University of Minnesota Libraries Response to request for comment on EPA-HQ-OA-2018-0259-0025 "Strengthening Transparency in Regulatory Science"

Date: August 28, 2018

Dear Acting Administrator Wheeler:

The University of Minnesota Libraries writes in response to the proposed rule change EPA-HQ-OA-2018-0259-0025 ("Strengthening Transparency in Regulatory Science"), published in the Federal Register on April 30, 2018.

The proposed rule proposes to support increased transparency, further reproducibility, and promote open science. Although the rule states that it "takes into consideration the policies or recommendations of third party organizations who advocated for open science," it does not take into consideration the intent behind calls for openness in scientific research.

We strongly support open science, but recognize that not all research data can and should be made available. The open science community supports the general principle that data should be "as open as possible, as closed as necessary" (European Commission Directorate-General for Research and Innovation. (2016). H2020 Programme: Guidelines on FAIR Data Management in Horizon 2020. Retrieved August 13, 2018 from http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

We are concerned that if the EPA only considers data that is open to the public in their decision-making process for regulations/policies, studies that were conducted with human subjects (due to the informed consent process) would be eliminated from consideration. Epidemiological data are the gold standard that drives and is critical to regulatory decisions, but cannot be made public or shared due to informed consent.

The proposal also allows the EPA to be less conservative with dose response models, opening up the possibility for additional harms to human health and the environment.

Supporters of open science recognize the need for some data to be "closed", particularly those which involve human subjects or personally identifiable information. That data are not made publicly available does not indicate that the data are not scientifically sound. Supporters of open science do not reject peer-reviewed studies simply because the underlying data cannot be made publicly available. As written, the proposed rule would prevent the EPA from using the best available science when drafting regulations. We urge the Acting Administrator to withdraw the rule.

In what follows, we respond to EPA's solicitation for comment on various aspects of the proposed rule.

EPA solicits public comment on:

I. *Which criteria the Agency should use to base any exceptions, including whether case-by-case exceptions may be appropriate.*

A. While we support transparency generally, the broad implementation of a rule enforcing sharing with only case-by-case exceptions is problematic. In order to share data, individual data sets must be considered independently as well as in context of other available data.

1. These caveats are noted within the documents listed in footnote 8. The documents the rule cites do not propose public access to all data, but public access when it is reasonable to do so. This is explicitly stated in EPA's Plan to Increase Access to Results of EPA-Funded Scientific Research, "While the Agency strives to increase access to its research results, it recognizes, consistent with the OSTP Memo, that Federal agencies have a responsibility to protect confidentiality and personal privacy, respect proprietary interests and property rights, and balance between the value of providing long-term access and its associated costs. It is important to recognize that some research data cannot be made fully available to the public but instead may need to be made available in more limited ways, e.g., establishing data use agreements with researchers that respect necessary protections. **Whether research data are fully available to the public or available to researchers through other means does not affect the validity of the scientific conclusions from peer-reviewed research publications.** [emphasis added]" (Environmental Protection Agency. (November 29, 2016). Plan to Increase Access to Results of EPA-Funded Scientific Research. Retrieved from https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf, p. 4)

B. As described later in this response, there would need to be exceptions for any data involving personally identifiable information, including both primary and secondary identifiers, or data where consent for broader sharing was not sought among initial participants at the time of enrollment. This would include the majority of data used by EPA to establish existing regulations.

C. The proposed rule gives the authority to make case-by-case exemptions for "significant regulatory decisions" to the EPA Administrator. Granting this level of authority to a single person does not allow for appropriate oversight. We would request that in the event that exemptions need to be reviewed that they are delegated to a committee of expert scientists who perform similar work.

II. *Whether and to what extent these requirements, or other provisions and policies, should apply to other stages of the rulemaking process, including proposed rules, as well as to other types of agency actions and promulgations, such as guidance.*

A. For all of the reasons outlined in this response, we do not believe the rule should

be applied.

