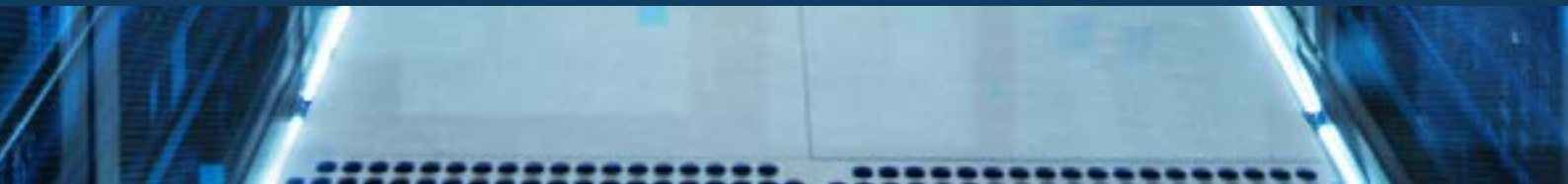
International Science Council

The global voice for science

THE NORMALIZATION OF PREPRINTS

ISC Occasional paper
March 2022



9 March 2022

AUTHOR: Luke Drury, Member of the Steering Group for the International Science Council's project on Scientific Publishing, with input from other members of the Steering Group.

SUGGESTED CITATION: Drury, L. 2022. The normalization of preprints. International Science Council, Paris. France. https://doi.org/10.24948/2022.02.

COVER IMAGE: gorodenkoff (iStock).

Work with the ISC to advance science as a global public good.

Connect with us at:

www.council.science
secretariat@council.science
International Science Council
5 rue Auguste Vacquerie
75116 Paris, France

www.twitter.com/ISC
www.facebook.com/InternationalScience
www.instagram.com/council.science
www.linkedin.com/company/international-science-council

INTRODUCTION

The last few years have seen an explosive growth in the use of preprints and the associated preprint servers by large sections of the scientific community. This article addresses the history of the preprint, its advantages and potential disadvantages, and concludes with some recommendations for how the growing acceptance of preprint posting should be handled within academia[1] and the changes in cultural norms (in other words its normalization) that this entails.

PREPRINT PREHISTORY

Contrary to popular belief preprints are not a consequence of the internet and digital technology (although their widespread adoption is). Before the digital preprint there existed paper preprints that were circulated through the postal system. They were a response to the unacceptably long delays in getting papers published through traditional journals in the days when papers could be held up in review, typesetting, proofing and finally distribution for up to two years or more on occasions. In response, those areas of science which were rapidly evolving, and had the necessary human and financial resources, resorted to sending preprints of articles that were due to appear to colleagues and the libraries of associated research centres in advance of formal publication, hence the name 'preprint'. This culture was particularly strong in particle physics and astrophysics, probably because as early instances of 'big science' they were already organized in large networks with a relatively small number of well-resourced research centres that could handle the logistics involved.

It is important to note that preprints arose to address the need for rapid communication within specific disciplinary communities and, at least initially, not as an open science initiative. On the contrary, the high costs of maintaining a formal preprint system based on secretarial assistance and involving typing manuscripts, large-scale photocopying machines (some occupying whole rooms), and significant postal expenses, in effect restricted it to a small number of elite institutions and the people working in them, the very antithesis of open science.

Two technological shifts radically changed this situation. The first was the release by Donald Knuth of his free mathematical type-setting system TeX in 1978. This defined a relatively simple way to encode even the most complicated mathematical formulae as a string of standard characters available on any computer as well as a sophisticated mark-up and macro language for formatting scientific papers. A document described in TeX (and more recently in one of its derivatives such as LaTeX) can be printed to a standard better than that provided by most commercial type-setting services and under the complete control of the author. And because the TeX source file is just a simple text file it can be easily and efficiently transmitted electronically and shared for re-use or text mining.

