

Accepted manuscript

# Recent developments in scholarly publishing to improve research practices in the life sciences

Stephen J Eglen ([orcid](#)), Ross Mounce ([orcid](#)), Laurent Gatto ([orcid](#)), Adrian M Currie ([orcid](#)), Yvonne Nobis ([orcid](#)).

## Introduction

In some ways, scholarly publishing has not changed much in the last ten years. Publishing in prestigious top-tier journals is still perceived as critical for career progression (especially gaining promotion and grants). Likewise, journal metrics continue to dominate in the evaluation of a paper's research, rather than the paper's contents [1,2]. Against the backdrop of highly competitive job and grant markets, factors such as these encourage narrow research agendas and tie researchers (particularly in early career) to placing work in exploitative publishers who draw significant funds from academic work. Further, standard publishing criteria, especially for instance on publishing statistically significant, positive results, creates biases across published studies. However, there are several reasons for optimism that the nature of scientific publishing will improve. Here we outline some recent developments that we believe will improve the working environment and career prospects for life scientists.

## Preprints

Since 1991 ArXiv [3] has become a standard tool for physicists to rapidly disseminate their research findings. Although on the surface ArXiv does not provide much more than a collection of PDFs grouped via topic, publishing there is now considered key for establishing priority in certain fields (subsequent journal publication is still the norm). It was initially assumed that biologists would not adopt a preprint culture: publishing a preprint might prevent subsequent publication in a top-tier journal, or leading to scooping by another group. There is some historic justification of these concerns: an NIH experiment in preprints was effectively halted in the 1960s by journals' refusal to accept preprints for submission [4].

BioRxiv, launched in 2013, has overcome these concerns. Researchers in such diverse areas as ecology, neuroscience and genomics are uploading preprints and choosing to share their work ahead of publication. There are many reasons for this usage:

- Sharing work before submitting to a journal allows for community feedback.
- Sharing work at the time of submission means that the community can read the work months (or years) before the work eventually appears in print.
- Journal editors browse bioRxiv and suggest relevant papers be submitted to their journal.
- BioRxiv preprints can be transferred rapidly to journal submission systems rather than going through (an often lengthy) direct submission to the journal.
- Several funding agencies, including NIH and UKRI, allows preprints to be listed on CVs and cited in grant applications.

Several other preprint servers are available, in particular [PeerJ Preprints](#), [OSF Preprint servers](#), and [preprints.org](#), although to date BioRxiv is the dominant repository. Unlike a few years ago, most journals in the life sciences no longer see prior appearance in a preprint repository as a block to formal publication.

## Overcoming the reproducibility crisis

According to recent surveys, life scientists across many domains believe there is a “reproducibility crisis” in science: i.e. many key findings in publications are either not independently verified, or fail verification when it is attempted [5]. The traditional publishing system must take some responsibility for these low levels of reproducibility, as authors feel under intense pressure to publish to avoid being scooped (see below). Further, space limitations imposed by print journals inhibit adequate method descriptions, with key details relegated to supplementary information which is rarely scrutinised to the same degree as the main paper. However, here we list four encouraging developments that should promote reproducibility.

**Preregistration papers** typically describe the introduction and methods sections of a study, and are peer-reviewed *before* the study is actually performed [6]. This allows reviewers to improve the study design and commits researchers to hypotheses that they wish to study along with their statistical analysis. Once the pre-registration study is approved, it is then published. After the research is completed, another paper can be submitted to the same journal which describes the results of the study using the pre-registered methods. (Additional findings can be reported, but are clearly marked as such.) Reviewer and editorial decisions on whether to accept the second paper is made on the technical correctness of the paper rather than the importance or novelty of the results. Preregistration is most prevalent today in psychology; The Center for Open Science Preregistration Challenge <https://cos.io/prereg/> is helping to popularise this notion more broadly. Initial analysis of results from preregistered papers indicates, perhaps reassuringly, a marked increase in null results reported [7].

