Contemporary Issue

The preprint debate: What are the issues?

CrossMark

Jaime A. Teixeira da Silva

P. O. Box 7, Miki-cho Post Office, Ikenobe 3011-2, Kagawa-ken 761-0799, Japan

ARTICLE INFO

ABSTRACT

The debate surrounding preprints is increasing. Preprint proponents claim that preprints are a way to shore up trust in academic publishing, that they provide an additional 'quality' screen prior to traditional peer review, that they can assist with the replication crisis plaguing science in part by making negative or contradictory results public, and that they speed up the publishing process because fundamental results can be presented early, serving as timely reports for the purposes of tenure or grant funding. Preprint skeptics and critics claim that preprints may represent a risk and a danger to quality-based academic publishing because they are documents that have not been carefully and thoroughly vetted prior to their release into the public domain. Thus, academics who cite invalid, poorly vetted, or false facts could cause harm, not unlike the unscholarly 'predatory' open access movement. Feedback on work from lesser-known groups, or on less glamorous topics, may be null or worse than from traditional peer review, annulling an initial key objective of preprints. Although there is no widespread empirical evidence or data yet regarding some of these issues, academics should be aware of the ideological, financial, and political tug-of-war taking place before deciding if they wish to publish their important findings as a preprint prior or simultaneous to submitting to a regular journal for peer review.

© 2017 Published by Elsevier B.V. on behalf of Director General, Armed Forces Medical Services.

Introduction: a brief overview of the evolution of the preprint market

The first preprint server, arXiv, launched in 1991, was used by physicists and mathematicians. It was designed as a platform to promote the discussion of unpublished results among academics, and also served to plant an ideological flag of one's academic ideas. In some cases, preprints are submitted simultaneously to a journal for traditional peer review, while in other cases, preprints may represent the final published version of a paper. In the latter case, some authors do not pursue the publication of their work initially presented as a preprint to a traditional peer-reviewed journal because that process can take, in some extreme cases, years to complete, making data sets 'old'. *Nature Proceedings*, a 5-year-old preprint experiment aimed at the field of biomedicine that terminated abruptly in 2012, showed that academics were not embracing preprints, possibly because they could not envision the scholarly merit of a document that had not been properly vetted by specialists through peer review. The underlying concern is that preprints might contain factually incorrect information. bioRxiv emerged in late 2013 as a preprint server to serve biology, and is in fact the fastest growing preprint

server, even though arXiv has the largest number of accumulated preprints. However, funding has entered the preprint equation, distorting the original academic objectives of preprints. For example, researchers funded by the Bill and Melinda Gates Foundation can use an exclusive platform to present their findings as a future (late 2017) preprint server, Gates Open Research. This is similar to researchers funded by the Wellcome Trust who have used an exclusive preprint server, Wellcome Open Research, since November of 2016. Both these preprint servers rely on the f1000Research technical platform, which is leased for a fee. Funding by independent groups and philanthropic organizations is in a boom. Select examples include the Chan Zuckerberg Initiative that funds biorXiv, the European Research Council that funds ArXiv, and the Laura and John Arnold Foundation that funds the 10 topics-based preprint servers hosted by the Center for Open Science. This injection of big money is causing a bullish preprint market to sprout. A massive push toward preprints, spurred by ASAPbio marketing, has caused preprints to become highly politicized, with visible tensions in the public domain.[1] Publishers such as Elsevier or MDPI have their own preprint servers, launched in May of 2017 and 2016, respectively, while SciELO, a primarily South American open access cooperation, is expected to soon launch its own preprint server. Scholars must appreciate the ideological background of an emerging, and potentially profitable, preprint market.

Academic and other issues in the preprint debate

An argument for the use of preprints is that it allows funders to observe the progress of a project in real time, allowing a more realistic opportunity to apply for, and obtain, tenure, funding or promotions. Since preprints can now be cited, through the use of a Digital Object Identifier (DOI), the issue of responsible citation and use of preprints lies in the hands of authors and editors who may choose to use and publish reference to work that has not been thoroughly vetted. Emilie Marcus, the CEO of Cell Press (Elsevier), spurred debate when she claimed that preprints should not be cited, thus not risking "pseudo-article sneaking into credibility through a back door".[2] Her line of view argued that preprints should be observed exclusively as a work in progress made open to the public for open feedback, either to improve the paper itself or the methodologies cited therein, but not to be mistaken with open peer review, which is a more formal and accountable process meant to detect errors prior to becoming a final citable and usable scholarly item. The risk of scooping intellectual ideas such as methods from a preprint is unlikely because a preprint offers time-sensitive evidence of an intellectual claim. However, could preprints be used for intellectual phishing, that is, an attempt to gather intellectual ideas from the public to improve a paper? Even though preprint servers such as biorXiv label preprints as 'not peer-reviewed', the fact is that no rigorous academic scrutiny takes place, with preprints being approved for release into the public domain within as little as 24 h after screening by an advisory board. There is little to prevent academics from citing such documents. Academics are weary of the false academic and 'predatory' Open Access (OA) publishing industry,[3] in which work is published that has not been peer reviewed or

