Preprints in the Spotlight
Establishing Best Practices, Building Trust

May 27, 2020

Oya Y. Rieger

ITHAKA S+R

Ithaka S+R provides research and strategic guidance to help the academic and cultural communities serve the public good and navigate economic, demographic, and technological change. Ithaka S+R is part of ITHAKA, a not-for-profit organization that works to advance and preserve knowledge and to improve teaching and learning through the use of digital technologies. Artstor, JSTOR, and Portico are also part of ITHAKA.

Introduction

Preprints have been getting a lot of attention recently. The COVID-19 pandemic—the first major health crisis since medical and biomedical preprints have become widely available online—has further underscored the importance of speedy dissemination of research outcomes. Preprints allow researchers to share results with speed, but raise questions about accuracy, misconduct, and our reliance on the "self-correcting" nature of the scientific enterprise. As scientists and health care professionals, as well as the general public, look for information about the pandemic, preprint services are growing in importance. So too are the policy decisions preprint platform leaders make.

Even before the crisis struck, it was clear that 2020 would be a year of reckoning for preprints. Last year saw a comprehensive review of the landscape by the Knowledge Exchange, several discussions at the Charleston Conference, and a two-day preprints event organized by NISO.[1] As the new decade opened, ASAPbio, in collaboration with the European Bioinformatics Institute (EMBL-EBI) and Ithaka S+R, spearheaded a preprints roadmap workshop designed to improve the current preprints infrastructure and processes for discoverability, reuse, and community trust.[2] In a virtual meeting in April 2020, the STM Association featured a long session on preprints, and since the pandemic, dozens of articles have appeared in the scientific and popular press about both the role of preprints in accelerating scientific communications and the associated concerns.[3]

---

[1] Andrea Chiarelli, Rob Johnson, Emma Richens, and Stephen Pinfield, "Accelerating Scholarly Communication: The Transformative Role of Preprints," Knowledge Exchange, 2019, https://doi.org/10.5281/zenodo.3357727; Charleston Conference; "Preprints: Why Librarians Should Care?" https://2019charlestonlibraryconference.sched.com/event/UXtN/preprints-why-librarians-should-care; "Hyde Park Debate, Resolved: Preprint Servers Have Improved the Scholarly Communication System," https://2019charlestonlibraryconference.sched.com/event/UXqQ/hyde-park-debate; "Open Access: The Role and Impact of Preprint Servers," *NISO on the Road*, NFAIS Foresight Event, November 14-15, 2019, https://www.niso.org/events/2019/11/open-access-role-and-impact-preprint-servers.

[2] "ASAPbio January 2020 Workshop: A Roadmap for Transparent and FAIR Preprints in Biology and Medicine," https://asapbio.org/meetings/preprints-roadmap-2020. Also see, Oya Y. Rieger, "Preprints in Biology and Medicine," Ithaka S+R (blog), January 30, 2020, https://sr.ithaka.org/blog/preprints-in-biology-and-medicine.

[3] Here are just a few of those articles: Justin Fox, "A Pandemic Moves Peer Review to Twitter," *Bloomberg Opinion*, May 5, 2020, https://www.bloomberg.com/opinion/articles/2020-05-05/coronavirus-research-moves-faster-than-medical-journals; Kim Tingley, "Coronavirus Is Forcing Medical Research to Speed Up," *The New York Times Magazine*, April 21, 2020, https://www.nytimes.com/2020/04/21/magazine/coronavirus-scientific-journals-research.html; Marcus Munafo, "What You Need to Know about How Coronavirus Is Changing Science, *The Conversation*, May 5, 2020, https://theconversation.com/what-you-need-to-know-about-how-coronavirus-is-changing-science-137641; "Scientific Research on the Coronavirus Is Being Released in a Torrent: Will That Change How Science is Published?" *The Economist,* May 7, 2020, https://www.economist.com/science-and-technology/2020/05/06/scientific-research-on-the-coronavirus-is-being-released-in-a-torrent; Jackie Flynn Mogensen, "Science Has

The purpose of this issue brief is to provide an overview of the preprint landscape in the first half of 2020 as we witness rapid changes to how they are perceived and utilized.[4] While we write within the context of a gathering and growing global pandemic, where preprints seem to be playing an essential role, the arc of how preprints are evolving is broader than our immediate public health needs. Preprints are no panacea, but as they have continued to develop in their own right they are putting useful pressure on some of the structures of traditional scientific publishing.

Expanding Uptake for Preprints

While preprints have been well-established in physics, mathematics, and computer science through arXiv, broader uptake has continued gradually, based on disciplinary ethos and communications cultures.[5] Almost a dozen new preprint services—using a variety of organizational models—launched last year. Several of these initiatives were initiated by society publishers. IEEE's TechRxiv supports preprints (or unpublished research) in electrical engineering, computer science, and related technology regardless of where authors eventually intend to submit.[6] Through a collaboration between the American Political Science Association (APSA) and Cambridge University Press, APSA Preprints provides a prepublication platform for early research outputs in political science and related disciplines.[7]

Another trend in the preprint space is the establishment of new national and regional platforms by open science advocates. For instance, in 2018 AfricArxiv was launched to improve the visibility of African science by helping researchers share their work quickly and foster collaboration.[8] It supports preprints, postprints, code and data, and welcomes submissions in all African languages. IndiaRxiv was established last year to manage a

---

an Ugly, Complicated Dark Side. And the Coronavirus Is Bringing it Out," *Mother Jones*, April 28, 2020, https://www.motherjones.com/politics/2020/04/coronavirus-science-rush-to-publish-retractions/.

