

November 2023

ACCESS TO SCIENCE AND SCHOLARSHIP: KEY QUESTIONS ABOUT THE FUTURE OF RESEARCH PUBLISHING

MIT FACULTY AUTHORS

Phillip A. Sharp (Chair), Institute Professor Emeritus, Department of Biology and Koch Institute of Integrative Cancer Research

William B. Bonvillian, Senior Director, Special Projects, Open Learning

Robert Desimone, Director, McGovern Institute for Brain Research

Barbara Imperiali, Class of 1922 Professor of Biology

David R. Karger, Professor of Electrical Engineering

Clapperton Chakanetsa Mavhunga, Professor of Science, Technology, and Society

OTHER AUTHORS

Amy Brand, Director and Publisher, The MIT Press

Nick Lindsay, Director for Journals and Open Access, The MIT Press

Michael Stebbins, Science Advisors, LLC

This unpublished report was authored by a self-selected group of MIT faculty and staff, along with an external consultant, who met periodically during the spring of 2023 to explore the scientific and scholarly communications landscape in relation to changes in publishing business models and public access policy. The views expressed in this report are the consensus views of the signed authors, and not an official statement on behalf of MIT.

CONTENTS

Acknowledgments 3

Executive Summary 4

Introduction 6

PART 1: Business Models in Journal Publishing: Looking Back and Ahead 11

Growth in Article Output 11

A Wave of Consolidation 15

Open Access 17

Research Funder Policies 28

Plan S 30

The Nelson Memo 32

Research Data 35

Vertical Integration in the Infrastructure of Research Communication 36

The Imperative for Experimentation 39

Diamond Open Access 40

PART 2: Open Questions on the Future of Academic Publishing in Six Key Areas Where More Research and Data Are Needed 43

Access and Business Models 44

Research Data 46

Preprint Publishing 47

Peer Review 48

Costs to Researchers and Universities 49

Infrastructure 51

Appendix A: Other Questions Generated by the Group 53

Access and Business Models

Research Data

Preprint Publishing

Peer Review

Costs to Researchers and Universities

Appendix B: Glossary 56

ACKNOWLEDGMENTS

The following individuals generously provided input during the drafting of this report: Claudio Aspesi (Independent Consultant), Chris Bourg (Director of the MIT Libraries), David Crotty (Senior Consultant, Clarke & Esposito) David Goldston (Director, Washington DC Office, MIT), Kevin Hogan (Technology Licensing Officer, MIT), John Inglis (Executive Director and Publisher, Cold Spring Harbor Laboratory Press), Lesley Millar-Nicholson (Executive Director, Technology Licensing Office, MIT), Aude Oliva (Schwarzman College of Computing, MIT), Heather Anne Paxson (Professor of Anthropology, MIT), Deb Roy (Professor of Media Arts and Sciences, MIT), and Erin S. Stalberg (Associate Director of Collections, MIT Libraries).

EXECUTIVE SUMMARY

The health of the research enterprise is closely tied to the effectiveness of the scientific and scholarly publishing ecosystem. Policy, technology, and market-driven changes in publishing models over the last two decades have triggered a number of disruptions within this ecosystem:

- Ongoing increases in the cost of journal publishing, with dominant open access models shifting costs from subscribers to authors
- Significant consolidation and vertical integration in the publishing industry, and a decline in society-owned subscription journals that have long subsidized scientific and scholarly societies
- A dramatic increase in what are increasingly termed “predatory journals” by industry analysts, with substandard peer review
- Decline in the purchasing power of academic libraries relative to the quantity and cost of published research

Shifting costs to researchers means less well-funded researchers and institutions may face new hurdles in getting their research published; this is also a concern for researchers in less developed nations. Consolidation and profit objectives in academic publishing have indirectly led to concerns about the quality of peer review in many journals and the selection of publishing options available to researchers. Ecosystemic changes in research publishing are proving especially challenging for smaller non-profit publishers, including university presses and professional societies, even as they embrace the principles of open and equitable access to research. Predatory journals are actively contributing to the growing problems of misinformation and mistrust in science. And the health of academic libraries fundamentally impacts research access for faculty, students, and other scholars.

Many in the research community remain unaware of the drivers of change in academic publishing and the potential consequences for the research enterprise. Although they are often left out of the conversation, researchers are directly affected by these developments. Decisions are being made now that could potentially reduce research budgets, increase researcher workloads, and alter publishing options and the reputational function that publishing has long played.

To illustrate how researcher behavior, funder policies, and publisher business models and incentives interact, part 1 of this report presents an historical overview of open access publishing. Part 2 of the report provides a list of key questions for further investigation to understand, measure, and best prepare for the impact of new policies related to open access in research publishing, categorized into six general areas: access and business models, research data, preprint publishing, peer review, costs to researchers and universities, and infrastructure.

Note, these questions are set out to provide a research agenda, not to recommend particular strategies or practices. Efforts to address these questions will help inform the implementation of open access policies, as well as future policy development.

INTRODUCTION

Publishing, broadly defined to include the diverse apparatus of scientific and scholarly communications, is a pillar of the academic and greater research enterprise. Publishing is how knowledge is curated and disseminated (for both research and teaching), and also how the quality and impact of new knowledge are assessed. The health of the research enterprise depends in no small part upon the health of the academic publishing ecosystem.

Digital media have radically transformed the dissemination of all information, impacting news, entertainment, personal and social communications, as well as science and scholarship. Several additional factors are accelerating change in research publishing models, among them:

- The urgency to discover and share potential solutions to global crises that range from climate change and the spread of infectious disease to mental illness and the implications of advanced AI for humanity
- The widespread adoption of preprint publishing models to speed research dissemination in many fields
- Pervasive digital and print piracy of paywalled research publications, which incentivizes publishers to adopt alternative business models
- Consolidation in the publishing industry, with fewer companies owning an increasing share of the journal publishing market
- An academic peer review system under strain (due to lack of appropriate incentives and credit for peer reviewers, inconsistent quality in peer review, and the absence of mechanisms for ensuring fairness in the selection of peer reviewers, among other factors)
- The imperative of equity and inclusivity in access to knowledge, and the resulting growth of diverse open access publishing models

Although the report does not address the impact of generative AI technologies, this too will accelerate change in science and scientific publishing.

Public access policies now being adopted by governments around the world (most notably in the U.S., the U.K., and the E.U.) are intended to remove paywalls from the outputs of publicly funded research, with the undisputed imperative of enabling globally equitable access to knowledge. The new public access directive from the U.S. Office of Science and Technology Policy (OSTP) takes effect in January of 2026 and, in the interim, requires each federal funding agency to develop its own plan for policy-compliant research awards. More specifically, the new policy requires that research publications be “made freely available and publicly accessible by default in agency-designated repositories without any embargo or delay after publication.”

This period prior to full implementation of the policy is an opportunity for the academic community to consider the practical implications of the policy and to raise questions that can inform the implementation of current public access policies as well as future policy development. Is open access changing who is conducting, publishing, or consuming research results? How will open access impact the availability of funds for research, the amount and kind of research being published, publishing models and strategies, researcher workflows, and how science is communicated to the general public? Most importantly, how can we ensure that the academic community's response to changes in policy ultimately serves to strengthen the research ecosystem?

We note that there are a range of perspectives on open publishing across academia, among them differing perspectives on how to balance the desirability for openness in the research enterprise with security, privacy, and needs for authors and their institutions to sometimes delay release of research products. There are also important differences across disciplines (and sub-

disciplines) in how “open” they already are. For example, in some disciplines, sharing preprints is a long-accepted norm, and in others, open sharing of data is common.

One voice that needs to be better represented in the future-of-publishing conversation is that of the researcher. Decisions are being made now that will potentially reduce researchers’ budgets, due to new publication costs, and increase their workloads, and also potentially alter the reputational function that publishing has long played in academic career advancement. There is as yet little practical understanding of how much time, effort, and funding open science will require of the average researcher.

Furthermore, there has been limited discussion of the technological, cultural, and business model innovations necessary to reduce costs and burden to researchers and improve the communication of scholarly works. To protect the health of the research enterprise, the impacts and trade-offs associated with open access publishing and open science practices need to be better analyzed and clearly understood. Ideally, changes in policy that promote increased access to research outputs should be evidence-driven and designed to strengthen the research ecosystem.

To illustrate how researcher behavior, funder policies, and publisher business models and incentives interact, a history of open access publishing is presented below, along with a list of critical questions that should be explored further.

Several key observations emerge:

1. Open access has been one factor driving consolidation in the publishing sector, and has created financial incentives for many publishers to increase the number of articles published. This trend towards “quantity over quality” risks reducing high-quality publishing options for

researchers (meaning, journals that adhere to established norms of peer review, curation, and editing). The future sustainability of research quality is now an issue, as is preserving a diversity of publishing options. The future of non-profit publishers, including scientific society and university presses, and the development and maintenance of new publishing pathways (e.g., preprint servers), will likely require research funders and universities to make (or increase) direct investments in the scholarly communications ecosystem.

2. Open access has been implemented in ways that have advantaged commercial publishers to the detriment of most scientific society and university publishers — although this has clearly not been the intention of policy makers. Non-profit publishers lack the financial resources to pivot quickly and to compete at scale. Without intervention by key stakeholders, open access will lead to more money flowing from universities, governments, and other research funders to commercial publishers. Money flows in the form of both “pay to play” (Gold) open access and purchased academic information technologies, thus tying academic reputation to the ability to pay publishers. A publishing ecosystem dominated by — or consisting only of — a few major commercial publishers would reduce the variety of publishing options and leave universities and researchers without choices on pricing and publishing approaches. It would also leave poorly-funded or unfunded researchers from the U.S. and abroad increasingly disadvantaged compared to well-funded researchers.
3. There is no one-size-fits-all model in academic publishing. Different fields of study have different cultures, communication norms, funding models, and publishing prerogatives. Hence, the future of research

publishing and open access should involve a variety of models (Green OA, Gold OA, Diamond OA, preprint publishing, etc.).

4. The fast-changing landscape is causing confusion among researchers. Many in the community remain unaware of the drivers of change in academic publishing and the potential consequences for the research ecosystem, underscoring the need for an in-depth, evidence-driven examination of key questions concerning the future of research publishing. Several guiding questions are provided in the last section of this report.

We acknowledge that this effort is limited by the absence of readily accessible, comprehensive data about the current state of scientific publication. We likewise acknowledge that there are a range of perspectives on open science and open publishing across academia, including on how to balance the desirability for openness in the research enterprise with security, privacy, and cost.

PART 1 – BUSINESS MODELS IN JOURNAL PUBLISHING: LOOKING BACK AND AHEAD

Scholarly communication is undergoing rapid change in which business models, platforms, and services are co-evolving. In journal publishing, quality has traditionally stemmed from serving readers and institutions well. In recent years, with the shift to open access and author-paid models, providing competitive services to authors as a primary customer base takes increasing precedence. This also impacts publishing practices in peer review, editing, and marketing.

Until the mid-1990s, journals were largely sold as print subscriptions to individuals, institutions, and other research entities. In the mid-1990s, journals started moving online, and access to content was typically provided through individual usernames and passwords at institutions. Over time, institutional sales shifted to site licenses, in which an institution's library purchases a subscription for the whole campus, and access is granted through IP address identification.

GROWTH IN ARTICLE OUTPUT

The last several decades have seen significant growth in the number of journals and articles published. There was a nearly fivefold increase in the number of articles produced annually from 1995 and 2022 (Figure 1), placing strain on library budgets. Contributing factors include increased global research funding (with huge growth in China for example, see Figure 2 below) combined with the continued emphasis on “publish or perish” in academia — with hiring, promotion, and funding decisions dependent on publication output. The growth in scholarly literature produced every year is also causing the number of peer review requests to skyrocket, raising concerns that the scientific community cannot sustain the pressure of increased requests to peer review publications

and that continued increases in peer review requests could potentially erode the overall quality of peer review.

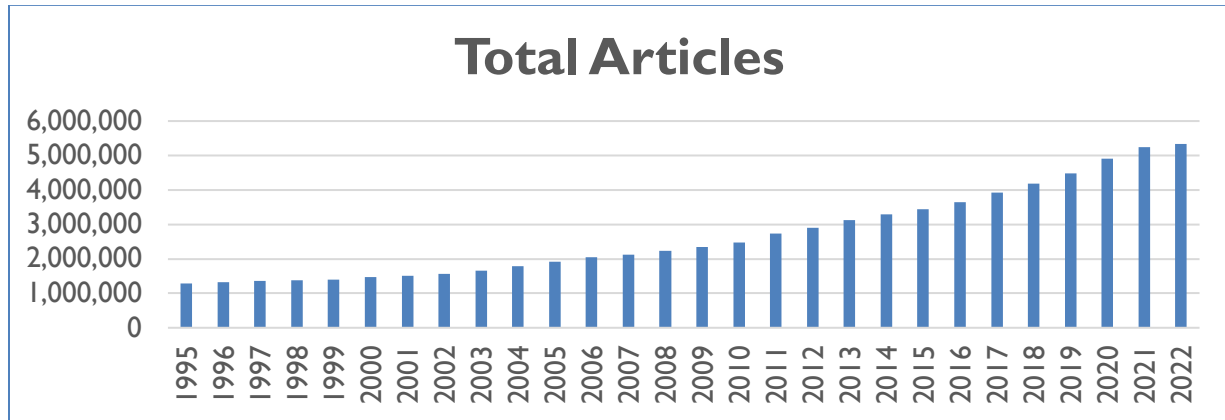


Figure 1. Source: Dimensions (Digital Science). Search limited to “articles”.