screened in detail for quality. The argument here is that preprints may represent a form of predatory OA behavior, despite their *caveat lector* (i.e., the reader should be aware and/ or cautious) label. Preprint proponents claim that preprints allow for the promotion of replications, confirmatory, contradictory, or negative findings, which generally tend to be marginalized by traditional journals and thus constitute an opportunity to present a wealth of 'lost' or 'hidden' data and information that should be available to academics and the public. But which preprint servers are valid and acceptable? Is there a risk that 'predatory' publishers may establish their own pay-to-publish preprint servers? Even though there is discussion underway about the creation of a centralized preprint service,[4] owners of current preprint servers for the biological sciences are showing little evidence of consolidation.

Are the so-called risks of preprints valid?

To counter the critics of preprints, several counter-arguments can be made. Since preprints carry a DOI, intellectual phishing or scooping can be proved, and intellectual pirates who violate priority claims can be punished accordingly. Even if preprints carry several versions prior to becoming a final published version, either as a preprint or as a paper in a scholarly journal, preprints must always be understood as an incomplete work in progress. Thus, responsible citation must rely on responsible interpretation of a preprint's content. Inculcating a culture of responsible use and citation will be difficult to achieve given the prominent nature of 'predatory' publishing, OA and traditional. A core challenge for preprints is how to harmonize the existence, and use, of peer reviewed work and non-peer reviewed work. In that sense, preprints can serve as a tool for post-publication peer review to refute erroneous literature. In a surprisingly opaque move in April–May of 2017, biorXiv expanded its range of papers that could be accepted for publication as a preprint, accepting preprints related to publishing policy. This action indicated that preprints are still in a highly fluid state of evolution, and new risks evolved as a result of this action by biorXiv: (a) will one day anything be published as a preprint, that is, how does one standardize an 'acceptable' quality threshold and filter relevant from junk or pseudoscience? (b) Why are letters, perspectives, or commentaries about academic issues, categories that are acceptable in many peer-reviewed journals, not acceptable by preprint servers such as biorXiv? (c) How do preprint servers ensure that there is voluntary public feedback on preprints, without biasing sensationalist preprints at the expense of less glamorous ones? (d) How can academics trust the owners of preprint servers when decisions are made in their best self-serving interests and possibly not in the best interests of academia?

Conclusions

Preprints are not just a hot topic. They offer an alternative academic platform to present data that might not be easily accepted in traditional publishing venues since they may show contradictory or refutable data.[5] They may also serve to

anchor intellectual claims as citable items. Several core assumptions, however, might not be true: not all academics will use preprints responsibly and not all preprint servers might be academically valid. Given the rising level of 'fake' in academics, it is to be expected that an element of fraud, deception and unscholarly behavior may also begin to enter the preprint landscape. In order to deal with this potential threat, preprint servers must have clear ethical and retraction policies in place, and these must be enforced. Academics must also be aware that there is a political and economic struggle in the emergent preprint market, with some of the most passionate preprint proponents also being some of the most passionate critics of status quo publishing, that is, preprints are a threat to the multi-billion dollar publishing oligopoly.[6] For example, there are no costs for authors to post a preprint on biorXiv, unlike exorbitant OA article processing fees charged by some leading OA journals. This rush to publish work as a free OA document with a citable identifier, the DOI, may also invite a wealth of bad, weak, or poor science. To reduce this risk, given the centrality of preprints in the open science movement, preprints should also have open data policies, that is, preprints cannot be published unless the data sets are also placed in the public domain.

Conflicts of interest

The author has none to declare.

REFERENCES

1. Teixeira da Silva JA. The preprint wars. *AME Med J*. 2017;2:74. http://dx.doi.org/10.21037/amj.2017.05.23.
2. Marcus E. *Let's Talk About Preprint Servers*. 2017. http://crosstalk.cell.com/blog/lets-talk-about-preprint-servers 03.06.16; Last Accessed: July 26.
3. Das AK. Publish and flourish: take the road less travelled!. *Med J Armed Forces India*. 2017;73:178–180.
4. Polka J. *New Developments and Plans for the Central Service RFA and Governing Body*. 2017. http://asapbio.org/category/central-service 10.05.17; Last Accessed: July 26.
5. Berg JM, Bhalla N, Bourne PE, et al. Preprints for the life sciences. *Science*. 2016;352:899–901.
6. Molteni M. *Biology's Roiling Debate Over Publishing Research Early*. 2017. https://www.wired.com/story/biologys-roiling-debate-over-publishing-preprint-research-early/ Last Accessed: 26.07.17.