



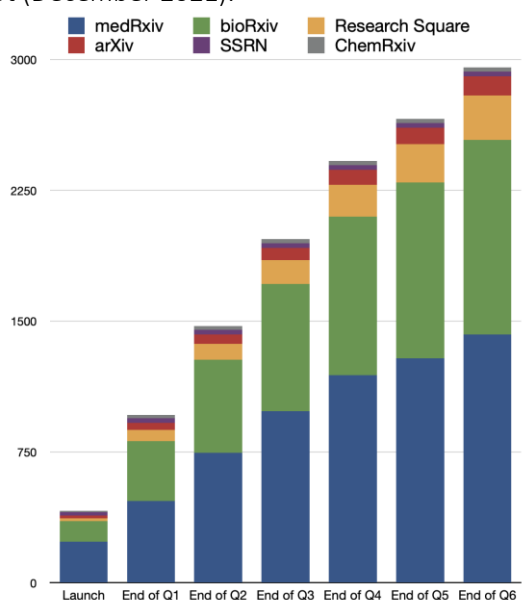
NIH Preprint Pilot – Q5/Q6 Status Update

On June 9, 2020, NLM [launched a pilot project](#) designed to test the viability of making full-text preprints resulting from NIH-funded research searchable in PubMed Central (PMC), with an accompanying citation discoverable in PubMed. Including preprints in PMC and PubMed is a way to enhance their discoverability.

Phase 1 of the NIH Preprint Pilot is focused on preprints resulting from NIH-funded COVID-19 research, curated from the [NIH iSearch COVID-19 Portfolio](#). Phase 1 impact assessment activities are ongoing to inform planning for subsequent phases of the pilot.

Preliminary Results – as of the end of the sixth quarter of the pilot (December 2021):

- Nearly 3,000 preprint records have been added to PMC and made discoverable in PubMed.
- Preprint records have been viewed 2.5 million times in PMC. Associated preprint metadata and abstract records have been viewed 2 million times in PubMed.
- Just over half of the preprint records in PMC and PubMed are matched to a published journal article. The likelihood of a preprint being linked to a published journal article increases somewhat with time (e.g., 60% of preprints posted in 2020 are linked to a peer-reviewed journal vs. 43% of those posted in 2021).
- Two-thirds were available under a Creative Commons (CC) license. NLM only converts the full-text of preprints for the CC-licensed papers. The remaining one-third of pilot preprints are metadata- and abstract-only records with links to the preprint server full text.



Q5/Q6 Implementation Details

- **Scope:** Preprints acknowledging NIH support and relating to COVID-19.
- **Preprint Server Sources:** arXiv, bioRxiv, ChemRxiv, medRxiv, Research Square.
- **Workflows:** New preprint records uploaded on a weekly basis. Checks for updated preprint and journal article versions run 1-2 times a week. Additional checks for withdrawn preprints run weekly.
- **User Interface:** Preprint records include preprint banners and citations, links to preprint servers via DOI and LinkOut, links to additional preprint versions, as well as resulting journal articles (when available). Improvements to search results display continue to be explored.

Outcomes

Faster dissemination

- Approximately 55% of preprint records in PMC have been linked to a published journal article.
- Based on a sample of 750 preprints included in the pilot and associated with a journal article, the average time from preprint posting to journal publication is 100 days; the longest period between preprint posting and journal publication in the sample was one year.

Increased discoverability

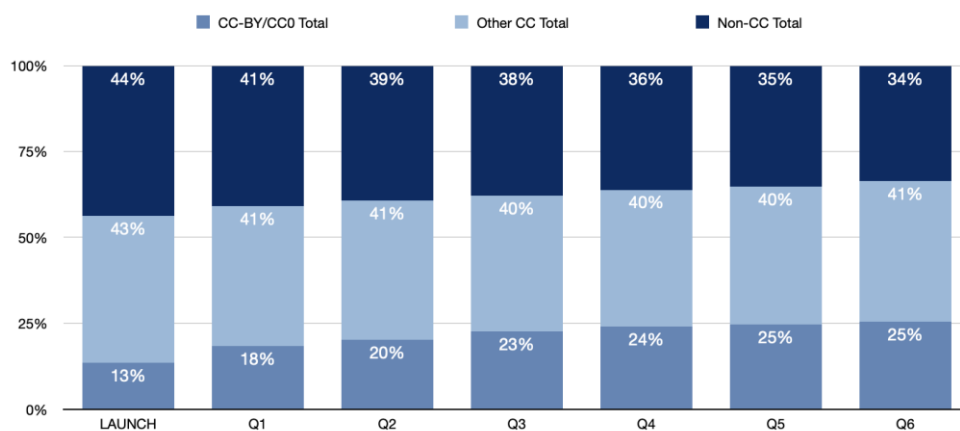
- More than two-thirds of readers who access preprint records in PMC come from a third-party search engine; availability of the preprint full text in PMC may play a role in increased discovery through these channels as PMC preprint views outpace PubMed preprint views.
- An analysis of PubMed search data in October 2021 found that 99% of preprint records were returned in search results and 98% were viewed.
- Others come to preprint records through PubMed search and social media.

Expanded access

- Forty percent of pilot preprints posted in 2020 remain unassociated with a journal article. The impact of preprints on expanding access to publicly funded research in addition to accelerating dissemination is an area for further exploration. Understanding the characteristics of and author motivations for “unpublished” preprints (e.g., null or negative results) could provide a more complete picture of the role of preprints in scholarly communications and communicating the results of Federally funded research.

Awareness of NIH-recommended preprint practices

- The percentage of preprint records posted under a Creative Commons license has gradually increased over the course of the pilot. A quarter of the preprints in the pilot are made available under an NIH-recommended CC BY (with attribution) or CC0 (public domain) license. The most common Creative License selected by authors is the most restrictive: CC BY-NC-ND (Non-Commercial, No Derivatives).



- More than three-quarters of the preprints available in full-text format include indicators of data sharing (i.e., a data availability statement or supplementary material).

Stakeholder Engagement

- A survey of PubMed and PMC users who accessed preprint records in July and August 2021 indicates that respondents are familiar with preprints, support their inclusion in PubMed and PMC, and do not find they diminish trust in NLM resources—some report increased trust.