[4] I am grateful to Jessica Polka (ASAPbio Executive Director) and Matthew Ismail (Editor in Chief, The Charleston Briefings; Director of Collection Development, Central Michigan University Library) for taking the time to review an earlier version of this brief.

[5] arXiv, https://arxiv.org/.

[6] TechRxiv, https://www.techrxiv.org/.

[7] APSA Preprints, https://preprints.apsanet.org/engage/apsa/public-dashboard.

[8] AfricArxiv is a digital archive for African research for working papers, preprints, accepted manuscripts (post-prints), and published papers with an option to link data and code, and for article versioning. A recent collaboration with PubPub, the open-source collaboration platform built by the Knowledge Futures Group, enables the hosting of audio/visual preprints to support multimedia research outputs. See https://info.africarxiv.org/.

preprints repository for India.[9] Illustrating the dynamic nature of the preprint space, although both of these new services were originally hosted on the Open Science Framework, IndiaRxiv recently collaborated with CABI to rebrand the service as "preprints for agriculture and allied sciences."[10]

Debate on the Value of Preprints

Although uptake is broadening, there continues to be some anxiety about if and how the preprint services are improving the scholarly communication ecology. For instance, Kent Anderson, Scholarly Kitchen founder and CEO of RedLink, has been vocal in expressing his concerns about how preprint services add abandoned or flawed work to the already fragmented distribution system.[11] At the heart of his apprehension is the potential risk of the public relying on preprints, not realizing that such works have not gone through a vetting process yet and may not be reliable. His concern has more immediacy because preprints are taking off, being used more widely by both researchers and the general public.

Social media has become one of the primary discovery channels for preprints, augmenting the worry about the risks associated with research that has not gone through peer review being circulated widely. Findings from an analysis of preprints posted on bioRxiv show that when research is shared early on with specific communities, they can make a significant impact rapidly regardless of the journal's prestige or citation impact.[12] Compared to their peer-reviewed and published versions in journals, the community-based nature of preprints coupled with social media promotion attracts more readers and online mentions.

Deposit and Usage

Although the number of preprint services is growing, another metric is the rate of preprint deposits. In 2019, arXiv received 156,000 new submissions and bioRxiv

---

[9] IndiaRxiv: https://indiarxiv.org.

[10] CABI's mission is to improve people's lives worldwide by providing information and applying expertise to solve problems in agriculture and the environment. See https://agrirxiv.org/.

[11] "Preprints: An Interview with Kent Anderson," NISO, October 2019, https://www.niso.org/niso-io/2019/10/preprints-interview-kent-anderson.

[12] Nicholas Fraser, Fakhri Momeni, Philipp Mayr, and Isabella Peters, "The Effect of bioRxiv Preprints on Citations and Altmetrics," bioRxiv; June 2019, https://doi.org/10.1101/673665.

accepted over 31,000 manuscripts.[13] Although this is not a systematic comparison, to provide a reference point, in 2017 there were over 400,000 papers published in the biological/biomedical sciences out of approximately 3.5 million journal articles published in 2018.[14] While there has been a significant increase in new preprints, preprints in biology represent approximately 2.4 percent of all biomedical publications.[15] The total number of papers in each preprint repository varies widely—arXiv (founded in 1991) includes 1,332,986 papers, RePEc (1996) includes 799,615, while SportRxiv (2017) includes 193, MindRxiv (2017) 194, and NutriXiv (2017) 67.

With the exception of a handful of preprint repositories such as arXiv, almost all services provide usage metrics to indicate the number of article downloads. Some services, such as COS preprints, include the number of views while others release details of social media interactions. SSRN (founded in 1994, 931,213 papers) compiles and makes available rankings of papers on a number of measures, such as downloads and citations.[16] Whereas arXiv's long-standing philosophical position is that because they cannot guarantee the accuracy of usage statistics (especially considering gaming), they do not make download/use statistics available at an individual author level to avoid potentially misleading data.[17] This example underscores the challenges associated in reporting reliable and consistent usage data. COUNTER enables the library community to monitor the use of electronic resources by facilitating the recording and reporting of online resource usage stats in a consistent and credible way. Adopting a similar standard would allow preprint platforms to report usage of papers to enable comparison and monitoring. For instance, the UK's Institutional Repository Usage Statistics initiative, a national aggregation service, provides COUNTER-conformant usage statistics for content downloaded from participating UK institutional repositories (IRs) to facilitate comparable, standards-based measurements. [18]

---

[13] bioRxiv's 2019 new submissions can be viewed through its website: https://www.biorxiv.org/search/limit_from:2019-01-01%20limit_to:2019-12-31%20numresults:10%20sort:relevance-rank%20format_result:standard; the arXiv 2019 new submissions are available in "arXiv Annual Update, January 2020," arXiv, January 16, 2020, https://arxiv.org/about/reports/2020_updatehttps://arxiv.org/about/reports/2020_update.

[14] Rob Johnson, Anthony Watkinson, and Michael Mabe, *The STM Report*, The International Association of Scientific, Technical and Medical Publishers, October 2018, https://www.stm-assoc.org/2018_10_04_STM_Report_2018.pdf.

[15] "Biology Preprints Over Time," ASAPbio, https://asapbio.org/preprint-info/biology-preprints-over-time.

[16] "SSRN Top 10,000 Papers," SSRN, May 17, 2020, https://hq.ssrn.com/rankings/Ranking_display.cfm?TRN_gID=10.

