

**International
Science Council**

The global voice for science

Business Models and Market Structure within the Scholarly Communications Sector

**ISC Occasional Paper
September 2020**



Business Models and Market Structure within the Scholarly Communications Sector

International Science Council Occasional Paper: 07 September 2020

Rupert Gatti¹

This paper was commissioned by the International Science Council (ISC) as a contribution to its project on the future of scientific publishing (<https://council.science/actionplan/future-of-scientific-publishing/>) and to its discussion paper *“Opening the record of science: making scholarly publishing work for science in the digital era”*, which is currently being debated by the ISC membership organisations as a basis for a campaign for reform.

The scholarly communications sector is undergoing a period of profound transformation. The emergence of digital publishing technologies, wide-spread demands for open access to research outputs, calls for more rapid dissemination of research findings and underlying data - especially in emergencies such as global pandemics, and the expansion of publishers operations into all aspects of the research cycle are requiring and inspiring innovative new process and practices. At the same time there is growing awareness and concern in scientific and scholarly communities about the business models and operating practices being adopted by publishers, particularly commercial publishers.

The International Science Council Discussion Paper *“Opening the Record of Science”* (Boulton et al. (2020)) highlights many of these concerns and identifies some fundamental principles for scholarship and scholarly communications to frame developments in the sector. This paper is designed to complement that paper, and similar discussions, by assessing the economic implications of current business models for scientific publishing, evaluating their advantages and disadvantages in relation

¹ Trinity College, Cambridge, CB2 1TQ, UK. email: irig1@cam.ac.uk. ORCID: orcid.org/0000-0003-0676-9024.

The author thanks Claudio Aspesi, Lucy Barnes, Geoffrey Boulton and Alessandra Tosi for very helpful feedback on previous versions of this paper. Of course any remaining errors and all opinions are the author's.

to the fundamental principles advocated, and proposing a range of models that would be compatible with those principles.²

The paper proceeds by first positioning the situation within the broader setting of how to effectively regulate digital markets. The dominant business model and industrial structure within scholarly communications at the end of the last century is then discussed, as a springboard from which to consider new business models that have arisen over the past twenty years and their likely implications for the sector. The paper concludes that there would be considerable benefit to the establishment of a permanent digital markets unit to monitor and assess ongoing developments in the scholarly communications sector and to coordinate and encourage “good behaviour” across all actors in the sector.

1. Regulating Digital Markets

Concerns about the impact of domination of an industry by one or a very few digital platforms is not unique to scholarly communications; antitrust regulators around the world are grappling with how to maintain control over digital industries to prevent monopolization and to ensure socially beneficial innovation and outcomes.³ There is widespread agreement amongst regulators and commentators that the traditional antitrust tools for assessing (and, where necessary, mitigating) the economic impact of actions by dominant firms are ill-suited to digital industries. Many goods in digital markets are free and highly valued by users (e.g. Google Search) but still may have such damaging consequences on an industry as a whole as to outweigh those benefits. Assessing the long-term consequences of an acquisition is extremely difficult in digital markets, and the length of time needed to impose legislative sanctions means that by the time an action is found to have been illegal or anti-competitive, the whole industry has moved on and it is too late to prevent the damaging repercussions.

² There have been numerous previous analyses of the economic impact of digital technology and open access business models on the scholarly communications sector, including Houghton & Vickery (2005), and Houghton & Oppenheim (2010). The main distinction between this and previous works is the explicit intention to position the analysis within the more recent dialogue surrounding competition and regulation of digital markets generally.

³ As I write this there is a US Congressional Committee assessing the actions of four of the largest digital companies: Alphabet (Google’s parent), Amazon, Apple, and Facebook. Numerous reports have been commissioned on this issue by national regulators: <https://research.chicagobooth.edu/stigler/events/single-events/antitrust-competition-conference/world-reports-digital-markets>

A recent report by the Digital Competition Expert Panel for the UK Treasury, headed by Jason Furman, identified five important characteristics of digital markets:

“1 The digital economy is creating substantial benefits.

2 In many cases, digital markets are subject to ‘tipping’ in which a winner will take most of the market.

3 Concentration in digital markets can have benefits but also can give rise to substantial costs.

4 Competition for the market cannot be counted on, by itself, to solve the problems associated with market tipping and ‘winner-takes-most’.

5 Government policy and regulation also has limitations.”⁴

The report recommends that regulators

“should take a forward-looking approach that creates and enforces a clear set of rules to limit anti-competitive actions by the most significant digital platforms while also reducing structural barriers that currently hinder effective competition. These rules should be based on generally agreed principles and developed into more specific codes of conduct with the participation of a wide range of stakeholders. Active efforts should also make it easier for consumers to move their data across digital services, to build systems around open standards, and to make data available for competitors, offering benefits to consumers and also facilitating the entry of new businesses. Implemented effectively, this approach would be more flexible, predictable and timely than the current system.”⁵

The international scholarly community should take these recommendations seriously for their own sector and take a leadership role in establishing such standards and norms internationally. Relying on national antitrust authorities to mitigate the potential damaging activities undertaken by large international publishing entities is not likely to be a realistic alternative.

