SPARC

Response to Call for Feedback on
"*Preparing for the Next Pandemic*" White Paper
Senate Committee on Health, Education, Labor & Pensions
Submitted: June 25, 2020

I am writing on behalf of SPARC (the Scholarly Publishing and Academic Resources Coalition), a membership organization of more than 240 academic and research libraries working to promote the expanded sharing of scholarship in the networked digital environment. We applaud the HELP Committee's leadership in creating an action plan for future pandemic preparedness and we thank you for the opportunity to submit additional recommendations.

We encourage the Committee to consider adding a foundational recommendation at the beginning of *Section 1: Tests, Treatments & Vaccines – Accelerate Research and Development.*

Recommendation: Congress and federal agencies should ensure that articles reporting on federally funded research and the data needed to validate/reproduce their results are made immediately accessible in an open and AI-ready, machine-readable format to scientists and the public.

Enabling researchers to build a knowledge base to help them rapidly identify and understand emerging infectious diseases is a critical first step in pandemic preparedness. When the COVID-19 crisis struck, there was a widespread assumption that researchers had ready access to all of the existing scientific articles on the coronavirus and related diseases. However, that assumption turned out to be incorrect and an important gap in our national readiness to effectively deal with the pandemic was identified.

When scientists attempted to access the full corpus of scientific papers related to COVID-19 to read, text and data mine, they discovered they could not – because no such collection existed. This meant that scientists could not quickly establish a shared baseline understanding of this new virus, thereby delaying the development of research routes for testing, treatments and vaccines. Most papers were behind paywalls on proprietary publisher platforms, and access had to be specifically requested by a group of National Science and Technology Advisors from 12 countries, including the Director of the U.S. Office of Science an[1]d Technology Policy. In a letter to the scholarly publishing community sent by group March 12, 2020, it stated:

> *"A topic of considerable interest is enhancing the ability of researchers and other stakeholders to access and re-use or text-mine all published articles on coronaviruses,*

[1] https://www.whitehouse.gov/wp-content/uploads/2020/03/COVID19-Open-Access-Letter-from-CSAs.Equivalents-Final.pdf

> *SARS-CoV-2, and COVID-19. This **timely access is critical**, as it allows researchers keep up with the rapidly growing body of literature and identify trends and relevant information in efforts to characterize this novel virus and address the associated global health crisis…Importantly, **this information should be in both human and machine-readable format** to allow for full text and data mining using artificial intelligence with rights accorded for research re-use and secondary analysis."*

The only way this group was able to gather the required articles was to request permission for their use from each publisher (a very time intensive endeavor), and then task the U.S National Institutes of Health (NIH) with converting them into a format that researchers could readily download and begin to analyze in the new CORD-19 Database[2], hosted by NIH's PubMed Central (PMC). While many publishers stepped up to make their articles openly available, some still have not. And some publishers have only allowed access up until the coronavirus crisis is deemed passed, and then the articles will once again be placed behind a paywall.

The US should never be in a position where its citizens, let alone its own government, does not have ready access to the outputs of the $65 billion research it has funded on behalf of taxpayers. We should not have to wait until there is a public health crisis to unlock research that could have prevented a pandemic in the first place. Articles reporting on science funded by the U.S. government should always be readily accessible to the public – particularly in the time of a pandemic.

The value of creating an open and machine-readable collection of scientific articles was immediately apparent. The CORD-19 database was established in March 2019, and within the first two weeks its articles had been accessed more **than 2 million times**. By June, PMC's coronavirus collection had grown to include the over 59,000 articles and had been accessed **more than 18 million times[3].**

There is a strong, consistent policy precedent already in place to allow for this recommendation to be quickly enacted:

- 2008: *The Consolidated Appropriations Act, 2008 (Division G, Title II, Section 218 of PL 110-161)* required all investigators funded by the NIH to submit an electronic version of their final peer-reviewed journal manuscripts to be made publicly available online no later than 12 months after the official date of publication.

- 2010: Requirement was extended to all LHHS/E agencies through the Consolidated Appropriations Act of 2010, and *The America COMPETES Reauthorization* Act called on the White House Office of Science and Technology Policy to coordinate with all federal science agencies to develop policies to ensure widespread public access to and long-term stewardship of the results of federally funded research.

[2] https://ncbiinsights.ncbi.nlm.nih.gov/2020/03/26/cord-19-a-new-machine-readable-covid-19-literature-dataset/
[3] https://sparcopen.org/news/2020/strong-community-response-to-free-scholarly-article-access-to-fight-covid-19/

- 2013: All U.S. federal agencies with research and development expenditures of $100 million or more annually were required to implement public access polices to their funded research by White House Executive Memorandum[4] signed by the Office of Science and Technology Policy Director John Holdren.

While these policies provided important steps forward, they allow up a to full year's delay in access to critical research, and do not require that be made available in a standard, machine readable, AI-ready format. Congress should adjust existing policy to remove the one-year waiting period and require all federally funded articles to be made immediately available in machine readable format. Such a step has strong support; organizations representing millions of Americans have written to OSTP this year calling for the enactment of a zero-embargo policy for federally funded research outputs.

Thank you for your consideration,

Heather Joseph
Executive Director, SPARC

4 https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf