

Open Science Reading Group
Meeting 002: Intro to Open Access
October 27th, 2020 | 2:00 – 3:00 PST

Coordinator: John Borghi, PhD

Reading:

Fraser, N., et al. (2020). Preprinting a pandemic: The role of preprints in the COVID-19 pandemic. *BioRxiv*. <https://doi.org/10.1101/2020.05.22.111294>

Supplementary Readings:

Penfold, N. C., & Polka, J. K. (2020). Technical and social issues influencing the adoption of preprints in the life sciences. *PLOS Genetics*, 16(4), e1008565. <https://doi.org/10.1371/journal.pgen.1008565>

Suber, P. (2012). What is Open Access? In *Open Access (pp 1-27)*. MIT Press. <https://archive.org/details/9780262517638OpenAccess/page/n13/mode/2up>

Discussion Topics for the Breakout Rooms

Poll Question

Have you done any of the following:

1. Read a preprint
2. Published a preprint
3. Cited a preprint

John's Notes

Open Access in Open Science

- Last week was Access Week, where we in the library took a very broad few of what that means.
- We talked about using open infrastructure (specifically ORCID) to manage your research profile, we had a demo of the data sharing platform Dryad, and I even had a workshop about preprints. If you missed any of this, we have recordings and slides we can share.
- Today's session is a discussion of open access publishing, with a particular focus preprints. We'll follow a similar structure as last time, I'll provide a brief introduction and then we'll move to breakout rooms. Like last time, I've prepared some potential discussion questions. But I think there's enough to talk about related to preprints and publishing in the time of COVID that we'll have a lively discussion.

What is open access?

- Last month's reading stated that open access is the most developed element of open science. I think that's right, but that doesn't mean there isn't confusing terminology,

space for ambiguity, or lots of current events to keep up with.

- Earlier this month, the Howard Hughes Medical Institute announced an open access requirement.
- Just last week, Nature and Max Planck Digital Library announced a big open access publishing agreement.
- There was, of course, a lot of precedent and conversation about what we now call “open access” earlier, but the open access as we talk about it in 2020 grew out of a series of statements in 2002 and 2003.
 - [Budapest Open Access Initiative](#) (2002)
 - [Bethesda Statement on Open Access Publishing](#) (2003)
 - [Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities](#) (2003).
- The basic idea of OA is simple: Make scholarly literature available online without price barriers and without most permission barriers.
 - Advantages - More people can read/use the work.
 - Disadvantages (for want of a better word) - The incentive structure of science (metrics, etc) and the economics of scholarly publishing (APCs, etc).
- But open access is not a binary choice, it is a continuum. The [HowOpenIsIt?](#) guide is a good illustration of this, demonstrating some of the different dimensions of OA.
 - Reader rights (when readers are able to read an article)
 - Reuse rights (what people can do with the article, including reusing figures)
 - Copyright (whether the authors hold it or the publisher)
 - Author posting rights (how the authors can share the article)
 - Automatic posting and machine readability (how the articles are being made available for use by both humans and algorithms)
- There are also many routes to open access.
 - Open access publishing - Journals that publish exclusively open access articles. Many (Gold OA) charge article processing charges to authors. Some (diamond OA) charge neither authors or readers.
 - PLOS, eLife, and Scientific Reports are examples.
 - Hybrid Publishing and Bronze OA - Occasions where articles are made free to read, though often not “open” in other ways. Oftentimes, articles become free after a period of time or if an author has paid a fee.
 - PNAS is an example of a hybrid journal - where some articles are free to read and some are not. Journal of Experimental Medicine is a bronze OA journal, where articles are made free to read six months.
 - There have been a lot of examples of bronze OA during COVID, where a journal or publisher temporarily makes articles free to read.
 - Transformative Agreements - Agreements made between a publisher and a library to transition from traditional/closed access to open access publishing. The new Nature agreement with MPDL is a transformative agreement. The negotiations between UC and Elsevier involved transformative agreements.
 -

- Self archiving (Green OA) - The practice of placing a version of an author's manuscript into a repository, making it freely accessible independent of the publisher. What this looks like in practice depends on what version of the manuscript you're sharing.
 - Postprints are often shared through an authors website, an institutional repository, or similar infrastructure. PubMed Central is actually an example of Green OA, being a repository of papers resulting from NIH funded work.
 - Preprints are often shared through preprint servers. Though there are older (and larger) preprint servers used in other disciplines, the two main preprint servers for biomedical science are BioRxiv and MedRxiv.

What is a preprint?

- A preprint is a version of a journal article (or other work) that has not yet gone through the peer review process. Posting preprints is an example of “self archiving” a paper.
- There is actually a long history of preprints in biomedical science.
 - Long history of informal sharing of work pre-publication by researchers in specific subdomains (see [Garvey & Griffith, 1967; 1971](#)). A famous example is the “[RNA Tie Club](#)” that began in the 1950s.
 - Information exchange groups at NIH during the 1960's were an organized effort to distribute preprints. They were very successful until shutting down, due to pressure from journal publishers (see [Cobb, 2017](#)).
 - Starting in 1969, NEJM (and soon many other biomedical journals) adopted the [Ingelfinger rule](#), a hard line against publishing work that had previously been published elsewhere.
 - The Ingelfinger rule was eventually [broadened](#) to be against reporting scientific results directly to the media before they are published in peer reviewed journals.
 - NEJM made their [policy against preprints explicit in 1995](#) (though they welcome preprints now).
 - The Ingelfinger rule was eventually [broadened](#) to be against reporting scientific results directly to the media before they are published in peer reviewed journals.
- In parallel: Librarians at SLAC develop Stanford Physics Information Retrieval System (SPIRES) in the 1960's to be a database for the physics literature including preprints. ([O'Connel, 2000](#)).
 - SLAC began to disseminate lists of preprints - Preprints in Particles in Fields - in 1969.
 - In 1991, SPIRES actually became the first website in North America - which shows how closely our current conversation about preprints and open access publishing is tied into the development of the internet.
 - The missing component of SPIRES was the ability to read the full text of preprints.
 - Enter arxiv. In 1989, Joanne Cohn (now at UC Berkeley) began an e-mail list for

sharing preprints related to theoretical physics. A few years later Paul Ginsburg created a repository that, in 2001, became known as arXiv.

- ArXiv is now managed by Cornell University library and includes preprints related to a large number of scholarly disciplines.
- There was a proposal in 1996 to make a large chunk of the biomedical research literature freely available - [E-biomed](#). Included provisions for sharing preprints. Eventually this proposal became PubMed Central.
 - PubMed Central (PMC) is a free digital repository that archives open access full-text scholarly articles that have been published in biomedical and life sciences journals.
 - Many open access journals have their full text included in PubMed Central as well as papers resulting from NIH funded research (under the NIH Public Access Policy).
 - With the addition of preprints to PubMed Central, we're moving closer to the initial proposal.
- BioRxiv launched in 2013 as a preprint server for biology.
- MedRxiv launched in 2019.

Preprints in a pandemic

- While overall adoption was low, preprints were used to disseminate results during both Ebola and Zika outbreaks ([Johansson et al. 2018](#)).
- COVID-19 has drawn a lot of attention to preprints.
 - Over 40% of the literature published about COVID has been work published in preprint form.
 - Preprints about COVID have been highly cited, widely discussed, and often reported on.
 - The posting of preprints about COVID has described alternately as a huge and beneficial change in how science is communicated and a major contributor to disinformation during the pandemic.
- But the entire system we use for scholarly communications has come under tremendous strain during the pandemic.
 - First COVID-related retraction - [Uncanny similarity of unique inserts in the 2019-nCoV spike protein to HIV-1 gp120 and Gag](#), Authors voluntarily withdrew the preprint.
 - Disinformation in a preprint - [Unusual Features of the SARS-CoV-2 Genome Suggesting Sophisticated Laboratory Modification Rather Than Natural Evolution and Delineation of Its Probable Synthetic Route](#). Likely fraudulent, almost certainly intentional disinformation. Notably submitted to a server without any screening processes.

Potential Discussion Prompts:

1. What are your perceptions of preprints? Do you read or publish preprints? Why or why not
2. What resources would be helpful to you as we all navigate the evolving landscape of open access and preprints? How can we help each other?
3. Much of the resistance to preprints in biomedical science has related to unintended consequences of disseminating results that have potential health implications before they go through peer review. How do you think that has manifested during the COVID-19 pandemic?
4. The reading described how COVID-19 has affected researcher behavior related to preprints (more first-time authors, shorter preprints, more downloads, wider sharing). Do you think this behavior is specific to the pandemic? What do you think the long-lasting effects of this will be (if any)?
5. The [NIH public access policy](#) has led to the majority of publications arising from NIH-funded projects to eventually be made available through PubMed Central. But what do you think about more institution-specific mandates? For example, policies like those at [Stanford's School of Education](#), [HHMI](#), or the [UC Open Access Policy](#) - both of which involve self-archiving.