1998 — 2000 (PY) (Average)				2008 — 2010(PY) (Average)				2018 — 2020 (PY) (Average)			
The number of papers				The number of papers				The number of papers			
Country/Region	Fractional counting			Country/Region	Fractional counting			Country/Region	Fractional counting		
	Papers	Share	World rank		Papers	Share	World rank		Papers	Share	World rank
U.S.	203,669	27.9	1	U.S.	246,188	22.7	1	China	407,181	23.4	1
Japan	64,752	8.9	2	China	107,955	10.0	2	U.S.	293,434	16.8	2
Germany	51,597	7.1	3	Japan	64,783	6.0	3	Germany	69,766	4.0	3
U.K.	51,053	7.0	4	Germany	58,095	5.4	4	India	69,067	4.0	4
France	37,657	5.2	5	U.K.	54,116	5.0	5	Japan	67,688	3.9	5
Italy	24,707	3.4	6	France	42,811	4.0	6	U.K.	65,464	3.8	6
Canada	24,320	3.3	7	Italy	36,858	3.4	7	Korea	53,310	3.1	7
China	22,549	3.1	8	India	35,150	3.2	8	Italy	52,110	3.0	8
Russia	22,351	3.1	9	Canada	34,913	3.2	9	France	45,364	2.6	9
Spain	17,140	2.3	10	Korea	31,650	2.9	10	Canada	43,560	2.5	10

Figure 2. Average papers per year per country. Digest of Japanese Science and Technology Indicators 2022, Chart 15¹

Although comprehensive data is not available, libraries anecdotally report that their budgets have remained relatively flat or declined relative to inflation over the last few decades. One study notes the decline of library spending as a percentage of university expenditures (Figure 3). (Note: This data covers 1982 to 2017, and more recent data on this metric does not appear to be publicly available.). Regardless of their budgets in absolute terms or how universities prioritize library spending, libraries whose budgets have remained flat

¹ https://www.nistep.go.jp/en/wp-content/uploads/NISTEP-RM318-SummaryE_R.pdf

effectively have less purchasing power in the face of continued growth in research publications.

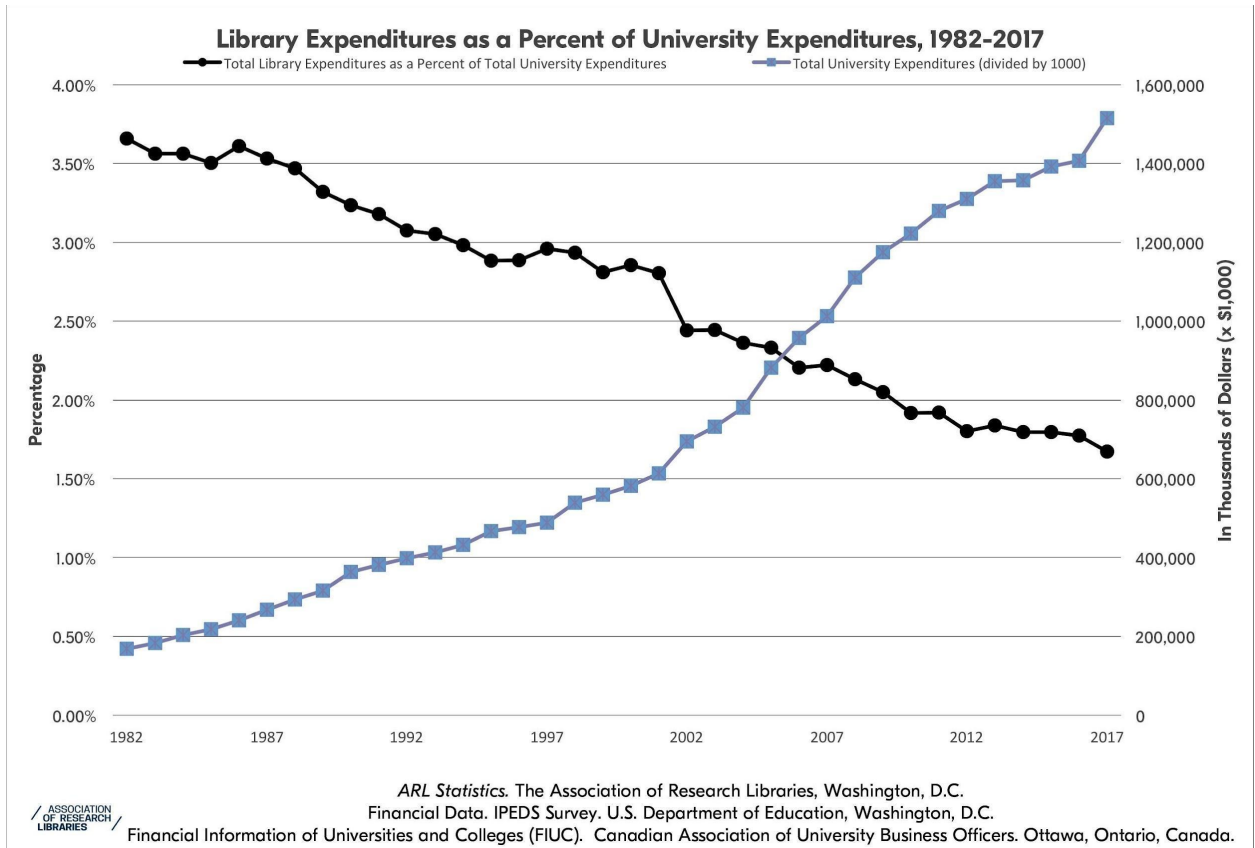


Figure 3. Library expenditures as a percent of university expenditures, 1982-2017, Association of Research Libraries²

This growth contributed to the rise in bulk pricing, and what has become known colloquially as “The Big Deal”. Rather than purchasing individual journals from a publisher at list price, libraries were offered annual access to a package of journals at a discounted price — often the entire catalog of journals from a given publisher.

The Big Deal made more research content available on participating campuses, and resulted in declining cost per article download. (Figure 4).

² <https://www.arl.org/wp-content/uploads/2020/03/ARL-Library-Expenditures.pdf>

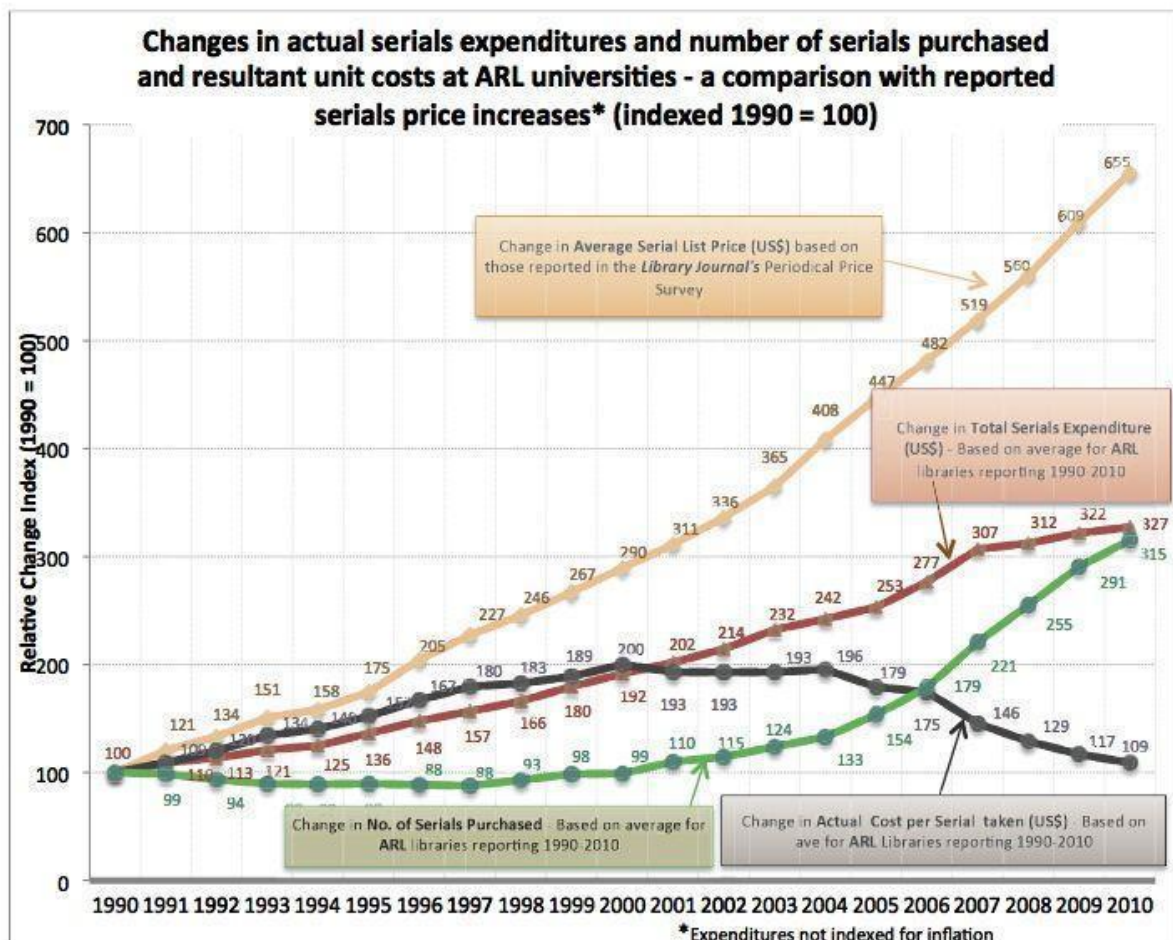


Figure 4. Changes in Expenditures, Gantz, P. (2013) Journal print subscription price increases no longer reflect actual costs. *Learned Publishing* **26**, p. 155-231³

Journal packages from larger publishers are expensive, and prices rise every year. Over time, an increasing proportion of library collections budgets have had to be allocated to purchasing journal access from large commercial publishers, leaving decreasing funds for books and subscriptions to journals from smaller publishers, including scientific societies and university presses.

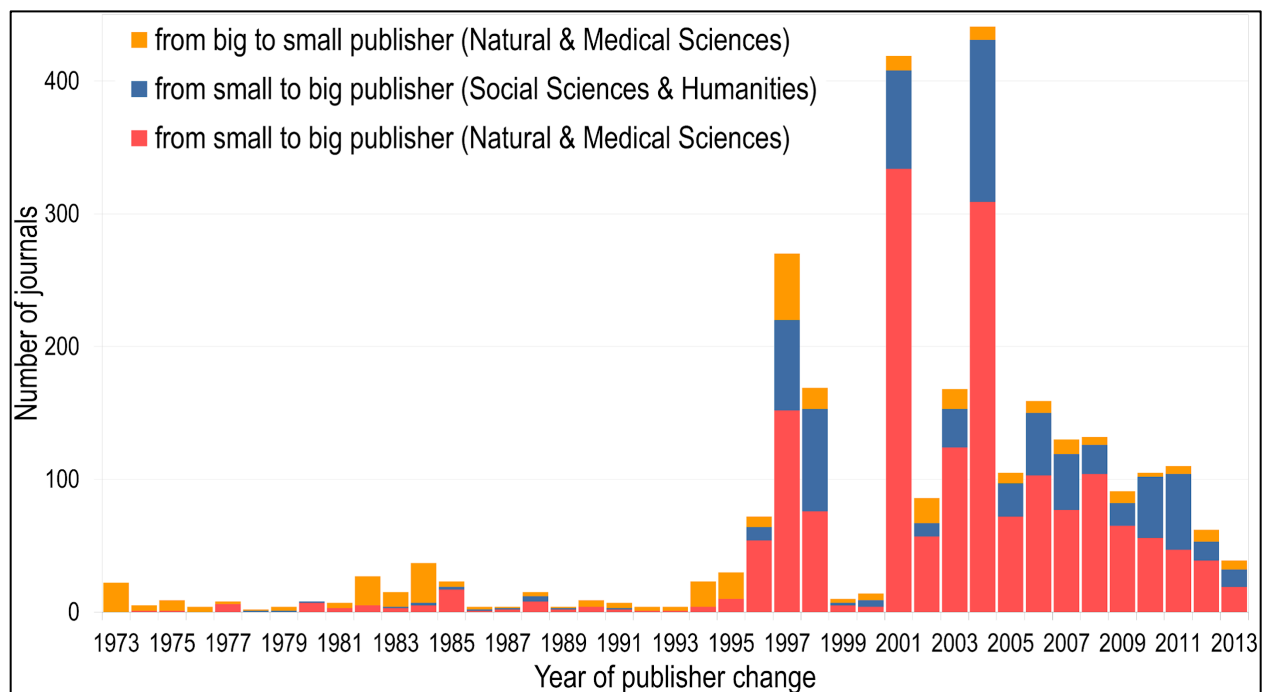
Before the Big Deal, a library could cancel individual journals that didn't see much use on campus, and redirect the savings elsewhere. Under the Big Deal,

³ <https://doi.org/10.1087/20130309>

library budgets effectively became locked in, since canceling an individual journal in a package is typically not an option. When it is possible, cancellation of one journal has little to no effect on the price of the overall package. As a result, journals from publishers outside of Big Deal packages have been the most vulnerable to cancellation, driving market consolidation. Many libraries are now canceling their Big Deal subscriptions, although the rate and scope of these cancellations have not been studied.

A WAVE OF CONSOLIDATION

The consolidation that resulted has been pronounced⁴. Individual journals, or entire publishing houses, were sold to the larger publishers, and many research societies with independent journals chose to sign Publishing Services Agreements in which they outsource their publishing operations to a larger publisher and become part of their sales packages (Figure 5).



⁴ <https://ospolicyobservatory.uvic.ca/2023/03/17/market-consolidation-and-scholarly-communications/>

Figure 5. Lariviere, et al. (2015) The Oligopoly of Academic Publishers in the Digital Era, *PLOS ONE*⁵

Scale became an essential component to success in the journals market⁶. Being part of the largest Big Deal packages extended a journal's reach and offered financial security. Further, by partnering with the largest publishers, smaller organizations received the benefits of scale through improved services and technologies that they couldn't afford on their own. For example, a small publisher might have one marketing person, but an international commercial publisher can offer access to a global network of hundreds of sales and marketing professionals. Furthermore, because the larger publishers purchase supplies and services in bulk, costs to produce a journal typically decrease for a society engaged in such a service partnership.

A study published in *PLOS ONE* in 2015 which analyzed papers published in 2013 describes an oligopoly – the top five publishers published more than 50% of papers (more than 70% in the social sciences)⁷. Those publishers were Elsevier, Wiley, Springer Nature, Taylor and Francis, and the American Chemical Society (ACS) in the sciences, and the same top four but with Sage Publications instead of ACS in the social sciences and humanities (Figure 6).

⁵ <https://doi.org/10.1371/journal.pone.0127502>

⁶ <https://scholarlykitchen.sspnet.org/2015/06/25/the-changing-nature-of-scale-in-stm-and-scholarly-publishing/>

⁷ <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502>

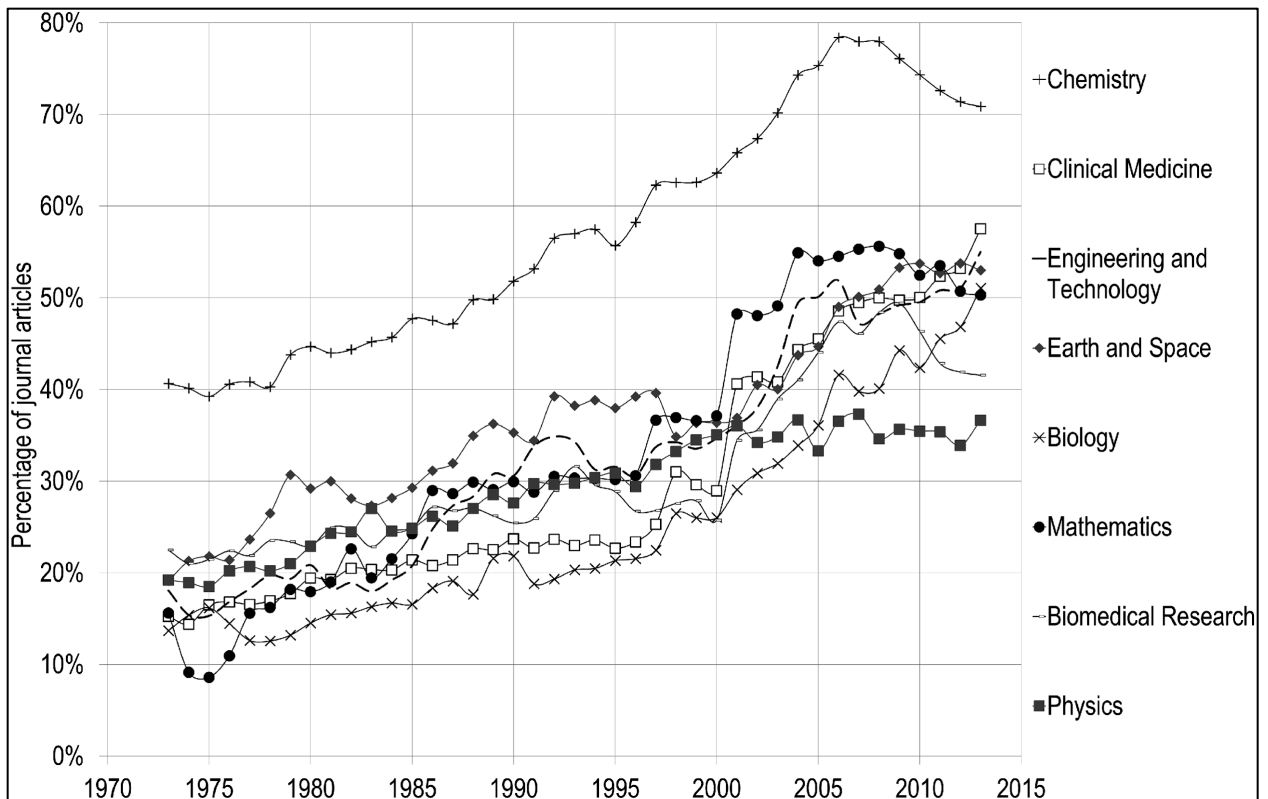


Figure 6. Percentage of papers published by the five major publishers, by discipline in the Natural and Medical Sciences, 1973–2013. Lariviere, et al. (2015) *The Oligopoly of Academic Publishers in the Digital Era*, *PLOS ONE*⁸

OPEN ACCESS

In parallel with the other ongoing shifts in the subscription journal market, the advent of open access (OA) to science and scholarship has been a major force driving change in academic publishing.

Open access (OA) is a set of principles and practices intended to result in the free and unencumbered distribution online of research publications. OA has a compelling rationale: to drive research progress by ensuring that everyone has access to published science and scholarship. But implementing OA in sustainable and equitable ways has long proved challenging.

⁸ <https://doi.org/10.1371/journal.pone.0127502>

In the U.S., the two dominant models for achieving open access in journal publishing are referred to as “Green” and “Gold”. In the Green OA model, the author deposits a version of the article on an institutional repository, preprint server, or funder platform such as PubMed Central (PMC), at no cost to author or reader. In compliance with publisher policies, the deposited work is typically the author’s final accepted manuscript (AAM) rather than the published version of record (VOR). This model functions in parallel with journal subscriptions, and hence does not necessarily relieve the financial pressure on institutional customers. However, to prevent the erosion of subscriptions, many publishers implemented embargoes that delay access to deposited versions for some period of time (typically 12 months). In addition, the Green OA model has raised concerns about “version confusion” in citation, although it is unlikely that this is a significant problem and the question has yet to be studied.

Gold OA publishing, on the other hand, results in the final version of the article being made freely available upon publication in the journal itself, typically after payment of an article processing charge (APC) by the author or their home institution. This eliminates most issues surrounding version control, but creates new challenges. BioMed Central, an independent organization before being sold to Springer Nature, introduced APCs in 2002. With the subsequent launch of *PLoS Biology* in 2003, the APC was cemented as the preferred business model for OA in many regions of the world for certain disciplines including the biomedical sciences. In the intervening years, several unintended consequences have become apparent.

Under Gold OA, the business model shifts from reader-centric to author-centric. Publishers are paid an APC for each article published, rather than relying solely on traditional subscription revenue. Thus, for-profit publishers adopting the APC-based Gold OA model have two basic ways to increase profits: publish more articles or cut costs. As a result, the Gold OA model has

created a perverse economic incentive for these publishers to increase the quantity of articles published within a given journal, potentially at the expense of quality. To reduce expenses, some OA journals do away with copyediting, for example, or minimize plagiarism and integrity checks. For a subscription journal, by contrast, the emphasis is on the subscriber's (institution and reader) perception of quality, hence the focus is on publishing only those articles that are likely to attract readers and increase the reputational value of the journal.

Until recently, OA was viewed by many publishers as an additive engine of growth rather than as a model to replace subscription revenues (Figure 7). Most subscription journals became "hybrid" journals that created a new revenue stream by offering an option for authors to pay for their papers to be published OA. But at the same time, fully Gold OA journals became increasingly attractive for both libraries and publishers because of the pressure on library budgets.

Gold OA journals could be funded through sources other than a library's collections budget, such as a university OA subvention fund or research grants. From the publisher perspective, OA journals could also become sustainable or profitable more quickly (assuming an adequate flow of article submissions) than a traditional subscription journal, which might take years to reach profitability through reputation building and the associated subscription growth.

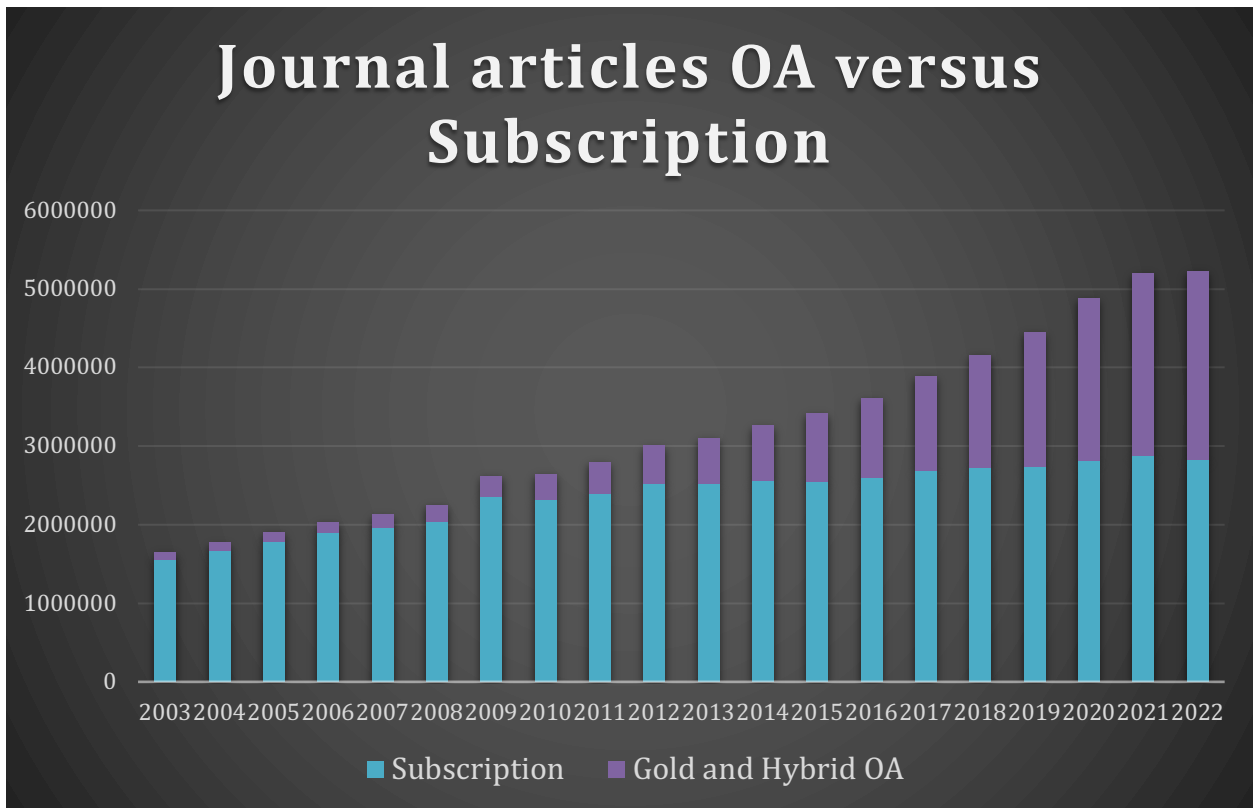


Figure 7. OA was largely additive until policies came into effect circa 2013. Data source: Dimensions (Digital Science). Search limited to “articles” AND “Gold OA” AND “Hybrid OA”, OR excluding “Gold OA” AND “Hybrid OA”.

Gold OA renders publishers’ revenues dependent upon a relatively small number of authors, rather than on a large number of readers. This creates an incentive not only to publish more articles but also to make APCs as high as possible. Journals with high APCs create equity issues, since not all authors can afford them, which is at odds with the OA goal of increasing equity. Further study of the impact of APCs on journal publishing choice is needed to determine if the shift to Gold OA is creating unintended equity challenges. If authors who are well funded or at wealthier institutions have better publishing options than researchers with less, or no, funding, then this publishing model fails to align with the principles of democratized access and participation.

Gold OA originally took hold in scientific fields in which research is generally well funded, but it has grown to encompass the humanities and social sciences

where funding levels are typically much lower. This disparity puts entire fields at a disadvantage. Gold OA further drives inequity by privileging researchers in wealthier countries and at wealthier institutions over those for whom paying an APC is impossible. Publishers' APC waiver programs do not satisfactorily address this fundamental inequity.

It is worth acknowledging that this exacerbates disparities that already exist in the research system. Virtually all research benefits from better grant funding, more and better equipment such as high-quality microscopes and powerful computers, more staff, and more experts hired to collect and analyze the data. In effect, wealthier countries and researchers with grants have always had a significant research advantage.

Gold OA can also be problematic at the institutional level. It is sometimes argued that there is currently enough money in the research system to support flipping a traditional subscription journal to Gold OA⁹. However, that money is not readily transferable. If the literature is made openly available, then costs for institutions that publish less but still access the literature go down, as costs are shifted to institutions that publish more papers. For example, Dimensions data show some 8,337 articles in 2022 listing an author affiliated with the Massachusetts Institute of Technology (MIT). If MIT was responsible for paying a \$3,000 APC for every paper with an MIT author in 2022, the cost would be over \$25 million (Figure 8). For Harvard, with 32,714 affiliated papers in 2022, the cost would be nearly \$97 million. Even if only a subset of these affiliated authors are responsible for paying the article's APC (which typically falls to the corresponding author), it would still represent a significant cost increase for the institution. Further study of how much individual universities would pay

⁹ https://pure.mpg.de/pubman/faces/ViewItemOverviewPage.jsp?itemId=item_2148961

annually for APCs and how that compares to current subscription spending is needed.

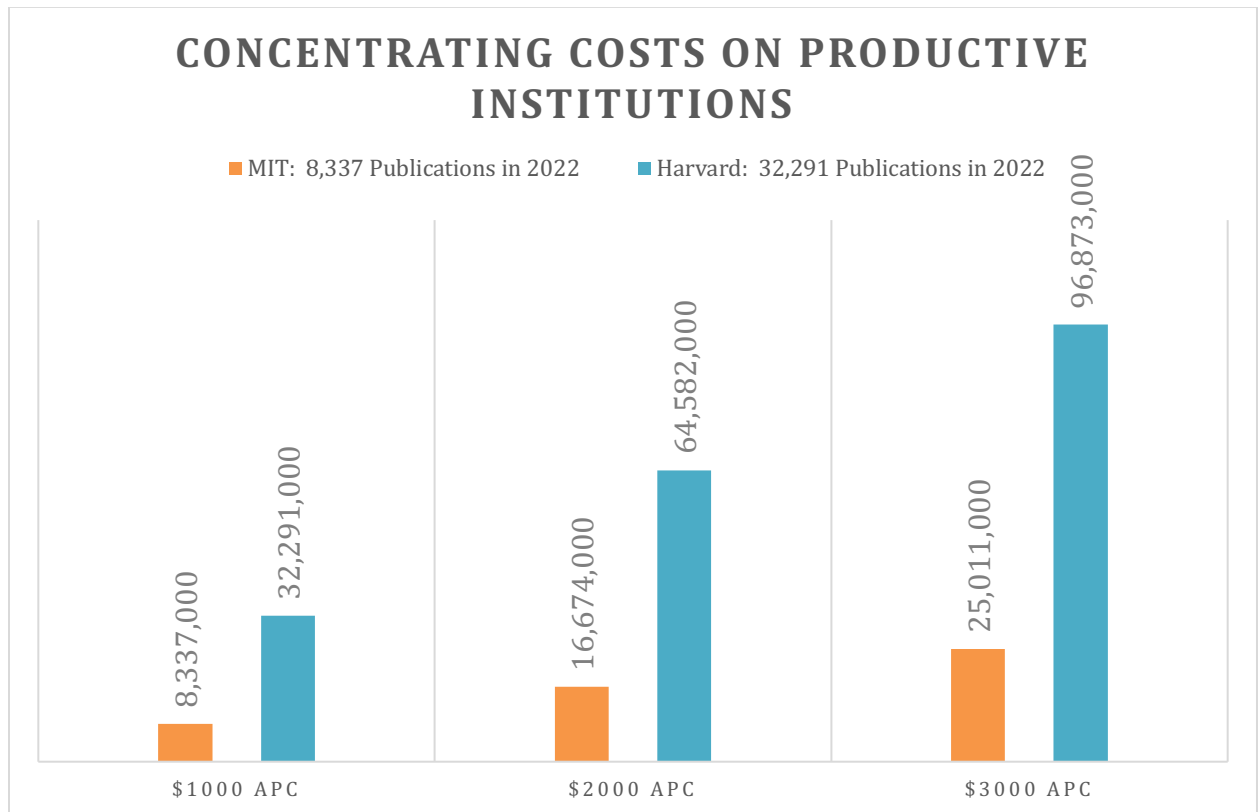


Figure 8. Productive institutions pay more under Gold OA. Data source: Dimensions (Digital Science). Search limited to “articles”.

As Gold OA has grown over the past two decades, it has increased the incentives for profit-oriented publishers to consolidate and to publish as many articles as possible, regardless of quality. Under Gold OA, the more articles a journal publishes, the more money it makes.

This has led to three new market trends. The first is the megajournal — first seen in *PLOS ONE*, but now surpassed by Springer Nature’s *Scientific Reports*. *Scientific Reports* is now the largest journal in the world, publishing more than 22,000 papers last year (Figure 9). These “sound science” journals review papers solely on the basis of accuracy — i.e., did the authors do what they said they did and are their conclusions supported by the data presented? No

judgment is made concerning the significance or novelty of the work. One positive impact of megajournals is that they surface a huge amount of incremental work that might otherwise not be made public. The resulting journals are, however, largely undifferentiated containers of content without any notable signals of the relative contribution papers make to a field.

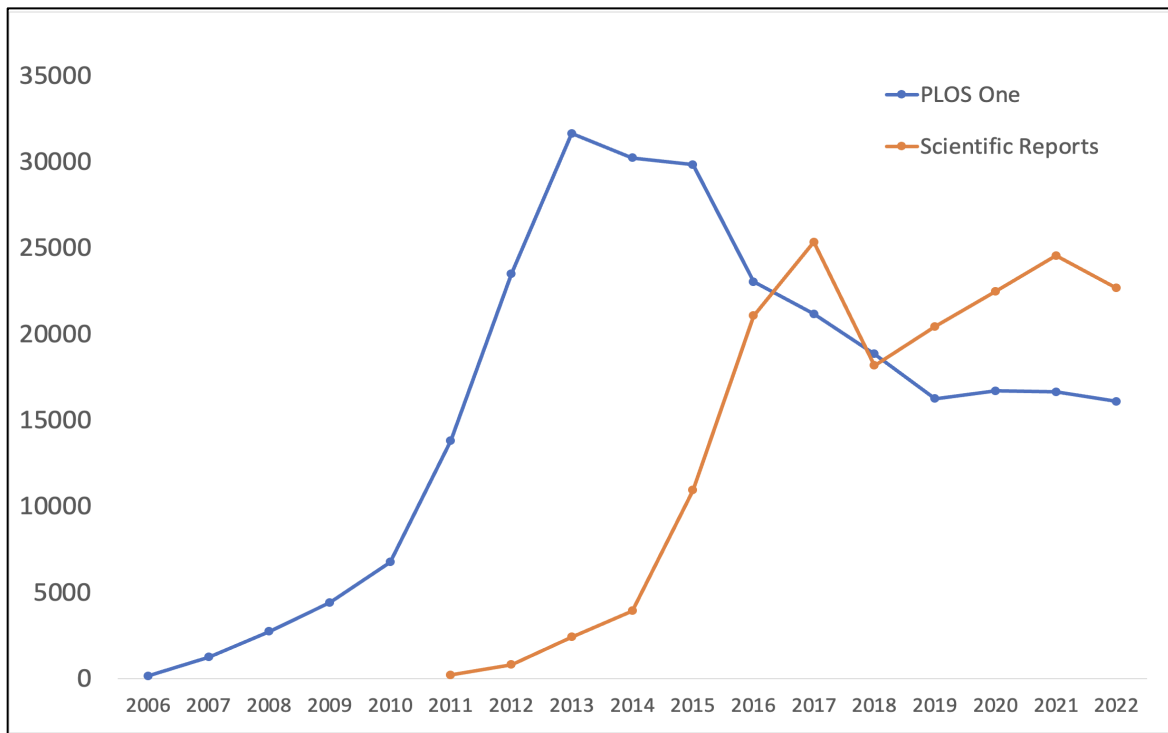


Figure 9. Megajournal publication volumes. Data source: Dimensions (Digital Science). Search limited to “articles”.

A second trend is the cascade or transfer model. Publishers incur costs for every paper that is submitted to their journal, but under the Gold OA model they are only paid APCs for the papers that are accepted and published. If the top journal in a given field rejects 95% of submissions, then authors of the accepted 5% of papers effectively have to bear the processing and review costs of those rejections, as well as the costs that their own papers incur. This fundamentally changes the economic position of flagship journals like *Nature* or *Cell*. Under a pure subscription model, flagship journals are high earners because so many people want to read the small number of high-quality articles

they publish. Under a Gold OA model they are far less valuable, even if the publisher charges a high APC, because they are expensive to run but publish relatively few articles.

This has led to the transfer strategy, in which the goal is to ensure that once a manuscript is submitted to one journal in a publisher's portfolio, it stays in that portfolio. This makes it more likely that the publisher will get paid an APC to recoup some of the costs of reviewing the article. The paper can be published in one of the portfolio's other titles after being rejected from the author's (more selective) target journal. In other words, the manuscript "stays in the family" of journals rather than leaving it to go to another publisher. This also provides some convenience to authors, who benefit from not having to resubmit their publication to a new journal; if the paper was reviewed, those reviews typically transfer with the paper, so that authors of rejected papers don't have to start over with a completely new set of reviewers. This, in turn, can reduce the time to publication.

In the Nature Publishing Group, for example, *Nature* sits at the top of the cascade, and rejected papers work their way down through Nature's subject-specific titles such as *Nature Medicine*, *Nature Genetics*, etc., until they reach *Nature Communications*, which serves as a catch-all for high-level research that does not make it into their other flagship publications. If it is not a good fit there, then the paper moves down to *Scientific Reports*, which collects all the rest that pass sound science review. Springer Nature is by no means alone in this strategy. It is being used by nearly every major journal portfolio (e.g., *Cell*, *The Lancet*, *JAMA*, etc.).

In response, some societies have followed the same strategy and launched additional journals to capture articles rejected from their flagship publications. This has allowed them to increase the quantity of articles they publish without

compromising the rigor and selectivity of their top journals. Anecdotally, however, many societies and smaller publishers report seeing significant submission declines at their subject-specific journals. Papers they may have received in the past following rejection from top-tier journals like *Nature* are now staying in the Springer Nature family of journals. There is no quantitative data on the scale of the cascade trend or its impacts on publishers that cannot take advantage of the model, because most journals consider submission data to be proprietary. How these shifts are impacting publishing trends, particularly the revenues of societies and other small presses, merits further investigation.

A third key trend is the creation of new publishers that are entirely OA. Without legacy publications to transition, it is easier for these publishers to optimize their practices to meet new market conditions. The two fastest growing OA publishers of this kind are MDPI and Frontiers. MDPI is now the third largest scientific journal publisher and Frontiers the sixth largest (Figure 10).

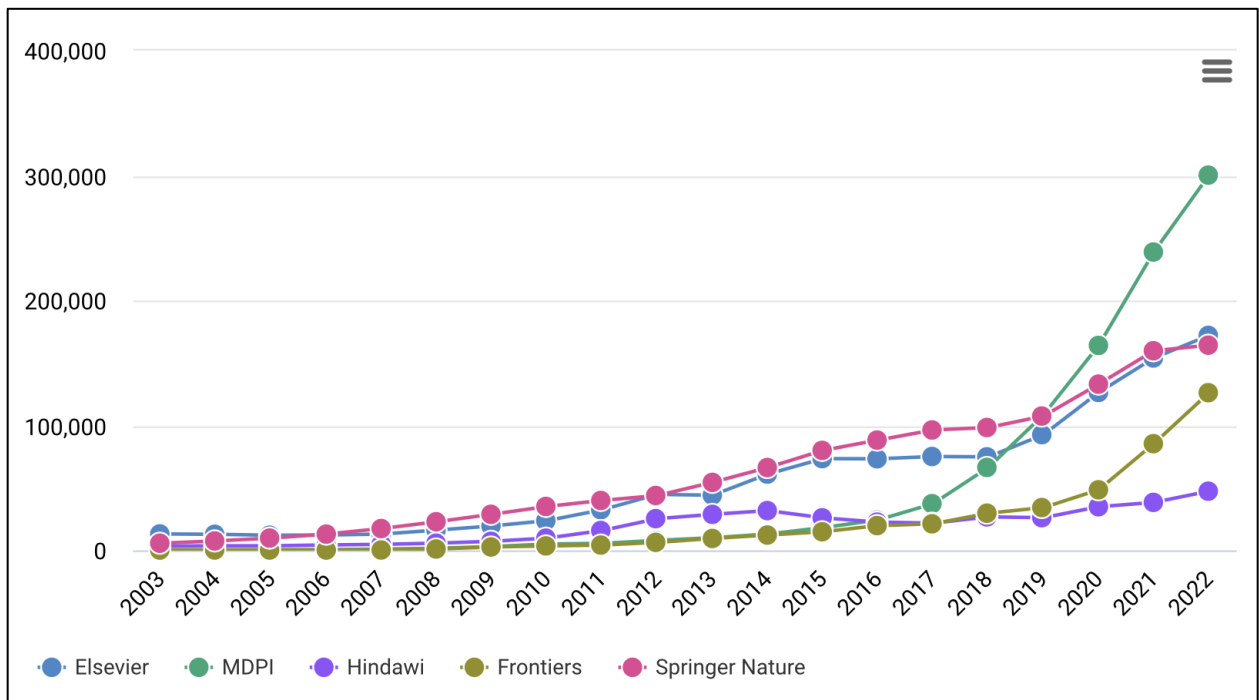


Figure 10. MDPI and Frontiers are the fastest growing OA publishers. Data source: Dimensions (Digital Science). Search limited to “articles” and “Gold OA” or “Hybrid OA”.

