable and indirect costs, including publishers' surplus. Excluding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

More recently, Lawson et al (2016) used a study of journal publication in the UK as the basis for a provisional visual model for analysing financial transparency around scholarly communication (Figure 5). The three main flows outlined are various financial flows to institutions (orange), and then two flows from institutions to publishers: subscription payments (blue) and APC payments (green). They also include another actor – the national negotiating body – and note that the decisions of individual researchers are almost entirely absent from the model; control over the flows is largely at the institutional or funder level, undertaken on behalf of the research community.

The authors acknowledge that some elements are missing from the picture, such as licensing revenues, page and colour charges and subscriptions from non-academic sources. They argue that a joined-up, systemic, publicly accessible picture of global financial flows around academic publishing would be of significant value to inform evidence-based deliberation, policy and action to shape scholarly communication systems.

21 http://charlotteinitiative.uncc.edu/
22 See http://www.cambridge.org/gb/academic/elements
23 Values from 2008; not inflated to current values.

Figure 5: Model of Financial Flows in Scholarly Publishing for the UK, 2014 (Source: Lawson et al, 2016)



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

2.5 Journal and article numbers and trends

There were about 33,100[24] active scholarly peer-reviewed English-language journals in 2018, collectively publishing some 3 million articles a year. Figure 6 shows the growth in the number of active, peer-reviewed journals recorded in Ulrich's directory between 2000 and 2013; over this period the rate of growth has increased from 3% a year in the early part of the century to 5-6% in the recent past.[25] At the time of writing, the CrossRef database includes over 97 million DOIs, of which 73 million refer to journal articles from a total of almost 60,000 journals. More broadly, Google Scholar was estimated to index between 100 and 160 million documents in 2014, including journal articles, books, and grey literature (Khabsa & Giles, 2014; Orduña-Malea et al., 2014), with volumes undoubtedly having risen further since that date. Meanwhile, the Web of Science 'Core Collection' included about 70 million article records as of June 2018, out of a total of 150 million items across all WoS databases.

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

[24] Ulrich's Web Directory listed 33,119 active scholarly peer-reviewed English-language journals as at August 2018. The count increases to 42,491 if non-English-language journals are included.

[25] Indexing delays mean data from more recent years tends to be unreliable and subject to change, so is excluded from this analysis.

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 11,655 journals from over 2,500 publishers included in Clarivate Analytics' (2018) *Journal Citation Reports* (JCR): these collectively publish 2.2 million articles, reviews, and other source items 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 Clarivate JCR are also on average substantially larger than those not included (with an average of 161 articles for journals in the Science Citation Index Expanded (SCIE) and 51 for the Social Sciences Citation Index (SSCI)). A further 7,500 journals are included in Clarivate's Emerging Sources Citation Index (ESCI), launched in late 2015.

Other abstracting and indexing services are intentionally broader in scope and include:

- Elsevier's Scopus, which covers 22,000 peer-reviewed journals from about 5,000 publishers. It contains over 69 million core records, from which are derived 1.4 billion cited references. Approximately 3 million new items are added each year.
- 1science's 1findr, which lists 87,000 academic/scientific journals and 90 million peer-reviewed articles. [26]
- Digital Science's Dimensions, which indexes over 90 million scholarly documents, of which about 85% are journal articles.[27]
- Informa's wizdom.ai, which includes 73,000 journals, and over 90 million publications.[28]

The number of peer reviewed journals published annually and still active had been growing at a steady rate of about 3.5% per year for over three centuries (see Figure 7), although the growth did slightly accelerate in the post-war period 1944–78. The growth rate of 5-6% seen in the last decade is therefore significantly above the long-term trend.

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 R&D expenditures and the number of scientific researchers in the world. This is illustrated in Figure 9, 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).

Again, current article growth appears to be higher than this long-term trend. Between 2006 and 2016, total world S&E publication output grew at an average annual compound rate of 3.9%; the total for developing countries grew more than twice as fast (about 8.6%) (NSB 2018).

[26] http://www.sciencemetrics.org/1findr/
[27] https://www.digital-science.com/products/dimensions/
[28] See https://www.wizdom.ai/.

Figure 6: Growth in the number of active, peer-reviewed English-language journals recorded in Ulrich's directory, 2000–2013

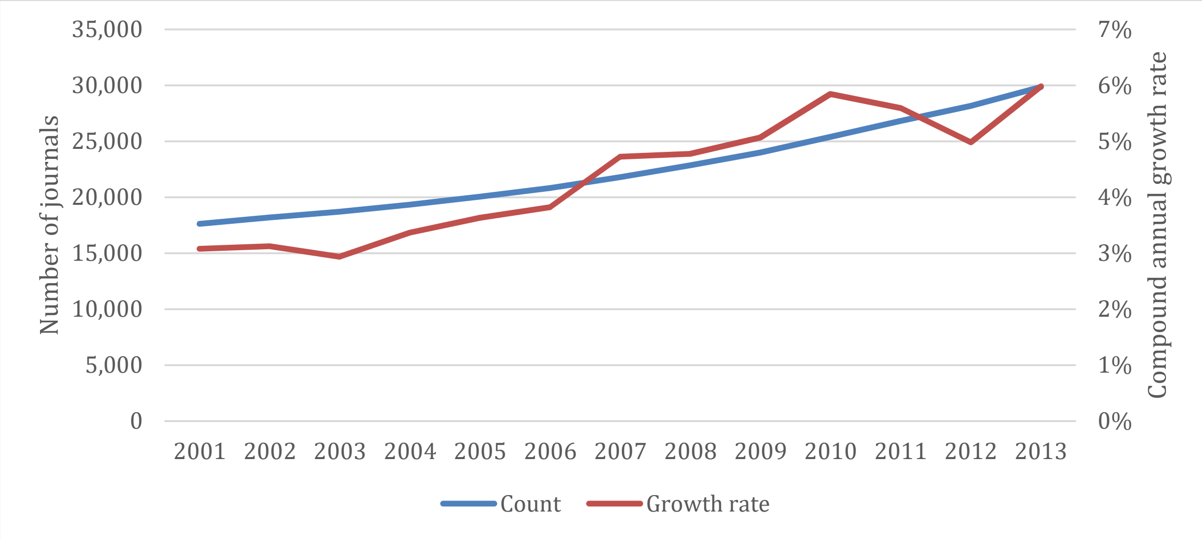


Figure 7: The growth of active, peer reviewed learned journals since 1665 (Mabe 2003)

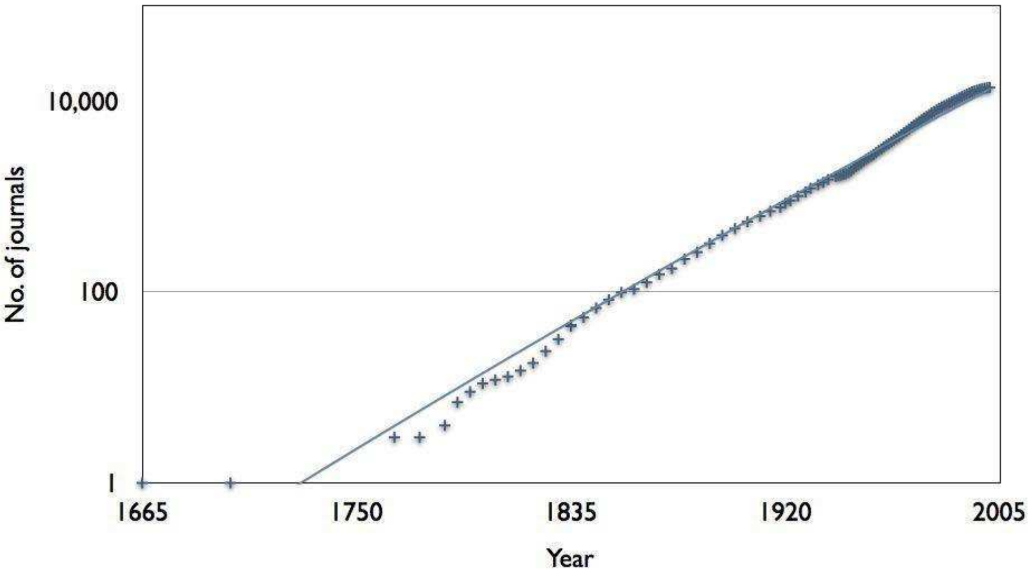


Figure 8: Articles indexed from academic & scientific journals – 1findr, Dimensions, Core + ESCI WoS and Scopus, 1975-2018 (Courtesy of Eric Archambault)

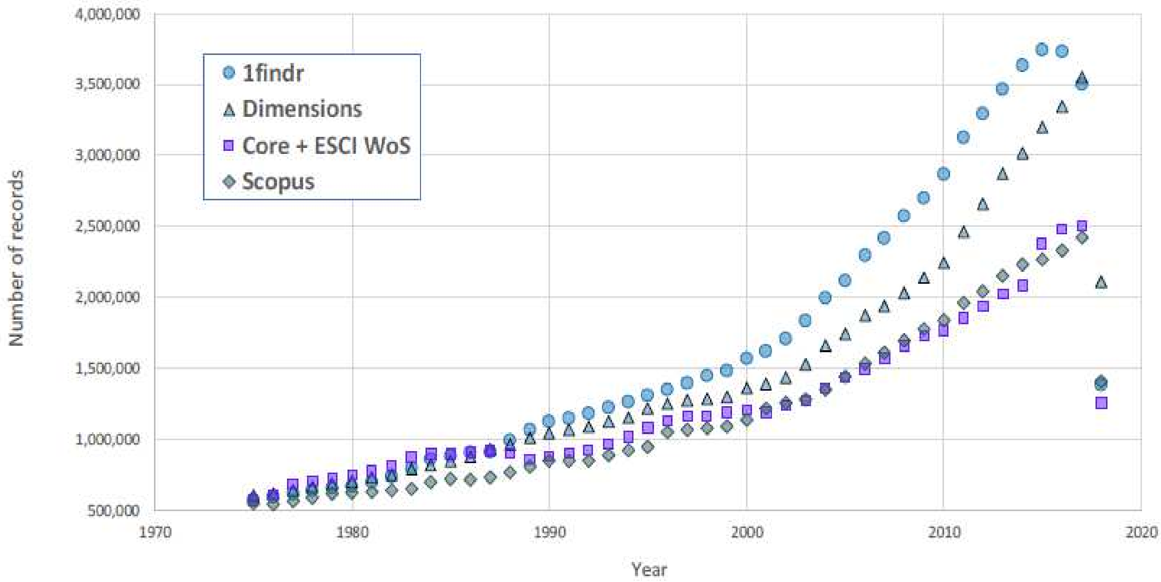
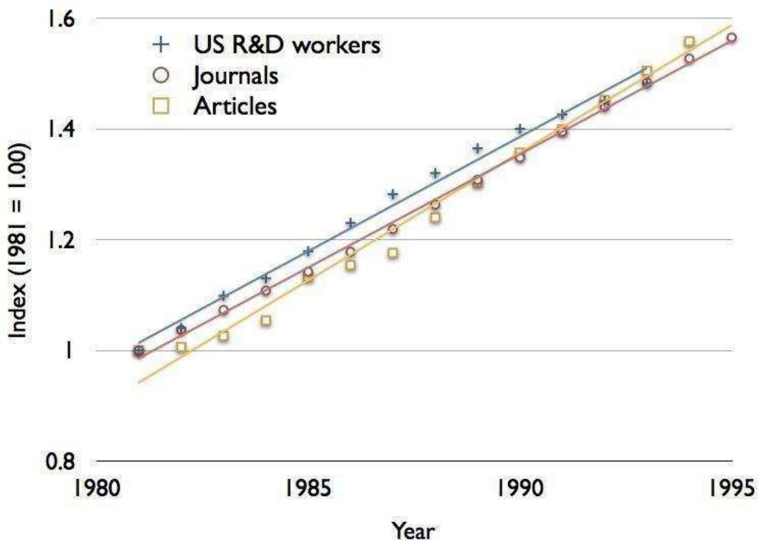


Figure 9: Relationship between numbers of researchers, journals and articles (Mabe 2004), using data from ISI and NSB)



2.5.1 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 2013 ALPSP report (Inger & Gardner) gave similar numbers to 2008, 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 reported 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.)

By 2016-17, 98% of serials subscribed to by UK academic libraries were received in electronic-only format, compared with 91% in 2009/10.[29] Nevertheless there is continuing demand for print 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. As a result, many journals still maintain their existing print editions, although there is a growing shift towards print on demand.

The transition to eBooks has been much slower, but there has been significant movement in recent years. Ebooks made up only about 17% of STM book revenues in 2011, but by 2016 the proportion was closer to a third. Growth rates are expected to slow from 6% over the period 2013 to 2016 to an estimated 4.1-4.6% from 2016 to 2020 (OC&C, cited in SpringerNature 2018), while textbooks may take longer to move largely to digital, although there is a lot of innovation in this area.

2.6 Global trends in scientific output

2.6.1 R&D expenditures

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

Global spending on R&D has generally grown faster than global GDP over the long term, rising from $522 billion in 1996 to $1.3 trillion in 2009 (NSB 2012) and an estimated $1.9 trillion in 2015 (NSB 2018, see Figure 10). The 2008/09 recession had a chilling effect on global R&D expenditures however: the annual growth for OECD countries for 2008–2012 was just half that for 2001–2008 (OECD 2014). Recent years have seen a return to growth, and government budget allocations for R&D increased (in real terms) by 2.5% in 2016 (OECD 2018b).

The large majority of this spending (90%) takes place in the three major economic regions of the world, North America, the EU and Asia. The USA spends the largest amount compared to other individual countries at $497 billion (a research intensity of 2.7% of GDP, well above the global average), with a 26% share of all global R&D. However, U.S. and European dominance is increasingly being challenged by China, whose expenditure overtook that of the European Union to reach $409 billion, or 21% of the global total, in 2015. The National Natural Science Foundation of China (NSFC) celebrated its 30th anniversary in 2016 and in that time the budget has grown from the equivalent of $10 million to almost $3 billion. This one body now funds around 10% of the world's scientific output by number of papers.

29 https://www.sconul.ac.uk/sites/default/files/documents/The_continuing_evolution_of_UK_academic_libraries.pdf

Governments have long seen spending on R&D as critical to innovation, growth and international competitiveness.[30] In recent years, the focus of scientific discovery has shifted towards problem-solving, in order to tackle pressing developmental challenges (UNESCO, 2015). As part of the United Nations' Sustainable Development Goals (2015-2030) developed and developing countries alike have pledged to substantially increase public and private R&D spending as well as the number of researchers by 2030.[31]

Across the world, the average proportion of national GDP spent on R&D was about 1.8% in 2015 (up from 1.7% in 2010), although there is (unsurprisingly) a wide range in this, from Indonesia's 0.1%, through India's 0.6%, Germany's 2.9% to South Korea's 4.2%, with the average for the OECD countries holding steady at 2.4%. The trend to increase relative spending on R&D looks broadly set to continue, though some have noted a decline in public commitment to R&D in many developed countries (Canada, UK, USA, etc), as opposed to a growing belief in the importance of public investment in R&D for knowledge creation and technology adoption in emerging and lower income countries (UNESCO, 2015).

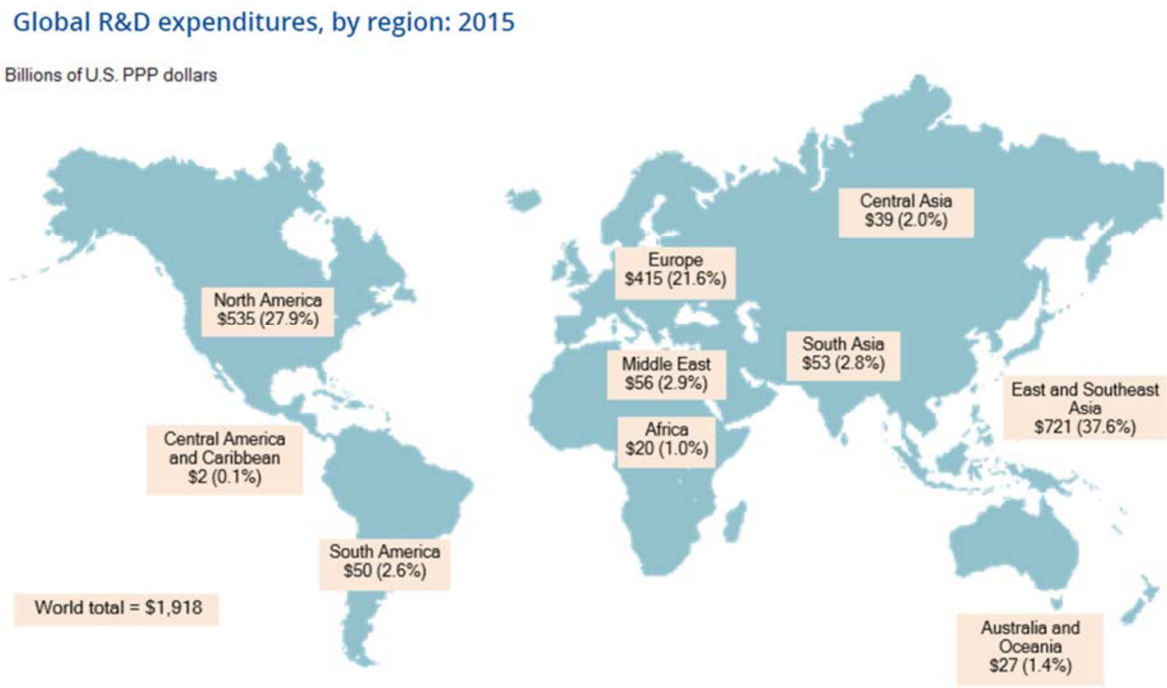
The growth in R&D spending in China has been particularly notable, tripling from 0.6% in 1996 to 1.7% of GDP in 2009, and over 2% in 2015. The latest estimates suggest China will meet the OECD average of 2.4% by 2020. Other East and South East Asian economies have also seen rapid growth, with the notable exception of Japan, which saw its total R&D expenditure fall by 3.3% in 2016 (OECD, 2018b).

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) (Elsevier 2017a).

[30] 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)

[31] https://www.un.org/sustainabledevelopment/sustainable-development-goals/

Figure 10: Global R&D expenditures, by region: 2015 (NSB, 2018)



PPP = purchasing power parity.

Note(s)

Foreign currencies are converted to dollars through PPPs. Some country data are estimated. Countries are grouped according to the regions described by *The World Factbook*, https://www.cia.gov/library/publications/resources/the-world-factbook/index.html.

Source(s)

National Science Foundation, National Center for Science and Engineering Statistics estimates, October 2017. Based on data from the Organisation for Economic Co-operation and Development, *Main Science and Technology Indicators* (2017/1), and the United Nations Educational, Scientific and Cultural Organization Institute for Statistics Data Centre, data.uis.unesco.org, accessed 13 October 2017.

Science and Engineering Indicators 2018

2.6.2 Role of industry

The majority of R&D expenditure is funded by industry: about 62% in the US, 48% in the UK, 66% in Germany and between 75% and 78% in China, South Korea and Japan. The fraction of R&D that is performed by industry is even higher, at a little over 70% in the US, for instance. This is not reflected in publishing data, however, with more than 75% of U.S. research papers originating from academic authors (NSB, 2018).

Most of the research included in the above expenditure figures is not basic science but more applied R&D. China spends only about 5% of its R&D funds on basic research, compared to 17% in the United States, and 84% in experimental development, versus 64% in the U.S. (NSB, 2018). The number of R&D centres in China has grown rapidly in recent years, reflecting the transformation of the economy from a labour-intensive and low-cost manufacturer to a technologically advanced and self-sustaining consumption-led economy. Foreign R&D centres have shot up from 24 in 1997 to an estimated 1,750 in 2018.[32]

In the US, perhaps surprisingly, the share of 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.

2.6.3 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 7.1 million full-time equivalent researchers for 2015, 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 2.5% on 2014, but below the long-term average annual growth rate between 2000 and 2014 of 3.2% (OECD n.d.).

The most recently available UNESCO data reports 7.8 million full-time equivalent researchers in 2013, up from 6.4 million in 2007 (a CAGR of 3.3%) (UNESCO 2015), while Elsevier's latest report for the UK government gives a lower figure, estimated at 6.66 million for 2014 (Elsevier 2017a). (The lower figures are based on the Frascati Manual definition of researcher, which is more tightly defined than UNESCO's "scientist and engineer".) OECD estimates there are about 8.3 researchers and engineers per thousand people in employment (8.1 in the EU, 9.1 US, 10 Japan). The World Bank, using a different measure, puts the total at 1,150 researchers in R&D per million people worldwide in 2015.[33]

Whichever definition is used the number of global researchers has been steadily growing over the longer term, at about 3-4% per year (although with short-term dips during economic recessions, most recently in 2009). The majority of this growth has been driven by emerging countries. One consequence of this is that China overtook the US in numbers of researchers in 2010, and is likely to pass Europe within the next year or two. Research density for China is now close to the world's average at 0.1% of total population: Israel is highest at about 0.8%. However, there are signs that the growth in Chinese researcher numbers is slowing, with the rate of increase dropping from over 10% per annum in the first decade of the century to a little over 5% since 2010 (OECD, n.d.).

2.6.4 Regional changes and the rise of China

On a purchasing power parity (PPP) basis, China overtook the US and EU to become the world's largest economy in 2015 or thereabouts.[34] Having overtaken the EU in R&D intensity in 2013 and in raw R&D spending in 2015, on current trends its research spending will exceed the US's by the early 2020s (OECD, 2018b; NSB 2018). A recent restructure of China's science administration will see the country's Ministry of Science and Technology (MOST) expanded and given more powers over the country's research and innovation drive (Sharma 2018).

---

[32] http://www.ashmoregroup.com/sites/default/files/article-docs/EV%20May%202015%20China%20R%20and%20D_0.pdf

[33] https://data.worldbank.org/indicator/SP.POP.SCIE.RD.P6

[34] The World Factbook puts China's GDP on a PPP basis at $23.1 trillion in 2017, versus $20 trillion for the EU, and $19.4 trillion for the US. See https://www.cia.gov/library/publications/the-world-factbook/rankorder/2001rank.html#ch

The consequences of the huge growth in research spending on research outputs are predictable. In 2016 China overtook the US to publish the largest annual number of research papers, with its share now at 18.6% according to Scopus data (NSB 2018). (Figures 11 and 12 show the trends in article outputs for 2010–2014 and 2006–2016 respectively.)

The likes of Harvard and Stanford may still produce the highest number of cited researchers, and the UK may rank second behind the US as a country, but China is gaining ground fast with the largest increase in 2017 – a 40% jump in the number of top cited authors. In subjects such as materials science and engineering, China has already overtaken the US.[35] The US still attracts more students from around the world, and American research is translated into more valuable IP, although Asia is becoming the epicentre for global innovation with the continent's companies outranking North American companies 45 to 36 in the top 100.[36]

Some broad trends can be discerned across virtually all countries, including a growing focus on research in the field of energy and green technologies (UNESCO, 2015). Otherwise, however, the research priorities of the major regional blocs tend to vary according to their economic needs and internal political pressures. In the US, for example, life sciences continues to receive almost 60% of funding (and the lion's share of this is for medical research), though computer science and engineering have seen higher average annual growth rates over the last decade (NSB, 2018). The EU is similarly focused primarily on biological 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. Priority projects for its 13th five-year plan, covering the period 2016-2020, include quantum communications and computation, brain research, cybersecurity, robotics, genetics, big data applications and new Arctic and Antarctic research stations.

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.

[35] See https://hcr.clarivate.com/2017-researchers-list/
[36] See http://top100innovators.clarivate.com/

Figure 11: Share of world articles 2010-2014 (Source: Elsevier 2017a)

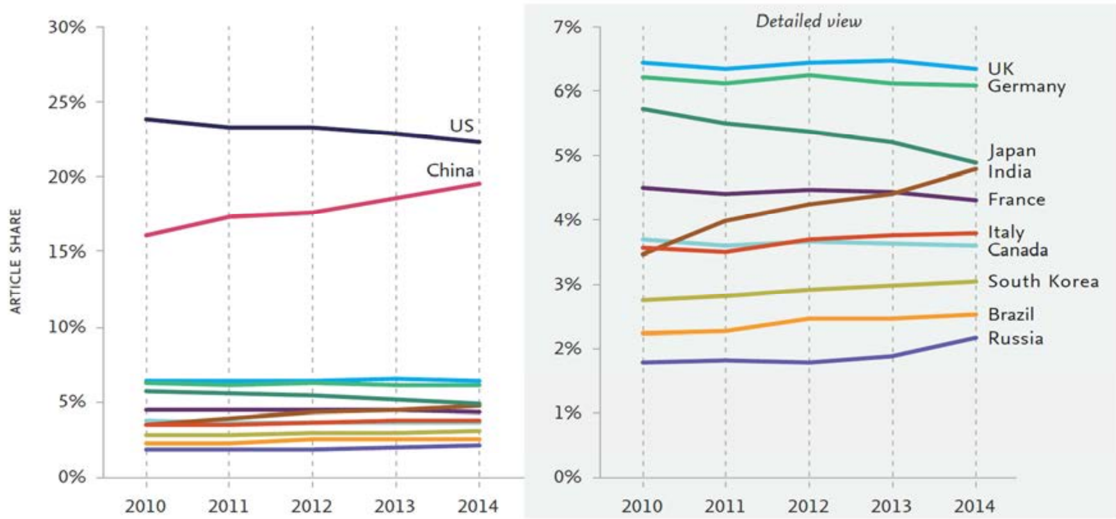
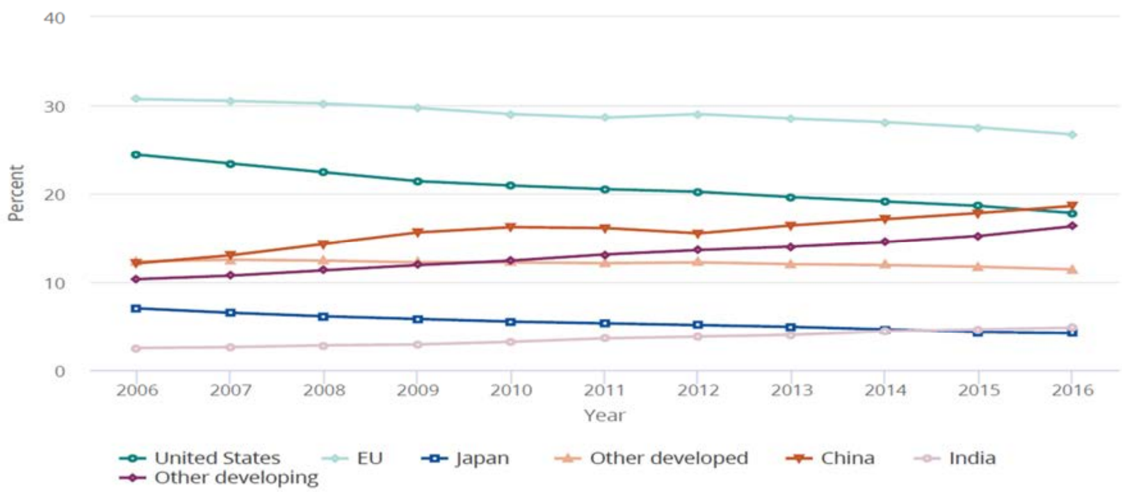


Figure 12: Global shares of article outputs 2006-16 (Source: NSB 2018)



2.6.5 Researcher mobility

The benefits of international collaboration for research are widely recognised (Adams, 2013), with university league tables adopting levels of internationalisation as a key indication.[37] 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,

[37] See, for example, https://www.timeshighereducation.com/student/best-universities/most-international-universities-world-2018

though mobility varies by country, with the UK and Canada having the lowest proportion of "sedentary" researchers (those not publishing outside their home country in the period 1996-2015) at 28%, compared to 62% in Japan and 78% in China (Elsevier 2017a). One trend in recent years is that of Chinese researchers returning to China after several years abroad. The returning 'haigui' (translated from Mandarin as 'homing turtles'), reflect China's improved standing in global research, and a greater confidence in the country's future.

2.6.6 Gender and diversity in research

There is a growing evidence base to suggest that diversity is a key driver of innovation (Hewlett et al, 2013), with others noting that knowledge production in the sciences is enhanced by attention to cultural diversity - specifically diversity of ideas, methods, populations, and sites of scientific practice (Medin and Lee, 2012).

Comparable data on the diversity of the scientific workforce within different countries and region is scarce (NSB, 2018). However, recent years have seen bodies such as UNESCO (2015) and the Global Research Council (2016) place particular emphasis on increasing the participation of women in the STEM fields as a means to drive innovation and achieve excellence in research. UNESCO (2015) estimates that only 28% of researchers around the globe are women, though an Elsevier (2017b) study found that more than 40% of researchers were female in nine of twelve (predominantly Western) countries and regions analysed.

The results vary substantially by field of research, with women still generally and markedly underrepresented in the physical sciences and amongst inventors listed in patent applications. The Elsevier study reports that women are less likely to collaborate internationally and across the academic and corporate sectors, but their scholarly output includes a slightly larger proportion of highly interdisciplinary research than men's. The findings also suggest that underrepresentation of either gender in a given field is associated with reduced likelihood to occupy lead author positions on a research paper in that field.

2.6.7 Collaboration and coauthorship

Collaboration is now the norm, with almost two-thirds of global articles having authors from multiple institutions (NSB 2018). However, the proportion of international collaboration globally has risen only slightly in recent years. In 1988, only 8% of all articles had international coauthors, and this figure had risen to 23% by 2009, but had increased only to 24% by 2016 (NSB 2018). Figure 13 shows the trends in the proportions of research articles with international coauthors. Interestingly the trend is not upwards for all countries, with India a notable exception.

Figure 13: Share of S&E articles internationally coauthored, by selected region, country or economy, 2006 and 2016 (Source: NSB 2018)

There was steady growth in the average number of authors per paper during the second half of the 20th century (Figure 14). A more recent analysis by the Economist (2016) suggests the trend has continued into the current century, with the average number of authors per paper in Scopus growing from 3.2 in 1996 to 4.4 in 2015.

Figure 15 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. 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 by 2015 a paper in Physical Review Letters listed 5,154 authors.[38] The trend has provoked debate over the nature of authorship, with some calling for the term "contributor" to be distinguished from "author" in such cases (See *Section 2.12.8 Typed citations and contributor roles*).

[38] https://www.nature.com/news/physics-paper-sets-record-with-more-than-5-000-authors-1.17567

Figure 14: Coauthorship patterns 1954 to 2000 (from (Mabe & Amin, 2002), using data from Thomson Reuter Science Citation Index)

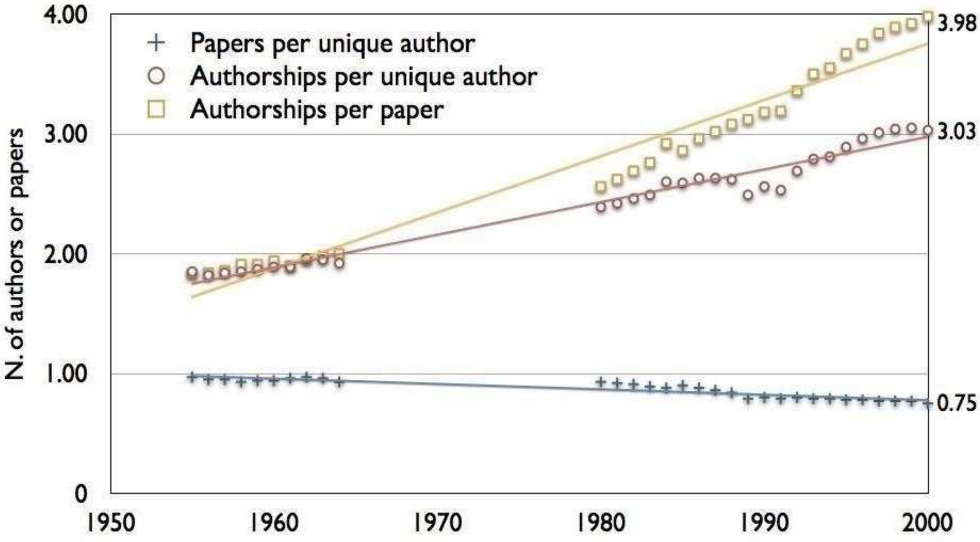
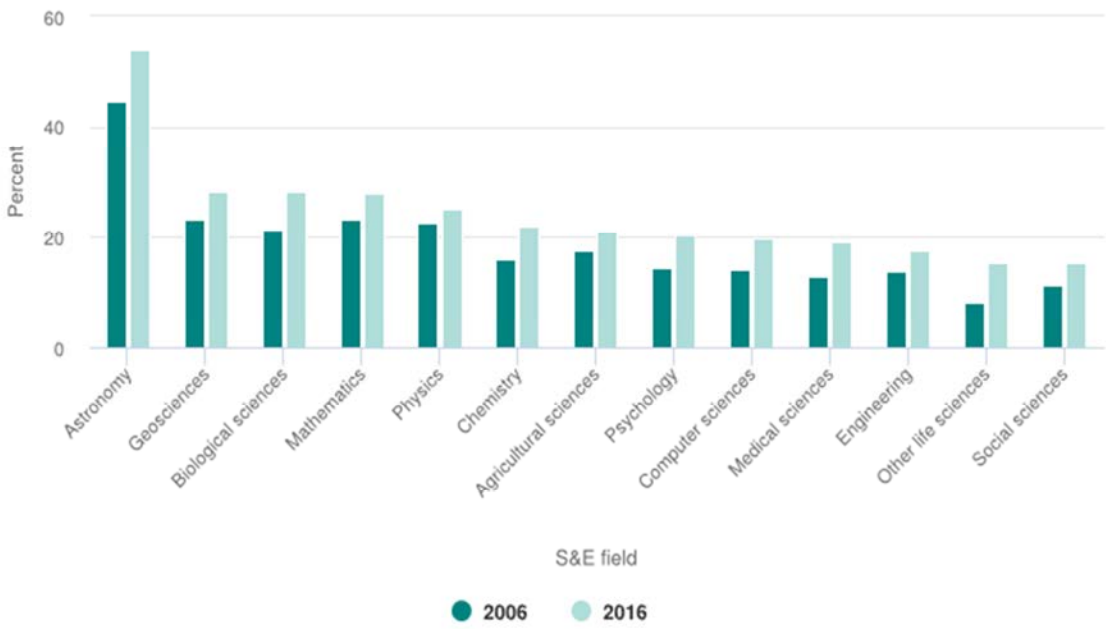


Figure 15: Share of world S&E articles with international collaboration, by S&E field; 2006 vs 2016 (Source: NSB 2018)



The US is a particularly important partner for international collaboration, with 39% of all internationally collaborative papers including at least one US-based coauthor (NSB 2018). The UK, France and Germany all have high levels of international coauthorship at around 50% of papers, and most major research-producing nations have seen an increase in levels of international collaboration in recent years. China, however, has seen only modest growth, while India has recorded a small decline (Elsevier 2017a).

There is a clear benefit to researchers from international collaboration 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 2.5 times as many citations as those with no collaborators beyond the lead institution (Elsevier 2017a).

2.6.8 Salami publishing and Least Publishable Unit

It is a widely held belief among the scholarly community (and often the publishers that serve them) that “others” (never themselves) are dividing up their results into ever smaller parcels so that they can claim more publications. This is stated as being due to the “publish or perish” phenomenon with tenure, grants and promotions dependent on publishing volume. There is no evidence at the macro scale for this. If it were true we would expect to see the ratio of papers published per unique author rising. It does not (Fanelli and Lariviére 2016). In Figure 13 the productivity of authors ranges from about 1.0 in the mid 1950s to 0.75 at the turn of the millennium, a decrease! It would appear that this myth is largely a perceptual one: collaboration as we noted above is clearly increasing, as is the number of coauthors on a paper. Authors are laying claim to more papers but not actually in aggregate producing more; it just looks as though this is the case because each can now cite their involvement in more multiply-authored publications.

2.7 Authors and readers

2.7.1 Authors

The global number of active researchers varies by definition used but is estimated to be between 6.5 and 7.8 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. As of August 2018, the Scopus database holds about 16 million author identifiers.

Scientific journal articles are written primarily by academics. Only 19% of US scientists and engineers are employed within the education sector, but academic institutions account for three-fourths of US S&E publication output (NSB 2018).

Work from Tenopir & King has 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).

2.7.2 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 Elsevier reports that ScienceDirect, the world's largest database of peer-reviewed primary scientific and medical research, has 14m monthly unique visitors.[39]

Meanwhile, the scientific social network Academia.edu reports having more than 63 million registered users, and over 20 million unique visitors a month (see *Scientific social networks*).

These overlapping author and reader communities can be illustrated as in Figure 16. 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.

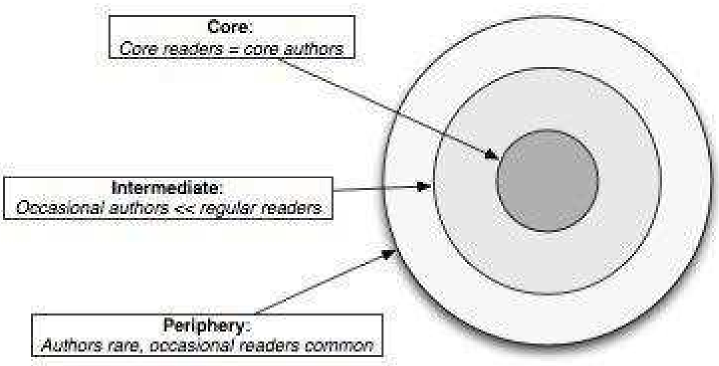
The PEER usage study (CIBER Research 2012a) found that over a six-month period almost every single article (99%) in the study was downloaded at least once from the relevant publisher website, and so was a very large majority, 74%, from a PEER repository. As the authors put it, "the scholarly literature 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) - though data from Ware (2016) suggests this may have increased somewhat in a recent years, with a median 5 hours (mean 9 hours) on each review.

[39] RELX Group Annual Reports and Financial Statements 2017.
https://www.relx.com/~/media/Files/R/RELX-Group/documents/reports/annual-reports/relx2017-annual-report.pdf

Figure 16: overlapping author and reader communities

About 4 million authors publish each year (Plume & van Weijen, 2014), out of a global population of approximately 8 million R&D workers (based on UNESCO figures)



2.8 Publishers

There are estimated to be of the order of 10,000 journal publishers globally: Crossref has over 10,500 membership organisations, publishing over 60,000 journals,[40] while the Scopus database covers 22,000 journals from over 5,000 publishers. The long tail of journals not covered by Scopus is likely to consist predominantly of publishers with just the one journal, as well as significant non-English language content.

The membership of main English-language trade and professional associations for journal publishers (ALPSP, SSP, OASPA 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 2,334 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.

2.8.1 Commercial publishers

The distribution of journals by publisher is highly skewed, with a tendency towards greater concentration over time. Analysis by Lariviére et al (2015) of the Web of Science (WoS)—including the Science Citation Index Expanded, the Social Sciences Citation Index and the Arts and Humanities Citation Index – indicates that the proportion of the scientific output published in journals under the ownership of large commercial publishers has risen steadily over the past 40 years, and even more so since the advent of the digital era – see Figures 17 and 18.

[40] Crossref membership is open to any organisation that wishes to share research outputs through metadata and persistent identifiers. This figure therefore includes organisations such as universities, funders and other bodies who would not necessarily self-identify as a publisher.

Table 1: The 10 largest English-language publishers, by number of journals

Publisher	*Number of journals*
SpringerNature	>3,000
Elsevier	2,500
Taylor & Francis	2,500
Wiley	1,700
Sage	>1,000
Wolters Kluwer (incl. MedKnow)	c.900
Oxford University Press	c.440
Hindawi	>400
Cambridge University Press	390
Emerald	>300

Figure 17: Percentage of Natural and Medical Sciences (left panel) and Social Sciences and Humanities (right panel) papers published by the top 5 publishers, 1973–2013 (source: Lariviére et al 2015)

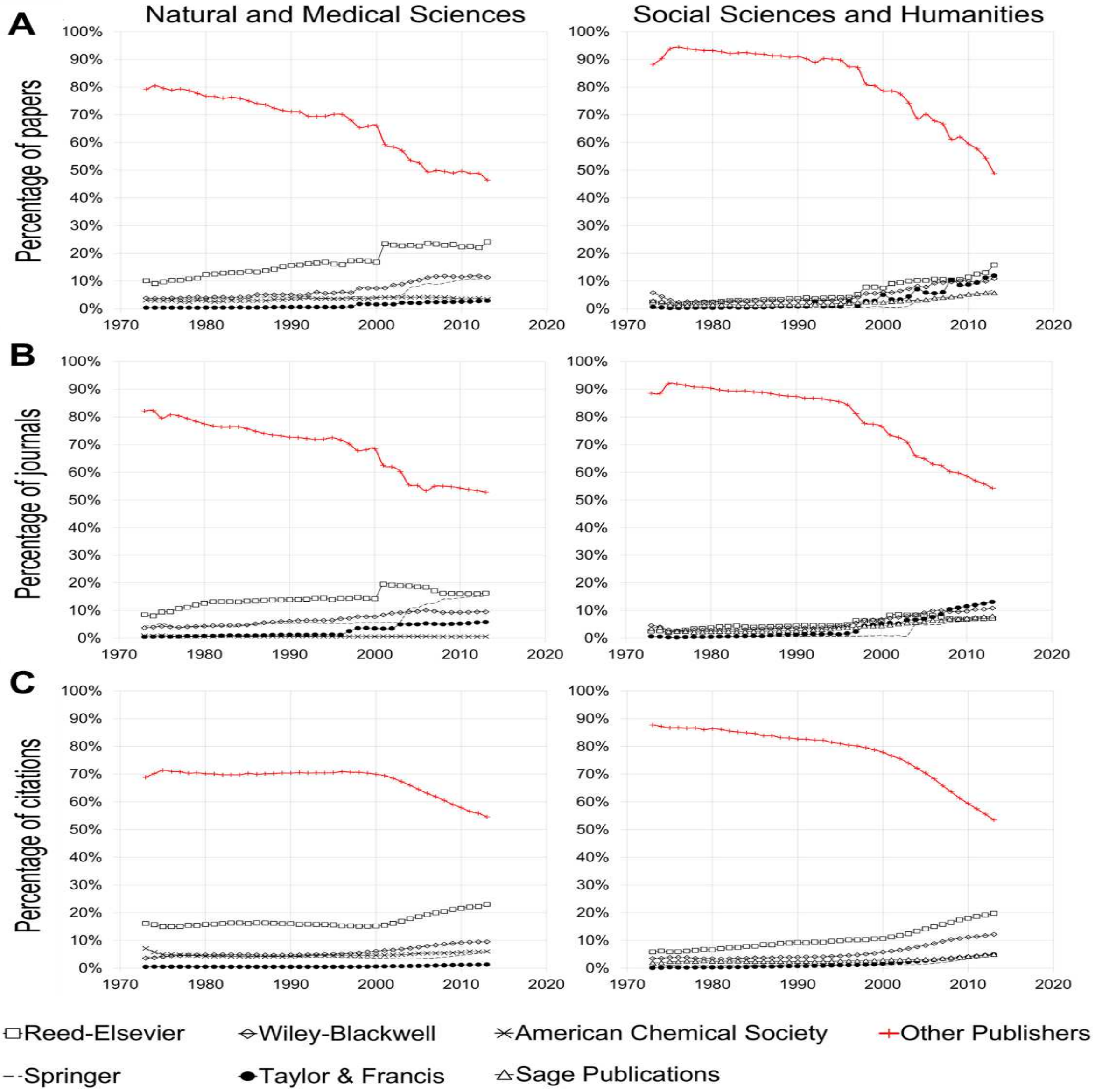
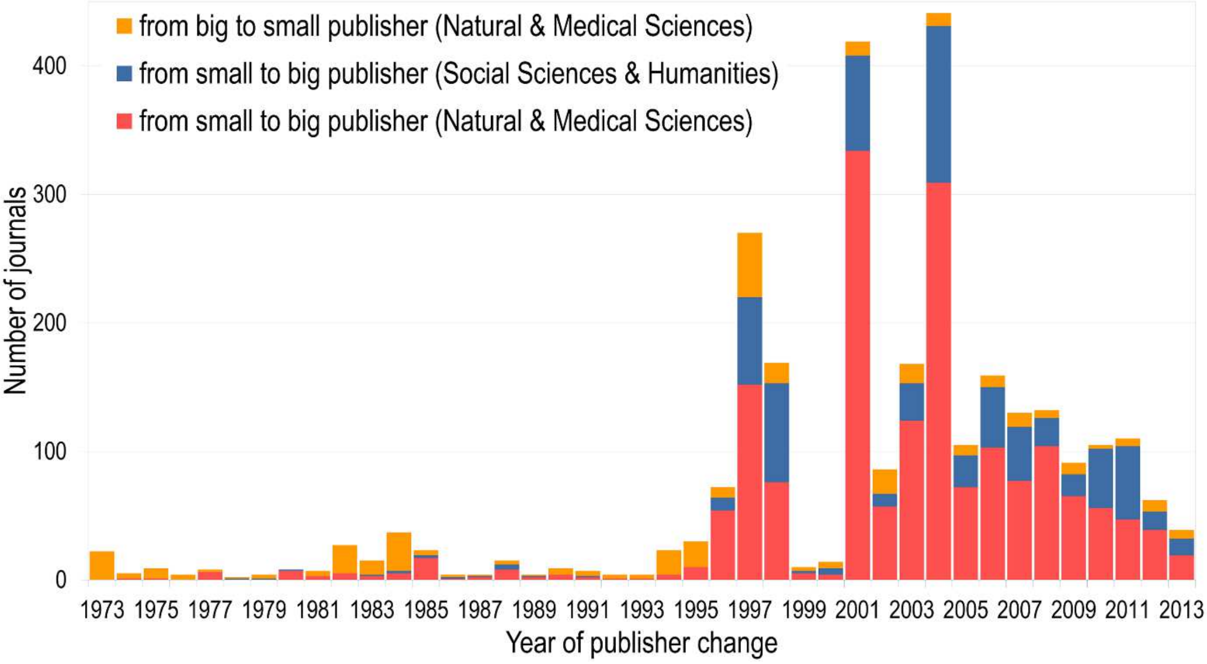


Figure 18: Number of journals changing from small to big publishers, and big to small publishers per year of change in the Natural and Medical Sciences and Social Sciences & Humanities (source: Lariviére et al 2015)



2.8.2 Learned Society Publishing

Learned societies embrace the advancement of their chosen field or discipline while at the same time supporting the professional development of their members through training, conferences and subsidies. The Europa World of Learning lists over 5,000 learned societies globally,[41] and a list of societies by country can be found online – albeit incomplete.[42] These organisations are often a key part of the national research environment.

Academic publications, and especially journals, often account for a significant part of the revenue generated by learned societies – though this varies across disciplines and tends to be less common in the arts and humanities. Revenue from publishing is typically used to subsidise other activities that further the development of the discipline and support society members. The Association of Learned and Professional Society Publishers (ALPSP)[43] has an international membership of 170 not-for-profit publishers, but there are many more societies globally who play an active role in publishing.

For example, Johnson and Fosci (2015) identified 279 learned societies that publish journals in the UK alone, most of which are in STM disciplines. Their publications (mostly journals) represented just over £300m or 26% of the societies' total revenue. Further analysis demonstrated that 80% of this £300m came from 28 societies. The long tail is probably still bigger in the US where the small number of larger learned societies are very large indeed. The largest learned society publisher, and one of the largest publishers of any sort, is the American Chemical Society.

Changes in the publishing environment - such as the wide availability of journals through the consortia publishing model (the "big deal") and the regulatory insistence on open access - have caused concern among learned societies (Johnson and Fosci 2015). Despite rising

[41] See http://www.worldoflearning.com/
[42] https://en.wikipedia.org/wiki/List_of_learned_societies
[43] https://www.alpsp.org/

membership numbers overall,[44] societies still rely on publishing for much of their revenue and publishing remains among the most valued functions for members. On the other hand, a journal's association with an established society is of only minor importance at best to authors in choosing a journal to publish in (Nature Publishing Group 2014). In this context, efforts are being made to ensure the economic viability of learned society publishers in a competitive market. Recently, a group of prestigious not-for-profit scientific membership societies launched the Scientific Society Publisher Alliance (SSPA),[45] an initiative focused on building awareness of and support for publication of scientific research by scientist-run scientific societies. The SSPA seeks to emphasize the value of publishing vital scientific research in scholarly journals managed, edited and peer-reviewed by working scientists.

Although partnerships existed back in the nineteenth century in the UK and Germany, the move to electronic journals and increasing competition led to a surge in partnering among the medium to smaller learned society journal lists. Wiley's acquisition of the two Blackwell companies in 2007 has made it the world's leading learned society publisher with over 700 partnerships, while Taylor & Francis, OUP and CUP also maintain significant society publishing programmes. It has become common for leading publishers to partner with Chinese society publishers in launching or further developing local English language journals.

Through partnerships, societies retain editorial input and a close relationship with their epistemic community, while commercial publishers increase their publishing revenue through their distribution channels and economies of scale. These advantages are partially mitigated for the currently free-standing learned society publisher by the presence of a very large and diverse third-party publishing services industry (Clarke 2015).