[17] It has been arXiv's long-standing philosophical position that because they cannot guarantee the accuracy of usage statistics (especially considering gaming), they do not make download/use statistics available at an individual author level to avoid potentially misleading data. See "Frequently Asked Questions on Public Statistics," arXiv, https://arxiv.org/help/faq/statfaq.

[18] See IRIS-UK, https://irus.jisc.ac.uk/. Recently Jisc and LYRASIS announced that they are joining forces to introduce IRUS in the United States. See, "Jisc and LYRASIS Help US Universities and Research Organisations Gather New Usage Insights," Jisc, May 5,

Sustainability

Simply put, sustainability is the capacity to endure; it entails long-term stewardship for the responsible as well as innovative management of preprint services. At the heart of this concept is the ability to secure resources (technologies, expertise, policies, visions, standards, and so on) needed to protect and enhance the value of a service based on a user community's requirements and vision.[19] A critical component of a sustainability plan is to consider this rich context and understand how a preprint service fits within the broader framework and how we can link related information and communities.

The premise of open scholarship is fostering broad and free access to knowledge to improve global research. However, every initiative, whether it is not-for-profit or behind a paywall, needs a business model to generate revenue to cover expenses. Many of the existing preprint services lack a scalable and transparent business model. Often they are operated by volunteer academic groups dependent on one-time funding, and in some cases based on gifts from foundations or contributions from libraries.

The availability of the Open Science Foundation Preprints from the Center for Open Science (COS) has significantly lowered the barrier to entry. The open source preprint software integrates OSF project infrastructure, allowing researchers to include data, code, and other supplemental information alongside preprints. With shared infrastructure and accelerating growth, OSF aims to allow an economy of scale. They forecast that it will cost $300,000 in 2020 to publish 33,650 papers ($6.81 per paper).[20] However, running a successful service requires a holistic sociotechnical and managerial approach. It is not sufficient to have access to readily available repository architectures, as is evidenced by the recent news that several COS-based preprint services are closing or struggling to operate due to a funding shortfall. This is partially in response to COS's decision to introduce fees to sustain the hosting service in the long term as grants from private foundations are not adequate. In 2020, COS started charging repository managers $1,000 a year (to be increased based on submission rate) to cover maintenance costs. This can be significant, especially for repositories run by volunteers

2020, https://www.jisc.ac.uk/news/jisc-and-lyrasis-help-us-universities-and-research-organisations-gather-new-insights-05-may-2020#.

[19] Oya Y. Rieger, "Sustainability: Scholarly Repository as an Enterprise," *Bulletin of the American Society for Information Science and Technology,* October/November 2013, https://asistdl.onlinelibrary.wiley.com/doi/full/10.1002/bult.2012.1720390110.

[20] David Mellor, "Conflict between Open Access and Open Science: APCs Are a Key Part of the Problem, Preprints Are a Key Part of the Solution," COS, January 21st, 2020, https://www.cos.io/blog/conflict-between-open-access-and-open-science-apcs-are-key-part-problem-preprints-are-key-part-solution.

with limited funds, such as INA-Rxiv and EarthArXiv.[21] Nevertheless, AfricArXiv's founders have made a plea for African and international institutional libraries, governments, foundations, and donors and initiated a crowdfunding campaign to help develop a financing strategy and roadmap for 2020 and beyond.[22] This is just an example to illustrate the sustainability challenges of community-based preprint services.

Hosted by Cornell University, arXiv was the first popular preprint server to be developed and maintained independently, largely driven by scientists in related disciplines with the motto of "science by the scientists." Although the service has established and supports several collaborations with commercial and society publishers, it continues to be run as an independent enterprise with heavy reliance on gifts and volunteer contributions. It is one of the few preprint services with a transparent business model. The service's annual revenues and expenses are reported each year in addition to providing a five-year business model.[23] Since 2010, arXiv's sustainability model has aimed to reduce arXiv's financial burden and dependence on a single institution by instead creating a broad-based, community-supported resource. The service is funded by the Simons Foundation ($400,000/year) and through a membership program. The libraries and research laboratories that are the heaviest institutional users pay between $1,000 and $4,400 a year in membership fees. Also, Cornell University provides a cash subsidy and makes an in-kind contribution of all indirect costs.

bioRxiv is financially supported by the Cold Spring Harbor Laboratory and Chan Zuckerberg Initiative (CZI). In 2018-2020, CZI contributed about $2 million to support the conversion of existing and new preprints on the bioRxiv service into the JATS XML format that will allow new forms of article enhancement and presentation.

It would be worthwhile to compare the operating costs of other major preprint servers and analyze their current funding sources and stability. This is especially important to build trust in their business operation and financial durability. As requirements for quality, discoverability, and innovation build up, so will the annual expenses for their maintenance and further development.

---

[21] Smriti Mallapaty, "Popular Preprint Servers Face Closure because of Money Troubles," *Nature*, February 13, 2020, https://www.nature.com/articles/d41586-020-00363-3.

[22] Jo Havemann, "Service Fees for OSF Preprint Hosting and Maintenance – AfricArXiv Continues its Services," AfricArXiv, February 14, 2020, https://info.africarxiv.org/osf-service-fees/.

[23] "arXiv Annual Update, January 2020," arXiv, January 16, 2020, https://arxiv.org/about/reports/2020_update; Oya Y. Rieger, "2018-2022: Sustainability Plan for Classic arXiv," arXiv, January 15, 2019, https://confluence.cornell.edu/display/arxivpub/2018-2022%3A+Sustainability+Plan+for+Classic+arXiv.