The second of course was the rise of the internet. It was an obvious step to replace the postal distribution of paper preprints with electronic distribution of the TeX source files. Initially this was done through simple e-mail distribution lists[2], but this was then automated to give the first preprint server which eventually became the arXiv system we know today. While everyone appreciates the importance of the internet in enabling preprint servers, the importance of TeX is often overlooked. Without a standard way of encoding a scientific paper as a text file it was much harder to distribute preprints and it is noteworthy that those fields (such as pure mathematics) that almost exclusively use TeX were among the first and most enthusiastic adopters of preprint servers. This is of course no longer the constraint it was in the early days when band-width and storage capacity were much more limited than is now the case.

Although not called preprints, we note that the tradition in economics of disseminating 'working papers' is closely analogous and most of what we have to say about preprints applies equally to working papers. Interestingly this is also a field that makes extensive use of TeX because of the high mathematical content of most economic papers. It is also a field that suffers from protracted peer review and a very small number of prestige publication venues so that the process of submission to publication can take several years.

RECENT DEVELOPMENTS

The success of arXiv in the physical sciences has stimulated the establishment of similar preprint servers in various disciplines; thus we now have ChemRxiv for chemistry, BioRxiv for biology, MedRxiv for health science, EarthArXiv for Earth science, AgriRxiv for agricultural science, SocArXiv for the social sciences, RePEc for economics etc. Essentially all areas of the natural and social sciences are now covered[3]. According to a recent survey article more than 40 preprint servers have been established in the last decade[4] covering both specific disciplines as well as geographical communities (e.g. the successful AfricArXiv for Africa, SciELO for Latin America, RINarxiv for Indonesia, and importantly ChinaXiv for China). However, some of the more specialized servers have only small numbers of posts and some have since closed down or no longer accept postings[5], emphasizing the importance of sustainable funding models.

Usage remains highest in the physical sciences (Fig 1) where arXiv has just passed two million posts[6] and is now the primary information source for many physicists. Over all areas of physics about 30% of papers are now first posted as preprints, and in the traditional areas of theoretical physics and astrophysics almost all work is now disseminated in preprint form. A 2017 study by the European Southern Observatory found that a remarkable 96% of the papers listed in its telescope bibliography database were available as arXiv preprints[7]. However, this is exceptional, and many disciplines have yet to adopt a preprint culture; overall it is estimated that somewhere between 4% and 6% of all scholarly publications are currently posted as preprints.

The big recent change has been the upsurge in health and life science preprints driven by the urgency of rapidly sharing information on COVID-19 related research[8]. In a rapidly

evolving pandemic it is clearly impossible to rely on the sclerotic publishing conventions of the past. This was recognized well before the current pandemic, and the World Health Organisation, back in 2015, at the time of the Ebola outbreak in West Africa, called for the rapid and early dissemination of research, a position it subsequently re-iterated in response to the Zika outbreak in Brazil.

