



SPARC Response to OSTP Request for Information to Improve Scientific Integrity Policies (86 FR 34064)

Submitted on behalf of SPARC (The Scholarly Publishing and Academic Resources Coalition) by Heather Joseph, Katie Steen, and Nick Shockey, July 28, 2021.

[SPARC](#) (the Scholarly Publishing and Academic Resources Coalition) is an advocacy organization working on behalf of more than 240 academic and research library members to make research and education open and equitable by design. We believe that sharing knowledge is a fundamental human right, and that strong open science policies that reflect this right will simultaneously increase scientific integrity while also increasing and diversifying participation in the scientific enterprise.

We thank OSTP for the opportunity to provide comments on federal scientific integrity policies and to make suggestions for their improvement.

Open Science is Foundational to Ensuring Integrity and Equity in the Scientific Enterprise

A strong open science policy should be a cornerstone of the Biden Administration's approach to improving scientific integrity. Open science can significantly advance numerous areas that impact public trust in research—including reproducibility, error and fraud detection, prevention of the suppression or distortion of scientific findings, and the advancement of equity in research.

The open communication of research outputs (including articles, data, software, code, and algorithms) can significantly boost scientific integrity by dramatically increasing the number of reviewers and enabling the use of new automated tools for detecting errors often missed by human readers.

While there are numerous examples of openness enabling a more powerful review of research results, a specific example is instructive. Researchers Michèle Nuijten and Sacha Epskamp developed a tool called *statcheck* that *Nature* described as “a spellchecker for statistics.”¹ The program automatically scans articles for statistical

¹https://www.nature.com/news/polopoly_fs/1.21049!/menu/main/topColumns/topLeftColumn/pdf/540151a.pdf?origin=ppub

results, recomputes the calculations and checks that the numbers match. In reviewing 30,717 Psychology papers to identify 16,695 that tested hypotheses using statistics, statcheck found at least one potential error in half,² demonstrating both the magnitude of the challenge and the utility of automated tools in meeting this challenge.

Nuijten and Epskamp's collaborator, Chris Hartgerink, extended the tool to automatically notify both the authors and readers of the potential error in papers analyzed by statcheck. When a statistical inconsistency was detected, Hartgerink's script automatically posted a message on [PubPeer](#), an open, non-profit platform where people share and discuss scientific articles.³ This extension of the statcheck script demonstrates how these tools can be applied not only to detect errors but also to begin to correct the scientific record through communication on open platforms.

Open science can be a powerful counter to the suppression and distortion of scientific findings. Openly licensing articles and data allows them to be distributed, shared widely, and mirrored in different repositories, avoiding a single point of failure where important results may be removed from a website or repository. Opening up the scientific evidence underlying government policies and recommendations can boost public confidence, particularly in difficult or politically-charged decisions. In a Joint Appeal for Open Science by CERN, OHCHR, UNESCO and WHO, the UN High Commissioner for Human Rights, Michelle Bachelet, specifically highlighted the importance of opening up research results to bolster public support for the measures necessary to combat both COVID-19 and climate change effectively.⁴

In the same statement, Commissioner Bachelet raises the opportunity that open science presents to "promote the inclusion of scholarship by people whose contributions and needs are too often overlooked."⁵ Well-constructed policies that support both open participation in and access to the results of science can boost public confidence in research by making it more representative of the public it seeks to serve.

By limiting participation to the select few who can afford to pay, the legacy, subscription-based model of distributing research findings is fundamentally at odds with the advancement of equity in research. The exorbitantly expensive "open" options now presented by the largest commercial publishers simply flip restrictions on access to restrictions on participation, locking out authors who cannot afford to pay to publish by

² <https://link.springer.com/content/pdf/10.3758/s13428-015-0664-2.pdf>

³ <https://www.vox.com/science-and-health/2016/9/30/13077658/statcheck-psychology-replication>

⁴ <https://www.ohchr.org/EN/NewsEvents/Pages/DisplayNews.aspx?NewsID=26433&LangID=E>

⁵ Ibid

charging fees that can exceed \$11,000 per article.⁶ To address the barriers presented by both paywalls and prohibitively expensive publication fees, OSTP should strengthen the 2013 Memorandum on [Increasing Access to the Results of Federally Funded Scientific Research](#) in the key areas described below that can put equity and the public interest at the core of U.S. open science policy.