If preprints are first-class research objects and not simply precursors to the eventual publication, then they merit more serious consideration for preservation to ensure long-term access. Digital preservation (a term used interchangeably with "archiving") refers to a range of managed activities to support the long-term maintenance of digital content, thereby ensuring that digital objects are usable and accessible over time. It should be seen as an important indicator of sustainability and involves more than bitstream preservation. Overall, there is no consistent information available about arrangements between preprint platforms and third-party preservation services providers such as CLOCKSS, Internet Archive, and Portico for the long-term management of digital assets. However, almost all preprint platforms in the medical and biomedical sciences expressed that they either have a preservation plan or are about to implement a strategy through Portico or a fund provided by the Center for Open Science (COS) to maintain read access for over 50 years.[24]

Emerging Business Models

An important sign of the maturation of the preprints marketplace is the closer integration and coupling of preprints with various stages of research workflows through new business models. There are some valuable models emerging that promise to be more financially durable and scalable. For instance, ChemRxiv is built on a partnership involving the five largest chemical science societies,[25] and APSA Preprints is a joint venture between the American Political Science Association and Cambridge University Press.[26]

In recent years, all the major publishing incumbents have made substantial investments in preprints. The first and most dramatic was Elsevier's acquisition of SSRN,[27] whose field-specific research networks have continued to grow since the purchase, and its later acquisition of bepress, whose Digital Commons repository contains numerous preprints

---

[24] Portico is a community-supported preservation archive that safeguards access to e-journals, e-books, and digital collections, https://www.portico.org/. COS established a $250,000 preservation fund for hosted data in case the service is curtailed or ceased. If activated, the preservation fund will preserve and maintain read access to hosted data for 50+ years of read access. See Sarah Bowman, "FAQs," OSF, https://help.osf.io/hc/en-us/articles/360019737894-FAQs#Backup-Preservation-Policy.

[25] ChemRiv: https://chemrxiv.org.

[26] APSA Preprints: https://preprints.apsanet.org/engage/apsa/public-dashboard.

[27] Roger C. Schonfeld, "Elsevier Acquires SSRN," *The Scholarly Kitchen*, May 17, 2016, https://scholarlykitchen.sspnet.org/2016/05/17/elsevier-acquires-ssrn/.

from institutional customers.[28] Research Square, which began as a journal-integrated preprint service in partnership with Springer Nature in 2018, has recently begun accepting multidisciplinary preprint submissions.[29] Research Square is an example of a journal-integrated preprint service that helps drive uptake and awareness. F1000 Research, acquired by Taylor & Francis in January 2020, combines the benefits of preprints with the ability to publish rapidly to support open research in a wide range of disciplines through editorial checks, open data support, and invited open peer review.[30] The platform provides services directly to research funders and institutions such as Wellcome, the Bill & Melinda Gates Foundation and the Health Research Board Ireland, as well as to other scholarly publishers including Emerald Publishing. In 2019, Wiley launched the Under Review service for authors when they submit to participating journals "to streamline the early sharing of research and open up the peer review process." It aims to seamlessly provide preprints at the same time of submission.[31] Publishers are pursuing a notable set of strategies in integrating preprints into their operations.[32]

Publishers have acquired many independent providers but there are other efforts as well. In 2019, SciELO, originally established in Brazil and now expanded to 16 countries, and Public Knowledge Project (PKP) announced a new collaboration to build a preprints server system that will be fully interoperable with Open Journal System (OJS) and other publishing systems such as the SciELO Network journals.[33] This collaboration demonstrates the importance of supporting diversity in scholarly communication and fosters the development of a multilingual system to strengthen the sustainability and affordability of open source software systems in line with editorial practices across the disciplines. In February 2020, PKP released a beta version of Open Preprint Systems

[28] Roger C. Schonfeld, "Elsevier Acquires bepress," *The Scholarly Kitchen*, August 2, 2017, https://scholarlykitchen.sspnet.org/2017/08/02/elsevier-acquires-bepress/.

[29] "Biology Preprints Over Time," ASAP Bio, https://asapbio.org/preprint-info/biology-preprints-over-time.

[30] Since peer review starts the moment preprints are published on F1000 Research, they are not considered preprints and cannot be submitted to other journals. However, if an author decides to discontinue peer review, the article is labelled with an explanation and may be considered equivalent to a preprint. See "F1000 Research joins Taylor & Francis Group," January 10, 2020, https://newsroom.taylorandfrancisgroup.com/f1000-research-joins-taylor-francis/# . Also see F10000 Research's website, https://f1000research.com/.

[31] "Frequently Asked Questions (FAQ) about Under Review," Authorea, https://support.authorea.com/en-us/article/preprints-on-authorea-with-the-under-review-service-17y6l8x/.

[32] Roger C. Schonfeld and Oya Y. Rieger, "Publishers Invest in Preprints," *The Scholarly Kitchen*, May 27, 2020, https://scholarlykitchen.sspnet.org/2020/05/27/publishers-invest-in-preprints/.

[33] "PKP and SciELO Announce Development of Open Source Preprint Server System," September 22, 2018, https://pkp.sfu.ca/2018/09/22/pkp-and-scielo-announce-development-of-open-source-preprint-server-system/.

designed to meet the growing role of preprints in open science anywhere, working across cultural and language differences. [34]

Preprints as First Class Research Outputs

There were some concerns early on that journals might reject articles that had previously been published as preprints. This is changing rapidly. With so many major publishers in the preprints business themselves, the vast majority allow preprint posting prior to submission to their journals.