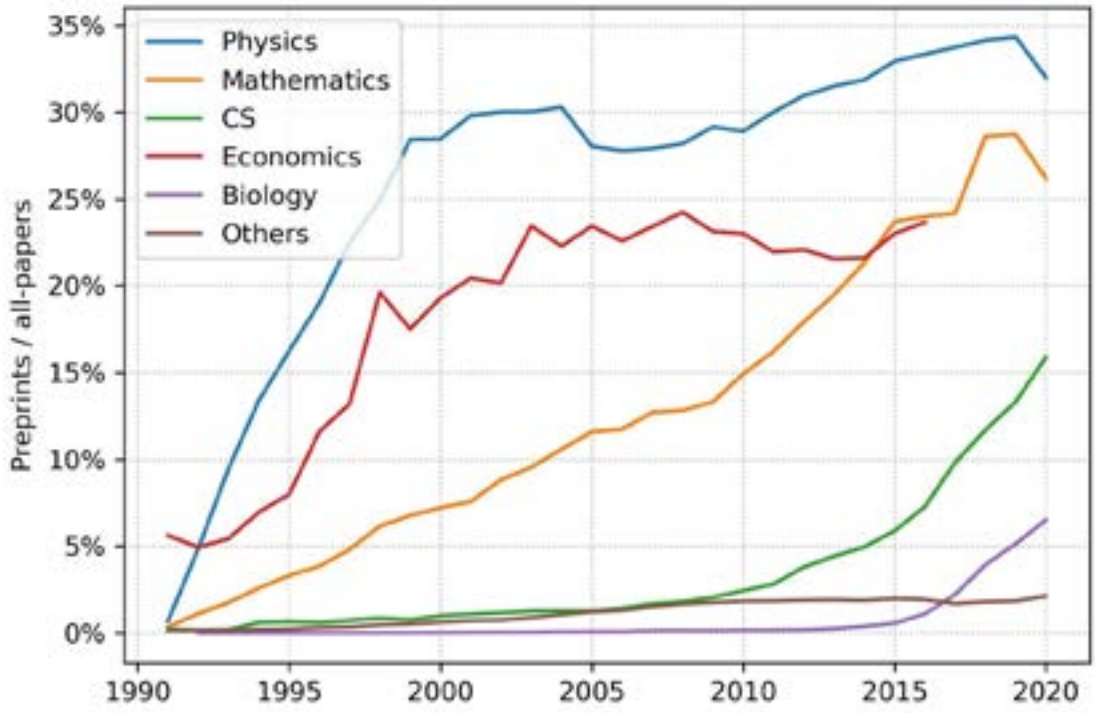


Fig 1 From Boya Xie, Zhihong Shen and Kuansan Wang arXiv:2102.09066v1. Percentage of all publications that are posted as preprints by discipline (CS is Computer Science).

ADVANTAGES OF PREPRINTS

The first and most obvious advantage of preprints is the one that gave rise to them in the first place, the need for rapid and efficient communication within disciplinary communities. Xie et al (loc cit) find that 'preprints help papers become accessible 7 months to 2.25 years earlier than peer-reviewed counterparts'. The advance of science should not be held back by unnecessary delays in publication, especially in rapidly advancing fields or where there is an urgent need for rapid progress (as recently in the case of COVID-19 related research). Dissemination to one's peer community in preprint form avoids this, allows for early peer community feed-back, and also serves a useful purpose in establishing priority should formal publication in a traditional journal be delayed (it may be an academic urban myth, but one does hear of malicious referees deliberately delaying acceptance of papers to allow time for their own rival papers to appear). In some ways preprints and preprint servers are the modern version of the early modern humanists' 'republic of letters' — networks for rapid knowledge exchange within distributed communities of researchers.

The second advantage of modern preprints as implemented on most server platforms is that it is easy to establish a record of versions as the paper evolves in response to

community feedback, especially if this is in the form of linked open peer reviews and comments. This provides a much clearer insight into how the ideas in the paper have evolved and, while useful now, will be invaluable to future historians of science. We have long accepted that books can and should go through multiple editions, and it is time that we replaced the outmoded concept of scientific papers having a unique 'version of record' with a similar 'record of versions' approach[9]. All too often issues are discovered with a paper after it has been published in the traditional system, and the journals with their static and immutable version of record do little to encourage corrections let alone retractions. With a record of versions this is much easier[10].

A third advantage is that many journals will now accept a link to a manuscript posted on a preprint server for submission, and there are new overlay journals that just add a peer review and indexing layer on top of the preprint server. This closer integration of more traditional models of publishing with preprint servers, and the emergence of hybrid systems such as the overlay journals, can be expected to continue[11].

A fourth advantage is that the ease of posting to a preprint server, and the rise of regional servers such as AfricArXiv and SciELO, can substantially increase the visibility of research in national languages and addressing local issues thereby making a valuable contribution to bibliodiversity in the spirit of the Helsinki initiative on multilingualism in scholarly communication.

But perhaps the greatest advantage is that anyone with an internet connection can download a preprint at effectively zero marginal cost, and if the record of versions includes the final 'author accepted manuscript' this is functionally equivalent to green open access. Indeed, in having access to the full source text, high resolution figures, corrections, comments, and other ancillary files the preprint version may in many cases be more useful than the journal version. If coupled with live links to all cited articles that are not restricted by copyright, to the data they contain, and to text and data mining applications, the resultant interoperability would create new opportunities for a more creative open science. These perspectives might currently seem distant, but they would be achievable if we were able to reform some of the restrictive practices of the dominant model of commercial publishing. As with other forms of open access there is clear evidence that this increases citation rates and impact. Xie et al (loc cit) find that papers with a preprint version attract on average three times more citations[12] than those without[13].