⁴ Furman J. et al. (2019), pp. 3-5.

⁵ Furman J. et al. (2019), p.2.

2.The twentieth-century scholarly publishing model: Reader-pays

At the most general level, publishers have four potential sources of income to support their operations: readers, authors, third-party institutions, and 'cross-subsidisation' through the sales of related services or products. This basic taxonomy of business models is used as the basis for the analysis and discussion that follows.

The dominant publishing business model throughout the twentieth century was "reader-pays". The publisher receives payment from readers - primarily via library subscriptions. While clearly this model served the industry well in many ways for many years, the primary disadvantage of this revenue model is that access to publications is only available to those who can afford the access fees. Very few individuals or institutions could afford to obtain access to all the works they or their researchers may have desire to read - so decisions about how best to ration access to publications is a requirement of this model. Those with the fewest resources can access very little research output, while members of the wealthiest institutions have the greatest access to scholarly works. This process increases and reinforces existing inequalities of opportunity across the scholarly sector as a whole.

Another feature of the reader-pays model is that the readers do not know the usefulness or value of the publication prior to paying for it. Such markets, where consumers purchase products with imperfect information about quality, are problematic in many areas (e.g. health services, second-hand cars etc.) and typically respond by providing third-party quality certificates or by firms establishing quality reputations to assist/reassure customers. In the scholarly sector we have both; a strong reputation and brand are commercially important for publishers, who are thus motivated to implement rigorous pre-publication peer-review processes. Making clear that the quality assessment is undertaken by scholars outside the publishing entity is an important part of the branding message. Two important characteristics develop: readers rely on the publishers' brand and reputation in deciding which content to purchase, and the publishers take responsibility for administering the quality checks prior to publishing. The fact that publishers utilise scholars in their assessment process means there is only a limited role for third-party 'quality assessors' that act independently of the publishing house. While branding holds advantages for resolving imperfect information about product quality, we will see that there are also negative implications in some settings.

The third feature of the reader-pays model is that the author doesn't pay.⁶ Because of this, institutions can be relatively relaxed about allowing the author to select the publisher for their work, and the author's primary considerations in making that decision need not be financial. This author-centric nature of the decision-making process reflected and reinforced a behavioural norm: that authors maintain independent control over the choice of publishing outlets for their research findings.

Very often it was the 'prestige' of the publisher that became the overriding consideration for authors. Thus the publishers with the most prestigious 'brand' attracted the highest-quality submissions - reinforcing their brand. In this setting, passing the peer-review process of a journal provided a signal that the research had exceeded a minimum quality assurance. Given that the minimum standards required by a journal were themselves correlated with the journals 'prestige', the system of publisher brands became a hierarchy of quality thresholds, which was then used as a proxy for quality rankings by tenure and funding assessment boards. The use of journal branding as a metric for assessing quality by tenure committees and grant-funding agencies further reinforced authors' incentives to care about the perceived prestige of the publisher when making their decisions.

A final characteristic of the twentieth-century publishing sector was the emergence of digital technologies in the latter decades which allowed significant cost savings for the largest publishers. Large investments in digital technology significantly reduced the per-unit production and distribution costs for publishers of sufficient scale. Smaller publishers, less able to afford large investments in digital technologies, found it increasingly difficult to match these cost reductions and many were either absorbed by, or outsourced their production and distribution processes to, the large commercial publishers who were able to afford such investment. It is not particularly surprising that it was commercial rather than non-profit publishers that took the lead in adopting these new technologies - typically, commercial entities have greater access to investment-financing opportunities through traditional financial markets than non-profits. Of course, the ability of commercial enterprises to access capital markets and make investments of this type can be seen as a good thing, as it allows for production efficiency gains throughout the industry. But the concentration of publishing activities towards commercial operators and away from mission-led non-profit publishing enterprises is an important characteristic.

⁶ In fact there typically were some author side payments as well - for the inclusion of plates and figures etc. - but these were relatively small in comparison and so had little impact on the business model.

To summarise, key features of the late twentieth-century scholarly publishing sector were:

1. Primarily a reader-pays business model.
2. Inequality amongst readers in access to research findings.
3. Scholarly quality assessment arranged by publishers through peer-review process rather than by third parties.
4. A behavioural norm that authors control decisions over publishing outlets.
5. Strong importance of publisher brand and reputation - impacting on the decisions made by readers, tenure committees, funding bodies, and authors.
6. Significant cost improvements could be achieved by large publishers through investments in digital production technologies.
7. Commercial entities were better able to take advantage of these new opportunities and adopt digital workflows, due to easier access to capital markets.

3. Publishing in the twenty-first century: Publishers as platforms operating in multi-sided markets

Two hugely important contemporary developments in scholarly communications, that are likely to determine its long-term future development, are:

a. increased calls for open access to research outputs, not only research findings but also for the underlying databases and research processes; and

b. recognition by the largest digital publishers that their business models will revolve around the provision of digital platforms with multi-sided markets, rather than 'traditional' publishing operations.