Several other signs indicate how funders are legitimizing preprints. For instance, Wellcome Trust and the Gates Foundation endorse publishing research as preprints in advance of formal publication in journals, and NIH encourages investigators to use interim products such as preprints to speed dissemination and enhance the rigor of their work. In terms of publishing infrastructure, Crossref now allows preprints to be assigned DOIs and enables them to be linked to the subsequent publishing history, and ORCID added preprint as a "work type" enabling authors to document their own preprints in their record. Google Scholar indexes preprints from all disciplines and makes them searchable on Google and Google Scholar. In another recent milestone, medRxiv was listed as one of the direct links on the Google Scholar homepage for articles about COVID-19, amongst publishers such as JAMA, Nature, Science, Elsevier, Oxford, and Wiley.

As services and platforms proliferate, discussions about preprints are shifting away from their challenges and benefits towards a community conversation on necessary standards and best practices to establish preprints as reliable and interoperable services with transparent policies. Also important is supporting accessibility and equality through features for multiple languages and visually-impaired readers. There also needs to be more emphasis on how preprints are being used and supported by scholars in developing countries. Another open question is the quality control networks formed by some preprint services (often composed on volunteer researchers) and the diversity of the composition of such expert groups who are in charge of filtering submissions.

---

[34] "The Road to Preprints (Part 1): Introducing Open Preprint Systems," PKP, February 24, 2020, https://pkp.sfu.ca/2020/02/24/the-road-to-preprints-part-1-introducing-open-preprint-systems/.

Research Ethics and the Integrity of Scholarly Record

In February 2020 a paper about coronavirus posted on bioRxiv was immediately withdrawn due to the fierce criticism of the study's methodology and arguments.[35] This led to a lively social media debate on the risks associated with non-vetted papers being distributed through preprints services and the value of the traditional publishing model in filtering out junk science. On the other hand, some news outlets declared that the quick dissemination and withdrawal of the paper was a good moment for science, illustrating the role preprint services can play in generating swift reactions from the scientific community and accelerating science.[36] This specific instance once again illustrates the importance of preprint curation and content policies and practices as quality indicators in order to establish preprints as a reliable scholarly format.

In the course of their preprints review in 2019, Knowledge Exchange identified over 60 platforms that can be used to store, share and, in some cases, comment on preprints.[37] Only a handful of them have a basic screening system in place to ensure that the materials submitted are relevant and valuable to the disciplines served, include ethics declarations (e.g., human subjects research), or scan for plagiarism. There is a growing emphasis on implementing transparent and systematic screening procedures both before and after submission, for instance for removing and withdrawing preprints after they are published. COPE (Committee on Publication Ethics) recently revised their 2018 preprints discussion document and will be releasing the new version to promote the discussion and adoption of various ethical aspects of publishing practices.[38] A recent study of biomedical and medical sciences preprints concludes that content screening policies and withdrawal or removal policies vary between platforms based on platform operation, ownership, governance, and financing.[39]

---

[35] The abstract page for the withdrawn paper can be found at: https://www.biorxiv.org/content/10.1101/2020.01.30.927871v1.

[36] Ivan Oransky and Adam Marcus, "Quick Retraction of a Faulty Coronavirus Paper Was a Good Moment for Science," *Stat*, February 3, 2020, https://www.statnews.com/2020/02/03/retraction-faulty-coronavirus-paper-good-moment-for-science/.

[37] Andrea Chiarelli, Rob Johnson, Emma Richens, and Stephen Pinfield, "Accelerating Scholarly Communication: The Transformative Role of Preprints," *Knowledge Exchange*, 2019, https://doi.org/10.5281/zenodo.3357727.

[38] COPE Council, "COPE Discussion Document: Preprints, March 2018, https://publicationethics.org/files/u7140/COPE_Preprints_Mar18.pdf.

[39] Jamie J. Kirkham, Naomi Penfold, Fiona Murphy, Isabelle Boutron, John PA Ioannidis, Jessica K Polka, and David Moher, "A Systematic Examination of Preprint Platforms for Use in the Medical and Biomedical Sciences Setting," bioRxiv, April 2020, https://doi.org/10.1101/2020.04.27.063578.

During the COVID-19 pandemic, both bioRxiv and medRxiv added new policies and cautionary remarks to emphasize that preprints are preliminary reports of work that have not been certified by peer review and should not be reported by the news media as established information. There are efforts to raise awareness about the need for all types of readers, including journalists, to examine preprints with caution and recognize that the findings presented are preliminary and have not been through a scientific vetting process yet.[40]

There are a number of open issues related to metadata standards and licensing terms. However, the most urgent one relates to the integrity of the scholarly record and how to indicate an article's preprint status transparently and accurately. What are the implications of assigning DOIs to preprints so that they are linked to the full publishing history? How do we present supporting data and materials? How do we link preprints to the relevant publishing infrastructure and long-term preservation systems?

What's a Preprint Anyway?