MDPI and Frontiers are highly efficient organizations that are optimized for article quantity. They employ large in-house staff to handle many of the tasks traditionally managed by academics serving as journal editors. This results in faster and more consistent turnaround times, which can help attract authors. Both publishers put enormous effort into recruiting authors, to the point that their “spamming” practices have damaged their reputations among researchers. Both have seen their share of controversy over editorial decision-making processes. A recent article noted the use of the term “MDPI Professors” (El Pais) to describe researchers whose CVs were populated with large quantities of such articles¹⁰.

MDPI and Frontiers drive growth of articles and revenue by putting out subject-focused “special issues.” (Figure 11). Huge numbers of individual researchers are recruited to guest-edit special issues. These guest editors are expected to recruit other authors to contribute to the special issue (each paying an APC for publication of their article). The numbers of special issues are staggering. MDPI is on track to produce 56,000 special issues in 2023, with some journals publishing 10 special issues per day, every day of the year.

¹⁰ <https://english.elpais.com/science-tech/2023-06-04/a-researcher-who-publishes-a-study-every-two-days-reveals-the-darker-side-of-science.html>

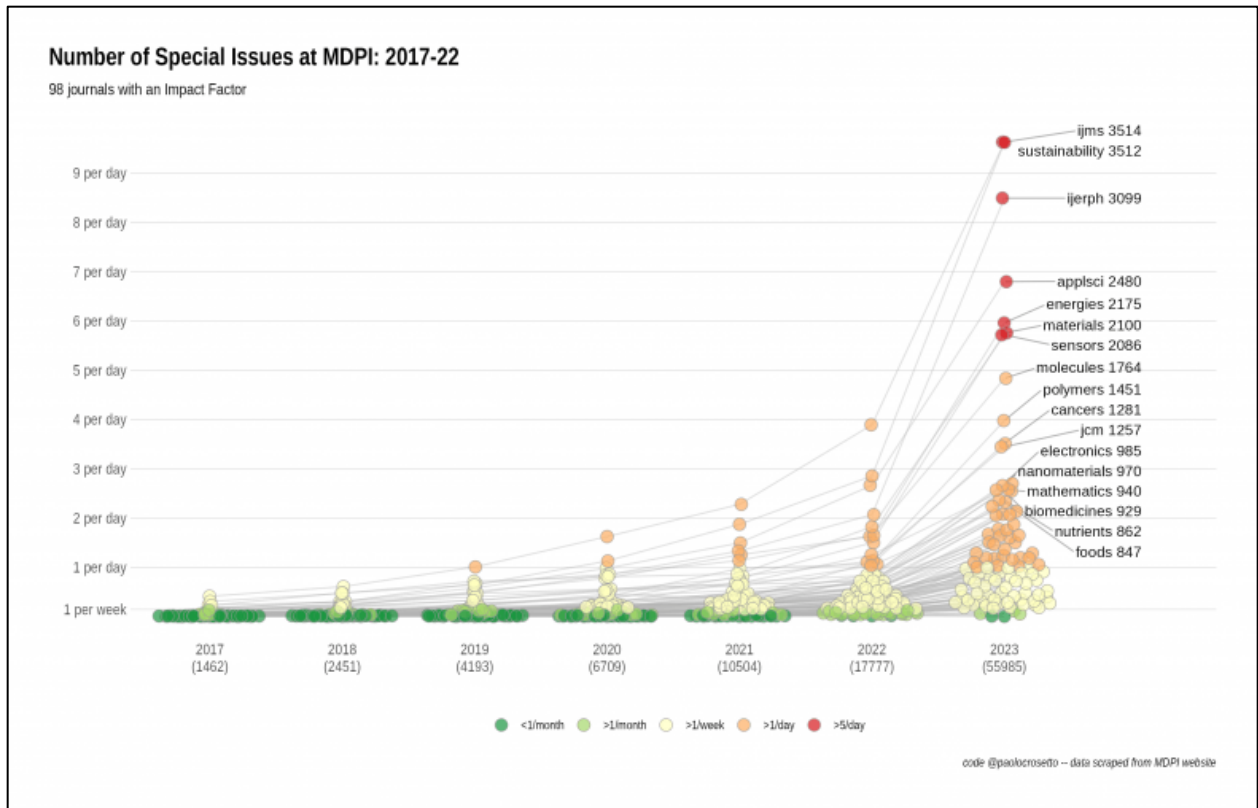


Figure 11. MDPI Special Issues. Source: Paolo Crosetti¹¹

A recent cautionary tale involves Hindawi, a fully OA publisher purchased by Wiley in 2021. According to Retraction Watch, much of what Hindawi was publishing was “fraudulent” — essentially, fake papers being sold to authors by “paper mills” for publication in special issues that had very little quality control^{12,13}. Wiley is in the process of retracting at least 1,700 articles and shutting down the corresponding journals¹⁴. Web of Science delisted 19 Hindawi

¹¹ <https://www.timeshighereducation.com/news/quality-questions-publishers-growth-challenges-big-players>

¹² <https://retractionwatch.com/2022/09/28/exclusive-hindawi-and-wiley-to-retract-over-500-papers-linked-to-peer-review-rings/>

¹³ https://en.wikipedia.org/wiki/Research_paper_mill

¹⁴ <https://scholarlykitchen.sspnet.org/2023/04/04/guest-post-addressing-paper-mills-and-a-way-forward-for-journal-security/>

journals from having Impact Factors, representing 50% of Hindawi’s published articles in 2022¹⁵.

The strategy employed by MDPI and Frontiers has been incredibly successful from a financial perspective. A comparison of the top 10 publishers by volume of articles published in 2012 versus 2022 (Figure 12) shows MDPI and Frontiers becoming two of the largest scholarly publishers (numbers 3 and 6, respectively).

Publisher	Publication Volume 2012	Publisher	Publication Volume 2022
Elsevier	514,157	Elsevier	782,636
Springer Nature	269,367	Springer Nature	448,643
Wiley	214,201	MDPI	300,366
Taylor & Francis	130,474	Wiley	243,955
Wolters Kluwer	81,276	Taylor & Francis	147,043
SAGE Publications	64,674	Frontiers	126,407
Oxford University Press	58,719	Wolters Kluwer	98,842
Trans Tech Publications	55,506	Oxford University Press	91,756
ACS	41,860	CABI Publishing	83,804
IEEE	37,499	SAGE Publications	79,310

Figure 12. Largest scholarly publishers by volume, 2012 and 2022. Data source: Dimensions (Digital Science; search limited to “articles”).

RESEARCH FUNDER POLICIES

In 2008, the US National Institutes of Health (NIH) started requiring the deposit of articles resulting from funded research in PMC, to be made available within 12 months of publication. At the time, publisher groups raised concerns about version control, pirating of papers, and undermining the sanctity of the editorial process, but they nonetheless began to deposit articles in PMC on behalf of funded authors as a service. There was little to no measurable impact on journal subscriptions as a result of the 2008 NIH policy, although it is difficult to

¹⁵ <https://www.science.org/content/article/fast-growing-open-access-journals-stripped-coveted-impact-factors>

measure, and the free availability of articles, even with delayed availability, may play a role in enabling the aforementioned Big Deal cancellations.

2013 saw two landmark policies: the Research Councils UK (RCUK) OA policy and the US White House Office of Science and Technology Policy (OSTP) Holdren Memo. RCUK was, at the time, the UK's main research funder (now subsumed by the UKRI). The RCUK policy required all articles supported by their funding to be published OA, with a strong preference for Gold OA routes to compliance. If the journal of choice did not offer Gold OA, or if the researcher had no funds for it, Green OA routes with a delay were acceptable. Rather than using research grant money to pay APCs, the RCUK made block grants to universities to cover these costs. The policy was largely successful in driving strong growth in OA in the UK, but it was also far more expensive than expected, particularly on the administrative side (e.g., managing payments and monitoring compliance)¹⁶. Because the UK took this action in isolation, UK universities had to continue to pay for subscription content from around the world. There was no offsetting of the additional costs.

In the US, the 2013 Holdren Memo extended the earlier NIH public access policy across all US federal science funders that spent more than \$100 million per year on research funding. The policy called for the same 12-month embargo on access with no reuse rights required. (This is a public access policy as defined in the glossary in Appendix B, not an open access policy.) As with the initial NIH policy, publisher groups pushed back but ultimately cooperated with the federal agencies. Many offered automatic deposit on behalf of authors, increasing compliance. The Holdren Memo policy does not appear to have

¹⁶ <http://www.research-consulting.com/new-report-highlights-9m-compliance-cost-of-uk-open-access-requirements/>

impacted journal publisher revenues, while public access to research content was significantly enhanced.

PLAN S

In 2018, an international consortium of research funding and performing organizations called “cOAlition S” launched Plan S. Plan S only covers about 5% of the scholarly literature, and yet it has had an outsized effect on the market as a major tipping point in the shift to OA¹⁷. Although it has been revised over time, Plan S essentially requires all research funded by coalition members to be published OA, and only in fully-OA journals. Research funding cannot be used to pay for OA in hybrid journals. Exceptions to this rule were added over time, largely in response to concerns that researchers would be shut out from publishing in the venue of their choice. Hybrid journals are now permissible, at least through 2024, but only within the “Transformative Agreement” framework described below.

Transformative Agreements (TAs) marked a significant shift in open publishing business models and the journals market (Figure 13). Under a TA — a “bigger Big Deal” — institutions pay not only for access to a publisher’s full portfolio of journals, but also pay to cover all the costs for authors at the institution to publish OA in that publisher’s journals. For this reason, TAs are sometimes referred to as “Read and Publish” or “Publish and Read” agreements.

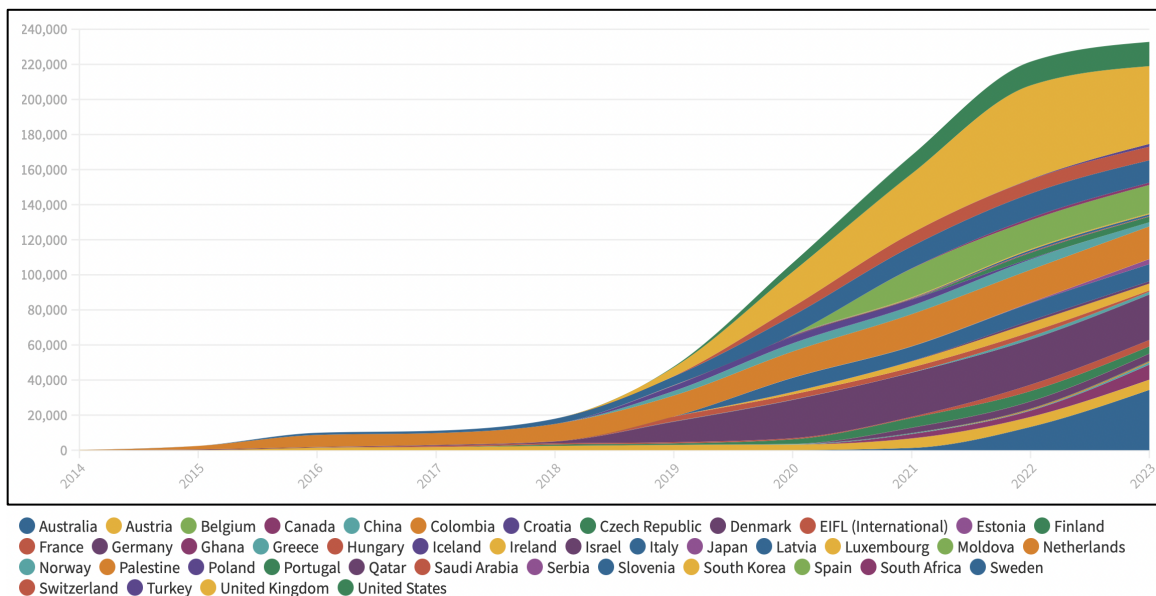
It requires a great deal of data collection and analysis to put together a TA. The deal parameters are time consuming to negotiate, meaning only publishers with sufficient resources can pursue them. TAs are also generally limited to larger publishers because they only make economic sense for the publisher and the institution if there will be a large number of readers and authors. If

¹⁷ <https://deltathink.com/news-views-plan-s-effects-2021-part-1-article-volumes/>

researchers from an institution publish only a handful of papers annually in a publisher's journals, it is hard to justify the work involved in negotiating a TA.

