



# The evolution, benefits, and challenges of preprints and their interaction with journals

Pippa Smart

PSP Consulting, Oxford, UK

### Abstract

This article presents the growth and development of preprints to help authors, editors, and publishers understand and adopt appropriate strategies for incorporating preprints within their scholarly communication strategies. The article considers: preprint history and evolution, integration of preprints and journals, and the benefits and disadvantages, and challenges that preprints offer. The article discusses the two largest and most established preprint servers, arXiv.org (established in 1991) and SSRN (1994), the OSF (Open Science Foundation) initiative that supported preprint growth (2010), bioRxiv (2013), and medRxiv (2019). It then discusses six different levels of acceptance of preprints within journals: uneasy relationship, acceptance of preprint articles, encouraging authors to preprint their articles, active participation with preprints, submerger by reviewing preprints, and finally merger and overlay models. It is notable that most journals now accept submissions that have been posted as preprints. The benefits of preprints include fast circulation, priority publication, increased visibility, community feedback, and contribution to open science. Disadvantages include information overload, inadequate quality assurance, citation dilution, information manipulation and inflation of results. As preprints become mainstream it is likely that they will benefit authors but disadvantage publishers and journals. Authors are encouraged to preprint their own articles but to be cautious about using preprints as the basis for their own research. Editors are encouraged to develop preprint policies and be aware that double-blind review is not possible with preprinting of articles and that allowing citations to preprints is to be encouraged. In conclusion, journal-related stakeholders should consider preprints as an unavoidable development, taking into consideration both the benefits and disadvantages.

### Keywords

Open science; Preprints; Publishing; Scholarly communication

**Received:** February 1, 2022

**Accepted:** February 8, 2022

**Correspondence to** Pippa Smart  
Pippa.smart@gmail.com

### ORCID

Pippa Smart

<https://orcid.org/0000-0002-5528-4704>

### Introduction

**Background:** Preprints are versions of articles made publicly available before traditional jour-

nal-based peer review. Preprints have been part of the publishing landscape for 30 years, but only recently have become a common means of communication for life science and medical authors. The year 2020 was a watershed period when the principle of publishing by preprint became a normalized process for research on the coronavirus disease 2019 (COVID-19) pandemic. However, many editors, publishers, and journals have viewed preprints with suspicion. It is reported from a survey in 2020 that only 28 of the 383 Asian academic society journals from 22 countries accept submissions that have already been posted on preprint servers. Equally only eight journals allow preprints to be cited in reference lists [1]. The study also found that half of the 118 Korean editors disagreed with the need for preprints. The concerns over preprints included a lack of scientific integrity, stealing ideas/scooping data, priority issues regarding research ideas, and copyright problems [2]. However, although there remains some resistance, there has been a growth of participation, experimentation, and integration of preprints within journals.

**Objectives:** This article aims to present the current advances in preprint-related developments to help journal stakeholders understand those issues and find a way to adapt to the growing preprint culture. Specifically, the article will present: how preprint servers have evolved; how journals are integrating with them; the benefits and disadvantages of preprints; and the opportunities and challenges that preprint posting poses for research, publishing, and the public.

## How Preprint Servers Have Evolved

The principle of preprints is to allow researchers to share their work before formal publication and digital means of sharing have been in operation since the 1960s [3]. However, the first official preprint server was arXiv.org, launched in 1991 at Los Alamos National Laboratory in the USA by Paul Ginsberg. The initiative followed other sharing initiatives between scientists, mostly in the discipline of physics, and the new service was designed to take advantage of the new World Wide Web, which was only a few years old [4,5]. Its focus was high energy physics when first launched. This repository has grown and developed over time into new areas, including economics, and articles on mathematics now comprise the largest part of its database. Currently, arXiv.org is based at Cornell University and hosts over two million articles with 16,000 new submissions each month. As with other preprint servers, it does not operate peer review, but it does scan each submission for suitability. This used to be done manually, but it now uses artificial intelligence to evaluate all submissions prior to posting on the server.

The second preprint server to launch was SSRN which was

launched in 1994 and now contains over one million articles (<https://www.ssrn.com>). This server was established by a pair of financial economists but was purchased by Elsevier in 2016. The purchase by a commercial publisher was criticized by the research community, who had previously considered it a community initiative, not appreciating that it had been run as a private enterprise for several years before the sale.

