

Biologists debate how to license preprints

Flood of online manuscripts generates confusion about terms for distribution and reuse.

Lindsay McKenzie

16 June 2017

Biology's zeal for preprints — papers posted online before peer review — is opening up a thorny legal debate: should scientists license their manuscripts on open-access terms?

Researchers have now shared more than 11,000 papers at the popular bioRxiv preprints site. But where some researchers allow their bioRxiv manuscripts to be freely redistributed and reused, others have chosen to lock them down with restrictive terms (see 'Licence confusion').

That split concerns Jessica Polka, the director of ASAPBio, a grass-roots organization that advocates for preprints. Polka wants authors to choose permissive licences that would allow anyone to share or adapt the work, even for profit. "We want to maximize the public good of preprints," she says. The US National Institutes of Health also encourages open-access preprint licences.

But some researchers are wary of the approach, and that needs exploring, Polka says. So on 16 June, ASAPBio announced the creation of a taskforce to help understand attitudes towards preprint licensing. "It's viewed as kind of a fraught issue," says Dick Wilder, associate general counsel at the Global Health Program of the Bill and Melinda Gates Foundation in Washington DC. Wilder is chairing the taskforce, which includes researchers, legal experts, funding agencies and publishers.

Open or closed

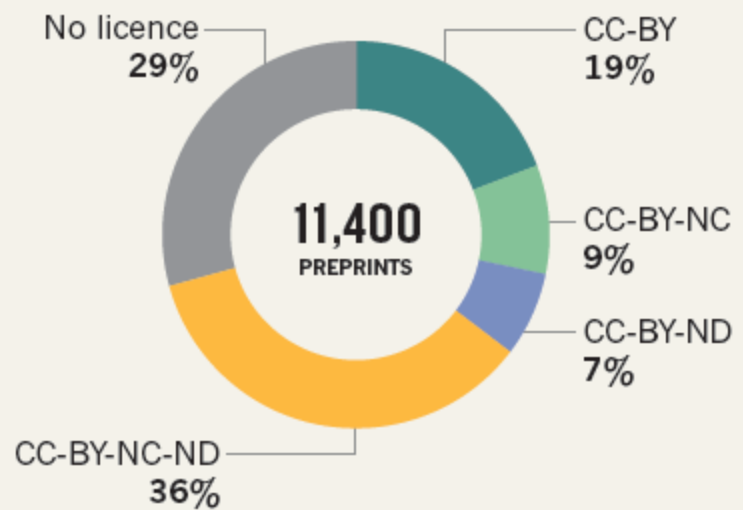
Authors automatically hold copyright on their preprint manuscripts. But by adding a licence to their preprint, they tell others how their work can be reused, and under what circumstances. At bioRxiv, for instance, scientists can choose from a spectrum of licences. This ranges from the most liberal CC-BY — which allows anyone to reuse work for any purpose, as long as they credit its source — to CC-BY-NC-ND, which bars commercial use and any 'derivative' works, such as translations or tools that distribute annotated versions of papers. (CC refers to Creative Commons, a non-profit organization in Mountain View, California, that constructed the licence terms.)

Scientists may be troubled by the idea of their preprints being repackaged and sold for profit, but Daniel Himmelstein, a data scientist at the University of Pennsylvania in Philadelphia and a member of the ASAPBio taskforce, says that he would welcome the free publicity if his work was shared this way. And he says that most entrepreneurs looking to reuse preprints are developers who could benefit researchers by creating new tools for displaying, interacting with or sharing the work.

According to statistics from bioRxiv, 29% of authors have decided to append no licence at all to their work. On the site, these are labelled: "All rights reserved. No reuse allowed without permission."

LICENCE CONFUSION

At the bioRxiv preprint server, scientists have chosen a mixture of licences for their preprints; around 29% have chosen none at all, which means they cannot be redistributed or reused without express permission.



©nature

CC, Creative Commons; BY, attribution; NC, non-commercial; ND, no derivative.

Source: bioRxiv

Saskia Hagenaars, a geneticist at Kings College London says that her team chose this option because “we don't want people freely using the non-peer reviewed versions of our papers”.

Himmelstein, who follows bioRxiv licensing on his blog Satoshi Village, says that he would like bioRxiv to remove that option. “I think it's counter to the minimum requirements of healthy science literature,” he says. Choosing no licence means that if bioRxiv disappeared, these papers could not be reposted elsewhere without the express permission of the authors, he adds.

Related stories

- BioRxiv preprint server gets cash boost from Chan Zuckerberg Initiative
- Legal confusion threatens to slow data science
- Researchers opt to limit uses of open-access publications

[More related stories](#)

Related stories

- BioRxiv preprint server gets cash boost from Chan Zuckerberg Initiative
- Legal confusion threatens to slow data science
- Researchers opt to limit uses of open-access publications

[More related stories](#)

Text-mining confusion

The bioRxiv website states that just by uploading papers to bioRxiv — no matter what licence is chosen — authors consent to text-mining of their work, a technique in which software crawls over thousands or millions of downloaded papers to pull out scientific insights, or to annotate key scientific terms. But researchers who want to text-mine preprints aren't always sure whether they can legally do this on restrictively licensed manuscripts, or whether they can redistribute the corpus of papers that they have been mining. John Inglis, the co-founder of bioRxiv, says that he's received queries about what is and isn't permitted, despite the site's stated terms.

At the physics preprint server, arXiv, the most popular option is a default licence that gives arXiv a non-exclusive right to distribute articles — and says little else. Paul Ginsparg, a physicist at Cornell University in Ithaca, New York, says that all arXiv preprints can be freely text-mined, relying on the principle that text-mining is ‘fair use’ of the papers and does not breach US copyright law. It's not entirely clear, however, whether scientists in other nations can depend on this principle; other jurisdictions have differing views on whether text-mining breaches copyright, notes Michael Carroll, who directs the Program on Information Justice and Intellectual Property at the American University in Washington DC.

One reason for researchers' hesitancy to choose open licences may be that some journals frown on them. Giulio Caravagna, a computational biologist at the University of Edinburgh, UK, decided not to openly licence his bioRxiv preprint because it gave his team "full rights to proceed further with submission to any journal that we want to target". The *Proceedings of the National Academy of Sciences USA (PNAS)*, for instance, says it will only publish papers arising from preprints that don't have CC licenses, because it feels that these are not compatible with its own licensing terms. But Himmelstein has found a dozen CC-BY preprints that led to work published in *PNAS* — and the publisher says it has never enforced its rule.

Jesse Bloom, an evolutionary and computational biologist at the Fred Hutchinson Cancer Research Center in Seattle, Washington, says that he didn't know about *PNAS*'s policy, but that if the journal were to reject his CC-BY preprints because of licensing terms, he'd view it as "their problem rather than ours". In his view, scientists and scientific journals "should focus on the quality of the work rather than spend their time worrying about publication licences".

Nature [doi:10.1038/nature.2017.22161](https://doi.org/10.1038/nature.2017.22161)