Thus, the normalization of preprints offers a relatively straightforward route to open science as envisaged by the recent UNESCO Recommendation on Open Science, which was endorsed by its 193 member states, and at the same time would open up new prospects for innovative applications. It is now 20 years since the Budapest declaration, which heralded the potential of the digital revolution to enable affordable open access to the record of science. Yet the mainstream commercial model of scientific publishing still dominates and impedes that vision. Its excessive costs discriminate against authors or readers or both, particularly in low- and middle-income countries, those in poorly funded institutions and those without institutional affiliations. It is difficult to see how UNESCO's vision of open science can be realized without deep reform of the current

mainstream publishing system; the normalization and expansion of the preprint domain may be the disruptive change that is needed.

A recent ISC report identifies eight essentials that must be maintained, and where there are deficits to be addressed by reform, if scientific publishing systems are to effectively serve the needs of 21st century science in ensuring that outputs are universally affordable for both authors and readers; that they carry open licenses; that openness to sceptical peer review is maintained; that the evidence for published truth claims is concurrently accessible for scrutiny; that the record of science is maintained for future generations; that the needs of different disciplines and communities are addressed (no one size fits all); that systems are flexible and adaptable to new needs and new technologies; and that systems are accountable to the international scientific community. The current commercial model fails on many counts. Preprint systems offer a route whereby these essentials can be maintained and developed where necessary.

PERCEIVED DISADVANTAGES OF PREPRINTS

The criticism that is most often levelled at preprints is that they have not been peer reviewed. In fact in their original form they often were peer reviewed. Many scientists would wait until the paper had been accepted for publication before posting the final manuscript version as a preprint with a covering note, 'to appear in', and specifying the journal, but this has largely fallen into abeyance and was never universal. But even without formal peer review organized through a journal, it is worth noting that many preprints will have been subject to varying levels of review and scrutiny. In the case of large experimental consortia any publication will generally have been subject to internal review by a publication committee and other members of the consortium before it is released as a preprint. Many laboratories and institutes operate similar internal review procedures, and in the case of publications by doctoral students and early career researchers any publications will usually have been vetted by their supervisors and mentors. Furthermore, even if it is well short of full peer review, most preprint archives operate some form of gate-keeping to exclude obviously inappropriate material[14]. As far as independent external peer review is concerned, 'review platforms' such as PreLights, Review Commons, PCI or PreReview may be arguably more reliable than conventional peer review as practiced by most journals[15][16]. These platforms crowd-source peer reviews of posted preprints from expert communities. The reviews are then reviewed and recorded in a structured format linked back to the original preprint. The result—a 'Refereed Preprint'—is even accepted by some journals without further review and offers an interesting and attractive alternative to the traditional journal model[17]. Interestingly plagiarized and fraudulent papers are almost never posted as preprints, presumably in part to avoid the early scrutiny by multiple peers which preprints attract.

The other problem with this criticism is that it greatly over-values the quality, value[18] and reliability of peer review as traditionally organized by journals[19]. We have all seen papers in even the most eminent of journals where we can only shake our heads and ask what the referees were thinking. And it is well known that some of the most damaging cases

of fraud and misinformation have appeared in long-established and reputable journals, supposedly after rigorous peer review, and have only been uncovered by critical voices in the community[20]. The most profound peer review is that applied after publication by the relevant scientific community to claims that are regarded as potentially important[21], but at the moment we have no way of properly signalling this. Ideally one would like to see verifiable and transparent markers of community esteem attached to publications, but in their absence we use crude bibliometric measures and journal reputation as proxies with all the well-known problems that this causes.