Alternative routes to maintaining engaged readerships and revenue are provided by digital platforms such as HighWire, Atypon (now owned by Wiley) and Silverchair. HighWire was formed by Stanford University Libraries in 1995 to help learned society publishers host peer-reviewed content online. The co-founder, the Stanford University Librarian, Michael Keller made it clear that this initiative was intended to stop learned societies having to go to the big commercial publishers to outsource their publishing. HighWire is now HighWire Press, a platform that hosts journals online and provides software solutions for its customers. Non-profit aggregators are available for smaller societies that are an ill-fit for these platforms, including Aggregant[46] and the non-profit BioOne, which offers database of 207 journals from 157 publishers and returns a surplus to the owners.[47]

2.8.3 University Press Publishing

Although Oxford University Press and Cambridge University Press are members of the Association of University Presses (AUP), formerly the Association of American University Presses, they are generally seen as not typical. The range of smaller university presses, now augmented by equivalents from other parts of the world, are on the whole known much more for their monograph lists where as a group they represent a significant historical force in the book field. That being said the AUP considers (2018) that 50% of its members do publish journals. Note that some members are not university presses but other not-for-profits who share the same standards. The big players (journal numbers in parenthesis) are:

- University of Chicago Press (73);
- Penn State University Press (71);
- Duke University Press (56);

[44] https://www.tandfonline.com/doi/full/10.1080/17521742.2015.1013807
[45] https://byscientistsforscience.org/
[46] https://accucoms.com/publisher-services/aggregagent/
[47] http://www.bioone.org/

- University of Toronto Press (40);
- University of Illinois Press (38);
- MIT Press (36);
- and University of California Press (32).

Only a minority of these are STM journals - Chicago has 7, MIT have 12, Penn State has no mainstream science journals, Duke has a cluster of 5 mathematical journals, Toronto has 7, Illinois has 1, and California has 5. Outside this group is Rockefeller University Press which publishes three established scientific journals and nothing else. There are exceptions. There is the Duke mathematics project Euclid bringing together books and journals. MIT present themselves as the only US University publisher whose list is based on science and technology and argue for their quality and cutting-edge nature. Collabra – Psychology, the official journal of the Society for the Improvement of Psychological Science, is a mission-driven Open Access (OA) journal from the University of California Press which shares a variable portion of revenue with reviewers and editors.[48]

Outside North America the model presented by the new UCL Press - fully open access, total subsidy by the university and free publication for UCL faculty - has been well publicised and gained a lot of interest. They are publishing or hosting journals but not (as yet) STM journals, though this may change with the planned launch of UCL's own OA megajournal.[49]

A report commissioned by Jisc (Adema and Stone 2017) identifies a "new wave" of university presses. Common characteristics are that they are open access (OA), digital first, library-based, and they often offer a smaller set of services than a traditional publisher, blurring the line between publisher and platform. In tandem, a small but notable number of academics and researchers have set up their own academic-led publishing initiatives, often demonstrating an innovative or unique approach either in workflow, peer review, technology or business model.

2.8.4 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). Most library publishers work with local academic departments, but more than half provide publishing services to third-party organisations such as learned societies and research institutes. As a generalisation, these library publishers do not take editorial responsibility for what is essentially a service activity. Note also that Library Publishers do not include University Presses even when the University Press reports into the Library.

Libraries surveyed by the Library Publishing Directory (2018) published a total of 442 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 reported has grown slightly since 2016 (when the figure was 436), and is likely to be an understatement, with the data biased towards North America, although this is slowly changing. For example, Open Journal Systems hosts many thousands of journals and bepress around 700 journals, many of which may be library-published. Faculty-driven journals, monographs, and student-driven journals make up the bulk of libraries' publishing output with the other two groups varying more year on year (Library Publishing Directory 2018).

[48] https://www.collabra.org/about/faq/
[49] https://www.ucl.ac.uk/news/news-articles/0118/ucl-launches-open-access-megajournal

Most library publishing 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. In 2018 the most-used platform on the list, Open Journal Systems, is used by 44% of responding programs, surpassing bepress's proprietary Digital Commons platform (41%). Three of the top four most-used platforms—OJS, DSpace, and WordPress—are open source. As pointed out in 2015 these 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. There is little change year on year in the overall picture. In 2018 there was more activity in digital preservation than there had been but there may be fewer services being offered – for example digitisation. It has to be noted that these figures are culled from answers to a question and the respondents are not identical year on year.

STM journals probably count for less than ten per cent of the output and not surprisingly these are small and little known.

2.8.5 The STM publishing workforce

A rough estimate is that the STM publishing industry employs an about 110,000 people globally, of which about 40% are employed in the EU. 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). 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). Meanwhile, a report for the UK Publishers Association estimated that 2,900 people were directly employed in UK academic journal publishing in 2015, representing 10% of the total publishing workforce of 29,000. This figure seems low. 15–18% of academic journal employees in the UK were estimated to be non-UK EU nationals with a further 2–6% from the rest of the world (Frontier Economics, 2017).

The issue of diversity and gender within the scholarly publishing workforce is one which has in recent years received greater attention and investigation. This can be seen by the Workplace Equity project's first initiative, *the Workplace Equity survey*, which aimed to collect data on the opportunities, practices and experiences felt by individuals working in the scholarly publishing industry.[50] The majority of the 1,182 who respondents reported that their employers had stated diversity values claimed that their own experiences within their workspace aligned with these. Nearly 75% of men and women report that they achieve some measure of work-life balance and are able to work in a supportive culture with paid family leave, flex time, emergency leave or state their diversity values. The majority of the workforce is young with the highest percentage in the 36-30 age group at 43%, closely followed by 34% in the 20-35 age group.

Qualitatively, the survey reports that older respondents were the least likely to consider their work environment inclusive. Meanwhile, 45% of women and 67% of those who identify as black or mixed/multiple race do not feel that they have equal access to opportunities for promotion, compared to 36% of men and 45% of those who identify as white. Women make up 67% of the lowest pay quartile, but only 41% of the highest, leaving women making on average, 19.8% less than male colleagues. Similar trends were seen in the 2015 *Scholarly Publishing Demographic Survey* conducted by Digital Science and associates.[51] The

[50] https://www.workplaceequityproject.org/

[51] https://www.digital-science.com/blog/news/scholarly-publishing-demographic-survey-reveals-major-diversity-challenges-in-scholarly-publishing-challengestm/

implications of a workplace which is not ethnically diverse are highlighted in a series of *Testimonies by People of Colour in Scholarly Publishing* (Scholarly Kitchen, 2018).

Recent data released by a series of UK publishers illustrates the extent of gender inequality with regard to pay (Meadows, 2018). From the data released, it was found that at best there was a gender pay gap of 12.6% (OUP) and at worst a pay gap of 40.4% (Elsevier). Many of the organisations include information about what they are doing or are planning to do to address their gender pay gap.

According to the Workplace Equity Project, action is needed in three key areas:

- sponsorship and advocacy;
- facilitating networks by partnering with industry organisations to encourage networking for early career colleagues; and
- challenging exclusion.

Schonfeld (2017) has outlined measures taking place in American scholarly publishing in the form of the University Press Diversity Fellowship Program, which aims to create a new, more diverse 'pipeline' of acquisitions editors.

2.8.6 STM publishing in China

China's ascent in recent years to become the dominant producer of scientific articles has been accompanied by increased investment in the country's publishing infrastructure. The Blue Book of China's STM Journal's Development (CAST 2017) lists 5,020 Chinese journals, of which 548 (10.9%) are English language. Roughly half of the latter are jointly published by Chinese institutes and foreign publishers, with SpringerNature having the largest share, followed by Elsevier and Wiley (Xu et al 2018).

The National Ministry of Finance has also implemented several targeted funding schemes in recent years in order to increase the international competitiveness of China's scientific journals. However, the academic publishing industry remains fragmented, with almost 96% of these titles belonging to a publisher that publishes just a single journal title (Montgomery and Ren 2018). Publishing is highly regulated, applying for a new licence to start a new journal title is difficult, and the price of STM publications in the China domestic market is extremely low. As a result, the sector does not show attractive returns and has so far attracted few private investors.

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, but that aspect of peer review is not covered here.)

The conduct and management of peer review is at the core of the business of publishing scholarly journals, and it continues to grow in scale. In 2013 the ScholarOne system was handling a total of 1.6 million original submissions per year (or 2.2 million including resubmissions) for 4,200 journals. Its rival Editorial Manager processed a total of 2 million manuscripts a year on behalf of 5,800 journals from over 250 publishers. By 2017 the ScholarOne system was handling a total of 2 million original submissions per year (or 2.9 million including resubmissions) for 6,700 journals. At a global scale, a recent survey (Publons 2018) suggests that some 13.7 million reviews were undertaken in 2016 to support the publication of 2.9 million articles.

The process of review varies from journal to journal but it typically involves (Jubb, 2016; D'Andrea and O'Dwyer, 2017):

- editorial staff who make initial checks that papers fall within a journal's scope and meet its standards as to format, research ethics, and so on;
- editors, who – with support from editorial boards and staff - select reviewers and make the final decision on which papers are published; and
- reviewers, who assess the quality of papers submitted to them, and make recommendations to the editors, usually based on a review template.

A range of studies over the last decade (Ware and Monkman 2008, Sense About Science 2009, House of Commons Science and Technology Committee 2010, Mulligan et al 2012, Publishing Research Consortium (PRC) 2015, Taylor and Francis 2015, Publons 2018) confirm that researchers at all stages of their careers (including early career researchers) remain strongly committed to the principle of peer review. But along with publishers and editors, they express many concerns about its practice. Hence there is growing support for change in order to tackle some of the major concerns relating to unfairness and bias, unwanted expense and delay, the burdens on reviewers, and the effectiveness of the process as a whole. The importance attached to addressing these issues is evidenced in the lively contributions to Peer Review Week,[52] which has run in September each year since 2015.

Strong support for continuity comes from a large-scale survey conducted by CIBER (Nicholas et al., 2015). This found that the key benefit of peer review was in providing the 'central pillar of trust', although value was also attached to its role in improving the quality of articles. 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 in choice of journal, alongside the Impact Factor. 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.

2.9.1 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 publishers' role has been to organise and manage the process, and more recently to develop or provide online tools to support it. Online submission systems are now the norm: Inger & Gardner (2013) reported that only 5% of publishers were without peer review systems (sharply down from 35% in a previous 2008 survey). Most publishers opted for one of the three market leaders, Editorial Manager (Aries), eJournal Press and ScholarOne (Clarivate Analytics). Aries was acquired by Elsevier in August 2018.

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

- Faster publication times: the systems can create a fully-linked version of the 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 authors' manuscripts can help editors identify reviewers and also help reviewers. Some

[52] https://peerreviewweek.wordpress.com/

publishers also provide editors with access to A&I databases to help in selecting reviewers as well as in making assessments. Newer artificial intelligence systems based on text mining can also integrate with online submission systems .

- Plagiarism detection: the Similarity Check system allows submitted articles to be compared to published articles and to articles on the web (see *Publishing ethics*).
- Integrated e-commerce: OA article publication charges, or page or colour charges can be managed using publishers' own systems or third-party plug-ins such as the RightsLink Author platform (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 Open Funder Registry data.

2.9.2 Purposes of peer review

The fundamental aim of peer review is to assess the quality of the research and the paper. But quality may be defined in a number of ways, and in practice review is used to serve at least four distinct but related purposes:

- To check for "soundness": whether the research has been performed to appropriate standards so that the findings and conclusions may be considered valid
- To help authors improve the quality of their research and/or its presentation
- To assess originality, significance, and broader interest
- To assess the 'fit' between a paper and the journal

The first purpose is the key one: seeking to ensure that only good science or scholarship gets published, and that work that does not meet acceptable standards does not enter the journal literature. Some "mega-journals" such as PLOS One review only for 'soundness' in this sense. Even for this limited purpose, however, there are some doubts (Jefferson, Rudin, Brodney Folse, & Davidoff, 2007) as to peer review's effectiveness. Nevertheless, the growth of "predatory journals" may have increased the need to distinguish journals and articles that are properly peer reviewed from those masquerading as such (see *Peer review certification* below).

Authors themselves attach great importance to the role of peer review in improving the quality of published papers. Over 90% of respondents to the PRC survey (2016) said that peer review had improved their own last published paper. Mulligan and Mabe (2011) report similar findings, though this varied a little by research discipline.

But filtering is important for readers too, particularly as the volumes of published papers continue to increase. Hence the importance that many journals attach to the third and fourth purposes, of checking for originality and significance, and for fit with the journal and the community it seeks to serve. There is strong debate about the usefulness (or the reverse) of the perceived hierarchies of journals in different fields and the effectiveness of filtering research findings in this way; but it remains widely accepted by researchers and administrators (Tenopir 2010; Ware 2011).

2.9.3 Peer review certification

Readers have 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, and even expert researchers encounter articles from outside their domain. Journals with otherwise good peer review do not always clearly label which articles have been peer reviewed and which not. To counter this problem, some initiatives propose certification of the peer review process at the article and/or journal level.

One such was PRE (Peer Review Evaluation) "a suite of services designed to support and strengthen the peer-review process". Its first service, PRE-val, verified for end-users that content had gone through the peer review process and provided information relevant to assessing the quality of the process. An article-level service, preSCORE was also proposed. In similar vein, the medical editor of BioMed Central called for a "kitemark" to identify research papers that have been peer reviewed by people with the necessary skills (Patel 2014). To date these developments have failed to gain traction within the community, and the PRE domain (http://pre-val.org) is no longer active.

2.9.4 Criticisms of peer review

It is commonly accepted that peer review is imperfect. Much of the criticism is focused in the biomedical and especially the clinical research community. Key criticisms (Ware 2011, Jubb 2016) are that peer review is:

- ***Ineffective*** in detecting error and/or fraud, as shown in the increasing numbers of papers that are amended or retracted post-publication. Studies have suggested (Fang et al, 2012) that high-status journals are no better than lower status ones in detecting flawed findings.
- ***Partial,*** in failing to ensure that wherever possible, published findings are reproducible, or that appropriate attention is paid in the review process to ensuring the soundness of underlying data and/or informatics, or of the statistical techniques employed.
- ***Unsuccessful*** in ensuring the best papers are published in the highest-status journals. Many papers published in the top journals receive few citations, or none at all; while studies suggest that papers published elsewhere (sometimes after rejection by the top journals) are often the most influential as measured by citation impact (Acharya et al 2014, Siler et al 2015).
- ***Unfair and biased*****,** in allowing editors and reviewers to make judgements biased against authors and papers on grounds of sex, race, nationality of field of study, or to fail to declare conflicts of interest, or even to use confidential information for their own purposes.
- ***Slow and costly,*** leading to long delays in publication, particularly when papers are submitted in sequence to several journals; and bringing ever-increasing cost burdens for funders, institutions and authors, as well as publishers.
- ***Distorting,*** in favouring the publication of papers likely to be of broad significance, rather than those reporting negative or confirmatory findings; or on the other hand systematically disfavouring innovative or interdisciplinary work.

Again, these criticisms come from researchers at all stages of their careers, though early career researchers tend to attribute responsibility for these problems to over-powerful editors. Ways to address the problems are at the heart of current discussions about peer review.

2.9.5 Types of peer review

The ways in which peer review is actually conducted varies considerably, in accordance with five main issues:

- ***Identifying reviewers and authors***, where there are three main variants
    - o Single-blind, the most common approach in many STM areas, where authors' identities are revealed to reviewers, but reviewers' remain anonymous. 84% of authors in the PRC survey (2016) had experience of this kind of review.

   - Double-blind, where both authors' and reviewers' names are concealed. This is common in the humanities and social sciences, and some STM journals have introduced it, at least as an option, in recent years; but clues to authors' identities may be difficult to conceal. 44% of authors in the PRC survey had experience of this kind of review.
   - Open, where both authors' and reviewers' names are revealed. Surveys suggest mixed views about this approach, with the most recent Publons survey (2018) suggesting that revealing their identities would make 42% of potential reviewers less likely to accept an invitation to review.
- ***Cascading or portable reports**.* Many publishers, keen to reduce the redundant effort of successive reviews of the same paper for different journals, offer the option for reports to be passed to another journal in their portfolio. It is not clear how widespread the practice has become, though Davis (2018) suggests that it may not work for all publishers. The cross-publisher Neuroscience Peer Review Consortium (nprc.incf.org) provides for journals to forward reports, if authors request it, to any journal in the consortium; but take-up is low. Members of the National Information Standards Organization (NISO) have approved the 'Manuscript Exchange Common Approach' (MECA) - a new initiative co-led by HighWire Founding Director John Sack. The project will see industry technology providers work together on a more standardized approach to the transfer of manuscripts between and among manuscript systems, such as those in use at publishers and preprint servers (see also *section 4.3.4*).[53]
- ***Interaction between reviewers and authors***. Some newer journals such as eLife and those published by Frontiers encourage authors to engage in online discussion with reviewers (whose names may or may not be revealed) or editors. There is evidence that reviewers are willing to change their minds during dialogue of this kind, and many publishers are beginning to show interest in such approaches.
- ***Publishing reviewers' reports***. Some journals have in recent years published the reports on which editors' decisions are based, as a service to both authors and readers. Again, reviewers' identities may or may not be revealed; and surveys again suggest mixed views with 49% of respondents to the Publons survey suggesting that open identities and reports would make them less likely to accept an invitation to review (see *Invitations for reviewers*). In August 2018 two biomedical funders — the UK Wellcome Trust and the Howard Hughes Medical Institute (HHMI) in Maryland — and ASAPbio, a non-profit organization that encourages innovation in life-sciences publishing, issued an open letter calling for the open publication of reviewers' reports (Polka et al, 2018). At the time of writing the letter had been signed by over 20 publishers, representing more than 100 journals.[54]
- ***Fully-open peer review**.* Some journals have adopted a controlled variant of the practice established as a result of the use of servers such as arXiv and BioRxiv, where 'pre-prints' are open for comment before formal publication. Journals such as PeerJ and those published by Copernicus Publications are posted after basic quality checks, and open for comment by named or anonymous reviewers. The papers may then be revised before formal submission for publication. F1000Research operates a similar system, with successive versions of papers and reports openly accessible on the platform.

[53] https://www.manuscriptexchange.org/
[54] http://asapbio.org/letter

Debates about the merits of these different approaches is likely to continue for some time. Some suggest that concealing reviewers' identities is the only way to ensure fully-candid reports, while others complain that it serves to mask sloppy reviewing and bad practice; and that the single-blind version in particular supports biased or even malicious reviewing. Similarly, while some suggest that revealing reviewers' names encourages more thoughtful and constructive reviews, others suggest that junior researchers in particular may be reluctant to have their names associated with negative reviews, or to be seen to criticise senior colleagues. PeerJ has found that only a small minority of those who submit negative reports are willing to release their names; and there is a suggestion that this may be particularly prevalent in medicine. And there are concerns that cascading reviews can trap authors into the portfolios of large publishers. Finally, many researchers at all stages of their careers express concern about the potential for reviews from people without relevant expertise or experience that may be implied by fully-open peer review; and established authors in particular may be reluctant to engage in the kind of online seminar Q and A about their findings that open review implies. There are also concerns, again especially in medicine, that fully-open reviewing may lead to the early 'publication' of flawed findings that could do real harm.

Nevertheless, recent years have seen a growth in support for greater transparency: the PRC survey found that 50–70% of researchers are supportive of, or prepared to accept, open review. This fell to 35–55% if openness included publishing signed reviews alongside the paper; and as we have noted, the most open options serve as a disincentive for many researchers to accept invitations to review (Publons 2018). But experiments and moves towards some form of greater transparency or openness in peer review are becoming increasing prevalent at journals and publishers including the BMJ, Biomed Central, the European Geophysical Union; Frontiers; and EMBO, which has made a strong case of the benefits of open review (Pulverer, 2010). Elsevier ran a pilot with five journals from 2012-2018, and has indicated that it plans to offer an open peer review option to all of Elsevier's journal editors (McCook 2018). Elife has recently (Patterson and Schekman 2018) introduced an experiment under which once an editor has invited a manuscript for full peer review, the journal is committed to publishing the work along with the reviewer reports, the decision letter, and the author response.

2.9.6 Post-publication peer review

Journals have long provided facilities for post-publication review through letters to the editor, review articles, and so on; and the BMJ has secured extensive use of its rapid response and e-letter services. But developments in publishing technology have opened up new possibilities. Some of these are related to the increasing use of pre-print servers, and also to the services of new initiatives such as Science Open,[55] which represent a radical shift from the 'review then publish' to a 'publish then review' model. Comments and ratings may take a number of forms:

- PLoS One is the best-known example of a journal which encourages readers to submit comments and ratings which are published alongside the article, but take-up is low, with fewer than 10% of articles receiving any comment, perhaps because time-pressed researchers have little incentive to devote time to such an exercise. PLOS has recently 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. These are of course similar to the questions posed to reviewers in conventional pre-publication review.

[55] www.scienceopen.com

- F1000 Prime is a more formal service, with a 'faculty' of researchers who provide ratings and evaluations of important articles in medicine and biology.
- Scholarly collaboration networks and sharing services such as Research Gate and Mendeley allow researchers to share bibliographies, and services such as PeerLibrary enable researchers to share reviews of published papers. Pub Peer allows accredited researchers to comment on papers, and contacts the authors when such comments are posted. The NIH launched in 2013 a pilot of a similar service allowing commenting on the 22 million articles in the PubMed database. Commenting was restricted to authors in PubMed who could both comment and rate the usefulness of other comments. PubMed's central position in biomedicine gave the approach significant credibility but, with comments submitted on only 6,000 of the 28 million articles indexed in PubMed, the NIH concluded in February 2018 that the low level of participation did not warrant continued investment in the project.[56]
- Many comments are posted on blogs and other social media, which may be influential but difficult to track.

The key problem with post-publication peer review is the relatively small number of papers that attract any comment or rating. Busy researchers have few incentives to engage, and there is a tendency to focus on controversial papers which attract negative comments. But there is growing interest in aggregating multiple "signals" of an article's 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*). Frontiers has developed algorithms to provide post-publication evaluations of this kind.

2.9.7 Independent peer review services

A more radical version of the cascade review discussed above is offered by a number of peer review services established in recent years, including Rubriq,[57] editage,[58] and Peerage of Science.[59] It is not clear how widespread the demand for such services might be, either from authors or publishers, and Davis (2017) has questioned the commercial prospects for such services.

2.9.8 Abuse of peer review

A few researchers have exploited loopholes in the peer review system to submit bogus reviews. 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. They were then able to propose themselves as reviewers for papers they submitted (Ferguson et al 2014). In one major case, a 14-month investigation by SAGE uncovered some 130 suspicious reviewer accounts and 60 papers articles were retracted,[60] while in 2015 Springer retracted 64 articles in 10 journals.[61] (See also *Publishing ethics*.) In 2017, 400 Chinese researchers listed as authors on some 100 now-retracted papers were reported to be facing disciplinary action for a massive peer-review fraud (Normile, 2017).

56 https://ncbiinsights.ncbi.nlm.nih.gov/2018/02/01/pubmed-commons-to-be-discontinued/

57 http://www.rubriq.com/

58 https://www.editage.com/publication-support/

59 https://www.peerageofscience.org/

60 https://retractionwatch.com/2014/07/08/sage-publications-busts-peer-review-and-citation-ring-60-papers-retracted/

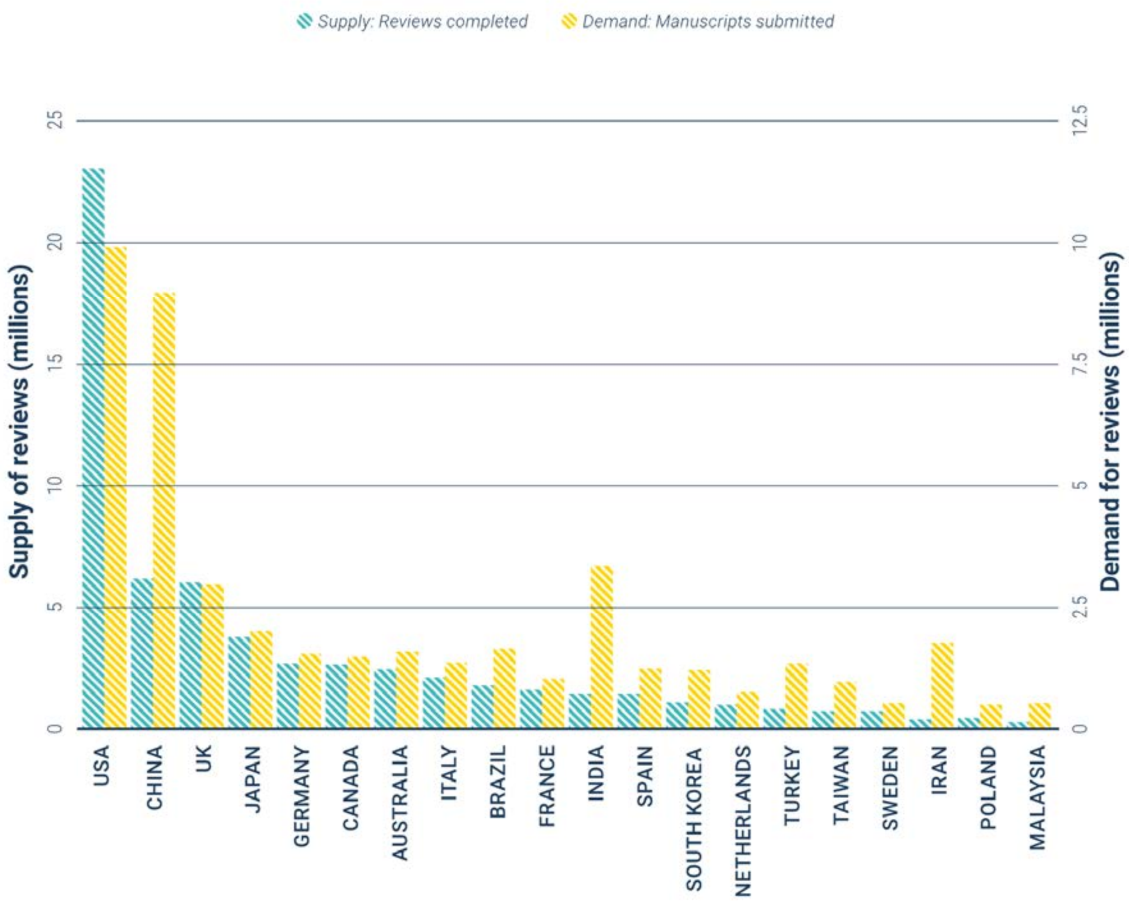
61 https://retractionwatch.com/2015/08/17/64-more-papers-retracted-for-fake-reviews-this-time-from-springer-journals/

As a result of such cases, some publishers have withdrawn the facility for authors to nominate reviewers, and enhanced the security of their systems.

2.9.9 Invitations to reviewers

The Publons survey shows that almost a third (32.9%) of all reviews are provided by researchers in the USA, who are responsible for only a quarter of all published journal articles. By contrast researchers in China provide 8.8% of reviews, while being responsible for a substantially-higher proportion (23.8%) of published articles. In broad terms, researchers in western Europe, North America, Australia and Japan together provide more than two-thirds (68.1%) of all reviews. Reviewers from emerging research nations and regions in the rest of the world (other than China) provided together under a fifth (19.1%) of reviews while being responsible for 29.3% of published articles.

Figure 19: Peer review supply and demand by region (Source: Publons 2018)



These disparities among reviewers reflect similar disparities among editors, who are of course responsible for inviting reviewers. More than 96% of editors come from established countries and regions (including China), with the USA by far the largest provider. The evidence suggests that editors from those countries tend to invite reviewers they know of in the same or similar countries; but that the rate of acceptance of invitations is higher for researchers in the emerging research nations and regions. It also suggests that predominantly male editors are more likely to invite male rather than female reviewers. All these findings suggest that they may be systemic biases in the peer review system. It is also often suggested that early career researchers are locked out of participating in peer review, though there is no robust evidence on this

But the Publons survey also indicated, as many publishers have noticed, increasing ‘reviewer fatigue’, with a rise from 1.9 to 2.4 in the number of invitations required in order to get one report submitted. Reviewers are also less likely to accept invitations where a journal operates an open peer model (Figure 20). In this context, the need to expand the pool of reviewers, and to provide them with good incentives, support and training, becomes ever more important.

Figure 20: The effect of journal review policies on peer review invitation acceptance rates (Source: Publons 2018)

	VERY UNLIKELY	UNLIKELY	MAKES NO DIFFERENCE	LIKELY	VERY LIKELY
Single-blind	2.9%	4.7%	19.3%	31.5%	41.7%
Double-blind	1.7%	2.9%	17.9%	29.6%	47.9%
Triple-blind	4.6%	9.2%	24.8%	26.1%	35.4%
Open identities	15.1%	27.3%	26.9%	20.6%	10.1%
Open reports	13.4%	24.5%	29.5%	22.4%	10.2%
Open identities and reports	18.6%	30.0%	26.3%	16.6%	8.6%
Open final-version commenting	13.8%	23.0%	32.2%	21.3%	9.7%

2.9.10 Incentives for reviewers

Peer review of journal articles has traditionally been seen as an integral part of researchers' professional obligations. Fees are almost never paid (one rare exception is *Collabra*, an open access journal publishing initiative from UC Press); but there are 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 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 have emerged, most notably:

- Publons offers a service whereby reviewers can post their peer review history online, and then showcase this as they choose (for instance by integrating it into their ORCID records and their CVs). The company was founded in New Zealand in 2012 and acquired by Clarivate Analytics in July 2017. By August 2018, Publons had recorded nearly 2.5 million reviews from 450,000 reviewers, covering over 25,000 journals. It also offers peer review training courses and awards for researchers and (as of July 2018) a reviewer search and matchmaking tool for journal.
- Elsevier held an open competition for ways to improve peer review; the winning entry proposed a "reviewer badges and rewards scheme". Reviewers could display badges (generated via Mozilla OpenBadges) on social media pages. In a second phase a "reviewer recognition platform" (RRP) was developed, which now covers some 2,000 Elsevier journals, with around 800,000 reviewer profiles. 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. Researchers can also register their interest for reviewing for their journal(s) of choice and indicate their preferred subject area(s) via the platform.

- ORCID launched peer review functionality in 2015. Organizations currently posting peer review activity data include: American Geophysical Union (AGU), F1000, IEEE, and Publons. Uptake to date has been limited, with less than 0.5% of ORCID iDs associated with peer review activities as of August 2018.

The need to enhance the discoverability and usability of peer review reports led Crossref to extend its services in 2017 to support the registration and retrieval of peer review content through its public API. A new deposit schema now enables the registration of peer reviews, referee reports, decision letters, author responses, and community comments. Metadata to characterize both the peer review type and stage—to accurately reflect the history of the review—can now be deposited and searched for.[62]

2.9.11 Time spent on peer review

Peer review involves a series of actions that take time: identifying and inviting reviewers; reviewers' accepting or rejecting the invitations; identifying new reviewers in response to rejections; producing review reports; evaluating the reports; and so on. Practice varies between disciplines, with overall review times measured in weeks (or less) for rapid-publication journals in life science disciplines; but they can be much longer (months, or more) in mathematics and in the humanities and some social sciences. In the PRC survey (PRC 2016) 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 by reviewers. In the PRC survey, the majority of reviews were completed by a more productive subset of reviewers who managed nearly twice as many reviews as the average. But across all respondents, reviewers reported spending a median 5 hours (mean 9 hours) on each review, and on average reviewed about 8 papers a year. In the Publons survey, reviews averaged 477 words in length (with a notable difference between longer reports from established research countries, and shorter ones from emerging nations). The elapsed time to submit a review report varied significantly across disciplines, but also by country and region, with a mean time between acceptance of the invitation and submission of the report of 19.1 days.

2.9.12 Overall costs of peer review

The notional global cost of peer review is substantial, albeit largely an estimate of academic time devoted to it rather than actual cash: an 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).

2.10 Reading patterns

The number of articles that university faculty members report reading per year had been steadily increasing over time (Tenopir 2007; Tenopir, King, Edwards, & Wu, 2009) but may have recently begun to decline (Figure 21). A range of sources support an estimate of around 250 articles per year for university academics, while non-university scientists read only about half as many (King & Tenopir, 2004). However, there are substantial differences

[62] See https://www.crossref.org/news/2018-06-05-introducing-metadata-for-peer-review/

in reading habits between disciplines and career stage, so apparent changes in overall reading volumes over time may simply reflect changes in the mix of survey respondents (see *Disciplinary differences*). A 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).

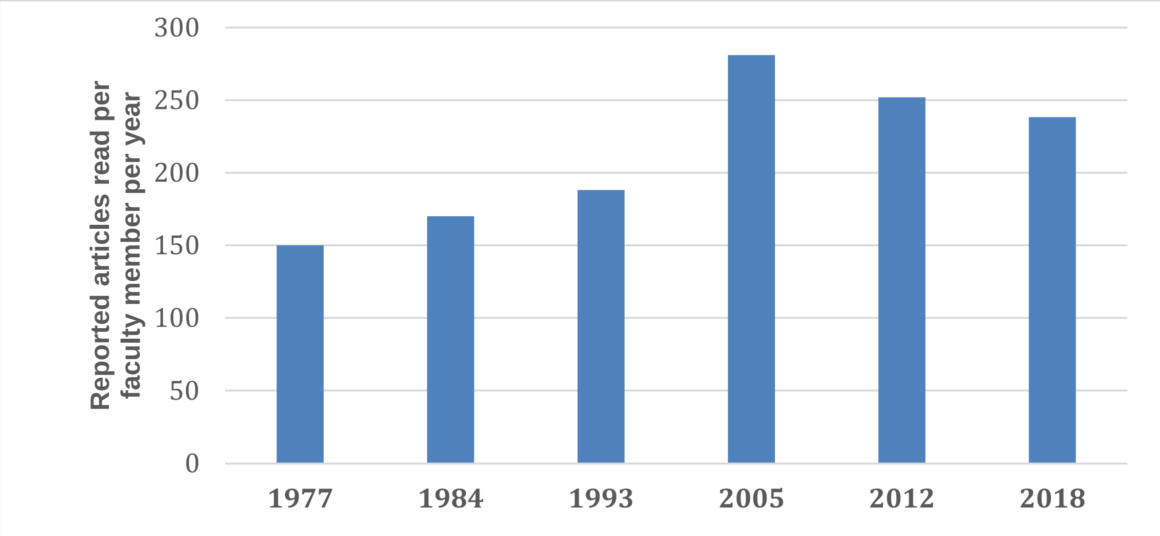
A 2008 international survey (Tenopir, Mays, & Wu, 2011) found that researchers in the sciences reported spending time reading scholarly content of between 12.3 hours/week (health sciences) and 15.3 hours/week (life sciences); while social science researchers said they spent a (somewhat implausible?)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

The average time spent reading a journal article remained at around 45–50 minutes between 1977 and the mid-1990s, before falling to just over 30 mins (Renear & Palmer, 2009). The latest data from Tenopir et al (2018, internal report) suggests it may have begun to increase again, which is consistent with the apparent fall in the number of articles read noted above. Researchers in the medical sciences, education and the humanities spend the least time per articles, while computer scientists and engineers spend the most. There are indications that the thoroughness of article reading may have declined, with reading times not increasing in line with the average length of journal articles increasing substantially (from 7.4 to 12.4 pages between 1975 and 2001). Only around a third of respondents to Tenopir et al's 2018 survey stated they read all of their most recent article "with great care".

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, only 1–3 pages are viewed, and half of visitors never come back. Researchers reported that only 40% said they had read the whole paper of the last "important" article they had read. Users will download articles for future reading or reference, but in follow-up interviews researchers reported that at least half the articles downloaded were never read (and this is likely to be an optimistic estimate). The CIBER authors argue that researchers in the digital environment have moved from vertical to horizontal information seeking and reading, that is, moving quickly over the surface from article to article ("bouncing, flicking, or skittering") rather than reading deeply. While the authors point to factors in the modern environment that encourage this behaviour (over-supply of articles; lack of discretionary time and more pressured workplaces; multitasking becoming the norm; social media conditioning us to accept fast information), they also suggest that researchers may always have read selectively and in snippets, and that the idea of in-depth scholarly reading as the norm was simply a myth.

Renear & 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), artificial intelligence, text and data mining, and the use of new metrics. These are discussed below (see *Technology in scholarly communication*).

Figure 21: Average number of articles that university faculty members reported reading per year (source: Tenopir et al. 2012b and Tenopir et al 2018, internal report)



2.10.1 Access and navigation to articles

The ways readers access and navigate to journal content on the web have consequences for publishers and librarians. Academics use a wide range of methods to locate articles, as illustrated in Figure 22. The growing importance in an online world of searching and parallel reduced importance of browsing is evident in this data. Asking colleagues remains an important strategy, however, albeit ranking behind browsing and searching.

Gardner and Inger's 2018 study (updating earlier 2005, 2008, 2012 and 2015 reports) focussed on citation searching, core journal browsing, and subject searching, and presented these findings:

- Readers are more likely 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, but around 55% of the time people found the article they needed via non-search behaviour.
- Specialist bibliographic (A&I) databases are still the single most popular option for readers searching for articles on a specific topic, with the decline in their use noted in the 2012 and 2015 reports having partially reversed in 2018.
- Use of both general web search engines and the academic search engines (Google Scholar, Microsoft Academic Search) has remained relatively stable since 2012, and it is only in the academic sector where journal readers use Google Scholar more than they do Google.
- Journal alerts have lost traction as a discovery method, but social media has become slightly more popular. Search alerts and bookmarks are still used, but to the lowest extent of those resources tested. -
- There were some notable differences between disciplines: for example, researchers in humanities and education research are much more likely to use the library web pages for article searching than those in physics and astronomy.
- Regional differences may also be significant: for example, people in Asia, Africa and South America think publisher websites have a similar level of importance to search engines while people in Europe and North America think they are far less important.

Furthermore, people in lower incomes countries use a wider range of search resources than those in wealthier countries, who seem to settle on just a few methods.

- Library discovery seems to have peaked in its importance rating, and is really only holding a strong position in Humanities, Education and Social Sciences. Librarians also behave quite differently to everyone else in search, preferring professional search databases and library-acquired resources. This may point to a continued significant gap between what librarians recommend, and how their patrons behave.

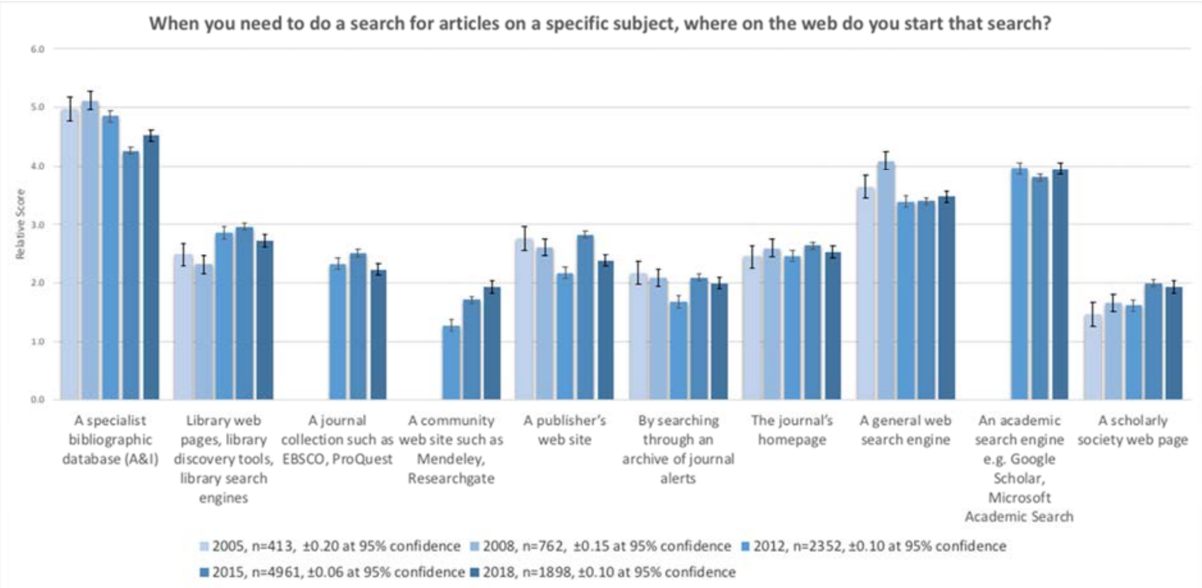
An interview study by Newman & Sack (2013) provides some useful qualitative background to the quantitative data on reading and information discovery behaviours:

> *"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 2015 survey of UK academics confirms this picture, with the leading ways to keep up with current research being attending conferences or workshops, reading materials suggested by other academics, and following the work of key academics. Electronic discovery tools (recommender systems, saved keyword alerts) were of much lesser importance, however compared with a previous survey in 2012, there was a substantial increase (from a low base) in perceived importance for "following other researchers through blogs or social media". There were corresponding decreases for "regularly skimming new issues of key journals" and "regularly skimming table of contents alerts of key journals (Ithaka S+R, JISC, & RLUK, 2016).

Early career researchers have been extensively studied recently by CIBER for the PRC (PRC 2017, 2018c). This has shown that Google Scholar is the first-choice way of finding articles and libraries do not feature highly as a mechanism.

Figure 22: Starting points for discovering latest articles – trend from 2005 to 2018 (source: (Gardner and Inger, 2018)



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 130 articles per year, science and technology titles are much larger at about 160 articles and social science and humanities much smaller at 51 articles a year. [63] 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 was based on a survey of UK academics but many of its findings remain broadly applicable. They 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 showing 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 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

[63] 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 Clarivate Analytics product.

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 differences in the reading and article-seeking behaviours between disciplines. For instance, the number of articles read by faculty members in medicine is much higher than that in the humanities (see Figure 23). These numbers will reflect both the relative importance of the journal article in the fields, the nature of what constitutes a "reading", and the complications of interpreting fields like medicine with a predominating practitioner component. Figure 24 illustrates differences in the ways readers find articles, with marked variance for instance in the importance of browsing.

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

There are also substantial 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). Eger and Scheufen (2018) surveyed over 10,000 academic scholars from all disciplines between 2012 and 2015. They found that disciplines tend to lean *either* towards gold open access or green access, and distinguish three different publishing cultures:

1. The *gold culture* (Biology and Life Science, Health Science, Agricultural Science and Earth and Environmental Science) with high usage rates for OA journals but little use of online repositories or self-archiving platforms.
2. The *green culture* (Physics and Astronomy, Mathematics and Statistics and Business and Economics) with little use of OA journals but strong use of repositories and other online platforms.
3. A *grey culture* (including Social Sciences, Technology and Engineering and Chemistry) with mediocre use of both gold and green open access.

The study found that publishing culture and attitude to OA is more likely to be driven by the respondents' field of research than by their country of residence. Similar disciplinary differences are evidence in Ithaka S+R's survey of UK academics (2016).

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 23: Average articles read per university faculty member per year (Source: Tenopir et al.2018, internal report)

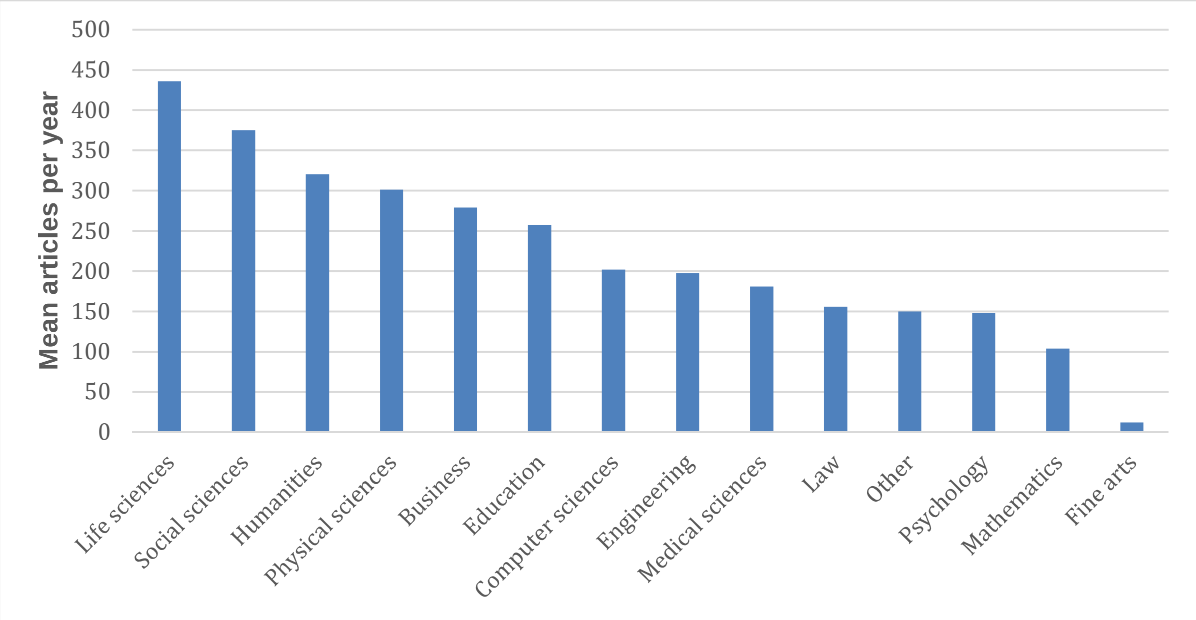
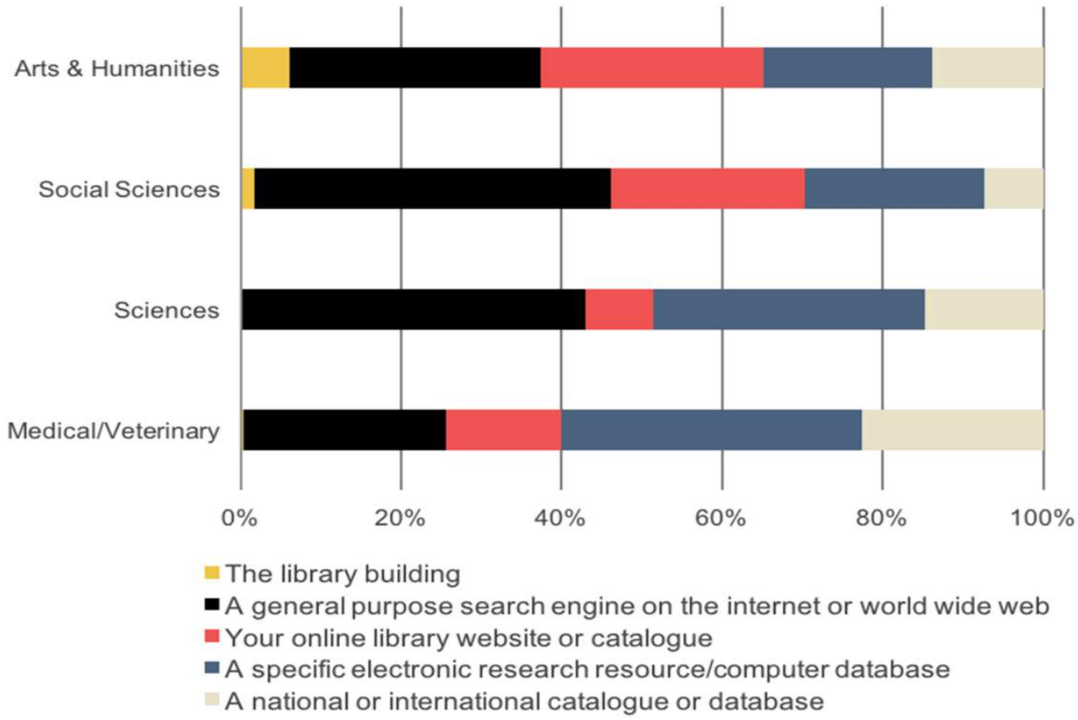


Figure 24: Subject differences in starting points for research in academic literature (Source: Ithaka S&R 2016)



2.12 Citations and the Impact Factor

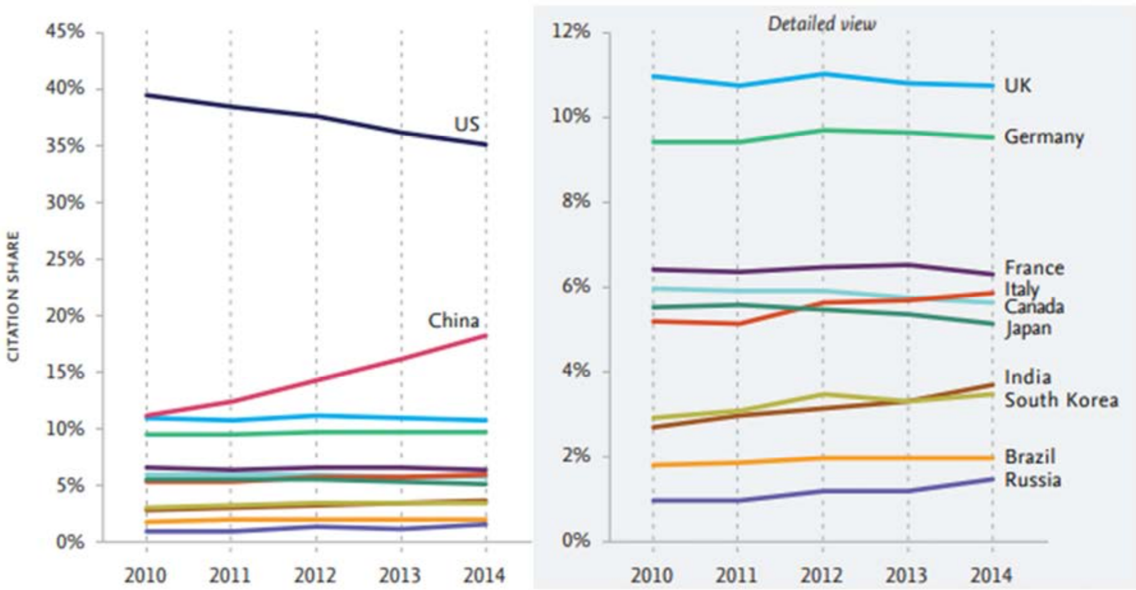
Citations are an important part of scientific articles, helping the author build their arguments by reference to earlier work without having to restate that work in detail. They also help readers enormously by pointing them to other related work (surveys show that this is one of the most popular ways authors navigate the literature, e.g. see Gardner and Inger, 2018). 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.