**Stronger data sharing policies and community expectations.** Both funders and journals are now making stronger statements about what research materials (data, computer programs, reagents) should be shared upon publication of the corresponding articles. Although these policies should increase data availability and reuse, compliance rates are fairly low [8]. Given that it can take considerable time and effort (for both researchers and journals) to ensure data is appropriately shared, these low-uptake rates are perhaps expected. To reward authors for this work, “data papers” (a paper that simply describes the data) are becoming more prominent, e.g. in journals like *Scientific Data* and *Gigascience*. Journals are also providing guidelines for authors to follow which should improve reproducibility and transparency [9].

**Reproducible manuscripts** are documents that contain the main text as well as the code to generate tables, figures and results [10]. However, even though researchers have been committed to reproducible research, the reproducibility of the final outputs were generally broken upon submission to journals. Researchers have released reproducible versions of their work in parallel to the journal articles [11,12]. Recently, some journals have moved closer to publishing reproducible manuscripts, by working towards a reproducible document stack [13] or supporting reproducible figures [14]. Further, some journals are experimenting with re-running of code “in the cloud” using services such as Cloud Ocean [15].

**Replicability studies** Given the (often intense) competition to be first to publish in some areas of biology, being “scooped” on publishing a particular result can be doubly damaging. Not only does someone else publish the result first, your manuscript is often no longer regarded as novel, and thus not worthy of publication by top-tier journals. However, given that science relies on the gradual accumulation of evidence over a large body of papers, such replication studies are valuable. In January 2018, *PLOS Biology* announced that they would consider for publication those papers that “confirm or extend a recently published study” [16]. In a similar vein, replicability studies can provide clear evidence to evaluate controversial findings [17].

## Other recent innovations of note

**ORCID** provides a persistent, unique digital identifier for researchers which can help in linking scholarly outputs to an individual and thus automatically curate a list of works, not just papers, in one place. Such identifiers should also reduce mistaken identity. Many journals now require that at least one author verifies their identity as author using ORCID [18].

**DORA** [19] is a declaration for individuals and institutions to commit to evaluating research based on its content rather than metrics. Most UK funders have signed, although only a few universities have currently signed. See also the Leiden Manifesto for Research Metrics (<http://www.leidenmanifesto.org/>). We expect most UK universities to now sign DORA, or equivalent, due to the Wellcome Trust's new policy (starting 2020) requiring institutions to sign. Assuming institutions take this seriously, evaluating papers by their content rather than where they are published should reduce the pressure to publish in top-tier journals.

**Published peer review reports.** Many journals now already, or have pledged to, provide greater transparency about the quality of peer review they provide by publishing reviewer reports alongside published articles. Notably two large open access publishers PLOS and MDPI are amongst those that are pledging to provide greater transparency from 2019 [20]. Publons (<https://publons.com/>) allows reviewers to 'claim' metadata records on their profile for peer reviewing and editorial work they have done. Publishing reviewer reports, whether signed or not, should increase transparency in the reviewing process.

**Post publication peer review.** A journal may immediately publish a paper upon submission; reviews are then sought for the preprint and made public. If sufficient reviewers support publication, the article is formally accepted and e.g. listed on Pubmed. Leading examples of this approach are F1000 Research, who provide the infrastructure for several institution- and funder-specific journals, such as *Wellcome Open Research* and *Gates Open Research*. This approach to publishing complements preprinting by ensuring the paper is publicly available whilst undergoing peer review.

**Format free submissions.** Journals have traditionally imposed strict formatting requirements for manuscripts. As editors at top-tier journals 'desk reject' most submissions before peer review, this leads to many wasted hours [21]. Gradually life science journals are now dropping these formatting requirements for initial submissions, instead allowing "format free" submissions [22], and hence saving researchers from tedious reformatting tasks.

## Funder mandates and compliance

Key funders in the UK have had policies in place supporting open access for many years. In particular, the Wellcome Trust has mandated Open Access for publications funded by them since 2006, with sanctions for non-compliance. Compliance rates (around 90%) are highest for Wellcome, as of October 2017 [23], with compliance for other main funders varying at 70-90%. Where work has been supported by relevant funding agencies, our experience to date is that funds have always been available to support Article Processing Charges (APCs).