What constitutes a preprint? What content types should be accepted and distributed by preprint services? These are perennial discussion topics at preprints forums. About 50-70 percent of the papers submitted are eventually peer-reviewed and formally published. A quick review of content types supported by a range of preprints revealed a long list, including preprints, e-prints, postprints, conference papers, working papers, reports, white papers, author accepted manuscripts (AAM), versions of record (VoR), literature reviews, book chapters, and even slide decks. For instance, F1000 Research supports a range of research articles, software, posters, and lecture slides. Nevertheless, as Jessica Polka argues, while having a consistent definition is desirable, the focus should be on, "in what contexts and to what degree I can trust this article?"[41] Others believe that getting the terminology right is important to indicate how research is communicated and build trust.[42] Take, for example, a recent article in *The Economist* that conflates open access

---

[40] Denise-Marie Ordway, "Covering Biomedical Research Preprints Amid the Coronavirus: 6 Things to Know," *Journalist's Resource*, April 2, 2020, https://journalistsresource.org/tip-sheets/research/medical-research-preprints-coronavirus/.

[41] Jessica Polka. "Why 'What Is a Preprint?' Is the Wrong Question," NISO, December 2019, https://www.niso.org/niso-io/2019/12/why-what-preprint-wrong-question.

[42] David Crotty, "When is a Preprint Server Not a Preprint Server?" *The Scholarly Kitchen*, April 19, 2017, https://scholarlykitchen.sspnet.org/2017/04/19/preprint-server-not-preprint-server/.

with preprints and proposes that science should use preprints in place of journals, eliding the fact that preprints are not peer reviewed.[43]

Many preprints are a pre-publication artifact. In some fields, this is seen as a benefit. In economics, for example, working papers serve a distinctly different function than the published version. In other fields, preprints are supplanted by the version of record once it is published. Preprint services generally hold preprints in perpetuity, offering in some cases to link to a subsequent version of record, but if the paper is never formally published this is impossible. Kent Anderson has raised a concern that as a result preprints that are reviewed but unpublishable, potentially because they are flawed or misleading, remain available.[44]

The critical issue is how to indicate that these contributions are works in progress and have not been peer reviewed and improved based on comments or suggestions from peer scientists and editors. There needs to be vigilant screening especially for content that has implications for public health and safety. The quality of papers posted on preprint services vary, so does the quality of journals and conference papers. Still such outcomes facilitate acceleration of communication and progress.

Libraries' Role in the Adoption and Sustainability of Preprints

As the service and technology landscape is getting richer, an important question relates to the role of library publishing and repository initiatives vis-à-vis preprints. Originally created to showcase and enhance the impact of institutional scholarly outputs, the value of institutional repositories has been evolving. Communications around institutional repositories thus needs to be reframed to highlight their role in the context of evolving public access mandates, the increasing importance of research data, and growing interest in accessing various versions of scholarly outputs. Research libraries continue to invest in building and developing institutional repositories and aspire to build publishing programs.[45] Are there any opportunities for research libraries to play a role in designing and maintaining a preprint services infrastructure? One of the potential synergies to

---

[43] "Scientific Research on the Coronavirus Is Being Released in a Torrent: Will that Change How Science Is Published?" *The Economist,* May 7, 2020, https://www.economist.com/science-and-technology/2020/05/06/scientific-research-on-the-coronavirus-is-being-released-in-a-torrent.

[44] Kent Anderson, "Does bioRxiv Fulfill Its Purpose?" *The Geyser,* August 8, 2019, https://thegeyser.substack.com/p/does-biorxiv-fulfill-its-purpose.

[45] "Next Generation Library Publishing Initiative," Educopia Institute, https://educopia.org/next-generation-library-publishing/.

explore is whether enabling interoperability among repositories with related and complementary content would reduce duplication and increase efficiencies. For instance, pushing copies of papers published by scientists to their home institution's repository would facilitate communication and exchange between preprint servers and institutional repositories. Libraries can also promote and raise awareness about ethics and integrity issues related to preprints, similar to the role they have played in regard to predatory publishing.

With the increasing drive for open science, libraries have led efforts to reform scientific publishing. After several years of planning, the Sponsoring Consortium for Open Access Publishing in Particle Physics (SCOAP3) was initiated in 2014 to convert paywalled high-energy physics (HEP) journals to open access by shifting the publishing costs to libraries and research institutions while preserving the value provided by publishers.[46] Although successful for a particular discipline, it has not garnered much uptake in terms of replicating the model for other disciplines. arXiv was seen as a potential beneficiary of redirected funding administered by the SCOAP3 consortium, potentially subverting a fraction of arXiv's operating costs based on HEP coverage. However, this vision has not materialized, further illustrating the fractured nature of efforts to ensure sustainable innovation in scientific communication.

Amid growing concerns about their long-term durability and agility, several research libraries are partnering in the Invest in Open initiative, which aims to optimize the benefits of open source and OA tools that facilitate scholarship, research, and education.[47] Preprint services are often mentioned within the context of the research library's interest in coordinating the creation and development of a reliable, scalable, and interoperable open scholarly infrastructure. Before arXiv was transitioned to Cornell's Computing and Information Science, Cornell University Library as steward of the service demonstrated a model for research library engagement with a preprint service by developing a transparent and accountable business and governance model for arXiv.[48]

---

[46] "SCOAP3 – Sponsoring Consortium for Open Access Publishing in Particle Physics," https://scoap3.org/. Also see "arXiv Business Model White Paper," Cornell University Library, January 15, 2010, https://arxiv.org/about/reports/whitepaper.

[47] Information about Invest in Open can be found at "Invest in Open Infrastructure Launches," May 14, 2019, https://investinopen.org/2019/05/14/ioi-launch.html. Also, see Oya Y. Rieger and Roger C. Schonfeld, "Beyond Innovation: Emerging Meta-Frameworks for Maintaining an Open Scholarly Infrastructure," Ithaka S+R, October 21, 2019, https://sr.ithaka.org/blog/beyond-innovation-emerging-meta-frameworks-for-maintaining-an-open-scholarly-infrastructure/.