2.12.1 International trends in citation

As with article publication patterns, the regional shares of citations are changing as a result of these globalisation pressures. From 2010 to 2014 the United States' and, to a lesser extent, Europe and Japan's, shares declined, while China and other Asian countries' shares increased (Figure 25)[64].

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 2004, 42% of citations to Chinese scientific articles came from outside China; by 2014, the proportion had dropped to 38%, suggesting China's expanding article output is being used mostly within China (NSB 2018).

Figure 25: Share of world citations 2010-2014 (Source: Elsevier 2017a)



2.12.2 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. However, different sources of citations data are liable to produce quite different citation counts for a given output. A comparison of highly-cited documents between Google Scholar, Web of Science and Scopus found significant gaps in coverage in the latter two databases, particularly in the humanities and social sciences (Martín-Martín et

[64] The apparent differences between these figures and the NSB 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; NSB used a subset of Web of Science containing 5087 journals.

al, 2018c). Further work in this area indicates that Google Scholar captures nearly all the WoS (95%) and Scopus (92%) citations and is far more likely to capture citations from non-journal sources, and in languages other than English. Nevertheless, Spearman correlations between citation counts in Google Scholar and WoS or Scopus are high (0.78-0.99). The results suggest that in all areas Google Scholar citation data is essentially a superset of WoS and Scopus, with substantial extra coverage (Martín-Martín et al, 2018d).

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 mean number of citations received in a given year by papers published in a journal over the two previous years.)

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: 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 26 shows that while the US remained broadly constant, China's share steadily increased between 2004 and 2014.

Concerns over the deficiencies of the Journal Impact Factor led to calls from a bibliometrician along with several high-profile editors and publishers for the publication of citation distributions (Lariviére et al 2016). Clarivate Analytics has responded to these concerns in its 2018 Journal Citation Reports by introducing a new journal profile page which includes the full citation distribution for the JIF calculation and geographic and institutional contributions.[65]

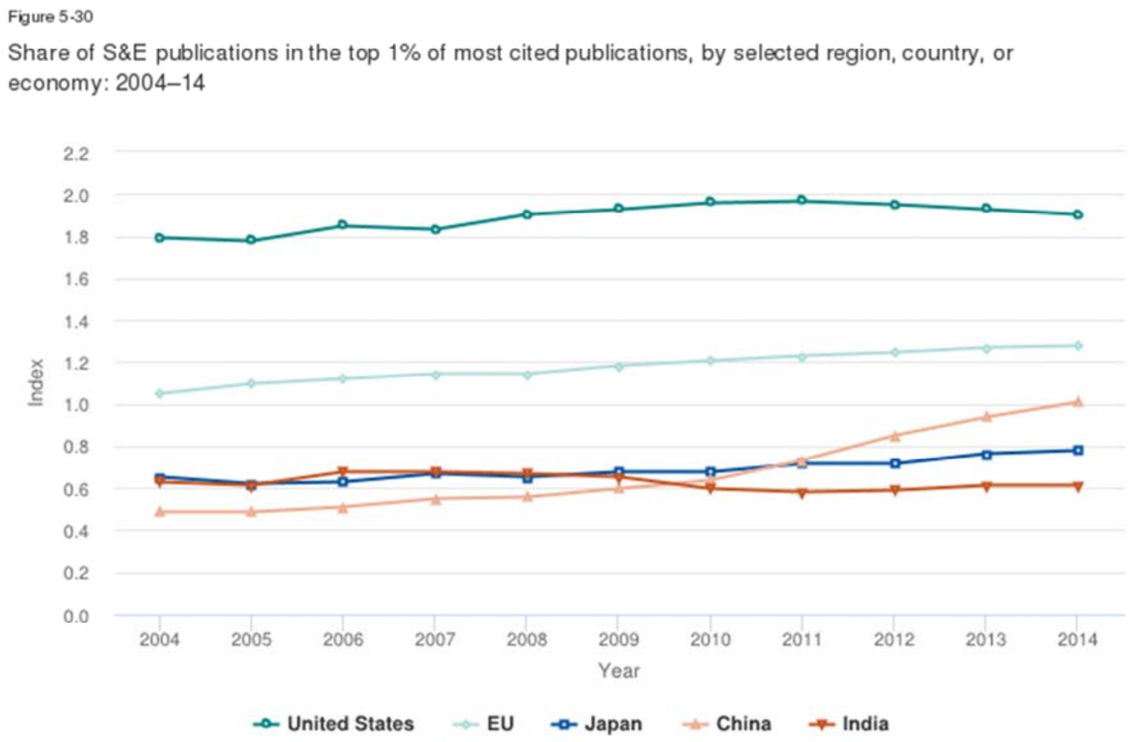
Average impact factors show considerable variation between subject fields, with the primary reason for variation being the average levels of coauthorship. Hence mathematics with coauthorship of 1.25 has an average Impact Factor of 0.5, while biology has coauthorship and Impact Factor both around 4. The fundamental and pure subject areas 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.

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).

[65] https://clarivate.com/blog/science-research-connect/the-2018-jcr-release-is-here/

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*).

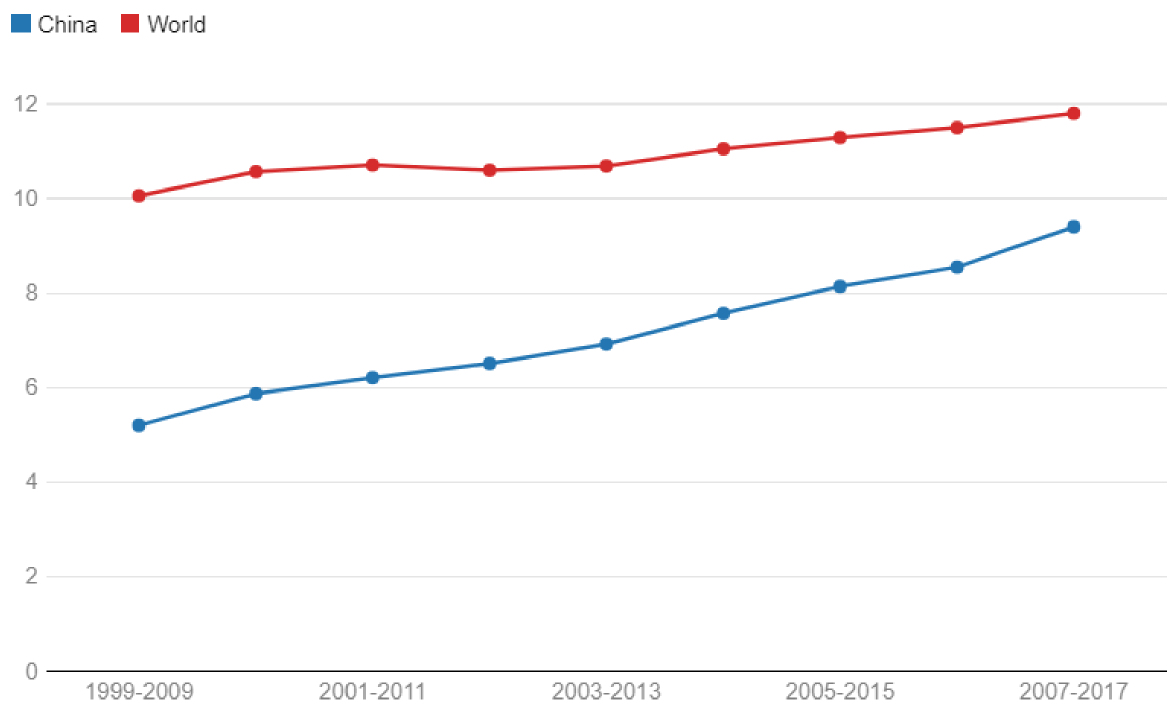
Figure 26: Share of S&E publications in the top 1% of most cited publications, by selected region, country or economy: 2004–14 (from: (NSB 2018)



2.12.3 Citation inflation

The number of citations is increasing faster than the number of publications. Figure 27 shows the trend in average citations per article from the Web of Science database for the period 1999–2017. The average for all countries has risen from about 10.1 in 1999 to 11.8 in 2017, while China has seen a rapid increase over the same period, from 5.2 to 9.4, but remains below the world average. 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 trend towards longer reference lists.

Figure 27: Citation inflation: increase in the average citations per article, for China and the World (Jia et al. 2018)



2.12.4 Effects of research assessment on researcher behaviour

Goodhart's Law[66] 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.

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/REF 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.

2.12.5 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:

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

- **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)[67] 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**,[68] 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.
- **The ResearchGate Score** is a number calculated by the platform ResearchGate according to undisclosed parameters but is believed to relate to quantities that are related to the interactions of the researcher on the platform. These quantities are believed to be publications, questions, answers, and followers.[69]

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.

More recently, EC3 Metrics (2018) published a ‘Periodic Table of Scientometric Indicators’, comprising a total 168 indicators grouped into five primary categories:

- basic indicators;
- webmetric indicators;
- bibliometric indicators;
- h-index based indicators; and
- altmetric indicators (see below).

In practice, use of the impact factor remains 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.

[67] http://www.scimagojr.com/

[68] http://googlescholar.blogspot.co.uk/2014/06/2014-scholar-metrics-released.html

[69] https://explore.researchgate.net/display/support/RG+Score

2.12.6 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[70] 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, and then Elsevier in 2017), PLOS Impact Explorer (based on Altmetric), PageCritic, and others.[71]

There are some preliminary indications that social media activity may predict citations, though the evidence is not strong (e.g. Eysenbach 2011, de Winter 2015). 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 the counting error: most articles will have data at the level of statistical noise and be indistinguishable from each other.

As part of its Open Science agenda, the Research and Innovation Directorate General of the European Commission appointed an Expert Group to report on next generation metrics. Their findings were published in 2017 (European Commission Expert Group on Altmetrics 2017) and they make a series of recommendations urging a middle path: better use should be made of existing metrics and transparency and accuracy are crucial.

2.12.7 The impact agenda

In the competition for funds with other governments departments and agencies, research funders have increasingly come under pressure to demonstrate the benefits they deliver. Recent years have therefore seen a growing emphasis on tracking and measuring the societal, cultural, economic and scientific impact of research.

In turn this has led to rising expectations being placed on institutions and individual researchers to demonstrate impact. The UK's Research Excellence Framework (REF), the Standard Evaluation Protocol in the Netherlands and Australia's National Innovation and Science Agenda (NISA) have all introduced new mechanisms to assess the impact and/or relevance of research over the last five years. Meanwhile, the European Commission's next Framework Programme, Horizon Europe, is expected to prioritise greater impact through mission-orientation and citizen involvement.[72]

---

[70] 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.

[71] See *http://altmetrics.org/tools/* for a current if incomplete list.

[72] https://ec.europa.eu/info/designing-next-research-and-innovation-framework-programme/what-shapes-next-framework-programme_en

In the US, the STAR METRICS collaboration was established to create a repository of data and tools that will be useful to assess the impact of federal R&D investments. Led by the National Institutes of Health (NIH) and the National Science Foundation (NSF), under the auspices of Office of Science and Technology Policy (OSTP), efforts since 2016 have focussed on the development of Federal RePORTER, a searchable database of scientific awards from Federal agencies.[73] A similar effort in the UK led to the development of the Gateway to Research website.[74]

Two initiatives aimed at supporting authors in tracking their own impact are:

- Kudos[75] 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. Launched in 2014, Kudos now has almost 250,000 researchers globally signed up to use its free tools for explaining and sharing information about their publications across multiple platforms and networks, then tracking the relative efficacy of those channels in terms of interest, reads and ultimately – citations – for their work. The free service for researchers is supported by a paid-for Kudos partner subscription service for a growing customer base of over 80 publishers and universities. A premium service is also currently under development.
- ImpactStory[76] 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 originally charged researchers a subscription of $60/year, but introduced a free version in 2016.

Publishers have also recognised the need to support authors in disseminating research results beyond the academic community, to professionals, industrial researchers, policy formers and the general public. Examples of efforts in this regard include Springer Nature's 'Grand Challenges' programme[77] and 'Change the World, One Article at a Time' campaign,[78] and Emerald's 'Real World Impact' campaign.[79]

2.12.8 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). The prospects for authors adopting such a structured process still, 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

[73] https://federalreporter.nih.gov/
[74] https://gtr.ukri.org/
[75] https://www.growkudos.com/
[76] https://impactstory.org/
[77] https://grandchallenges.springernature.com/
[78] https://www.springernature.com/gp/researchers/campaigns/change-the-world
[79] https://www.emeraldpublishing.com/real-impact-awards/

describing contributor roles that could be used in STM journals (Allen, Scott, Brand, Hlava, & Altman, 2014; Meadows 2014).

The CRediT taxonomy seeks to regularize these issues and is dealt with in a recent paper by Brand et al. (2015).

2.12.9 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.

2.12.10 Changes in citation behaviours

In addition to the trends outlined above, two 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.

2.12.11 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 NSB 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 23% between 2003 and 2016, with foreign articles drawing more citations in USPTO patents (54%) than US articles (46%) (NSB 2018). The majority of cited articles fall into a small number of fields: biological sciences (34%), medical sciences (24%), computer sciences (12%), engineering (11%), chemistry (9%), and physics (8%)

2.12.12 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 2) 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 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). Release 4 is the current Code of Practice and the requirement for COUNTER-compliance (COUNTER 2018). The effective date for compliance with Release 5 is January 2019.

The Code defines the publication and usage period as two concurrent years: that is, the usage factor for 2019/2020 will be based on 2019/2020 usage data for articles published in 2019/2020. The Usage Factor: Journals (UFJ1) is defined as "the Median Value in a set of ordered full-text article usage data (i.e. the number of successful full text article requests) for a specified Usage Period of articles published in a journal during a specified Publication Period." 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 is reported annually as an integer (greater precision is deprecated because the level of variation means there is a lot of statistical noise). It integrates articles-in-press from the accepted manuscript stage, and incorporates 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 are 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 full-text articles only.

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 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. Downloading 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 2: Article downloads by country, 2010 and 2014 (source: (Elsevier 2017a)

Country	*Article downloads as a proportion of global total (%, 2010)*	*Article downloads as a proportion of global total (%, 2014)*
USA	30.2	28.9
China	11.4	16.8
UK	9.8	9.9
Germany	6.9	7.0
Japan	4.8	4.2

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 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 key source for reinvestment and innovation, but the profit margins reported by the major commercial publishers have been heavily scrutinised and debated in recent years (van Noorden, 2013; Lariviére et al, 2015).

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).

Recent work by CEPA for the Publishing Research Consortium (PRC 2018a) sought to create a much more accurate model of the economics of journal publishing. Its main findings were that intangible costs such as editorial activities were much higher than tangible ones, such as production and sales and distribution, and were key drivers in per article costs. It is these editorial activities that generate the most value. At a high level, economies of scale seem to apply since the largest journals on the whole have the smaller costs per article. The report says a number of insights were generated by the study:

- It is important to consider the full range of activities involved in academic publishing, including those which may be less tangible (e.g. editorial relationship management).
- The activities involved in publication, their associated costs, and the subsequent quality of published material, varies considerably between different journals/articles. Therefore, a 'one-size-fits-all' approach may be insufficient, or even misleading.
- In addition to the significant differences in quality between journals, demand for journals is linked to value generated, which may not necessarily be related to costs.

2.13.1 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 2016 shows total costs (including overheads) of $39 million for about 26,000 articles published, giving an average of $1,500 per article (compared with only $1088 per article in 2014). 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 2016)
- eLife's financial statements 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). These costs fell significantly as article volumes increased, with the 2017 cost per article estimated at £3,085 (Patterson and McLennan 2016). Of those costs, £1,287 represented the fixed costs, with marginal

costs of £1,798 are incurred for every published article. eLife introduced a publication fee of $2,500 in January 2017, intended to cover the marginal costs of £1,798 and also contribute a small amount towards its fixed costs, which are predominantly funded by its sponsors, the Howard Hughes Medical Institute, the Max Planck Society, and the Wellcome Trust.

- At the other end of the cost scale, Hindawi was reported in *Nature* as publishing 22,000 articles 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; and
- PeerJ said their average costs were in the "low hundreds of dollars" per article

Much lower figures than these have been quoted elsewhere. Brembs (2015) estimated that the per article costs of publication by SCieLo are between $70 and $600 per OA article, depending on the services provided. Meanwhile Bogich and Ballesteros (2016) found that the services required for scholarly communication were available for a price ranging between $69 and $318 per article. Furthermore, they postulate that access to software solutions could reduce the marginal cost of scholarly communication to as little as $1.36 to $1.61 per article. Neylon (2015) offers a critique of these figures, noting that hosting, discovery, labour and development costs are excluded from the calculation. He estimates the base level cost of a modern platform for a journal publishing 50-500 articles a year providing submission through publication in JATS XML, where the selection (and risk of mistakes) are loaded heavily on the academic community, at around $450-550.

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 only around 10–12% of its cost base (Aspesi 2012).

2.13.2 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). Rates of increase have slowed in recent years, with OC&C estimating a compound annual growth rate for academic journals of 2.2% between 2011 and 2016. Nevertheless, journals have outperformed the overall growth in library expenditure over the same period, with the latter growing at only 1.7% per annum (cited in Springer Nature 2018).

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 from a high of 3.7% in 1984 to just 1.8% in 2011.[80] Nor have funding levels improved in recent years, with libraries in both North America and Europe facing budget cuts in 2017 (Publishers Communication Group 2017). This is partly attributable to efficiency gains (e.g. bundled and consortium-based purchasing, other shared services, outsourcing of cataloguing 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.

2.13.3 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). Data for more recent years was not available at the time of writing.

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.

[80] See http://www.libqual.org/documents/admin/EG_2.pdf

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; Eger and Scheufen 2018).

In *New journal publishing models: an international survey of senior researchers* Rowlands & Nicholas (2005) 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.

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].

2.14.1 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).

It should however be noted that in some countries, most notably China, researchers are rewarded for publishing more directly than by just tenure and impact. Cash rewards ranging from $30 to as much as $165,000 per paper have contributed to China’s rapid increase in output over the last couple of decades and a shift in publishing from internal journals to international ones (Quan et al 2017). One consequence has been the development of a ‘paper-broker’ industry which has been well-reported: hundreds of papers have been retracted from the journals of several leading publishers.

2.14.2 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 2015 survey of UK academics (Ithaka S+R et al., 2016), and 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).

2.14.3 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, an 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 openly 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/Jisc/RLUK 2015 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., 2016).

- 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. Work by Eger and Scheufen (2018) demonstrates the significant relationship between field of research and attitudes to open access (see section 2.11 *Disciplinary Differences)*, meaning generalisations across all fields should be treated with care. Nevertheless, some highlights include the following:

- that there are benefits of open access seems 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 of 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, biased 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

A recent author survey of authors published in Gold and APC and Gold Hybrid journals, conducted for the Publishing Research Consortium (PRC 2018b), shows that academics continue to have mixed attitudes towards OA publishing. On the one hand, publishing Gold OA is perceived as giving wider exposure to academic articles and more than half of those who published in a Hybrid OA journal via payment of an APC would have selected another journal were no OA option available. On the other hand, OA plays a minor role in the selection process, and there is great resistance among authors to the use of discretionary budgets for APC payments. Interestingly, offsetting deals appear to have a stimulating effect on OA publication: for some authors they raised awareness of OA, while for others they increased their appetite for Gold OA publishing in the future (PRC 2018b).

2.14.4 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).

2.14.5 Attitudes towards social media and Open Science

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 was based on a relatively small, local sample of researchers, but did 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.

2.15.1 Committee on Publication Ethics

The Committee on Publication Ethics (COPE)[81] 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 guidelines. All COPE members are expected to follow its Code of Conduct and Best Practice Guidelines for Journal Editors, and Code of Conduct for Journal Publishers which have recently been merged into a single document called "core practices" to be found on their website.

2.15.2 Other organisations with an interest in publishing ethics

The International Committee of Medical Journal Editors (ICMJE)[82] 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)[83] 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).

2.15.3 Similarity Check and other automated detection tools

Similarity Check[84] 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. Similarity Check 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.

81 http://publicationethics.org/
82 http://www.icmje.org/
83 http://www.wame.org/
84 https://www.crossref.org/services/similarity-check/

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 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).

2.15.4 Predatory Journals, Beall's List and Think, Check, Submit!

One of the biggest concerns of recent years has been the growth in the number of so-called predatory journals. These often promote themselves to potential authors through bulk, sometimes SPAM emails, frequently have fictitious editorial boards and in many cases use the Gold Open Access model to get money upfront before an author can detect whether their article has been subjected to any peer review whatsoever. Shen and Bjork (2015) estimated that there were 8,000 such journals in 2014, generating some $75 million in revenues, while Cabells' Blacklist included 9,179 journals verified as predatory as of August 2018.[85] Perhaps still more concerning is the threat these journals pose to the integrity of scholarly communication, and the tendency for legitimate open access journals to be tarnished by these practices.

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). 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).

One of the first approaches to tackling the emergence of predatory publishers was the creation of a blacklist of titles by the University of Colorado librarian Jeffrey Beall (Beall's List)[86]. While this approach drew attention to the issues it was flawed by a somewhat inconsistent review policy for correcting the listing of otherwise acceptable journals as "predatory" and the potential legal issues to which Beall and his institution could have been subjected. Beall's list was discontinued in early 2017, and Cabells, publisher of a longstanding journal directory, has sought to fill the gap, introducing new journal blacklist and whitelist products later that same year.[87] The Chinese Ministry of Science and Technology (MOST) has also signalled its intention to establish a blacklist of 'poor quality' scientific journals, including domestic and interntional titles (Cyranoski 2018).

The response of the legitimate publishing industry has been two-fold: firstly the key associations of journal publishers strengthened their codes of conduct and grounds for admission of members; secondly they grouped together with other interested parties to address the problem in a different way to a blacklist. Think!Check! Submit! was created as a resource for scholars to allow them to ask pertinent questions of any publication so as to ascertain whether it observed the best practices or not.[88]

[85] https://www2.cabells.com/blacklist
[86] https://beallslist.weebly.com/
[87] https://www2.cabells.com/
[88] https://thinkchecksubmit.org

The Directory of Open Access Journals also responded by cleaning its database of journals and publishers that did not meet criteria similar to those of the Open Access Scholarly Publishers' Association (OASPA), after discovering that at least 900 suspect journals were included (Anderson 2014b).

Further information on the history of predatory publishing, the publishing industry's response, and recent journalistic investigations into the scale of the problem can be found in Crotty's (2018) retrospective on the Scholarly Kitchen Blog.

2.16 Copyright and licensing

A robust copyright (or more generally, intellectual property) regime that is, and is seen as 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 the complete copyright to the publisher (while retaining certain defined rights) or grant the publisher the sole and exclusive licence to exploit a set of defined rights (about two-thirds of large publishers now prefer this grant of exclusive 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 authors (peer review, copy-editing, kudos etc.). In the case of open access books and journals, authors typically retain copyright, but grant the publisher the sole and exclusive licence to publish and release their scientific works or articles under a Creative Commons licence or similar (see below) which allows use and reuse but imposes conditions, such as attribution of the authors, which depend on copyright. Thus, OA is entirely consistent with the traditional copyright regime and in principle was always possible depending on market demand.

Copyright and other IP law (such as patent law) seeks to establish a balance in law between granting monopoly rights to the creator/innovator (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.

2.16.1 Copyright reforms

The most recent reviews of copyright in a pre-Brexit UK and the EU (the changes adopted and implemented by the UK Government subsequent to the Hargreaves report,[89] and the EU's Digital Single Market Strategy[90]) cover the topics raised by the digital environment that are relevant under any regime:

[89] For the UK, see the IPO's information last updated on 18 November 2014: https://www.gov.uk/government/publications/changes-to-copyright-law. How the UK will be impacted by Brexit in terms of which EU laws will still require to be implemented in the UK, remains an open question. Publishers are well advised in any case to treat the UK as a separate territory as of March 2019 for licensing purposes, while the Republic of Ireland would be expected to remain a territory licensable as part of the EU after that date.

[90] On 14 June 2017 the EU adopted a regulation on cross-border portability of online content services, which aims at ensuring that consumers who buy or subscribe to films, sport broadcasts, music, e-books/journals and games can access them when they travel in other EU countries. By end 2018, the EU is expected to adopt a second and more comprehensive set of rules in a directive modernising copyright. The said directive will focus on wider online availability of content across the EU, adapting exceptions and limitations to the digital world, and achieving a well-functioning copyright market place

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 Convention's 3-step test: exemptions must be confined to a special case; they must not interfere with the normal exploitation of the work; and may not unreasonably prejudice the legitimate interests of the rights-holders

Exceptions currently under review include: portability of purchases access across borders, exceptions for education, research, text and data mining, access of cultural heritage and the inclusion of disabled persons.

2.16.2 UK pre-Brexit copyright changes and changes in the making in the Republic of Ireland

In the UK, a number of the key recommendations made by Hargreaves that were relevant to publishers have now been implemented (Hargreaves 2011; Intellectual Property Office (IPO) 2014). The UK IPO commendably has released easy-to read guidelines for copyright owners, licensors, as well as for users and consumers. For STM the following exceptions stand out:

- 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 to "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*)
- a very limited copyright exception for format-shifting came into effect on 1 October 2014 despite objections from a variety of rights-holders and a concomitant judicial review. The exception was meant to cover personal copying, but was struck down as over-reaching and for want of sufficient consultation by the High Court.
- A notable trend in both EU and UK copyright legislation is to provide that exceptions cannot be altered or more closely defined by way of binding contracts, even where online uses are concerned. This could in practice hamper the development of a "Single" Digital Market, as exceptions continue to differ from EU territory to EU territory and will require great attention when devising online licensing agreements spanning multiple countries or cross-border.

The Republic of Ireland announced the publication of the Copyright and Other Intellectual Property Law Provisions Bill 2018 on the 13th March 2018. The Bill, explanatory memorandum and impact assessment[91] suggest far-reaching changes to Irish law. The Bill is expected to progress through the Houses of the Oireachtas in the coming months, and can be followed on their website.[92]

---

between technology platforms and copyright holders, as well as allowing authors and publishers to continue benefiting jointly from copyright, unhindered by a 2013 EU Court of Justice judgment. On 13 September 2017 a Directive and a Regulation implementing the Marrakech Treaty in the EU were adopted benefitting people who are blind, visually impaired, or otherwise print disabled increasing print material in accessible formats, including adapted audio books and e-books, from across the European Union and the rest of the world.

[91] https://dbei.gov.ie/en/Legislation/Copyright-and-Other-Intellectual-Property-Law-Provisions-Bill-2018.html

[92] https://www.oireachtas.ie/en/bills/bill/2018/31/

2.16.3 US and other territories

In the US too there is an active debate on the need for copyright reform and the Congressional record is slowly but steadily building up to a significant reform. Maria Pallante's words qua the US Register of Copyrights (as she then was) correctly summarized as follows the aims of an eventual US reform: "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."

In the absence of law-making through Congress, US courts – like the Court of Justice of the EU, shape copyright law and deal with striking the balance between technology and content, encouraging science and new knowledge vs wide access to state of the art and existing knowledge.[93]

2.16.4 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. More recent surveys of authors confirm that confusion about copyright and their retained rights persist (e.g. Taylor & Francis 2014). A 2017 author survey conducted by Kudos, in partnership with 10 publishers, found that 83% of respondents agreed or strongly agreed that publisher / journal copyright and sharing policies should be respected, but 60% agreed or strongly agreed that they should be allowed to upload articles regardless of publisher / journal policies.[94] Gadd (2017) attributes this 'cognitive dissonance' in authors' behaviour to a tension between 'copyright culture' (adhering to copyright policies) and 'scholarly culture' (sharing papers with their peers). Authors perceive greater overlap between the two than is necessarily the case, and are liable to side with the latter where the two are in conflict.

---

[93] See for instance US Federal Court for the Federal Circuit, Oracle America Inc., vs Google Inc., 50 F.3d 1381, 750 F.3d 1376, Google's use of the Java API packages was not fair as a matter of law; the district court's decision was reversed, and the case remanded to the district court for damages.

[94] https://blog.growkudos.com/2017/04/04/author-sharing-survey/

In one sense the challenge for a sound copyright law has not changed over time (from the beginnings of enlightenment in 1709 to the present day and the dawn of perhaps artificial intelligence): how to provide a system sufficiently flexible to encourage innovation and new actors, yet also strong enough to protect the works of authorship so central to the purpose of progress and to what makes science science. However, the interpretation of that protection in a digital age continues to prove challenging for authors, publishers and legislators alike.

2.16.5 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, who offer their own list of model licences here: https://support.ringgold.com/cdo-useful-info/). 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 included 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, with sublicences for archives, databases, and SHEDL. The licences underwent a major review in 2017, with changes to the definition of an 'Authorised User', strengthened wording on post-cancellation access', expansion of permitted uses, and new schedules on: industry standards; service levels; and obligations on the Publisher for agreements that include an OA publishing element. 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, last revised in November 2014, as well as those created by various large consortia. http://liblicense.crl.edu/licensing-information/model-license/

2.16.6 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).

2.16.7 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 author's perspective, the typical arrangement is for them to retain the copyright but to sign a sole and exclusive licence agreement with the publisher allowing the latter to issue the work to the general public under the specified open access licence chosen by the authors. The publisher also receives the right by way of grant to licence in parallel and/or subsequently the work in question by way of any other licence that may subsequently be demanded in the marketplace or find favour with commercial or non-commercial users, although other arrangements are possible.

The licences most frequently used for gold open access journals are those offered as part of the Creative Commons licensing structure. 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.[95] 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 particular concerns.

2.16.8 Text and data mining rights

Text and data mining (TDM) has been identified as an important and growing way of using STM content as well as other content, including structured data associated with STM content. It is discussed in more detail under *Technology in scholarly communication* but deserves an entry within this Copyright section because the rights issues remain under active debate and in flux, particularly within Europe, where the potential role of a copyright exception for TDM has been much-discussed.

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.[96]

An alternative 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).[97]

Publishers have since issued and progressively updated statements of commitment to facilitating TDM for non-commercial use (STM 2017a and b). 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*.

2.16.9 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

[95] http://www.stm-assoc.org/open-access-licensing/

[96] There are, for example, over 600 publishers with more than 1000 papers published in PubMed since 2000, clearly an infeasible number for most people to negotiate with.

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

readable format within the resource itself. This already occurs to some extent with certain types of media file, notably music and videos for online sale. In these arenas it is often associated with digital rights management (DRM) arrangements, but this is not necessary: the licences can simply assert ownership and specify allowed downstream uses and licensing requirements. STM also developed orphan and out-of-commerce safe harbour statements to which many members have signed up. Under these "safe harbours" researchers, libraries and their patrons are benefitting from a shield of limited fully exempt liability as long as they have carried out a diligent search to ascertain the orphan work or "out-of-commerce" work status of a work.

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. Journals were archived and preserved by National Libraries in most countries and usually through the legal deposit of publications even in the Netherlands where it was not compulsory.

Before the turn of the century it became clear that electronic journals were going to become the normative form of journal content and there was a lot of research on methods of preservation (Watkinson 2003). 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. LOCKSS and CLOCKSS have to a large extent dealt with this problem (see below). The literature on this topic is difficult to understand if it is not appreciated that outside the US National Libraries are almost always assumed to be the place of preservation for print or digital. Inside the US the discussion revolves around the concept of trusted repositories.

The main solutions currently in use are as follows:

Portico is a not-for-profit preservation service for scholarly content, initially as a JSTOR project before spinning out as an independent organisation. It offers a permanent managed archive of journal (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. Portico currently has 1014 participating libraries, 553 publishers, and preserves more than 200,000 e-journals, 1,245,000 e-books, and 200 digital collections. Senior leaders from Portico shared their perspective on the digital preservation landscape in a recently published paper for UKSG Insights (Wittenberg et al. 2018).

The e-Depot at the Koninklijke Bibliotheek was one of the earliest and best-known players. 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 KB is planning an upgraded e-Depot system with even faster processing and larger capacity functionalities in the coming policy period. The KB assumes that national libraries can take joint responsibility for preservation of the worldwide published heritage, including STM publications, and will cooperate with Portico in a back-up role. The current e-

Depot/KB system contains approximately 27 million articles of e-journals (largely STM) at this moment.

LOCKSS (Lots of Copies Keeps Stuff Safe). As the name suggests, it works on the principle of redundancy, similar to the way that multiple print journal holdings provide security. The Global LOCKSS Network, based at Stanford, allows libraries to collect and store local copies of subscribed content under a special licence (more than 500 publishers have given permission for their content to be preserved in the LOCKSS system). The software allows each library server continually to compare its content with others and thus identify and repair any damage.

CLOCKSS (Controlled LOCKSS) is a sustainable, collaborative, non-profit organisation of some 250 scholarly publishers and 300 research libraries, using the LOCKSS technology. CLOCKSS is governed by a Board comprised of 12 leading academic libraries and 12 leading scholarly publishers. As of mid-2018, the CLOCKSS Archive included the electronic versions of 30 million journal articles and over 20,000 journal titles, 75,000 books, and many types of supplementary material.

The British Library (BL) is charged through the Legal Deposit Act with collecting the cultural and intellectual outputs of the UK for posterity. The Legal Deposit Libraries (Non-Print Works) Regulations 2013 has since come into force, extending legal deposit to include electronic publications, whether offline or online (Akeroyd et al. 2018). The BL's non-print legal deposit work has seen the major UK STM publishers transition journal deposit to electronic over recent years. 30 publishers accounting for c.7500 titles are involved, with some depositing directly with the BL , others via Portico.

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. For these smaller publishers see below - the numbers in this category are almost certainly an underestimate.

The Keepers Registry,[98] based at the University of Edinburgh, reports on the holdings of 13 preservation archives, including CLOCKSS, the Global LOCKSS Network, Portico, eDepot, and others. Finding out where a journal is preserved is seen as very important, though the emphasis is usually on the small number of journals which are abandoned by their publishers in some way

Lynch (2016), the Executive Director of the Center for Network Information (CNI) gave a presentation in 2016 which is much quoted[99] and in it he picked out two ongoing failures. These were the long tail of smaller journals, particularly smaller journals owned and run by smaller journal societies, and also the large number of new open access journals especially originating from outside publishing experience. A collaborative project among CRL, Portico, and CLOCKSS to jointly target the preservation of more "long-tail" journals[100] , has not been sustained. However a recent grant from the Mellon Foundation to the Internet Archive looks a more promising approach[101]. Moreover there are signs that Portico are still making relevant plans.[102]

98 https://thekeepers.org/
99 https://www.cni.org/about-cni/staff/clifford-a-lynch/talks-interviews
100 https://www.crl.edu/events/webinar-pursuing-long-tail-elusive-publishers
101 https://www.librarytechnology.org/pr/23277
102 https://www.librarytechnology.org/pr/23277

Larger Open Access publishers like, for example, BMC make their archiving and presentation policies open[103] but in general in the Open Access wider world the word “archiving” tends to mean the “self-archiving” of Green Open Access in repositories which rarely have plans for long term preservation. A contrary example is the journal Internet Archaeology[104] which works with the UK Archaeology Data Service.

However, a much bigger issue is the archiving and preservation of publications, especially journals, which exist in forms other than text. The taxonomy in the Kenny report that preceded the Act of 2003 specifically mentions multimedia journals.[105] They existed from earlier in the 1990s but they were not mainstream. Indeed the SuperJournal project (Pullinger and Baldwin 2002) from 1995 which sought to explore use multimedia articles in journals found no takers among the author community.

Now however the picture is different. For example, many journals offer facilities for inclusion of video, as well as other supplementary materials. At one time there was a distinction between additional and essential materials but this no longer seems to be relevant to authors and readers. One very large publisher deposits with Portico its digital journals along with the supplementary material they hold. It appears that this additional material is not currently being ingested though it is passed on to client libraries.

The challenge seems to be that there are two lots of players in the preservation world -- services focused on digital content collections (e.g., Academic Preservation Trust, MetaArchive, Digital Preservation Trust -- DPN) and services focus on "publications," mainly textual (e.g., Portico, CLOCKSS). This means that journals or books that combine narrative and data fall between the stools, or only one aspect of their output is properly preserved and connections get broken between the components.

2.18 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 2,900 to 7,200 (Creaser et al., 2006). SCONUL figures show continued growth in UK access over the last decade, with an average of 3.1 serial titles purchased per full-time equivalent (FTE) user in 2016-17, with

[103] https://www.biomedcentral.com/getpublished/indexing-archiving-and-access-to-data/journal-archiving
[104] http://intarch.ac.uk/about/index.html
[105] http://www.bl.uk/aboutus/stratpolprog/legaldep/report/index.html #Appendix A (iv)

98% of these received in electronic-only format - compared to an average of 1.1 serial titles purchased per FTE user in 2009-10, of which 91% were received in electronic-only format.[106]

2.18.1 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 tend to suffer from methodological weaknesses to a greater or lesser extent (Meadows, Campbell, & Webster, 2012).

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.

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.”*

2.18.2 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

106 https://www.sconul.ac.uk/sites/default/files/documents/The_continuing_evolution_of_UK_academic_libraries.pdf

engagement; closer linkages between research and innovation; economic growth; improved efficiency in the research process; and increased returns on the investments made in research.

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

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

In 2016 STM started the RA21 (Resource Access for the 21st Century)[107] project with the active support of some of its larger members, both commercial and not-for-profit. Publishers, libraries, and consumers have all come to the understanding that authorising access to content based on IP address no longer works in today's distributed world. The RA21 project hopes to resolve some of the fundamental issues that create barriers to moving to federated identity in place of IP address authentication, by looking at some of the products and services available in the identity discovery space today, and determining best practice for future implementations going forward. In scope is the creation of a set of recommended best practices around identity discovery and authentication, and engaging publishers, librarians, and other interested parties in the implementation of those best practices. RA21 is now a joint STM/NISO initiative with the governance of all stakeholders established.

RA21 has also brought out a RA21 Position Statement on Access Brokers.[108] The RA21 position is that Access Broker tools are workarounds to solve a fundamental access problem that has plagued users of scholarly information resources for many years, whereas RA21 is working on a long-term, structural solution to the core problem using industry standards and leveraging years of investment from the academic sector in federated identity management infrastructure.

Lack of clarity in writing is a barrier, erected by authors, and has not always been actively discouraged by publishers according to Anderson (2017). As he states, there are no easy answers: "[For] some disciplines and with some kinds of studies, it may be wise to change the way in which studies are written up; in others, it may be wise to leave the writing alone, but add a layer of explanation or summary. And some kinds of studies may resist effective or accurate simplification at all."

Finally, there is a growing concern about the problems of accessing scholarly publications for those potential readers with physical, learning, or cognitive challenges. It is generally agreed that in this sense "accessibility" is not a priority for most publishers in spite of legal obligations in most jurisdictions (Conrad 2018a). For Conrad making contact "accessible" is intrinsically part of the mission of publishers. The position of STM has been clear for many years at least in relation to some specific groups of users (STM 2010). A recent issue of the journal Learned Publishing brings the practical opportunities up to date (Conrad 2018b).

[107] https://ra21.org/
[108] https://ra21.org/index.php/what-is-ra21/ra21-position-statement-on-access-brokers/

2.18.3 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, but 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 outlined 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.

In the past few years new services have launched offering users who lack institutional subscriptions the ability to rent or purchase journal articles at prices lower than the full "pay per download" price on the publisher site. Providers include DeepDyve, Proquest Udini, ReadCube Access, and RightFind (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. To date, uptake of these services appears limited.

2.19 Access in developing countries

The Research4Life programmes[109] are collaborations between UN agencies, STM publishers, universities and university libraries, philanthropic foundations and technology partners. The partnership's original goal was 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 six programmes that collectively provide some 8,500 institutions in 118 developing world countries with free or low cost access to some 20,000 journals, 69,000 books and 120 other information resources from some 180 publishers:

109 http://www.research4life.org

- 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 some 10,000 journals from 60 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 the United Nations Environment Programme, offers access to the environmental literature with some 11,500 journals. Subjects include environmental chemistry, economics, law, 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, and is aimed at promoting the integration of developing and least developed countries into the global knowledge economy.
- GOALI (Global Online Access to Legal Information) was launched in March 2018 in partnership with the International Labour Organization and gives users access to a wide range of essential legal information for their research, scholarship, teaching, studies, advocacy, and work. The programme seeks to enhance legal scholarship and practice, and strengthen legal frameworks and institutions.
- DAR (Digital Access to Research) is a collaboration between Research4Life and the new UN Technology Bank, which was formally inaugurated on 4 June 2018. DAR is the Bank's first in-country effort which, over the next 10 years, plans to develop a tailored outreach and training programme to promote use of Research4Life in each of the world's least developed countries (LDCs).

Based on a matrix of development indicators including GNI per capita, the UNDP's Human Development Index, and the WHO's Healthy Life Expectancy Data, the programmes offer free access to the poorest countries and very low cost access ($1500 per institution for the complete package) to other developing world countries.

Other schemes include:

- HighWire Press offers free access for developing countries to a list[110] of about 140 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 SRKS scheme ended in 2018 and its replacement programme, SERKS (strong and equitable research and knowledge systems), will embrace their AuthorAID and local Journals Online (JOLS) activity, but will significantly downsize the licensing and access component which was the cornerstone of the previous PERII and SRKS five year programmes.

[110] http://highwire.stanford.edu/lists/devecon.dtl

- EIFL (Electronic Information for Libraries)[111] 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, and schemes to support bandwidth improvement. Support is also provided for authors, for instance through INASP's AuthorAid programme.[112]

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, and the Journal Publishing Practices and Standards (JPPS) was established – and is managed - by African Journals Online (AJOL) and INASP to provide detailed assessment criteria for the quality of publishing practices of Global South journals.[113]

[111] http://www.eifl.net
[112] http://www.authoraid.info/
[113] https://www.journalquality.info/en/

3. Open Access

Open access refers to the making available of published scholarly content (such as journal research articles, monographs and conference proceedings) in online digital copies, free of charge at point of use, free of most copyright and licensing restrictions, and free of technical or other barriers to access (such as digital rights management or requirements to register to access).

Its definition in these terms can be traced to the 2002 declaration of the Budapest Open Access Initiative, which is increasingly regarded as canonical within the open access community.[114] Nevertheless, it remains a contested term (Anderson, 2018), though as Peter Suber observes: 'all of the major public definitions of OA agree that merely removing price barriers, or limiting permissible uses to "fair use" ("fair dealing" in the UK), is not enough'.[115] The removal of price barriers alone is referred to as "gratis OA",[116] while in cases where at least some permission barriers are removed as well the term "libre OA" is used (Suber, 2012).

Strictly speaking, open access is a property of the written outputs of a specific research project, rather than the publication that hosts such outputs. In other words, it is a property of an article rather than a journal, or of an individual monograph rather than a book series. This section will largely discuss open access to journal articles, which has attracted the greatest interest from policymakers, and accounts for the lion's share of OA market revenues. Monographs, conference proceedings and other research outputs will be discussed separately but share many of the same challenges as journal articles, as well as others unique to the format in question.

The different approaches to open access can be considered in terms of *what* is made open, *when* it is made open and *where* it is made open. Taking a peer-reviewed journal article as a reference, three "*what"* stages may be distinguished:

- **Stage 1 - Author's original**: this is the author's un-refereed draft manuscript for consideration by a journal, also called a preprint.[117] Authors usually retain copyright over this version of the manuscript.
- **Stage 2 - Accepted manuscript**: this is the author's final refereed manuscript accepted for publication by a journal, sharing in its imprimatur and containing all changes required as a result of peer review – from this stage, copyright is typically assigned to the publisher, or an exclusive licence is granted.
- **Stage 3 - Version of record**: this is the final published, citable article available from the journal's website.

In terms of timing (the "*when"*) there are also three options: prior to (formal) publication, immediately on publication, and at some period after publication (an "embargo" period).

Although articles can be made available on several different platforms (the "*where*"), it is again useful to distinguish between three main options:

---

[114] http://www.budapestopenaccessinitiative.org/read

[115] http://www.earlham.edu/~peters/fos/overview.htm

[116] "Gratis OA" is functionally equivalent to "public access", as used in the White House Office of Science and Technology Policy's memo *Expanding Public Access to the Results of Federally Funded Research:* *https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf*

[117] For definitions of common terminology on journal article versions, see the NISO typology: http://groups.niso.org/publications/rp/RP-8-2008.pdf

- **Publisher's platform**: the version of record is branded and published on the journal's website or any other publisher-owned platform. When this happens immediately on publication, this is referred to as 'Gold OA', where it occurs at a later date it is termed 'delayed OA'.
- **Open access repository**: a version of a manuscript can be deposited in a repository and made publicly available either prior to or immediately on publication, or after an embargo period. Where the accepted manuscript or version of record is made available in this way, it is commonly referred to as 'Green OA' or 'self-archiving'. Repositories may be institutional, subject-specific, national, or international.
- **Other online postings**: Versions of articles may be posted online in a variety of other locations, including author's personal websites, academic social networks and filesharing sites. These postings fall outside the traditional definitions of 'gold' and 'green OA', and in many cases contravene article licensing terms and copyright law. Nevertheless, they are widespread, and often prove easier for readers to discover than deposits of the same articles in OA repositories (UUK, 2017).

3.1 Defining openness

The combination of what version of the article is published open access, when it is published and where it is published has created over the years a complex and contested typology of open access models. For the purposes of this report, we adopt a typology of access that distinguishes between two main types of open access ('gold' and 'green'), based on where the publication is made available, and further differentiates between the business models used for gold OA ('APC', 'no-APC' and 'hybrid', which are discussed further below).

Nevertheless, it is helpful to acknowledge that the dividing line between 'open' and 'closed' access is not always clear. The DART Framework for Open Access evaluates openness based on four dimensions: discoverability, accessibility, reusability, and transparency (Anderson et al 2016). It posits that openness can be conceptualized as a spectrum, with the minimum attributes for a research output to be considered "open" being that it is discoverable and freely accessible at the point of use. Beyond this baseline, degrees of openness (i.e. scale of less open to more open) occur based on more or less nuanced attributes.

Table 3 - The DART Framework for Open Access

Dimension	*Attributes include*	*Description*
Discoverable	• Indexed by search engines • Sufficient, good quality discovery metadata • Links • Persistent unique identifiers • Explicit rights statements • Open and widely used standards (for all of the above attributes)	This may be the most fundamental baseline condition of open (meaning that if an object is not discoverable, it is not open). However, there is a wide range here, including open with bad metadata or links and no or faulty identifiers.
Accessible	• Free (in terms of cost) to all users at point of use, in perpetuity • Downloadable (binary) • Machine-readable (binary) • Timeliness of availability (spectrum)	Generally drives whether we currently consider something to be open, although many variations exist (taking into account embargoes and other conditions).
Reusable	• Usable and reusable (including commercial uses)	Openness is advanced by having fewer restrictions on

	• Able to be further disseminated • Modifiable	reuse, dissemination and modification.
Transparent	• Peer review • Impact metrics • Transparency in the research process (based on the Center for Open Science TOP Guidelines),[118] including data transparency (metadata and level of availability), and software (including version and operating system/hardware) • Research design and analytical methods (plus software and versions), including citation standards, pre-registration of studies and of analysis, and replication • Author transparency (funding source, affiliations, roles, other disclosures such as conflict of interest)	Serves the research lifecycle, given that outputs of research become inputs. Some of the factors that affect transparency include the software used, inclusion of data, the transparency of the peer review process and analytical methods, and more.

Martín-Martín et al (2018a) have proposed a conceptual model of open access rooted in the concepts of legality and sustainability expressed in van Leeuwen et al (2017). It features six dimensions of open access: authoritativeness, user rights, stability, immediacy, peer-review and cost (Figure 28). Yet another model is used by the Open Access Spectrum Evaluation tool to quantitatively score journals' degrees of openness.[119]

Figure 28: A conceptual model of open access (Source: Martín-Martín et al, 2018a)

[118] https://cos.io/our-services/top-guidelines/
[119] See http://www.oaspectrum.org/

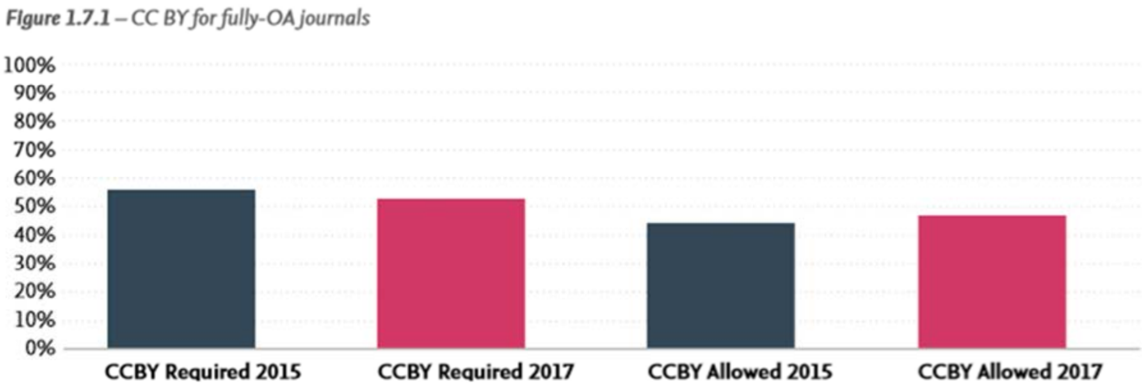
3.1.1 Open access licensing

As seen above, reusability is one of the key attributes of open access. While gratis OA grants access rights to all, the question of what reuse rights are included in libre OA is more complex. A growing number of research funders and organisations now require not just that some version of funded research articles are made freely available, but that they are licensed using the Creative Commons Attribution (CC-BY) licence to facilitate redistribution and reuse with the fewest restrictions (e.g. RCUK, 2012).

