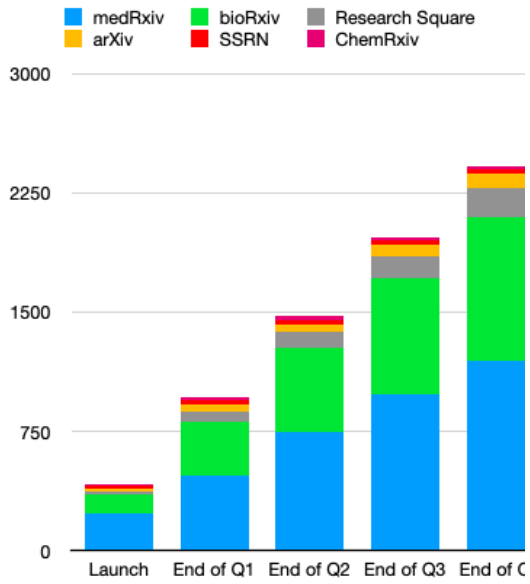
NIH Preprint Pilot – Q4 Status Update

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. The pilot has been running for one year (4 quarters).

Preliminary Results – as of the end of the fourth quarter of the pilot:

- Nearly 2,500 preprint records have been added to PMC and made discoverable in PubMed, up from approximately 2,000 at the end of Q3. Half have been posted to the preprint server medRxiv.
- Preprint records have been viewed more than 1.8 million times in PMC. Associated preprint metadata and abstract records have been viewed approximately 1.5 million times in PubMed.
- Approximately 44% of the preprint records in PMC and PubMed are matched to a published journal article, up from 40% at the end of Q3.
- Preprints included in the pilot have been cited approximately 12,000 times.
- Availability of preprint records under a Creative Commons (CC) license held steady from Q3 to Q4.



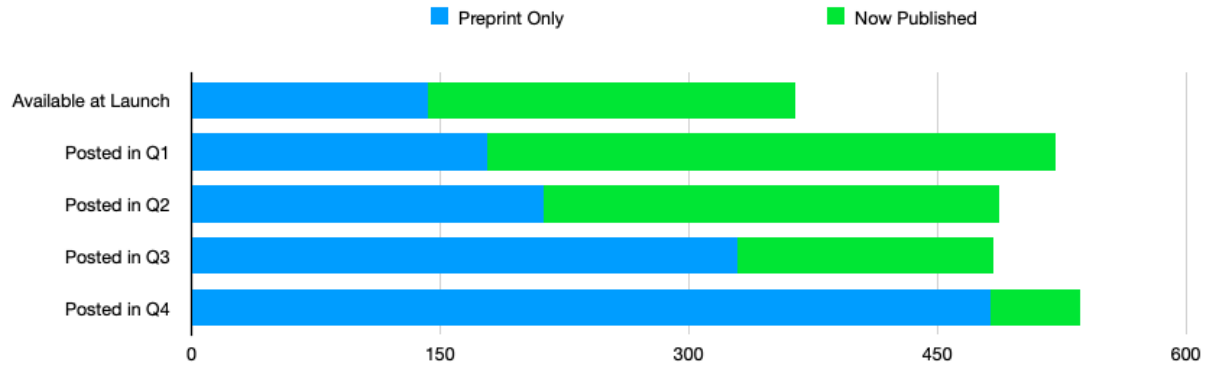
Q4 Implementation Details

- **Scope:** Preprints acknowledging NIH support and relating to COVID-19.
- **Preprint Server Sources:** arXiv, bioRxiv, ChemRxiv, medRxiv, Research Square, and SSRN.
- **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 Additional checks for withdrawn preprints run weekly.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 44% of preprint records in PMC have been linked to a published journal article.
- Based on a sample of 750 preprints included in the pilot, the average time from preprint posting to journal publication is 100 days; the longest period between preprint posting and journal publication in the sample is one year.



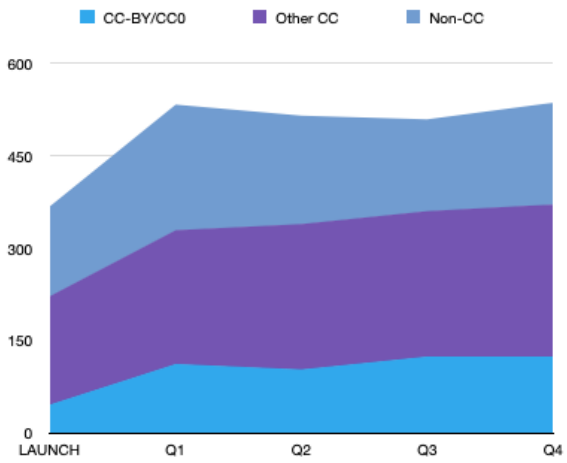
Increased discoverability

Traffic to preprints in PMC and PubMed comes primarily from third-party search engines, direct search in PubMed, or social media links.

- In PMC, more than 4,500 users viewed nearly 6,000 preprint records on an average day in Q4.
- In PubMed, preprint record views increased in Q4, with an average of 5,200 users viewing more than 7,000 preprint records each day.

Awareness of NIH-recommended preprint practices

- The share of preprint records posted under a Creative Commons license has increased over the course of the pilot. Application of the NIH-recommended CC-BY or CC0 license leveled off from Q3 to Q4.
- 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 on Preprints

- Facilitated focus groups with researchers, clinicians, librarians, and journalists to better understand different perceptions of preprints among stakeholder groups.
- Provided updates on the NIH Preprint Pilot at events hosted by NISO and the Society for Scholarly Publishing.