One unintended consequence of these policies has been that most traditional journals have established a "hybrid" model of publishing, with APCs that on average exceed those in pure Open Access (OA) journals [24]. This hybrid model of publishing has so far shown little signs of disappearing. The success of OA publishing however has meant that government-provided funds can often no longer cover all APCs and UK institutions are beginning to restrict the choice of journals for which APCs will be paid-to in order to best optimise the allocation of limited financial resources. We understand and are supportive of institutions that do not allow APCs to be paid for hybrid-OA, or for particularly expensively priced OA

journals - some publishing options are simply exploitative of the system and authors may need protection from them.

However, the OA publishing world is due to change dramatically in 2020 with the recent announcement of “plan S” [25], a European initiative to enforce OA, cap APCs and not financially support hybrid journals. Whilst we support the notions underlying plan S, its success will depend on further implementation details that are currently under discussion. To date, gold OA, where authors typically pay often quite large APCs, is seen as the predominant way of meeting funders’ OA mandates. However green OA (publishing your author-accepted manuscript on a suitable server) is a viable alternative to making work freely available. Finally, a relatively new model of diamond OA, where there are no fees either to read or publish papers, is being explored. This approach has been successfully used in mathematics, where costs are kept low by hosting the published papers on ArXiv. With the rise of preprint servers in the life sciences, we look forward to the emergence of similar low-cost overlay journals in the life sciences.

## Concluding remarks

Current publication practices can often lead early career researchers to be ‘Bullied into Bad Science’ (<http://bulliedintobadscience.org/>). We have outlined several recent developments that we hope present alternatives to the traditional hierarchy of scholarly publishing. These developments should help reduce the pressure on early career researchers that they currently face in the “publish or perish” culture. We encourage the adoption of the above open practices to help create a more ethical research environment.

## Conflicts of interest

SJE, RM, LG and AMC are members of the Bullied into Bad Science campaign. SJE is on the editorial board of *Scientific Data* and an affiliate of *BioRxiv*. LG is an ASAPbio ambassador.

## Acknowledgements

SJE thanks Magdalene College Cambridge for financial support.

## References

1. Brenner, S. (1995). Loose end. *Curr. Biol.* 5, 568. DOI: [10.1016/S0960-9822\(95\)00109-X](https://doi.org/10.1016/S0960-9822(95)00109-X).
2. Simons, K. (2008). The misused impact factor. *Science* 322, 165. DOI: [10.1126/science.1165316](https://doi.org/10.1126/science.1165316).
3. Ginsparg, P. (2017). Preprint déjà vu: An FAQ. arXiv [cs.DL]. Available at: <https://arxiv.org/abs/1706.04188>.
4. Cobb, M. (2017). The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLoS Biol.* 15, e2003995. DOI: [10.1371/journal.pbio.2003995](https://doi.org/10.1371/journal.pbio.2003995).
5. Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454. DOI: [10.1038/533452a](https://doi.org/10.1038/533452a).

