Making Science More Open Is Good for Research—but Bad for Security

The open science movement pushes for making scientific knowledge quickly accessible to all. But a new paper warns that speed can come at a cost.

Grace Browne Apr 22, 2022 7:00 AM

Photograph: Paul Taylor/Getty Images

For decades, scientific knowledge has been firmly shut behind the lock and key of eye-wateringly expensive journal paywalls. But in recent years a tide has been turning against the rigid, antiquated barriers of traditional academic publishing. The open science movement has gained momentum in making science accessible and transparent to all.

Increasingly journals have published research that's free for anyone to read, and scientists have shared their data among each other. The open science movement has also entailed the rise of preprint servers: repositories where scientists can post manuscripts before they go through a rigorous review by other researchers and are published in journals. No longer do scientists have to wade through the slog of the peer-review process before their research is widely available: They can submit a paper on bioRxiv and have it appear online the next day.

But a new paper in the journal *PLoS Biology* argues that, while the swell of the open science movement is on the whole a good thing, it isn't without risks.

Though the speed of open-access publishing means important research gets out more quickly, it also means the checks required to ensure that risky science isn't being tossed online are less meticulous. In particular, the field of synthetic biology—which involves the engineering of new organisms or the reengineering of existing organisms to have new abilities—faces what is called a dual-use dilemma: that while quickly released research may be used for the good of society, it could also be co-opted by bad actors to conduct biowarfare or bioterrorism. It also could increase the potential for an accidental release of a dangerous pathogen if, for example, someone inexperienced were able to easily get their hands on a how-to guide for designing a virus. "There is a risk that bad things are going to be shared," says James Smith, a coauthor on the paper and a researcher at the University of Oxford. "And there's not really processes in place at the moment to address it."

While the risk of dual-use research is an age-old problem, "open science poses new and different challenges," says Gigi Gronvall, a biosecurity expert and senior scholar at the Johns Hopkins Center for Health Security. "These risks have always been there, but with the advances in technology, it magnifies them."

To be clear, this has yet to happen. No dangerous virus or other pathogen has been replicated or created from instructions in a preprint. But given that the potential consequences of this happening are so catastrophic—like triggering another pandemic—the paper's authors argue that even small increases in risk are not worth taking. And the time to be thinking deeply about these risks is now.

During the pandemic, the need for preprint servers was thrown into sharp relief—crucial research could be disseminated far more quickly than the traditionally sluggish journal route. But with that, it also means that "more

people than ever know now how to synthesize viruses in laboratories," says Jonas Sandbrink, a biosecurity researcher at the Future of Humanity Institute at the University of Oxford and the other coauthor of the paper.

Of course, just because research is published in a journal instead of a preprint server doesn't mean it's inherently risk-free. But it does mean that any glaring dangers are more likely to be picked up in the reviewing process. "The key difference, really, between journals and the preprint server is the level of depth that the review is going into, and the journal publication process may be more likely to identify risks," says Smith.

The risks of open publishing don't stop at biological research. In the AI field a similar movement toward openly sharing code and data means there's potential for misuse. In November 2019, OpenAI announced it would not be openly publishing in full its new language model GPT-2, which can independently generate text and answer questions, for fear of "malicious applications of the technology"—meaning its potential to spread fake news and disinformation. Instead, OpenAI would publish a much smaller version of the model for researchers to tinker with, a decision that drew criticism at the time. (It went on to publish the full model in November of that year.) Its successor, GPT-3, published in 2020, was found to be capable of writing child porn.

Two of the biggest preprint servers, medRxiv, founded in 2019 to publish medical research, and bioRxiv, founded in 2013 for biological research, publicly state on their websites that they check that "dual-use research of concern" is not being posted on their sites. "All manuscripts are screened on submission for plagiarism, non-scientific content, inappropriate article types, and material that could potentially endanger the health of individual patients or the public," a statement on medRxiv reads. "The latter may include, but is not limited to, studies describing dual-use research and work that

challenges or could compromise accepted public health measures and advice regarding infectious disease transmission, immunization, and therapy."