Digital technologies and the emergence of the world wide web has brought down the cost of distributing research outputs to an additional reader to something very close to zero. In this situation it becomes very difficult to justify rationing access to scholarly outputs (as required in the reader-pays model) and the inequalities associated with that process - particularly when the research itself has been funded by public expenditure. 'Artificially' denying access to the findings of research that

the public has paid for is difficult to justify, particularly if there are alternative ways of paying for the 'first copy' costs that might be feasibly introduced.⁷

But what are the alternative business models for publishers to obtain the revenue required to recoup operating costs and, for commercial publishers, to maintain the profit margins demanded by investors if the price charged to readers is set to zero? As we have seen earlier, there are three remaining primary sources: authors, institutions, or other products.

3.1 Author-pays model

Probably the most obvious solution is for publishers to shift from a reader-pays model to an author-pays model - and this has become the dominant business model adopted by commercial publishers for open access publications.

One immediate implication of transitioning to an author-pays model is that it is now the authors, as opposed to the readers, who face restrictions and inequality in access to publishing outlets. Researchers operating within less financially endowed institutions or scholarly systems are now less able to place their papers in the most expensive outlets - which are also likely to be the outlets with the highest quality reputation. This may also dramatically change the relative treatment of researchers within institutions: it may no longer be the case that all researchers in a single institution have the same access to both publications and publishers as their peers; institutions or research systems with budget constraints may now need to decide who to allow to publish their findings in expensive (prestigious) outlets. Of course this impacts further on promotion and funding opportunities for the less financially advantaged researchers - both between, but now also within, research institutions and structures.⁸

Branding remains critically important for publishers in this model, as it is the strength of their brand that allows publishers to charge higher fees to authors⁹ - but the mechanisms for sustaining those

⁷ The International Science Council discussion paper identifies as a Key Principle: "There should be universal open access to the record of science, both for producers and consumers, with no unnecessary barriers to participation, in particular those based on ability to pay, institutional privilege, language or geography." Boulton G. et al. (2020), p2. Of course open access does not only mean free to access; it also involves freedoms to download and reuse the publication as well. But I concentrate here on the free-to-access component of the term.

⁸ Houghton, J. W. & Vickery, G. (2005), Burchardt, J. (2014).

⁹ Nina Schönfelder (2020) shows that author charges for open access publication increase with the journal's citation measure (SNIP), with greater responsiveness observed in pure open access journals than in hybrid journals.

brands are now different. For publishers, brand now matters primarily because of the impact it has on author submission decisions. Why should authors care about a journal's brand? Two important structural features persist from the earlier model:

- a. the publisher's peer review process provides a quality threshold 'stamp'; and
- b. appointments/funding processes continue (or are believed by authors to continue) to use the perceived hierarchy of 'stamps' as a proxy for research quality when making their own assessments about an individual researcher.¹⁰

The persistent association between journal brand and perceived research quality sustains high profit margins in the industry, and ties authors into following publishing norms they often dislike. Weakening this association by creating alternative mechanisms for assessing research quality will reduce the ability of publishers to systematically maintain significant markups over production costs in their pricing structure. But weakening the behaviour norms that have developed around that association is difficult, and significant profit margins continue to be sustained by publishers using this model. Initiatives such as the San Francisco Declaration on Research Assessment (DORA)¹¹ encourage appointments committees to ignore publishing outlets in their decision making process, and have been adopted by many universities and faculties worldwide. Another development has been the emergence of article-specific citation measures rather than journal-specific measures, and it is probably no coincidence that this approach has been most obviously championed by Google, a company without entrenched brand association in this sector.¹²

3.2 Institutionally supported publishing infrastructure

A second business model developed in response to the transition away from a reader-pays model is the creation of institutionally funded digital publishing infrastructures. In this setting, neither the reader nor the author pays the publisher, but the publishers' operations are supported either directly or indirectly via institutional funding. There is a long history of universities financially supporting

¹⁰ There may be other aspects to publisher 'brand' that are also valued by authors, such as the quality of copy-editing, or speed of publication process - but these are more closely related to the production process itself (and so open to competition by others) rather than being used as a signal for the quality of the research published. It is the distortions generated by that signalling component that is the primary concern here.

¹¹ <https://sfdora.org/>

¹² Google adopted the calculation of H-Indexes for researchers, based on the citation rates of their individual articles. <https://en.wikipedia.org/wiki/H-index>

publishing operations - often through the creation of university presses - but here the focus is specifically on institutional support for digital publishing technologies.

3.2.1 University publishing services

Over two thirds of the 15,000 open access journals listed in the Directory of Open Access Journals (DOAJ) do not charge either readers or authors to publish.¹³ In many cases, these are small journals supported by the technical infrastructure provided by a specific institution. While this type of support has existed for small non-profit publishers for many years, there has been a growth in interest for university and library publishing services in recent years. A key characteristic of this approach is that research published this way is peer-reviewed via the publishing outlet - typically prior to publication.

