Allison Langham-Putrow and Amy Riegelman

Discovery and scholarly communication aspects of preprints

Sources for online information

Preprints are not a 21st-century invention, but there has been a considerable uptick in the creation of preprint repositories and the attention to the role that preprints play in open science. Note, the definitions of *preprint* have evolved over time, lack consensus, and vary in terminology used across disciplines. For the purposes of this article, we will be referring to a preprint as a version of a manuscript that is self-archived and shared publicly before publication in a scholarly journal. The status of a preprint can exist at several points on the scholarly continuum including, but not limited to, an author's early draft, a submitted manuscript under review, or an accepted manuscript.

Preprints are classified as grey literature and green open access. Prior to electronic communications, early versions of manuscripts were circulated by authors via mailing services and informal sharing among colleagues. After the creation of Los Alamos National Laboratory's (LANL's) arXiv in 1991, preprints were commonly referred to as *e-prints*. Confusing matters, the term *e-prints* encompasses preprints as well as post-prints made publically available on a server.

For the purposes of this paper, the authors are not treating preprints as synonymous with working papers. This has been disputed, but important distinctions exist as working papers are not necessarily green open access due to privileged access, paywalls, and/or embargo periods on some platforms. As explained by Cameron Neylon et al., a *preprint* in physics means something different than a *working paper* means in economics.[1] Whereas the physics community cites preprints and considers preprints as establishing priority, a working paper operates under a different standing as a work in progress. It is important to acknowledge that language has evolved oddly as scholarly communication has transitioned awkwardly from print to electronic.

The purpose of preprints is to increase the speed at which research results are disseminated. They are not a way to bypass peer review—they bypass delays resulting from the peer review process.[2] They have clear benefits to the authors, as preprints allow authors to stake a claim in their research by putting a "time-stamp" on their ideas.[3] The National Institutes of Health (NIH) cites both of these reasons for sharing preprints in a March 2017 Notice allowing NIH-funded researchers to cite preprints as products of NIH funding and cite them in further grant applications. It identifies additional benefits of sharing preprints: the ability to obtain feedback and offset publication bias.[4]

Allison Langham-Putrow is scholarly communications librarian, email: lang0636@umn.edu, and Amy Riegelman is a social sciences librarian, email: aspringe@umn.edu, at the University of Minnesota

Some criticisms of preprints include the fear of being scooped, plagiarism, and individuals making medical decisions and other life choices based on unrefereed (potentially unverified) research findings. A counterpoint to the latter is that flawed research has and will likely continue to be published by highly reputable peer review journals.[5] Being penalized by journal publishers for sharing data or results before submitting a manuscript is a common concern regarding preprints. The "Ingelfinger rule"[6] is often cited as a reason to not share a preprint; however, publishers are changing their stances to encourage the sharing of preprints.[7]

Preprint sharing, particularly in health-related research, is not without its critics, who worry of negative consequences of waiting to review research until after publication. However, as Amy Harmon reported, ". . . some #ASAPbio [a non-profit organization that promotes sharing of preprints as quickly and openly as possible] advocates argue that since the rise of the Internet, biologists have been abdicating their duty to the public—which pays for most academic research—by not sharing results as quickly and openly as possible."[8]

In this article, readers will be informed about where to locate preprints, the general infrastructure of repositories, preprint platforms for journal clubs, and where to locate journal and publisher policies regarding preprints.

Servers, repositories, and platforms

This list does not include institutional repositories, which often host self-archived preprints.

- **arXiv.** Initially the server was hosted at LANL and was limited to physics research but now includes several other disciplines, including mathematics, computer science, quantitative biology, quantitative finance, statistics, electrical engineering and systems science, and economics. arXiv was created by Paul Ginsparg in 1991, and in 2001 Ginsparg and the repository moved to Cornell University. arXiv is considered a trailblazer for electronic preprints. It is currently funded by Cornell University, the Simons Foundation, and contributions from 233 member institutions. *Access:* https://arxiv.org.
- **bioRxiv.** Launched in 2013, this repository contains research from the life sciences and is maintained by Cold Spring Harbor Laboratory (CHSL). bioRxiv facilitates direct transfer of submissions to participating journals. This means authors are able to submit manuscripts to participating journals by transmitting files as well as metadata, from the bioRxiv platform. They refer to this as "B2J." *Access:* www.biorxiv.org.
- **ChemRxiv.** The scope of this preprint repository is chemistry and related fields. It is a collaborative initiative including the following groups: the American Chemical Society, the German Chemical Society, and the Royal Society of Chemistry. *Access:* https://chemrxiv.org.
- **MedRxiv.** Launched in June 2019, this health sciences preprint server is a collaborative effort between CSHL, BMJ, and Yale. Submissions are screened to deter potential risk to public health. *Access:* medrxiv.org.

OSF preprints

Several partner repositories are hosted on the Open Science Framework (OSF) preprint platform. The partner repositories vary by discipline, region, governing structure, and screening process. The repositories and their disciplines are listed below.