6. Nosek, B.A., Ebersole, C.R., DeHaven, A.C., and Mellor, D.T. (2018). The preregistration revolution. *Proc. Natl. Acad. Sci. U. S. A.* *115*, 2600–2606. DOI: [10.1073/pnas.1708274114](https://doi.org/10.1073/pnas.1708274114).
7. Warren, M. (2018). First analysis of “pre-registered” studies shows sharp rise in null findings. *Nature*. DOI: [10.1038/d41586-018-07118-1](https://doi.org/10.1038/d41586-018-07118-1).
8. Federer, L.M., Belter, C.W., Joubert, D.J., Livinski, A., Lu, Y.-L., Snyders, L.N., and Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. *PLoS One* *13*, e0194768. DOI: [10.1371/journal.pone.0194768](https://doi.org/10.1371/journal.pone.0194768).
9. Nosek, B.A., Alter, G., Banks, G.C., Borsboom, D., Bowman, S.D., Breckler, S.J., Buck, S., Chambers, C.D., Chin, G., and Christensen, G. *et al.* (2015). Promoting an open research culture. *Science* *348*, 1422–1425. DOI: [10.1126/science.aab2374](https://doi.org/10.1126/science.aab2374).
10. Buckheit, J.B., and Donoho, D.L. (1995). WaveLab and reproducible research. In *Wavelets and statistics Lecture notes in statistics*. (Springer New York), pp. 55–81. DOI: [10.1007/978-1-4612-2544-7\\_5](https://doi.org/10.1007/978-1-4612-2544-7_5).
11. Gatto, L., and Christoforou, A. (2014). Using R and bioconductor for proteomics data analysis. *Biochim. Biophys. Acta* *1844*, 42–51. Available at: <http://www.bioconductor.org/packages/release/data/experiment/vignettes/RforProteomics/inst/doc/RforProteomics.html> DOI: [10.1016/j.bbapap.2013.04.032](https://doi.org/10.1016/j.bbapap.2013.04.032).
12. Breckels, L.M., Mulvey, C.M., Lilley, K.S., and Gatto, L. (2018). A bioconductor workflow for processing and analysing spatial proteomics data. *F1000Res.* *5*. Available at: <https://github.com/lmsimp/bioc-pRoloc-hyperLOPIT-workflow> DOI: [10.12688/f1000research.10411.2](https://doi.org/10.12688/f1000research.10411.2).
13. Reproducible document stack from eLife Available at: <https://elifesciences.org/labs/7dbeb390/reproducible-document-stack-supporting-the-next-generation-research-article>.
14. Bioconductor gateway on f1000 research Available at: <https://f1000research.com/gateways/bioconductor>.
15. Nature research journals trial new tools to enhance code peer review and publication Available at: <http://blogs.nature.com/ofschemasandmemes/2018/08/01/nature-research-journals-trial-new-tools-to-enhance-code-peer-review-and-publication>.
16. PLOS Biology Staff Editors (2018). The importance of being second. *PLoS Biol.* *16*, e2005203. DOI: [10.1371/journal.pbio.2005203](https://doi.org/10.1371/journal.pbio.2005203).
17. Aizawa, S. (2016). Results of an attempt to reproduce the STAP phenomenon. *F1000Res.* *5*, 1056. DOI: [10.12688/f1000research.8731.2](https://doi.org/10.12688/f1000research.8731.2).
18. ORCID open letter on workflows Available at: <https://orcid.org/content/requiring-orcid-publication-workflows-open-letter>.
19. San Francisco Declaration on Research Assessment (DORA). Available at: <https://sfedora.org/>.
20. Open letter on the publication of peer review reports Available at: <http://asapbio.org/letter>.
21. Budd, J. (2017). Publishing: Reformatting wastes public funds. *Nature* *543*, 40. DOI: [10.1038/543040e](https://doi.org/10.1038/543040e).
22. Khan, A., Montenegro-Montero, A., and Mathelier, A. (2018). Put science first and formatting later. *EMBO Rep.* *19*. Available at: <https://asntech.github.io/format-free-journals/> DOI: [10.15252/embr.201845731](https://doi.org/10.15252/embr.201845731).
23. Larivière, V., and Sugimoto, C.R. (2018). Do authors comply when funders enforce open access to research? *Nature* *562*, 483. DOI: [10.1038/d41586-018-07101-w](https://doi.org/10.1038/d41586-018-07101-w).

24. Pinfield, S., Salter, J., and Bath, P.A. (2017). A “gold-centric” implementation of open access: Hybrid journals, the “total cost of publication,” and policy development in the UK and beyond. *Journal of the Association for Information Science and Technology* 68, 2248–2263. DOI: [10.1002/asi.23742](https://doi.org/10.1002/asi.23742).

25. Schiltz, M. (2018). Science without publication paywalls: cOAlition S for the realisation of full and immediate open access. *PLoS Biol.* 16, e3000031. DOI: [10.1371/journal.pbio.3000031](https://doi.org/10.1371/journal.pbio.3000031).