A second type of university service involves the use of university or subject-specific repositories for the dissemination of research. Posting research papers in repositories is typically far more cost efficient than traditional publishing, and researchers in some disciplines (eg mathematics and astrophysics) have been using repositories (such as arXiv) as their primary route for the sharing and discovery of new results for some time - although this is often combined with subsequent publication of the paper by a 'traditional' journal, often explicitly for promotion purposes. Repositories form the foundations for the so-called "Green OA" publishing process. In recent months we have seen repositories being widely adopted for sharing results about COVID-19, with some observers suggesting that this experience will transform the behaviour of scholarly dissemination practices in the future.¹⁴

Repositories are typically associated with the online publication and dissemination of research findings prior to an external peer-review process. Consequently there remains a separate peer review/quality assessment service to be undertaken. As we have noted, in some cases this assessment is undertaken by 'traditional' publishers, but increasingly other means have evolved, such as so called 'overlay' journals and alternative post-publication peer-review processes.¹⁵

¹³ Search conducted on 4 August 2020 shows 15,047 open access journals listed on the Directory of Open Access Journals, and with 10,927 reported as not charging authors.
https://doaj.org/search?source=%7B%22query%22%3A%7B%22filtered%22%3A%7B%22filter%22%3A%7B%22bool%22%3A%7B%22must%22%3A%5B%7B%22term%22%3A%7B%22_type%22%3A%22journal%22%7D%7D%5D%7D%7D%2C%22query%22%3A%7B%22match_all%22%3A%7B%7D%7D%7D%7D

¹⁴ Kupferschmidt K. (2020).

¹⁵ Brown, J. (2010).

3.2.2 Public Digital Infrastructures to support publishing activities

One of the dangers of the new digital technologies is that they are expensive to implement and maintain (both financially and in terms of technological expertise) and so are really only viable if operated at scale. Thus digital technology, while clearly bringing down production costs, can also act as a major barrier to entry for small or innovative scholarly groups. The diversity of ideas and research processes is a critically important component of the scholarly sector, and would be a significant loss if it could not be maintained.

Creating open infrastructures to support smaller and new initiatives to engage with the international research community without requiring each to implement all the necessary digital processes themselves can be viewed as an enabling technology, with benefits not only to the individual initiatives but also to the wider research community.

An example of such an enabling infrastructure is [SciELO](#), an international network of over 1700 journals¹⁶ - primarily, but not exclusively, based in South America. There are several important technological features of this network. From the outset it has been designed as a modular platform, developed using open source software and adopting open standards and protocols between operational units.¹⁷ These features mean the journals or participating institutions do not all need to use identical technologies to be able to benefit from the network, nor do all journals need to implement all components. It also means that the technologies can be expanded and updated incrementally - rather than requiring wholesale and potentially costly technological updates and simultaneous adoption by multiple small players - and mitigating against participant lock-in to old technologies.

A second example is [Open Journal Systems](#) (OJS) which is a free and open source system for creating and managing a journal, built by the Public Knowledge Project (PKP), that is being used by over 10,000 journals worldwide. This is a system that can be reasonably quickly and easily installed on a local server and allows smaller journals and new entrants the opportunity to utilise powerful digital journal management and digital dissemination technologies.

¹⁶ <https://scielo.org/en/journals/list-by-subject-area>

¹⁷ It may be worth noting that open source code is not in itself sufficient to create an open platform. Android is built using open source code, however it has been effectively controlled by Google through technical dependencies and closed protocols between some key components of the system (Google Play Services).

The key point to note here is that these open digital infrastructures can be viewed as a ‘public infrastructure’, which can significantly lower barriers to entry into the scholarly communications sector and have a substantial impact on the sector by encouraging and sustaining a large and diverse range of mission-led initiatives - a group with great value to scholarly activity.

We should also recollect that non-profit entities are naturally at a disadvantage in raising capital to finance digital platforms such as these. Funding from national research programmes is likely to be one of the few feasible financing options available.¹⁸

3.2.3 Direct funder publishing initiatives:

Research funding agencies are increasingly developing their own publishing outlets specifically for the research they have funded. Examples include the Wellcome Trust and the Gates Foundation,¹⁹ while the EU has recently awarded a tender to create a digital platform for EU funded research.²⁰ Interestingly, F1000 - the relatively small technology supplier for the infrastructure used to support all these initiatives - was recently acquired by Taylor and Francis,²¹ one of the ‘Big Four’ commercial scholarly publishers, an issue we address below.

3.3 Multi-sided markets model

The third major development in scholarly publishing business models has been the recognition by the largest publishers that they can use their digital publishing platforms to simultaneously operate in multiple different product markets, and so use the platform to create profits through the provision of products not typically associated with publishing.