From bioRxiv's outset, biosecurity risks were always a concern, says Richard Sever, one of bioRxiv's cofounders and assistant director of Cold Spring Harbor Laboratory Press. (Sever was a peer reviewer of Smith and Sandbrink's paper.) He jokes that in the early days of arXiv, a preprint server for the physical sciences launched in 1991, there were worries about nuclear weapons; with bioRxiv today the worries are about bioweapons.

Sever estimates bioRxiv and medRxiv get about 200 submissions a day, and every one of them is looked at by more than one pair of eyes. They get "a lot of crap" that is immediately tossed out, but the rest of the submissions go into a pool to be screened by practicing scientists. If someone in that initial screening process flags a paper that may pose a concern, it gets passed up the chain to be considered by the management team before a final call is made. "We always try to err on the side of caution," Sever says. So far nothing has been posted that turned out to be dangerous, he reckons.

A few papers have been turned away over the years because the team thought they fell into the category of dual-use research of concern. When the pandemic arrived, the issue became all the more urgent. The two servers published more than 15,000 preprints on Covid-19 by April 2021. It became an internal wrangle: Do the high life-or-death stakes of a pandemic mean they are morally required to publish papers on what they call "pathogens of pandemic potential"—like Sars-CoV-2—which they might have traditionally turned away in the past? "The risk-benefit calculation changes," Sever says.

But while bioRxiv and medRxiv have taken steps to deeply consider whether their output may pose a biosecurity risk or compromise public health advice,

other servers and repositories may not be as fastidious. “Data and code repositories are pretty much fully open—anyone can post whatever they want,” Smith says. And Sever makes the point that if they do turn away a paper, it doesn’t mean it can’t end up online elsewhere. “It just means they can’t put it online with us.”

In their paper, Smith and Sandbrink make recommendations to safeguard against potential biosecurity risks. For instance, when researchers post data and code in repositories, they could be required to make a declaration that that data isn’t risky—though they acknowledge that this requires a level of honesty one wouldn't expect from bad actors. But it is an easy step that could be implemented right away.

On a longer timescale, they recommend following the model that’s been used in the sharing of patient data, such as in clinical trials. In that situation, data is stored in repositories that require some form of access agreement in order to gain entry. For some of this data, the researchers themselves don’t actually ever get to see it; instead it gets submitted to a server that analyzes the data away from the researchers and then sends back the results.

Finally they advocate for preregistering your research, already a pillar of open science. Put simply, preregistration means writing down what you intend to do before you do it, and keeping a record of that to prove that you actually did it. Smith and Sandbrink say it could offer an opportunity for biosecurity experts to assess potentially risky research before it even happens and give advice on how to keep it secure.

But it’s a tough balancing act to achieve, Sandbrink admits, in avoiding over-bureaucratizing the process. “The challenge will be to say, how can we make things as open as possible and as closed as necessary, whilst also ensuring equity and ensuring that it’s not just the researchers at Oxford and

Cambridge that can access these materials." There will be people around the globe whose credentials might be less clear, Sandbrink says, but who are still legitimate and well-intentioned researchers.

And it would be naive to pretend that a paywall or journal subscription is what impedes nefarious actors. "People who want to do harm will probably do harm," says Gabrielle Samuel, a social scientist at King's College London whose research explores the ethical implications of big data and AI. "Even if we have really good governance processes in place, that doesn't mean that misuse won't happen. All we can do is try to mitigate it."

Samuel thinks mitigating risky science doesn't begin and end at the publishing stage. The real issue is that there's no incentive for researchers to carry out responsible research; the way scientific journals and funding bodies have a tendency to favor new, exciting research means the more boring, safer stuff doesn't get the same support. And the hamster-wheel nature of academia means scientists "just don't have the capacity or chance of being able to have the time to think through these issues."

"We're not saying that we want research to go back to a model of being behind paywalls, and only being accessible to very few individuals who are privileged enough to be able to afford access to those things," Smith says. But it's time for open science to be reckoning with its risks, before the worst happens. "Once something is publicly available, fully, openly—that is a pretty irreversible state."

- You should always question the default settings
- *Battle Kitty* stretches the limits of Netflix's tech
- The rise of brand-new secondhand EVs
- 👁️ Explore AI like never before with our new database
- 🎧 Things not sounding right? Check out our favorite wireless headphones, soundbars, and Bluetooth speakers