B. The use of valid scientific research data should be fundamental in the decision-making process. However, the analysis and interpretation of research data are a complicated and nuanced undertaking which requires in-depth expertise and a range of perspectives. Data should only be used in rulemaking where it has been subject to rigorous evaluation and interpretation by a range of impartial, qualified experts. This process is already undertaken during the scholarly peer review process.

C. Encouraging openness with data going forward is appropriate. However, it will never be reasonable to expect that all data would be made publicly available, since it could necessarily include private health information about individuals that is protected by the Health Insurance Portability and Accountability Act (45 C.F.R. § 164 (2016).

D. We concur with the Association of American Medical Colleges (AAMC), Association of American Universities (AAU), Association of Public and Land-grant Universities (APLU), and Council on Governmental Relations (COGR) in their July 11, 2018 statement "Re: Docket Number EPA-HQ-OA-2018-0259-0025, Strengthening Transparency in Regulatory Science," which states that "[e]ven the strongest and most sincere supporters[1] of the open science movement have recognized that there is value in research for which underlying data are not made publicly available and acknowledge an imperative to leverage all science to develop policies and regulations." (https://www.aamc.org/download/490086/data/aamceparesponsetonprm7-11-18.pdf)

III. *Whether a narrower scope of coverage would be appropriate, such as only final regulations that are determined to be "major" under the Congressional Review Act, or "economically significant" under EO 12866.*

A. The proposed rule cites Exec. Order No. 13563, 76 FR 3821 (Jan. 21, 2011). ''Our regulatory system must protect public health, welfare, safety, and our environment while promoting economic growth, innovation, competitiveness, and job creation. It must be based on the best available science.'' However, "best available" may not be compatible with a requirement for public access. Applying a rule that prohibits the use of "best available" data simply because the data cannot be made publicly available is inappropriate for any regulation, particularly those that are considered "major" and will be expected to have "major" effects on the environment.

B. In order to rely on the best evidence to develop guidelines and regulations, EPA could conduct systematic reviews to critically evaluate all available published and unpublished research. This is established practice in federal agencies, notably the Agency for Healthcare Research and Quality (AHRQ) with the Department of Health and Human Service and their Evidence-based Practice Centers (EPC)

[1] Ioannidis, J.P.A. (2018). All science should inform policy and regulation. *PLOS Medicine. 15*(5): e1002576. https://doi.org/10.1371/journal.pmed.1002576.

Evidence Based Reports. Such reviews may not involve querying the initial raw data, but instead incorporate critical appraisal of research methods to assess for bias and determine the validity of evidence. To further enhance the EPA's commitment to transparency, the full research methods of any review undertaken to develop guidelines or rules should be registered in advance as a protocol and be open to public comment.

1. We note that the EPA currently conducts systematic reviews, and provide a short list of examples to demonstrate that EPA researchers are familiar with systematic review and recognize its validity and importance:

    a) U.S. EPA Office of Chemical Safety and Pollution Prevention. Application of Systematic Review in TSCA Risk Evaluations. (2018). EPA Document# 740-P1-8001. Available at: https://www.epa.gov/sites/production/files/2018-06/documents/final_application_of_sr_in_tsca_05-31-18.pdf

    b) U.S. EPA. Thayer, Kris. (2017). Systematic Review for Chemical Assessments: Core Elements and Considerations for Rapid Response. Available at: https://www.epa.gov/sites/production/files/2018-04/documents/epa_comptox_cop_v1_nov_16_2017_v4.pdf

    c) U.S. EPA. Systematic Review of Chloroprene. U.S. Environmental Protection Agency, Washington, DC, EPA/635/R-09/010F, 2018. Available at: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=339504

    d) U.S. EPA. Systematic Review Protocol for the IRIS Chloroform Assessment (Inhalation) (Preliminary Assessment Materials). U.S. Environmental Protection Agency, Washington, DC, EPA/635/R-17/486, 2018. Available at: https://cfpub.epa.gov/ncea/iris_drafts/recordisplay.cfm?deid=338653

IV. *Whether certain categories of regulation should be excluded from coverage, such as those that merely reaffirm an existing standard, or some other category.*

A. There is potential for this rule to be used to erase essentially all existing regulations, due to issues in making the data used in their creation publicly available. Because it is unethical to expose participants to likely harms, such as environmental contaminants previously found to cause illness, these data could not be recreated and additional studies on the topic could not be conducted (e.g., it would be unethical to do a randomized control trial to study the effects of consuming lead chips on children).