However, publishers take a widely varying approach to what reuse rights they allow upon making articles open access. Many publishers switched from the CC-BY-NC to the CC-BY licence as their default for open access articles to align with funder and institutional OA requirements. By dropping the “non-commercial” restriction, publishers forego revenues from commercial reuse of their publications (such as reprints for pharmaceutical companies and other rights income, which are an important source of income for medical journals); to compensate for lost revenue, some publishers have responded by charging more for CC-BY licensing than for CC-BY-NC.

UUK (2017) found that all the major publishers of fully-OA journals allow or require the publication of articles under a CC BY licence, meaning the version of record can be freely shared on other sites and platforms. The balance between those making this a requirement versus those merely allowing CC BY is almost evenly split. By contrast, 90% of hybrid journals allow the use of a CC BY licence (sometimes depending on the payment of an additional fee) while only 10% require it. CC BY is not a requirement for any of the fully-OA and hybrid titles popular with UK authors in the arts, humanities and social sciences; but it is a requirement for OA articles in half or more of such journals in science, technology, engineering and medicine (STEM) subjects.

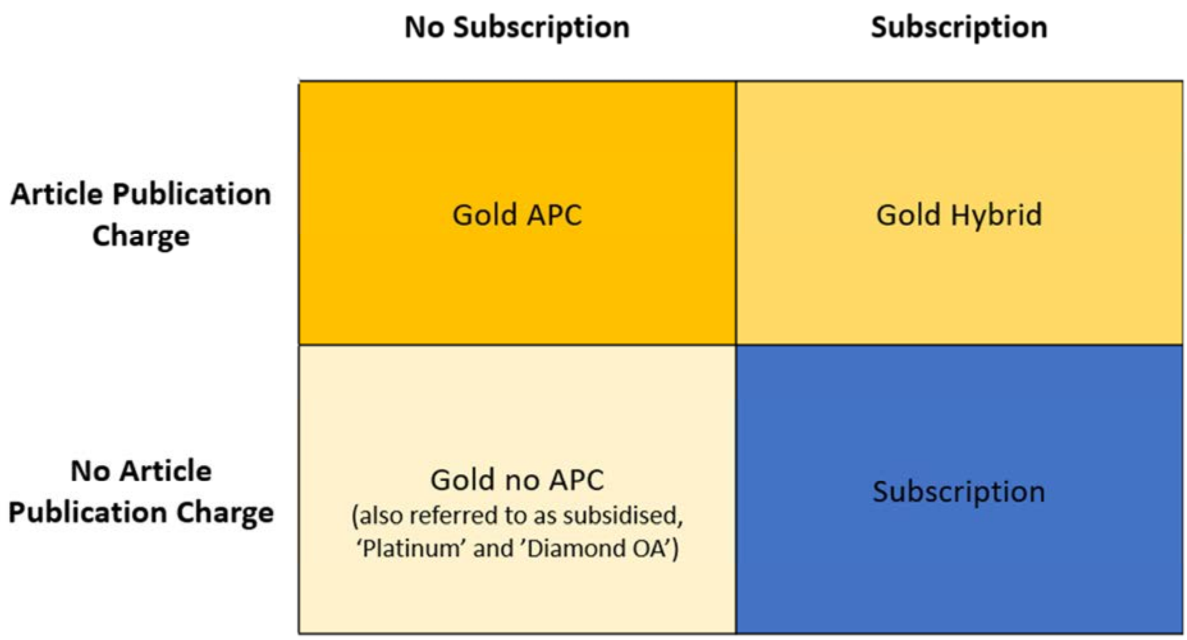
Figure 29



3.2 Open access publishing

This section considers the various business models employed to publish scholarly articles free of access restrictions on publishers' platforms, as summarised in Figure 30 below.

Figure 30 - Overview of open access business models



3.2.1 Gold APC and author fees

A growing number of journals make their entire contents freely available immediately on publication (*full open access journals*) using supply-side payments. Generally, these payments take the form of an article publication charge (APC) levied by journals at the point of publication, and it is up to authors (or frequently their research funder or institution) to make the payment. System-wide efficient systems for payment and tracking of APCs remain a work-in-progress (Pinhasi et al, 2018), but a number of publishers have developed dedicated systems for this purpose, such as Wiley's Open Access Dashboard.[120] Meanwhile others have adopted third-party solutions such as the Copyright Clearance Center's RightsLink OA Agreement Manager.[121] Increasingly, these systems seek not only to simplify payment mechanisms, but also to standardise and streamline the collection of associated metadata to enable monitoring and compliance. Publishers have also introduced alternative APC payment systems to reduce the high volume of low-value transactions associated with the APC model. These include institutional memberships, which entitle institution's authors to publish for free or at a reduced rate, and prepayments, which allow for discounted advance purchase of a defined number of APCs. Moreover, some journals vary APC prices by differentiating between basic publishing service and additional paid-for services (such as rich media, longer mss and so forth).

Publishers also use a range of other author-based charges, summarised in Table 4. Gold APC and other supply-side publishing models provide immediate universal access to scholarly content using a business model where revenue scales in line with increases in research output. However, arriving at a level of APC pricing which is considered sustainable

120 https://authorservices.wiley.com/open-science/open-access/for-institutions-and-funders/account-dashboard.html

121 http://www.copyright.com/wp-content/uploads/2018/02/OA-Agreement-Manager-Product-Sheet-1.pdf

for both research organisations and publishers, as well as being responsive to market forces, is hard, and has been an obstacle to even wider adoption. Further challenges arise from the fact that a shift from subscriptions to APCs would result in a redistribution of costs from a large number of research consumers to a smaller number of research producers, as the Pay It Forward project has demonstrated (University of California Libraries, 2016). Waivers or discounts are also needed to ensure authors in low-income countries are not excluded from the publication process through their inability to meet the cost of APCs. These issues are considered further in section 3.6 *Transitioning to a sustainable open access market*.

Table 4 – Author-based payment models

Model	*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
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
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
APCs supported by third party	Often intended as transitional support rather than a permanent model Discounted (or zero) APCs Supported by societies, institutions, foundations, etc.	Some BMC transfers-in; MedKnow; Versita (De Gruyter Open); eLife (at present)
Submission fees	Non-refundable fee payable on submission regardless of outcome of peer review, typically low value (e.g. USD 20-90 per article) A potentially viable model for high-rejection rate journals	Cultural Anthropology; Hereditas; JMIR (submission fees are notably more common in subscription-based journals than gold OA journals)

Flipped journals

Converting a journal from subscriptions to open access is known as "flipping" the business model. Following the example set by Nucleic Acids Research in 2005,[122] the Open Access Directory[123] now lists 280 examples of journals that flipped their business model. Meanwhile, a review of Elsevier journals suggested that there were 432 immediate OA titles out of 2,642 journals (16.4%), but that only 42 journals flipped to an OA business model between 2013 and 2017.[124]

A recent study (Solomon et al 2016) indicates that flipping is a long process that requires careful planning and preparation. The low number of flipped journals can be explained by a number of factors. First of all, the business case for flipping the journal is not always clear. Conditions that can push journals to consider flipping include: a modest subscription revenue; expected longer term growth in authorship than in subscriptions; higher rejection rates; attractiveness to authors; available funding for OA in the discipline; the volume of existing hybrid articles; and the ratio of current revenues to published articles (Jones 2014a).

Secondly, publishers must consider the consequences of flipping for their authors and readers. For instance, flipping to a Gold APC business model would likely reduce the number of submissions from authors, for instance, because they had not budgeted for APC costs in their research grants. In the short term the shortfall in supply may even decrease article views and citations, and therefore the journal impact factor (Solomon et al 2016). Engaging with the community to understand how authors and funders will react to journal flipping, and what readership can be gained, is therefore a prerequisite to making a decision to flip a journal.

Finally, a number of additional variables affect individual decisions on whether or not to flip a scholarly journal (adapted from Solomon et al, 2016):

- The type of organization that owns and controls the journal
- The transition mode (direct or via some intermediate stage)
- The publishing platform (in-house or outsourced)
- The way in which the funding of the converted journal is envisaged.
- Pressure from research funders, institutions and authors.

3.2.2 Gold - no APC

Another common OA publishing model is so called 'Gold no-APC' (also referred to as 'Platinum'[125] and 'Diamond OA'[126]): open access journals that do not use publication charges. Over 70% of the full OA journals listed in the Directory of Open Access Journals do not use publication charges, and although the number of articles published in gold no-APC journals is likely to be lower (e.g. Dallmeier-Tiessen et al 2010), the scale of this publishing model is nevertheless significant. It is particularly popular in China, where 91% of OA journals published by Chinese learned societies did not charge APCs in 2013 (Montgomery and Ren 2018), and in the humanities and social sciences, where research funding is much lower than in the experimental sciences (Edwards 2014) and academics are resistant to the APC model (Mandler 2014).

---

122 https://academic.oup.com/nar/pages/Open_Access_Initiative

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

124 https://figshare.com/articles/Elsevier_embargo_periods_2013_2015/1554748/11

125 https://www.martineve.com/2012/08/31/open-access-needs-terminology-to-distinguish-between-funding-models-platinum-oagold-non-apc/

126 https://scholarlykitchen.sspnet.org/2017/06/01/diamond-open-access-societies-mission/

No-APC journals most commonly rely on sponsorships from institutions (research performing organisations, research funders, libraries, learned societies, museums, hospitals, for-profit or non-profit organisations, foundations, government agencies and so forth). The subsidies provided by an institution can be financial or in-kind (facilities, equipment, or personnel). They cover a substantial part of the journal's publishing costs and may be provided on a one-off or continuing basis. As Anderson (2018) has observed, the downside of this model is the opportunity cost to the sponsor of allocating scarce resources to publishing, in preference to other activities.

An important variation of the sponsorship model is consortia funding (also known as library partnership funding). This involves the creation of *ad hoc* library consortia for the collective funding of open access publishing. For instance, the journal Paleontologia Electronica is an open access publication owned by a non-profit organization (Coquina Press) which is funded by a consortium of learned societies and IT support from several research organisations.[127] Other consortia take on much more ambitious projects with the aim of publishing open access research from a wide range of journals, generally within a discipline or field of research.

For example, the Open Library of the Humanities (OLH) uses the library partnership subsidy model to fund its PLOS-inspired humanities megajournal platform. OLH is outspoken about its objective of making research in the humanities (which is generally unfunded) more affordable by providing an alternative to both subscriptions and APCs. Annual charges are between $600 and $2500 per library, depending on size and geographic region. Assuming 200 participating libraries and more than 250 articles are published per year, the cost per published article per institution is around $3.70.[128]

Another notable example is SCOAP3 (Sponsoring Consortium for Open Access Publishing in Particle Physics)[129], a consortium led by CERN that supports the conversion of the leading titles in high-energy physics from subscriptions to open access. SCOAP3 centrally pays publishers from a common fund at CERN, to which libraries, library consortia, research institutions and funding agencies jointly contribute. All articles funded by SCOAP3 appear in the SCOAP3 repository upon publication, alongside the publishers' own platforms. To control costs, SCOAP3 and the publishers agree a fixed maximum yearly payment for the entirety of articles published in each journal, commensurate with an estimated number of articles.[130] The initiative has been largely successful as most leading titles have been converted, following a process of soliciting library pledges and a subsequent tender. However, after the first phase (2014-2017), prolonged negotiations with publishers for a second phase (2017-2019) saw two journals drop out and three new ones join. These complications, for a consortium involving just a handful of journals, make it hard to see how such a model could easily be scaled up. SCOAP3 now involves 11 journals from six publishers and has published 19,116 OA articles since its inception in 2014. Articles funded by SCOAP3 will be available open access in perpetuity, under a CC-BY licence, while publishers will reduce their subscription fees accordingly.

[127] http://palaeo-electronica.org/owner.htm

[128] https://about.openlibhums.org/wp-content/uploads/LPS.pdf

[129] http://scoap3.org/

[130] https://scoap3.org/phase2-journals/

Table 5

Model / strategy	*Description*	*Examples*[131]
Sponsorship (no-APC models)	Sponsors cover costs with no intention to adopt APCs	The Journal of Electronic Publishing (JEP) The Berkeley Planning Journal Electronic Transactions on Numerical Analysis (ETNA) Clinical Phytoscience (Springer) Asia & the Pacific Policy Studies (Wiley)
Consortia funding	Creation of new funding consortia for the collective funding of open access publishing	Open Library of the Humanities SCOAP3 OpenEdition Knowledge Unlatched (monographs)

The sustainability question for no-APC journals and platforms

Gold no-APC journals and platforms' reliance on sponsorships, subsidies, volunteer labour and other kinds of external support makes it difficult to scale their operations, as higher submissions result in increased costs that are not necessarily matched by increased revenues. Individual publications therefore often rely on digital platforms that either host Gold no-APC journals or publish OA articles directly, whose economies of scale and stable funding streams provide greater financial stability (Johnson et al 2017).

Many gold no-APC journals and platforms struggle to develop a sustainable and scalable business model. One of the unresolved issues with this model is the "free-rider" problem, the fact that libraries have access to a research publication irrespective of their contribution. Many libraries are keen to support open access initiatives, but without some restraining features (e.g. multi-year contracts), such contributions are at risk of falling prey to budgetary constraints, especially where the alternative is to cut subscriptions to other, equally valuable, content.

Aside from grants and subsidies, no-APC journals and platforms have deployed a variety of revenue sources, including advertising, commercial reprints, subscriptions to print editions, crowdfunding and premium services. Among them, the sale of additional services under a freemium model has demonstrated some scalability. The OECD (Organisation for Economic Cooperation and Development) uses freemium open access to publish its research results (books, working papers, datasets and journal articles), while OpenEdition has adopted the same model for access to electronic resources in the humanities and social science. The basic service makes content available in a read-only version and includes discovery, citation, sharing and embedding tools. Subscription to premium services gives access to copy-paste, download, local printing, and file download in various formats (PDF, Excel, ePub). Premium subscribers may also benefit from downloadable MARC records, usage reports, customer support and on-site training. To date the freemium model has not any gained traction amongst STM journals, however.

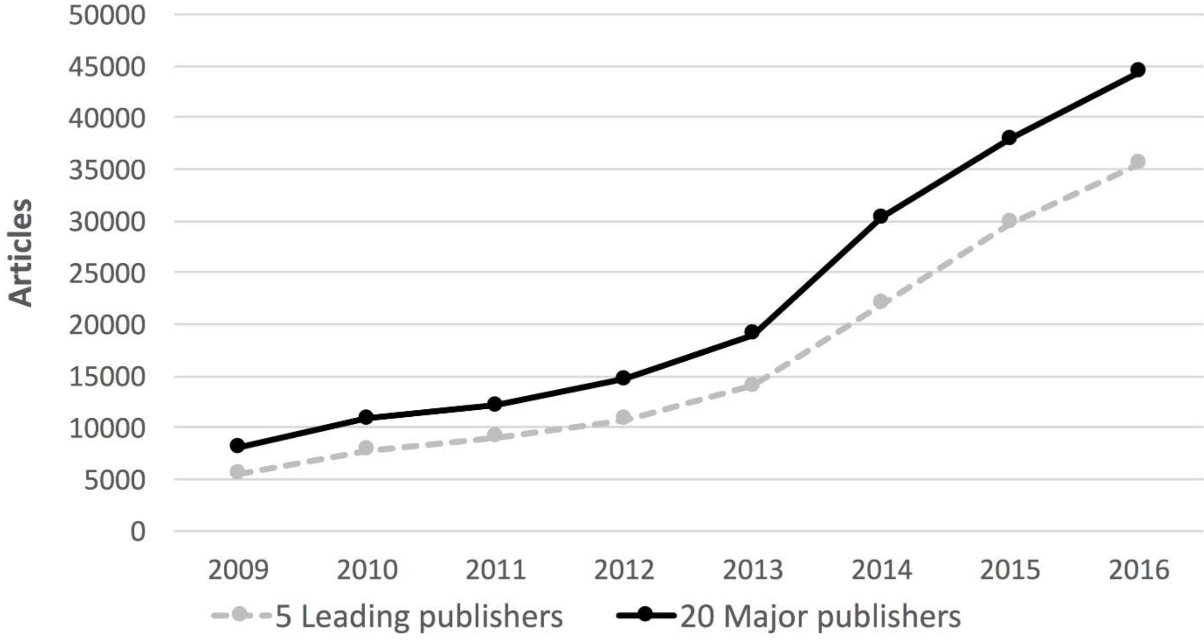
131 For a longer list of examples see http://oad.simmons.edu/oadwiki/OA_journal_business_models

3.2.3 Gold – Hybrid

The Gold - Hybrid model combines supply-side payments at article level (in the form of APCs levied on authors, institutions or funders) with demand-side payments at journal level (paid for by libraries). From the perspective of publishers, the model provides a relatively low-risk way for established subscription journals to make at least some of their content open access without jeopardising existing revenue models. Gold – Hybrid allows the market (i.e. authors, or their funders) to decide what value they place on open access, but is often overlooked in surveys of the landscape, which tend to focus on journal level data.[132]

Nearly all the major journal publishers, both commercial and not-for-profit, now offer hybrid schemes. A recent UUK analysis shows that 45% of the journals listed in Scopus now use a Gold - Hybrid business model (UUK 2017). Bjork (2017) estimates that, between 2009 and 2016, the number of journals offering the hybrid option increased from around 2,000 to almost 10,000 and the number of hybrid articles grew from 8,000 to 45,000. Nevertheless, shifting from a subscription to a hybrid business model has not generally resulted in widespread uptake of the OA option by authors, perhaps reflecting the relatively low level of author interest in OA (see section 2.14.3 *Author perceptions of, and attitudes towards open access*). Piwowar et al (2018) found that only 4.3% of articles indexed in Web of Science from hybrid journals are made open access.

Figure 31 – The evolution of hybrid articles (Bjork 2017)



Older studies reveal that uptake of hybrid articles differs substantially across disciplines. Bird (2010) showed that while overall uptake was 6%, this ranged from 2% in the humanities and social sciences, through 4% in medicine, 6% in mathematics, to 10% in the life sciences. Individual journals, however, had much higher percentages (for instance, Nature Communications had 40% uptake of hybrid articles, which no doubt was a major consideration behind the subsequent decision to make the title fully open access).

[132] See for example the European Commission's Open Science Monitor: https://ec.europa.eu/info/open-science/open-science-monitor/trends-open-access-publications_en, which at the time of writing does not fully reflect the level of hybrid OA, and Walt Crawford's (2018) definition of 'serious OA' as limited to those titles included in the Directory of Open Access Journals.

Journals employ a variety of revenue sources. For instance, The BMJ makes its research articles immediately available through APC payment (around £3000/$4800 per article) but requires a subscription to access other "value added" content such as commissioned review articles, editorials etc. Moreover, in 2016 The BMJ Journal generated £2.77 million revenue from product advertising, reprint sales and commercial sponsorship (12% of total revenues).

Offsetting deals

A common objection to the hybrid APC model is what has been termed 'double dipping' - the perception that publishing a hybrid journal means getting paid twice, once by the author in return for the publishing service, and then again by the subscriber in return for access to the article (Anderson, R., 2018). Publishers have long refuted this, with many issuing formal policies or statements outlining their approaches,[133] while Kent Anderson (2017) has argued that the multiple revenue streams involved in hybrid publishing can act to spread costs and decrease the burden on any single payer. In practice, while publishers have globally discounted the subscription rates of journals which also benefit from APC revenues for a number of years, uneven take up of hybrid to date has seen some countries and institutions experience increased costs (UUK 2017). Meanwhile the corresponding savings on global subscriptions are widely distributed and may be obscured by price changes arising from inflationary pressures, increasing article volumes and a range of other factors.

In response, several publishers have entered into local offset agreements designed to reduce the overall cost faced by research organisations or consortia. Under an offset agreement, open access publication costs are offset by lower subscription costs. There are different approaches to achieving this. Some offset agreements reduce the cost of APCs and some reduce the amount an institution pays for a subscription in proportion to the amount it pays for APCs. Some publishers offer credits against future APCs when subscriptions are taken out; others offer credits against future subscription payments when APCs are paid; a third approach bundles subscriptions with future APCs for modest additional payments.

These approaches to offsetting have been monitored in the UK through a review of offsetting deals between six major publishers (Wiley, Taylor & Frances, Springer, SAGE, Institute of Physics and Royal Society of Chemistry) and a consortium of 38 UK universities. Lawson (2015, 2016)[134] found that all deals focused on the total combined expenditure and estimated the total value of offsetting for the UK higher education sector at £8 million in 2016 (an average of 13.8% of the total publication costs). However, the combined costs of subscriptions and APCs rose by an average 11% per annum between 2013 and 2016 (UUK 2017) - suggesting that the value of offsetting for participating institutions should at best be considered *cost avoidance* rather than *costs saving*.

The Jisc analysis (Lawson 2016) also showed that deals providing offset credits for APC spend claimed against subscription costs (such as the Wiley deal) benefit research-intensive institutions more, whereas offset deals that provide both a discount on APCs and access to a large collection of journals (such as the Springer deal) are beneficial for less research-intensive institutions as well.

[133] See for example those of Elsevier: https://www.elsevier.com/about/policies/pricing#Dipping, Oxford University Press: https://academic.oup.com/journals/pages/open_access/oxford_open_faqs#eleven and Sage: https://uk.sagepub.com/en-gb/eur/subscription-pricing-for-hybrid-journals

[134] The data collection exercise, managed by Jisc, aimed to identify the "Total cost of ownership" of scholarly communication

Table 6 UK offsetting deals and discounts on publication (adapted from Lawson, 2016).

Publisher	*Deal*	*Total cost of publication*	*Offset against total cost of publication*
Wiley	25% discount on APCs plus 'offset credits' on total APC spend	£13,486,424	4.2%
Taylor & Frances	75% discount on APCs	£7,728,645	5.3%
Springer	APC waiver when publishing on 1,600 Springer journals and access to 2,500 journals	£6,712,913	36%
SAGE	APC reduced to £200 per article	£3,982,992	3.5%
IOP	90% of APC expenditure is offset	£1,417,690	9%
RSC	APC waiver on all RSC journals	£1,085,259	30%

Approaches to collective negotiations have been tried elsewhere, with mixed success. Notably, the Alliance of Science Organisations in Germany commissioned an initiative to conclude nationwide licensing agreements for the entire portfolio of electronic journals from major academic publishers from the 2017 licence year. The initiative, known as Projekt DEAL,[135] is now supported by a large consortium of German Universities and aims to relieve the financial burden on individual institutions through collective negotiation. Its approach marks a decisive change in academia-publisher subscription negotiations.

Although Projekt DEAL has reached an impasse in the negotiations with Elsevier at the time of writing, as has a similar initiative in Sweden,[136] it has inspired other European countries to explore similar approaches. In December 2017, the French Ministry of Higher Education, Research and Innovation, EDP Sciences and Couperin.org the national consortium for academics in France, signed a 5-year open access deal. The deal allows authors from French institutions participating in the agreement to publish their articles under an open access CC-BY license in a large selection of EDP Sciences journals regardless of whether their institution has a current subscription to the publisher.[137] In January 2018, the Finnish university consortium FinELib signed a three-year agreement with Elsevier for nationwide subscription access and open access. The deal applies to 13 Finnish universities, 11 research institutions and 11 universities of applied sciences, and has a total value of around

[135] https://www.projekt-deal.de/about-deal/
[136] https://www.timeshighereducation.com/news/german-and-swedish-libraries-shrug-elsevier-shutdown
[137] http://tagteam.harvard.edu/hub_feeds/119/feed_items/2335383

€27 million. The deal grants subscription access to around 1,850 journals and a 50% discount on APC cost.[138]

Offsetting has yet to gain significant momentum in North America, where the decentralised nature of higher education results in no common policy or incentive structure to govern such developments (Schonfeld, 2018). Nevertheless, a read-and-publish agreement was signed in 2018 between MIT and the Royal Society of Chemistry, and the University of California Libraries (2018) have established an Offsetting Task Force to pursue offsetting agreements, including with “big deal” publishers like Wiley, SpringerNature, and Elsevier.[139]

3.2.4 Delayed (open) access

Under this approach, 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 access policies, though for many OA advocates the delay, and the ability of the publisher to revoke access at any time, means this approach cannot be considered 'true' OA. The best known proponents of delayed access 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 over 2.4 million full-text articles that are open access, of which the majority are from delayed access journals.[140]

The delayed OA business model depends on the embargo period being long enough not to compromise subscription sales, so is typically adopted in rapidly developing and competitive fields. A 2013 study 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 average citation rates that are twice as high as 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).

A more recent study by Piwowar et al (2018) reaffirms the significance of delayed OA but notes that such articles are frequently made free-to-read, but without an open licence. They refer to this content as ‘bronze open access’, and report that in 2016 bronze articles accounted for more than 40% of all research articles. Because access can also be time-limited or restricted to a certain version of the article (such as pre-print or accepted manuscript), it is questionable whether these articles represent a *sustainable* source of open access literature.

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. The arguments on these points are essentially the same as those applied to self-archiving, with the caveat that delayed access to the version of record seems more likely to lead to cancelled subscriptions than access to authors' accepted manuscripts.

3.2.5 Open access publishing platforms

As well as categorising business models, it may also be helpful to describe the main types of OA publishing models. Most open access articles are published in ‘standard’ topical research journals using an OA or hybrid business model. OA business models are increasingly the

[138] http://finelib.fi/finelib-and-elsevier-agreement-access-to-scholarly-journals-and-50-percent-discount-of-article-processing-charges/
[139] https://libraries.mit.edu/news/royal-society-chemistry-3/27769/
[140] http://highwire.stanford.edu/lists/freeart.dtl

norm for new topical journals. Older and more established titles tend to use the hybrid business model, whilst 'flipped' journals are still relatively uncommon.

However, the spread of open access has also been boosted by the development of alternative publishing platforms. This section will discuss two examples of such platforms: megajournals and funders' platforms.

Megajournals

A 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 to have had significant widespread impact. Wakeling et al (2017a) identify some common characteristics of megajournals.

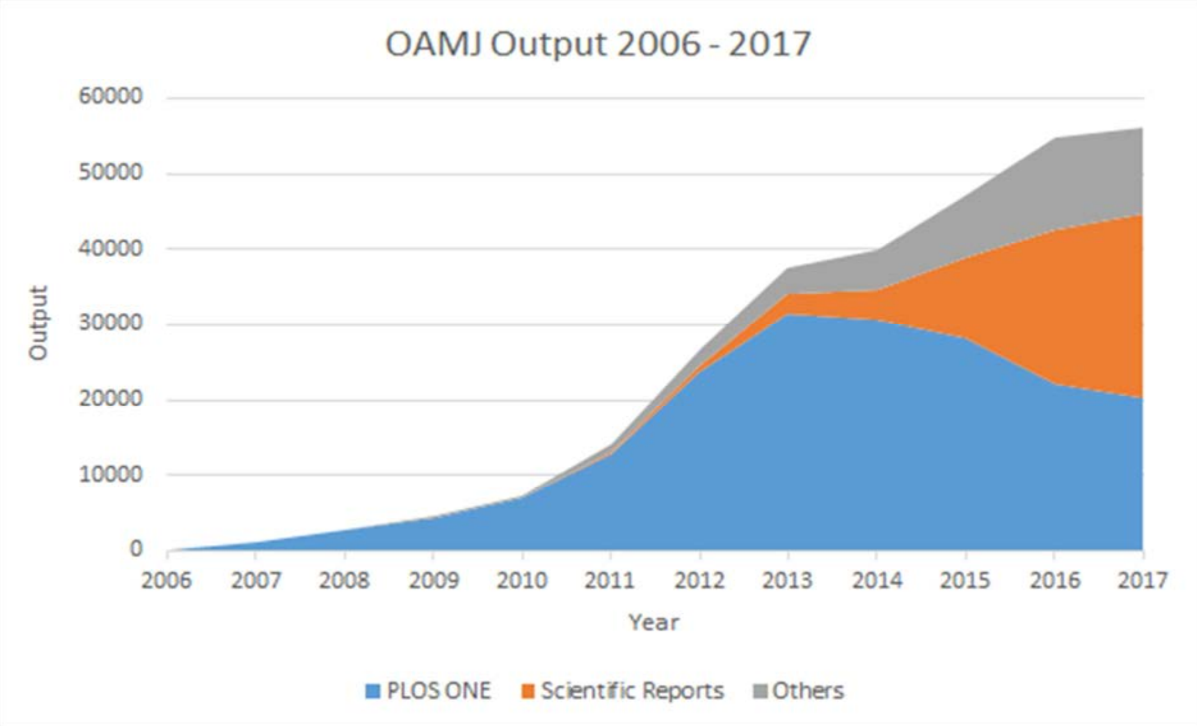
First, megajournals have a broad subject scope, encompassing either multiple disciplines or a single large discipline such as medicine or physics (the authors note the emergence of 'mini megajournals' that have a narrower disciplinary scope). Secondly, they use a rapid "non-selective" peer review based 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. Megajournals typically have a policy of keeping review straightforward, for instance avoiding where possible requests to conduct additional experiments and resubmit. In addition, the model has been associated with the 'cascade' peer review model, with the resubmission of articles rejected by selective journals to the same publisher's megajournal. Wakeling et al (2017a) noted a clear consensus that the selling point of the cascade model for authors was the efficiency of the submission process.

Third, they note that a large article output is often perceived as a key trait of megajournals, but is in fact secondary as not all megajournals are 'big'. In 2017, the combined outputs of 11 prominent mega-journals totaled 56,152 articles (see fig 32 below). However, while PLoS and Scientific reports publish several thousand articles each year, most other mega journals only publish a few hundred articles (Spezi et al 2016).

Fourth, megajournals are open access and charge relatively low APCs (reflecting a 'lighter' peer review process and a business model based on quantity). Spezi et al (2016) note that mega-journals' average APC fee of US$1,300 is more than the average APC for fully OA journals but substantially less than that of top ranking fully OA journals (ranging from US$2,500 to US$5,000) or hybrid journals (US$3,000).

There are examples of megajournals adopting alternative revenue models, such as PeerJ's $99 pay-once-publish-for-life membership plan. However, PeerJ has recently pivoted towards a more typical traditional APC and institutional-support business model, in conjunction with an editorial-driven publishing model (Davis, 2018).

Figure 32 Open access megajournal output 2006-2017 (Updated from Spezi et al, 2016, using Scopus data)



The success of the megajournal model has led to widespread emulation by other publishers. Figure 32 shows a rapid growth in article outputs across a sample of 14 megajournals, although there are signs the rate of increase is now slackening. Furthermore, the growth rate of individual titles differ, with Scientific Reports marking the most rapid growth while PLOS One has seen a sharp decline in overall output, partly explained by a lower acceptance rate.[141] Evidence to date suggests the megajournal is unlikely to become the dominant form some commentators predicted in the early part of this decade.[142] Bjork (2015) has observed that a lack of motivated, unpaid reviewers , and a lack of branding for authors, are likely to constrain further growth of the megajournal format. Nevertheless, the constant increase in research output around the world suggests that we will see a continued increase in the number of megajournals over time (Anderson, 2018).

Open access publishing platforms

A recent development in the open access landscape is that of open access, open peer-review scientific publishing platforms. These platforms allow for articles to be published first and peer reviewed after publication by invited referees, and make the peer reviewer's names and comments visible on the site. As part of an open science model, the data behind each article are also published and are downloadable. OA platforms are intended to accelerate the publication of open access articles and open data.

[141] https://scholarlykitchen.sspnet.org/2017/04/06/scientific-reports-overtakes-plos-one-as-largest-megajournal/?utm_source=feedburner&utm_medium=email&utm_campaign=Feed%3A+ScholarlyKitchen+%28The+Scholarly+Kitchen%29

[142] Peter Binfield, then publisher of PLOS One, predicted in 2011 that ‘in 2016, almost 50% of the STM literature could be published in approximately 100 mega journal’, see https://www.slideshare.net/PBinfield/ssp-presentation4

The primary example is F1000Research, an organisation that created a platform for life scientists, offering immediate publication of articles and other research outputs without editorial bias. The organisation then expanded its offering creating F1000Prime, which publishes article recommendations in biology and medicine from around 6,000 scientists and clinical researchers and over 5,000 junior researchers. Author uptake to date has been limited, with the platform publishing less than 700 articles in the 2016 and 2017 years, according to Scopus data. Nevertheless, F1000Research has attracted high levels of interest, particularly from research funders (see below). Emerald became the first publisher to partner with F1000, announcing in July 2018 that it would develop its own open research publishing platform – Emerald Open Research – using the technology, infrastructure and editorial services provided by F1000.

Funder publishing initiatives

A quite different approach is represented by the launch (in late 2012) of the journal *eLife*[143] 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 (see also *2.13.1 Open access and possible cost savings*).

Building on the technology of open access platforms such as F1000 Research, research funders such as the Gates Foundation and the Wellcome Trust have also started creating proprietary publishing platforms as a way of disseminating their own research. Wellcome Open Research[144] aims to publish articles within a week after submission following a basic check by an editorial team. The article is then available for immediate viewing and citation. At the same time an open peer review system and user commenting ensure the quality of the research. Reviews are published alongside the article, which can then be revised by the author for further review. Currently the platform hosts 238 articles from Wellcome-funded researchers, also accepting methods, software tools systematic reviews and data notes.

In recent months, other funders such as the Irish Health Research Board have adopted the F1000 platform, while the European Commission has issued an invitation to tender for a publishing platform for scientific articles as a service for Horizon 2020 beneficiaries. Such platforms offer the potential to increase OA uptake, control costs of OA, lower administrative burden on researchers, and demonstrate funders' commitment to fostering open practices. However, their long-term success will depend on funders resolving concerns over potential conflicts of interest, difficulties of scale, potential lock-in and issues of the branding of research (Ross-Hellauer et al, 2018).

3.2.6 Open Access Books

The penetration of open access in the academic book market is considerably slower compared to scholarly journals for two interconnected reasons. On one hand is a supply-side problem: digitisation in this market is an incomplete revolution, with many academic books and monographs still not available in a digital form. On the other hand is a demand problem, with university libraries and academics still largely relying on print versions. Nevertheless, changes in reading technology and (such as enhanced eBooks) and the digitisation of research workflows is changing the market. Increase book digitisation has spearheaded a gradual move towards an acceptance of open access among authors (Watkinson 2016).

Springer has created a portfolio of over 200 peer-reviewed, fully open access journals and over 500 open access books covering all areas of science, technology, medicine, the humanities and social sciences, called SpringerOpen. A report commissioned by the

143 http://www.elifesciences.org/

144 https://wellcomeopenresearch.org/

publisher showed a correlation between open access and as much as seven times more downloads (Springer Nature 2017).

Despite strong year-on-year growth, the OA book market is still less than 1% of all scholarly and professional e-book publishing: according to some estimates there were only around 10,000 titles in 2016, with humanities and social sciences (HSS) accounting for almost three quarters of all OA books published. In the US, support for humanities publishers from the Mellon Foundation has greatly advanced open access among humanities book publishers. Among the more successful OA book publishers are The University of California Press[145] and the University College London Press (UCL Press), launched as an open-access only press and now offering services to other presses with the same aims.[146] Among commercial publishers, Ubiquity Press has an important role in supporting university and society-based publishing and driving growth in the sector. Ubiquity also works outside English-speaking countries where growth in open access publishing among not-for-profits is slower.

Commercial publishers also produce open access books. Pioneers like Bloomsbury Academic originally built their business models on the idea that sales of print books would be increased by offering a digital open access equivalent.[147] However, the dominant business model became charging Book Publishing Charges (BPC). A typical BPC is currently around £10,000. Palgrave, which is the humanities and social sciences imprint of Springer Nature, have a similar arrangement: their BPC is £11,000. Springer, Taylor & Francis/ Routledge also have OA book options.

These charges and the current lack of funding at this level for book authors is hampering the spread of open access in the academic book market. In the UK, renewed attention has been paid on the subject since the Research Excellence Framework has confirmed that monographs submitted for assessment must be open access. The British Academy (2018) has explored this problem at length without any clear conclusion. The solution that is currently viewed as most attractive is by pooling library resources to provide alternative publishing paths for OA books. Started by Knowledge Unlatched in 2012, this model relies on libraries paying to make a book open access on a dedicated publishing platform.[148] A number of publishing platforms for open access books have since emerged in the market, each with its own distinct business model. Among them, the French platform OpenEdition has acquired a leading position with a catalogue of over 6,000 e-books in the humanities and social sciences, most of which are open access. It sustains its operation through a freemium model, whereby books are made freely available online but libraries can choose to pay for premium services such as downloadable file formats. Other platforms in the sector have smaller catalogues, ranging from a few hundred to just over 2,000, but the platform sector is steadily growing its catalogue and readership.

3.3 Open access via self-archiving

The "Green" route to open access is by self-archiving the accepted manuscript of the article, and making it publicly accessible, either immediately upon publication or at the end of an embargo period set by the publisher. Self-archiving has no independent business model, and it relies on the assumption that making accepted manuscripts freely available will not compromise the sales of versions of record.

Self-archiving simply consists of depositing the full text of the accepted manuscript in an open repository or other open archive. Where the full text is not deposited, the metadata of the article is deposited alongside a link to the full text (e.g. as hosted on another repository

[145] https://www.luminosoa.org
[146] http://www.ucl.ac.uk/ucl-press
[147] https://creativecommons.org/2008/10/20/an-interview-with-frances-pinter-of-bloomsbury-academic/
[148] http://www.knowledgeunlatched.org

or on the journal's website). Self-archiving is traditionally done by the author but is increasingly undertaken by others on their behalf, such as publishers, librarians or research support staff.

Typically, repositories are run by research institutions to preserve and present their own research (institutional repository), or they aggregate and present relevant research in a given discipline (subject repository). Additionally, repositories can be created by research funders and government agencies for the purpose of aggregating and presenting the research they have funding. Other online postings by authors in academic social networks, on their personal websites and on third-party websites are not typically considered as 'OA self-archiving', but have become an increasingly important means of authors sharing content in recent years.

3.3.1 Global uptake of Green OA

OpenDOAR estimates that a total of 3,520 repositories are currently active and curated.[149] It also found that this number has significantly increased over the past four years, continuing a trend that had already been observed in the years prior to 2014. Growth was recorded in each of the following categories (data as of April 2018):

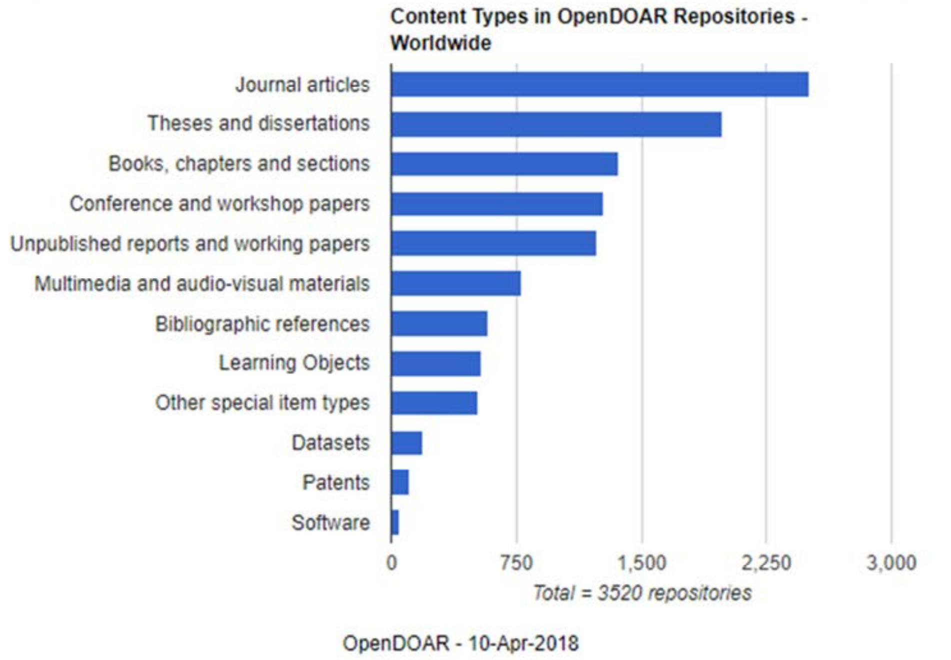
- Institutional repositories: 85.9% (3,023 repositories, up from 2,257 in 2014)
- Disciplinary or subject-based repositories: 8.6% (304 repositories up from 296 in 2014)
- Aggregating: 3.1% (108 up from 98 in 2014)
- Governmental: 2.4% (85 up from 77 in 2014)

In absolute terms, growth was lowest among disciplinary repositories and highest among institutional ones, with as many as 766 new repositories added to the database since December 2014.

Repositories contain a wide variety of content types, including: journal articles, theses, book chapters, working papers, conference papers, monographs and data. However, while most repositories listed in OpenDOAR (2,505) accept journal articles, only half this number contain non-peer reviewed publications (books, conference papers, working papers etc).

[149] http://v2.sherpa.ac.uk/opendoar/

Figure 33: repository content by type (source: OpenDOAR)



3.3.2 Institutional repositories

An institutional repository (IR) 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. author's original and accepted manuscripts), and digital versions of theses and dissertations, but it might also include other digital assets generated in the course of normal academic life, such as administrative documents, course notes, or learning objects. IR software enables the article metadata to be harvested by dedicated search engines such as OAIster or Google Scholar. This distributed search allows users to find articles of interest regardless of which institutional repository hosts them.

The three main objectives for having an institutional repository are:

To increase the visibility of 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).

To help researchers comply with research funder open access mandates.

The spread of institutional repositories dates back to the early 2000s, with the development of Eprints software at Southampton and the launch of DSpace at MIT. As seen above, the total number of IRs has grown rapidly since then and so have the number of article deposits and downloads (UUK 2017). However, article downloads from institutional repositories are still very low relative to levels of traffic on publisher platforms and academic social networks. Whereas a few years ago researchers had no or little knowledge of, or experience with, institutional repositories and were unfamiliar with self-archiving opportunities (Singeh 2013, Wallace 2012), more recent studies showed that academics are becoming familiar with the concept but remain disengaged from the post-publication process (Odell et al 2017). Harnad (2015) argues that IRs are created more in reaction to downward pressure from research funders or institutional management than upward pressure from authors, and the adoption of institutional mandates is necessary to achieve higher deposit rates.

For researchers, IRs are arguably less attractive than subject repositories and social networking sites, which are often used as discovery tools bringing clear visibility benefits within their epistemic community. By contrast, IRs are rarely used in research and they are often not optimised for discoverability. Author incentives to archive their research in such repositories are therefore limited. 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 who see their role in terms of scholarly communication and publishing (Albanese 2009, Poynder 2016).

3.3.3 Subject repositories

Subject repositories have been around for much longer than institutional repositories. Björk reviewed the status of subject repositories in 2013, concluding that they catered to a strong market demand when they first emerged but the development of Internet search engines 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, attracting growing numbers of citations both within and beyond the core disciplines they serve (Li et al, 2015).

This section summarises three examples of subject repositories with wide variations in scope, function and cost - arXiv, RePEc, and PubMed Central. arXiv contains mainly authors' accepted manuscripts; RePEc is essentially an indexing service over some 1,600 repositories; PMC is a highly centralised database, an 'electronic library' that converts various input formats into structured XML and clear PDF for digital preservation.

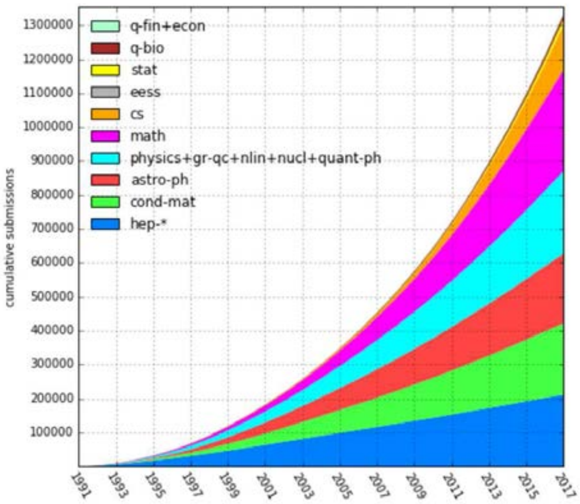
arXiv

One of the earliest subject repositories was arXiv, [150] established in 1991 at Los Alamos by Paul Ginsparg and now hosted by the Cornell Computing and Information Science, following its move from the Cornell University Library in September 2018. arXiv (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 other areas such as mathematics, computer science, quantitative biology, statistics or economics albeit with less comprehensive coverage. It currently holds over 1.3 million preprints, receiving around 10,000 submissions per month.

As the arXiv has grown (Figure 34) its host organisation (now Cornell, originally LANL) has struggled to sustain the funding requirements. In January 2013 arXiv therefore introduced a new funding model[151] 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 for a five-year period. The membership costs are matched with $300,000 per year by the Simons Foundation to keep membership fees low. arXiv currently counts over 210 institutions as its members. Beginning in 2018, arXiv added a fourth source of revenue in the form of grant funds from foundations and agencies to assist with special projects.[152]

[150] http://www.arxiv.org/
[151] See http://arxiv.org/help/support
[152] https://confluence.cornell.edu/display/arxivpub/2018-2022%3A+Sustainability+Plan+for+Classic+arXiv

Figure 34: Growth in arXiv; physics and maths remain the most important subjects (Source: arxiv.org, accessed 17th April 2018)[153]



RePEc

RePEc (Research Papers in Economics)[154] 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 2.3 million research pieces from 2,800 journals and 4,500 working paper series. It differs from arXiv in several ways: first, it is a decentralised (and volunteer-based) bibliographic database rather than a centralised repository, integrating content from some 1,900 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.

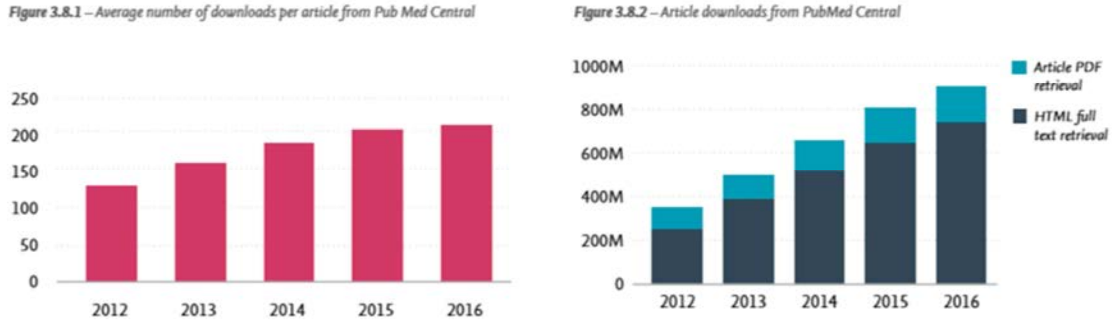
PubMed Central

A subject repository of great current interest to publishers is PubMed Central (PMC), currently the largest single subject repository. The number of articles available on PubMed Central rose by 56% from 2.8 million in 2012 to 4.4 million in 2016. But the number of downloads rose even faster by 157%, as shown in Figure 35. As a result, the average number of downloads per article rose from 127 to 209.

153 https://arxiv.org/help/stats/2017_by_area/index

154 http://repec.org/

Fig. 35 - Article downloads and downloads per articles on PubMed Central (UUK, 2017)



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 or in some cases the full text, which can be made available immediately (for full open access journals) or after an embargo period (for delayed open access journals). PMC has also worked with publishers to digitise back content, which must then be made freely available. Since 2004, PMC has taken accepted manuscripts from authors for archiving in support of the NIH funding policy discussed below (see *3.4.1 Funders' policies on open access*).