Digital platforms have emerged to dominate many digital markets - and, indeed, such platforms are the primary focus of the antitrust concerns raised at the start of this piece. The key features of a

¹⁸ In recent years we have seen initiatives such as Invest in Open Infrastructure (<https://investinopen.org>) and SCOSS (<https://scoss.org/>) advocating for coordinated funding of these infrastructures, and we have seen direct funding for initiatives developing such infrastructures from public research funding bodies such Research England (e.g. COPIM (<https://www.copim.ac.uk/>)) and the European Funding Council (e.g. OPERAS <https://operas.hypotheses.org/>).

¹⁹ <https://www.nature.com/news/gates-foundation-announces-open-access-publishing-venture-1.21700#:~:text=The%20Gates%20Foundation%20has%2C%20similarly,by%20its%20in%2Dhouse%20editors.>

²⁰ https://ec.europa.eu/info/news/european-commission-awards-contract-setting-open-access-publishing-platform-2020-mar-20_en

²¹ <https://newsroom.taylorandfrancisgroup.com/f1000-research-joins-taylor-francis/>

digital platform and some of the strategies that commercial platforms employ to enhance the profits made from them are summarised below. These strategies often differ significantly from those adopted by firms in 'traditional' markets.²²

First, it is extremely valuable for a platform to have a group of people who are persistent (or captive) users, as other retailers may be willing to pay the platform a fee to enable them to 'access' those users. In fact, attracting users to the platform might be so valuable that it is worthwhile providing a highly valued product for free solely to attract more users and so gain greater revenue from other companies wishing to sell items to those users.²³ This, for example, is Google's strategy in providing valuable products such as Google Search or Google Maps to users for free. By having users engage with these free products, Google is able to sell associated advertising space on their webpages to other companies - and the profits Google achieves from the advertising activity alone far exceeds anything they could expect to earn from the sale of the products themselves²⁴.

In the case of scholarly publishers, the captive users are those researchers and readers that visit their platform to access the articles published there. To the extent that the publisher can prevent the articles being available anywhere else, they are able to sustain a captive (and therefore valuable) audience on their platform.²⁵ There are significant returns to scale in this activity: Increasing the quantity of content in any specific subject area will increase the regularity with which an individual researcher visits the platform, and increasing the diversity of content will increase the number of researchers visiting the platform. Both are clearly advantageous - so the value of the platform increases with both the breadth and depth of the content on the platform.

A second strategy for obtaining profits from a platform is either to provide or acquire products or services with which the same population of users interact. This becomes particularly valuable if

²² Levin, J. D. (2013), Rysman M. (2009)

²³ Weyl E.G. (2010), Gao M. (2018).

²⁴ In 2019 Alphabet (the parent company of Google) made US\$134.8bn (84% of their total revenue) from advertising. <https://www.statista.com/statistics/633651/alphabet-annual-global-revenue-by-segment/>

²⁵ Publishers typically host digital content on their own servers rather than allowing libraries to host the content directly so, unlike printed works, perpetual access to digital content purchased by the library through a subscription is not assured. Commercial publishers remain particularly concerned about any 'leakage' of traffic from their site when content is hosted elsewhere. Concern over loss of control of content with open access publications is another reason for publishers to technically tie the content to other features on the site. See also Moore S.A. (forthcoming).

dependencies can be introduced between the two products - making it difficult for the users to select alternatives. So, for example, a publisher may acquire a university repository or data management service and build a dependency between their repository service and their publishing service. In this way, authors of papers published on the platform can easily integrate their publications with the underlying database using the publisher's database product, but it will be more difficult to do this using a different database provider. If the publishing platform is relatively large, and the dependencies sufficiently great, the publishing platform could obtain a significant competitive advantage over any competing supplier of database solutions and so make significant profits in that sector, while reinforcing further usage of their publishing platform. The publishing platform is in a position to make profits from the sale of the database solutions - even though they are not charging either authors or readers for the publication itself.²⁶

Notice that for both the authors and the scholarly community generally, the integration of the functionality and the interaction between published output and the underlying data is a good thing, that should rightly be applauded. The problem here lies with the dependencies built into the integration: when smooth functioning requires specific software solutions. These dependencies may be technical - encoded into the software - or financial - the products are jointly provided as part of a bundle sold together (or a 'big deal' as they are typically referred to in the publishing sector).

Product bundling and technical dependencies of this nature are generally frowned upon by antitrust regulators - and most of the large digital providers have fallen foul of these concerns at some point in their history. But, as we have seen, antitrust regulators find it very difficult to assess and respond to this behaviour. One strategy particularly valued by digital platforms is to acquire new and innovative products and solutions early in their life cycle - while they are relatively small - and then either integrate them into their own service or close them down; in either case they successfully protect their own product from being disrupted. This is a particularly difficult strategy for regulators to assess when considering merger and acquisition cases.²⁷

In recent years, all of the four largest publishing platforms have acquired products and services designed for researchers at all stages of the research cycle.²⁸ RELX Group (the parent of Elsevier) now

²⁶ Aspesi et al. (2019) provide a detailed and sobering account of acquisition strategies of large publishers within scholarly communications and education resources sectors

²⁷ Report for the UK Competition and Markets Authority (2019).