Thus there is undoubtedly a peer review problem, but it is not specific to preprints. In many ways the problem is that we have failed to recognize that while in the past the most resource intensive part of scholarly communication was physical printing and distribution, now that everything is digital the balance has radically changed. Production and dissemination is now trivial with virtually zero marginal costs, and in consequence publication, in the sense of making public, has moved upstream in the research workflow and increased enormously in volume. This is one of the core ideas of open science, that we publish not just final summaries, but preliminary and intermediate products such as data, protocols, software etc. The resource intensive step is now peer review, with estimated costs well in excess of 2 billion USD annually being borne by the scientific community rather than publishers[22], suggesting that we need to re-organize our processes and values to better reflect this. Although artificial intelligence tools may help somewhat, it will simply not be possible in an open science future to peer review everything to the same standard and arguably we should focus peer review on those cases where it is most needed.

The issue of not being peer reviewed is most acute in the case of medical and health-related research (although similar concerns could be expressed about any research with major and immediate political and social consequences, in economics or sociology for example)[23]. As we have sadly seen several times during the pandemic, great harm can be done by the publication of misleading and false claims. Peer review can catch some of these, but is fallible and over-valued: It is no philosopher's stone that can turn sloppy work into science gold, and it does not allow authors to abrogate their ethical responsibilities to research integrity. We are right not to uncritically trust work that has not been peer reviewed, but equally we should not assume that everything that claims to have been peer reviewed has in fact been thoroughly reviewed by competent experts. The tendency by some science journalists to treat 'has been peer reviewed' as equivalent to fact-checking and endorsement by the scientific community is dangerous. Arguably preprints, explicitly subject to peer community review, better reflect the contingent and preliminary nature of much research.

A practical disadvantage of preprints at the moment, and a consequence of our over-valuation of journal peer review, is that they are still in many quarters considered to be not quite proper forms of research output. The most crass example of this was the recent (thankfully now reversed) decision by the Australian Research Council in a recent funding round to disallow all applications that cited preprints[24]. But without going to this extreme it is all too common to see instructions for funding applications, or for career progression, state that only peer-reviewed journal publications should be listed.

This is problematic. On the one hand it goes against the spirit of modern developments in research assessment (e.g. DORA) which require us to recognize a broader range of research outputs and not just the traditional journal article. Perhaps more importantly, if we are assessing research for funding or career progression, it is illogical to exclude evidence of recent research which may be the more relevant information on which to base a decision. This is particularly an issue for early career researchers who cannot rely on a long back-catalogue of published work to establish their credentials; for them the early visibility afforded by preprints may be crucial.

Another problematic issue is the sustainability of preprint servers and the related question of the long-term curation of their content. But one could equally, and with more reason perhaps, ask the same questions of the commercial publishers. It is true that, like all shared common good infrastructures, the preprint servers are dependent on third party support and voluntary contributions; but the costs of running the servers are minuscule compared to the costs of traditional journal subscriptions or article processing charges[25], and they are now so essential to the efficient conduct of research, that it is difficult to believe that they will not continue to be supported by funders and users. There is clearly more than sufficient money in the global science system to support a major preprint effort, it is just that too much of it is locked into large and often ruinously expensive 'big deals' with commercial publishers. As far as long-term curation is concerned the commercial publishers offer no better guarantee; this is an issue that needs to be addressed by something like the internet archive and underlines the need to develop a better funding model for preserving the corpus of knowledge as a public good.

CONCLUSIONS AND RECOMMENDATIONS

It is clear that preprints can and do perform a valuable service in accelerating the advance of science and are here to stay. Furthermore, we can anticipate that their use will continue to expand into new disciplinary areas as well as geographical and linguistic communities. This will require adjustments in the cultural norms we apply to scholarly and scientific communication, and crucially also in how we assess research[26].

The first change is that we all need to accept that posting preprints is now a normal part of modern research culture and an important enabler of greater efficiency, visibility and integrity in research. Preprints should be seen as just one part of a coherent move towards more open science, where we use modern technology (extraordinarily cheap digital storage and virtually free communication) to open up the research process and share a broader range of research outputs at earlier stages of our work. A corollary is that citing preprints should be encouraged as evidence of recent relevant research and adoption of open science practices in research assessments; the still all too common practice of requiring applicants to only list peer-reviewed journal publications is to be deprecated.