The rise of TAs led to a further wave of market consolidation. Many scientific societies entered into new partnerships with larger publishers to avoid being left out. This wave of market consolidation has not been thoroughly studied; data is needed to understand the impact on smaller, independent, and scientific society publishers. cOAlition S is now retrenching, and seeking alternatives to pay-to-publish models in achieving the broader goals of open access¹⁸.

It is also worth noting that TAs likely create an incentive for researchers to choose to publish in paid open access journals that their libraries subsidize over those their libraries do not subsidize. These non-subsidized outlets include most journals independently published by scholarly societies and university presses. Data on the scale of this phenomenon is not publicly available, but would elucidate the impact that TAs have had on publishing choice.



¹⁸ <https://www.researchprofessionalnews.com/rr-news-europe-infrastructure-2023-6-alternatives-to-dysfunctional-open-access-model-sought/>

¹⁹ <https://esac-initiative.org/about/transformative-agreements/agreement-registry/>

THE NELSON MEMO

The “Nelson Memo”, named for then-Acting OSTP Director Alondra Nelson, was released in August 2022 with the aim of enabling free, immediate, and equitable access to federally funded research²⁰. The Memo directs federal agencies to remove the 12-month embargo for public access to federally funded research papers and, further, mandates immediate public release of data collected using federal funding. The Nelson Memo requires federal agencies to have new policies in effect by January of 2026. It also extends the Holdren Memo public access requirements to all federal agencies, not just science agencies that spend more than \$100 million on funding annually.

Specifically, the Nelson Memo requires that the author’s accepted manuscript (AAM) version or, if possible, the published version of record (VOR) of any paper listing federal funding be deposited in an agency-designated repository and made publicly available immediately upon publication. Again, this is “public access,” and no specific licensing requirements, such as the use of a Creative Commons license, are required. As with the Holdren Memo, no additional funding has been provided to support the policy, and federal agencies have not been provided with formal guidance on how to cover the costs of Gold OA publishing from their research budgets.

As a result, Gold OA publication costs for researchers are presumably going to be paid by their universities or come directly out of their research grants. Without additional publishing subsidies, a laboratory that publishes 10-20 papers per year under this model could lose the equivalent of a postdoc’s salary, or a significant amount of equipment or reagents; whereas an alternative to such subsidies could be lower-priced publication. With funding diverted from performing experiments to paying for publication, it is possible

²⁰ <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf>

that less research will be done in exchange for more access to the results of federally funded research. It remains an open question whether that reduction will be balanced out by research progress enabled by immediate public access to papers or other benefits of open access. (The same trade-off would exist if agencies fund Gold OA costs without having their budgets increased accordingly.)

Independent analyst Christos Petrou estimates that the Nelson Memo will cover around 200,000 papers per year, about a third of total output from US researchers²¹. However, those 200,000 articles are not evenly distributed. Some journals and fields will see little impact (e.g., a mathematics journal where federal funding is scarce, versus an oncology journal where the majority of papers carry some level of NIH funding). Because grant funding is increasingly competitive and tends to fund high-quality research from top laboratories, the Nelson Memo could impact researchers publishing in high-end, selective journals more strongly.

While the Nelson Memo allows compliance through Green OA, it is not yet clear how significant a role that model will play in researchers' and publishers' response to the Nelson Memo. Compliant Green OA articles would be made available simultaneously in subscription journals and for free online in agency (or other) repositories. Publishers that do not trust subscriptions to hold up under such circumstances will likely attempt to promote Gold OA instead. Springer Nature, for example, has stated that all federally funded authors will be required to pay an APC to publish in their journals.²² Only a few journals, particularly those that publish significant amounts of non-research paper content (e.g., *Science*) or those more reliant on licensing rights and advertising

²¹ <https://scholarlykitchen.sspnet.org/2022/09/13/guest-post-quantifying-the-impact-of-the-ostp-policy/>

for revenues (e.g., *Journal of Clinical Oncology*), have stated that they prefer the Green OA model.

For now, we exist in a liminal space²². While the research community in principle embraces a future of increasing open access to articles, data, code, and other research products, how current policies will be implemented remains to be seen. Without further policy guidance, or interventions to support future diversity in the publishing ecosystem, we are likely to see continued rapid growth of publishing models that emphasize quantity and efficiency over quality control. This may be good for some publishers, but it is likely not what's best for science or the broader research community.

For the moment, journal publishing is fragmented. OA is growing, but does not represent the whole market; there are still entire fields and parts of the world that are likely to stay with the subscription model for the time being. China, given that country's large and growing research output, would have to increase spending enormously if their authors were required to publish under Gold OA²³. Subscriptions are not going away on January 1, 2026 when the Nelson Memo goes into effect. Over time, it will be important to track what proportion of a journal's articles has to be OA, on average, before subscription sales decline, particularly from libraries, which make up the majority of subscription revenue.

The largest commercial publishers are committed to APC-based Gold OA models and are now facing competition from large born-OA publishers in terms of submissions. Mid-sized publishers such as Wolters Kluwer and Oxford University Press (the world's largest university press) are in growth mode,

²² <https://scholarlykitchen.sspnet.org/2022/01/13/life-in-a-liminal-space-or-the-journey-shapes-the-destination/>

²³ <https://www.ce-strategy.com/2023/06/an-interview-with-china-stm-publishing-expert-nicko-goncharoff/>

launching new journals and signing partnerships with scientific society journals in an effort to achieve the scale necessary to survive an OA market. Smaller and independent publishers are in crisis, with many signing agreements to outsource their publishing operations to larger commercial publishers or facing an uncertain future. The volatile market has made larger publishers more risk-averse, making the terms offered to society journals less favorable (for example, with respect to revenue sharing). Some societies and researcher editors are abandoning commercially-owned journals and moving to smaller non-profit publishers due to the high APCs put in place by large publishing houses and the requirement to publish more papers. For example, the entire editorial board of a top neuroscience journal published by Elsevier resigned in April of 2023 to start a new journal with the MIT Press²⁴.

There is indeed growing researcher awareness of, and unhappiness with, the unintended consequences of the author-pays APC model for OA. An earlier example of this trend is the journal *Quantitative Science Studies*, which launched at the MIT Press in 2019 after the editorial board of *Journal of Informetrics* (Elsevier) resigned to protest high APCs and restrictive policies related to the free distribution of abstracts and reference lists²⁵. Despite this upheaval and the absence of an impact factor at the outset, the new non-profit journal's submissions and reputation grew quickly, and it is now ranked #1 out of 77 journals in the 2022 Emerging Sources Citation Index (ESCI) Information & Library Science category with an impact factor of 6.4.

RESEARCH DATA

The new OSTP policy also requires federal funding agencies to develop policies obligating researchers to make the data underlying their published findings

²⁴ <https://www.nature.com/articles/d41586-023-01391-5>

²⁵ <https://direct.mit.edu/qss>

"freely available and publicly accessible by default at the time of publication."²⁶

The widespread availability of open, reusable data democratizes research by decoupling the ability to access and analyze research data from having the resources to collect that data. It also promotes trust and transparency in research.

This shift in policy will require universities and individual researchers to share the data related to research publications on or before publication. How, where, and at what cost are not yet clear. There are also situations in which the sharing of data publicly is not possible, for example, for federal security and due to patient privacy concerns. Hence, the policy leaves open questions about how federally funded researchers will comply, and how universities should plan and budget accordingly.

Note: we acknowledge that research data is a huge and multifaceted topic. This section of the report is short because the group did not bring in outside expertise on the topic of research data during the course of our discussions. However, the group did raise several questions related to data sharing that are included below.

VERTICAL INTEGRATION IN THE INFRASTRUCTURE OF RESEARCH COMMUNICATION

Scholarly and scientific communication relies on tools, standards, and platforms for hosting, editorial management, analytics, and more. Some of this infrastructure is open-source or community owned, but more of the highly utilized technology is proprietary. Mergers and acquisitions over the last several years have resulted in a handful of large publishers owning much of the relevant infrastructure.

As open access publishing grows, major commercial publishers have expanded into associated areas, largely by acquiring scholarly infrastructure, services, and data analytics. Many of the core tools used to publish and access research

²⁶ <https://www.whitehouse.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-access-Memo.pdf>

results have, consequently, gone from being independently owned to being part of commercial publishing houses and technology companies. In their 2019 book chapter *Vertical Integration in Academic Publishing: Implications for Knowledge Inequality*, Chen, Posada and Chan report on a detailed analysis of the mergers and acquisitions of three of the five major academic publishing companies.

Elsevier, Wiley, and Taylor and Francis each have a long history of acquiring other publishers and established journals. More recently, these companies have made significant acquisitions of tools and services that function across the knowledge production lifecycle. For example, Aries, and their Editorial Manager submission and peer review system, is now owned by Elsevier. Wiley now owns the Atypon platform, host to over 100,000 publications, along with J&J editorial services and most recently, the eJournal Press submission and peer review system. Wiley has stated that nearly half of the world's peer-reviewed research at some point goes through Wiley-owned platforms²⁷.

Figure 14 depicts the results of Chen et al.'s analysis of the various Elsevier academic services and how they influence institutions and individuals' decision making.

²⁷ <https://edscoop.com/wiley-jay-flynn-open-access-research/>

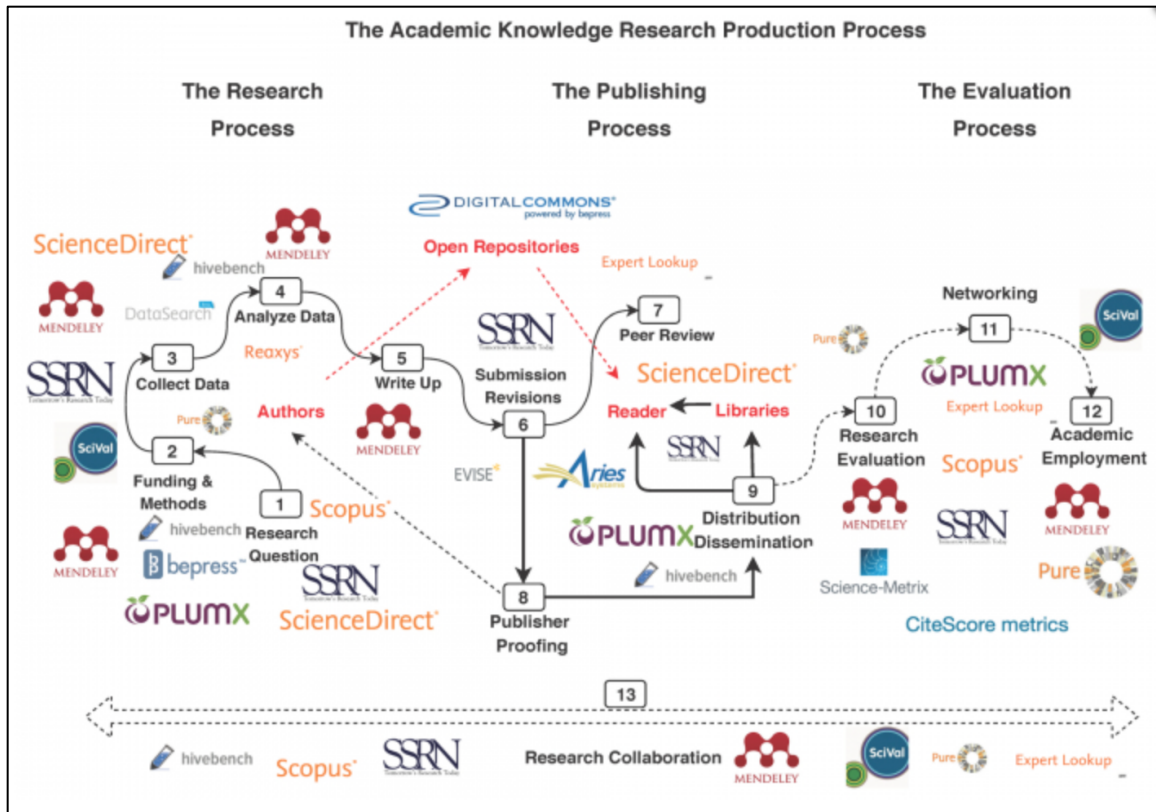


Figure 14. Elsevier presence throughout the research lifecycle. Source: Figure 5 in CHEN, George ; POSADA, Alejandro ; et CHAN, Leslie. *Vertical Integration in Academic Publishing: Implications for Knowledge Inequality*. In: *Connecting the Knowledge Commons – From Projects to Sustainable Infrastructure : The 22nd International Conference on Electronic Publishing – Revised Selected Papers* Marseille : OpenEdition Press, 2019.²⁸

In 2022, Elsevier acquired Interfolio, a system that enables academics to collect and manage critical data for academic hiring, review, promotion and tenure. As Aspesi and Brand argued, the growing ownership of data analytics, hosting, and portal services by large scholarly publishers may enhance publishers' ability to lock in institutional customers through combined offerings that condition (or discount) one product or service upon purchase of another²⁹.