²⁸ Posado & Chen (2018) detail acquisitions by RELX, Wiley and Taylor and Francis up until 2017, showing acquisitions by RELX and Wiley have been dominated by academic services rather than content over the

describes itself not as a media company but as an information and digital analytics company, having successfully reduced its dependence on print sales from over 50% to under 10% in the last 10 years - with the vast majority of revenues now derived from digital products and services.²⁹

Posada & Chen (2018) show the positioning of RELX acquired products over the past few years throughout the research life-cycle:

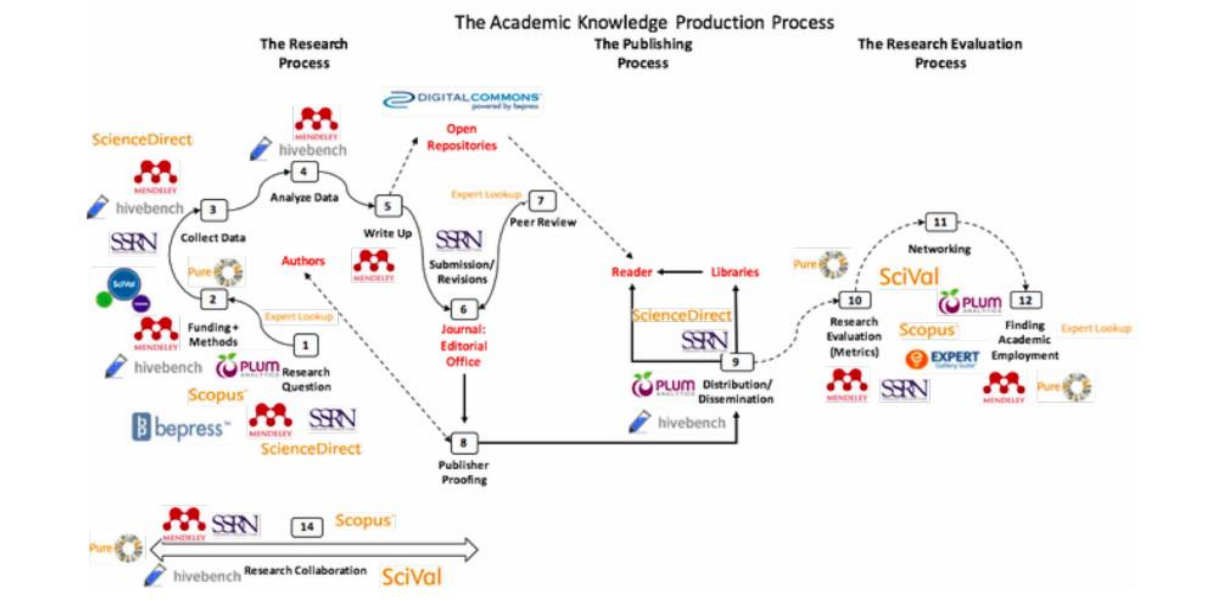


Figure 1: Elsevier’s Presence Throughout the Research Lifecycle.³⁰

Increasingly Elsevier is looking to tie analytics and digital services into their existing publishing agreements with universities. A report by SPARC commenting on this year’s deal between Elsevier and a national coalition of Dutch Universities notes that “The general terms of the deal included a zero increase in total spending for access to subscriptions and on publishing articles in open access, in exchange for the broad adoption of Elsevier’s data analytics tools in Dutch universities.”³¹ The report goes on to observe that there are a number of competitors in the analytics market that will be

preceding decade. Springer Nature is owned by the private company Holtzbrinck Publishing Group, which has similarly developed a broad portfolio of academic services, particularly within its Digital Science Group.

²⁹ <https://www.relx.com/our-business/our-stories/transformation-to-analytics>

³⁰ Posada A. & Chen G. (2018). Figure 5, p. 6.

³¹ Aspesi, C. et al. (2020).

unable to compete with such a joint publishing/analytics platform and are likely to be driven out of the market.

One fundamental way of mitigating technical dependencies, and one that lies at the heart of the success of the World Wide Web, is to require interoperability between competing systems - which can be achieved through the creation and monitoring of open data formats and protocols for data exchange, and the creation of open metadata.³²

A third potential source of revenue for publishing platforms comes from the analysis of usage data. For many platforms this is the most valuable resource of all - and is sometimes described as the 'new oil' in the digital age. There are numerous ways this usage data can generate commercial gains. Firstly, it can be used to identify new trends in the operations and activities of users, thus allowing the platform to have an advantage in refining their products or introducing new services. Secondly, it can be analysed and repackaged for other users. A particular example in the research sector would be to provide third parties (universities, appointment boards, research funders) with comparative assessments of the quality of research by various scholars. Having an integrated platform across the entire scholarly cycle will enable the platform to report on and compare how frequently a scholar uploads research data, the quality of the data uploaded, the time taken between data upload and analysis, who their co-authors are, and how much each person actually contributed to any jointly authored papers, how many journals the scholar submits to prior to publication (author rejection rates), and much more besides.