Since 2010 there have been many new preprint servers launched, several supported by the Center for Open Science which launched the OSF initiative (Open Science Foundation, <https://osf.io>) as a platform to support the hosting of preprint collections. The largest change in the history of preprint servers came in 2013 with the launch of bioRxiv, followed by medRxiv in 2019. Both are hosted at Cornell alongside arXiv.org. Although very much smaller than SSRN and arXiv with approximately 200,000 articles, they both experienced tremendous growth during the pandemic as researchers used preprints as a rapid means of sharing research on COVID-19. The statistics of uploads and downloads can be found at <https://rxivist.org/stats>.

It is worth noting that although preprint servers do not undertake peer review, they each operate a screening procedure before adding articles to the sites. This ranges from a scant check, performed by an individual or using artificial intelligence, to a fuller check performed by medRxiv, which operates probably the most stringent checks before posting, since it is aware of the potential dangers of making available misleading or inaccurate medical research (<https://www.medrxiv.org/submit-a-manuscript>).

Research suggests that not all preprints are finally published in peer reviewed journals (or other accredited outlets), although it is impossible to accurately assess the number that are. A 2021 article suggested that 60% of preprints are never published and that preprints represent 4% of the published literature [6]. Other research suggests that 30% of preprints from bioRxiv remain unpublished [7].

## How Journals Are Integrating with Preprints

The relationship between journals and preprints is extremely varied, and has been divided into six steps to acceptance.

**Uneasy relationship:** The Ingelfinger rule was named in the late 1960s after the editor in chief of the *New England Journal of Medicine*. Franz Ingelfinger stated that the journal would only publish items that were entirely new, novel, and had not previously been published or where a press release about the research findings had been made. His argument was that the journal wanted all rights of first publication. Many journals still feel this way, although in recent years it is notable that many have changed their opinions. Of interest, although

many journals now accept preprints, the research community appears to be somewhat unaware of this, and in 2013 it was reported [8] that 60% of German scientists believed that no journal would accept a submission that had appeared in a preprint server. This data is now quite old, and it would be interesting to run the same survey to see if opinions have changed alongside journal policies.

**Acceptance:** While individual publishers and journal editors may still feel somewhat uncomfortable with preprints, there has been a general policy change, and many now accept these articles—whether willingly or not. Several publishers have made statements saying that journals should allow submissions of preprints (for example, <https://authorservices.wiley.com/author-resources/Journal-Authors/open-access/preprints-policy.htm>).

**Encouragement:** Moving further, several journals now actively encourage authors to preprint their articles—going one step beyond stating that preprints are accepted. Some journals even offer to upload articles to preprint servers on behalf of the authors following submission. For example, PLOS authors are asked if they have uploaded onto a preprint server and, if not, if they would like the journal to do this on their behalf. In this case, articles are uploaded onto bioRxiv or Med Rxiv (<https://journals.plos.org/plosone/s/preprints>).

**Participation:** In a linked-in world, there is increased collaboration between different services, which is the same within the journal landscape. Some journals have changed from “opt in” (i.e., “would you like us to upload your article onto the preprint server?”) to opt-out (i.e., “we will upload your article onto the preprint server unless...”). For example, the *Lancet* family of journals upload all submissions, unless the author opts out with a valid reason, onto SSRN—the preprint server owned by its parent, Elsevier ([https://www.thelancet.com/journals/lancet/article/PIIS0140-6736\(18\)31125-5/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS0140-6736(18)31125-5/fulltext)). In another example, the journal *eLife* will only accept submissions that have already been made available on a preprint server such as medRxiv or bioRxiv (<https://elifesciences.org/inside-elifesciences/00f2f185/preprints-and-peer-review-at-elifesciences>). Allied to this approach is a linking service introduced onto the preprint server ResearchSquare affiliated with Biomed Central (BMC) journals. This service, called “In Review” links with the submission system used by BMC. As articles go through peer review, they are tracked and the status of the article is reported against the preprint on ResearchSquare (<https://www.researchsquare.com/publishers/in-review>).