B. For studies involving human subjects, participants agree to the study based on specific terms for how they will be treated and how their data will be used (informed consent). In order to make data from these studies openly available, all

participants would need to be informed and agree to allow their data to be used in this way. This would be a challenge for many of the longitudinal studies that EPA regulations are based on.

C. Additionally, the informed consent agreements that participants sign indicate the period of time in which their data will be stored. If this period has passed, the researchers were obligated to destroy this data. We reiterate that not having the raw data available does not in any way negate the validity of a study that has undergone peer review.

V. *Whether the provisions of the proposed rule should apply to individual part adjudications, enforcement activities, or permit proceedings when EPA determines that these provisions are practical and appropriate and that the actions are scientifically or technically novel or more likely to have precedent-setting influence on future actions*

A. We ask EPA to define what they would consider "practical." If they have precedent-setting influence, we suggest that studies could highlight their methods, but there cannot be a carte blanche expectation that raw data will be released.

VI. *The definitions of "pivotal regulatory science," and "dose response data and models" and how to implement such definitions.*

A. We are concerned that the rule does not recognize existing definitions for these terms. Until EPA has internally agreed to definitions for these terms, requirements for what data can be used should not be instituted.

B. "Dose response data and models" are well-defined concepts. The dose-response relationship, "the relationship between the degree of response of the biological system and the amount of toxicant administered," is in fact considered to be "the most fundamental and pervasive concept in toxicology" (Casarett, L. J., Doull, J., & Klaassen, C. D. (2008). Casarett and Doull's toxicology: The basic science of poisons. New York: McGraw-Hill.). The controversial aspect of these data and models are the underlying assumptions made, particularly in terms of low-dose toxicity.

1. Historically, the EPA rightly has been conservative in its assumptions, meaning that it errs on the side of protecting human health by assuming that where there is insufficient evidence to rule out harm, one must assume harm. This is particularly true when considering regulations that will have an effect on human health and safety. We believe that the approach EPA has historically taken is appropriate and should be maintained.

2. The Rule notes the "growing empirical evidence of non-linearity in the concentration-response function for specific pollutants and health effects," but does not provide any reference to that evidence.

3. The reference to "considering the breadth of dose response data and models used" would seem to reflect a change to the EPA's traditional

practice of opting for a protective stance.

4. Dose-response relationships are used to predict the likelihood and magnitude of negative health consequences for individuals or broader populations. Conservatism assumes that individuals may be exposed to risk at lower doses. Conservative assumptions are pivotal, since, as the EPA notes in its Exposure Factors Handbook (U.S. EPA. Exposure Factors Handbook 2011 Edition (Final Report). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-09/052F, 2011.), conservative assumptions "provide the basis for measures protective of human health."

VII. *How to incorporate stronger data and model access requirements into the terms and conditions of cooperative agreements and grants.*

A. In response to a memorandum from the White House Office of Science and Technology Policy in 2013 ("Expanding Public Access to the Results of Federally Funded Research"), EPA, along with more than 20 other federal departments and agencies, developed a Plan to Increase Access to Results of EPA-Funded Scientific Research (https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf). Two sections that address current sharing of information are copied below:

1. "EPA has a long history of collaboration in scientific research and is a leader in providing access to environmental information to encourage better decisions and a more informed public. Transparency is a core EPA value. The Agency already makes publicly available much of the Agency's scientific and technical work, including information that supports regulatory decisions. For example, EPA provides all materials and scientific information supporting each regulation in public dockets, which are publicly available for comment at www.regulations.gov. In addition, the Agency maintains an enterprise dataset metadata catalog (the Environmental Dataset Gateway (EDG) at https://edg.epa.gov/metadata/catalog/main/home.page) through which thousands of EPA datasets are publicly available. EPA's Enterprise Information Management Policy (EIMP), adopted March 3, 2015, codifies the Agency's approach to facilitating access to data held in EPA information systems. Appendix D describes, in more detail, a small sample of EPA's extensive ongoing efforts to be transparent and to increase access to environmental information." (page 4)

2. "**While the Agency strives to increase access to its research results, it recognizes, consistent with the OSTP Memo, that Federal agencies have a responsibility to protect confidentiality and personal privacy, respect proprietary interests and property rights, and balance between the value of providing long-term access and its associated costs**. It is important to recognize that some research data cannot be made fully available to the public but instead may need to be made

available in more limited ways, e.g., establishing data use agreements with researchers that respect necessary protections. Whether research data are fully available to the public or available to researchers through other means does not affect the validity of the scientific conclusions from peer-reviewed research publications." [emphasis added]

B. We strongly support the implementation of the OSTP memorandum and EPA's *Plan to Increase Access to Results of EPA-Funded Scientific Research* (https://www.epa.gov/sites/production/files/2016-12/documents/epascientificresearchtransperancyplan.pdf). The language emphasized above reflects EPA's understanding of the need for maintaining balance between the desire to public access to data and the responsibility to protect data providers their subjects. The proposed rule extends the intent for public access beyond reasonable expectations.

VIII. *How it can build upon other federal agencies' policies regarding grantee and cooperator requirements for data access and data sharing.*

A. We are unsure what is meant by this statement. As of January 2017, the EPA, along with more than 20 other federal agencies and departments, created a plan for public access to the results of research they fund (Holdren, JP. (January 9, 2017). Retrieved August 6, 2018 from https://obamawhitehouse.archives.gov/sites/default/files/microsites/public_access-report_to_congress-jan2017-final.pdf). Enforcement of this policy would increase the amount of data that can be made openly available.

B. Neither EPA, nor the federal government as a whole, funds all relevant science for protection of human health and the environment. Data resulting from research funded through sources other than EPA may be "best available" science and per EPA's current directives, should form the basis of new regulations. However, this research would not be subject to federal data access and sharing requirements and it is unreasonable to expect that anyone doing research that may one day be important for regulatory science would anticipate the importance for this future use. These researchers would not be subject to current public access plans and may be unable to prepare their data in a way that it could be made publicly available, whether because their agreements with human subjects do not permit this or because a lack of funding would prevent them from taking the steps necessary to enable the data to be made public.

IX. *Suggestions for a platform that would enable the Agency to implement the provisions of this proposal related to increasing public access to EPA-funded data.*

A. Aggregating sensitive health data onto one platform would unnecessarily increase risks to human subjects. Data breaches at the Office of Personnel Management (OPM) (https://www.opm.gov/cybersecurity/cybersecurity-incidents/) and other commercial organizations (e.g., Equifax, Yahoo) demonstrate that no platform can ever be completely secure.

B. Numerous issues will need to be considered if a platform is developed for providing public access to EPA-funded data, especially if the goal is long-term preservation. Guidelines would need to be established as to how long the data would be held; it may be unreasonable to suggest these data will be available forever. Further, such a repository would need to provide mechanisms for file format migration and would need to ensure that data remain un-corrupted and readable. These are time-intensive and resource-heavy activities and may require dedicated staff to ensure they are completed. Ideally, dedicated data curators would curate the data upon deposit to check that adequate documentation has been provided such that the data can be understood and checksums completed to make sure no data is missing or there are errors.