²⁸ <https://books.openedition.org/oep/9068>

²⁹ <https://www.science.org/doi/10.1126/science.aba3763>

THE IMPERATIVE FOR EXPERIMENTATION

Some smaller not-for-profit publishers are in a state of rapid experimentation, developing new models. These are works in progress, looking for solutions to compensate for the negative consequences of the APC model. Small non-profits are well placed to take on this experimentation because they are at risk in the market, are eligible for grant funding to pursue innovative ideas, and because they do not have the pressure of constantly needing to meet shareholder expectations.

Many of these new models are risky because they rely on the goodwill of, and voluntary spending from, the library community. For example, *Subscribe to Open* is a model wherein the publisher sets a threshold of subscription levels for a journal (or a collection of journals) and, if enough libraries subscribe, the journal (or collection) is made OA for that year.³⁰

Other models rely on community approaches, like the Open Library of the Humanities, which solicits donations from supporter libraries each year. One concern about these types of models is that there is no penalty imposed on free riders. If you stop paying, it's likely that you will still have full access to the journal as long as it endures, so why pay for something that is otherwise free? If a library has to make a budget cut, it is easier to stop spending on something that they will retain access to anyway. It remains to be seen whether this will happen in practice, and many of these models hedge against the free rider problem by providing additional benefits to participating libraries.

Other models include the University of California's "multi-payer" model, in which the university pays a portion of the APC and requires researchers to cover the rest out of grants, if they have them. Right now, participation is

³⁰ <https://subscribetoopencommunity.org/>

voluntary — a funded researcher can choose not to contribute³¹. It is unclear how well the model will be accepted if authors are required to pay part of the fee.

PLoS has introduced a Community Action Publishing model that has been described as “subscribe-to-publish.”³² It spreads the costs of publication out over all authors on a paper, rather than asking for one (large) fee from the corresponding or lead author. In addition, the number of papers published at an institution in the preceding three years is averaged, and an annual charge to the institution is determined based on a given institution’s publishing rate. The downsides of this model include the enormous amount of ongoing data crunching, and the year-to-year variance. So too, the model does not support authors who are not affiliated with institutions under one of these deals (or not affiliated with any institution at all). Hence, this model may not be viable for a publisher with few journals or a low article volume, as it could prove difficult to engage libraries in the effort required.

DIAMOND OPEN ACCESS

Under Diamond OA, access is free for readers *and* publishing remains free for authors. Diamond OA relies on a combination of subsidy (e.g., through an endowment, direct university or society support, or grant funding), and unpaid volunteer time. This model can work, up to a certain volume. If a journal has a small number of submissions, an academic editor can reasonably be expected to manage the review process as well as publication. But as a journal grows, the workload may become too significant to be handled by volunteers alone.

³¹ <https://osc.universityofcalifornia.edu/uc-publisher-relationships/negotiating-with-scholarly-journal-publishers-a-toolkit/an-introductory-guide-to-the-uc-model-transformative-agreement/>

³² <https://plos.org/resources/community-action-publishing/>

A few society journals have adopted Diamond OA, notably among them *Cultural Anthropology*, the prestige journal of the Society for Cultural Anthropology (SCA), a section of the American Anthropological Association.³³ Diamond OA since 2014, *Cultural Anthropology* is currently published on an instance of Open Journal Systems, a low-cost journal hosting platform.³⁴ The journal's modest budget includes salary and benefits for two half-time editorial positions, with costs shared by the SCA and in-kind contributions from the individuals' home institutions. To support its share of the budget, the SCA recently ran a successful campaign to secure funding from LYRISIS's Open Access Community Investment Program³⁵, for an initial five-year commitment. The journal's financial future remains uncertain but nonetheless demonstrates a viable non-profit alternative to the Gold APC model.

Diamond OA has worked on a larger scale in some regions, including South and Central America where national funders have built cooperative models such as SciELO.³⁶ These large-scale, multimillion-dollar models have promise, but may be easier to build in areas that do not have established publishing infrastructure and expectations already in place. One question about this model concerns innovation: in a cooperative model with no competing journal publishers, there may be less incentive to invest in improvements. Many of the new tools valued by research funders and researchers (e.g. persistent identifiers, and paper mill and image manipulation detection tools) have been instigated by large publishers. If competition is removed or reduced, how will that impact the drive towards ongoing improvements in publishing infrastructure, in service to the research ecosystem, and how will those technological advances be funded?

³³ <https://journal.culanth.org/index.php/ca>

³⁴ <https://pkp.sfu.ca/software/ojs/>

³⁵ <https://www.lyrasis.org/content/Pages/oacip.aspx>

³⁶ <https://scielo.org/>

As the transition to open access continues, more experimentation and study is needed, particularly to monitor the impact of different models on researchers and the research enterprise. There is no one-size-fits-all model in academic publishing. Different fields of study have different cultures, publishing norms, and funding models. Hence, the future of research publishing and open access will continue to involve a variety of models.

A publishing ecosystem dominated by a few major commercial publishers would reduce the diversity of publishing options and leave universities and researchers without choices on pricing and publishing approaches. It would also leave less well-funded or unfunded researchers from the U.S. and abroad increasingly disadvantaged compared to well-funded researchers. Many in the research community remain unaware of the drivers of change in academic publishing and the potential consequences for the research ecosystem, as well as for society publishers (many of whom have relied on journal subscription revenue to subsidize other society activities) and university budgets. There is now a need for an in-depth, evidence-based examination of key questions concerning the future of quality research and its availability.

PART 2 – OPEN QUESTIONS ON THE FUTURE OF ACADEMIC PUBLISHING IN SIX KEY AREAS, WHERE MORE RESEARCH AND DATA ARE NEEDED

The health of the research enterprise is closely tied to academic publishing practices and systems. As detailed above, policy-driven changes in publishing business models over the last two decades have triggered a number of disruptions in this ecosystem that raise concerns about their impact on the research enterprise. Because different research fields have differing publishing practices and funding mechanisms, it is critical to the health of the overall research enterprise that stakeholders help construct and sustain a publishing ecosystem that encompasses a variety of access and business models (Green OA, Gold OA, Diamond OA, pre-print publishing, etc.) and a diversity of publishers and knowledge infrastructures (by which we mean, systems and technologies that support researchers in the generation, sharing, and maintenance of scientific knowledge).

In the discussions that led to the drafting of this report, we focused on defining key questions for investigation to understand, measure, and best prepare for the impact of open access in research publishing, as well as broader open science practices such preprint publishing, open peer review, open software, and open protocols. These questions are intended to spur research that will help inform future policies and their implementation. We raise questions in six areas:

- Access and business models
- Research data
- Preprint publishing
- Peer review
- Costs to researchers and universities
- Infrastructure

Below we include the five key questions in each area that the group considered to be high priority. Appendix A contains additional questions that our conversations generated.

ACCESS AND BUSINESS MODELS

- 1. How do we assess whether open access policies are measurably improving equitable access and the overall research enterprise?** The Nelson memo is designed to increase equity of access. How will equitable access be defined and measured over time, to determine if the policy change has had the desired impact in terms of access to scholarship? Is open access changing who is conducting, publishing, or responding to research? Is there any indication of different research questions/directions or any other change in research because of open access? Does access equate to impact and beneficial use by society, and how can that be quantified? Does “equitable access” apply to authors as well as readers of the literature?
- 2. What new metrics or indicators should be developed to assess academic reputation?** Gold OA may have the unintended consequence of tying academic reputation (conveyed by the prestige of a journal and associated citation metrics) to publisher APC payments, since more prestigious journals tend to have higher APC charges. Is cost per citation a valuable metric for assessing the value of publishing in a particular journal? What new indicators of the value of one’s research should be developed to assess academic reputation?
- 3. How is paid open access impacting the publishing output of less well funded researchers, fields, institutions, and countries?** Is the shift to Gold OA creating unintended equity issues and, if so, how can that be quantified? How will open access models, in particular APCs, impact research disciplines that attract fewer and smaller grants (especially in the humanities and social sciences)? What does an OA system that does *not* mostly benefit well-funded labs, and institutions that can afford to publish their results in their

journals of choice, look like? Is it possible to measure the extent to which authors from places and disciplines that typically receive less funding are excluded from publishing in journals with high author fees (controlling for quality of submitted work)?

4. **How is open access policy impacting scientific and scholarly societies?** How will new open access policies impact scientific and scholarly societies that have relied on subscription revenue to subsidize other activities? What challenges and opportunities are there for these societies in developing and adapting to new business models, and how would such transitions be funded? How is publishing's transfer strategy to keep submissions within a single company's family of journals impacting the revenues of scientific societies and other small publishers? More generally, what is the role of for-profit publishers in research dissemination? Do they benefit the current ecosystem beyond what can be provided by society and non-profit publishers? If they are generally causing harm to academic publishing, what steps might be taken to make non-profit alternatives successful, and to help disciplines that currently depend on for-profit publishers to cut free from that dependence?"
5. **How is the Gold open access model impacting industry consolidation?** The journal publishing industry has been consolidating, with larger publishers acquiring established journals and related academic infrastructure, as well as launching new open access journals. Reduced competition may lead to increased costs for researchers and universities. How will new federal policies impact industry consolidation and what metrics should be used to measure industry trends? Could current trends have been predicted? How should the rate of industry consolidation and its impact on publication quantity, quality, and cost be tracked and made publicly available?

RESEARCH DATA

- 1. How can we estimate the overall cost of research data sharing over time?**
How much data is the research enterprise generating, how can this be measured, and how is this output changing over time? How will the overall cost of implementing research data sharing be estimated and how much will it cost stakeholders to implement broad data sharing policies?
- 2. What metrics should be used to measure the broad impacts of research data sharing over time?** How do we measure the impacts of open science policies on intellectual property protection and management, innovation, and technology licensing at universities? What measures could be developed to assess the impact of open data sharing on scientific progress and commercialization? How will new data sharing policies impact intellectual property and the commercial impact of research discoveries, and how should those be measured over time?
- 3. How should universities and researchers prepare for required research data sharing?** How can federal agencies work with universities to ensure the longevity of data, especially when grants run out or researchers change institutional affiliation or retire? How can stakeholders assess if researchers have the expertise in data management needed to make the data they deposit truly useful to the community? What data management and curation training should be developed and required of researchers and how will that training differ by field? How can universities help their researchers acquire this expertise?
- 4. What standards for exemptions to sharing research data should be developed?** What exemptions are acceptable for delaying the sharing of data that underlies a submitted/published paper? Can these exemptions be standardized across federal agencies? Across fields? Is there value in developing a cross-agency standard for data sharing?
- 5. What tools and technologies should be developed to make the most of broad research data sharing practices?** Should technologies that enable

research data to be re-used without exposing or publishing the underlying data be part of open data sharing practices? This is particularly important for health data that potentially includes identifying patient/subject information. It is also important when considering the impact of generative AI.

PREPRINT PUBLISHING

1. **How have preprint sharing practices impacted different fields?** How are preprints being read in different fields, and by whom? How has growth in preprint publishing impacted peer review, speed of publication, and rates of journal publishing in different fields? What percentage of preprints go on to be formally published in peer reviewed journals?
2. **How has preprint sharing impacted scientific outcomes and public communication of science?** What measures should the community develop to assess how increasing access to preprints is impacting scientific outcomes (e.g., accuracy, reproducibility, etc.)? How does public access to preprints that have not been peer reviewed impact public understanding and media coverage of scientific research?
3. **What standards should be developed for preprint sharing?** There are, thus far, no common standards in preprint archiving for vetting papers prior to posting, linking preprints to journal versions of papers that may result, removing preprints when final papers are retracted, or when preprints themselves are retracted, verifying that a preprint has the same content (modulo formatting) at the reviewed journal version, etc. What are the desirable standards for preprint publishing and how should they be developed, deployed, and enforced? What does the community need by way of improved discovery tools to make searching across disparate preprint archives most efficient in the research process?
4. **What are the measurable impacts from sharing preprints and author manuscripts on citation practices?** Have preprints and the availability of

author manuscripts through Green OA models created significant citation confusion associated with multiple versions being available? How much do preprints and Author's Accepted Manuscript (AAM) versions of papers differ from Versions of Record (VOR) of papers (wherever they may be posted)? Does this vary among fields, and what impact has this had on the scientific enterprise?

5. **How can we assess differences in scientific or public impact between preprint sharing and OA article publishing (within a journal)?** Are there measurable differences in scientific impact between articles published immediately on preprint servers and articles that are made available through open access after full peer-review (and not first deposited in a preprint archive)?