Effectiveness of current agency policies

The [2010 OSTP Memorandum](#) on Scientific Integrity identifies “facilitating the free flow of scientific and technical information” and “establishing principles for conveying scientific and technical information to the public” as foundational to scientific integrity in government, but we have largely fallen short in implementing effective policies that build on these key principles. The 2013 OSTP memorandum on [Increasing Access to the Results of Federally Funded Scientific Research](#) was designed to address that gap, and directed federal science agencies to develop plans to improve public access to the results of their funded research.

While this Memorandum resulted in federally-funded research being made available more rapidly and broadly than before, as we saw during the COVID-19 crisis, it has fallen far short of providing scientists and the public with what is needed: a free, immediately accessible, fully machine-readable collection of articles reporting on taxpayer-funded research.

Eight years after the issuance of the 2013 Memorandum, many articles reporting on taxpayer-funded research are not readily available to most scientists (or to the public) for a full year after publication. They are still difficult to locate, housed on publisher’s proprietary websites, and published in formats and under licensing terms that make them impossible to use fully through text and data mining.

In fact, in November 2019, the Government Accountability Office (GAO) released a report on [Additional Actions Needed to Improve Public Access to Research Results](#) which reviewed the progress of 19 federal agencies on the implementation of their public access plans as directed by the memorandum. According to the report, 11 agencies still do not fully ensure that researchers comply with public access policies, and 7 agencies have not taken steps to make data underlying publications findable and accessible.

Rather than focusing on these compliance challenges, we have the opportunity—and the imperative—to revamp the 2013 Memorandum to establish a strong, uniform policy

⁶<https://www.forbes.com/sites/madhukarpai/2020/11/30/how-prestige-journals-remain-elite-exclusive-and-exclusionary/?sh=af9deb94d48f>

that ensures equitable open access to and participation in U.S. federally-funded scientific research. We encourage OSTP to consider the following recommendations:

Recommendation 1: Update the 2013 Memorandum on Public Access to Federally-funded Research Outputs

To boost scientific integrity, improve research assessment, and increase equity, we recommend the federal government update OSTP's 2013 Memorandum to require a consistent policy across federal agencies that ensures articles reporting on all basic and applied federally-funded research and the data, software, code, and algorithms needed to validate and/or reproduce their results are made immediately available with no embargo period to the public, at no cost, under an open license, and in an AI-ready, machine-readable format.

Specifically, the policy memorandum should be updated to require that:

1. All articles reporting on federally-funded research and the underlying data and tools needed to validate their conclusions (including software, code, and algorithms) should be made freely available online to the public immediately upon publication.
2. Articles should 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.
3. Articles and underlying data should be made available under an open license or be published as part of the public domain, specifically a Creative Commons Attribution 4.0 International (CC BY) license or similar for articles and CC0 for data.
4. To minimize the cost to researchers and increase contributions from more diverse voices, a copy of a researcher's final accepted manuscript or final published article should be made available via a digital repository maintained by a U.S. federal agency or in an open, nonproprietary repository designated by the agency that ensures long-term open access to and preservation of these articles.
5. Underlying data, software, and code should be made available via digital repositories maintained or approved by a U.S. federal agency. (Further details on the desirable characteristics of data repositories were submitted by SPARC to OSTP in an earlier submission to this RFI available [here](#).)
6. 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 and adhere to findable, accessible, interoperable, and reusable ([FAIR](#)) principles.

