A web of tidings

Preprints provide an efficient way for scientific communities to share and discuss results. We encourage authors to post preprints on arXiv, bioRxiv or other recognized community preprint platforms.

Nature journals have welcomed submissions that also exist as preprints for over two decades¹, but we still hear surprisingly often that scientists are unsure about our policy. To highlight more clearly that we recognize the important role of preprint posting in the process of scientific discourse, *Nature Machine Intelligence* offers authors the option to add a link back to their arXiv or bioRxiv preprint from the published paper, visible to all readers. Readers can find two examples of this function in the current issue — the Articles by [Brendon Lutnick et al.](#) and [William Severa et al.](#)

A well-known story is that the first preprint server arXiv had its origin in high-energy physics. Paul Ginsparg, a physicist at Los Alamos National Laboratory in New Mexico, decided to launch an electronic bulletin for sharing unpublished papers with colleagues and friends². It was 1991 and the main mode of exchanging papers was mailing paper copies to one another. Even though it hardly seems long ago for some of us, this was a time of big, noisy desk computers, floppy disks and a world without a wide web.

But with the next information revolution around the corner and the rapid uptake of the world wide web during the rest of the decade, arXiv was well positioned to serve a physics community eager to benefit from the rapid flow of information. It expanded quickly, taking other fields on board: computer science, mathematics, economics and several more. At arXiv's 20-year anniversary in 2011 Paul Ginsparg noted³, presumably with a mixture of pride and concern, that no community that had adopted arXiv had renounced it. The server's growth has been unstoppable and currently a staggering 1.5 million preprints exist on arXiv. In 2018, 140 thousand new submissions were posted, up by 14% from 2017.

Different scientific communities have different ways of disseminating and

Credit: Gary Waters/Alamy Stock Photo

discussing their findings prior to peer-reviewed publication: preprint posting has become popular in many fields, while others stick to conferences or other platforms. The reasons scientists give for posting preprints centre on being able to share results without delay with experts and getting feedback from the community, putting a time-stamp on the work to establish priority, and increasing visibility.

Already in 2011, the increase in arXiv preprint posting in the field of computer science was noted. The growth continued and in 2018, the field accounted for 26% of new submissions, second after physics. There is no doubt that the rapid interactions allowed by preprint sharing contributed to the quick advances in the field of machine learning in the past decade. But the daily output in this field and related ones has made it a challenge for anyone trying to keep up with latest developments. In 2016, arXiv began planning for a significant overhaul³, first to improve the infrastructure to deal with the high volumes of preprints, but also to offer a more user-friendly

interface. External tools to browse and filter content now also exist and a popular one is 'arxiv sanity preserver'⁴, which attempts to tame the flood of preprints, thereby 'preserving the sanity' of anyone trying to keep up with new papers on arXiv. This tool filters by subject terms and popularity, provides a useful interface for skimming through content and offers several ways to sort papers. It is running live on the web for the subjects in computer science and statistics closely associated with artificial intelligence and machine learning — from a quick count on www.arxiv-sanity.com, currently close to a 100 preprints are posted per day in these topics.

Some other fields, such as the life sciences, have been slower to embrace open preprint sharing. However, [bioRxiv](#) came into existence in 2013, and quickly became highly popular. Unlike arXiv, it offers a commenting function and provides metrics on article usage and attention. [ChemRxiv](#) launched in 2017 to serve the chemistry community.

With the year-on-year growth in scientific output, an ongoing challenge is how to filter papers, provide quality control, integrate papers with data, code and other tools, and make papers reach intended and new audiences. Such questions, tangled up with the future of science publishing, have been discussed for at least two decades. While there are no simple answers, the availability and wide acceptance of central repositories for preprints ensures open scientific discourse. □

Published online: 11 February 2019
<https://doi.org/10.1038/s42256-019-0027-2>

References

1. *Nature* **434**, 257 (2005).
2. Ginsparg, P. *Nature* **476**, 145–147 (2011).
3. Van Noorden, R. *Nature* **534**, 602–602 (2016).
4. *GitHub* <https://github.com/karpathy/arxiv-sanity-preserver> (2018).