PEER REVIEW

1. **How can we make the type of peer review a journal article has undergone more transparent, and will this help promote trust in science?** What standards and tracking capabilities should be developed to signal the quality and type of peer review a journal has used, and what metrics could be used to assess the quality of research peer review? Would these measures help identify predatory publishers and journals (where "predatory" describes publishers that exploit authors by charging APCs but don't provide the editorial and publishing services, such as peer review, that are associated with legitimate publishers)?
2. **What experiments and technologies should be developed to improve the quality and efficiency of peer review?** With respect to research integrity, among other concerns, what experiments in peer review could help inform the deployment of new and better peer review models? What AI tools could be developed to accelerate and improve the peer review process? What new technologies could help reviewers become more efficient and increase the quality of their reviews?

3. **What steps can be taken to ensure the long-term viability of quality peer review in academic publishing?** With the dramatic increase in article publication rates over the last two decades, peer review burnout is a problem for the entire scholarly communications ecosystem. How much time are researchers dedicating to peer review? How can we ensure that reviewers are representative of their fields? What could universities and funders do to encourage high engagement and appropriate credit for peer reviewing, especially for early career scholars?
4. **What models of publisher-independent, cost-effective *closed* peer review could be developed?** Many journals are piloting or using forms of open peer review, but “closed” peer review remains more common. Indeed, one of the major functions of journals is to manage closed peer review, to hide the identities of the reviewers in order to avoid social pressures that can bias review results. What models of publisher-independent, cost-effective closed peer review could or should be developed?
5. **What standards should be established for ensuring the quality and veracity of research data submitted to journals for publication?** Now that researchers will be required to share the data underlying the conclusions in their publications, should that underlying data also be subject to peer review? If so, how would standards for peer review of data be developed and implemented? How will this impact publication cost? As the quantity of data explodes, how will peer review scale? With the existing peer review system already under stress, how would it cope with such increased volume of material for review?

COSTS TO RESEARCHERS AND UNIVERSITIES

1. **What models should the research community adopt to track how much funding is going toward publication costs over time?** Should the flow of university and federal funds to publishers according to business model and publisher size/type be monitored? If so, how? Is open access expenditure

reducing the availability of funds for research? Who is paying APCs and other charges? Is the total cost of publication to research entities changing, or just being redistributed? Should standards for reporting publication spending be developed for both grant recipients and for university allocations of grant overheads? How much will it cost researchers and universities over time to pay to publish their research under different open access models, including APC-based and Read-and-Publish models? How is this best estimated, and how do costs compare to current expenditures?

2. **How should the research community monitor shifts in the amount and type of research publishing costs over time?** As more journals shift to APC models, the cost to read journals is moving from libraries paying for subscriptions to authors who pay APC charges to make their papers freely available at the time of publication. At the same time, libraries that engage in Read-and-Publish agreements are now carrying publication costs. How do the costs of subscriptions currently compare with the future total costs in APC charges for libraries? For authors? For universities? How much does the APC model concentrate costs on research-intensive universities?
3. **What are the most effective ways for universities and researchers to manage changing publication costs?** Historically, publishing was mostly free to authors, and accessing literature was paid for by library collections budgets. Most major research libraries have now converted at least some of their collections budget to cover some publication costs, whether via an APC fund available to institutional authors, via Read-and-Publish contracts that cover APCs in bulk, or via direct subsidies to support open access ventures. How will/should stakeholders determine the most effective ways to fund the costs of publication? What are the implications of different models for university stakeholders (budget offices, provosts, departments, scholars, libraries, etc.)?
4. **How should universities track the costs of open access over time and what are possible mechanisms for public sharing of such data?** How can the

manner in which universities are paying for “supply side” open access be tracked over time? Will it mostly come from grant funds, other university sources, or from dollars saved in library budgets if and when subscriptions are canceled? How will federal agencies track changes in APCs over time to keep track of the cost of publishing? How are savings/cost increases at institutions reflective of research productivity?

5. **How should researchers and universities budget for compounding costs associated with data deposition and storage over time?** The cost of storing data, along with the creation and maintenance of metadata over time, will increase with the amount of data shared and archived. How should researchers budget for long-term data storage and metadata maintenance in grant proposals? Is it possible to match the requirement for indefinite/long-term data storage to the short-term nature of grant proposals?

INFRASTRUCTURE

1. **What data management and storage infrastructures are necessary to ensure long term access, discoverability, and integrity of research data over time?** Is current data storage and management infrastructure in the U.S. sufficient for the purpose of managing and storing data into the foreseeable future? Data sharing infrastructure is siloed and often field specific. For example, there are ~1,300 biomedical research databases alone. How does the potential cost of maintaining a fragmented system compare with the cost and benefit of creating and supporting a federated/distributed infrastructure system?
2. **What open or community-owned research publishing infrastructure is needed to help support the research ecosystem?** How can we measure investments in essential infrastructure for scholarly communication on the part of different stakeholder groups (funders, publishers, universities, etc.)? How has vertical integration of publishing companies — meaning, the

growing ownership of data analytics, hosting, and portal services by large scholarly publishers — impacted researchers and research communication?³⁷ Once built, how can infrastructure be maintained and improved over the long term?

3. **What infrastructure should be developed to support compliance with new public access requirements?** How well-prepared are researchers, in terms of knowledge and skills, to comply with new requirements? What new technology and human services should be developed to ensure that researchers can easily comply with public access requirements? What are the appropriate infrastructure and support services that universities should be putting in place to aid researchers in complying with new policy requirements?
4. **How can we track and learn from successful examples of academy-owned publishing infrastructure?** What are the leading examples of widely used open or community-owned infrastructure, developed through grant funding, by nonprofits, or through university consortia? What can we learn from these examples to inform the future development of academy-owned publishing infrastructure?
5. **How can we ensure the longevity and stability of code sharing infrastructure?** What are the current infrastructures available to researchers for sharing code, and how well do they meet the needs of ease of use, longevity, and stability? How can we create a federated infrastructure for software and code that ensures longevity and stability? How can this infrastructure account for evolving code and updated contributions? How should the costs and benefits of this effort be assessed?

³⁷ <https://www.science.org/doi/10.1126/science.aba3763>

APPENDIX A: OTHER QUESTIONS GENERATED BY THE GROUP

ACCESS AND BUSINESS MODELS

1. What are the key opportunities going forward for scholarly societies and university presses to serve the research enterprise, given the changing publishing landscape?
2. Lariviere et al., studied market consolidation up to 2013, before many of the major publication policies from governments and funders went into effect.³⁸ What has happened since then and how much has the market consolidated? How many smaller/independent society journals remain in the market? How much are different publishers of different sizes (and approaches) growing? How much of that growth is through partnering with existing journals, including society journals, and how much is through launching proprietary competitor journals to those societies? How has the impact of publishing in various sectors of the market changed over the past ten years?
3. What impact will the changes in scholar's behavior/choices have on the evolving landscape? For example, what are the broader and longer-term impacts of editorial boards resigning from for-profit journals and starting new journals with non-profit publishers? What about scholars electing to publish their work with non-profit publishers, who typically charge lower APCs (or provide Diamond OA journals), thus reducing the overall cost of publishing to the university and keeping the funds "in the system"? If authors choose to publish their papers in commercial journals because they feel the journals offer higher quality services, should universities try to incentivize them to change their behavior?
4. How will scholarly book publishing be impacted by open access policies? Will such policies create disincentives to apply for federal funding if the

³⁸ <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0127502>

product will be a book that must be made available through open access?

5. There has been a steady increase in the rate of research papers being published over time, including a dramatic increase in publications overall. This growth has been driven in part by the volume incentive to journals of author-funded open access, and in part by the incentive system for scientists to publish more often, along with the opportunity to publish smaller contributions. How do we balance the contribution of more substantial publications with the contribution of publishing less well-developed research results faster, and how do we measure these changes? Has the Gold open access model affected the publication quality of research papers given the incentive to publish more articles?
6. How do journal offerings of services such as highly selective peer review, professional editors, copyediting, integrity checks, etc. compare between subscription and OA journals? How have some commercial and highly selective nonprofit journals managed to keep their citation rates so high? Are there lessons to be learned by journals with lower impact factors? Are there correlations in the amounts of services offered and the price level of the APC charged?

RESEARCH DATA

1. How can standards best be developed and maintained for common data types? Common standards (file types, metadata, etc.) for data types greatly increase efficiency of reuse, allowing for cross comparison between experiments and different research groups. How can community-driven standards be developed and supported as they evolve?
2. Should the funding agencies require in data management plans that grant applicants explain how the data will be stored forever? Will it be necessary to create some form of "storage annuity" that will allow researchers to pay up front for indefinite storage? What are best

- practices for federally funded researchers in budgeting for research data storage and access over time?
3. How will established intellectual property be impacted if data are reused in a new publication? In generative AI systems?
 4. What is the role, present and future, for libraries as data repositories?
 5. What are best practices for teaching researchers -- especially early career researchers -- good data (and code) sharing and stewardship practices?
 6. Will agencies specify Creative Commons or Open Source licenses for data and code publications?

PREPRINT PUBLISHING

1. How can we assess the financial impacts to journal publishers from expanded preprint publishing access?

PEER REVIEW

1. How can the scientific community better communicate to its own broad membership, and to the wider world, which journals are reliable or of particularly high quality? What role do networks of trust and endorsement play?
2. There are different types of open peer review. For example, some open peer review involves the reviews being shared openly while the identity of reviewers remains protected. Other open peer review practice opens the identity of reviewers as well. What measures can we use to assess the pros and cons of different types of open peer review and recommend best practices?

COSTS TO RESEARCHERS AND UNIVERSITIES

1. How will university costs for patenting be impacted, if data sharing compliance requires researchers to seek patents before commercial partnerships (with patent sponsorship) are solidified?

APPENDIX B: GLOSSARY OF TERMS

Article processing charge (APC). A fee charged to authors to allow for immediate, unrestricted access to the full version of a publication. APCs are paid by the author, the author's funding body, or their institution.

Big Deal. Bundled online subscriptions to academic journals are sold under a Big Deal at prices lower than the sum of their individual prices.

cOAlition S. cOAlition S is a group of private funding agencies and national research funding organizations, with the support of the European Commission and the European Research Council (ERC), whose goal is to ensure all the research they fund is made open access immediately on publication.

Creative Commons (CC By) licenses. CC licenses include six different Creative Commons licenses that allow users to distribute, remix, adapt, and/or build upon the author's work depending on the specific license, so long as the user gives credit to the original author.³⁹

Diamond open access journals. Diamond open access journals typically receive financial support from one or more institutions and organizations and thus do not have article processing charges (APCs) or other fees for publication.

Embargo Period. An embargo period is the time between the formal publication of an article and when it becomes freely available online.

Gold open access. Gold open access papers are immediately and freely accessible via the journal on the publisher's website as the version of record, often (but not always) via a fee paid by or on behalf of an author. They are published under a Creative Commons license.

Green open access. Green open access papers are freely accessible outside of the formal journal, often via an open access repository.

³⁹ <https://creativecommons.org/about/cclicenses/>

Hybrid journals. Hybrid journals are subscription-based journals that provide authors with an option to publish their papers as Gold open access, accompanied with a Creative Commons license.

Mega journal. Mega journals publish a broad variety of research without judging the perceived importance of it. Instead, they look purely at the soundness of the research. Well-known mega journals include PLOS ONE and Scientific Reports.

Open access. Open access (or OA) is the practice of providing free and unrestricted online access to research publications. Open access research is made available for free in perpetuity with limited restrictions on how readers can share and re-use the content.

Open Access Journal. An open access journal is a peer-reviewed publication that makes all of its articles freely available online without the need for subscriptions. It includes Gold open access journals, which charge APCs, and Diamond open-access journals, which usually do not include additional charges and rely on institutional funding.

Plan S initiative. Initiated in 2018 by cOAlition S, a global consortium of research funding bodies, Plan S requires all scholarly publications that are funded by its member organizations to be made immediately open access.

Predatory publishing. Predatory publishing is an exploitative publishing model to intentionally take advantage of the academic need to publish by charging authors while not conforming to the normal peer review process for individual articles.

Preprint. Preprint papers are posted by authors to a preprint server where they can be viewed freely prior to peer review and formal publication. While preprints have been shared for decades in the physics, math, and computer science communities, the practice of posting preprints in other scientific disciplines is relatively new. Typically (but not always), preprinted

publications are published twice: first in an open repository for the purposes of establishing intellectual priority and communicating with colleagues, and, secondly, in a journal for the purposes of peer review, broader dissemination of the work, and reputational advancement, among other traditional functions of journals.

Public access. Public access refers to the requirement placed on recipients of federal funds to make research results freely available. It does not address copyright or choice of open access business model.

Read and Publish agreement. (See Transformative Agreements below) In Read and Publish agreements the publisher receives payment, typically from an institution or consortium, for reading and publishing by authors from that institution or consortium in that publisher's journals.

Transformative Agreements. Transformative Agreements (Sometimes called Read and Publish agreements) are made between publishers and research institutions or consortia. They include provisions to cover the costs of Gold OA content published by researchers based at the institution, negating the need for APCs when those researchers publish their work Gold OA. They allow affiliated authors to make the final version of their article Gold open access as well as providing access to subscription content for the library users.

Version of record. The full-text HTML and formatted PDF of the final edited publication posted to a journal website.