At the time of writing there were 4.8 million research articles hosted on PMC, of which 1.94 million were in the open access subset (the others are freely available but not open access in the sense used by PMC - released under a Creative Commons licence permitting redistribution and reuse). Europe PubMed Central,[155] formerly UK PubMed Central, is based on PubMed Central with some additional services and functionality (McEntyre et al., 2011). PubMed Central Canada was launched in October 2009, and provided a location for Canadian Institutions of Health Research (CIHR)-funded researchers to deposit peer-reviewed articles. It was permanently taken offline on February 2018, on the basis that only 4% of author manuscripts arising from CIHR research had been deposited since its creation, and the time and resources needed to upgrade the system to meet Government of Canada web and security standards were prohibitive. All manuscripts were copied the National Research Council's digital repository.[156]

3.3.4 Aggregators and indexing services

CHORUS

CHORUS – Clearinghouse for the Open Research of the United States – is an aggregator (meta-repository) that collects the metadata of documents resulting from federally funded research. It provides an access layer on top of existing publishing platforms that links to freely accessible journal articles. It was formed by a group of publishers and service providers, as a non-profit public-private partnership to develop a service that would enable funding agencies to meet the OSTP requirements.[157]

155 http://europepmc.org/

156 http://www.cihr-irsc.gc.ca/e/50728.html

157 The OSTP requires federal agencies in the U.S. with more than $100M in research expenditures to develop plans to make the published results of federally funded research freely available to the public within one year of publication. http://guides.libraries.psu.edu/c.php?g=431943&p=2949762

CHORUS provides five core functions: identification, discovery, access, preservation, and compliance with policy requirements. The service depends on the Crossref Open Funder Registry, a standard way to report funding sources for published scholarly. When adopted by publishers, it allows papers funded by federal agencies to be identified via CHORUS and to be made freely accessible to the public on the publishers' sites. Publishers considering participating in CHORUS will have to decide whether to make available the final version of record or the accepted manuscript (Cochran 2014) and if they want to choose to impose an embargo period of up to twelve months on the publication. CHORUS also allows agencies to create discovery portals to their content and offers "dashboards" to enable all stakeholders to monitor public-access compliance. In addition, CHORUS partners with Portico, CLOCKSS and other services to ensure the long-term preservation and public access to federally-funded papers.

At the time of writing, eight US Departments and agencies, plus the Japan Science and Technology Agency, have signed agreements with CHORUS, with pilots underway with others.[158] The largest US agency, NIH is committed to its own platform, PubMed Central. 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 approach is offered by SHARE[159] (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 (APLU). SHARE addresses the need for preservation and reuse of, and access to, research outputs. In contrast to CHORUS, which assumes that OA content will remain under the functional control of publishers, it rests on the idea that the content will be under the control of institutional repositories.

The service architecture consists of four layers: a notification service; registry; discovery; and mining and reuse. The first layer, "SHARE Notify", distributes notifications about research release events, in the form of a concise set of metadata to stakeholders such as funding agencies, sponsored research offices, institutional repositories, and disciplinary repositories. It also enables users to set up a feed to receive notifications upon the release of research matching specified criteria based on searches of SHARE's data set. Like CHORUS, SHARE does not store copies of research outputs but will maintain a registry of content that will subsequently support a discovery layer, including OSF Preprints: The Open Preprint Repository Network, launched in 2016.

RCAAP

The RCAAP (Repositórios Científicos de Acesso Aberto de Portugal) portal aims to collect, aggregate and index Open Access scientific contents from Portuguese institutional repositories.[160] It currently indexes 150,000 documents from 129 resources. The FCT (Fundação para a Ciência e a Tecnologia) - the Portuguese public agency that supports science, technology and innovation – currently requires all research funded by them to be made available in one of the open access repositories hosted within RCAAP. Embargo periods are possible, but articles should be made immediately available on RCAAP. In cases where the publisher does not allow self-archiving of the published version, the author may deposit his final version (Accepted Manuscript). [161]

---

[158] https://www.chorusaccess.org/resources/chorus-funder-participants/

[159] http://www.arl.org/focus-areas/shared-access-research-ecosystem-share

[160] https://www.rcaap.pt/about.jsp

[161] https://www.fct.pt/acessoaberto/

RCAAP also provides support for the setting up of institutional repositories and a hosting service for institutional repositories (SARI). For research affiliated institutions in the national scientific system who do not have their own repository, RCAAP offers the *Common Repository* to archive their results. In addition a repository validator ensures that all repositories listed comply with RCAAP standards.[162]

CAS IR Grid

The CAS IR Grid (Chinese Academy of Sciences Institutional Repository Grid) is an integrated platform linking some 100 existing institutional repositories operated by CAS research institutes, providing author self-archiving for high-quality publishing papers. Set up to improve visibility, discoverability and ease-of-use of the OA resources archived by individual institutional repositories, the Grid also assists CAS in monitoring OA compliance. English-language journal articles account for one third of the full text items (Montgomery and Ren 2018) and at the time of writing, the CAS IR grid contained almost 1.2 million resources.

OpenAIRE

The EC-funded OpenAIRE project has led a number of initiatives to create a connected open access infrastructure in Europe. In 2009, the first OpenAIRE project was funded to enable researchers to deposit their FP7 and ERA funded research publications into Open Access repositories. After a three-year pilot, the project was expanded in 2011 to attract data providers from domain-specific scientific areas. The updated platform, branded OpenAIREplus, created a heavy-duty and collaborative service for the cross-linking of peer-reviewed publications and associated datasets. The project also created a set of guidelines for the creation of national repositories and stimulated the development of OpenAIRE-compliant open access infrastructure in several European countries.

The fourth phase of the OpenAIRE project – OpenAIRE-Advance – is currently underway, aiming to consolidate and optimize existing services, to support citizen-science and to work together with the EOSC-hub project. OpenAIRE also developed, documented and published open APIs[163] allowing developers to access the metadata information space of OpenAIRE programmatically. Moreover, OpenAIRE is running the open source infrastructure framework D-NET Software Toolkit which assists developers in the construction and maintenance of aggregative data infrastructures (i.e. repository aggregators) through services for the collection, processing and provision of metadata and files.

COAR

COAR (Confederation of Open Access Repositories) is an international association with over 100 members and partners from around the world, aiming to bring together repository networks and the repository community.[164] Inter alia, COAR intends to raise the level of interoperability between repositories to ensure an open science infrastructure. Initiated in 2009, the COAR Interoperability Project led a number of initiatives including controlled vocabularies for bibliographic metadata, open metrics (to measure the impact of scholarly publications) and the alignment of repository networks to create one seamless global repository network, culminating in the COAR Roadmap: Future Directions for Repository Interoperability.[165] More recently, the *COAR Next Generation Repositories Working Group* has sought to identify the core functionalities needed to position repositories as the foundation for a distributed, globally networked infrastructure for scholarly communication. Its

162 http://projecto.rcaap.pt/index.php/lang-en/sobre-o-rcaap/enquadramento
163 api.openaire.eu
164 https://www.coar-repositories.org/about/coar-ev/
165 https://www.coar-repositories.org/activities/repository-interoperability/

conclusions are outlined in the report Next Generation Repositories: Behaviours and Technical Recommendations of the COAR Next Generation Repositories Working Group.[166]

SciELO

SciELO (Scientific Electronic Library Online) is not a conventional repository but a bibliographic database and a digital library of open access journals. 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 (now Clarivate Analytics) for the integration of the SciELO Citation Index into Web of Science.

Redalyc

Redalyc[167] (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 tools, aggregators and indexing services

A number of services have been developed to improve discovery and use of specifically open access content. This can be done by making metadata available in an aggregator (directory for articles or repositories) so the articles are easily accessible. While aggregators let users browse through their indexed entries, other services will present the user with a notification if an article they are looking for is available via open access. Tables 7-9 give an overview of the services available.

Table 7: Aggregators and databases with searching and browsing functions

Service and launch date	*Description*	*Supporting organization type and location*	*Records*
OAIster (2002)	Union catalog of millions of records that represent open access resources	Nonprofit, Global	> 50m records
BASE (2004)	Search engine for academic web resources	Academic institution, Germany	> 131m documents
CORE (2010)	Aggregator of OA content across different systems such as repositories and OA journals	Academic institution, UK	>134m OA articles
ScienceOpen (2013)	Search and discovery platform that puts research in context by leveraging smart filters, topical collections and input from the academic community	Private organization, Germany and USA	>44m article records
Scilit (2013)	Database of scholarly works, including workflows and services for publishers and data providers	Private organization, Switzerland	>111m articles (19.6m OA articles)

166 https://www.coar-repositories.org/files/NGR-Final-Formatted-Report-cc.pdf
167 http://www.redalyc.org/

Service and launch date	*Description*	*Supporting organization type and location*	*Records*
Paperity (2014)	Multidisciplinary aggregator of OA journals and papers	Private organization, Poland	> 1,7m OA articles
Dissemin (2017)	Service detecting papers behind paywalls and inviting their authors to upload them to an open repository	Nonprofit, France	>89m OA articles
1findr (2018)	Discovery platform including search functions, OA content, custom metrics and library integration services, also covering library subscriptions	Private organization, Canada	>90m records (>27m OA articles)

Table 8: Browser extensions that enable access to paywalled literature

Service and launch date	*Description*	*Supporting organization type and location*	*Records*
Open Access Button (2013)	Browser extension to discover OA copies of paywalled academic literature	Nonprofit, UK	Not applicable
Lazy Scholar (2014)	Browser extension to discover free copies of paywalled academic literature, also covering library subscriptions	Personal project, USA	Not available
Google Scholar Button (2015)	Browser extension to access Google Scholar from any web page and find full texts on the web, also covering library subscriptions	Private organization, USA	Not available
Kopernio (2017)	Browser extension to discover OA copies of paywalled academic literature, also covering library subscriptions	Private organization, UK	Not available
Unpaywall (2017)	Browser extension to discover OA copies of paywalled academic literature, including library integration services	Nonprofit, USA	>19.6m OA articles

Table 9: Online directories for OA

Service and launch date	*Description*	*Supporting organization type and location*	*Records*
DOAJ (2003)	Online directory indexing and providing access to OA journals and articles	Nonprofit, Sweden	>3.2m articles
OpenDOAR (2005)	Directory of academic open access repositories	Nonprofit, UK	>3,700 repositories
DOAB (2013)	Directory of academic open access books	Nonprofit, Netherlands	>12,000 peer-reviewed books and chapters

3.3.5 Other repositories

Over the past few years, new types of repositories have emerged. In this section, we discuss the Social Science Research Network and two examples of funder repositories: a funder-backed repository accepting research from all sources (Zenodo); and a national repository where all national researchers are expected to archive their research (Digital.CSIC).

Social Science Research Network (SSRN)

SSRN can be classified as a subject repository, but its partnerships with a large fraction of 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 PubMedCentral. Founded in 1992, the platform today hosts over 800,000 research papers across 30 disciplines, though it has its roots in the long-standing preprint culture in economics. Authors can upload their papers to SSRN as green OA, but publishers and institutions are allowed to charge fees for downloading their SSRN papers, hence SSRN is only partially OA, and uses a different model to other subject repositories (Li 2015). SSRN was acquired by Elsevier in May 2016, and since then has pursued a deeper integration with Elsevier's other research products, particularly Mendeley.

Zenodo

Zenodo - launched in 2013 - is a general purpose open access repository, supported by the European Commission but available to all. Developed by CERN and the EC's OpenAIRE project, Zenodo aims to capture every step taken by researchers to ensure connectivity in the scientific community. Therefore, it accepts not only publications but also images, data, software, videos and every other digital artifact in connection with the research process, imposing no requirements on format, size or licence.[168] As at April 2018 the repository hosted 370,000 artifacts, of which 140,000 were publications, across over 1,800 communities. 94% of all items are publicly available without restrictions or embargoes. Restrictions can be imposed on various objects during the review process to ensure safe

[168] http://about.zenodo.org/

communication between researchers and reviewers. Zenodo makes research results citable (citation information is also passed onto scholarly aggregators) and integrates them through OpenAIRE into existing reporting lines to funding agencies.

Digital.CSIC

Digital.CSIC is the institutional repository of the Spanish National Research Council (CSIC), Spain's largest institution dedicated to research. It aims to preserve and provide access to CSIC research outputs. Founded in 2008, it contains 155,000 records from around 150 institutions and 1,300 collections. 62% of all records are open access as of 2018. Digital.CSIC is the result of the signing of the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities of CSIC in 2006 and allows all CSIC researchers to self-archive their work in the repository.[169]

3.3.6 Other factors affecting self-archiving

Two factors potentially affecting the spread of self-archiving are costs and uncertainty over the version of articles to be deposited, as well as other forms of outputs.

Costs of repositories

Swan (2016) identified two kinds of repository costs: the costs of building and running repositories (dissemination costs) and the costs of storing content and associated content migration and other technical procedures involved in long-term archiving (storage and archiving costs).

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), although the validity of these estimates has been heavily criticised in some quarters. There is no doubt, however, that large disciplinary repositories are substantially more expensive than institutional ones. For instance, the UK Data Archive cost around £3.43 million in 2010-2011 alone while the National Institutes of Health has estimated the cost of administering its self-archiving policy via PubMedCentral at around $4 million.[170] But economies of scale may permit lower per-article costs in larger repositories. For example, Cornell University Library estimated the 2014 annual running costs for the (highly automated) arXiv at $886,000, less than $10 per new article deposited. Meanwhile, the Croatian portal Hrcak (which hosts almost 185,000 full text articles) was developed and run for several years as a voluntary project by the University of Zagreb's Computing Centre and supported with just a €20,000 grant.

No updated studies on the cost of repositories have been forthcoming in recent years. 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, software upgrades and repository maintenance treated as sunk costs and not accounted for separately, and expenditures divided between multiple departments. From available estimates, however, it is clear that the largest cost item for repositories is staff (around 71% for the UK Data Archive and 82% for arXiv), while storage and computing infrastructure accounts for a small percentage of total costs (5% in the case of arXiv).

Article versions, data and other files

Another 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). As seen above, authors can self-archive either the author's original or the accepted manuscript, or in some cases both, whereas few publishers permit archiving of the version of record, unless the article is published under an open access licence. Most funder and

[169] http://digital.csic.es/dc/faqs/#faq3

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

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), thus creating further confusion. 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 is valuable here in distinguishing the version of record from other versions (see *Versions of articles* above).

3.4 The drive for open access

3.4.1 Funders' policies on open access

Uptake of open access has been driven significantly in recent years by interventions from and policies of research funders. Pressure from the top, combined with the growth and maturity of digital publishing technology, has spurred the development of credible open access publishing options for authors. Policy development has continued over the past few years with no sign of slowing down.

Open access policies have been adopted by a growing number of research funders and research organisations. The SHERPA/Juliet website[171] listed (as of July 2018) 147 research funders with an open access policy, including the UK Research Councils, the Wellcome Trust, the Howard Hughes Medical Institute, the National Institutes of Health, the European Research Council, the Chinese Academy of Sciences, the DFG and the Fraunhofer in Germany, and the Australian Research Council. The website is likely to underestimate the global number of funder policies, with UK and European funders over-represented in the dataset. Of the funder policies included, 122 had self-archiving provisions (varying from requiring to just encouraging open access archiving) and 104 had open access policies.[172]

In addition to research funders, many research organisations have also adopted their own policies. The Eprints/ROARMAP website[173] recorded 716 full institutional and 75 sub-institutional mandates in July 2018. High-profile institutions adopting mandates include Harvard, MIT, UCL, ETH Zurich, Fraunhofer-Gesellschaft, and the University of California.

European Union

In many respects, European countries and the European Union have been at the forefront of open access and open data policy development globally. The Amsterdam Call for Action on Open Science, the result of a conference on open science hosted by the Netherlands' EU Presidency in April 2016, contained several ambitious proposals to remove barriers to open science (including by changing the way research is evaluated and increasing transparency of publishing costs) and to further develop open access policies in Europe. It set a target of achieving 100% open access in Europe by 2020 and called on research funders to develop open access plans and create monitoring infrastructures and services.

In May 2016, the European Union Competitiveness Council endorsed the call for full open access to scientific publications in Europe by 2020.[174] The ambitious goal was reiterated in subsequent EC policy documents,[175] and complemented by a European Commission decision to make open access an obligation for its Horizon 2020 grantees. This was reinforced by the appointment of a special envoy on open access, tasked with making concrete policy recommendations to achieve full and immediate open access to scientific

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

172 It should be noted however, that some US funders classified as "requir(ing) open access archiving" in SHERPA/Juliet do not require the granting of reuse licences, and thus these policies might be better described as 'public access' rather than 'open access' policies.

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

174 http://www.consilium.europa.eu/media/22779/st09357en16.pdf

175 http://data.consilium.europa.eu/doc/document/ST-9526-2016-INIT/en/pdf

publications by 2020, a goal that is unlikely to be met.[176] Additionally, the Commission has issued a tender for the creation of an Open Research Publishing Platform that would offer the opportunity to Horizon 2020 beneficiaries to make their scientific publication immediately open access.[177] Under the leadership of the European Commission, and with support from pan-European initiatives such as OpenAIRE and others, many national research funders in Europe have been developing and implementing open access policies over the past decade.

In September 2018, 11 national research funding organisation, with the support of the European Commission and the European Research Council (ERC), announced the launch of cOAlition S,[178] an initiative to make full and immediate Open Access to research publications a reality. It is built around Plan S, which consists of 10 principles based on the key principle that:

"After 1 January 2020 scientific publications on the results from research funded by public grants provided by national and European research councils and funding bodies, must be published in compliant Open Access Journals or on compliant Open Access Platforms."

The plan includes a number of measures that are more radical than those seen in previous OA policies, including authors' retention of copyright over publications, and a statement that the hybrid model of publishing is not compliant with the stated principles. The plan covers all scholarly outputs, but recognises that the transition to OA for monograph will be achieved over a longer timeframe. At the time of writing, the application of some of the principles remains open to interpretation, and it remains too early to assess its long-term impact.

United Kingdom

Following the publication of the Finch report (Finch Working Group 2012), the United Kingdom has been one of the most active European countries in the development of open access policy. The two arms of its dual support research funding system, now combined under the umbrella of UK Research and Innovation (UKRI) have however taken somewhat different approaches. The Research Councils UK policy[179] was developed during 2012 and came into effect in April 2013. Unlike most other funder policies, it included a stated preference for Gold over Green open access, all else being equal and funds available. In order to cover the cost of article publication charges a block grant was provided to universities and eligible research organisations. A 2014 review of compliance with the policy in the first 18 months from its entry into force revealed that 94% of reporting institutions had exceeded the 45% open access target set by RCUK for the first year of implementation.

By contrast, the open access policy developed by HEFCE/Research England (and its equivalent funding bodies for Scotland, Wales and Northern Ireland) as part of the UK's Research Excellence Framework (REF) requires author to self-archive their research, albeit with an exception for those adopting immediate publication via the gold route. The 2021 REF OA policy requires all research outputs to be deposited as soon after the point of acceptance as possible, and no later than three months after this date. Concerns over institutions' ability to implement this requirement in practice led to an exception being introduced which allows outputs unable to meet this deposit timescale to remain compliant if they are deposited up to three months after the date of publication.

UKRI has recently launched a review of both the RCUK and REF open access policies, to run in parallel with a similar review by the Wellcome Trust. Although the reviews remain underway, there are indications that one area for consideration will be the increasing cost of

[176] https://horizon-magazine.eu/article/open-access-scientific-publications-must-become-reality-2020-robert-jan-smits_en.html
[177] http://ted.europa.eu/TED/notice/udl?uri=TED:NOTICE:141558-2018:TEXT:EN:HTML
[178] https://www.scienceeurope.org/coalition-s/
[179] http://www.rcuk.ac.uk/research/openaccess/policy/

open access, particularly with respect to hybrid journals.[180] As a signatory to Plan S (see above), it appears that UKRI will seek to phase out its support for hybrid OA over the coming years, with Wellcome highly likely to follow suit.

Other European countries

Across Europe, the vast majority of national research funders either have a dedicated open access policy in place or are aligned to an overarching national open access policy. A desk-based review of 40 European funders undertaken by the authors on behalf of SPARC Europe (internal report) found that 26 include a mandatory requirement to either immediately publish or self-archive the publication, while the remaining funders encourage doing so. Of those requiring or encouraging self-archiving, the vast majority accept an embargo period of either 6 or 12 months, with the largest group accepting 6 months for STM disciplines and 12 months for arts, humanities and social sciences. Twenty-nine funders allow for both delayed open access through self-archiving and immediate open access through the payment of APCs, whilst the remaining 11 funders only support self-archiving.

United States

The US has seen a number of developments in national legislation and policy designed to increase access to research. However, US support for OA publishing using the Gold OA model has been limited, and funder policies tend to require 'public access' as distinct from 'open access'. This means compliant outputs are made free-to-read, often after an embargo period, but are not free of copyright and licensing restrictions. Public access therefore accords greater freedom to authors (and their publishers) to choose how and in what ways their work will be reused.

The US was the first country to adopt a national access mandate with the Consolidated Appropriations Act 2008, the legislative basis for the public access policy of the National Institutes of Health (NIH) - the largest biomedical research agency in the world.[181] In February 2013, the White House's Office of Science and Technology Policy (OSTP) issued a policy memorandum (the OSTP Directive on Public Access) directing all federal agencies with R&D expenditures of over US$100 million to develop a plan to support increased public access to the results of research funded by the Federal Government.[182] More than twenty U.S. Federal Agencies have since developed public access plans.[183] Agencies had the option of building their own dedicated repositories, using, cloning or extending the NIH's PubMed Central repository; using SHARE; or using CHORUS.

While the OSTP memorandum remains in force, advocates of open access remain concerned that, as an executive order, it could be rescinded at any point. Longstanding efforts to enshrine public access requirements in law have therefore continued by means of two key bills, the Public Access to Public Science Act (PAPS) and the Fair Access to Science and Technology Research Act (FASTR).

In particular, the FASTR bill[184] incorporates the requirement for Federal agencies to develop 'public access' policies relating to research conducted by employees of that agency or from funds administered by that agency. FASTR reiterates that all agencies with research budgets greater than $100 million are required to make research outputs – specifically, "any results published in peer-reviewed scholarly publications that are based on research that directly

[180] https://www.timeshighereducation.com/news/uk-research-funders-target-hybrid-open-access-charges - survey-answer

[181] For further details see NIH Public Access Policy Details.

[182] United States Office of Science and Technology Policy (OSTP). (2013). Memorandum for the heads of Executive Departments and Agencies.

[183] For an updated list, see: https://www.science.gov/publicAccess.html

[184] https://www.congress.gov/bill/115th-congress/house-bill/3427

arises from Federal funds" – freely available with a maximum delay of 12 months following publication. The bill does not specify how this is to be achieved but required agencies to develop plans to meet these requirements.

Moreover, research institutions in the US typically enjoying a higher degree of autonomy and market-orientation than in Europe (Labaree 2007). There is no centralised higher education authority and US public research universities may receive as little as 10% of their revenues from the public purse.[185] As a consequence, the ability of government policymakers to exert influence over institutional policy is limited. Purchasing power is also more distributed in North America, with large numbers of independent library consortia conducting separate licensing negotiations with publishers.[186] The concept of a national open access strategy, such as those adopted by a number of European countries in the recent past, thus has little currency in the US. Widespread take-up of immediate OA publishing is likely to require bottom-up adoption by US academic libraries, but to date the appetite for this appears low.[187]

However, many thought leaders in the open access movement are based in the US, and there are substantial advocacy initiatives in support of immediate open access led by organisations such as SPARC and PLOS. There is also significant support for immediate open access amongst charitable bodies such as the Andrew W. Mellon Foundation, the Ford Foundation and the Bill and Melinda Gates Foundation. For example, the latter's open access policy,[188] which came into effect for all new funding agreements as of January 2015, has one of the strictest mandates. The policy requires immediate open access to both research outputs and the underlying data resulting from research funded by the Foundation, which no longer accepts embargoes and mandates the use of a CC-BY 4.0 licence or equivalent.

China

Over the past few years, there have been notable developments on open access in China. In 2014, the National Natural Science Foundation of China (NSBC) and the Chinese Academy of Sciences (CAS) introduced a mandate for all researchers to deposit their papers into online repositories and make them publicly accessible within 12 months of publication. In May 2015 NFSC launched an open access repository to support implementation of its policy statement. The Chinese Academy of Sciences developed two open access portals: the Institutional Repository Grid of Chinese Academy of Sciences, with content from 102 repositories, and the China Open Access Journal Portal with content from hundreds of journals.[189] Other research funders are also supporting open access. The National Science and Technology Report Service has created an OA digital collection while the national Ministry of Science and Technology (MOST) is developing its own open-access policy.

The focus on self-archiving of current OA policy in China aligns it with the US. However, recent statements by Chinese representatives have indicated support for the OA 2020 movement, and for experimentation with subscription agreements that would also cover OA papers authored by an institutions' researchers. In this respect, China's goals appear to be similar to those of the EU when it comes to the dominant English-language publishers, with a shared desire to improve market competitiveness, reduce costs, and enable affordable

[185] Ibid.

[186] The International Coalition of Library Consortia (ICLC) lists more than 100 library consortia from North America, compared with only 44 in Europe.

[187] An analysis of 'redirectable library expenditures' for 13 North American institutions in the 2013 year by the Pay It Forward project found that OA memberships and APC payments represented less than 1% of total expenditure, with the balance relating to subscription costs for in-scope materials. See University of California Libraries (2016), Pay It Forward, p.59.

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

[189] www.oaj.cas.cn

participation by all (Zhang 2016). However, as Montgomery and Ren (2018) observe, the drivers for open access in China are very different to those in the Western world. With its largely state-controlled publishing sector, China has faced no equivalent of the 'serials crisis' and Chinese-language scholarly content is already readily accessible for most Chinese scholars. Open access policy developments are thus framed primarily by a desire to increase quality and transparency, and to strengthen the country's knowledge infrastructure.

3.4.2 Open access' social and economic impacts

The policy drive behind open access is built on the assumption that the free circulation of scientific research outputs is likely to contribute to economic competitiveness, innovation and success. The rationale for open access relies in part on the characterisation of scientific knowledge as a global public good, which should be disseminated freely for the wider benefit of society.[190] Yet understanding what specifically these benefits are, and how significant they are, has proven complicated.

Over the years, there have been many attempts to study the impacts of open access, including the system-wide effects for scholarly communication and (more controversially) the wider economic impacts. These studies have often been criticized due to the methodological constraints that limit the accuracy of their findings. Criticisms have focused on the inclusion of non-cash items (such as estimates of researchers' time saved by improved access or the increase in productivity resulting from open access), which are notoriously difficult to prove and adequately measure. The uncertainty affecting the estimates has also been associated with accusations of implicit bias which could result in an overestimation of the positive impacts of open access (Anderson 2014c). Despite doubts about their accuracy, the studies have nevertheless had the merit of highlighting some of the mechanisms by which open access could contribute to social and economic development.

Tennant et al (2016) undertook a comprehensive review of the literature on the academic, social and economic impacts of open access. They concluded that the evidence points to:

- a favourable impact of OA on the scholarly literature through increased dissemination and re-use;
- substantial benefits to research- and development-intensive businesses
- the potential for OA to foster the development of stable research ecosystems.

One of the main benefits, as Tennant et al note, is cost-savings for taxpayer-funded universities. A 2008 report (RIN 2008) estimated that under a system in which 90% of articles were published using author fees, savings across the system would be about £560m, split almost equally between publishers and universities. However, costs and benefits would fall unequally across institutions - with research-intensive institutions paying 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. A Jisc report (Houghton et al., 2009) published the following year 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. The study also estimated increased economic returns to UK public-sector R&D arising from increased access might be worth around £170m, although this rests on hard-to-test assumptions about the levels of current access and the marginal rate of return to any increased access.

In the US, the Pay it Forward project (University of California Libraries, 2016) echoed the findings of the 2008 RIN report, concluding that institutional research publishing productivity

[190] Stiglitz, J. (1999). Knowledge as a public good, in Kaul, I., Grunberg, I., & Stern, M.A. (1999). Global Public Goods: International Cooperation in the 21st Century. OUP, New York.

will determine whether the library journal budget can cover publication fees. Less research-intensive institutions could entirely fund their authors' publishing costs from the library's budget, while more research-intensive institutions would likely have a significant funding gap.

3.4.3 Open access impacts on use

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 18 years[191].

Impact on usage (downloads)

In the era of digital publishing, the usage of research has changed substantially. Researchers operate in a context of widely available and generally unrestricted information, and they have access to more published research than ever before. Researchers' behaviour has therefore changed, with a substantial reduction in the time a researcher dedicates to each publication. Research patterns have changed and rely increasingly on discoverability of, and convenient access to, research outputs.

In this context, there is evidence that open access leads to increase downloads. A 2011 randomized controlled trial found that OA articles were downloaded significantly more often than paywalled articles, with HTML downloads roughly doubling and PDF downloads increasing by 62% (Davis 2011; Davis & Walters, 2011). In 2015, Wang and colleagues found that the increased usage of open access articles is more observable over time: whereas non-OA papers "only have a short period of attention" open access articles are used for a much longer timeframe (Wang et al. 2015). Although the trend in the extended usage of open access articles is clear, however, the evidence that higher downloads are resulting in more *active usage* and impact (e.g. citations and media presence) is less strong.

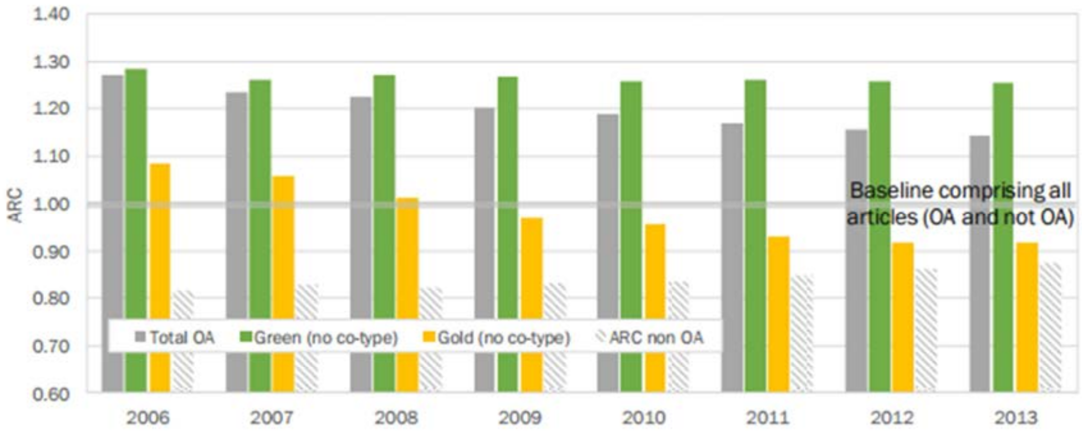
Open access citation advantage (OACA)

Many studies conducted to date have found that there is at least some increase in citations of open access papers over paywalled ones, though the observational nature of the findings means they remain open to question. A recent meta-analysis (McKiernan et al. 2016) of 70 studies registered in the SPARC Europe database of citation studies found that 46 (66%) had an OA citation advantage, 17 (24%) had no advantage, and 7 (10%) were inconclusive. The study also suggests that the open citation advantage seems to vary substantially across disciplines, with the highest OA citation advantage found in applied and physical science.

The relatively few large-scale studies available suggest that OA gives a citation advantage, but provide very different estimates. An analysis of over 200,000 articles from multiple disciplines found a 40% citation advantage for open access papers (Archambault et al. 2014). A study by Mikki (2017) looking at the total publication output of Norway (as indexed in Cristin) showed that open access articles receive on average, twice as many citations as paywalled ones on Google Scholar. By contrast, Piwowar et al (2018) suggest that open articles receive 18% more citations than otherwise expected. They draw their finding from a very large sample of 200,000 open access and 100,000 paywalled articles considered to be representative of the literature as a whole. A study from Science-Metrix (2018) shows that, while a citation advantage is apparent for all types of open access, self-archived papers appear to have significantly more citations than papers published using gold OA. Impact analyses in the report are based on the average of relative citations (ARC).

[191] For example, the Open Access Citation Advantage Service (OACA) by SPARC (continued until 2015) is a large source of papers in this area.

Fig 36: Scholarly impact by OA type, for papers published between 2006 and 2013, as measured in Q3 2016 (Source: Science-Metrix 2018)



The same study also shows that OACA is especially strong for Arts and Humanities (62% increase), and very significant for social sciences (33% increase) whilst still being significant for health, natural and applied sciences (16% to 19% increase).

Fig 37: Impact of open access publications by OA type, at the level of scientific domains (2010) (Source: Science-Metrix 2018)

Domain	All papers	All OA	Green (no co-type)	Gold (no co-type)	Non OA	Ratio ARC OA/Non OA
WOS	**1.00**	**1.19**	**1.26**	**0.96**	**0.84**	**1.42**
Applied Sciences	1.00	1.16	1.33	0.69	0.89	1.30
Arts & Humanities	1.00	1.62	2.00	0.67	0.82	1.99
Economic & Social Sciences	1.00	1.33	1.29	0.74	0.76	1.75
Health Sciences	1.00	1.19	1.19	1.11	0.81	1.48
Natural Sciences	1.00	1.17	1.26	0.80	0.85	1.39

Note: Citation advantage is the ratio of ARC OA/ARC non-OA. Color gradient is applied against the world level, with values above colored in green and those below in red, with stronger intensity of the gradient indicating a larger departure from world reference.

Source: Prepared by Science-Metrix using the Web of Science (Clarivate Analytics) and the 1science database

While many studies thus appear to show advantages in citations for open access articles, the findings have been subject to intense criticisms. One criticism focuses on selection bias, or the theory that authors select their best work for open access publication, and therefore that it is the quality of the work which leads to citations (Craig et al., 2007). At the same time, Berg (2010) argues that well-funded studies – which can afford to pay for APCs - are more likely to receive more citations than poorly funded studies. At the heart of these issues is the fact that the vast majority of studies constitute observational research, making it nigh on impossible to eliminate the effect of other variables (Crotty, 2018). These criticisms have been contested (Gargouri et al. 2010) but question marks remain, with some older studies, most notably a randomised, controlled trial undertaken by Davis (2011), finding evidence of increased downloads and a wider audience, but no clear citation advantage.

In summary, while a link between open access and citations is apparent, some uncertainty remains over whether there is a causal relationship, and the extent of any advantage arising. Nevertheless, recent analysis by Pollock and Michael (2018b) shows that an increasing number of fully OA publications are attaining higher impact factors at faster rates than their subscription and hybrid counterparts. Whether this is due to a citation advantage, or is simply evidence of fully OA journals maturing, remains open to debate.

In practice, the methodological limitations discussed above are frequently overlooked in press releases and headlines derived from studies in this field. OA advocates, academic scholarly networks and publishers alike have all found it beneficial to highlight findings that point to an OA citation advantage.[192] It is perhaps unsurprising, therefore, that a significant and growing proportion of authors believe that OA journals are more highly cited than subscription journals (Taylor & Francis, 2014).

Broader impact of open access publications

While citations are still the top-ranking measurement for the success of an article, alternative metrics have been developed. The extended use of social media and knowledge exchange platforms like Wikipedia led to new ways of interacting with research outputs. Projects such as Altmetric or ImpactStory have developed analytics systems to capture non-traditional metrics, some of which have subsequently been acquired by publishers. For instance, Elsevier acquired Plum Analytics and now provides PlumX Metrics score to their readers. PlumX Metrics measures usage, capture, social media mentions and citations with different subcategories to calculate an overall outreach score for each article.

There is an argument that open access has a positive effect on the broader societal impact of research, as papers can be more easily shared on social networks and picked up by the traditional media. A 2014 internal review of Altmetic.com found that open access articles published in Nature Communications generated significantly more tweets – including tweets from people who tweet research semi-regularly – and attracted more Mendeley readers than paywalled articles from the same publication (Adie 2014). As noted by Tennant et al. (2016), 42 of the top 100 articles of 2015 as listed on Altmetric.com were open access.[193] This is a higher proportion than the total number of open access articles published in the same year, therefore suggesting a correlation between open access and higher non-traditional metrics.

Despite these initial indications, however, more data is needed before concluding that open access has a *significant* effect on the societal impact of research. It is still unclear how much of the actual traffic around a research paper altmetrics are picking up, since the data used varies from metric to metric. Moreover, many researchers communicate interesting finds or share links via email – so called ‘dark social’ traffic (Madrigal 2012) that is notoriously hard to measure.

3.5 Market penetration of open access publications

Björk (2011) identified three phases in the development of the open access publishing industry: first, a phase of voluntary open access (often led by an individual scholar) in the 1990s, followed by a period in which long-established journals started publishing open access electronic versions of their article alongside the printed version, and a third phase that saw the introduction of OA as a business model from 2002. Over the past decade, we have arguably entered a fourth phase in the expansion of open access, which has seen most traditional scholarly publishers adopting (to a greater or lesser extent) open access publishing practices, and new open access journals becoming well-established within their disciplines. This expansion is reflected in the growing number of OA journals and articles.

[192] See, for example, SPARC Europe, https://sparceurope.org/open-data-citation-advantage/, Academia.edu, https://medium.com/academia/academia-edu-citations-and-open-science-in-action-4a24a6376573, and SpringerNature https://group.springernature.com/gp/group/media/press-releases/report-shows-value-of-hybrid-journals-to-research-community/15894508.

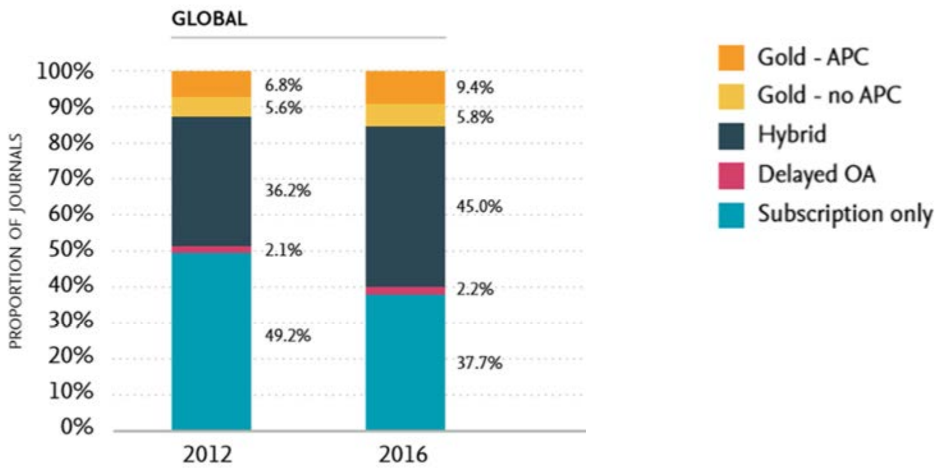
[193] https://www.altmetric.com/top100/2015/

3.5.1 Open access journal numbers

The number of open access journals listed by the Directory of Open Access Journals[194] was 11,811 as of July 2018 (of which 9,172 were published in English); this represents an increase of 1,720 over the 3 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).[195] 70% of journals in the DOAJ do not charge author-side fees (Crawford, 2018). Ulrich's Directory lists 12,618 peer reviewed OA journals at the time of writing, of which 7,817 are English-language.

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 some 22,500 peer-reviewed journals, of which 3,600 or 16% are open access. A further breakdown of Scopus journals by publishing model in 2012 and 2016 is shown in Figure 38.

Figure 38: Proportion of journals in the Scopus database y publishing model (Source: UUK 2017)



The Web of Science includes some 963 OA journals in its Science Citation Index Expanded (12% of the total) and 166 in its Social Science Citation Index (7%), and 84 in the Arts and Humanities Citation Index (6%). The greater prevalence of OA journals amongst newer publications is reflected in the Emerging Sources Citation Index, which lists 2,095 OA journals, representing 28% of the total.

3.5.2 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 *Scientific Reports*, which published over 24,000 articles in 2017), 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

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

[195] 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, which slowed growth in the number of OA journals listed substantially (Marchitelli et al 2017).

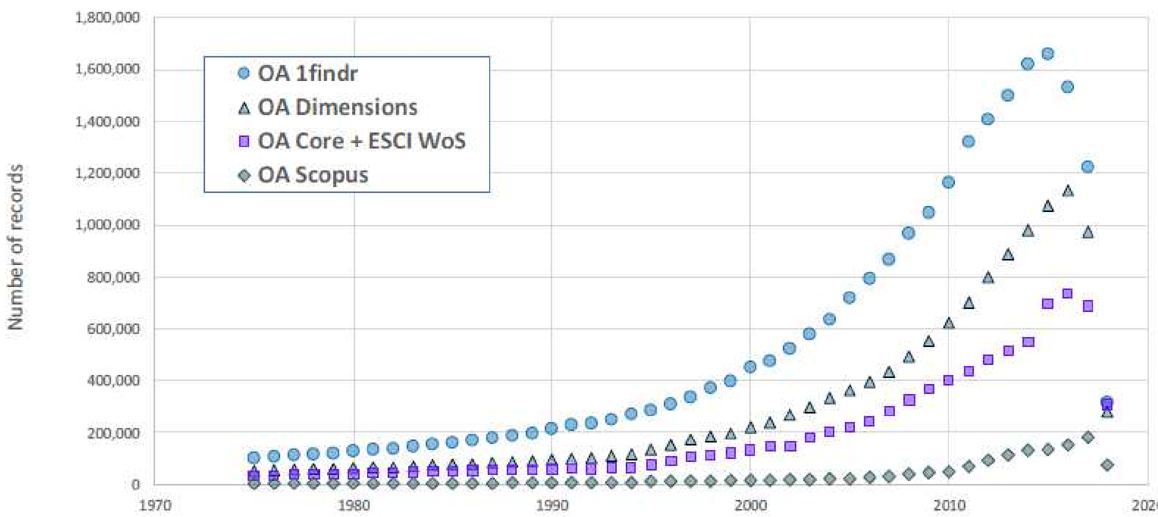
of their different estimates hard or impossible. Broadly speaking the four categories of articles counted are:

- 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);
- Delayed: articles in subscription journals made openly available after an embargo period;
- Green: copies of article versions available in institutional or other repositories; often embargoed for a period following the publication date; may exist in multiple archived versions and in multiple copies on different repositories; and
- Other: strictly speaking these are not open access but freely available articles, some legitimate (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, by reference to the new breed of discovery and analytics tools such as 1Findr and Dimensions, or by using specialised search engines and bots. Figure 39 illustrates the range of results available from four of the most common indexes and analytics tools, although the upward trend is consistent across all sources.

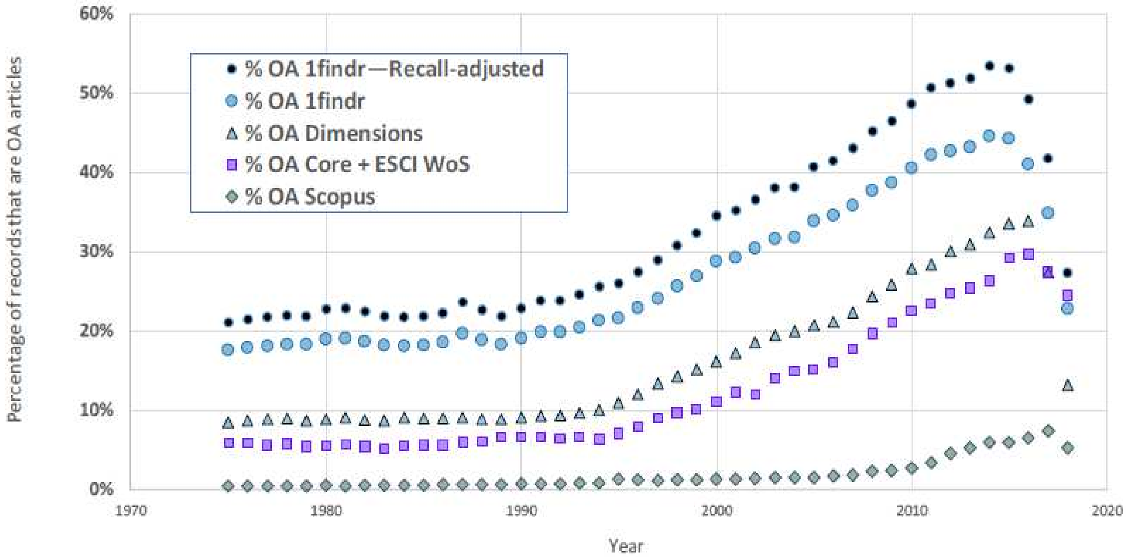
Figure 39 – OA articles indexed from academic & scientific journals, 1975-2018 (Courtesy of Eric Archambault)



3.5.3 Estimating the percentage of open access articles

Estimates of the percentage of content that is open access vary for the reasons outlined above. Newer content is more likely to be OA, with 30-45% of articles identified as open access by the Web of Science, Dimensions and 1findr databases in 2015, as shown in Figure 40.[196]

Figure 40 – Percentage of OA articles in academic and scientific journals: 1findr, Dimensions, Core +ESCI WoS and Scopus, 1975-2018 (Courtesy of Eric Archambault)



Several dedicated studies looking at levels of OA in the years 2014-2016 have also returned figures for OA content in the region of 30%, while others report substantially higher figures, in excess of 50% in some cases (see Table 10). Allowing for the fact that the EC's Open Science Monitor does not include hybrid OA, the consensus view suggests that roughly 15-20% of new articles were immediate (gold or hybrid) OA by 2016, compared with less than 5% in 2006, and 12% in 2011 (Van Noorden 2012a). In addition to gold OA, most studies include a combination of green, delayed/bronze (free-to-read but unlicensed) and other online postings in order to arrive at an overall percentage figure.

The use of these terms is somewhat inconsistent between studies, and all indices and tools are liable to underestimate the total OA article numbers due to the difficulty of correctly identifying and categorising openly available articles.[197] 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, and the proportion of OA content that may be missed. Where results are adjusted accordingly, this tends to result in much higher figures being quoted, as reflected in the 'recall-adjusted' figure for 1finder shown in Figure 40, above, and the Science-Metrix figure of 55% OA content in Table 10. Piwowar et al (2018) also estimate that their figures are likely to undercount OA by approximately 30%.

[196] At the time of writing, Scopus content is not reliably tagged as OA, though this is expected to change with the planned integration of Impactstory's Unpaywall database, see https://www.elsevier.com/connect/elsevier-impactstory-agreement-will-make-open-access-articles-easier-to-find-on-scopus

[197] Automated measurement systems are usually assessed for both *precision* and *recall*. The precision rate shows how much of the information captured by the measurement system is relevant in the reference system. The recall rate shows how much of the relevant information in the reference system is captured by the measurement system.

Table 10: Open access article shares reported by selected studies (see text for details and qualifications)

	EC Open Science Monitor[198]	*Bosman and Kramer (2017)*	*Universities UK (2017)*	*Piwowar et al (2018)*	*Science-Metrix (2018)*	*Martín-Martín et al (2018b)*
Primary data source(s)	Scopus + DOAJ, ROAD, CrossRef, PubMed Central, OpenAIRE	*Web of Science + Unpaywall*	Scopus + Google	*Web of Science + Unpaywall*	*Web of Science + 1Findr*	*Web of Science + Google Scholar*
Publication year considered	*2016*	*2016*	*2016*	*2015*	*2014*	*2014*
Gold (total)	14.4%	-	19%	16.7%	23%	11.6%
Gold OA	14.4%	-	15%	11.2%	-	10.1%
Hybrid OA	-	-	4%	5.5%	-	1.5%
Delayed OA	-	-	3%	-	-	1.1%
Bronze OA	-	-	-	10.8%	-	12.6%
Green OA	13.9%	-	-	10.4%	31%	10.5%
"Other OA" (total)	-	-	11%	-		20%
All OA	**28.3%**	**29%**	**33%**	**37.8%**	**55%**	**55.8%**

Several authors note the importance of both sustainability and legality when considering whether free-to-read content should be included in estimates of open access levels (van Leeuwen et al 2017; Martín-Martín 2018b). While more than 50% of content may be available to read on the web for free, and discoverable via search engines such as Google, there is often no guarantee that it will remain so in the future, nor that it has been shared legitimately. On the basis that such content does not meet the traditional definition of open access, a balanced assessment is that roughly one third of the scholarly literature was available OA in 2016.

Science-Metrix (2018) also report the percentage of OA across disciplines for publication year 2014, per OA type (as measured in late 2016). This shows the overall increased prevalence of green OA over gold OA, except for Health Sciences where gold OA is more widespread.199 A very low percentage of articles in the Arts & Humanities are made open access through either self-archiving or OA publishing.

Table 11. Percentage of OA across scientific domains for publication year 2014, per OA type, as measured in Q3 2016 (source: Science-Metrix 2018)

[198] https://ec.europa.eu/info/open-science/open-science-monitor/trends-open-access-publications_en#open-access-to-publications. Accessed August 2018.