X. *Methodologies and technologies designed to provide protected access to identifiable and sensitive data, such as individual health data, and on commenters[‘] experience with the use of such methodologies and technologies and their strengths and limitations.*

A. We cannot state strongly enough that there is no such thing as truly anonymized data. “De-identified” and “anonymous” data can be re-identified. This has been clearly shown by researchers at the University of Melbourne, who were able to re-identify individual patients based on de-identified data published by Australia’s federal Department of Health (Culhane, C., Rubinstein, B.I.P., & Teague, V. (2017). Health data in an open world. https://arxiv.org/abs/1712.05627).The Data Privacy Lab at Harvard University lists numerous projects in which they explore the possibility of identifying participants in “anonymized” systems or studies. As an example, Dr. Latanya Sweeney was was able to re-identify 42% of supposedly anonymized participants in the Personal Genome Project (Sweeney, L., Abu, A., & Winn, J. (2013). Identifying Participants in the Personal Genome Project by Name (A Re-identification Experiment). https://arxiv.org/abs/1304.7605)

B. We reiterate that government systems and public sector organizations with highly personal data have been breached and there is no reason to believe any system could be completely secure.

XI. *How to balance appropriate protection for copyrighted or confidential business information, including where protected by law, with requirements for increased transparency of pivotal regulatory science.*

A. Data with intellectual property or confidentiality concerns should not be shared. The sharing of data containing patentable information or trade secrets should not be required due to the potential negative economic implications of such a requirement.

XII. *Whether there are other compelling interests besides privacy, confidentiality, national and homeland security that may require special consideration when data is being released.*

A. The aforementioned intellectual property considerations must also be noted.

B. As research is increasingly the product of international collaboration, differing

regulations and requirements worldwide must also be considered.

C. Where secondary data analysis is being conducted, data use agreements prohibit broader dissemination of data, which would also prohibit data from being released publicly.

XIII. *How the prospective or retrospective application of the provisions for dose response data and models or pivotal regulatory science could inadvertently introduce bias regarding the timeliness and quality of the scientific information available.*

A. Retrospective application and the process for this, is not described within the text of the rule.

B. Retrospective application of the provision would directly introduce bias regarding the timeliness and quality of the scientific information available, as applying the rule would remove most data from consideration.

C. No data sharing rule should ever be applied retroactively. Enforcement of data sharing for previously completed projects would introduce undue financial burden on principal investigators and may be largely impossible. IRB-approved studies include provisions regarding disposal of data following study completion. Researchers abiding by these provisions would have securely deleted and disposed of all raw data in the allotted time following project completion. The retroactive imposition of this rule would invalidate well-conducted research to the detriment of public health.

XIV. *How to address a circumstance in which EPA has a statutory requirement to make a determination for which scientific information publicly available in a manner sufficient for independent validation does not exist.*

A. Scientific papers are part of the scholarly record that have undergone peer review, typically double blind peer review, to ensure their quality. This well-established practice is the standard for ensuring unbiased, rigorous vetting of scholarly content prior to publication.

XV. *Any additional implementation challenges not discussed in this notice that commenters may be aware of as well as suggestions for addressing them.*

A. The inability to make data secure makes it unreasonable to expect that health data would be made public. No system is ever fully secure.

XVI. *Whether the disclosure requirements applicable to dose response data and models in the proposed rule should be expanded to cover other types of data and information, such as for example economic and environmental impact data and models that are designed to predict the costs, benefits, market impacts and/or environmental effects of specific regulatory interventions on complex economic or environmental systems.*

A. If the economic and environmental impact data and models described here would be developed by EPA scientists, they would seem to be subject to the federal Open Data Policy. In general, economic and environmental impact data do not necessarily include data from individual humans and therefore do not pose the

challenges we have outlined in our above response.

B. However, again, there are issues with mandating sharing when i) the research was conducted without intent to inform a rule; ii) was conducted without a pre-existing mandate to share data; iii) was conducted a long time ago and accessing the data for the research is impossible. There will always be reasons why not all data can be shared and a strict mandate for public access will needlessly remove important research from consideration.