**Submerger:** An extension of participation is a close alliance of preprint servers, journals, and the peer review service. “Review Commons” was launched as a partnership between ASAPbio and the journal EMBO. Authors who upload their articles onto bioRxiv can request a review from the EMBO

editorial office. EMBO manages peer reviews to obtain two reviews which are posted against the article on bioRxiv. Once the reviews have been received, authors can revise and submit to one of 17 participating journals which use the reviews as the basis for accepting or rejecting the revised article (<https://www.reviewcommons.org>). This service is limited as it relies on the goodwill and the capacity of both EMBO editorial offices, and so is not ultimately scalable.

**Merger and overlay:** The final step in acceptance between journals and preprint servers is merger. There are a few examples of these. *F1000 Research*, now owned by Taylor & Francis, was the first journal to create a model that merged both the immediacy of the preprint server with the selectivity of journals. Submitted articles are published immediately after a quick evaluation, including a plagiarism check, etc. Reviewers are then invited, and their reviews are posted against the article. The reviewers make the publishing decision, and if there is sufficient consensus, the article is considered “Accepted” or “Rejected.” If accepted, the article goes into various indexes, including Scopus and MEDLINE (<https://f1000research.com/about>). Authors can revise their articles at any point, and all revisions are linked to previous and subsequent versions—each undergoing the same check and review process.

Another model is that of an overlay journal, and there are a few of these—mostly linked to arXiv. Under this model as used by the *Open Journal of Astrophysics*, authors whose works are on arXiv can submit to the journal which provides traditional peer review. If the article is accepted, it is considered a published article within the journal. However, only a summary is provided on the journal site, and the final article must be posted onto arXiv. In this way, the journal accredits the articles it “accredits” but leaves all articles on arXiv as the preferred host.

## Benefits and Disadvantages of Preprint Servers

### Benefits of preprints

There are several benefits of preprints, many of which have been extensively discussed [9,10] and which include the following.

**Fast circulation:** When results are needed quickly, preprints will always be a faster means of communication than journals that have more stringent—and time consuming—checks, including peer review before publication.

**Priority publication:** Posting a preprint gives authors priority over ideas and results, which may be important if there is competition between different laboratories or departments.

**Increased visibility:** Preprints are currently all free to view and published under open access licences which enables greater dissemination without any access barriers. They also often have greater visibility on search engines than other

smaller platforms such as institutional or personal websites.

**Community feedback:** Most preprint repositories allow for feedback, and in some communities this is a valuable means of obtaining community comments.

**Author publication control:** Authors decide where and when to post their research without having to hope that editors will accept their articles, thus giving them greater control over their own work.

**Open science:** Preprints support open science in several ways: not only do they currently publish under open licenses, but they also allow for different versions to be posted, including linking to data sets, working papers, and other documents to help ensure that the complete picture of research can be obtained.

**Democratic process:** There is a history of bias within journal publications [11] and discrimination regarding which articles are published; however, preprint servers are not subject to such selectivity and offer a more democratic publishing process.

**Version access:** It can be important in some areas to see the history of an article and how it has developed through different iterations. Preprints can provide this history by providing access to the original submission before peer review and final publication.

### Disadvantages of preprints

Although there are obvious benefits to preprint publication, there are counterbalancing challenges that should be considered [12,13]. These include the following.

**Information overload:** There is concern that the increase in scientific literature is causing problems for researchers due to the amount of time required to sift and discover relevant quality information. Preprints increase the volume of publications and, therefore, the information overload.

**Quality assurance and trust:** There is no scrutiny of preprints such as peer review before posting, they therefore cannot provide any quality assurance or trust signaling regarding the content of articles. This lack of quality control may not be appreciated by some people to whom a preprint site such as bioRxiv appears to be the same as a journal and to whom, therefore, the same quality and trust values are applied.

**Reputation damage:** It can be argued that authors are protected by the peer review system since it hopefully prevents the publication of poor quality articles and helps authors improve their articles before they are made publicly available. Preprints provide no such protection and allow the publication of very poor quality articles that may harm an author's reputation.