This should be an issue of considerable concern and debate within the scholarly community. A primary issue is that no matter how good or bad an assessment measure may be, if specific measures are applied in settings authors care about (e.g. promotion or funding decisions) they have little alternative but to engage with them, and so with the platform providing those metrics. If promotion boards or funders rely on the statistics provided by a single platform, authors will only be willing to submit research via services provided by (or that score well within) that platform - and the platform is now in the enviable strategic position of controlling the entire research cycle for those scholars. In the process, it is now the platform which defines what is 'acceptable' research and which research

³² In many industries there are examples, such as CrossRef and BISG, where competing companies work together to create common standards. But these industry groups will be successful only when there is common commercial gain from interoperability. In this section we are referring to dependencies introduced to prevent interoperability - so industry groups cannot be relied upon to monitor or self-regulate these activities.

processes 'count' in assessments - something the scholarly community may feel reluctant to delegate to a commercial publishing platform.

There are numerous instances of this already. For example, many funders and universities evaluate researchers' performance using data and reports obtained via indexing services such as Scopus or Web of Science. Of course, such reports necessarily ignore any outputs not indexed by these services - so apparently those outputs are not to be valued in the assessment process. But these indexes are owned by commercial operators - who make it extremely difficult for many publishing outlets to join their index. New outlets, innovative outlets, outlets in many developing countries, book publishers, and many more find it difficult or impossible to be "accepted". Of course, making it difficult for new entrants supports the business strategy of the controlling publisher, and is detrimental to innovation in the scholarly communication sector generally. In this light, the announcement that Elsevier are launching a new International Centre for the Study of Research "tasked with examining and advancing the evaluation of research across all fields of knowledge production"³³ should be viewed with critical concern by the scholarly community.³⁴

Ensuring that the processes used for scholarly assessment are openly defined and transparently applied, that diversity in assessment processes are enabled, and that control over the underlying data used in any assessment is itself not subject to commercial 'capture' are all important components of any solution that preserves scholarly independence in the assessment process.

4. What can the scholarly community do to maintain an innovative, vibrant and open scholarly communications sector?

As noted in the introduction, the behaviour beginning to be observed within the scholarly publishing sector is familiar to many other sectors - so we can look to the general recommendations made in other areas to help answer this question. For example, the UK report referenced in the introduction goes on to recommend the establishment of a national 'digital markets unit' with a series of specific tasks:

"First, it would develop a code of competitive conduct, with the participation of stakeholders. This would be applied only to particularly powerful companies, those deemed

³³ Elsevier (2019)

³⁴ Sarah Lamden (2019) documents how some major publishers are selling researcher data collected from university library systems to immigration authorities in some countries.

to have 'strategic market status', in order to avoid creating new burdens or barriers for smaller firms.

Second, the digital markets unit would be charged with enabling greater personal data mobility and systems with open standards where these tools will increase competition and consumer choice. [...]

Third, the digital markets unit would be able to advance data openness where access to non-personal or anonymised data will tackle the key barrier to entry in a digital market, while protecting privacy.”³⁵

These are transferable concepts, and as such can be seen as a set of recommendations for the international scholarly community to consider implementing as well. Of course such a unit would not have the judicial or legal powers that may be associated with a national antitrust agency, however, that does not mean it could not be extremely influential in determining the structure and behaviour of the scholarly communications sector going forward, providing the key stakeholders, national funding bodies and universities were prepared to support it. The largest scholarly publishers are international organisations that are not particularly constrained by local needs or legal structures. But the international scientific and scholarly community has the potential collectively to apply significant pressure on the publishing activities of even the largest publishers. Creating a body that could identify and highlight concerning and anti-competitive activities by publishers, and facilitate coordinated responses by institutions globally when they negotiate contracts with publishers, could be a powerful force. Such a digital markets unit would provide a useful reference point and mechanism for building consensus amongst institutions and countries that are negotiating large and often complex multifaceted contracts with commercial publishing platforms. The costs of creating and operating such a unit is likely to be small in comparison to the global social and financial impact it could have for the scientific and scholarly community.

Finally, we should note that the scholarly community has a set of policy tools not available to regulators: the ability to support and direct the development of open scholarly research and publishing infrastructures that can operate alongside commercial platforms. As was noted in section 3.2, such infrastructure can directly help to sustain independent research and publishing activities; but they also provide viable alternatives that limit the market power and markups of commercial

³⁵ Furman et al. (2019), pp. 5-6. The report goes on to make a number of other recommendations, but the remainder are tailored more specifically for national antitrust authorities.

platforms, and provide the mechanisms for creating and maintaining open standards, protocols, data models, researcher privacy practices and data analytic processes across the international research sector. All of this would help ensure interoperability between the large commercial digital platforms, and mitigate the possibilities of technological tie-in and anti-competitive behaviour by dominant players. Clearly the development of these infrastructures require local funding and need to be responsive to local needs - but there is the potential for a major international multi-disciplinary scholarly association to take a leadership role in facilitating and helping to coordinate the actions and investments of national and disciplinary specific initiatives.