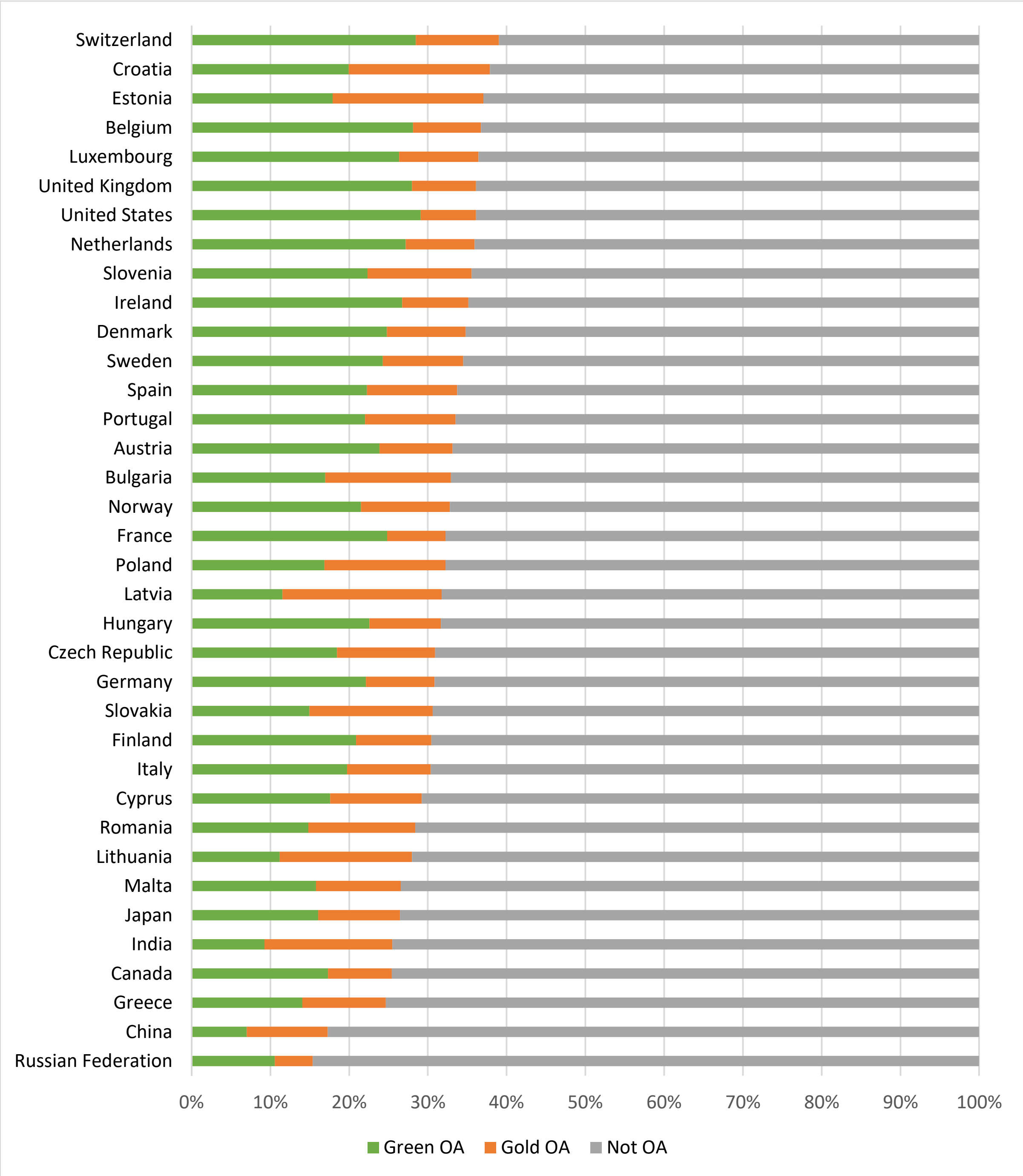
[199] The true proportion of green content is often obscured by the fact that many studies and OA discovery tools report content that is available in both gold and green OA forms as gold.

	Total OA	Green	Gold	Both Green & Gold	Undetermined
WoS	**55%**	**31%**	**23%**	**7%**	**12%**
Health Sciences	59%	30%	33%	9%	10%
Natural Sciences	55%	37%	15%	5%	12%
Applied Sciences	47%	29%	13%	3%	12%
Economic & Social Sciences	44%	21%	8%	1%	21%
Arts & Humanities	24%	9%	7%	1%	9%

Note: Percentages are based on the total number of publications and not only open access publications.
Source: Prepared by Science-Metrix using the Web of Science (Clarivate Analytics) and the 1science database

The European Commission's Open Science Monitor uses a limited range of data sources, and thus presents a conservative picture of the proportion of OA content. However, it effectively demonstrates the wide variations in levels of OA at country level. Countries such as Switzerland and Croatia achieve more than twice the levels of OA seen in China and the Russian Federation (Figure 41).

Figure 41: Percentage of open access publications (gold and green) by country (Source: EC Open Science Monitor, reference date April 30th 2018)



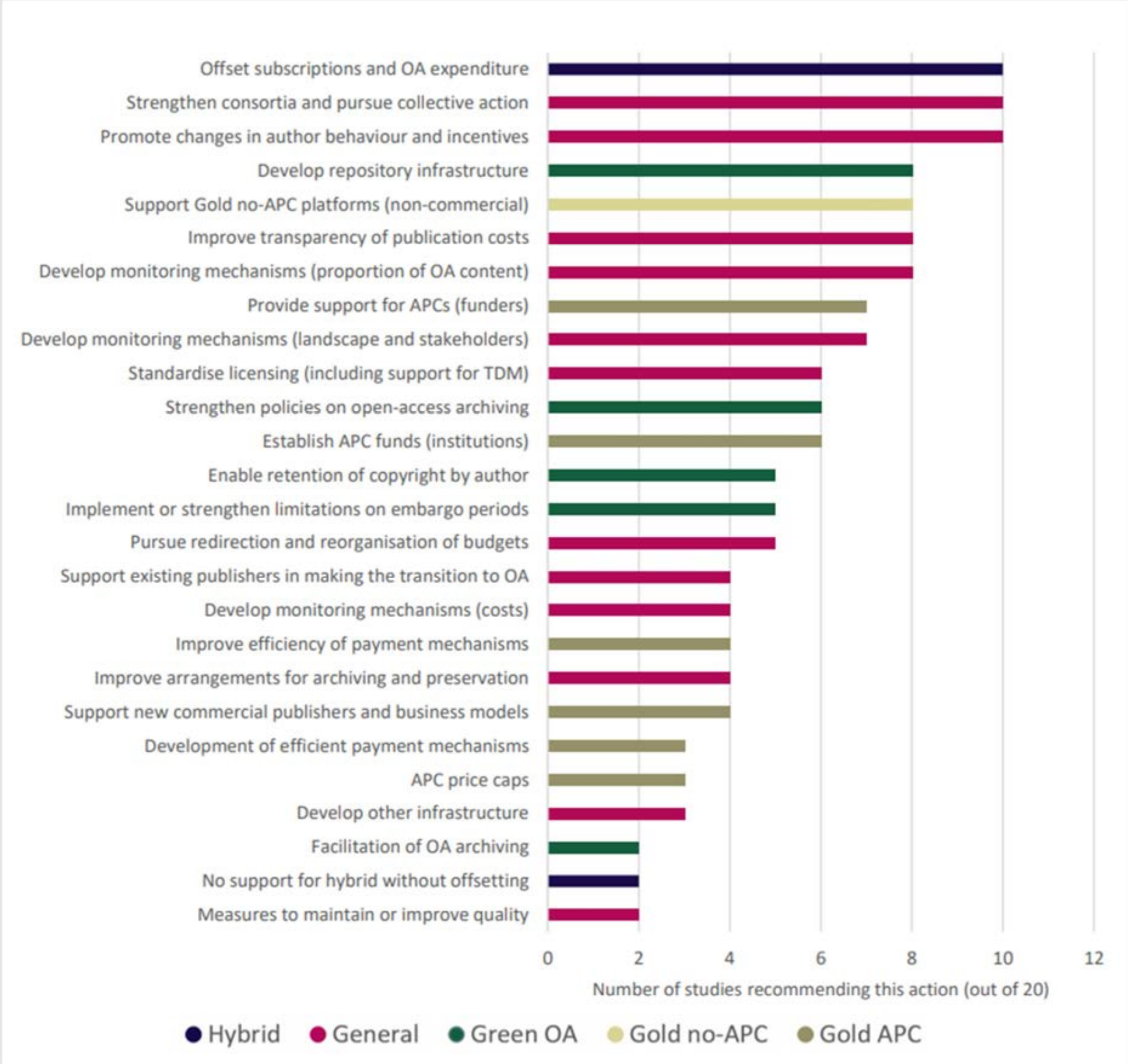
3.6 *Transitioning to a sustainable open access market*

The actions of policy makers and the publishing market make it clear that the open access debate has now moved on to *how* to make it sustainable and *how* to manage the transition. Sustainability implies a price equilibrium that leads to optimal continued access to high-quality scientific research. A sustainable market therefore balances the interests of the suppliers of publishing services (publishers and learned societies) with those of beneficiaries (researchers, research organisations, research funders and the public at large). Key sustainability questions include:

- 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, and (in the case of a flipped model), how will funding increases be determined?
- 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 adopting 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?

A wide range of interventions have been proposed to accelerate the transition, as shown in Figure 42. Offsetting, strengthened consortia and changes to author behavior and incentives have received the most support in recent years, particularly in Europe, but there remains no international consensus on the best way forward.

Fig. 42 - Meta-analysis of recommendations across 20 studies on the transition to open access (Johnson et al, 2017)



3.6.1 OA2020

Offsetting lies at the heart of the OA2020 initiative established at the 12th Berlin Open Access conference in 2015. The transformation envisioned by OA2020 is founded on a White Paper, published by the Max Planck Digital Library in April 2015, arguing that there is already enough money within journal publishing to allow for a transition to open access that will be – at a minimum – cost-neutral. The initiative endorses various ways of implementing open access, but asserts that a 'smooth, swift and scholarly oriented transition' relies on converting resources currently spent on journal subscriptions into funds to support sustainable OA business models. What is less clear is how long the exponents of OA2020 expect "resources currently spent" to remain sufficient, how future rates of increase might be determined, and how costs, and cost increases, are to be distributed.

At the time of writing, the OA2020 expression of interest had been signed by 109 scholarly organisations, but these are primarily from Europe, with only nine signatories in the US.[200]

[200] https://oa2020.org/mission/ - eois

STM publishers have given the plan a cautious welcome, noting that the proposed redirection of financial flows will be a complex endeavour, with both winners and losers, and that a mixed economy of different models is likely.[201]

3.6.2 APCs and sustainability

A major bone of contention for open access publishing is the cost of APCs. The question of what constitutes an appropriate price level for APCs encapsulates many of the tensions that underpin debates around sustainability.

A 2014 study of over one hundred thousand articles published in 1,370 journals (Björk and Solomon 2014) found that fully OA journals charged an average of $1,400 per APC, while hybrid journals for the six biggest publishers had APCs in excess of $2,700. Delta Think's Open Access Data & Analytics Tool, which holds information on pricing trends for over 14,000 journals, suggests the average APC list price for fully OA journals in 2018 is now just under $1,600 USD (median around $1,500), and for hybrids just under $2,900 (median $3,000).[202]

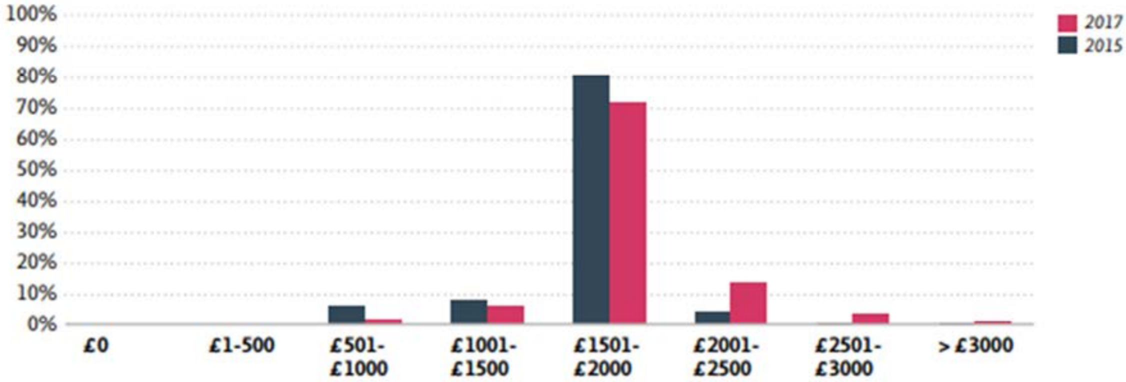
A recent study commissioned by Universities UK (UUK, 2017) also found that fully-OA journals are substantially cheaper than hybrid journals. The study looked at APCs levied in a range of open access journals from 32 publishers. Almost half of the journals either do not levy an APC or charge less than £500 ($630). Around 90% of journals have APCs of less than £1500 ($1,900), and in almost 99% of journals APCs cost less than £2000 ($2,550). By contrast, over 70% of hybrid journals charge between £1,500 ($1,900) and £2,000 ($2,550).

The study also showed that APC median prices rose substantially between 2015 and 2017, although there was variation between disciplines, [203] and the tracking of charges is complex due to exchange rate fluctuations variations, with APCs frequently charged and reported in different currencies.

Fig. 43 - APC price bands in sterling for full-OA journals from 32 publishers (from UUK 2017)

[201] https://www.stm-assoc.org/2017_06_13_2017_03_STM_Presentation_13_Berlin_OA_conference.pdf
[202] http://deltathink.com/open-access/oa-data-analytics-tool/
[203] The rise was the most substantial in arts and humanities (around £220/$280), followed by social science journals (around £180/$230 increase), physical sciences and engineering (around £150/$190), medicine and life sciences (showing a mere £20/$25 increase).

Fig 44 - APC price bands in sterling for hybrid OA journals from 32 publishers (from UUK 2017)



Concerns have been raised about not just the high cost of APCs, but also their distribution. A 2016 study found that the top 10 publishers, responsible for publishing most high-profile hybrid journals, make up 77% of APCs paid by UK institutions with the remaining 98% of publishers sharing the rest (Shamash 2016). A number of 'born-OA' publishers have also been acquired by subscription publishers in recent years, with examples including:

- 2008 - acquisition of BioMedCentral by Springer in 2008.
- 2011 - Medknow PVT Ltd. acquired by Wolters Kluwer.
- 2012 - Versita aquired by De Gruyter.
- 2013 - Nature Publishing Group acquired a controlling interest in Frontiers.
- 2016 - purchase of the journal portfolio of Liberatas Academics by Sage in 2016.
- 2016/17 - acquisition of Dove Medical Press and Coaction Publishing by Taylor and Francis/Informa.

Independent OA publishers often lack the relationships and access to library budgets available to subscription publishers, and so risk being excluded from the moves to negotiate new 'big deals' that include OA.

At the same time, Siler et al (2018) identified a correlation between institutional rank and publication in APC-based journals, suggesting that the cost of APCs, especially among the most prestigious hybrid journals, may be pushing authors working at lower-ranked institutions to publish in closed/paywalled outlets. By contrast, academics from higher-ranked (and better resourced) institutions were found to pay relatively higher fees for gold and hybrid OA publications. The authors concluded that this is leading to “new professional hierarchies” as open-access models become more popular, though the link remains unproven.

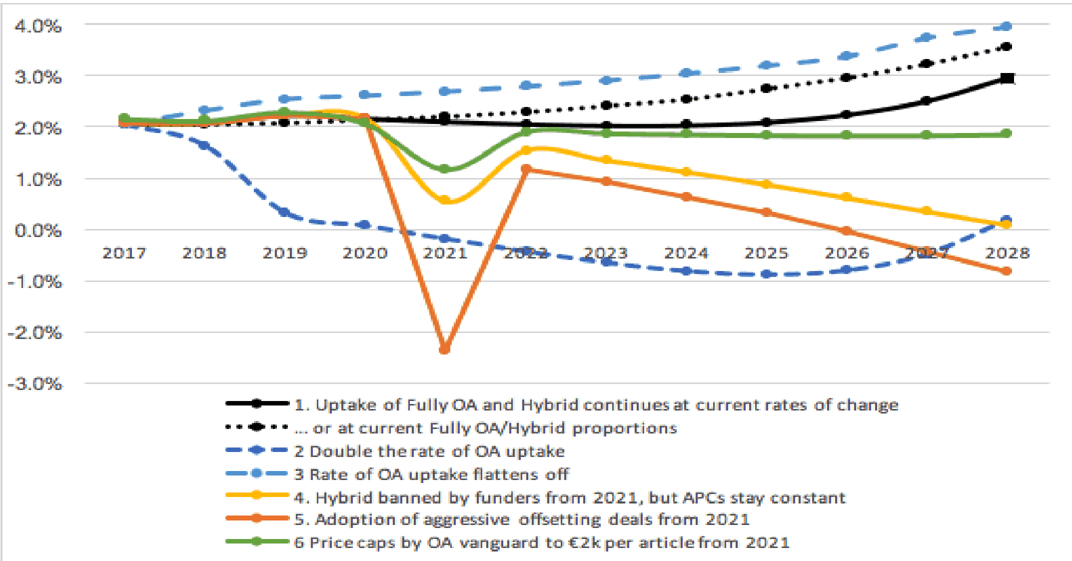
APCs are especially unaffordable for researchers working in low-income countries. Incentives to publish in prestigious international journals are equally strong in developing countries, where a single APC of $2,000 is equivalent to many months of a researcher’s salary. A number of initiatives are trying to tackle this unbalance (see also *2.19 Access in developing countries*). Most full open access journals grant researchers based in developing countries fee waivers or discounts that are, at least in part, ‘priced into’ the subscriptions and APCs paid by research organisations in developed countries, and partly supported by governmental actors. However, practical difficulties in obtaining waivers can make the process impracticable for researchers, therefore making open access publishing an uneven playing field for developing country researchers.

Many open access advocates maintain that APCs are far too high and point to the ability of some journals to be financially sustainable at low APC rates. However, journals have substantially different cost bases – with a strong, though imperfect, correlation between a journal's prestige, the proportion of submissions it accepts for publication and its publishing cost base. On the assumption that the market dynamic will be to lower prices, the concern of some industry commentators is that pressure to lower costs could lead to corners being cut and quality reduced (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. At the same time, pressures on revenues and thin margins could increase pressures on editors or publishers to reduce scientific standards to accept more articles.

Consideration of APC prices tends to overlook the not insignificant costs of scholarly infrastructure provided by publishers, which remains largely underwritten by subscription revenues. The financial obstacles are further exacerbated by the potential loss of the significant revenues that publishers earn from channels other than library subscriptions in an OA world – including licensing revenues and corporate subscriptions.[204]

As a recent SpringerNature whitepaper has observed, a rapid shift to a fully OA model would entail significant cost, time, risk and disruption for the whole research ecosystem, and thus incumbent publishers continue to favour an 'orderly evolution' (Mithu et al, 2018). Analysis by Pollock and Michael (2018a) indicates that the most radical changes could come from wholesale moves to offset big deals (Figure 45). However, even if these happened at scale over a short period of time (an unlikely scenario), they estimate that the worst-case effects would be to slow overall market growth to just below zero in the long term. This reflects the fact that revenue per article is already declining under the subscription model, and is likely to be roughly in line with APCs of €2,500 ($2,800) in 10 years' time.

Figure 45: Effects of OA scenarios on annual growth in the scholarly journals market over time (Pollock and Michael 2018a)



Source: Delta Think Open Access Data & Analytics Tool market data, US Science and Engineering Indicators 2018, Open APC data, Delta Think Analysis.

[204] For example, the BMJ recently disclosed that it earns 8.7% of its income (£6.76m; €7.6m; $9m) from product advertising, commercial sponsorship, and the sale of article reprints. See https://www.bmj.com/content/359/bmj.j4930

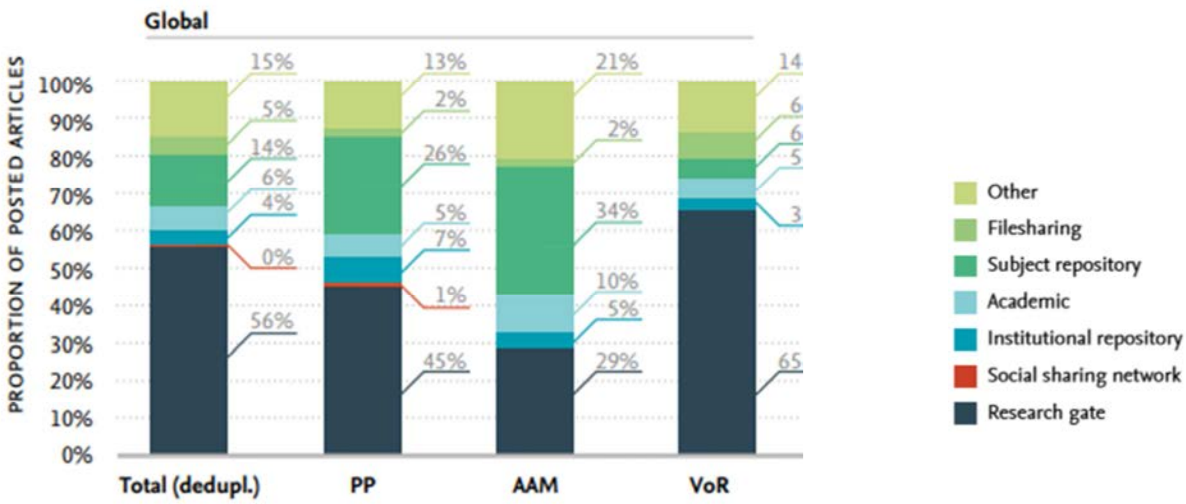
3.6.3 Effect of self-archiving on journals

There have been longstanding concerns that the self-archiving of articles could, in time, lead to widespread cancellation of subscriptions. Most publishers allow authors to archive versions of their articles on the web, although many of these policies were originally introduced on the understanding that the archiving would not be systematic. The STM Association has stated that Green OA can be consistent with fostering a viable system for funding publication of authoritative communications about research, but has emphasised the need for flexible and appropriate embargo policies.[205]

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 posted to subscription revenues). Policies will also vary to reflect specific funders' requirements; for instance very few publishers will not allow deposit to PubMed Central. There is evidence that publishers' OA policies are influenced by the national OA policy environment within which the publisher operates, notwithstanding the global nature of the journals market (Gadd et al, 2018). For example, some UK-based publishers, including Emerald and the Microbiology Society, have adopted more liberal green open access policies following introduction of a deposit mandate under the Research Excellence Framework (REF).

A recent study by Universities UK (2017) found that, overall, ResearchGate has by far the largest proportion of posted articles, higher than all other sources combined.[206] Among accepted manuscripts, subject repositories have the largest number of articles, followed by ResearchGate and other sources (such as author's websites). By contrast, institutional repositories – whose number has been increasing more rapidly than any other repository - have a very small proportion of articles at just 4% of the duplicated total (Figure 46).[207]

Figure 46: Proportion of subscription articles posted online globally, by document version and location



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

[205] https://www.stm-assoc.org/2017_05_01_STM_Position_Green_OA.pdf

[206] It should be noted that Sci-Hub was not included in scope of the study.

[207] Since postings may occur in different versions and at different locations, the study's figures are presented as proportions of the duplicated totals and so add to 100%

indefinitely) will increasingly choose to rely on the self-archived version rather than subscribe to the publisher's version.

A review of the impact of short embargoes on publishers by the UK Publishers Associations provides some support for this hypothesis, citing a small number of cases where journals subscriptions were cancelled following a reduction in embargo periods.[208] A now-dated study by SIS for the Publishing Research Consortium (Beckett & Inger, 2006) 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 and a report by the British Academy (Darley et al 2014), have demonstrated the preference of researchers for the version of record, at least for some stages of the research publishing cycle.

3.6.4 Sci-Hub

Concerns about copyright infringement are perhaps best exemplified by the great popularity of Sci-Hub. Sci-Hub is a website with around 60 million academic papers and which serves over 400,000 requests per day (Himmelstein et al, 2018). Founded by a Kazakhstani graduate student, but obscurely funded, the site bypasses publisher paywalls to download copies of scholarly articles through educational institution proxies onto its own server and makes those copies freely available to the public. A recent study found that Sci-Hub's repository contained 69% of all scholarly articles with DOIs, while average coverage of the most cited journals exceeded 90% (ibid.). Sci-Hub is thus in persistent and large-scale infringement of copyright and has been declared illegal by several courts. Despite that, and despite routinely having its domain taken down, the site still experiences high levels of usage among researchers from developing and developed countries alike. There are indications that researchers use Sci-Hub out of convenience, even where they have legitimate access to content (Bohannon, 2016). Not dissimilarly to illegal file-sharing sites for music and videos, Sci-hub illustrates the challenge of protecting copyright for digital publications. Its success has prompted calls for publishers to support legal open access to prevent the growth of illegal file-sharing.

3.6.5 Scholarly Communications Licence

In 2008, the Harvard University Faculty of Arts and Sciences adopted a policy that purported not only to require scholars to deposit their works in open access repositories, but also to grant the university nonexclusive copyright licences to archive and publicly distribute all faculty-produced scholarly articles.[209] Over 70 other universities, predominantly but not solely in the US, have since adopted similar policies.[210] These policies typically include provision for institutions to waive application of the licence for a particular article or delay access for a specified period of time upon request from the author. As Priest (2012) has noted, the Harvard-style "permission mandate" has the potential to be more disruptive than its "deposit mandate" cousin, turning exclusive rights to copy, distribute or re-use of another's work on their head by automatically granting broad permissions absent the author's written objection. However, in practice authors often default to the most restrictive stated practice, even in the face of institutional OA policies (University of California Libraries, 2018).

[208] See https://www.publishers.org.uk/_resources/assets/attachment/full/0/24490.pdf
[209] See https://osc.hul.harvard.edu/policies/fas/
[210] See http://cyber.harvard.edu/hoap/Additional_resources#Policies_of_the_kind_recommended_in_the_guide

The debate over the potential impact of such licences on publishers has been reignited in the recent past by proposals for a UK Scholarly Communications Licence, an iteration of the Harvard licence adapted for UK law.[211] Advocates of the licence argue that it would allow authors and institutions more control over the research publications which they produce, and offers the potential for 'one-step' compliance with the various funder policies to which UK academic authors are subject (Baldwin and Pinfield, 2018). However, it was been criticized as monolithic and inflexible in some quarters (Wulf and Newman, 2017), with the UK Publishers Association raising concerns over the potential administrative burden arising from waiver requests; the way the SCL seeks immediate non-commercial re-use rights for all UK research outputs; and the limit it could place on the choice of researchers over where to publish.[212]

These debates look set to continue over the coming years, particularly with the recent announcement of the Plan S principles, which seek to ensure that authors retain copyright of their publication with no restrictions (see *Section 3.4.1 Funders' policies on open access*).

---

211 See http://ukscl.ac.uk/

212 See https://www.publishers.org.uk/policy-research/copyright-and-ip/scholarly-communications-licence/

4. Technology in Scholarly Communication

Technology is driving 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 presents some challenges. This chapter discusses trends that are important to scholarly publishing at the moment of writing. For updated reviews of new technology trends and their impact on the STM publishing industry, readers can follow the reports of the STM Future Lab Committee on 'Technology Trends'.[213]

Recent topics covered by the reports cover a broad range of subjects.[214] Among others, the report "Entering the AI Era" (2018) highlights the impacts and opportunities that artificial intelligence has for the worlds of scholarly communication in general and STM journal publishing in particular.[215] The topic has also been explored in other work, and an in-depth summary is available on the STM website.[216] Other trends include the increasing pressure to leverage big data and user analytics to improve user workflows and commercialisation, "de-siloing" products in order to use all available data sources, and adopting "platform services" for journal articles and other scholarly content. Another authoritative source of information on these topics comes from Outsell, whose reports are especially useful for STM platform providers (see Outsell 2017).

4.1 Recent initiatives and organisations

4.1.1 Open Science

The Outsell (2017) report observed that there is an increasing need to decide how to deal with "Open Science". This is now an even more important driver of change, and potentially transformation of the current system of scholarly communication.

The EU's FOSTER portal defines Open Science as "the practice of science in such a way that others can collaborate and contribute, where research data, lab notes and other research processes are freely available, under terms that enable reuse, redistribution and reproduction of the research and its underlying data and methods".[217] Open Science is broader than open access because it is concerned with openness throughout the research cycle and not just open access for the publication produced at the end of research. STM (2016) have adopted a similarly broad view of Open Science as including:

- Open access to scholarly publications.
- Open Data, intended as the ability of researchers to make their research data FAIR (Findable, Accessible, Interoperable, and Re-usable).
- Research metrics that measure the research endeavour and underpin assessment and evaluation.
- Citizen science, co-developed by researchers and the wider community.
- Research integrity, especially in the peer review process.

The various components of Open Science are discussed throughout this report: open access, is discussed at length in section 3; Open Data is discussed in section 4.2; research metrics are discussed in sections 2.12.2-6 and 4.8.

213 https://www.stm-assoc.org/standards-technology/future-lab-committee/
214 See for a summary: https://www.stm-assoc.org/standards-technology/tech-trends-2021/
215 https://stm-assoc.org/standards-technology/tech-trends-2022
216 https://www.stm-assoc.org/2018_04_20_2018_STM_Tech_Trends_Philadelphia_Conference.pdf
217 https://www.fosteropenscience.eu/foster-taxonomy/open-science-definition

4.1.2 Open Scholarship Initiative

OSI[218] is a United Nations-backed partnership between research universities, publishers, government agencies, and 15 other key stakeholder groups in scholarly communication from around the world. Over 380 senior leaders from 24 countries and 250 institutions are currently part of this effort.

The OSI summit group recognises that until a global system is created that puts the needs and concerns of researchers front and centre, the research community will look upon open mandates with disdain, resignation and confusion. Accordingly, they are working to make find a way to make Open benefit research and be enthusiastically embraced by researchers.

In 2018 OSI will begin issuing a series of issue briefs, derived from its work and conversations to date, offering guidance to the global scholarly communication community on issues of importance. These briefs will be endorsed by UNESCO and circulated as global guidance. The first brief will be on defining the meaning of open---OSI endorses an "open spectrum" that includes a wide variety of approaches to and states of open, not just a static "open access" endpoint (see Section 3.1 *Defining openness*). Their hope is that these briefs will serve as important tools for bringing together the global stakeholder community in common action, and moving progress forward (hopefully rapidly) on a number of key issues.

4.1.3 Force 11

FORCE 11[219] is an older organisation dating from 2011. It describes itself as a community of scholars, librarians, archivists, publishers and research funders that has arisen organically to help facilitate the change toward improved knowledge creation and sharing.

FORCE11 has grown from a small group of like-minded individuals into an open movement with clearly identified stakeholders associated with emerging technologies, policies, funding mechanisms and business models. It organises a well-regarded conference and three useful calendars covering digital events, STM events and scholarly communication events. Its 2011 manifesto, now available in an updated form, remains a realistic analysis of the new opportunities and challenges.[220] It focuses its attention on the need for tools for researchers as producers of STM outputs, enhanced products for researchers as consumers, and tools and services for reputation management. FORCE 11 has been closely involved in a number of the specific issues discussed elsewhere through groups on Data and Software Citation, Reports Ideas and Projects (RIAP), Resource Identification, and Scholarly Commons.[221]

4.1.4 Center for Open Science

COS[222] is a rather different type of organisation, with a mission to enable open and reproducible research practices worldwide. Its Executive Director and Co-Founder, Brian Nosek, is a professor in the Department of Psychology at the University of Virginia. Reproducibility is a special problem in the behavioural sciences and it is thus not surprising that COS concentrates their wide range of services in these disciplines. At the same time their preprint servers collect preprints across a much wider range, and the Open Science Framework (OSF), operated by COS, is a free and open source project management repository that supports researchers across their entire project lifecycle.

[218] http://osiglobal.org/
[219] https://www.force11.org/
[220] https://www.force11.org/about/manifesto
[221] https://www.force11.org/groups
[222] https://cos.io/

4.1.5 Joint Roadmap for Open Science/Scholarly Tools

The JROST community, formed in early 2018, comprises a group of individuals and organisations building nonprofit, open-source tools for scholarship and publication.[223] These groups are actively building and delivering solutions, and thus have product roadmaps that they can speak authoritatively to. It is also critical that researchers participate because their needs and perspective drives the demand for what is being done.

4.1.6 Government and national initiatives on Open Science

Many countries have governmental initiatives under the banner of Open Scholarship/ Open Science and there are some which are relevant in this context. Not all initiatives are covered but where they are of wider scope they may be discussed elsewhere in the Report. At an international level part of the declared policy of the G20, under the heading “Breaking a New Path for Growth”, visualises the promotion of open science and the facilitation of appropriate access to publicly funded research results on findable, accessible, interoperable and reusable (FAIR) principles as part of the drive towards innovation-driven growth and the creation of innovative ecosystems.[224] In 2017 the G7 Expert Group on Open Science articulated their ambition to:[225]

- Foster a research environment in which career advancement takes into account Open Science activities, through incentives and rewards for researchers, and valuing the skills and capabilities of the Open Science workforce
- Enable all researchers are able to deposit, access and analyze scientific data across disciplines and on international scales. [That] Research data management adheres to the FAIR principles whereby data is findable, accessible, interoperable, and reusable.

4.1.6.1 Open Science policies in the European Union

At a policy level the EU has taken an active interest in Open Science for some years. Under Carlos Moedas, the current Commissioner for Research, Science and Innovation, the emphasis is on the three “O”s - Open Innovation, Open Science, Open to the World. Open Science is closely linked to innovation and Europe’s competitiveness in the World [European Commission 2017] Another initiative is the EC Open Science Policy Platform which is formed to provide advice on open science policy and which meets regularly. The EC maintain a site which provides news, events, publications related to Open Science[226].

As discussed in section 3.4.1, the European Commission and the European Research Council (ERC) recently spearheaded the launch of Plan S, a set of principles on open access endorsed by 11 national research funding organisation. These principles are likely to have a profound effect on open science across Europe and beyond - especially with regards to scholarly outputs.

4.1.6.2 Implementation of EU policies on Open Science at a European and National level

The EU has a number of programmes and projects intended to bring together member countries to take an active view of Open Science. A useful source of information is the OpenAire blog[227]. Particularly relevant is the emphasis on using repositories as central to the

[223] https://jrost.org/
[224] http://europa.eu/rapid/press-release_STATEMENT-16-2967_en.htm
[225] www.g7italy.it/sites/default/files/documents/G7%20Science%20Communiqué.pdf
[226] https://ec.europa.eu/research/openscience/index.cfm
[227] https://blogs.openaire.eu

progress of Open Science. OpenAire is aimed to support the implementation of the EC and ERC Open Access policies and has considerable influence among member countries.

The European Open Science Cloud (EOSC) is at the centre of another area of infrastructural development: a roadmap is in position. EOSC is a cloud for research data in Europe. Background, policy information, events and publications related to the EOSC is provided via this link which provides the latest picture in 2018[228].

Another example is the 'Science with and for Society' programme[229] will be instrumental in addressing the European societal challenges tackled by the broadly based Horizon 2020 initiative and it is intended to build capacities and develop innovative ways of connecting science to society. It will make science more attractive (notably to young people), increase society's appetite for innovation, and open up further research and innovation activities.

A typical smaller enterprise is an Open Software introductory course for early career researchers. This is part of a general "Foster" project reaching out to this audience[230].

Member countries have varied in the extent to which they work with the EU on these projects and to what extent they have their own attitudes to Open Science. The UK has been closely involved with the development of Open Access as we have seen elsewhere. The Jisc, which provides digital services and solutions for the higher and further education and skills sector set out their programme in 2015[231]. However on the whole the UK government has concentrated on the development of the Finch process as explained elsewhere in this report Germany too has concentrated on Open Access[232].

Very recently Open Science, which in France seems to comprise the humanities too, has been taken out of the hands of the National Library and handed to new national committee Comité pour la Science Ouverte (COSO) under the auspices of the Ministry of Higher Education and Research. It is intended that researcher involvement should be greater than has currently been the case.

4.1.6.3 Open Science policies and implementation outside the EU

As a generalisation the US differs from most other countries in not having a central body which can explore and provide policies under an Open Science banner. They have also been much concerned with Federal policies regarding public access. However, a new report from the US National Academies seeks to provide a national roadmap (National Academies of Science Engineering and Medicine 2018). The report is aimed at overcoming barriers and moving toward open science as the default approach across the research enterprise. It explores specific examples of open science and discusses a range of challenges, focusing on stakeholder perspectives. It is intended to provide guidance to the research enterprise and its stakeholders as they build strategies for achieving open science and take the next steps

The government of Japan set out its approach to open science in the document *Promoting Open Science in Japan 2015 (*Cabinet Office Government of Japan Expert Panel on Open Science 2015). The document places an emphasis on open knowledge as a tool to promote

[228] https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
[229] https://ec.europa.eu/programmes/horizon2020/en/h2020-section/science-and-society
[230] https://www.fosteropenscience.eu/learning/open-source-software-and-workflows/#/id/5abf67d9dd1827131b90e6bd
[231] http://www.jisc.ac.uk/blog/open-science-many-hands-make-light-work-17-aug-2015
[232] https://innovationpolicyplatform.org/content/germany-open-science-country-note

innovation, and covers both research publications and research data. It advocates for a top-down approach with centralised mechanisms for open science.[233]

4.1.7 Reproducibility and integrity

The lack of reproducibility of scientific research published in journals is increasingly perceived as a serious problem (sometimes called the “reproducibility crisis”). Because the challenges to reproducible science are systemic and cultural (A Manifesto for Reproducible Science), the literature has tended to focus on theoretical solutions to what is still an ill-defined challenge (Casadeevall 2010).

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 for a short while the launch of Pubmed Commons to support open discussion on published articles – discontinued in early 2018. A more recent initiative is the Transparency and Openness Promotion (TOP) guidelines[234] which spring from an article in 2015 (Nosek 2015). These guidelines include eight modular standards, each with three levels of increasing stringency. Journals select which of the eight transparency standards they wish to adopt for their journal, and select a level of implementation for each standard. These features provide flexibility for adoption depending on disciplinary variation, but simultaneously establish community standards.

Reproducibility is a complex, multi-dimensional problem with roots deep in the research process. However, it is also affected by some aspects of publishing such as: incentives and pressures for early publication; selective publication of positive findings; and weak challenge of statistical analysis in peer review. Even the language in which the concept is expressed causes difficulties because many associated terms are not standardised. Johnson notes that lack of reproducibility may have more to do with inadequate study designs than bad science per se or scientific misconduct,[235] while other commentators warned against over-emphasising reproducibility problems.[236]

Publishers and individual 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 *Section 4.2 Research Data*); encouraging or requiring the sharing of computer code as well as research data; and strengthening peer review, for instance by adoption of reviewer checklists 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; and semantic mark-up of entities like reagents and antibodies to ensure unique identification.

A survey of early career researchers conducted as part of the Harbingers project (PRC 2017) indicated that one of the key priorities in ensuring reproducibility is the ability to repeat a research experiment and test the results. In this light, most respondents suggested the following:

a) making dataset and supplementary materials available online and easily/openly accessible;

---

[233] http://blogs.nature.com/scientificdata/2016/04/15/reflections-sdjpn16/

[234] https://cos.io/our-services/top-guidelines/

[235] Valen E. Johnson Revised standards for statistical evidence PNAS November 26, 2013. 110 (48) 19313-19317; http://www.pnas.org/content/110/48/19313.full

[236] Daniele Fanelli -Opinion: Is science really facing a reproducibility crisis, and do we need it to? PNAS March 13, 2018. 115 (11) 2628-2631; http://www.pnas.org/content/115/11/2628

b) making articles (and especially the methods section) more detailed and extended;
c) using videos to explain methodology; and
d) being transparent and engaged in answering methodological questions from peers.

4.2 Research data

4.2.1 Data as an enabler of scientific discovery

A wide range of bodies, including publishers, funding agencies, individual research organisations and groups of researchers have been actively promoting and supporting developments in the domain of research data. STM publishers have been committed to the principle of sharing raw data along with publications since 2007, date of the Brussels Declaration.[237] This was complemented by a 2012 joint statement between DataCite and STM publishers, which included best-practice recommendations to "make research data easier to find, link to, reuse and cite".[238]

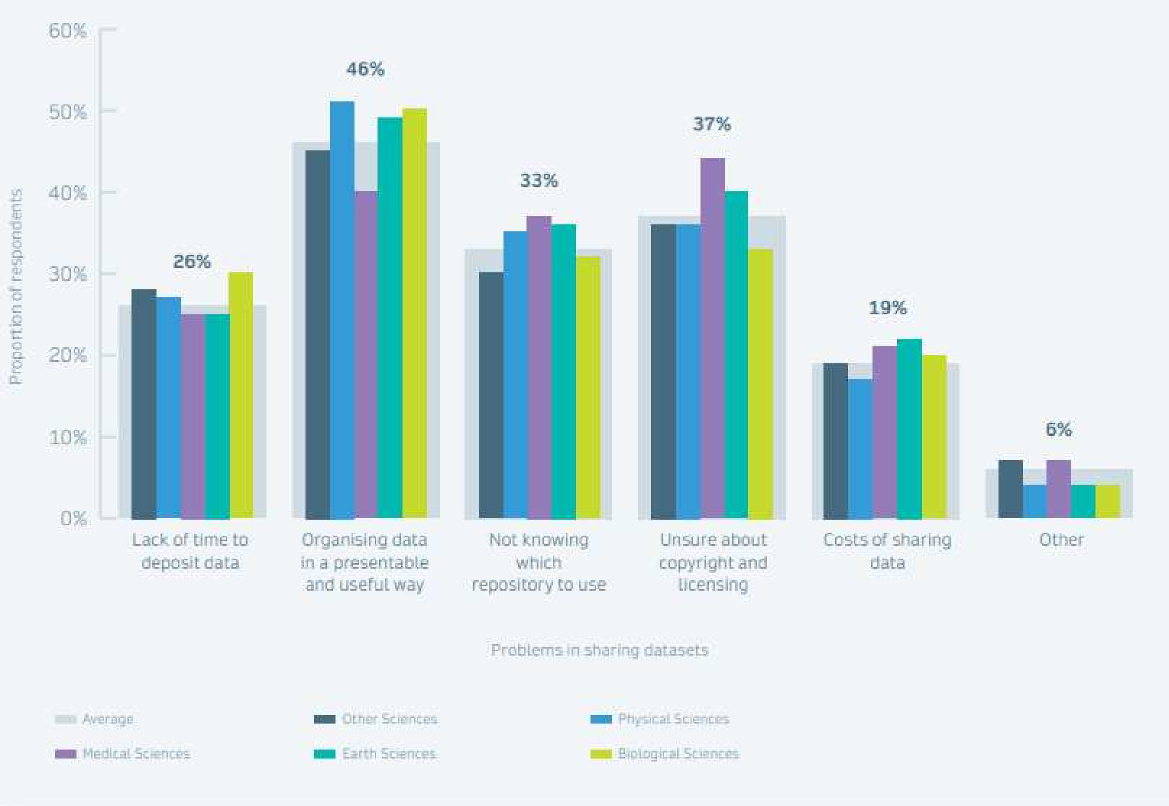
4.2.2 Barriers to data sharing

For the most part, the cultural changes necessary to enable widespread dissemination of research data have been slow to occur, although a few disciplines, such as genomics and astronomy, have well-established sharing practices. To be usable, information must be documented, sorted, curated, shared and preserved (Corti et al. 2014), which places a burden on researchers but also on publishers and infrastructure providers. The former need to provide information in a suitable format, while the latter must provide technological solutions to enable research efforts.

The challenges of organising data and uncertainty over copyright and licensing of data appear to be the greatest barriers to data sharing, but lack of time, funding and knowledge of available solutions are also significant factors (Figure 47). Fears that data may be misused, or that researchers will find themselves scooped are also commonly cited in the literature, though appear to be of secondary importance to the practical challenges of data sharing (Stuart et al 2018).

[237] https://www.stm-assoc.org/public-affairs/resources/brussels-declaration/
[238] http://www.stm-assoc.org/2012_06_14_STM_DataCite_Joint_Statement.pdf

Figure 47: Barriers to sharing datasets in different subject areas (Source: Stuart et al 2018)



4.2.3 Infrastructural challenges

The availability of unprecedented quantities of data offers new opportunities but also poses infrastructural challenges, as follows:

- Data users, including researchers, are increasingly interested in machine-readable information. This has immediate implications for data discoverability but also for practical applications such as text and data mining (see section 4.8).
- Every year, new outlets to deposit data are created, including online repositories and data journals. At the time of writing, the re3data registry[239] holds details of almost 2,150 data repositories, and the number continues to increase. The wide range of options complicates the need for interoperable infrastructure, e.g. to allow the creation of robust metrics or links between datasets and the scientific outputs they underpin.
- Researchers' ability to analyse large-scale datasets may be dependent on access to high performance computing facilities. Individual researchers or research groups can work with small or medium datasets, but beyond a certain threshold, advanced computational approaches are required. This is frequently the case in fields such as high-energy physics, astronomy and biology.

To improve the provision of data-related services, STM publishers may collaborate with existing infrastructure providers (Inchcoombe 2017). This relieves them from the burden of developing ad-hoc and, possibly, duplicated services and is proving an effective way to split

[239] https://www.re3data.org/search

efforts during the article publication process. While, in several cases, infrastructure providers are private organisations, in others public actors are involved. A recent example is the European Open Science Cloud (EOSC),[240] which is expected to provide EU researchers with an environment including free, open services for data storage, management, analysis and re-use across disciplines.

4.2.4 Depositing data

Data deposit by academic researchers takes widely different forms, with requirements ranging from depositing a custom-made dataset with a detailed explanation to sharing figures, tables, materials and methods which are not required to enable an understanding of their research outputs but can offer further insights to readers (Mooney 2016). Options include:

- **Online repositories** do not require peer-review and assign DOIs in minutes: this route to data deposit is the most flexible and need not involve the publication of an article. Key sources to identify suitable online repositories are re3data[241] and FAIRsharing.[242] The choice between publicly-funded (e.g. Zenodo)[243] or private initiatives (e.g. figshare)[244] contributes to an already-complex environment.
- Depositing **data along with a publication** is possible with a wide range of journals. These may partner with existing infrastructure providers (e.g. Springer Nature's collaboration with figshare, (Inchcoombe 2017)) or use in-house services (e.g. Elsevier's Mendeley Data, which is embedded in the article submission workflow).[245]
- Publishing in a **data journal** (see examples in Table BBB) is the preferred pathway when authors need to explain in detail how data has been created or processed. Datasets may be either deposited in the same journal or in a dedicated data repository. Where data journals link to external datasets there are often minimum requirements for the third-party hosting (e.g. Geoscience Data Journal specifies the repository must be able to mint a DOI).[246]

Nature's Scientific Data[247] maintains a list of recommended repositories with high standards in terms of data access, preservation and stability. These are split by subject and link to re3data and FAIRsharing entries.

It should be noted that depositing research data is becoming increasingly important, as major search engines such as Google are entering this market (Castelvecchi 2018). The release of Google Dataset Search,[248] for instance, is expected to contribute to the success of the open data movement by making research data more discoverable to all players in the landscape. Notably, the release of this tool also prompted some organisations holding data (e.g. repositories) to standardise their metadata and align it to the search engine's requirements.

---

240 https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
241 https://www.re3data.org/
242 https://fairsharing.org
243 https://zenodo.org
244 https://figshare.com
245 https://www.elsevier.com/authors/author-services/research-data/mendeley-data-for-journals
246 http://rmets.onlinelibrary.wiley.com/hub/journal/10.1002/(ISSN)2049-6060/about/author-guidelines.html#dataset_submission
247 https://www.nature.com/sdata/policies/repositories
248 https://toolbox.google.com/datasetsearch

Table 12: Data journals (examples)[249]

Journal	*Publisher*
Biodiversity Data Journal	Pensoft
Dataset Papers in Science	Hindawi
Earth System Science Data	Copernicus
Ecological Archives – Data Papers	Ecological Society of America
F1000 Research	Science Navigation Group
Genomics Data – Data in Brief papers	Elsevier
Geoscience Data Journal	Wiley
GigaScience	BGI/Oxford University Press
International Journal of Robotics Research	SAGE
Journal of Open Archaeology Data (JOAD)	Ubiquity Press
Scientific Data	Springer Nature
Data in Brief	Elsevier
Chemical Data Collections	Elsevier

4.2.5 Data citation

Data citation and re-use are inevitably correlated: citing data makes it more discoverable, but a range of other motivations exist in the research data landscape. These include data attribution, data connection, impact, sharing and reproducibility (Silvello 2018).

In 2014, a working group reflecting the multi-stakeholder nature of the data landscape prepared the Joint Declaration of Data Citation Principles.[250] The declaration set out eight principles for the purpose, function and attributes of citations. Citing datasets in journal articles is in some respects relatively straightforward, since it simply extends existing citation practices. However, the process of linking data to articles is not as simple. Callaghan (2014) recommended an investigation into a metadata brokerage service that could simplify the process of passing information between data repositories and journals. Today, Crossref[251] and Datacite[252] fill this gap and provide automated workflows that track and link data and publications, along with initiatives such as Scholix that aim to facilitate the sharing of information about the links between data and literature.[253] The use of ORCID is also simplifying the process of linking datasets (and, of course, research outputs more broadly) to their creators.

Overall, it appears that researchers are increasingly citing datasets. Digital Science et al. (2016) found that nearly 70% of researchers value a data citation as much as an article citation, with a further 10% valuing a data citation more than an article citation. Follow-up work by Digital Science (2017) confirmed these figures and added that citations to non-traditional research outputs are increasing every year.

Nevertheless, Silvello (2018) identifies a number of unresolved issues in the field of data citation, including:

[249] For an extensive list, see https://www.wiki.ed.ac.uk/display/datashare/Sources+of+dataset+peer+review
[250] https://www.force11.org/datacitationprinciples
[251] https://www.crossref.org/blog/linking-publications-to-data-and-software/
[252] https://www.datacite.org
[253] http://www.scholix.org/about

- The *identification* problem – How can “a single resource, a subset of resources, and an aggregation of resources” be easily identified?
- The *completeness* problem – How can appropriate and informative citations be created when data is extracted from a complex and evolving database?
- The *fixity* problem – How can infrastructure providers guarantee that cited data will be accessible in their cited form?