Recommendation 2: Support the policy update with federal funding for the development and ongoing sustainability of critical research infrastructure

Any policy update should be supported by federal investment in research infrastructure so that scientists can quickly and openly disseminate knowledge and engage with other researchers and the public on their findings. OSTP should direct agencies to establish new and leverage existing funding mechanisms, such as partnerships with philanthropies, research institutions, and scholarly societies to support critical community-driven, open infrastructure—including article, data, software, and code repositories.

Recommendation 3: Engage with the academy on transitioning to open access models that center equity and inclusion

Perpetuating a system that does not welcome voices from historically underrepresented populations has consequences that endanger long-term trust and integrity in science. Incentivising (intentionally or unintentionally) existing power dynamics in the scientific community hinders the sharing of research results and favors exclusivity in science. Current subscription and APC-based pay-to-publish approaches restrict researchers at less-resourced institutions from reading the latest research or publishing their own research because fees to access or publish are too high. More equitable open science practices (including use of open repositories) invite a greater diversity of voices into scientific discussions from the scientists that conduct the research, to the research participants, to the public consumers of research outputs.⁷ We recommend that OSTP support compliance solutions that center equity of participation in scientific communication, and encourage the research community to develop business models that better support open and equitable access to and participation in science.

Recommendation 4: Ensure full reuse through consistent data practices and licensing

Strong reproducibility and replicability practices are key to addressing any improper interference or errors in research and data collection. Underpinning these practices is the ability to access and reuse the data, software, code, and algorithms of a scientific finding so researchers may analyze each other's work to validate the results. Full reuse requires both the article and data to be machine-actionable. This will differ by scientific discipline, but it is imperative that implementation at the agency level incentivize the incorporation of data experts in research design and require an open license for final manuscripts and the underlying data needed to reproduce the results.

Recommendation 5: Re-think research assessment metrics and practices

⁷ <https://journals.sagepub.com/doi/10.1177/1475725719869164>

The current research incentive system is overwhelmingly skewed to reward a single research output: publication of articles in high impact factor journals. Reliance on publications as the sole proxy for quality erodes trust in science and further entrenches inequities in the scientific communication system. We recommend that OSTP support the expansion of incentives to include outputs beyond the journal article (e.g. data, code, preprints, etc.) and encourage the use of more qualitative factors for impact such as influence on policy and practice by inviting supporting language on research output sharing in all Federal calls for applications, grant evaluation guidelines, and grant reporting systems. For example, the Howard Hughes Medical Institute (HHMI) [asks researchers](#) to provide a written impact statement for each journal article they reference and an overview of their training and service activities.

Recommendation 6: Expand the practice of preregistration

The practice of preregistration of a research study design, before data is collected, has been demonstrably successful in the clinical trials environment and should be considered for expansion to other disciplines. The routine practice of preregistration improves research rigor and integrity by including unpublished studies in the scholarly record, and helps to address potential biases that interpretation of data so often has on a study's stated objectives. Similarly, protocol registrations allow researchers to publish research procedures and methods, often in a machine-readable format. We recommend that OSTP consider endorsing practices such as preregistration and protocol registration as critical aspects of research communication, which facilitate reuse and replication of research and reduce inefficiencies in the research ecosystem.

Recommendation 7: Agency workforce development

Training the workforce at federal science agencies on open science practices that improve both rigor and transparency is central to ensuring scientific integrity and specifically, the open sharing of research with the public. For example, many scientists 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 for research sharing. Agency staff and, particularly, grant officers have the potential to standardize when and how scientists share their results from federally-funded research. They can also be helpful in ensuring the results are shared in a way that maximizes reuse to allow for reanalysis and collaboration.

SPARC appreciates the opportunity to provide our feedback on this critical issue, and stands ready to work with the White House Office of Science and Technology Policy in its laudable efforts to improve federal scientific integrity policies.

Please contact Katie Steen (katie@sparcopen.org) with any questions.