**Corrections to the scholarly record:** Journals are usually good at correcting errors via errata and retractions. Still, these corrections are frequently not communicated in the preprint, and

people are likely to remain unaware that an article has been corrected or retracted.

**Citation dilution and problems:** Journals may find that their citations are reducing as people read and cite the preprint in preference. Equally, some journals may experience an increase in citation even when their readership has not increased. This variation can happen because some journals do not permit citation to preprints, so the author will cite the journal publication when they have read and used the preprint to support their assertions.

**Promoting bad science and political influence:** Because of the lack of quality control and validation of articles before posting, there is the opportunity for bad actors to use preprint servers to post misleading, possibly fraudulent, science. This may be a simple error and mistake, but it could also serve political aims rather than objective science.

**Inflation of results and impact:** Related to the problem of promoting bad science is the potential for authors to use preprint servers to inflate their research. This can be done by salami publications—deliberately creating multiple articles from a single piece of research, each one of them only incrementally adding to the others. Some authors may also use inflated language and skew the discussion and key findings to make a small discovery appear to be more important. Journal editors often moderate this type of inflation, but preprint servers have no such checks in place.

**Version control:** The articles appearing on preprint servers are usually the first version which is changed and hopefully improved and clarified through peer review before final publication. However, readers may not appreciate the changes wrought by the journal and assume that the version in the preprint server is a final—or adequate—version for future research.

### Opportunities and Challenges among Stakeholders

**2020—a watershed year?:** From the launch of the World Wide Web, there has been concern about the proliferation of poor-quality science, readily available without the filter provided by journals. For example, in 2016, the *New England Journal of Medicine* published an editorial stating “On the Internet, speed and simplicity often displace depth and quality, especially on complex subjects” [14]. These concerns led to reluctance in the life sciences and biomedical area to adopt preprints, worried about the potential for dubious findings being made available to the public in a way that made them appear credible. There was also concern that some factions could use preprints to promote particular viewpoints without valid scientific evidence.

However, the pandemic has changed this viewpoint and has increased acceptance of preprints as a valid means of rapid

communication. At the end of 2020, *JAMA* issued an editorial considering the benefits and disadvantages of preprints and concluded that they provide greater benefits to scholarly communication than challenges [15].

Therefore, it can be assumed that preprints are here to stay and that they will become embedded within the scholarly landscape. However, it is important that all stakeholders are fully cognisant of the challenges that remain, and make informed decisions about engagement with preprints.

**Authors and researchers:** Authors must ensure that their use of preprints does not undermine their own reputations. However, the use of preprints is unlikely to prevent acceptance in a journal although they should check if the journal has a stated policy. Preprints usually provide more benefit to authors than problems. Although peer review is an imperfect system and allows for errors to be published [16], researchers need to be fully aware of the lack of quality control over anything appearing in a preprint server, and to use such publications with caution. They should also ensure correct citation to their sources whether they are grey literature (e.g., preprints) or peer reviewed publications.

**Editors:** Journal editors need to have a policy regarding their interaction with preprint servers. There are a few issues to take into account: first, double-blind review is impossible to maintain if authors have posted their article in a preprint server; second, authors should be allowed to cite preprints—to prevent inaccurate citations; and third, editors need to consider what value they add to submissions and how they can differentiate what is published in their journal from the preprint server.

**Publishers:** The existence of preprints leads to the question of how journals will continue when preprint servers are likely to play a greater role in the dissemination of scholarly information. Therefore, it is important for journal publishers to closely consider their own financial and publishing models and how these will change in a world where preprints may become the publication of choice for authors and readers—and increasingly supported by grant funders.

## Conclusion

It has been 31 years since the launch of arXiv.org in 1991. Although this preprint server was confined to the physics and mathematics fields, there are now preprint servers in all academic fields. Preprint servers are popular as preparatory work before submission to journals. Journals adopt a range of attitudes to preprint ranging from uneasy relationships to final merger and overlay. Preprints can benefit authors and editors, but they also have some disadvantages, including information overload, insufficient quality assurance, political influence,

and outsized impact. Since the explosion of use during the COVID-19 pandemic, preprints have become more accepted and mainstream. All journal-related stakeholders need to recognize the challenges that preprints pose and make informed decisions about engagement with them. Journal editors and publishers should have a publicly-available preprint policy. Editors should consider two critical challenges before accepting preprint submissions. First, a double-blind review is impossible because reviewers can find the manuscript and authors on preprint servers. Second, authors should be allowed to cite preprints to maintain a more accurate citation of their sources.

