August 14, 2018

The Honorable Andrew Wheeler
Administrator
Environmental Protection Agency
1200 Pennsylvania Avenue NW
Washington, DC 20460

RE: COMMENT ON PROPOSED RULE, "STRENGTHENING TRANSPARENCY IN REGULATORY SCIENCE" (EPA-HQ-OA-2018- 0259)

Dear Administrator Wheeler,

The Open Scholarship Initiative (OSI) is an international effort working in partnership with the United Nations Educational, Social and Cultural Organization (UNESCO) to improve the transparency and openness of scholarly communication around the world. OSI is a very diverse group, including over 400 leaders from 24 countries, 250 institutions and 18 scholarly communication stakeholder groups—publishers, universities, researchers, libraries, open science groups and more. As such, we have broad expertise and a wide variety of perspectives on what the future of open science (and more broadly, open scholarship) should look like and how we can most realistically get there from here. Our diversity and expertise are both crucial parts of trying to develop workable solutions to the global future of open and transparent science.

With regard to this particular EPA proposal, I would like to offer you some of the observations that OSI participants have made in the hopes that this information will help better inform your policy making efforts. These are not official positions that everyone in OSI endorses—just some of the observations that have been shared by our participants:

1. The authority to make data public to the extent envisioned in this proposal, may not lie with the EPA: This proposal concludes that the EPA has appropriate sources of statutory authority for this regulation; you have asked for comment on whether additional or alternative sources of authority are appropriate. The concern raised by some in our group is not necessarily whether the EPA's authority is appropriate or complete, but whether it is realistic and sufficient to affect the kind of changes you're looking for. It is understood that your proposal is to modify EPA practices, not science publishing practices, but the impact on science publishing—or at least the collision it sets up with existing publishing practices—will be significant. As you know, most US research funding comes from NIH, NSF, DoD, DOE and NASA. The researchers who receive funding from these agencies adhere to the publishing policies set forth by these agencies—which include open and transparency policies—which in turn work in parallel with the US public access program policies managed by OSTP, and with a wide variety of publishing, open and transparency policies put forth by universities, publishers, private funders, foreign governments, and other groups. It's unclear how the EPA will be able to influence this existing symphony of public and private regulation, at least for the research it doesn't fund. The EPA certainly has a right to try, but

without broad agreement amongst other stakeholder groups, it's unlikely that any one stakeholder in this space will be able to promulgate or enforce regulations that will have such a broad impact.

2. The proposal could impact research practices in unintended ways: Closely related to this first point, some in our group are concerned that what this proposed rule will actually do (albeit unintentionally) is create a new category of "EPA usable research." What does it mean to make research "usable" or "accessible"? In what format? At what stage of research? Will data standardization be required (otherwise, usability can be quite limited)? On what time tables, with what funding and on whose authority? What protections will be applied for the use and reuse of data? And what might be the potential broader effects on research in general if we don't know exactly which forthcoming studies will be potentially of interest to the EPA? Will this cause all science to shift to an EPA-standard of openness? You have requested comment to several of these questions, so it's clear you are aware of the complexity here, but in our experience, the answers are, unfortunately, not simple, not one-size-fits-all, and we dare say not even known at this point. Trying to find these answers is certainly worthwhile, but this inquiry will be a significant and lengthy undertaking.
3. This proposal may be insufficient to address privacy concerns: As you know, there are well-established requirements and procedures for protecting the privacy rights of research participants (45 CFR, Pub. L. 93-348, etc.). These rights are also encoded in the international Helsinki Declaration of the World Medical Association. Coming up with new guidelines for releasing private data, even if the intent is to mask this data and make it available only to government scientists, will require review and approval by the US (the FDA and other agencies) and the global research community. For older studies, improving data access to the degree necessary to pass the openness requirements of this proposed rule could end up being almost impossible given that consent forms written 10-20 years ago (to say nothing of forms written prior to the Internet) were largely silent on the issue of broad sharing of data. Institutional Review Boards cannot release private data without consent, and getting this consent from study participants where the PI has long since retired and participants have either died or moved multiple times will be extremely challenging and costly at best, and effectively impossible at worst, therefore possibly rendering "old" studies unusable.
4. This proposal conflates the need for transparency and openness with misperceptions about "secrecy": The word "secrecy" isn't explicitly mentioned in the proposed rule but it is mentioned in the EPA press release announcing it and also alluded to in various passages throughout the rule. The concern expressed by some in our group is that the dynamics in science referred to as secrecy are in fact misperceptions of how science—as a complex process involving many deliberate steps, many people, many interests, and much discussion—actually works, and that the issues attributed to this secrecy are being conflated with the separate and distinct issues related to openness and transparency. To the latter, the foundations of science have always required and will continue to require vast amounts of openness and transparency. Still, everyone in science acknowledges there is room for improvement, which is why the systems for handling transparency and fact-checking communication (peer review, journal publishing, conferences, replicability, statistical analysis, etc.) are slowly changing for the better. Perhaps where the conflation comes in about secrecy is that, at the study level, there is a tolerance of and even need for certain research practices that are not entirely open (including but not limited to the privacy concerns mentioned above). However, this will continue to be a reality even as science becomes more open—there will be a continued need to preserve the R&D incentives of drug companies by keeping certain data private, for instance, and to recognize that researchers want to avoid being scooped (to preserve their job security and future funding) by releasing their data too soon. Currently, a new generation of digital native researchers is entering science and is torn between being more open and sticking with the communication norms that work best for their career paths. Many tensions exist at many levels, and many difficult challenges are ahead. Overall, the global research community is working hard to make the transformation to more openness and transparency but it's a long way from getting there and we're really only at the starting edge of discussing how some of these changes will be able to

happen (especially with regard to open data). The animating concern in this global push for transparency and openness is to benefit science and society. Along the way, though, we also need to be careful to ensure that the openness and transparency improvements we make end up benefitting research and not harming it.

5. This proposal needs additional safeguards in the final decision process: Some of our participants feel that as written, this rule will allow the EPA to ignore sound science in developing its regulations. An EPA rule encouraging making data publicly available and independently verifiable might be welcome as long as there were assurances that sound science that didn't meet the EPA standard (for a whole variety of reasons) wouldn't be ignored. The proposed rule as written, however, leaves too much discretion in the hands of the administrator with regard to the final ruling. There are probably any number of minor adjustments that could address these concerns—for instance, holding public hearings for each study to be questioned by the EPA might allow scientific findings to be vetted without ruling them out entirely. Without such adjustments, though, this rule as written could potentially have far-reaching consequences for how the EPA develops regulations, and for public health and the environment.
6. An alternative, short-term approach could achieve similar results without affecting science: Perhaps an alternative, short-term approach would be for the EPA to work diligently with investigators to try to get the data they need. This alone could prove challenging and may involve providing funding to try to reconsent study participants from work done 20 years ago, for instance. Any broader and more ambitious open efforts can then be given more time to consider.

I hope this information is helpful and well-received. We stand ready to help you in any way we can.

Sincerely,

Glenn Hampson
Program Director, OSI
www.osiglobal.org

The observations expressed herein are those of the author and do not represent the opinions or policies of OSI participants or their institutions, trustees, officers, or staff.