AfricArXiv, literature from African scientists
AgriXiv, agriculture and allied sciences
Arabixiv, Arabic open science repository
BodoArXiv, medieval studies
EarthArXiv, earth sciences
EcoEvoRxiv, ecology, evolution, and conservation
ECSarXiv, electrochemistry and solid state science and technology
EdArXiv, education research

engriXiv, engineering

FocUS Archive, focused ultrasound research

Frenxiv, literature from French scholars

INA-Rxiv, literature from Indonesian scholars

indiarxiv, literature from Indian scholars, maintained by Open Access India

LawArXiv, legal scholarship

LIS Scholarship Archive, library and information science

MarXiv, ocean and marine-climate sciences

MediArXiv, media, film, and communication studies

MetaArXiv, research transparency and reproducibility

MindRxiv, the mind and contemplative practices

NutriXiv, nutritional sciences

PaleoXiv, paleontology

PsyArXiv, psychological sciences

SocArXiv, social sciences

SportRxiv, sport, exercise, performance, and health

ThesisCommons, theses

Access all OSF partner repositories: https://osf.io/preprints/.

- **PeerJ Preprints.** Launched in 2013, this multidisciplinary preprint service is complementary to PeerJ, which is an open access megajournal. On September 3, 2019, PeerJ announced its plan to stop accepting preprints as of September 30, 2019. The archive will remain accessible.[9] *Access:* https://peerj.com/preprints/.
- **Preprints.org.** Launched in 2016, this platform is multidisciplinary and maintained by Multidisciplinary Digital Publishing Institute. *Access:* https://preprints.org.
- **Research preprints: Server list.** The list of preprint repositories in this column is not meant to be comprehensive but instead to give readers an idea of some of the options. Alternatively Martyn Rittman from Preprints.org maintains a spreadsheet with timely updates of characteristics of existing preprint servers. *Access:* http://z.umn.edu/PreprintRepositories.

Discovery

The following are a few resources researchers can use to find preprints.

- **Europe PMC.** Europe PMC began indexing preprints in 2018.[10] Preprints are index using Crossref searches on specific DOI prefixes. It currently indexes preprints from bioRxiv, PeerJ Preprints, ChemRxiv, and F1000Research. Preprints are labeled as such and crosslinked to the peer-reviewed versions once they are published. *Access:* https://europepmc.org/.

Google Scholar. Google Scholar results may include links to preprints, although there are some idiosyncrasies with how preprints are linked. OSF provides several reasons why a preprint may not appear in Google Scholar results, including issues with author names.[11] And although Google Scholar's policy [12] is to set the primary link to the publisher version, there are cases in which the record of an article links only to the preprint, even after the article has been published in a journal. This has been a known issue caused by a lag in time between publication in a journal and Google Scholar updating its record.[13] Check the preprint server record to determine if a final version has been published. *Access:* https://scholar.google.com/.

- **search.bioPreprint.** The University of Pittsburgh Health Science Library Service released this preprint search tool in 2016, introducing it through a preprint posted on bioRxiv.[14] search.bioPreprint conducts a federated search of the Quantitative Biology section of arXiv, bioRxiv, F1000Research, PeerJ Preprints, preprints.org, and Wellcome Open Research. It allows results to be filtered by topic or source, but it also uses IBM Watson Explorer to offer a "remix" feature that reclusters search results to identify secondary topics. search.bioPreprint also offers a bookmarklet that allows the user to highlight a word or phrase on any webpage and search on it. *Access:* https://www.hsls.pitt.edu/preprint.

• **SHARE and OSF Preprints.** SHARE is a collaboration between the Center for Open Science and the Association of Research Libraries that was established to make research, particularly that hosted in institutional repositories, widely discoverable by improving the standardization of metadata. SHARE is hosted by COS and is what powers the search functionality of OSF Preprints. *Access:* https://share.osf.io/discover, https://osf.io/preprints/.

Infrastructure

• **Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) and the Santa Fe Convention.** OAI-PMH developed as a result of a 1999 meeting to address issues with interoperability of preprint servers and other digital repositories, which resulted in an interface/protocol called the Santa Fe Convention.[15] OAI was founded the following year in recognition of the fact that the same issues of interoperability affected libraries, museums, and publishers. Version 2.0, the current version, was released in 2002. Crossref, which has allowed registration of preprints since 2016, makes its metadata available through OAI-PMH. *Access:* https://www.openarchives.org/pmh.

• **ASAPbio: Surveying the landscape of products and services for sharing preprints.** ASAPbio maintains a list of preprint server products and services available to server operators. The list identifies the product or service, the owner or developer, and a list of known users. Open source products identified include OSF, ePrints, and a preprint platform that the Public Knowledge Project is developing for SciELO. FigShare and Atypon[16] have partnered with a number of publishers and scholarly societies. *Access:* https://asapbio.org/preprint-products.

Platforms for preprint journal clubs

• **PREreview (Post, Read, and Engage with preprint reviews).** PREreview is hosted by Authorea. Journal clubs can collaboratively write reviews of preprints and issue a DOI to their review, allowing the review to be cited and indexed by search engines. PREreview is working with bioRxiv to make reviews discoverable from the preprint server. *Access:* https://www.prereview.org/.