## Conflict of Interest

Pippa Smart has been an editorial board member of *Science Editing* since 2014. She was not involved in the review process. Otherwise, no potential conflict of interest relevant to this article was reported.

## Funding

No funding was received for this work

## References

1. Choi YJ, Choi HW, Kim S. Preprint acceptance policies of Asian academic society journals in 2020. *Sci Ed* 2021;8:10-7. <https://doi.org/10.6087/kcse.224>
2. Yi HJ, Huh S. Korean editors' and researchers' experiences with preprints and attitudes towards preprint policies. *Sci Ed* 2021;8:4-9. <https://doi.org/10.6087/kcse.223>
3. Till JE. Predecessors of preprint servers. *Learn Publ* 2001;14:7-13. <https://doi.org/10.1087/09531510125100214>
4. Ginsparg P. It was twenty years ago today. arXiv:1108.2700v2 [cs.DL] [Preprint]. 2011 [cited 2022 Jan 20]. Available from: <https://arxiv.org/abs/1108.2700>
5. Ginsparg P. Lessons from arXiv's 30 years of information sharing. *Nat Rev Phys* 2021;3:602-3. <https://doi.org/10.1038/s42254-021-00360-z>
6. Xie B, Shen Z, Wang K. Is preprint the future of science? A thirty year journey of online preprint services. arXiv:2102.09066v1 [cs.DL] [Preprint]. 2021 [cited 2022 Jan 20]. Available from: <https://arxiv.org/abs/2102.09066>
7. Anderson KR. bioRxiv: trends and analysis of five years of preprints. *Learn Publ* 2020;33:104-9. <https://doi.org/10.1002/leap.1265>
8. Peters HP. Gap between science and media revisited: scientists as public communicators. *Proc Natl Acad Sci U S A* 2013;110 (Suppl 3):14102-9. <https://doi.org/10.1073/pnas.1212745110>

9. Sarabipour S, Debat HJ, Emmott E, Burgess SJ, Schwesinger B, Hensel Z. On the value of preprints: an early career researcher perspective. *PLoS Biol* 2019;17:e3000151. <https://doi.org/10.1371/journal.pbio.3000151>
10. Chiarelli A, Johnson R, Pinfield S, Richens E. Preprints and scholarly communication: an exploratory qualitative study of adoption, practices, drivers and barriers [version 2; peer review: 3 approved, 1 approved with reservations]. *F1000Research* 2019;8:971 <https://doi.org/10.12688/f1000research.19619.2>
11. Dubben HH, Beck-Bornholdt HP. Systematic review of publication bias in studies on publication bias. *BMJ* 2005; 331:433-4. <https://doi.org/10.1136/bmj.38478.497164.F7>
12. Leopold SS. The dangers of undercooked science and a hungry public. *The Seattle Times* [Internet]. 2021 Nov 19 [cited 2022 Jan 20]. <https://www.seattletimes.com/opinion/the-dangers-of-undercooked-science-and-a-hungry-public/>
13. Mullins M. Opinion: the problem with preprints. *The Scientist* [Internet]. 2021 Nov 1 [cited 2022 Jan 20]. Available from: <https://www.the-scientist.com/critic-at-large/opinion-the-problem-with-preprints-69309>
14. Campion EW, Scott L, Graham A, Prince JM, Morrissey S, Drazen JM. NEJM.org: 20 years on the web. *N Engl J Med* 2016;375:993-4. <https://doi.org/10.1056/NEJMe1610607>.
15. Flanagin A, Fontanarosa PB, Bauchner H. Preprints involving medical research: do the benefits outweigh the challenges?. *JAMA* 2020;324:1840-3. <https://doi.org/10.1001/jama.2020.20674>
16. Tennant JP, Crane H, Crick T, et al. Ten hot topics around scholarly publishing. *Publications* 2019;7:34. <https://doi.org/10.3390/publications7020034>