4.2.6 Software citation

As interest in the sharing and citation of research data has grown, there is increasing recognition of the critical role of software in modern research. This is reflected in a growing number of citations of code or software, which traditional publishing systems have struggled to serve in the past (Digital Science et al. 2017). Building on the FORCE11 working group focused on data citation, the FORCE11 Software Citation Working Group issued a set of six software citation principles. The group’s work focussed on making software a citable entity in the scholarly ecosystem (e.g. by uploading software to a suitable data repository) and identified a range of use cases that assume the existence of a citable software object, typically created by the authors/ developers of the software. The FORCE11 Software Citation Implementation Working Group[254] continues to work with relevant stakeholders (publishers, librarians, archivists, funders, repository developers, other community forums with related working groups, etc.) to:

1. endorse the principles
2. develop sets of guidelines for implementing the principles
3. help implement the principles
4. test specific implementations of the principles.

A significant initiative in the area of software is Code Ocean,[255] a cloud-based computational reproducibility platform. Code Ocean has partnered with Taylor & Francis[256] and Cambridge University Press[257] in order to incorporate its services into publishing workflows. This is expected to improve how readers and end users can understand, visualize and reuse the code, but also to enable better software citation via the inclusion of the metadata of linked articles.

4.2.7 FAIR data principles

The FAIR principles[258] for scientific data management and stewardship aim to make data:

- **Findable,** by leveraging metadata and persistent identifiers;
- **Accessible**, through free and open communications protocols;[259]
- **Interoperable,** by using controlled vocabularies, implementing machine-readability and including references where appropriate;

[254] https://www.force11.org/group/software-citation-implementation-working-group
[255] https://codeocean.com
[256] https://codeocean.com/press-release/taking-the-journal-article-to-the-next-level-taylor-francis-partner-with-code-ocean
[257] https://codeocean.com/press-release/cambridge-university-press-and-code-ocean-announce-partnership
[258] https://www.force11.org/group/fairgroup/fairprinciples
[259] It should be noted that accessibility does not necessarily equate to openness: access to data may need to be constrained due to legitimate concerns around privacy, national security, or commercial interests.

- **Reusable,** by highlighting clear licence statements that enable the greatest possible reusability.

The principles are increasingly used as a basis for policy development in countries across the world, and have an important relationship to questions of reproducibility. A recent report (Bruce R 2018) takes stock of how far FAIR principles are supporting open science in the UK and how they are understood and adopted by the research community.

4.2.8 Data sharing and journal policies

As with open access to journal articles, research funders are playing an important role in mandating the open sharing of research data. Funders have introduced policies (or tightened existing ones) to require the deposit and sharing of research data. For instance, Cancer Research UK recognises that the value of data is often dependent on its timeliness and, therefore, requires data to be released no later than the acceptance for publication of the main findings from the final dataset.[260] Recipients of Horizon 2020 are not only bound to share their data but are also required to ensure it is FAIR (see above).[261]

As the prevalence of research data policies from research organisations funders increases, publishers and editors are also paying more attention to standardisation and the wider adoption of data sharing policies. Work to establish a requirement for data access statements to accompany journal articles has had an impact in some areas (Murphy et al. 2017); but their use remains far from comprehensive. Meanwhile, several major publishers have moved to adopt a standardised research data policy framework, and the Research Data Alliance (RDA) Interest Group on Data Policy Standardisation and Implementation is defining a common set of journal data policy requirements.[262] Data-level metrics analogous to article-level metrics have also emerged over the past few years, including altmetrics and the Data Citation Index.[263] These serve to provide feedback on data usage, views and impact (Lin 2014).

4.2.9 Initiatives in the area of research data

The Research Data Alliance (RDA) was started in 2013 by the European Commission, the American National Science Foundation and National Institute of Standards and Technology, and the Australian Department of Innovation. The remit of the organisation in the research data landscape is broad (Treloar 2014b): they aim to improve the sharing of data across barriers (e.g. national, disciplinary, producer/consumer) by the means of working groups and interest groups involving experts from all over the world. The RDA works at the infrastructure level, including hardware, software, content and format standards and human actors.

The following groups are of particular relevance to STM publishers:

- Publishing Data interest group[264]
- Publishing Data Services working group[265]
- Publishing Data Bibliometrics working group[266]

[260] https://www.cancerresearchuk.org/funding-for-researchers/applying-for-funding/policies-that-affect-your-grant/submission-of-a-data-sharing-and-preservation-strategy/data-sharing-guidelines
[261] http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm and http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm
[262] https://www.rd-alliance.org/groups/data-policy-standardisation-and-implementation
[263] http://wokinfo.com/products_tools/multidisciplinary/dci/
[264] https://www.rd-alliance.org/groups/rdawds-publishing-data-ig.html
[265] https://www.rd-alliance.org/groups/rdawds-publishing-data-services-wg.html

- Publishing Data Workflows working group.[267]

Other significant initiatives in the area are supported by CODATA,[268] Science Europe,[269] the Open Research Funders Group[270] and the International Science Council.[271]

4.3 Standards, identifiers and conventions

The technical infrastructure that is crucial to the smooth functioning of the scholarly communication ecosystem is made possible by observance of a range of standards and convention. STM (the Association) has been very active in the development and maintenance of major parts of this infrastructure.

4.3.1 Crossref

Crossref[272] makes research outputs easy to find, cite, link, and assess. They are a not-for-profit membership organization that exists to make scholarly communications better. Crossref is one of the most successful examples of co-operation across the publishing community. It is run by the Publishers International Linking Association Inc. (PILA) and was launched in early 2000, following a decision at the previous Frankfurt Book Fair, as a cooperative effort among publishers to enable persistent cross-publisher citation linking in online academic journals. There are now 2,000 voting member publishers who represent 4,300 societies and publishers, including both commercial and not-for-profit organizations.

Crossref interlinks millions of items from a variety of content types, including journals, books, conference proceedings, working papers, technical reports, and data sets. Linked content includes materials from Scientific, Technical and Medical (STM) and Social Sciences and Humanities (SSH) disciplines. When members register their content with Crossref it collects both bibliographic and non-bibliographic metadata, which is processed so that connections can be made between publications, people, organizations, and other associated outputs. The metadata is preserved for the scholarly record. It is also made available across a range of interfaces and formats so that the community can use it and build tools with it, with the expense paid for by Crossref member publishers.

Most services provided by Crossref, or into which Crossref has a major input, are based on Digital Object Identifiers (DOIs). The DOI is not managed by Crossref but by a separate body, the International DOI Foundation (IDF). This not-for-profit membership organisation is the governance and management body for the federation of Registration Agencies providing Digital Object Identifier (DOI) services and registration, and is the registration authority for the ISO standard (ISO 26324) for the DOI system. Other organisation also use DOIs for their work, most notably DataCite.

Crossref offers a wide array of services to ensure that scholarly research metadata is registered, linked, and distributed. Significant services offered by Crossref which has been undergoing significant changes in the provision of new services and changes in the names of others:

- The Funder Registry is discussed separately [see below]

[266] https://www.rd-alliance.org/groups/rdawds-publishing-data-bibliometrics-wg.html
[267] https://www.rd-alliance.org/groups/rdawds-publishing-data-workflows-wg.html
[268] http://www.codata.org/
[269] http://www.scienceeurope.org/
[270] http://www.orfg.org/
[271] https://council.science/
[272] https://www.crossref.org/

- Bringing new content types into the more formal system of citation and linking – Examples are Preprints and Peer Review Reports which are both becoming more important in scholarly communication
- Becoming an independent source of citation data as recommended by the Initiative for Open Citations[273]
- Offering CrossRef's text and data mining tools (originally Prospect) which supplies 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.
- Providing Crosscheck now called Crossref Similarity Check is a service for editors who want an extra check. The service helps members to actively engage in efforts to prevent scholarly and professional plagiarism by providing their editorial teams with access to Turnitin's powerful text comparison tool, iThenticate

Crossref has recently announce that all its metadata is to be preserved in the CLOCKSS.

Moreover the ongoing thinking within Crossref reaches out into a number of new areas through the work of Crossref Labs,[274] and it is a lead on Metadata 2020 (*Section 4.3.7*).

4.3.2 Funder Registry

The Open Funder Registry (formerly FundRef) is a taxonomy of grant-giving organisations worldwide. This is provided in the form of a freely-downloadable file with a CC-0 licence, which includes funders and their unique identifiers (i.e. DOIs). Donated by Elsevier, the Registry of some 18,000 funders is updated monthly.

The Funder Registry is used to enable "clear, transparent and measurable information on who funded research and where it has been published". It can be used:

- by publishers to analyse the sources of funding supporting their authors and to ensure compliance with funder policies;
- by research funders to track the published results of their grants;
- by research organisations to monitor the publications by their staff;
- by the general public to understand how R&D funding is used.

Publishers play a major role in maintaining funder information: this can be gathered through journal submission systems and acknowledgements sections in articles, and then standardised by matching it against the Funder Registry. This allows publishers to add metadata fields describing funder name, funder id and grant number to their Crossref deposits.[275]

At the time of writing, over 2.8 million unique DOIs include funder data, which represents a significant increase over the figure of 386,000 noted in the 2015 STM report.

4.3.3 National Information Standards Organization (NISO)

NISO is a US standards organisation with an international role which extends well beyond "standards" as usually defined. Its Recommended Practices play a major role in journal

[273] https://i4oc.org/
[274] https://www.crossref.org/labs/
[275] See https://www.crossref.org/blog/global-persistent-identifiers-for-grants-awards-and-facilities/.

publishing. It is recognised as a secure place to hold and develop codes of practice which have been initiated elsewhere. Important standing committees for journal publishers and journal users include:

- the Standardized Markup for Journal Articles;[276] and
- the Tracking Link Origins Working Group, which will develop a NISO Recommended Practice to allow libraries, publishers and other content providers to accurately track the sites/platforms from which incoming links originate when they pass through a link resolver.

NISO also works closely with ICEDIS (EDItEUR) on the ONIX standard as it applies to journals.

4.3.4 Manuscript Exchange

MECA is a new initiative in the field of manuscript exchange – the process of taking an article which has been submitted to one journal and transferring it for submission to another journal. MECA stands for Manuscript Exchange Common Approach and was launched in 2017.[277] This is not the first experiment with review transfer across publishers; the Neuroscience Peer Review Consortium was launched in 2008 as a framework for sharing of reviews between publishers, and the progress / fate of portable peer review initiatives such as Axios, Rubriq and Peerage of Science have been widely covered (see also *section 2.10*). A new NISO working group will continue the work of MECA.[278]

4.3.5 Transfer Code of Practice

The UKSG Transfer Code of Practice 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. The establishment of this code of practice and its general acceptance among larger publishers was a result of cross-sector interactions under the aegis of the United Kingdom Serials Group (UKSG) who handed it over to be managed by NISO.[279] Phillpotts and colleagues provide more background on the evolution of the Code (Phillpotts, Devenport, & Mitchell, 2015). In late 2018 an enhanced transfer alerting service is due to be hosted by the ISSN International Centre.[280]

4.3.6 ICEDIS

ICEDIS (the International Committee on EDI for Serials) is a special interest group for subscriptions, serials and library supply. It was at one time a separate organisation but is now part of EDItEUR, an international standards body particularly concerned with supply chains across publications in general. ICEDIS is responsible for the governance of those standards within its area of competence, which fall into two main groups: descriptive formats to communicate key items of metadata about subscription resources and transactional formats to support trading activities. All of the metadata formats currently supported utilize

[276] https://www.niso.org/standards-committees/jats
[277] www.manuscriptexchange.org
[278] https://www.niso.org/press-releases/2018/05/niso-launches-new-project-facilitate-manuscript-exchange-across-systems
[279] https://www.niso.org/standards-committees/transfer
[280] https://journaltransfer.issn.org

the ONIX XML approach. Transactional messaging standards that EDItEUR supports span several different formats.[281]

4.3.7 Metadata 2020

Metadata 2020[282] is a collaboration that advocates richer, connected, and reusable, open metadata for all research outputs, which will advance scholarly pursuits for the benefit of society. This is an advocacy organisation not a standards organisation which brings together publishers, funders, researchers and librarians. It advocates that richer metadata fuels discovery and innovation; that connected metadata bridges the gaps between systems and communities; and that reusable, open metadata eliminates duplication of effort.[283].

4.3.8 Identity and Disambiguation

Unambiguously identifying the author(s) of a given publication has always been a challenge. Common headaches in the publishing sector include authors with identical names, different arrangements or transliterations of the same name and authors changing names (e.g. upon marriage). Digital author identifiers – unique numbers or alphanumeric codes assigned to individuals as a form of authority control – are a way to support publishers and researchers alike. The most common digital author identifiers are:

- ORCID: arguably the most widespread initiative in terms of author identification, ORCID (Open Researcher and Contributor iD) is a non-profit collaboration involving participants from across the research and scholarly communication worlds (846 organisations, including research organisations, publishers, funders, professional associations and other stakeholders in the research ecosystem). ORCID is a platform-agnostic initiative whose mission is to connect researchers, their contributions and their affiliations. ORCID iDs can be obtained by registering online and include information on education, employment, funding and published works. In addition, ORCID includes an API for system-to-system communication.
- ResearcherID: the ResearcherID system has been developed by Thomson Reuters and is connected to the Web of Science database. ResearcherIDs can be obtained by registering online and provide citation statistics including the h-index.
- Scopus Author ID: unlike the previous two, the Scopus author ID does not need to be created manually. Every author with articles indexed in the Scopus database will be automatically assigned a Scopus Author ID, which includes citation information and the h-index similarly to ResearcherIDs.

At the time of writing, ORCID has issued over 5 million IDs. While comparable figures are not available for ResearcherID and Scopus Author ID, both can be connected to an author's ORCID iD, if desired.

Since 2015, an increasing number of journals and publishers (for a total of over 3,000 journals to date) have started requiring ORCID iDs in their publication workflows. Increasingly, this enables authors to submit papers, and then see their ORCID record automatically updated as their work is published, registered with Crossref, and enters the global citation network. ORCID offers significant benefits in terms of reduced reporting burdens for researchers and enhanced discovery processes, and lays the foundation for trust in a digital research environment.

[281] https://www.editeur.org/59/ICEDIS/
[282] http://www.metadata2020.org
[283] https://scholarlykitchen.sspnet.org/2017/09/06/much-ado-metadata-2020/

4.3.9 International Standard Serial Number (ISSN)

The International Standard Serial Number (ISSN) is an eight-digit serial number used to uniquely identify a serial publication. The ISSN is especially helpful in distinguishing between serials with the same title. ISSN are used in ordering, cataloguing, interlibrary loans, and other practices in connection with serial literature. Other systems are based on ISSNs and it is crucial for publishers to obtain one. The ISSN identifies the publication as such, in reference to its title and its medium. However, it should be noted however that ISSNs can be granted for any content that is ongoing and it is given in some countries not only to series of monographs and magazines but such as outputs as scientific blogs. The ISSN International Centre is in Paris and (uniquely perhaps) the whole structure is based on an initial treaty between UNESCO (also in Paris) and the French government – a treaty which has subsequently been accessed by many governments who usually delegate powers under its terms to their national library. The ISSN International Centre thus assigns ISSN to publications issued by international organizations and by publishers located in countries with no ISSN National Centre. Since January 2018, the ISSN International Centre has charged a fee for ISSN assignment.

The new ISSN portal[284] offers an access interface tailored to various ISSN users. A "freemium" model has recently been adopted, combining free data on the one hand and more complete fee-based data on the other. Non-registered users can browse the free portal, with access to a subset of ISSN data –including ISSN, key title, title proper, country of publication, and medium– which has been made available under a CC-BY-NC-SA (Creative Commons Attribution-NonCommercial-ShareAlike) licence.

The ISSN International Centre has partnered with NISO to offer a new Transfer Alerting Service interface to publishers, who shall state their title transfer by means of the ISSN portal beginning in September 2018. The TAS website has been revamped and shall shortly replace the existing one while retaining the same free services.

4.4 Technology trends

Fashions in technology trends change – semantic enrichment was central, now AI and Blockchain are the latest focus.

4.4.1 Data, analytics and artificial Intelligence

AI in the form of Machine Learning is behind a number of innovations in the last few years, both developed directly by publishers and in tools which will be considered below. Viewed in terms of the Gartner hype cycle,[285] it appears that say that AI has emerged from the trough of disillusionment phase and now has reached the area of realistic expectations and real applications. This is helped by three things:

- Large bodies of data are now available, as well as huge bodies of digital content for text mining;
- Much more computer power at low price; and
- Better software and self-learning algorithms.

A good way to consider this whole area is to think about what the drivers are and what publishers are hoping to achieve. The current underlying driver is about becoming more data-driven, and many publishers are investing in the development of data warehouses,

[284] https://portal.issn.org
[285] https://www.gartner.com/en/research/methodologies/gartner-hype-cycle

dedicated analytics teams, and business intelligence (BI) tools. In layman's terms, organisations with AI capabilities are able to build on these data resources to support strategy development and to provide end users with an improved service. Use cases for AI in STM publishing are still emerging, but the ambition is that it will:

- increase efficiencies in production;
- support better editorial decisions (e.g. commissioning);
- improve the effectiveness of marketing and sales activities;
- deliver better search and discovery of content;
- support personalisation;
- improve the value of content assets; and
- help publishers better understand their users.

Both semantic enrichment (getting more data) and AI (getting more data and using it more effectively) support this agenda. The bottom line is that publishers should not invest in AI but invest in a business problem.

Companies and other initiatives that are actively exploring the use of AI within scholarly publishing include UNSILO,[286] Ictect,[287] Iris.ai[288] Semantic Scholar[289] and Meta,[290] while the new breed of web-scale discovery and analytics tools are also making growing use of AI technologies (see *Section 4.7 Tools, apps and new services for funders and institutions*).

4.4.2 Semantic web and semantic enrichment

It could be argued that semantic enrichment in the traditional sense is being superseded by machine learning in combination with human editorial insight. There are probably few publishers who have seen a return on investment in semantic enrichment, other than those who have sold or licensed their taxonomies

It is convenient to distinguish between the semantic web, and the use of semantic technologies and semantic enrichment of content. Semantic technologies should now be seen within the content of AI. Semantic enrichment is focused on enriching content with data, the end result being a content set with associated extra data. This is typically achieved by analysing the content to understand more about what it contains e.g. identifying drugs / people / locations. That data is then typically used by a different process to deliver an outcome e.g. relatedness. AI is a much broader term, but in many cases in publishing it starts with a similar first step of identifying data in content. The AI algorithm then uses that data to deliver an output, this can be as simple as a subject category or it can be something more complicated like a prediction of the impact of a journal article. In many cases AI is supported by human input to train the machine i.e. machine learning, but in addition to the human training data the computer needs a large quantity of additional data to feed the algorithm. This additional data typically comes, in publishing, from automated analysis of the content.

What has moved on is the assumption that a taxonomy is required in these processes. The more recent content analysis approaches (in semantic enrichment and AI) use more statistical and grammatical analysis, rather than analysis against a taxonomy or ontology. This makes them more flexible and potentially more fine-grained in their output. It also

[286] https://site.unsilo.com/site/
[287] http://www.ictect.com/
[288] https://iris.ai/
[289] https://www.semanticscholar.org/
[290] https://meta.org/

removes the need for the upkeep of such taxonomies and ontologies. There are cases where the use of a taxonomy or ontology are still appropriate, but this should no longer be the assumed starting point.

We are seeing a significant increase in automated content enrichment by the majority of publishers, with these technologies used as specific solutions for distinct tasks e.g. extracting entities like locations to deliver faceted search. They are being talked about less now as they become mainstream technologies in the publishers toolbox (e.g. natural language processing).

Publishers have for many years manually enriched content e.g. keywords, but they are now using technologies to augment or replace some of these manual processes. Today the discussion is more about business benefits, whilst semantic enrichment talks more about data extraction (but may be one of the tools to achieve the business benefit).

4.4.3 Blockchain

The situation with Blockchain (a technology almost entirely associated with Bitcoin until recently) is different. Exactly what Blockchain is and does is not altogether clear to many. One carefully reasoned explanation helps (van Rossum 2017)

However a commentator on the blog Scholarly Kitchen is extremely dubious about what he calls "The hype over blockchain — a new technology that has generated what one observer has called a "mind virus" — (Anderson 2018).

The presentation on STM Tech Trends 2022 cited above examines the use of Blockchain to aid the big problem of loss of trust in science. The question is whether Blockchain is going to be robust and fast enough to solve trust issues and ensure authenticity

There are signs of more enthusiasm. Blockchain in Healthcare Today is a new (2018) journal and the initial articles give potential use cases.[291]

Some major publishers have joined an initiative, announced in March 2018, which focuses on the problems of research reproducibility, recognition of reviewers and the rising burden of the peer-review process. The project will develop a protocol where information about peer review activities (submitted by publishers) is stored on a blockchain. This will allow the review process to be independently validated, and data to be fed to relevant vehicles to ensure recognition and validation for reviewers.[292] This project is organised by Digital Science.[293]

An ambitious project (http://artifacts.ai/) which has a strapline "building the ledger of record for research" is devoted to the needs of the research community.

It would seem that these technologies may well need to succeed, if they are to succeed, in other sectors first before there is the investment needed for the academic publishing niche to become available.

[291] https://blockchainhealthcaretoday.com/index.php/journal

[292] https://www.digital-science.com/blog/news/taylor-francis-and-cambridge-university-press-join-blockchain-for-peer-review-project-blockchainforpeerreview/

[293] https://www.blockchainpeerreview.org/about-the-project/

4.5 Tools, apps and new services

4.5.1 Publishing platforms and APIs

The functionality of publishing platforms is a recurring theme of this report. For many years after journals went online the emphasis was on hosting, an appropriately passive term, which implied that the platform for online journals was the equivalent of a warehouse/distribution centre – mostly concerned with taking orders and sending out issues. Over the last few years most publishers of any size have gradually tied together the various work flows from submission to post publication curation and discovery

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 is 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. In 2018 some analysts consider that whereas API's are valuable for interoperability and product development (and effective API's could mean that a publisher isn't held hostage by one platform), the growing value of API's will be enabling access for text and data mining (TDM). A concrete example is the new Text and Data Mining Site[294] launched by Springer Nature in 2018 with a number of open APIs at its core.[295]

4.5.2 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, in these circumstances, that mobile access to information, tools and services has the potential to create huge benefit.

For a broadly-based publisher typically usage is estimated at 65-75 % desktop vs. 25-35% mobile. Nevertheless, it is standard for platforms to be optimised for mobile. Research on use by laboratory researchers of mobile phones in the UK and the US demonstrates a lot of use but only when away from the desktop, and not by preference because many of the journals they read are not optimized for mobile (or so they report) (PRC 2018c). By contrast, for a leading medical publisher with a large clinical readership mobile usage has become culturally and institutionally embedded within everyone's DNA.

[294] https://www.springernature.com/text-and-data-mining
[295] https://dev.springernature.com/

The latest Meeker report[296] shows that the rate of increase in mobile (and internet) penetration is falling after internet penetration hit 50% of the global population – but the interesting thing about mobile is that it just is and must be *part* of the publisher strategy – it is no longer optional.

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 year-on-year. Gardner and Inger (2018)'s survey shows 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, and in low income compared to high income counties, but still very much a minority activity.

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

Although most of the current interest is generated by the rapidly expanding tablet market, there seem 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 (which can include annotation and highlighting, adding papers to bibliographic systems, and tagging, through to perhaps creating presentations or other new content). In the future there will be ever-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.).

[296] https://www.kleinerperkins.com/perspectives/internet-trends-report-2018

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). 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).

4.5.3 Publishers as workflow providers

It has been one of the axioms of the internet that players should partner with other (specialists) for maximum efficiency and penetration and this is what publishers have done. They have tended not to build but rather use the software of other smaller players.

There has been a change in the last few years in that some big players have concentrated on building, with a complete set of workflow offerings potentially being established by a "duopoly" of Elsevier and Digital Science (Schonfeld 2017b). Some of Schonfeld's key points include:

- Content is giving way to workflow.
- Should the research university outsource more core scholarly infrastructure? And if so to whom, and under what terms?
- The largest publishers are becoming workflow providers. What strategies are available for those left behind?

According to Schonfeld, for Elsevier, "integration" is key. It is building scientific workflow services on individual products and platforms like Mendeley, hivebench, SSRN, Digital Commons, Scopus, and SciVal. To his list can be added bepress, acquired in 2018, together Elsevier's existing Pure, ScienceDirect (for a long time), and Plum Analytics products.

For Digital Science (not a publisher, though owned in part by Holzbrink Publishing Group) there is a different philosophy. These are a portfolio of autonomous companies related to the central company in various ways though benefitting from a productive environment. Each company from Altmetrics, through Dimensions, Figshare, Overleaf, ReadCube and Symplectic have been sold as a suite to universities but mostly do individual deals with publishers and with researchers themselves. Schonfeld admits there is no end-to-end solution here – yet. Most if not all of the Digital Science are mentioned elsewhere in the text of this Report

Schonfeld offers a strategy for those publishers "left behind"on the basis that now the "competitive environment is qualitatively different" (Schonfeld 2017c). There are no silver bullets but a range of common sense suggestions including various types of partnerships such as a not-for–profit alliance usually involving a central provider or maybe a decentralised collection of lightweight tools.

4.5.4 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 developments mentioned in the last report have continued to exert an influence on subsequent initiatives. This is only a selection of what publishers have offered their authors or users and misses out in particular specialized tools for specific disciplines. It also assumes the standard journal article reporting primary research and is not concerned with the review article. The basic model can be stretched a lot – for example the ongoing changes of the construct from F1000Resarch.

Registered Reports: are a form of empirical article in which the methods and proposed analyses are preregistered and reviewed prior to research being conducted. This format is designed to minimise bias in deductive science, while also allowing complete flexibility to conduct exploratory (unregistered) analyses and report serendipitous findings. This definition is taken from the author guidance prepared for the journal *Cortex*[297] which was the first example of an idea which has been taken up by other publishers. For the origin in 2013 see an explanation by the editor.[298]

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. The Article of the Future was rather more than just an enhanced HTML format from the start (Aalperberg 2012). Its bigger impact was its innovative and useful introduction of elements that are separately mentioned later, such as dynamic figures and data visualizations (interactive tables, interactive maps, and 3D visualizations). In many of these cases, Elsevier was actually the first to offer these features beyond "experimentation" and into a fully operation process and workflow and both within and outwith Elsevier there are continuing developments. Some of these features are part of and/or central to the article whereas others are additional to the article.

Visual Abstracts and videos of figures are being used more in spite of the costs, which include a fair amount of hands on production/editorial work. Figure 360 for Cell Press journals is one of these developments flowing from the Article of the Future, enabling better discovery, understanding and outreach of some main points. These are Author-narrated videos of select figures.[299] Visual (or 'graphical') abstracts ("VAs") have been used in some journals for many years, in a number of disciplines and with a wide range of production quality. Such experiments are not confined to major commercial companies. The American Journal of Nephrology for example asks authors to submit a schematized version of their abstracts.

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 is ReadCube. In 2014 the fact that it was based on the PDF was seen as surprising but now needs no defence. ReadCube is part of the Digital Science portfolio of companies. Depending on how its offering is looked at it is not only a reference manager, but also a sharing and discovery mechanism. In 2018 it claims over 40 million users and partners with many major publishers including Taylor & Francis and Springer Nature.[300] If seen as a reference manager, Mendeley is perceived as the main competition.

Article viewing and sharing: ReadCube's Content Sharing Initiative allows users to share subscribed content with non-subscribers via a special link. Since the last edition Wiley and Sage have also taken up this technology as well as Springer Nature. The Wiley approach is explained here.[301] The pilot initiated in 2017 enables sharing to non-subscribers. Meanwhile,

297 https://www.elsevier.com/__data/promis_misc/PROMIS%20pub_idt_CORTEX%20Guidelines_RR_29_04_2013.pdf
298 https://www.elsevier.com/editors-update/story/peer-review/cortexs-registered-reports
299 https://www.cell.com/figure360
300 https://www.readcube.com/press
301 https://www.wiley.com/WileyCDA/PressRelease/pressReleaseId-130882.html

the publisher-optional Kudos Shareable PDF is intended to support comparative, copyright compliant sharing across multiple Scholarly Communication Networks. [302]

3D PDF: Publishers in fields heavily reliant on 3D information – earth sciences, geophysical, geospatial, engineering, medical scanning, etc. – are starting to adopt the 3D PDF format. Again Adobe Acrobat, now in their 25th year, seems to have been able to offer enhancement via PDF based services after so many prophecies that their offering would soon be superseded.[303]

Article versions. The taxonomy of article versions has already been discussed (*Section 2.2.2*) and mention has been made there of platform developments that allow articles to be updated or expanded, while rigorously preserving the original version(s) and its publication record. As mentioned there, F1000Research have worked closely with Crossref in how to handle a more fluid concept of the journal. 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. It has been clear for some time that publishers need to be more serious about their stewardship of the article and especially changes after the version of record. How this is to be handled is a work in progress

The hub and spokes mode. This is the concept behind another new format, combining an article with its peer review reports. Many initiatives can be found on some form of Open Peer Review.[304]

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. The use of streaming media is discussed in 4.5.4.

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, Springer Nature, PLOS) use a service offered by Figshare for storage and visualisation of such datasets.[305] Figshare visualizes (2018) over 1000 different file formats in its supplementary data Viewer, lowering the barrier at submission stage to what file types can be received as supplementary data. The recent work by CIBER with early career researchers (PRC 2018) demonstrates that ECRs in the UK and the US have a high level of recognition of Figshare and also a preference for depositing data as supplementary data.

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. This service is now re-branded as Research Elements.[306]

[302] https://blog.growkudos.com/2018/05/22/s-pdf-doubles-sharing/
[303] https://theblog.adobe.com/evolution-digital-document-celebrating-adobe-acrobats-25th-anniversary/
[304] See for example https://www.elsevier.com/reviewers-update/story/innovation-in-publishing/is-open-peer-review-the-way-forward
[305] An example of the use of Figshare for this purpose, but directed at researchers, is to be found here: https://research.unimelb.edu.au/infrastructure/doing-data-better/figshare/features-roadmap
[306] https://www.elsevier.com/authors/author-services/research-elements

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.

4.5.5 Streaming Media

In a previous section the problems of long term preservation of "dynamic" scholarly content is raised as one not yet attended to by libraries (*Section 2.17*). This section provides the evidence for the need. It is based mainly on a survey by Renew Consultants (2018) who interviewed organizations involved in scholarly communication to explore their use of video and audio. The key conclusions arising were:

- engaging with video and audio is critical to future success, as it may bridge the gap between the activities we undertake now and the way that future generations are likely to choose how to consume their information; and
- video and audio is not new; it is just that the STM publishing industry has only just started to really grapple with how it might fit within the landscape – or, perhaps, how it might fit within the video and audio landscape.

Among the case studies were a number of STM journals publishers who broadly speaking were either encouraging or allowing authors to submit video summaries, admitting video as supplementary materials, or (most interestingly) including video as central to the article.

4.5.6 Open Annotation

Open annotation, a new open specification for web-based annotation, offers the potential for richer types of commentary and discourse to be supported in a layer sitting over journal (and other academic) content (Carpenter 2013). Approval of open annotation by the W3C as a web standard in February 2017 changed everything by establishing a foundation upon which interoperable systems could be built. Naydenov and Staines (2018) explain the advantages of annotations over comments

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 from the content annotated website or that it installations of any new software on its part. 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 locations within pages, right down to the sentence or word 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 (I Annotate). 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.

4.5.7 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. This is an outline description of a range of services offered both by publishers and alternative platforms such as HighWire but it is certainly not exhaustive:

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.) Peerwith[307] provides a peer-2-peer marketplace enabling academics to access a wide range of experts and services.

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. There are still a big range of tools sometimes concentrated in an offer which links an online editorial system with a production one. An example is Aries Systems.

Information and alerts: Tracking and status reporting during production; citation alerts following publication. These functions are often combined.

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). Then there is Impact Vizor™ - visual analytics for content value and its companion Usage Vizor™ gives publishers immediate insight into article-level, institutional, and turnaway usage so that they can make more informed, timely, and evidence-based editorial and business decisions about their content.

4.5.8 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.

Overleaf (originally WriteLaTeX) now part of Digital Science, offers such a service, which in 2014 claimed to have over 150,000 users at more than 1000 institutions and over 2 million documents created. In mid-2018 2 million users are claimed to be "enjoying the easiest way to create, collaborate and publish online".

Other academic online writing tools include Authorea (which was acquired by Atypon/Wiley in mid-2018), Fidus Writer and shareLaTeX. The Authorea platform offers publishing

[307] https://www.peerwith.com/

services in addition to writing. 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. These are all slightly different offerings with none having the impact of Overleaf to date.

At present all these services are used by a tiny minority of scientists. This may change with publisher endorsement and integration. For instance, there is an offer to publishers of 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.

4.6 Tools, apps and new services for funders and institutions

There is a growing market for services built on STM publishing information in the form of research analytics: research information management systems linked to analytics tools. According to OC&C (cited in Springer Nature 2018), approximately 14% of the materials/content budget of academic libraries is currently allocated for abstracting and indexing databases and platforms, and the academic spend market for these databases and platforms is expected to continue to show robust growth over the coming years.

The idea behind such services 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.[308] CRIS systems integrate information on an institution’s 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 who have traditionally been 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; Clarivate Analytics, whose CRIS Converis (previously AVEDAS, acquired by Thomson Reuters and integrated with its Research in View service), and InCites Benchmarking and Analytics suite draw on Web of Science data; and Digital Science, which has long had a presence in the market through its ownership of Symplectic. 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[309]; and data licensing for internal analysis. There are also some non-commercial national-level initiatives such as METIS (Netherlands) and CRIStin (Norway). In North America, there has historically

308 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

309 An interesting example is the series of reports Elsevier have prepared for the UK Department of Business, Energy and Industrial Strategy on the international competitiveness of the UK research base (e.g. Elsevier 2016)

been a greater focus on Faculty Activity Reporting (FAR) Systems than CRIS systems, but interest in the latter is now growing.[310]

New entrants to the analytics market in recent years include:

- 1Science, a sister company to Science-Metrix, offers a suite of products based on a curated collection of 90 million articles. The company claims to offer significantly improved coverage of journals outside the West, thereby providing a more comprehensive discovery and analytics solution than Scopus or Web of Science.
- Dimensions brings together data from six of Digital Science's portfolio companies ReadCube, Altmetric, Figshare, Symplectic, Digital Science Consultancy and ÜberResearch, aiming to offer a unified view of research from input to output to impact. The current product, which was relaunched in early 2018, combines a citation database, a research analytics suite, and modern article discovery and access functionality. The methodology of construction of the Dimensions dataset and user interface is outlined in Hook et al (2018).
- wizdom.ai offers research intelligence for researchers, institutions, funders and publishers through interactive dashboards and visualisations for decision making. Using AI to interconnect publications, funding, patents and clinical trials, it aims to provide holistic view of global research activity. Owned by Informa, wizdom.ai is a sister company to Taylor & Francis, and was developed by the team behind the colwiz reference manager (now wizdom.ai research assistant).

This new generation of analytics tools reflect a growing tendency to see research data as a commodity, with value being generated instead from the algorithms used to enrich, interpret and surface that data to meet a range of user needs.

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.

4.7 Social media and scholarly collaboration networks

This section is concerned with social media and scientific collaboration networks. They provide a function that is as important if different from conventional channels of scholarly communication such as e-mails.

4.7.1 Social Media

During the last decade 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

[310] Popular FAR systems in the U.S. include Digital Measures and Academic Analytics, see https://conservancy.umn.edu/bitstream/handle/11299/188632/Bakker%20McBurney%20Poster%20ALA.pdf?sequence=1&isAllowed=y

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.

Other studies 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.

The most recent authoritative review on scholarly use of social media and altmetrics is by Cassidy Fujimoto and colleagues (Sugimoto 2017) There is a lot of literature reviewed and from particular studies mentions of use of specific social media are recorded rather than the extent of the use by (for our purposes) researchers in science disciplines

It was the view expressed in the last edition of the STM Report that 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. Scientific social networks have grown very rapidly (see below). 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 play a part in content discovery and sharing.

In the last few years however there is some evidence that use of social media by researchers in general has not increased as much as might have been anticipated. The reasons for this lack of change probably in part reflect the comments made in the last decade by Crotty (see above). Donelan (2016) based on research done in 2014 summarised: "Social media are not currently viewed by all STEM academics as an essential, or in fact necessary, tool for carrying out their daily tasks".

A recent survey on discovery (Gardner and Inger 2018) finds that Social Media (which for this purpose included Mendeley and Research Gate as well) have become significantly more important in all subject areas since 2012 but in the sciences usage for this purpose has now peaked.

Social media is a broad term and different use cases need to be analysed separately. Here is one definition from Donelan: in her research, the term 'social media tool' includes blogs, Twitter, social networking sites (such as Facebook and professional based sites such as LinkedIn), media sharing sites (such as YouTube, Flickr and SlideShare) and social bookmarking sites (such as Delicious) (Donelan 2016). For our purposes in this sub-section we are concerned primarily with Twitter and with Facebook while other social networking sites are considered below

What about media sharing sites and social bookmarking sites? There is no doubt that journal authors use the three media sharing sites mentioned and many routinely put up presentations on YouTube and SlideShare. As far as we know social bookmarking is not a practice of the majority of STM researchers.

Recent Ciber work has shown that Twitter is used by researchers in much the way that it is used by other groups. As of 2016, Twitter had more than 319 million monthly active users. Researchers within the general embrace of Twitter are part of their own communities and whether or not such communities are active in use of the medium for professional purposes will to a large extent depend on use by individuals. Twitter provides information about new work in a specific field and enables publicity for one's own work. There is evidence that some early career researchers drop in and out of engagement with Twitter depending on how busy they are and whatever demands in their research they have at any particular time (PRC 2018).

Many researchers come across Twitter at conferences where its use is sometimes encouraged but perhaps less so than it once was (Sugimoto et al 2017). Studies listed certainly concur in confirming that only a small minority of those attending a conference use Twitter in this way.

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.

When Chinese researchers are asked whether or not they communicate with their research colleagues using social media they routinely say that they do not but when WeChat is mentioned they admit to using for that purpose this Chinese multi-purpose messaging, social media and mobile payment app. It is quite different in its scope from anything developed outside China. In 2018 WeChat has hit one billion monthly users for the first time, the owner of the Chinese messaging app has revealed. The vast majority of its users are based in China, where it's known as Weixin, and where the recent Lunar New Year boosted it past the billion milestone[311]. In 2017 a major university[312] suggested that posts on social media and other un-peer-reviewed activity might be taken into account in promotion or similar decisions There is no evidence in 2018 that any other institutions followed this lead.

A recent survey by Springer Nature described in a blog[313] is primarily concerned with social media as a source of information as well as a way of communicating. In the Nature survey conducted in 2014, the most-selected activity on both ResearchGate and Academia.edu was simply maintaining a profile in case someone wanted to get in touch (68%). This year's survey revealed also that over three quarters of respondents stated that they use social media and SCNs for discovering and/or reading scientific content (Nature's 2014 study 33%). This is not the conclusions of the recent ongoing study of ECRs by CIBER where their behaviour at least in the UK and the USA is currently more aligned to the 2014 results (PRC 2018).

Another result of the Springer Nature survey is that "50% of professional users said they accessed Facebook on a daily basis". It is not clear whether access is concerned with personal or professional matters. The CIBER results suggest that many ECRs avoid using

[311] https://www.bbc.co.uk/news/business-43283690
[312] https://www.nature.com/news/top-chinese-university-to-consider-social-media-posts-in-researcher-evaluations-1.22822
[313] http://blogs.nature.com/ofschemesandmemes/2017/06/15/how-do-researchers-use-social-media-and-scholarly-collaboration-networks-scns

Facebook for professional purposes but they do engage with local or professional groups on Facebook. Sugimoto quotes evidence for significantly less professional than personal use: indeed when Facebook use for professional purposes is investigated specifically only a quarter use this medium.

Sugimoto surveys a range of articles that have looked at the use of blogs by researchers both as producers and as consumers. What she has found seems to confirm the relatively small number of bloggers compared with those who consume them, particular in medicine (Sugimoto 2017). It is not surprising that ECRs in the Harbingers project (PRC 2017) are less likely to blog during the three years of the project. They claim lack of time. They associate blogs with outreach and would like to do more blogging in the future. There is no doubt that blogging is perceived as something legitimate to do. Researchers may (and do) make use of other social media for a variety of purposes sometimes with encouragement of their publishers and/or as part of the programme proposed by services like Kudos.

4.7.2 Social Collaboration Networks (SCN)

One of the most significant developments in the STM landscape is the rapid rise of SCNs in the academic environment, most notably ResearchGate and Academia.edu. SCNs are platforms aimed at connecting researchers with common interests. Users create profiles and are encouraged to list their publications and other scholarly activities, upload copies of manuscripts they have authored, and build connections with scholars they work or co-author with. By 2014, awareness of the networks was already high, especially for ResearchGate in STM fields and Academia.edu in social sciences and humanities (Van Noorden 2014c).

Within the last few years ResearchGate has become by far the most important for researchers in the sciences. It is a ‘pure’ social networking site whose mission is to “connect researchers and make it easy for them to share and access scientific output, knowledge, and expertise." Launched in 2013, ResearchGate has over 15 million users (as of April 2018) and stores tens of millions of articles. It needs to be mentioned that other outputs are uploaded by researchers such as presentations at conferences. As seen in Section 3.1, a free-to-read version of an article is more likely to be found on ResearchGate than in all open access repositories combined. Much of this content is posted without permission and contrary to the STM Principles (STM 2016). Recent research has shown that, at least among early career researchers, at least 80% in the sciences are likely to have a presence (PRC 2017) but this does not necessarily mean active engagement.

Academia.edu is also meant as a platform to share research papers, and the company's mission is to “accelerate the world's research." The website was launched in 2008 and reportedly has 36 million users, but fewer articles than ResearchGate. Academia is probably still used more by social sciences and researchers in the humanities but in general it is a less welcoming site and there is pressure to pay for a premium service that is off-putting to many researchers.

It is worth pointing out research already mentioned rather surprisingly unveils a significant usage of LinkedIn by many early career researchers. They consider that a profile on this professional site is worth having not just by those considering a career in industry or in government

In response to the for-profit nature of Academia.edu and ResearchGate, some academics have created ScholarlyHub, a non-profit open access repository that gives access to academic papers, research projects and researchers. The platform aims to become a member-run and owned SCN that aggregates research, teaching and other professional resources. The membership of the advisory board is slanted towards the humanities and social sciences and as yet there is no evidence of any serious take-up

Moreover, other services are exploring the potential of social networks as a vehicle to gather and disseminate research. Mendeley is a popular reference manager software produced by Elsevier that allows sharing research papers and online collaboration among over eight million researchers. Another network, colwiz (collective wizdom), launched in 2011 and provided interactive digital collaboration and free reference management services for researchers in academia, industry and government globally. They also developed the ACS Chemwork platform for the American Chemical Society. In 2013, Taylor & Francis incorporated colwiz's interactive PDF reader into their journals platform and in 2017 its parent, Informa, acquired the whole company. In 2016, the company also developed the wizdom.ai research intelligence product (see section 4.7 *Tools, apps and new services for funders and institutions*). At the time of writing, colwiz functionality was being merged into wizdom.ai to develop an intelligent research assistant under the wizdom.ai brand.

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

The popularity of SCNs is perhaps an indication of the way in which authors prefer to share their articles. However - unlike open access repositories - academic social networks do not routinely check for copyright compliance, and therefore much of their content is illegally posted and hosted (Jamali, 2017).

Uncertainty over the copyright status of academic papers hosted on social networking sites raises concerns over the persistence of such content (Chawla, 2017) and the ethics of ASN services themselves (Fortney & Gonder, 2015). It seems likely that ECRs at any rate are less likely to make their research publications available for immediate downloads rather than invite an invitation to share.

4.8 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 Johnson, Fernholz and Fosci, 2016; Clark, Jensen, & Campbell, 2014; and Smit & van der Graaf, 2011).

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 *2.16.8 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; limited incentives for researchers to use the technique 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 their comments that they have never been approached by a client about text mining” (Inger & Gardner, 2013). A 2016 study of the use of TDM for public research in the UK and France suggested relatively little had changed in the intervening years, despite the introduction of a copyright exception for TDM in the UK in 2014 (Johnson, Fernholz and Fosci 2016).

The challenges associated with TDM are being addressed via 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. The statement has been subsequently updated, with most recent version dating from 2017 (STM 2017a).
- CrossRef's text and data mining tools (originally Prospect): 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) offers a service targeted at life science companies. RightFind XML for Mining provides access to approximately 10 million articles in XML content from more than 60 STM publishers with normalised metadata, and consistent licensing terms for mining the content for internal research. The system reduces the necessity for one-off licensing negotiations, along with the associated administration costs, while providing additional royalties to rightsholders when their content is used for textmining.

4.9 Developments in preprint use and preprint servers

Preprints are commonly defined as the original manuscripts that an author has not yet submitted for peer-review; however, dozens of potentially conflicting and overlapping definitions are currently in use (Tennant et al, 2018).

When a preprint is posted, authors maintain full copyright and there is an underlying expectation that the work will be submitted as a peer-reviewed articles in the future. This, however, is not always the case, and it is estimated that about 20% of preprints posted in arXiv[314] (a preprint server) are not published in journals.[315]

Preprint servers represent the online platforms or infrastructure designed to host preprints. They may include a combination of peer reviewed and non-peer reviewed content, from a variety of sources and in a range of formats. Sharing and using preprints has been common among researchers since before the internet era, when CERN managed a repository for manuscripts in the physical sciences. This idea was first adapted online by arXiv, which also functions as a repository (see *Section 3.3.3*). An overview of some significant players in the preprints landscape is provided in Table 13.

Table 13: Overview of significant platforms in the preprints landscape.

Platform (year founded)	*Description*	*Type and URL*
arXiv (1991)	ArXiv is a preprint server in the fields of mathematics, physics, astronomy, electrical	Preprint server https://arxiv.org

314 https://arxiv.org

315 https://blog.scielo.org/en/2017/02/22/scielo-preprints-on-the-way/#.W5omK8CZ2Hs

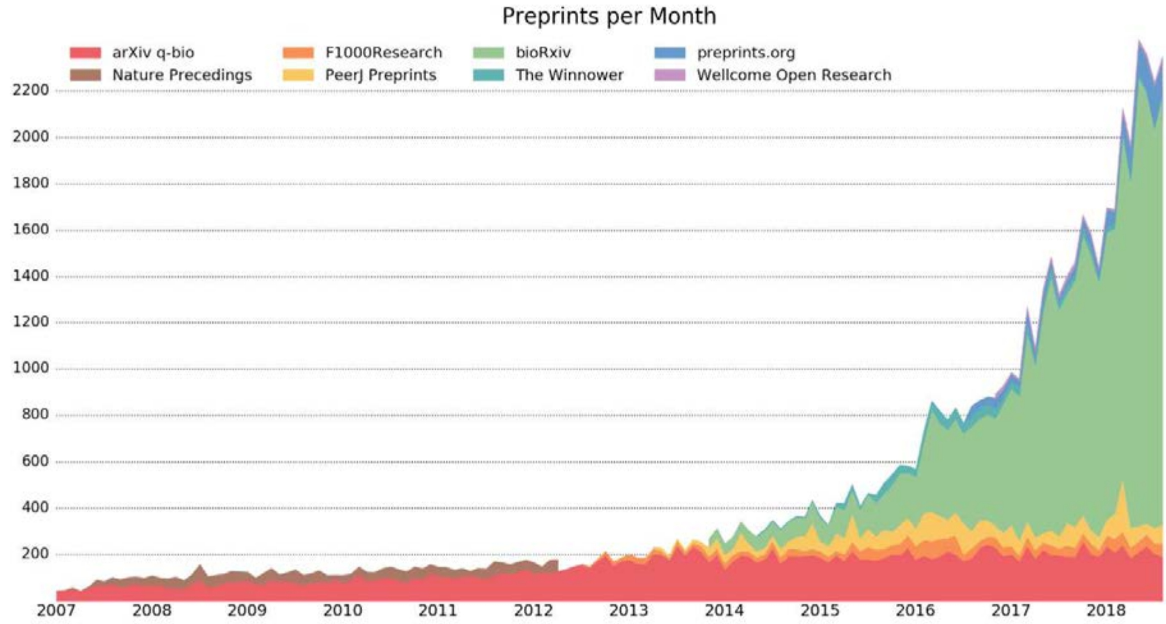
	engineering, computer science, quantitative biology, statistics and quantitative finance. In some fields, such as mathematics and physics, almost all articles are posted on arXiv.	
SSRN (1994)	Started as a social science network, SSRN now also covers biology, chemistry and even medicine. Owned by Elsevier, it is also integrated into a larger workflow offering including Mendeley and incorporates Sneak Peak from Cell Press.[316]	Preprint server https://ssrn.com/en/
bioRxiv (2013)	BioRxiv, a preprint repository specializing in life sciences, doubled the number of preprints listed over the last year and received funding in 2017 from the Chan Zuckerberg Initiative to expand the preprint server and add more software tools.	Preprint server https://www.biorxiv.org
PeerJ Preprints (2013)	PeerJ is an OA peer-reviewed journal in the fields of biological sciences, environmental sciences, medical sciences and health sciences. The platform now includes an option to post preprints on their dedicated PeerJ Preprints page. It is possible to submit bioRxiv preprints for peer review at PeerJ.	Preprint server https://peerj.com/preprints/
OSF Preprints (2016)	The Open Science Framework (OSF) Preprint search engine hosted by the Center for Open Science (COS), listed a total of 30 preprint service providers, with over 2.2 million searchable preprints as at 13 September 2018.	Preprint indexing service https://osf.io/preprints/
Preprints.org (2016)	Preprints is a non-profit platform supported and funded by MDPI. It is a multidisciplinary platform, covering all subjects from arts and humanities to mathematics and computer science.	Preprint server https://www.preprints.org
ChemRxiv (2017)	ChemRxiv,which is co-owned, and collaboratively managed by the American Chemical Society (ACS), German Chemical Society (GDCh) and the Royal Society of Chemistry. The chemical community has long been hostile to preprints (Carà et al 2017), so the initiative represents a litmus test for the adoption of preprints across a wider range science	Preprint server https://chemrxiv.org

316 https://www.elsevier.com/en-gb/solutions/ssrn

	disciplines.	

