SPARC Response to Federal Register Request for Information 85 FR 9488: Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting from Federally Funded Research.

Submitted on behalf of SPARC (The Scholarly Publishing and Academic Resources Coalition) by Heather Joseph, Executive Director, SPARC, May 6, 2020

I am writing today on behalf of SPARC (the Scholarly Publishing and Academic Resources Coalition), a membership organization of more than 240 academic and research libraries promoting the expanded sharing of scholarship in the networked digital environment. We thank the Office of Science and Technology Policy for your efforts to convene a substantive discussion on the importance of ensuring broad public access to the results of federally funded research.

Question # 1 "What current limitations exist to the effective communication of research outputs (publications, data, and code) and how might communications evolve to accelerate public access while advancing the quality of scientific research?"

Scientific research is critical to the well-being of society. The U.S. government annually invests ~$65 billion in taxpayer dollars into research, with the expectation that the results will be made widely available in order to speed discoveries, turn breakthroughs into treatments and cures, and improve the lives of the American public.

Yet the reality is that the articles reporting on this research are not widely available to the public. Once the purview of not-for-profit publishers and university presses, scientific journal publishing is now dominated by large commercial companies. Research articles are locked behind expensive journal subscriptions, which routinely cost thousands of dollars per journal – and up to $50 for temporary access to a single article.

SPARC's member libraries invest tens of millions of dollars in journal subscriptions each year to serve the needs of their researchers, faculty, and students, but there is no library that can even come close to affording access to all of the journals that they are asked to provide. In fact, right now, there is a growing trend towards canceling commercial journal subscriptions, as libraries examine the true value this model provides – and increasingly opt not to support 30-40% publisher profit margins in favor of exploring new, more values-aligned ways for researchers to share their results. The recent decision by the University of California System to cancel its $11 million Elsevier journal package highlights this trend. In just the past month, the State University of New York (SUNY) System also opted to cancel its $7 million Elsevier package, and the University of North Carolina and Iowa State University also followed suit.

Access is only part of the problem. Even when researchers can get to them, articles often are not in formats or on platforms where they can be used in the ways that are needed. As our nation's experience with COVID-19 highlights, it is critical for scientists to be able to text mine these papers and use machine learning or AI to fully unlock their value. They cannot perform this kind of analysis when articles are on proprietary platforms and are not in open and machine-readable formats.

The need for better access to the data resulting from taxpayer funded research also must be addressed. The current system makes it routine for scientists to hold onto data for years without sharing it while they wait for papers to be published. Part of this problem is driven by an outmoded incentive system, but part is also driven by the lack of consistent standards. Even when researchers do share data it is often in formats that make it hard to use or lacks the computer code and tools needed to interpret it. Even the data needed to verify or reproduce the results of published articles is often unavailable. This erodes trust in scientific research at a time when it is of particular importance.

Our nation has the opportunity to optimize the system of scientific communication. There is no more compelling illustration of the need for an open system than the current pandemic. As COVID-19 emerged, the very first thing scientists did was rush to make any and all information on the emerging disease openly available. Researchers released the genetic sequence of the virus, posting it in GenBank for their colleagues around the world to access. They began openly sharing preprints, data, code, and other insights through open, online platforms so that anyone could immediately get to work on understanding the disease and start innovating towards testing, treatments, and vaccines.

This unfolded essentially in real time, with one exception. When scientists wanted to access the corpus of previously published scientific papers related to COVID-19, they could not – because no such collection existed. Most of the papers were locked away in individual publisher collections, and access had to be specifically requested by a group of National Science and Technology Advisors from 12 countries – including the U.S. The group wrote:

> *"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, 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."*

While many publishers stepped up quickly to make their articles openly available, some did not. Some have given only temporary access and will re-paywall content once the coronavirus crisis is deemed passed.

As Americans, we should never again be in a position where the U.S. government does not have ready access to the outputs of the research it has funded on behalf of taxpayers. Articles

reporting on science funded by the U.S. government should always be readily accessible to the public. The results of research funded by taxpayers should not be kept locked behind glass that is only broken in the case of an emergency.

Question #2: "What can Federal Agencies do to make taxpayer funded research results, including peer-reviewed author manuscripts, data, and code funded by the Federal Government, freely and publicly accessible in a way that minimizes delay, maximizes access, and enhances usability? How can the Federal Government engage with other sectors to achieve these goals?"

The federal government should act without delay to implement a strong national policy ensuring that taxpayers receive immediate, barrier-free access to the full results of the scientific research that their tax dollars have funded. In doing so, the Trump Administration can end the needless delays that researchers currently face and provide all citizens with hope that issues that most directly affect them and their families are being worked on with the same urgency and efficiency as the coronavirus.

At minimum, this policy should require:

- All articles reporting on federally funded research and the corresponding data and tools needed to validate their conclusions (software, code, etc.) should be made freely available online to the public immediately upon publication.

- Articles must be made available in open and machine-readable formats that fully enable productive reuse including text/data mining and computational analysis. Title II of the Foundations for Evidence-Based Decision Making Act provides guidance on this.

- Articles and data should be made available under an open license or be published as part of the worldwide public domain (specifically a Creative Commons Attribution 4.0 International (CC BY) license or similar for articles, CC0 for data).