The second change we need to make, not just with preprints but more generally with open science, is to re-emphasize the need for researchers to behave in an ethical and

responsible manner. In areas where there is a risk of misunderstanding by the public, or where research can be anticipated to have significant social, medical and economic consequences, there is an enhanced onus on researchers to observe principles of research integrity and social responsibility. Open science does not mean that everything has to be freely shared, and reasonable restrictions on access can be imposed in cases where this is justified. But even when access restrictions are imposed, and even when research has been approved by ethical committees, it is still the primary responsibility of the researchers concerned to make sure that their work is presented in a proper context and with appropriate caveats. Fundamentally the more open the science, the greater the responsibility to communicate it responsibly.

Thirdly we need to rethink how we organize and record peer review, and more generally how peer community acceptance or rejection is signalled. The current journal-based system is over-burdened, fallible, inefficient, widely mis-understood and introduces perverse incentives[27]. But this is a problem of research assessment generally and not of preprints per se. We need to move away from selection for novelty and more towards selection for quality; one suggestion is to focus more on methodological issues of reproducibility, data quality and data availability as advocated by Michael Barber (loc cit). Technology offers opportunities to automatically and efficiently screen preprints prior to posting as demonstrated by a screen[28] of COVID preprints by the Automated Screening Working Group[29]. This approach is not intended to replicate peer review. Instead, it uses sophisticated tools, including natural language processing, to screen a preprint on a range of indicators designed to detect weaknesses in research methodology and, particularly, to assess the probability that the reported research could be replicated or reproduced, thereby facilitating post-publication review of novelty and impact. Preprint servers and the research communities they serve should be encouraged to develop and refine such tools with appropriate discipline specificity and then deploy them routinely.[30] Indeed, use of such tools by authors prior to dissemination would have a significant effect on the quality of papers whether posted to a preprint server or submitted to a journal.

Finally, and crucially, we need sustained long-term investment by funders and research performing organizations in community-governed not-for-profit repositories such as arXiv, Zenodo and the exemplary Latin American La Referencia. Science, and the metadata about who is generating and using it, is too valuable to be allowed to fall exclusively into the hands of commercial entities; it must be curated as a public good for the benefit of this and future generations. A corollary here, of course, is support for a rights retention policy; authors should retain basic rights to their own work. Ideally, we need a global system of federated publicly funded and community governed repositories for open science products, including preprints, backed up by a long-term curation strategy. Long ago we accepted the idea of national copyright deposit libraries as public institutions dedicated to the curation of printed materials. We need a modern equivalent for digital knowledge, and just as a library is useless without catalogues and indices, we need intelligent, platform-agnostic search engines that can help us navigate this global digital virtual library. If we can achieve this, we will have taken a major step towards realizing the vision of science as a global public good, open to participation by all.

REFERENCES

1 For a similar discussion see Garisto, D. 2019. Preprints Make Inroads Outside of Physics. APS News Vol. 28, N. 9. http://www.aps.org/publications/apsnews/201909/preprints.cfm.

2 Feder, T. 2021. Joanne Cohn and the Email List That Led to ArXiv, Physics Today. https://doi.org/10.1063/PT.6.4.20211108a.

3 ASAPbio maintains a list of mainly life science preprint servers https://asapbio.org/preprint-servers, and the Confederation of Open Access Repositories plans to launch a preprints directory https://www.coar-repositories.org/news-updates/ccsd-and-coar-announce-plans-to-launch-preprint-directory/.

4 Xie, B., Shen Z., Wang, K. 2021. Is preprint the future of science? A thirty year journey of online preprint services. https://arxiv.org/pdf/2102.09066.pdf

5 Mallapaty, S. 2020. Popular Preprint Servers Face Closure Because of Money Troubles. Nature 578, no. 7795, pp.349–349. https://doi.org/10.1038/d41586-020-00363-3.