[48] A decision was made in 2018 to transition arXiv to Cornell's Computing & Information Science (an academic division) to increase ties with the computing and information science community, and to continue advancing innovations in scientific communication. See Oya Y. Rieger, "Transition FAQ: Move to Cornell Computing and Information Science," Cornell University, January 8, 2019, https://confluence.cornell.edu/display/arxivpub/Transition+FAQ%3A+Move+to+Cornell+Computing+and+Information+Science.

EarthArXiv (founded in 2017, hosting 1,500 papers) recently announced a partnership with the California Digital Library (CDL) to transition the earth sciences preprint service from the Center for Open Science to the eScholarship Publishing program at the CDL in order to ensure sustainability.[49] CDL will host EarthArXiv using Janeway, an open source publishing platform developed by the Centre for Technology and Publishing and the Open Library of Humanities at Birkbeck University of London. EarthArXiv's Advisory Council will maintain ownership and control over the preprint server, while the eScholarship Publishing team will contribute to the development, support, and maintenance of the Janeway platform.

Library-community led Confederation of Open Access Repositories (COAR) is spearheading efforts to promote the design of a distributed global networked infrastructure for repositories, on top of which layers of value-added services, such as overlay journals over existing preprint services, can be deployed.[50] Although this is an attractive vision, its viability will be determined by the global research community's ability to commit resources and align their priorities towards a common vision. Research libraries are mission-driven organizations with a commitment to facilitating enduring access to scholarly information. However, given their competing priorities and strained resources, it is not clear how they will contribute to the development of an open infrastructure to support preprint services.

The Potential Role of Preprint Services in a Public Access Compliance System

There is tremendous enthusiasm for enabling broader access to publicly funded research. However, the public access policies that delineate the grant conditions of funding organizations are diverse and complex. Although the common denominator is to

---

[49] CDL will host EarthArXiv using Janeway, an open source publishing platform developed by the Centre for Technology and Publishing and the Open Library of Humanities at Birkbeck University of London. See Justin Gonder, "EarthArXiv Announces New Partnership with California Digital Library to Host Earth Sciences Preprint Service," Office of Scholarly Communication, University of California, May 20, 2020, https://osc.universityofcalifornia.edu/2020/05/eartharxiv-announces-new-partnership-with-cdl/. Illustrating the plurality of approaches, in 2018 the American Geophysical Union and Atypon developed the Earth and Space Science Open Archive (ESSOAr) to accelerate the open discovery and dissemination of earth and space science with support from Wiley. Currently it includes 910 preprints and 1,500 posters. See "About ESSOAr," https://www.essoar.org/about.

[50] An example of a library's involvement in establishing an overlay journal is the collaboration of two scientists with the Queens' University in Canada to launch a peer-reviewed mathematics overlay journal built entirely on articles contained in arXiv. The relatively low costs of running the journal are being covered by the library, which is also providing administrative support. See "Progressing the Overlay Journal Model," COAR, 2019, https://www.coar-repositories.org/news-updates/progressing-the-overlay-journal-model/.

maximize the dissemination of funded research and make research outputs freely available to the public in a given time interval, the supporting systems are largely manual and involve resource-intensive processes. Given this landscape, some question whether preprint services could play a role in facilitating and ensuring compliance. Given the immature state of public access infrastructure, there are a number of open questions. Do preprint services have the functionality to support compliance? What are the potential risks for preprint services to take on this unfunded mandate? For instance, arXiv's Feedback on the Guidance on the Implementation of Plan S stresses that arXiv will need to identify additional resources to implement some of the changes required to make the service Plan S compliant, taking into consideration both one-time development and ongoing maintenance costs.[51] COAR is in the process of conducting a survey to assess the ease with which different repository platforms can comply with the Plan S mandates and recommended technical criteria.[52] The aim is to identify any challenges for repository platforms, related to compliance with Plan S, so we can communicate these issues with cOAlition S, and develop plans to help address these issues.

Overlay journals provide another context for considering the role of preprints vis-a-vis public access mandates. Over the past 20 years, the concept of overlay publishing, or layering journals on top of existing repository platforms, has developed from a pilot project idea to a handful of successful implementations.[53] However, it has not reached a critical mass yet.

User Experience

Early and fast dissemination continues to be a key motivation for sharing preprints. Given that peer review takes on average about four months, publishing an early version of a paper allows peers to recognize ongoing contributions and increases readership. One of the enablers of this speedy dissemination is the low entry-barrier for submission. Most preprint platforms require minimal metadata as scholars upload their papers, making the process rather easy and quick. This is a double-edged sword as this practice promotes ease-of-use while limiting processes that rely on rich metadata, such as consistent recording of version information.

---

[51] Erick Peirson, Oya Y. Rieger, Steinn Sigurdsson, and Licia Verde, "arXiv's Feedback on the Guidance on the Implementation of Plan S," *arXiv.org* (blog), January 2019, https://blogs.cornell.edu/arxiv/2019/02/04/arxivs-feedback-on-the-guidance-on-the-implementation-of-plan-s/ For information about Plan S, see "About Plan S," https://www.coalition-s.org/.

[52] "COAR and cOAlition S Supporting Repositories to Comply with Plan S," COAR, October 7, 2019. https://www.coar-repositories.org/news-updates/coar-and-coalition-s-supporting-repositories-to-comply-with-plan-s/.