Overall, there are a number of indicators of significant growth in the area of preprints. For example, the Crossref database shows an increasing take up from an admittedly low start in 2016, when DOIs were first offered for preprints.[317] Meanwhile the Prepubmed service shows a very rapid increase in uploads over recent years, most notably for bioRxiv (Figure 48).

Figure 48: Preprint uploads per month for selected servers (source: www.prepubmed.org, accessed 4 September 2018)



With the numbers of preprints rising, tools and service have emerged to address the need of enhanced discoverability. In addition to servers, other platforms and services have started hosting and using preprints – two are particularly worthy of consideration:

- **Open research platforms:** while most manuscripts uploaded to preprint servers are later published in a journal, open research platforms such as F1000 offer the option to publish both preprint and version of record on the same platform. Preprints are immediately published on the platforms, and then reviewed by peers using an open peer review process. The final peer-reviewed article and the reviewer's comments are then published alongside the preprint.
- **Overlay journals**: these are online journals that do not do not host manuscripts.[318] Overlay journals provide links to preprints hosted in a repository/server and provide peer review services ranging from an editorial introduction to a full report. Overlay journal provide most of the services of a regular journal (peer review, audience curation, discoverability) except for preservation. Many overlay journals are free of charge for both the author and the reader, because their running costs are much smaller compared to regular journals. – with the overall cost per submitted paper as low as $15. Today, overlay journals (long ago proposed) have in most cases yet to

[317] https://www.crossref.org/blog/preprints-growth-rate-ten-times-higher-than-journal-articles/

[318] For an introduction to Overlay Journals see http://discovery.ucl.ac.uk/19081/1/19081.pdf

make any mark, though an exception in terms of quality if not quantity of articles may be Discrete Analysis founded by the Fields medallist Timothy Gowers.[319]

There are other examples of new developments which are relevant though not central. The open access database, library and OA publisher SciELO, which currently hosts 74,000 OA articles, is planning to launch a preprint server in mid-2018. Funding bodies like the NIH, Cancer Research UK and BBSRC are also actively encouraging their authors to share their preprints before publication, with the NIH explicitly allowing "interim research products" (i.e. preprints) to be cited anywhere other research products are cited.

Other than supporting the open science culture, the posting of preprints is seen as beneficial for a range of reasons, including:

- the scope to receive early feedback from the community;
- rapid dissemination;
- an increased pace of discovery and reuse;
- a reduced risk of "scooping";
- easier text and data mining;
- the exposure to a wider audience; and
- the opportunity to reduce predatory publishing.

At the same time, concerns exist when it comes to preprints, particularly around the risk of reduced rigour and reproducibility. This appears to be important in cases where preprints are picked up by media outlets and articles or blog posts are published based on non-peer-reviewed content. Critics of preprints feel that this approach to publishing may lead to confusion and distortion of the public's understanding of science, and even pose a risk to public health. However, the debate is lively, and many supporters notes that science journalists are used to consulting with experts before publishing stories and preprints would make little difference to current practices.

Another issue with preprints is the so-called 'Ingelfinger Rule', which originally stipulated that The New England Journal of Medicine (NEJM) would not publish findings that had been published elsewhere, in other media or in other journal. Once widely adopted, today a wide range of journals will now consider preprints for publication.[320] Nevertheless, many authors remain concerned that the posting of a preprint may preclude acceptance by their journal of choice.

Furthermore, some publishers fear that the updated version of a preprint, which is somewhat similar to a peer-reviewed article, may be a business threat to their practices.[321] Whether such a risk will materialise remains to be seen, but the abovementioned rise of overlay platforms that peer-review preprints (e.g. biOverlay)[322] offers a view on what the landscape might look like if traditional journals were to be bypassed. Concerns have also been raised over the loss of citations from journals to preprints servers, with well over 8,000 citations to bioRxiv reported on Web of Science.

Further growth in preprints and preprint servers seems likely but will need to be accompanied by clarification of the relationship between preprints and peer-review, and the development of new community norms for licensing and citation of preprints. How the role of preprints impacts on the role of journal articles has yet to be established and, as we have seen, all sorts of models are proposed, some of which may not prove to be sustainable.

---

319 http://discreteanalysisjournal.com

320 https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy

321 https://www.highwirepress.com/sites/default/files/documents/London2018/13%20-%20Preprints%20and%20the%20Journal%20Ecosystem%20-%20Katherine%20Brown.pdf

322 https://www.bioverlay.org/post/welcome/

5. References

Aalbersberg,I et al (2012) *Elsevier's Article of the Future* Insights – 25(1), March 2012

Adie, E. (2014). "Attention! A Study of Open Access vs Non-Open Access Articles." *Altmetric* (blog). October 23, 2014. https://www.altmetric.com/blog/attentionoa/

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. (2013). Collaborations: The Fourth Age of Research. *Nature*, May 29, 2013. https://doi.org/10.1038/497557a.

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/

Adema, J. and Stone, G. (2017). *Changing publishing ecologies; A landscape study of new university presses and academic-led publishing.* Jisc. Retrieved from http://repository.jisc.ac.uk/6666/1/Changing-publishing-ecologies-report.pdf

Akeroyd, J. et al. (2018) Research to support the British Library's work on emerging formats. Retrieved from; http://ciber-reserach.eu/download/20180201-British_Library-Emerging_Formats-Final_Report.pdf

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). *The Scholarly Kitchen Blog.* Retrieved from http://scholarlykitchen.sspnet.org/2014/10/21/updated-80-things-publishers-do-2014-edition/

Anderson, K. (2017). "Creating a Safety Net - Why Double-Dipping Is the Wrong Term and the Right Approach." *The Scholarly Kitchen Blog.* Retrieved from https://scholarlykitchen.sspnet.org/2017/11/20/creating-safety-net-double-dipping-wrong-term-right-approach/.

Anderson, K [2018] "Can Blockchain Withstand Skepticism? An Inquiry" *The Scholarly Kitchen* retrieved from https://scholarlykitchen.sspnet.org/2018/04/05/can-blockchain-withstand-skepticism-an-inquiry/].

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

Anderson, R., Denbo, S., Graves, D., Haigh, S., Hill, S., Kalfatovic, M., Kaufman, R., Murray-Rust, R., Shearer, K., Wilder, D., & Wise, A. *(2016) Report from the "What Is Open?" Workgroup*. http://dx.doi.org/10.13021/G8XK5R

Anderson, R. (2017). When the Wolf Finally Arrives: Big Deal Cancellations in North America. Retrieved from: https://scholarlykitchen.sspnet.org/2017/05/01/wolf-finally-arrives-big-deal-cancelations-north-american-libraries/

American Libraries. *The Scholarly Kitchen (blog)*. Retrieved from https://scholarlykitchen.sspnet.org/2017/05/01/wolf-finally-arrives-big-deal-cancelations-north-american-libraries/.

Anderson, R (2017a). Access vs. Accessibility in Scholarship and Science. *The Scholarly Kitchen* retrieved from https://scholarlykitchen.sspnet.org/2017/11/06/access-vs-accessibility-scholarship-science/

Anderson, R. (2018) *Scholarly Communication: What Everyone Needs to Know*. Oxford University Press.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.

Baldwin, J., and Pinfield, S. (2018) "The UK Scholarly Communication Licence: Attempting to Cut through the Gordian Knot of the Complexities of Funder Mandates, Publisher Embargoes and Researcher Caution in Achieving Open Access." *Publications* 6, no. 3: 31. https://doi.org/10.3390/publications6030031.

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

Berg J. 2010. Measuring the scientific output and impact of NIGMS grants. *NIGMS Feedback Loop Blog*. Retrieved from https://loop.nigms.nih.gov/2010/09/measuring-the-scientific-output-and-impact-of-nigms-grants/

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. (2015) "Have the 'Mega-Journals' Reached the Limits to Growth?" *PeerJ* 3: e981. https://doi.org/10.7717/peerj.981.

Björk, B. (2017) Growth of hybrid open access, 2009–2016. *PeerJ* 5:e3878 https://doi.org/10.7717/peerj.3878

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/

Bogich, T., Ballesteros, S., Berjon, R. (2016) On the marginal cost of scholarly communication science.ai.2016

Bohannon, J. (2016) "Who's Downloading Pirated Papers? Everyone." Science | AAAS. Retrieved from http://www.sciencemag.org/news/2016/04/whos-downloading-pirated-papers-everyone

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.

Bosman, J., and Kramer, B. (2018). "Open Access Levels: A Quantitative Exploration Using Web of Science and OaDOI Data." *PeerJ*. *[preprint].* https://doi.org/10.7287/peerj.preprints.3520v1.

Brand, A., Altman, M., Hlava, M., Scott, J. (2015) Beyond authorship: attribution, contribution, collaboration and credit *Learned Publishing* 28.151-155

Brembs, B. (2015). "What Goes into Making a Scientific Manuscript Public?" Bjoern.Brembs. Blog (blog). Retrieved from. http://bjoern.brembs.net/2015/06/what-goes-into-making-a-scientific-manuscript-public/

Bruce, R and Cordewener, B. (2018) *Open science is all very well but how do you make it FAIR in practice?* LSE Impact Blog July 2018 retrieved from http://blogs.lse.ac.uk/impactofsocialsciences/2018/07/26/

British Academy (2018) Open access and monographs: Where are we now? 2018 *British Academy*

Cabinet Office Government of Japan Expert Panel on Open Science *Promoting Open Science in Japan* 2015 retrieved at http://www8.cao.go.jp/cstp/sonota/openscience/150330_openscience_en1.pdf]

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.

Carà, P. D., Ciriminna, R. and Pagliaro, M. (2017). Has the Time Come for Preprints in Chemistry? *ACS Omega*, 2 (11), pp 7923–7928 **DOI:** 10.1021/acsomega.7b01190

Carlin, L. (2017). Delta Think Releases Updated Open Access Market Sizing Analysis*. Delta Think* (blog). Retrieved from https://deltathink.com/delta-think-releases-updated-open-access-market-sizing-analysis/

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/

Casadevall,Arturo and Ferric C. Fang (2010) *Reproducible Science* Infect. Immun. December 2010 vol. 78 no. 12 4972-4975 retrieved from http://iai.asm.org/content/78/12/4972.full] R

CAST (2017) *2017 Blue Book of China's STM Journal Development* China Association for Science and Technology

Castelvecchi, D. (2018) Google unveils search engine for open data. Nature 561. http://doi.org/10.1038/d41586-018-06201-x

Chan L, Kirsop B, Arunachalam S (2011) Towards Open and Equitable Access to Research and Knowledge for Development. PLoS Med 8(3): e1001016. https://doi.org/10.1371/journal.pmed.1001016

Chawla, D. (2017). Publishers take ResearchGate to court, alleging massive copyright infringement. *Science News.*

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 Ltd. (2011). The Journal Usage Factor: Exploratory data analysis. Retrieved from http://ciber-research.eu/CIBER_news-201103.html

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/

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 Blog.* Retrieved from http://scholarlykitchen.sspnet.org/2013/07/15/game-of-papers-elife-bmc-plos-and-embo-announce-new-peer-review-consortium/

Clarke, M. (2015).*The Changing Nature of Scale in STM and Scholarly Publishing.* The Scholarly Kitchen blog retrieved from https://scholarlykitchen.sspnet.org/2015/06/25/the-changing-nature-of-scale-in-stm-and-scholarly-publishing

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

Conrad, L. (2018a). Roadblocks to Accessibility *The Scholarly Kitchen blog* retrieved from https://scholarlykitchen.sspnet.org/2018/01/25/roadblocks-to-accessibility/

Conrad L. and Kasdorf B. (2018b). Making accessibility more accessible to publishers *Learned Publishing 2018; 31; 3-4*

Corti, L. (2014) *Managing and Sharing Research Data* Sage Publications

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 (2018) The COUNTER Code of Practice retrieved from https://www.projectcounter.org/code-of-practice-sections/general-information/.

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

Crawford, W. (2018). *GOAJ3: Gold Open Access Journals 2012-2017*. Livermore, California: Cites & Insights Books. Retrieved from https://waltcrawford.name/goaj3.pdf.

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/

Crotty, D. (2018). Revisiting: Six Years of Predatory Publishing. *The Scholarly Kitchen Blog*. Retrieved from https://scholarlykitchen.sspnet.org/2018/08/14/revisiting-six-years-predatory-publishing/.

Cyranoski, D. (2018) “China Introduces Sweeping Reforms to Crack down on Academic Misconduct.” News. Nature. https://doi.org/10.1038/d41586-018-05359-8.

Czujko, R., & Chu, R. (2015). *Scholarly and scientific publishing workforce in the U.S*. AAP/AIP. Retrieved from https://publishers.org/our-markets/professional-scholarly-publishing/professional-and-scholarly-publishing-23-billion

D’Andrea, R., and O’Dwyer, J.P. (2017). “Can Editors Save Peer Review from Peer Reviewers?” PLOS ONE 12, no. 10: e0186111. https://doi.org/10.1371/journal.pone.0186111.

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

Darley, R., Reynolds, D. and Wickham, C. (2014) “Open Access Journals in Humanities and Social Science.” British Academy. Retrieved from https://www.britac.ac.uk/open-access-journals-humanities-and-social-science.

Davis, P. M. (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. (2014) PeerJ Waives APCs and Pivots Again. *The Scholarly Kitchen* (blog), March 9, 2018. Retrieved from https://scholarlykitchen.sspnet.org/2018/03/09/peerj-waives-apc-pivots/.

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

Davis, P.M. (2013). “Public Accessibility of Biomedical Articles from PubMed Central Reduces Journal Readership—retrospective Cohort Analysis.” *The FASEB Journal 27 (7):* 2536–41. https://doi.org/10.1096/fj.13-229922

Davis, P. M. (2017). Portable peer review: RIP *Scholarly Kitchen.* Retrieved from https://scholarlykitchen.sspnet.org/2017/09/25/portable-peer-review-rip/

Davis, P. M. (2018). Not Every Publisher Can Support A Cascade Journal. *The Scholarly Kitchen.* Retrieved from https://scholarlykitchen.sspnet.org/2018/01/24/not-every-publisher-can-support-a-cascade-journal/.

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

Digital Science (2016) The State of Open Data 2016. Retrieved from: http://digital-science.com/resources/portfolio-reports/state-open-data/

Digital Science (2017) The State of Open Data 2017. Retrieved from: http://digital-science.com/resources/portfolio-reports/state-open-data-2017/

Donelan, Helen (2016) Social media for professional development and networking opportunities in academia: Journal of Further and Higher Education Volume 40, 2016 - Issue 5 retrieved from https://www.tandfonline.com/doi/full/10.1080/0309877X.2015.1014321

EBSCO (2018) Serials Price Projections – 2018. Retrieved from https://www.ebscohost.com/promoMaterials/EBSCO_2018_Serials_Price_Projections.pdf?_ga=2.4129336.946664490.1535281102-1743670085.1532728515

EC3metrics. (2018). "Periodic Table of Scientometric Indicators." *El profesional de la información* (blog). http://www.elprofesionaldelainformacion.com/notas/tabla-periodica-de-indicadores/.

Economist, The. (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

Economist, The. (2016). Why Research Papers Have so Many Authors. *The Economist*, Retrieved from https://www.economist.com/science-and-technology/2016/11/24/why-research-papers-have-so-many-authors.

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

Eger, T., and Scheufen, M. (2018). *The Economics of Open Access: On the Future of Academic Publishing*. Cheltenham: Edward Elgar 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. (2013). *International comparative performance of the UK research base -2013*. Retrieved from https://www.elsevier.com/__data/assets/pdf_file/0016/53026/Elsevier_BIS_2013_web_Dec2013-2.pdf

Elsevier. (2017a). *International comparative performance of the UK research base – 2016*. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/660855/uk-research-base-international-comparison-2016.pdf

Elsevier (2017b). *Gender in the Global Research Landscape Report*. Retrieved from https://www.elsevier.com/research-intelligence/resource-library/gender-report

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

European Commission Expert Group on Altmetrics (2017). Retrieved from https://ec.europa.eu/research/openscience/pdf/report.pdf

European Commission (2016). Proposal for a DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on copyright in the Digital Single Market, COM (2016) 593 final https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52016PC0593

European Commission (2017) Directorate-General for Research and Innovation *Europe's future.Open innovation, open science, open to the world : reflections of the Research, Innovation and Science Policy Experts (RISE) High Level Group.* Retrieved from https://publications.europa.eu/en/publication-detail/-/publication/527ea7ce-36fc-11e7-a08e-01aa75ed71a1.

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

Fanelli D, Larivière V (2016) Researchers' Individual Publication Rate Has Not Increased in a Century. PLoS ONE 11(3): e0149504. https://doi.org/10.1371/journal.pone.0149504

Fang, F. C., et al. (2012), Misconduct accounts for the majority of retracted scientific publications, PNAS, 109(42), doi:10.1073/pnas.1212247109.

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

Ferwerda, E, et al (2018) A landscape study on open access (OA) and monographs. *Knowledge Exchange* retrieved from http://repository.jisc.ac.uk/6693/5/Knowledge_Exchange_summary_-_landscape_study_on_OA_and_Monographs_revised_July_2018.pdf

Finch Working Group. (2012). *Accessibility, sustainability, excellence: How to expand access to research publication*. Retrieved from https://www.acu.ac.uk/research-information-network/finch-report-final

FORCE 11. (2011). *FORCE 11 manifesto: Improving future research communication and e-scholarship*.

Fortney, K., Gonder, J. (2015). A social networking site is not an open access repository. Office of Scholarly Communication. University of California.

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

Frontier Economics (2017) *The contribution of the publishing industry to the UK economy: A report for the Publishers Association.* Retrieved from https://www.publishers.org.uk/policy-research/the-value-of-uk-publishing/

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

Gadd, Elizabeth. (2017) *Guest Post - Academics and Copyright Ownership: Ignorant, Confused or Misled?* The Scholarly Kitchen, October 31, 2017. Retrieved from:

https://scholarlykitchen.sspnet.org/2017/10/31/guest-post-academics-copyright-ownership-ignorant-confused-misled/.

Gadd, Elizabeth, Jenny Fry, and Claire Creaser. "The Influence of Journal Publisher Characteristics on Open Access Policy Trends." *Scientometrics* 115, no. 3 (June 1, 2018): 1371–93. https://doi.org/10.1007/s11192-018-2716-8.

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

Gardner, T and S. Inger (2018) *How Readers Discover Content in Scholarly Communications.* Renew Publishing Consultants

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/

Global Research Council. (2016). "Statement of Principles and Actions Promoting the Equality and Status of Women in Research," Retrieved from https://www.globalresearchcouncil.org/fileadmin/documents/GRC_Publications/Statement_of_Principles_and_Actions_Promoting_the_Equality_and_Status_of_Women_in_Research.pdf
.

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

Goodman. Steven N, Daniele Fanelli and John P. A. Ioannidis *What does research reproducibility mean?* Science Translational Medicine 01 Jun 2016: Vol. 8, Issue 341, pp. 341 ps12)

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

Harnad, Stevan (2015) Open Access: What, Where, When, How and Why. In: Ethics, Science, Technology, and Engineering: An International Resource eds. J. Britt Holbrook & Carl Mitcham, (2nd of Encyclopedia of Science, Technology, and Ethics, Farmington Hills MI: MacMillan Reference). https://eprints.soton.ac.uk/361704/1/ESTEarticle-OA-Harnad.pdf

Hewlett, S.A., Marshall, M., and Sherbin, L. (2013) “How Diversity Can Drive Innovation.” Harvard Business Review. Retrieved from. https://hbr.org/2013/12/how-diversity-can-drive-innovation.

Himmelstein, Daniel S., Ariel Rodriguez Romero, Jacob G. Levernier, Thomas Anthony Munro, Stephen Reid McLaughlin, Bastian Greshake Tzovaras, and Casey S. Greene. (2018). “Research: Sci-Hub Provides Access to Nearly All Scholarly Literature.” Elife. Retrieved from https://doi.org/10.7554/eLife.32822.

Hook, D. W., Porter, S. J., and Herzog C. (2018). “Dimensions: Building Context for Search and Evaluation.” *Frontiers in Research Metrics and Analytics* 3. Retrieved from https://doi.org/10.3389/frma.2018.00023.

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/

Inchcoombe, S. (2017) The changing role of research publishing *Insights* **30**(2) 10-16. Doi: 10.1629/uksg.355

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

Ithaka S+R, Jisc, and RLUK (2016). *UK Survey of Academics 2015*. Retrieved from http://www.sr.ithaka.org/publications/uk-survey-of-academics-2015/

Ithaka S+R. (2017). *US Library Survey 2016*. Retrieved from http://www.sr.ithaka.org/wp-content/uploads/2017/03/SR_Report_Library_Survey_2016_04032017.pdf.

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

Jamali, H.R. (2017). Copyright compliance and infringement in ResearchGate full-text journal articles. *Scientometrics* 112(1):241-254.

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

Hepeng, J. (2018) “China’s Citations Catching Up.” *Nature Index* (blog). Retrieved from https://www.natureindex.com/news-blog/chinas-citations-catching-up.

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/

Johnson, R. & Fosci, M. (2015). On shifting sands: assessing the financial sustainability of UK learned societies. *Learned Publishing*, 28: 274–281. doi:10.1087/20150406.

Johnson, R., Fernholtz, O., Fosci, M. (2016) Text and data mining in Higher Education. Research Consulting. Retrieved from: https://adbu.fr/competplug/uploads/2016/12/TDM-in-Public-Research-Revised-15-Dec-16.pdf

Johnson, R., Fosci, M., Chiarelli, A., Pinfield, S. & Jubb, M. (2017). Towards a Competitive and Sustainable OA Market in Europe - A Study of the Open Access Market and Policy Environment, *OpenAire*.

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

Jubb, M. (2016). “Peer Review: The Current Landscape and Future Trends". *Learned Publishing*. Retrieved from https://onlinelibrary.wiley.com/doi/full/10.1002/leap.1008.

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

Kaiser, J. (2017). "Are Preprints the Future of Biology? A Survival Guide for Scientists." *Science | AAAS*. September 21, 2017. http://www.sciencemag.org/news/2017/09/are-preprints-future-biology-survival-guide-scientists

Keller, M. and Prusiner, S. (2018). "Citations Must Default to the Online Publication Date." News. Nature, June 27, 2018. https://doi.org/10.1038/d41586-018-05387-4.

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

Labaree, D. (2007). Understanding the Rise of American Higher Education: How Complexity Breeds Autonomy. http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=C47E9F487804F7A16F2CA7E195319554?doi=10.1.1.485.6425&rep=rep1&type=pdf

Larivière V, Haustein S, Mongeon P (2015) The Oligopoly of Academic Publishers in the Digital Era. PLoS ONE 10(6): e0127502. https://doi.org/10.1371/journal.pone.0127502

Lariviere, V., Kiermer, V., MacCallum, C.J., McNutt, M., Patterson, M., Pulverer, B., Swaminathan, S., Taylor, S. and Curry, S. (2016). "A Simple Proposal for the Publication of Journal Citation Distributions." *BioRxiv*, 062109. https://doi.org/10.1101/062109.

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

Lawson, S. (2015). Report on offset agreements: evaluating current Jisc Collections deals, Year 1 – evaluating 2015 deals. *Jisc*, October 2016.

Lawson, S., Gray, J., Mauri, M. (2016). "Opening the Black Box of Scholarly Communication Funding: A Public Data Infrastructure for Financial Flows in Academic Publishing." *Open Library of Humanities* 2, no. 1. https://doi.org/10.16995/olh.72.

van Leeuwen, T.N., Yegros-Yegros, A. and Costas, R. (2017). "Developing Indicators on Open Access by Combining Evidence from Diverse Data Sources." In *Proceedings of the 2017 STI Conference*. https://arxiv.org/ftp/arxiv/papers/1802/1802.02827.pdf.

Library Publishing Directory, 2018. https://librarypublishing.org/wp-content/uploads/2017/03/LPC_LPDirectory2018.pdf

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/

Li, X., Thelwall, M., & Kousha, K. (2015). The role of arXiv, RePEc, SSRN and PMC in formal scholarly communication. Aslib Journal of Information Management, 67(6), 614-635. https://doi.org/10.1108/AJIM-03-2015-0049

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

Mabe, M & Watkinson, A. (2019 - forthcoming).*Journals* in the History of the Book in Britain volume 7 *The Twentieth Century and Beyond* edited by Nash, Andrew et al Cambridge University Press

Madrigal, A. C. (2012). "Dark Social: We Have the Whole History of the Web Wrong." *The Atlantic*, 12 October. https://www.theatlantic.com/technology/archive/2012/10/dark-social-we-have-the-whole-history-of-the-web-wrong/263523/

Mandler, P. (2014). Open access: A perspective from the humanities. *Insights: The UKSG Journal*, *27*(2), 166-170. doi:10.1629/2048-7754.89

Marchitelli, A., Galimberti, P., Bollini, A., & Mitchell, D. (2017). Helping journals to improve their publishing standards: a data analysis of DOAJ new criteria effects. *JLIS.it, 8*(1), 1-21.

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/

Martín-Martín, A., Costas, R., van Leeuwen, T., & López-Cózar, E. D. (2018a). Unbundling Open Access dimensions: a conceptual discussion to reduce terminology inconsistencies. *[preprint]* https://doi.org/10.31235/osf.io/7b4aj

Martín-Martín, A., Costas, R., van Leeuwen, T., & López-Cózar, E. D. (2018b) "Evidence of Open Access of Scientific Publications in Google Scholar: A Large-Scale Analysis." *Journal of Informetrics* 12, no. 3: 819–41. https://doi.org/10.1016/j.joi.2018.06.012.

Martín-Martín, A., Orduna-Malea, E. & Delgado López-Cózar, E. (2018c). Coverage of highly-cited documents in Google Scholar, Web of Science, and Scopus: a multidisciplinary comparison. *Scientometrics*. https://doi.org/10.1007/s11192-018-2820-9

Martín-Martín, A., Orduna-Malea, E., Thelwall, M. and Delgado López-Cózar, E. (2018d) "Google Scholar, Web of Science, and Scopus: A Systematic Comparison of Citations in 252 Subject Categories." *SocArXiv*, August 15, 2018. https://doi.org/10.31235/osf.io/42nkm.

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

McCook, Alison. (2018) "What Happened When Elsevier Tried Open Peer Review? And Which Field Says 'No, Thanks?'" *Retraction Watch* (blog). Retrieved from https://retractionwatch.com/2018/07/03/what-happened-when-elsevier-tried-open-peer-review-and-which-field-says-no-thanks/.

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

McKiernan, E. C., Bourne, P. E., Brown, C.T., Buck, S., Kenall, A., Lin, J., McDougall, D., et al. (2016). "How Open Science Helps Researchers Succeed." *ELife*, 5 July. https://doi.org/10.7554/eLife.16800

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. (2018) A Woman's Worth: Examining the Gender Pay Gap among UK Scholarly Publishers. *Scholarly Kitchen Blog*. Retrieved from

https://scholarlykitchen.sspnet.org/2018/04/17/women-like-men-cheaper-examining-gender-pay-gap-among-uk-scholarly-publishers/

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

Medin, D. and Lee, C. (2012). "Diversity Makes Better Science." *APS Observer* 25, no. 5 Retreived from https://www.psychologicalscience.org/observer/diversity-makes-better-science.

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

Mikki, S. (2017). Scholarly publications beyond pay-walls: increased citation advantage for open publishing. Scientometrics, 113(3), 1529-1538. https://doi.org/10.1007/s11192-017-2554-0

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

Mithu, L., Draux, H., and Walker, J. (2018). "Assessing the Open Access Effect in Hybrid Journals," June 29, 2018. https://doi.org/10.6084/m9.figshare.6396290

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

Montgomery, L. and Ren, Xiang (2018) Understanding open knowledge in China Cultural Science Journal 10(1). 17-26. Retrieved from : http://doix.org/10.53334/CSCI.106

Mooney, H. (2016) Scholarly communication and data. In: Kellem, L. M., Thompson (eds) Data Librarianship, pp195-218, ACRL

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

Munfao, M et al (2017)*A manifesto for reproducible science.* Nature Human Behaviour volume 1, Article number: 0021

Murphy, F. (2014). Data and scholarly publishing: The transforming landscape. *Learned Publishing*, *27*(5), 3-7. doi:10.1087/20140502

Murphy, F. et al. (2017) Key components of data publishing *I. J. Digital Libraries* 18(2).77-92

National Academies of Sciences, Engineering, and Medicine. 2018. *Open Science by Design: Realizing a Vision for 21st Century Research*. Washington, DC: The National Academies Press. https://doi.org/10.17226/25116.]

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

Naydenov, A and H.Staines (2018) *The Time for Open and Interoperable Annotation is Now* The Scholarly Kitchen Blog retyrieved from https://scholarlykitchen.sspnet.org/2018/08/28/all-about-open-annotation

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

Neylon, Cameron. (2015)"The Marginal Costs of Article Publishing – Critiquing the Standard Analytics Study." Science in the Open (blog). Retrieved from //cameronneylon.net/blog/the-marginal-costs-of-article-publishing-critiquing-the-standard-analytics-study/.

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

National Science Board. (2012). *Science & Engineering Indicators 2012*. Retrieved from http://www.nsf.gov/statistics/seind12/

National Science Board. (2014). *Science and Engineering Indicators 2014*. Arlington, VA: NSB 14-01. Retrieved from http://www.nsf.gov/statistics/seind14/

National Science Board (2018). *Science and Engineering Indicators 2018*. NSB-2018-1. Alexandria, VA: National Science Foundation. Available at https://www.nsf.gov/statistics/indicators/

Normile, D. (2017). "China Cracks down after Investigation Finds Massive Peer-Review Fraud." *Science* | *AAAS*. Retrieved from http://www.sciencemag.org/news/2017/07/china-cracks-down-after-investigation-finds-massive-peer-review-fraud.

Nosek, B.A. et al (2015) Promoting an open research culture Science 26 Jun 2015: Vol. 348, Issue 6242, pp. 1422-1425

Odell, J., Palmer, K., & Dill, E. (2017). Faculty Attitudes toward Open Access and Scholarly Communications: Disciplinary Differences on an Urban and Health Science Campus. *Journal of Librarianship and Scholarly Communication*, 5(General Issue), eP2169. http://dx.doi.org/10.7710/2162-3309.2169

OECD. (n.d.). *OECD.StatExtracts: Main Science and Technology Indicators.* Retrieved from 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

OECD (2018a). *Main Science and Technology Indicators Database, March 2018*. Retrieved from www.oecd.org/sti/msti

OECD (2018b). *Release of Main Science and Technology Indicators - Latest estimates of R&D investment in OECD and major economies*. Retrieved from http://www.oecd.org/sti/DataBrief_MSTI_2018.pdf

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. (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

Outsell (2017) *STM Platform Providers* Outsell Company Analyses 2017

Outsell. (2018a). *Science, Technology and Healthcare: 2018 - Market Size, Share, Forecast and Trend Report*. Retrieved from https://www.outsellinc.com/product/science-technology-and-healthcare-2018/

Outsell. (2018b). *Five Steps Not to Take as you Embrace the Cloud* Outsell

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

Patterson, M. and Schekman, R. (2018). “Scientific Publishing: A New Twist on Peer Review.” ELife. https://doi.org/10.7554/eLife.36545.

Patterson, M., and McLennan, J. (2016) “Setting a Fee for Publication.” Inside ELife (blog), Retrieved from https://elifesciences.org/inside-elife/b6365b76/setting-a-fee-for-publication

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

Pinhasi, R., Blechl, G., Kromp, B., & Schubert, B. (2018). The weakest link – workflows in open access agreements: the experience of the Vienna University Library and recommendations for future negotiations. Insights, 31, 27. DOI:http://doi.org/10.1629/uksg.419

Piwowar, H., Priem, J., Larivière, V., Alperin, J.P., Matthias, L., Norlander, B., Farley, A., West, J. & Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6:e4375. https://doi.org/10.7717/peerj.4375

PLOS. (2016). *PLOS progress update 2016*. Retrieved from https://www.plos.org/annual-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/

Polka, J. K., Kiley, R., Konforti, B., Stern, B. and Vale, R. D. (2018) "Publish Peer Reviews." *Nature* 560, no. 7720: 545. https://doi.org/10.1038/d41586-018-06032-w.

Pollock, D., and Michael, A. (2018a). "News & Views: Potential Impact of Open Access Uptake Scenarios." *Delta Think* (blog). Retrieved from https://deltathink.com/news-views-potential-impact-of-open-access-uptake-scenarios/.

Pollock, D., and Michael, A. (2018b). "News & Views: Evaluating Quality in Open Access Journals." *Delta Think* (blog). Retrieved from https://deltathink.com/news-views-evaluating-quality-in-open-access-journals/.

Poynder, Richard. (2018). "Open and Shut?: Q&A with CNI's Clifford Lynch: Time to Re-Think the Institutional Repository?" Open and Shut? (blog). Retrieved from https://poynder.blogspot.com/2016/09/q-with-cnis-clifford-lynch-time-to-re_22.html.

PRC (2016) PRC Peer Review Survey 2015 (Mark Ware Consulting). Retrieved from http://publishingresearchconsortium.com/index.php/134-news-main-menu/prc-peer-review-survey-2015-key-findings/172-peer-review-survey-2015-key-findings

PRC (2016b) Early Career Researchers: The Harbingers of Change Year One (Ciber). Retrieved from http://publishingresearchconsortium.com/index.php/137-news-main-menu/ecrs-year-one-key-findings/175-early-career-researchers-year-one-2016-key-findings

PRC (2017) Early Career Researchers: The Harbingers of Change Year Two (CIBER). Retrieved from http://publishingresearchconsortium.com/index.php/prc-documents/prc-research-projects/71-early-career-researchers-the-harbingers-of-change-report-on-year-two-2017/file

PRC (2018a) Economics of Academic Publishing (CEPA)

PRC (2018b) Hybrid OA and Author Choice (Research Consulting and Pleiade Management Consulting) Retrieved from http://publishingresearchconsortium.com/index.php/prc-documents/prc-research-projects/78-hybrid-oa-and-author-choice-executive-summary

PRC (2018c) Early Career Researchers : The Harbingers of Change Year Two (CIBER) forthcoming

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

Priest, E. (2012) "Copyright and the Harvard Open Access Mandate." *Northwestern Journal of Technology and Intellectual Property* 10, no. 7: 377. Retrieved from: https://scholarlycommons.law.northwestern.edu/njtip/vol10/iss7/1/

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

Proquest. (2016). *Academic Library Book Purchasing Trends.* Retrieved from http://contentz.mkt5049.com/lp/43888/438659/D187_Ebooks_Aquisition_whitepaper_v5.pdf

The Publishers Association. (2016). "PA Market Report: China Journals." Retrieved from http://www.pabookshop.co.uk/index.php?option=com_jw_bookshop&task=showPublicationDetail&call_task=showPromoted&id=215&limit=5&limitstart=0&Itemid=1410

Publishers Communication Group. (2017) "Library Budget Predictions for 2017," Retrieved from http://www.pcgplus.com/wp-content/uploads/2017/05/Library-Budget-Predictions-for-2017-public.pdf

Publons. (2018). *Global State of Peer Review*. Retrieved from https://publons.com/community/gspr

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/

Pullinger, D., and Baldwin, C. (2002) Electronic Journals and User Behaviour: Learning for the Future from the Superjournal Project deedot press, Cambridge

Quan, W, Chen, B., and Shu, F. (2017) "Publish or Impoverish: An Investigation of the Monetary Reward System of Science in China (1999-2016)." *Aslib Journal of Information Management* 69, no. 5: 486–502. https://doi.org/10.1108/AJIM-01-2017-0014.

Rapple, Charlie (2017) MECA – A New Manuscript Exchange Initiative The Scholarly Kitchen retrieved from https://scholarlykitchen.sspnet.org/2017/08/17/meca-new-manuscript-exchange-initiative/

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

Renew Publishing Consultants (2018), *A Landscape Review of Streaming Media in Scholarly Publishing*, 2018. Retrieved from http://renewpublishingconsultants.com/wp-content/uploads/2018/08/Streaming-Landscape-2018-Renew-Publishing-Consultants-FINAL.pdf

Research England. (2018) *Monitoring Sector Progress towards Compliance with Funder Open Access Policies*. Retrieved from https://re.ukri.org/documents/2018/research-england-open-access-report-pdf/.

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 (2011e) (Cann A) *Social Media A guide for Researchers:* retrieved from http://www.rin.ac.uk/system/files/attachments/social_media_guide_for_screen_0.pdf],

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

Ross-Hellauer T, Schmidt B, Kramer B. (2018) Are funder Open Access platforms a good idea? PeerJ Preprints 6:e26954v1 [not peer-reviewed] https://doi.org/10.7287/peerj.preprints.26954v1

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

Rowlands, I., Nicholas, D., Williams, P., Huntington, P., Fieldhouse, M., Gunter, B., et al. (2008). "The Google Generation: The Information Behaviour of the Researcher of the Future." *Aslib Proceedings*, 60 (4): 290–310. https://doi.org/10.1108/00012530810887953

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.

Scholarly Kitchen (2018) On Being Excluded: Testimonies by People of Colour in Scholarly Publishing. *Scholarly Kitchen Blog*. Retrieved from https://scholarlykitchen.sspnet.org/2018/04/04/excluded-testimonies-people-color-scholarly-publishing/

Schonfeld, R.C. (2017) Diversifying the Intellectual Leadership of Scholarly Publishing. *Scholarly Kitchen Blog.* Retrieved from https://scholarlykitchen.sspnet.org/2017/06/29/diversifying-intellectual-leadership/

Schonfeld, Roger C (2017a) "Strategy & Integration Among Workflow Providers" *The Scholarly Kitchen Blog* retrieved from https://scholarlykitchen.sspnet.org/2017/11/07/strategy-integration-workflow-providers

Schonfeld, Roger C (2017b) “Workflow Strategy for Those Left Behind” *The Scholarly Kitchen Blog* retrieved from https://scholarlykitchen.sspnet.org/2017/12/18/workflow-strategy-left-behind-context/.

Schonfeld, Roger C (2017c) “Workflow Strategy for Those Left Behind” The Scholarly Kitchen Blog retrieved from https://scholarlykitchen.sspnet.org/2017/12/19/workflow-strategy-left-behind-options/

Schonfeld, Roger C. (2018) “Will the European Big Deal Contagion Spread to North America?” *The Scholarly Kitchen Blog*. Retrieved from https://scholarlykitchen.sspnet.org/2018/06/18/will-european-contagion-spread/.

Science-Metrix. (2018) Science-Metrix, Analytical Support for Bibliometrics Indicators Open access availability of scientific publications. http://www.science-metrix.com/sites/default/files/science-metrix/publications/science-metrix_open_access_availability_scientific_publications_report.pdf

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

Serghiou S, Ioannidis JPA (2018). Altmetric Scores, Citations, and Publication of Studies Posted as Preprints. *JAMA.*2018;319(4):402–404. https://doi.org/10.1001/jama.2017.21168

Shamash, K. (2016). Article processing charges (APCs) and subscriptions: monitoring open access costs. Retrieved from https://www.jisc.ac.uk/reports/apcs-and-subscriptions.

Sharma, Yojana. (2018). “Science Ministry Expands Power over Research Funding - University World News.” *University World News*. Retrieved from http://www.universityworldnews.com/article.php?story=2018032117551180.

Shen, C., and Björk, B. C. (2015). “‘Predatory’ Open Access: A Longitudinal Study of Article Volumes and Market Characteristics.” BMC Medicine 13, no. 1. https://doi.org/10.1186/s12916-015-0469-2.

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

Silvello, G. (2018) Theory and practice of data citation *J Amer Soc Info Sci Tech* 68(1).6-20

Siler, K., et al. (2015), Measuring the effectiveness of scientific gatekeeping, PNAS., 112, 2, doi:10.1073/pnas.1418218112.

Siler, K., Haustein, S., Smith, E., Larivière, V., Alperin, J.P. (2018). Authorial and institutional stratification in open access publishing: the case of global health research. *PeerJ* 6:e4269 https://doi.org/10.7717/peerj.4269

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

Simba. (2017a). *Global Scientific & Technical Publishing 2017-2021*. Retrieved from https://www.simbainformation.com/Global-Scientific-Technical-10845051/

Simba (2017b). *Global Medical Publishing 2017-2021*, Retrieved from https://www.simbainformation.com/Global-Medical-Publishing-11208060/

Singeh, F. W., Abrizah, A., Karim, N. H. A. (2013) What inhibits authors to self-archive in Open Access repositories? A Malaysian case, *Information Development* Vol 29 (1)/

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

Solomon, D. J., Laakso, M., Björk, B. (authors) & Suber, P. (editor). (2016). "Converting Scholarly Journals to Open Access: A Review of Approaches and Experiences".

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

Spezi, V., Wakeling, S., Pinfield, S., Creaser, C., Fry, J. & Willet, P. (2016). Open-access mega-journals: The future of scholarly communication or academic dumping ground? A review https://www.emeraldinsight.com/doi/full/10.1108/JD-06-2016-0082

Springer Nature (2017) The OA Effect: How Does Open Access Affect the Usage of Scholarly Books? *Springer Nature white paper* Retrieved from https://media.springernature.com/full/springer-cms/rest/v1/content/15176744/data/v3

SpringerNature. (2018). Prospectus dated April 25, 2018

STM (2010) *STM Guidelines: access for people with print disabilities*. Retrieved from https://www.stm-assoc.org/2010_09_12_STM_Guidelines_access_for_people_with_print_disabilities.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. (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/

STM. (2016). Voluntary principles for article sharing on scholarly collaboration networks retrieved from https://www.stm-assoc.org/2015_06_08_Voluntary_principles_for_article_sharing_on_scholarly_collaboration_networks.pdf

STM position on Open Science (2016) retrieved from https://www.stm-assoc.org/2016_10_31_STM_Open_Science_Position.pdf

STM. (2017a). 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 https://www.stm-assoc.org/2017_05_10_Text_and_Data_Mining_Declaration.pdf

STM. (2017b). *Text and data mining: Building a health and sustainable knowledge ecosystem for Europe*. Retrieved from https://www.stm-assoc.org/2017_12_20_2017_12_STM_Text_and_Data_Mining_Summary.pdf

Strieb, K., & Blixrud, J. (2013). *The state of large-publisher bundles in 2012*. ARL. Retrieved from http://publications.arl.org/rli282/13

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M. and Astell, M. (2018). "Whitepaper: Practical Challenges for Researchers in Data Sharing," Springer Nature. https://doi.org/10.6084/m9.figshare.5975011.v1.

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

Suber, P. (2008). Gratis and libre open access. *SPARC Open Access Newsletter*, 124.

Suber, P. (2012). *Open access*. Cambridge, Mass: MIT Press.

Sugimoto CR, Work S, Larivière V, Haustein S (2017). Scholarly use of social media and altmetrics: A review of the literature. Journal of the Association for Information Science and Technology 2017; 68(9): 2037-62.

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/

Swan, A. (2016). The costs and benefits to the research community of Open Access: A briefing paper. http://pasteur4oa.eu/sites/pasteur4oa/files/resource/Costs%20of%20OA%20final_0.pdf

Taylor & Francis. (2014). *Open access survey June 2014*. Retrieved from http://www.tandfonline.com/page/openaccess/opensurvey/2014

Tennant, J., Bauin, S., James, S., & Kant, J. (2018, May 17). The evolving preprint landscape: Introductory report for the Knowledge Exchange working group on preprints. http://doi.org/10.17605/OSF.IO/796TU

Tennant J.P., Dugan J.M., Graziotin D. *et al.* (2017) A multi-disciplinary perspective on emergent and future innovations in peer review [version 3; referees: 2 approved]. F1000Research 2017, 6:1151. https://doi.org/10.12688/f1000research.12037.3

Tennant, J. P., Waldner, F., Jacques, D.C., Masuzzo, P., Collister, L.B., and Hartgerink, C.H.J. (2016). "The Academic, Economic and Societal Impacts of Open Access: An Evidence-Based Review." *F1000Research*, 5 April: 632. https://doi.org/10.12688/f1000research.8460.1

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., Allard, S., Douglass, K., Ayinoglu, A. U., Wu, L., Read, E. (2011b) Data sharing by scientists. *PLoS ONE* **6**(6):e21101

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

Tenopir, C., King, D. W., Christian, L., Voluntine, R. (2012b) Scholarly article seeking, reading and use *Learned Publishing* 28: 93-105

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. (2014a) "The Life-Sciences as a Pathfinder in Data-Intensive Research Practice." Retrieved from https://www.slideshare.net/atreloar/the-lifesciences-as-a-pathfinder-in-dataintensive-research-practice?from_action=save.

Treloar, A. (2014b). 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/

UNESCO. (2015). UNESCO Science Report: Towards 2030. Retrieved from http://unesdoc.unesco.org/images/0023/002354/235407e.pdf

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

University of California Libraries. (2016). "Pay It Forward: Investigating a Sustainable Model of Open Access Article Processing Charges for Large North American Research Institutions," Retrieved from http://icis.ucdavis.edu/wp-content/uploads/2016/07/UC-Pay-It-Forward-Final-Report.rev_.7.18.16.pdf.

University of California Libraries. (2018). "Pathways to Open Access," Retrieved from https://libraries.universityofcalifornia.edu/about/initiatives/scholarly-communication.

Universities UK (2017). Monitoring the Transition to Open Access. *Universities UK*, December.

Unsworth, J. (2017). eBooks – The Next Big Deal? *eBooks and the Academic Library Market: The Charlotte Initiative Open Conference*. Retrieved from http://people.virginia.edu/~jmu2m/charlotte.ebook.talk.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/

van Rossum, J (2017) *Perspectives on a New Paradigm for Scholarly Communication* Digital ScienceVerstak, 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

Wakeling, S., Spezi, V., Fry, J., Creaser, C., Pinfield, S. & Willett, P. (2017a) Open access megajournals: The publisher perspective (Part 1: Motivations), *ALSP*, doi: 10.1002/leap.1117.

Wakeling, S., Spezi, V., Fry, J., Creaser, C., Pinfield, S. & Willett, P. (2017b) Open access megajournals: The publisher perspective (Part 2: Operational realities), *ASLP*, doi: 10.1002/leap.1118.

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

Wang, X., Liu, C., Mao, W. & Fang, Z. (2015). "The Open Access Advantage Considering Citation, Article Usage and Social Media Attention." *Scientometrics* 103 (2): 555–64. https://doi.org/10.1007/s11192-015-1547-0

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

Watkinson, A. (2003). Securing authenticity of scholarly paternity and integrity. Retrieved from: http://ciber-reserach.eu/download/Watkinson-Securing_Authenticity-2003.pdf

Watkinson, A (2016) The Academic Book in North America: Report on attitudes and initiatives among publishers, libraries, and scholars 2016. *Academic Book of the Future* Retrieved from https://academicbookfuture.org/academic-book-north-america-watkinson/

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.

Winter, J. C. F. de. (2015) "The Relationship between Tweets, Citations, and Article Views for PLOS ONE Articles." *Scientometrics* 102, no. 2: 1773–79. https://doi.org/10.1007/s11192-014-1445-x.

Wittenberg, K., Glasser, S., Kirchhoff, A., Morrissey, S., & Orphan, S. (2018). Challenges and opportunities in the evolving digital preservation landscape: reflections from Portico. *Insights*, *31*, 28. DOI: http://doi.org/10.1629/uksg.421

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

Wulf, K., and Newman, S. (2017) "Missing the Target: The UK Scholarly Communications License." *The Scholarly Kitchen Blog.* Retrieved from https://scholarlykitchen.sspnet.org/2017/07/26/missing-target-uk-scholarly-communications-license/

Xu, J., Wang, J. Y., Zhou, L. and Liu, F. (2018) "Internationalization of China's English-Language Academic Journals: An Overview and Three Approaches." *Learned Publishing* 0, no. 0. https://doi.org/10.1002/leap.1198.

Xuemei, L., Thelwall, M. and Kayvan Kousha. (2015) "The Role of ArXiv, RePEc, SSRN and PMC in Formal Scholarly Communication." *Aslib Journal of Information Management* 67, no. 6: 614–35. https://doi.org/10.1108/AJIM-03-2015-0049.

Zhang, X. (2016). Open Access in China, COAR Annua, Meeting & General Assembly 2016. https://www.coar-repositories.org/files/160413-Open-Access-in-China-new.pdf

Zuckerman, H and Merton, R. K. (1971) Patterns of Evaluation in Science. *Minerva* 9, no. 1: 66-100

In honour of Karen Hunter

1945 – 2018

Printing and distribution
sponsored by Elsevier

ELSEVIER