6 Garisto, D. 2022. ArXiv.Org Reaches a Milestone and a Reckoning. Scientific American. https://www.scientificamerican.com/article/arxiv-org-reaches-a-milestone-and-a-reckoning/.

7 U. Grothkopf, D. Bordelon, S. Meakins, E. Emsellem. 2017. On the Availability of ESO Data Papers on arXiv/astro-ph. The Messenger 170, pp. 58–61. https://doi.org/10.18727/0722-6691/5056

8 Fraser, N., Brierley, L., Dey, G., Polka, J.K., Pálfy, M., Nanni, F., et al. 2021. The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. PLoS Biol 19(4): e3000959. https://doi.org/10.1371/journal.pbio.3000959

9 For an alternative view see Hinchliffe, L. J. 2022.The State of the Version of Record. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2022/02/14/the-state-of-the-version-of-record/.

10 See discussion of retractions in Michael Barber's occasional paper for the International Science Council, Strengthening Research Integrity, https://council.science/wp-content/uploads/2020/06/2021-11-Research-integrity.pdf; and for a criticism of how retractions work in practice at the moment in Avissar-Whiting, M. 2022. Downstream Retraction of Preprinted Research in the Life and Medical Sciences. MetaArXiv. https://doi.org/10.31222/osf.io/xdekq.

11 Johnson, R. and Chiarelli, A. 2019."The Second Wave of Preprint Servers: How Can Publishers Keep Afloat?. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2019/10/16/the-second-wave-of-preprint-servers-how-can-publishers-keep-afloat/.

12 See also Fu, Darwin, Y., and Hughey, J.J. 2019. Releasing a Preprint Is Associated with More Attention and Citations for the Peer-Reviewed Article. P. Rodgers and O. Amaral (eds). ELife 8 (December): e52646. https://doi.org/10.7554/eLife.52646.

13 For a more critical view see: Davis, P. 2018. Journals Lose Citations to Preprint Servers. The Scholarly Kitchen. https://scholarlykitchen.sspnet.org/2018/05/21/journals-lose-citations-preprint-servers-repositories/. However his argument is that citations are split between the preprint and journal version, and not that total citations are reduced.

14 See this interesting review for details on how arXiv handles this issue. Ginsparg, P. 2021. Lessons from ArXiv's 30 Years of Information Sharing. Nature Reviews Physics, pp. 1–2. https://doi.org/10.1038/s42254-021-00360-z.

15 Enago. 2020. Will Crowd-based Peer Review Replace Traditional Peer Review? https://www.enago.com/academy/will-crowd-based-peer-review-replace-traditional-peer-review/

16 See also: Franco Iborra, S., Polka, J., Puebla, I., 2022. Guest Post: Preprint Feedback is Here – Let's Make it Constructive and FAST, The Scholarly Kitchen, https://scholarlykitchen.sspnet.org/2022/02/28/guest-post-preprint-feedback-is-here-lets-make-it-constructive-and-fast/, and O'Sullivan, L., Ma, L. and Doran, P., 2021. An Overview of Post-Publication Peer Review. Scholarly Assessment Reports, 3(1), p.6. http://doi.org/10.29024/sar.26

17 See i.a. discussion in Brembs, B., Huneman, P., Schönbrodt, F., Nilsonne, G., Susi, T., Siems, R., Perakakis, P., Trachana, V., Ma, L., and Rodriguez-Cuadrado, S. 2021. Replacing Academic Journals. https://doi.org/10.5281/zenodo.5526635.

18 Two separate recent studies have compared a large number of preprints with the subsequently fully published versions. One study used manual comparisons (Brierley, L., Nanni, F., Polka, J.K., Dey, G., Pálfy, M., Fraser, N., et al. 2022. Tracking changes between preprint posting and journal publication during a pandemic. PLoS Biol 20(2): e3001285. https://doi.org/10.1371/journal.pbio.3001285), and another used machine learning algorithms (Nicholson, D.N., Rubinetti, V., Hu, D., Thielk, M., Hunter, L.E., Greene, C.S. 2022. Examining linguistic shifts between preprints and publications. PLoS Biol 20(2): e3001470. https://doi.org/ 10.1371/journal.pbio.3001470). The conclusions were that a very large proportion (83-93%) of the preprints underwent no major changes when fully published. The rigour of the first study has been contested in a paper (https://thegeyser.substack.com/p/data-missing-study-may-be-useless) that suggests ways in which the conclusions may be invalid. These are important issues for publishing and need to be rigorously explored.