[53] "How Journals Are Using Overlay Publishing Models to Facilitate Equitable OA," *Scholastica* (blog), October 25, 2019, https://blog.scholasticahq.com/post/journals-using-overlay-publishing-models-equitable-oa/.

From the users' perspective, preprints help level the playing field; there is no cost to either submit or read a preprint. This is especially essential for many scholars from institutions or countries that cannot afford subscriptions. Also, preprints allow the dissemination of scholarship that is not historically covered by traditional journals, including negative results, baseline but otherwise routine data, preliminary findings, methods and protocols, and short reports from projects.

With increasing emphasis on transparent, fair, and fast reviews, some preprint services are piloting annotation or commenting capabilities to allow readers to provide feedback. For instance, most medical and biomedical preprints have commenting features.[54] But early findings indicate that such commenting and annotation features are seldom used, raising questions about the affordances of such technologies in facilitating valuable online discussions.[55] For instance, based on a sample of nine services, a 2018 study identified only 135 Hypothes.is annotations across more than 9,000 papers.[56] From a technological perspective, a range of applications are currently available that support open review. However, the intriguing and "tricky" parts are much more in the sociology of science domain that involves human factors, especially those related to the reputation, fairness, power dynamics, bias, civility, and the qualifications of the participants in open review. These problems are not insurmountable, yet they certainly require the careful development of policies, procedures, and workflows that can ensure a trusted and useful environment for open review and annotation. An arXiv user study, which garnered 36,000 responses, revealed that users unanimously urged vigilance when approaching any changes and cautioned against turning arXiv into a "social media" style platform.[57]

ASAPbio is serving an important gateway role by acting as an awareness building and advocacy organization. For instance, they recently released a directory of preprint server policies and practices relevant to life sciences, biomedical, and clinical research to compare the scope, ownership, indexing, and preservation of fifty preprint platforms. A recent investigation indicates that 39 percent of journals surveyed do not provide clear information on whether preprints can be posted, and roughly 75 percent of journals have

---

[54] Jamie J. Kirkham, Naomi Penfold, Fiona Murphy, Isabelle Boutron, John PA Ioannidis, Jessica K Polka, and David Moher, "A Systematic Examination of Preprint Platforms for Use in the Medical and Biomedical Sciences Setting," bioRxiv, April 2020, https://doi.org/10.1101/2020.04.27.063578.

[55] Tom Narock and Evan Goldstein, "Quantifying the Growth of Preprint Services Hosted by the Center for Open Science," *Publications* 7, no. 2: 44 (2019), https://doi.org/10.3390/publications7020044.

[56] Ibid.

[57] Oya Y. Rieger, Gail Steinhart, and Deborah Cooper, "arXiv@25: Key Findings of a User Survey," Cornell University, July 2016, https://arxiv.org/abs/1607.08212.

no clear policy on citation of preprints. Researchers, especially the early career ones, would benefit from having clear journal policies in regard to preprint practices.

Looking Ahead

Preprints are here to stay and we should be prepared to see an increasing uptake. As the academic community's understanding of the value and risks associated with preprints continue to expand, there are a number of issues that will benefit from deeper exploration.

For instance, although we have seen the proliferation of preprints in different disciplines, the science-based fields continue to get more uptake and, perhaps in part for that reason, are subject to greater scrutiny. What are the opportunities, impediments, and open issues in preprints that focus on social science and humanities disciplines?

As stressed earlier, another open issue is that there are not yet any proven models to secure the financial stability of preprint services. Unlike other publication types that recover costs through article processing charges or by offering value-added services, the main premise of preprint services is that they support the open and free dissemination of scholarship with no restrictions. It is complicated to find a way to monetize preprints in order to create reliable and systematic revenue sources, although some consider offering fee-based value-added services on top of free access to preprints. Emerging publisher-led models that integrate preprints into their publishing workflows as early stage research will also need to address how to fund the additional efforts.

The concept of the scholarly record is broadening. Understanding how ideas evolve into knowledge is critically important. There is an increasing emphasis on sharing various outputs from the initial investigation to the final dissemination stage. The challenge is interconnecting different nodes of the publishing ecosystem where preprints, peer-reviewed articles, supporting data and code, and related comments, amendments can be discovered and interpreted by a range of researchers from a diverse range of settings. This is easier said than done as the scholarly communication infrastructure entails well-established policies, practices, and workflows that take time to adjust. Such an adjustment requires the alignment of many stakeholders including researchers, publishers, technology providers, standard developers, and funders.

As Ithaka S+R's deep dives indicate, disciplinary and subfield communities continue to share scholarly communication practices, but not all disciplines lend themselves to

preprints.[58] It is indisputable that preprints have taken off and we should expect further growth in this domain. Nevertheless, we need to be mindful as technologies may have negative or unintended consequences, with the potential impact on how we create, validate, share, and archive knowledge. We must consider the sustainability requirements upstream and remember that the services we are now experimenting with and creating have vital long-term implications.

Ultimately, what will determine the future of preprints is the level of uptake by different communities and the scholarly enterprise's capacity to support yet another mode in facilitating communication and exchange of ideas. Open access does not entirely remove fees and access limitations, but it replaces and reconfigures them for the key stakeholders in the scholarly communication endeavor. Given the trend toward interdisciplinary and multi-institutional collaborations, perhaps one of the meta questions about the future of preprints is how some of the current individual efforts that focus on specific research communities will align eventually to avoid a fragmented preprints ecology.

---

58 See "Research Support Services" for a list of Ithaka S+R's reports on the research practices of scholars in different disciplines, https://sr.ithaka.org/our-work/research-support/.