- A copy of a researcher's final accepted manuscript or final published article should be made available via either a digital repository maintained by a U.S. federal agency (NIH's PubMed Central is a robust, cost-effective option) or in an open, non-proprietary repository designated by the agency that ensures long-term open access to and preservation of these articles.

- Corresponding data and software should be made available via a digital repository maintained or approved by a U.S. federal agency. Further details on the desirable characteristics of such data repositories were submitted by SPARC to OSTP in an earlier submission to this RFI available here.

- All other non-classified data not directly attributable to a publication, including associated metadata, should be made available to the public as soon as possible under findable, accessible, interoperable, and reusable (FAIR) terms and conditions.

We encourage journal publishers to experiment with a variety of business models that support open and equitable access to and participation in science. SPARC and the wider library community have a long history of partnering with publishers to develop innovative business models (including the recent "Subscribe to Open" model) and collective support models, and we are eager to continue and expand this work.

We ask OSTP to encourage federally funded researchers to utilize platforms that enhance the speed and transparency of scientific communication – including preprint servers and research funder platforms such as Gates Open Research. These platforms are being heavily used by researchers in the fight against the coronavirus, and we encourage OSTP to consider collaborative efforts with private research foundations to support them.

It is important that OSTP continue to explore mechanisms to reward federally funded researchers for openly sharing their research outputs. We applaud OSTP's commitment to collaborating on this important effort through the recent joint meeting of the White House JCORE group and the National Academies of Science Roundtable on Realigning Research Incentives and believe that this collaboration could play an important role in ensuring the ultimate success of this policy.

As we have seen with COVID-19, time is of the essence. OSTP should establish this policy immediately. For the past 15 years, the U.S. government has moved deliberately and incrementally towards a national policy and has ample experience and data to justify the need for and benefits of final implementation. The U.S. should join the European Union, Canada, and other leading nations that have already established strong national open access policies in order to promote advances in science and technology, encourage innovation and economic growth, and improve the public good.

While the need to establish this policy is urgent, we recognize that it cannot be fully implemented overnight. We recommend a transition period of up to 18 months for stakeholders to prepare implementation plans. SPARC and our member libraries are firmly committed to working with the federal government and all stakeholders to support a smooth and effective implementation process.

We are eager to work with academy-friendly players – particularly scholarly societies – to develop financial risk-mitigation strategies to smooth their transition to providing open access to this content. Libraries are uniquely positioned to do so, as library dollars account for more than 75% of the average journals subscription revenue, which plays a significant role in supporting the overall operations of many scholarly societies.

Question #3: "How would American science leadership and American competitiveness benefit from immediate access to these resources? What are potential challenges and effective approaches for overcoming them?"

Science sets the pace for economic growth. What starts in the lab ends up in the pharmacy, in the factory, and on the farm. How quickly science progresses directly impacts America's ability to innovate, and open sharing accelerates the progression of research. As a result, governments around the world are making open the default for their publicly funded

research—directly tying these policies to their national innovation agendas. In this global context, the Administration has the opportunity to bring the U.S. back to the forefront in defining how science will be conducted in the networked digital environment, while simultaneously achieving other key goals such as making research AI-ready.

Openly available research is viewed more frequently, cited more regularly, and built upon more often than closed research. The Human Genome Project provides a powerful case in point, generating an estimated $965 billion in economic activity on a $3.8 billion taxpayer investment. This outsized return on research investment was only possible because anyone could build on the results without cost or permission, whether in a corporate lab or a garage.

Conversely, the current delay and difficulty in accessing the results of research negatively impacts businesses, entrepreneurs, patients, and the public – all of whom are eager for the latest science to be translated into materials that the manufacturing sector can use to make tests, treatments, and vaccines.

The move to work openly in order to accelerate COVID-related research affirms the advantage of openness in accelerating research broadly. Researchers are choosing fast, open channels such as preprint servers to share the results of research out as quickly as possible; and some publishers are innovating by creating new, collaborative mechanisms to provide peer review in as close to real time as possible, while ensuring that rigor and reproducibility remain paramount. Efforts like these provide excellent blueprints for publisher services in a fully open access era.

Through initiatives such as the Virus Outbreak Data Network (VODAN), the research community is collaborating to make COVID-related data openly available, FAIR-compliant and machine readable, and ready for analysis by research teams around the world. As this transition to open, collaborative, and data-intensive research is accelerated by COVID-19, a strong national open science policy will provide the foundation needed for continued American leadership in science.

Data demonstrates that a repository-based policy can be realized in a cost-effective manner. Representing more than half of the total amount that the U.S. government spends on research, the NIH has estimated that the cost for making all of its funded articles publicly available through PubMed Central is less than $5 million per year, only 1/90th of 1% of the NIH's overall budget, with the costs remaining relatively flat over the past decade. By contrast, recent research has shown that there is already evidence of hyperinflation in the Article Processing (APC) market, with costs nearly doubling over the past decade.

We applaud OSTP's extensive efforts to convene stakeholders in the consultation process while considering a strong national open access policy, and encourage you to continue these efforts throughout the implementation process. We stand ready to collaborate towards ensuring that U.S. federally funded research can achieve the maximum public good.