19 For an interesting discussion see Heesen, R., and Bright, L.K.2021. Is Peer Review a Good Idea? British Journal for the Philosophy of Science 72, no. 3: 635–63. https://doi.org/10.1093/bjps/axz029.

20 The most egregious recent example must be Andrew Wakefield's paper in the Lancet falsely linking the MMR vaccine to autism https://en.wikipedia.org/wiki/Lancet_MMR_autism_fraud; see also https://www.buzzfeednews.com/article/stephaniemlee/elisabeth-bik-didier-raoult-hydroxychloroquine-study

21 Interestingly there are some examples of famous papers, including Murray Gell-Mann's 1961 Nobel prize-winning "Eightfold way" paper and, more recently, the three papers by Grigory Perelman solving the Poincaré conjecture, for which he won the Fields medal and Clay Millennium prize, which were immediately recognized for their significance by the peer community and only circulated as preprints without appearing in a traditional journal.

22 Aczel, B., Szaszi, B. and Holcolmbe, A.O.. 2021. A billion-dollar donation: estimating the cost of researchers' time spent on peer review, Research Integrity and Peer Review 6:14 https://doi.org/10.1186/s41073-021-00118-2 found that the total time reviewers globally worked on peer reviews was over 100 million hours in 2020, equivalent to over 15 thousand years. The estimated monetary value of the time US-based reviewers spent on reviews was over 1.5 billion USD in 2020. For China-based reviewers, the estimate is over 600 million USD, and for UK-based, close to 400 million USD.

23 Watson, C. 2022. Rise of the Preprint: How Rapid Data Sharing during COVID-19 Has Changed Science Forever. Nature Medicine. https://doi.org/10.1038/s41591-021-01654-6.

24 Watson, C. 2021. Australian Funder Backflips on Controversial Preprint Ban. Nature. https://doi.org/10.1038/d41586-021-02533-3.

25 The basic operating costs of a preprint server, such as those hosted by the Centre for Open Science, are less than $10 per preprint, see https://www.cos.io/products/osf-preprints. The big commercial publishers quote article processing charges that are typically many thousands of dollars, two to three orders of magnitude higher.

26 For a critical view see e.g. Altman, M., Cohen, P.N. and Polka, J. 2021. Preprints and Pandemics: Interventions into the Dynamic System of Scholarly Communication. MetaArXiv. https://doi.org/10.31222/osf.io/6nzhe.

27 See e.g. Arvan, M., Bright, L.K. and Heesen, R. 2022. Jury Theorems for Peer Review. The British Journal for the Philosophy of Science. https://doi.org/10.1086/719117.

28 Weissgerber, T., Riedel, N., Kilicoglu, H. et al. 2021. Automated screening of COVID-19 preprints: can we help authors to improve transparency and reproducibility?. Nat Med 27, 6–7. https://doi.org/10.1038/s41591-020-01203-7

29 https://scicrunch.org/ASWG

30 The ASWG tools are used by journals of the the American Association for Cancer Research, see https://www.scicrunch.com/news/2020/8/11/sciscore-to-launch-a-pilot-with-the-american-association-for-cancer-research-to-help-authors-improve-rigor-and-reproducibility-in-their-published-work

International Science Council

The global voice for science

Work with the ISC to advance science as a global public good.

Connect with us at:

www.council.science
secretariat@council.science
International Science Council
5 rue Auguste Vacquerie
75116 Paris, France

www.twitter.com/ISC
www.facebook.com/InternationalScience
www.instagram.com/council.science
www.linkedin.com/company/international-science-council