Preparing for Preprints

Bernd Pulverer

Preprints reduce delays in sharing research results and increase the amount and diversity of data available to the scientific community. Support of this communication mechanism through appropriate policies by journals, funders and institutions will encourage community engagement. Widespread adoption would benefit both individual scientists and research, and it might improve publishing in scientific journals. Preprints are one step towards an Open Science future.

Preprint promise

Scientific research is a communal activity and progress depends on cooperation—the proverbial “my work is built on the shoulders of giants...” should be complemented with “... and continual discussion and collaboration with colleagues”. For a long time, scientific conferences have provided a forum for exchanging emergent research. Traditionally, this exchange happened before publication—presenting retrospectives of published work at conferences would have been considered gauche. Now, competitive pressures and accelerating discovery rates risk turning collaborative workshops into marketing platforms to showcase research achievements. Preprints offer a formal extension of conferences for the open exchange of prepublication research, with the benefit of a global reach and the formal documentation of research output.

In the physical sciences, preprints have been *de rigueur* for a quarter of a century—the majority of research across a wide spectrum of disciplines is first posted on [arXiv](#) as non-peer-reviewed manuscripts (Ginsparg, 2016); in engineering and computer science, conference proceedings fulfill a similar role. Thus, more than 100,000 research

manuscripts annually on [arXiv](#) are open to comments from colleagues, which fosters collaboration and helps scientists to improve manuscripts before they are submitted to a peer-reviewed journal.

Because preprints are stably archived, dated, and citable, they provide formal evidence for research activity, often well in advance of the published paper. Given the widespread use of [arXiv](#), researchers are aware of what colleagues are working on and can adapt their own research accordingly. If someone exploits privileged information to their own advantage, the community will know about it as the evidence is there for all to see. In physics, preprints are therefore seen as leveling the playing field, rather than providing a competitive advantage to those who keep their findings close to their chest.

[arXiv](#) coexists symbiotically with journals: Most research posted as a preprint is ultimately destined for formal peer review and publication (Ginsparg, 2016). Often, preprints are first revised and reposted as a new version either before or in parallel with journal submission. As a result, more mature work is submitted to journals and the peer review process is consequently more straightforward. Publishing is distinctly less painful in physics. Journals continue to thrive as they constitute the formal “point of record” for a research project and carry weight in research assessment.

Open science

Peer review is a powerful quality assurance mechanism. Undoubtedly, the system is subject to stresses in the biological sciences both on account of the exponential expansion of the number of research papers and the increasing pressure to publish in only a small subset of journals that meet ambitious bibliometric targets set by research institutions and funders (Pulverer, 2015a).

Valuable results that do not make it into prestigious journals fall below the radar of research assessment and are often not visible to other researchers. Much research is never published in the first place—this can include perfectly robust and reproducible negative data, confirmatory data, and data at odds with published findings. It can also include descriptive datasets and other resources. A fundamental tenet of hypothesis-driven research is that the hypothesis is tested experimentally. This can and should result in many hypotheses failing, but this does not negate the value of the results. Often these datasets are in fact particularly compelling, yet such “negative” data do not succeed in the selection process of journals that focus on impactful science.

It is entirely understandable that journals that aim to reach a broad readership select research of broad interest that presents a sizeable conceptual advance. However, if reliable data that do not fit these editorial criteria remain invisible, the efficacy of scientific research will suffer. Scientists will repeat unreported experiments, while others rely on results that may not stand the test of time, because contradictory data are not shared. The literature is littered with such papers and the argument that “everyone in the community knows which papers are not to be trusted” will only ever apply to the most visible findings as more than 1 million papers are published every year.

To be sure, there are many well-respected community journals happy to publish high-quality data of more specialist interest. There are also journals that have explicitly dropped “conceptual advance” as an editorial criterion, a concept first applied by *PLOS One* which has since spread rapidly, partially because it can function as an engine of open access publishing business models. Furthermore, preprints add to a number of existing platforms that allow

sharing of more or less structured data, such as [Dryad](#), [figshare](#), [Zenodo](#), and [BioStudies](#) (McEntyre *et al.*, 2015). Incentives to engage with these platforms remain limited—the ball is in the court of funders and institutions to change their research assessment processes if they value this research (Pulverer, 2015a).

Solving the time discontinuum

Preprints diminish the delay between discovery and dissemination of results. Journal-based publishing struggles to keep up with the rapid speed of biomedical research. This is most apparent where speed is dictated by public health, for example, to address the latest pandemic threat. A number of journals, including this one, have committed to release relevant data openly and with minimal delay (Isom, 2016).

Still, most research papers need many months to navigate the editorial process. The problem is compounded by pressure to publish at all costs in high-impact journals, leading to systemic serial submissions. This time gap is not merely frustrating to researchers, but can also compromise the novelty of their work and lead to redundant research. The EMBO publications and a few others have managed to minimize these delays by applying peer review only to manuscripts with a real chance of publication, by restricting revision to realistic, essential, and well-defined experimentation, and by offering manuscript transfers to other journals where appropriate (Pulverer, 2014). Nonetheless, it is essentially impossible to publish in a high-quality peer-reviewed publication in less than a handful of months without compromising on quality. Preprints address this problem by moving the initial dissemination of findings closer to their discovery.

Commenting and quality control

A central promise of preprints is that it improves shared research on account of constructive commenting in an open form or via direct peer-to-peer communication (Ginsparg, 2016). The biology preprint platform [bioRxiv](#) reports that 10% of preprints receive open comments, but most discourse happens via social media and e-mail, and one-third of the preprints are revised and 60% subsequently published.

The jury is still out on how to best ensure preprints do not allow dissemination of misleading, ethically compromised, plagiarized, false, or outright dangerous information. Currently, minimal checks are applied at [bioRxiv](#) to ensure posts conform to basic standards and [arXiv](#) is working on automated screening (Ginsparg, 2016). Clearly, applying peer review to weed out flawed contributions would defeat the purpose. Peer review by selective journals will remain an important quality assurance mechanism; manuscripts posted on [bioRxiv](#) automatically forward-link to papers published in *EMBO Press* journals as the point of record.

Staking claims or salami slicing

Preprints are formally citable and provide a time stamp for claims of discovery. All EMBO publications encourage citation of preprints where appropriate in the reference list. We are working on a standard to ensure that citations to preprints are on the same footing as references to peer-reviewed papers, while also signalling “non-peer reviewed—proceed with due caution”. In our view, it is better to share science in two forms: *carefully peer reviewed* and *not peer reviewed* than the “*modest peer review and quality control*” that many journals can muster now in the face of pressure to publish more at lower costs.

What protects preprinters from being scooped? Conversely, what prevents preprinters from posting work prematurely to claim a discovery? In physics, the time stamp of preprints protects from scooping and the community disfavors cooperation with those who exploit the system (Ginsparg, 2016). As Ron Vale, initiator of the [ASAPbio](#) coalition for preprints, notes “why would I post lower quality work that could compromise my reputation?” In contrast to presenting unpublished work at conferences to an unknown audience who often tweet results in real time, preprints formally document research advances.

Extended scooping protection

Many reputable journals allow or [encourage](#) preprint posting of primary research.

EMBO publications will not consider preprints in the evaluation of the conceptual advance of submitted manuscripts and encourages posting before—or in parallel

with—submission to the journal. Reviews, commentaries, and versions of articles that have been revised to address referee comments and accepted for publication or published, should not be posted.

The EMBO publications have long promulgated their *scooping protection* policy for submitted manuscripts and during revision. EMBO Press has now gone one step further to [extend scooping protection](#) to the day of posting the preprint version of the manuscript (<http://emboj.embopress.org/authorguide#preprintservers>). We hope this will encourage researchers to post their work as a preprint. We hope in particular that it encourages younger scientists to have the confidence to post preprints – for them the stakes are highest, but they also gain the most as preprints allow them to document research formally and in time for the next application for a grant or position.

Back to the future

Preprints may become an important part in an Open Science future, encouraging the open and transparent sharing of research findings with minimal delay. Preprints that are broadly visible and that matter for research assessment will depressurize the rush to publish at all costs, allowing quality journals to invest time and effort to add reliability and reproducibility assurances to research findings through careful peer review and prepublication quality control and curation processes (Pulverer, 2015a,b; Vale & Pulverer, 2016). If preprints should attain the dominant role they have in physics, publishing papers in journals may remain attractive only in journals that add real value to the scientific communication process. This may spur investment or contract the publishing landscape, freeing resources.

In need of carrots

Why is everyone using preprints in physics, while anxiety still trumps in biology? Even though physical disciplines can be more contained than sprawling fields such as “cancer biology”, competition can be as high and funding as competitive. One crucial difference is that physicists have grown up with [arXiv](#): it is universal and fully embedded in the scientific process. Since everyone is using it, the playing field is level.

A crucial incentive to post preprints ought to be that researchers are thus also

able to view proportionally much more data from colleagues. There is a clear net benefit in sharing one's own work more systematically if only enough others do it.

Nonetheless, without appropriate incentives, biomedical researchers will remain reluctant to share results. To get to this point, it will be necessary to provide tangible incentives. EMBO Press's "extended scooping protection" is one.

It is ultimately up to funders and institutions—EMBO encourages applicants for fellowships to include preprints (see http://www.embo.org/documents/LTF/LTF_Guidelines_for_Applicants.pdf).

Scientific exchange has to become more cooperative and open. Even in a hypercompetitive environment, this will benefit everyone. Let us see whether the biosciences community can pull together to bolster prepublication exchange and thus catch up with physics. It will take reassurances and incentives from journals and funders to support those scientists who "risk it".

References

- Ginsparg P (2016) Preprint Déjà Vu. *EMBO J* 35: 2620–2625
- Isom H (2016) Sharing data during Zika and other global health emergencies. <https://wellcome.ac.uk/news/sharing-data-during-zika-and-other-global-health-emergencies>

- McEntyre J, Sarkans U, Brazma A (2015) The BioStudies database. *Mol Syst Biol* 11: 847
- Pulverer B (2014) EMBO Press – a new way to publish. *EMBO J* 33: 1–2
- Pulverer B (2015a) Dora the brave. *EMBO J* 34: 1601–1602
- Pulverer B (2015b) Reproducibility blues. *EMBO J* 34: 2721–2724
- Vale R, Pulverer B (2016) Are biologists ready for preprints? *EMBOencounters* 33: 7 http://www.embo.org/documents/news/encounters/EMBO_encounters_issue33.pdf