• **PubPeer.** PubPeer presents itself as "The online journal club." Users can provide reviews on any type of manuscript (including preprints) that has a DOI or arXiv ID. It allows for anonymous commenting, which they suggest has led to the exposure of "unexpected levels of research misconduct."[17] *Access:* https://pubpeer.com/.

Journal and publisher perspectives

Preprint policies vary by publisher and journal.

• **Committee on Publication Ethics (COPE).** COPE provided a number of recommendations in a Discussion Document in March 2018:[18] journal editors will need to determine whether to consider for review work that has been posted to a preprint platform and decide which, if any, versions of the article the authors may post during peer review or after article acceptance. These decisions have implications on issues, such as anonymizing peer review. COPE recommends that publishers have clear policies on copyright and license requirements and notes that preprint platforms should provide a clear description of the license applied to preprints. *Access:* https://publicationethics.org/resources/discussion-documents/preprints.

• **International Committee of Medical Journal Editors (ICMJE).** ICMJE addresses preprints in their recommendations documentation. The recommendations state that authors should not submit the same manuscript to two different journals, but that journals can consider publishing articles that were posted as preprints. The preprint server must clearly identify the work as a preprint that has not been peer reviewed and include conflict of interest statements.

ICMJE places the responsibility for amending the preprint to identify the final, published version of the article on the authors. *Access:* http://www.icmje.org/icmje-recommendations.pdf.

- **List of academic journals by preprint policy.** A Wikipedia page lists preprint policies for various journals and publishers. The information is provided in two tables. For publishers with unified policies, the table lists the policy type, policy text, policy on preprint licenses (if stated), and a reference. Journal entries list the publisher, policy type, policy text, policy on preprint licenses, and provides a link to a reference. *Access:* https://en.wikipedia.org/wiki/List_of_academic_journals_by_preprint_policy.
- **SHERPA/RoMEO.** This initiative uses color-coding to identify the self-archiving

policy of journals. In SHERPA/RoMEO, the term preprint refers to the version of an article that has not been peer reviewed and "postprint" refers to the version of the article after peer review revisions have been completed. Both "green" and "yellow" coded journals allow for preprint sharing. *Access:* http://www.sherpa.ac.uk/romeo/index.php.

Notes

1. Cameron Neylon, Damian Pattinson, Geoffrey Bilder, and Jennifer Lin, "On the Origin of Nonequivalent States: How We Can Talk About Preprints," *F1000Research* 6 (May 2, 2017): 608.

2. Peter Suber, *Open Access* (Cambridge, MA: MIT Press, 2012).

3. Ibid., 63.

4. Office of Policy for Extramural Research Administration, "Reporting Preprints and Other Interim Research Products," NOT-OD-17-050, 2017, Grants.nih.gov/grants/guide/notice-files/NOT-OD-17-050.html (accessed August 25, 2019).

5. "Key MEDLINE Indicators," https://www.nlm.nih.gov/bsd/bsd_key.html (accessed August 25, 2019).

6. E. Marshall, "Franz Ingelfinger's Legacy Shaped Biology Publishing," *Science,* vol. 282, no. 5390 (1998): 861.

7. "Springer Nature Journals Unify Their Policy to Encourage Preprint Sharing," *Nature* 569, no. 7756 (2019): 307.

8. Amy Harmon, "Handful of Biologists Went Rogue and Published Directly to Internet," *New York Times*.

9. Jason Hoyt, "PeerJ Preprints to stop accepting new preprints Sep 30th 2019," PeerJBlog, September 3, 2019, https://peerj.com/blog/post/115284881747/peerj-preprints-to-stop-accepting-new-preprints-sep-30-2019/.

10. Maria Levchenko, "Preprints in Europe PMC: Reducing friction for discoverability, Europe PMC [blog], July 11, 2018, http://blog.europepmc.org/2018/07/preprints.html.

11. Sara Bowman, "Why isn't my preprint showing in Google Scholar? Preprint FAQs," OSF.io 2019, https://help.osf.io/hc/en-us/articles/360019930493-Preprint-FAQs.

12. "Publisher Support." Google Scholar, https://scholar.google.com/intl/en/scholar/publishers.html#policies.

13. "The Google Scholar preprint bug redux," *Serial Mentor,* 2015, https://serialmentor.com/blog/2015/10/8/Google-Scholar-bug-r.

14. Carrie L. Iwema, John LaDue, Angela Zack, and Ansuman Chattopadhyay, "search. bioPreprint: A discovery tool for cutting edge, preprint biomedical research articles," bioRxiv (2016): 052563, https://www.biorxiv.org/content/10.1101/052563v1.full.

15. Marshall Breeding. "Understanding the protocol for metadata harvesting of the open archives initiative," *Computers in Libraries* 22, no. 8 (2002): 24-29.

16. Wiley purchased Atypon in 2016.

17. PubPeer Foundation, "About PubPeer," PubPeer.com, https://pubpeer.com/static/about (accessed August 25, 2019).

18. Committee on Publication Ethics, "Preprints," Publicationethics.org, https://publicationethics.org/resources/discussion-documents/preprints (accessed August 25, 2019).