Without coordinated international action of this nature by the research community it seems inevitable that scholarly research processes will, over the next few years, become dominated by a small number of large commercial digital platforms.

References

- Aspesi, C., Allen, N. S., Crow, R., Daugherty, S., Joseph, H., McArthur, J. T. & Shockey, N. (2019). "SPARC Landscape Analysis." March 29. doi: [10.31229/osf.io/58yhb](https://doi.org/10.31229/osf.io/58yhb)
- Aspesi, C., Allen, N. S., Crow, R., Daugherty, S., Joseph, H., McArthur, J. T. & Shockey, N. (2020). "2020 Update: SPARC Landscape Analysis & Roadmap for Action." , June 22. <https://doi.org/10.31229/osf.io/2pwft>
- Brown, J. (2010). "An Introduction to Overlay Journals." Repositories Support Project: UK. <https://discovery.ucl.ac.uk/id/eprint/19081/1/19081.pdf>
- Boulton G. et al. (2020). "Opening the Record of science: making scholarly publishing work for science in the digital era" Discussion Document for International Science Council. July.
- Burchardt, J. (2014). "Researchers Outside APC-Financed Open Access: Implications for Scholars Without a Paying Institution." *SAGE Open*. <https://doi.org/10.1177/2158244014551714>
- Elsevier (2019) "Elsevier Announces the International Center for the Study of Research." *PR Newswire*, June 19 2019. <https://www.prnewswire.com/news-releases/elsevier-announces-the-international-center-for-the-study-of-research-300871135.html>
- Furman J. et al. (2019). "Unlocking Digital Competition: Report of the Digital Competition Expert Panel." March. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/785547/unlocking_digital_competition_furman_review_web.pdf
- Gao M. (2018) "Platform pricing in mixed two-sided markets" *International Economic Review*, **59(3)** August, pp.1103-1129. doi:[10.1111/iere.12298](https://doi.org/10.1111/iere.12298)
- Houghton J. & Oppenheim C. (2010) "The Economic Implications of Alternative Publishing Models" *Prometheus*, **28(1)**, pp. 41-54. doi:[10.1080/08109021003676359](https://doi.org/10.1080/08109021003676359)
- Houghton J. & Vickery M. (2005). *Digital Broadband Content: Scientific Publishing*. OECD Publishing, Paris. doi:[10.1787/9789264065901-en](https://doi.org/10.1787/9789264065901-en)

- Kupferschmidt K. (2020) "A Completely New Culture to Doing Research" *Science* Feb. 26, 2020. doi:[10.1126/science.abb4761](https://doi.org/10.1126/science.abb4761)
- Lamdan, S. (2019). "Librarianship at the Crossroads of ICE Surveillance." *In the library with the lead pipe*. <http://www.inthelibrarywiththeleadpipe.org/2019/ice-surveillance/>
- Levin, J. D. (2013) "The economics of internet markets" In Acemoglu D., Arellano M. & Dekel E. (Eds.), *Advances in Economics and Econometrics: Tenth World Congress* (Econometric Society Monographs, pp.48-75). Cambridge: Cambridge University Press. doi:[10.1017/CBO9781139060011.003](https://doi.org/10.1017/CBO9781139060011.003)
- Moore, S.A. (Forthcoming). "Individuation through infrastructure: Get Full Text Research, data extraction and the academic publishing oligopoly", *Journal of Documentation*. doi: [10.1108/JD-06-2020-0090](https://doi.org/10.1108/JD-06-2020-0090)
- Posada A. & Chen G. (2018). "Inequality in Knowledge Production: The Integration of Academic Infrastructure by Big Publishers." ELPUB 2018 Proceedings, June, Toronto, Canada. doi: [10.4000/proceedings.elpub.2018.30](https://doi.org/10.4000/proceedings.elpub.2018.30)
- Rysman M. (2009) "The Economics of Two-Sided Markets" *Journal of Economic Perspectives*, **23(3)** Summer, pp.125–143. doi: [10.1257/jep.23.3.125](https://doi.org/10.1257/jep.23.3.125)
- Schönfelder, N. (2020). "Article Processing Charges: Mirroring the Citation Impact or Legacy of the Subscription-based Model?" *Quantitative Science Studies*, **1(1)**, pp. 6–27. doi: [10.1162/qss_a_00015](https://doi.org/10.1162/qss_a_00015)
- UK Competition and Markets Authority (2019). "Ex-post Assessment of Merger Control Decisions in Digital Markets: Final Report." May. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/803576/CMA_past_digital_mergers_GOV.UK_version.pdf
- Weyl E.G. (2010) "A Price Theory of Multi-Sided Platforms." *American Economic Review*, **100** September, pp.1642–1672. doi:[10.1257/aer.100.4.1642](https://doi.org/10.1257/aer.100.4.1642)