UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C. 20460

OFFICE OF THE ADMINISTRATOR
SCIENCE ADVISORY BOARD

September 30, 2019

EPA-SAB-19-005

The Honorable Andrew R. Wheeler
Administrator
U.S. Environmental Protection Agency
1200 Pennsylvania Avenue, N.W.
Washington, D.C. 20460

Subject: Consultation on Mechanisms for Secure Access to Personally Identifying Information (PII) and Confidential Business Information (CBI) Under the Proposed Rule, *Strengthening Transparency in Regulatory Science*

Dear Administrator Wheeler:

EPA's Science Advisory Board held a public teleconference on August 27, 2019, and conducted a consultation with EPA staff on mechanisms for secure access to personally identifying information (PII) and confidential business information (CBI) under the proposed rule, *Strengthening Transparency in Regulatory Science*.

The Science Advisory Board Staff Office has developed the consultation as a mechanism to provide individual expert comments for the EPA's consideration early in the implementation of a project or action. A consultation is conducted under the normal requirements of the Federal Advisory Committee Act (FACA), as amended (5 U.S.C., App.), which include advance notice of the public meeting in the Federal Register.

No consensus report is provided to the EPA because no consensus advice is given. Individual written comments were requested from all members of the Science Advisory Board. The EPA's charge questions for the consultation are provided in Enclosure A and the individual written comments that were received from EPA Science Advisory Board members are provided in Enclosure B.

We thank the EPA for the opportunity to provide advice on secure access to PII and CBI under the proposed rule.

Sincerely,

/s/

Dr. Michael Honeycutt, Chair
EPA Science Advisory Board

Enclosures

NOTICE

This report has been written as part of the activities of the EPA Science Advisory Board (SAB), a public advisory group providing extramural scientific information and advice to the Administrator and other officials of the Environmental Protection Agency. The SAB is structured to provide balanced, expert assessment of scientific matters related to problems facing the Agency. This report has not been reviewed for approval by the Agency and, hence, the contents of this report do not necessarily represent the views and policies of the Environmental Protection Agency, nor of other agencies in the Executive Branch of the Federal government, nor does mention of trade names of commercial products constitute a recommendation for use. Reports of the SAB are posted on the EPA Web site at http://www.epa.gov/sab.

U.S. Environmental Protection Agency
Science Advisory Board

CHAIR
Dr. Michael Honeycutt, Division Director, Toxicology Division, Texas Commission on Environmental Quality, Austin, TX

MEMBERS
Dr. Rodney Andrews, Director, Center for Applied Energy Research, University of Kentucky, Lexington, KY

Dr. Hugh A. Barton, Independent Consultant, Mystic, CT

Dr. Barbara Beck, Principal, Gradient Corp., Cambridge, MA

Dr. Deborah Hall Bennett, Professor and Interim Chief, Environmental and Occupational Health Division, Department of Public Health Sciences, School of Medicine, University of California, Davis, Davis, CA

Dr. Frederick Bernthal, President Emeritus and Senior Advisor to the Board of Trustees, Universities Research Association, Washington, DC

Dr. Bob Blanz, Chief Technical Officer, Arkansas Department of Environmental Quality, North Little Rock, AR

Dr. Todd Brewer, Senior Manager, Partnership Programs, American Water Works Association, Denver, CO

Dr. Joel G. Burken, Curator's Professor and Chair, Civil, Architectural, and Environmental Engineering, College of Engineering and Computing, Missouri University of Science and Technology, Rolla, MO

Dr. Janice E. Chambers, William L. Giles Distinguished Professor and Director, Center for Environmental Health Sciences, College of Veterinary Medicine, Mississippi State University, Mississippi State, MS

Dr. John R. Christy, Distinguished Professor of Atmospheric Science and Director of Earth System Science Center, University of Alabama in Huntsville, Huntsville, AL

Dr. Samuel Cohen, Professor, Pathology and Microbiology, University of Nebraska Medical Center, Omaha, NE

Dr. Louis Anthony (Tony) Cox, Jr., President, Cox Associates, Denver, CO

Dr. Alison C. Cullen, Associate Dean and Professor, Daniel J. Evans School of Public Policy and Governance, University of Washington, Seattle, WA

Dr. Otto C. Doering III, Professor, Department of Agricultural Economics, Purdue University, W. Lafayette, IN

Dr. Susan P. Felter, Research Fellow, Global Product Stewardship, Procter & Gamble, Mason, OH

Dr. Joseph A. Gardella, SUNY Distinguished Professor and John and Frances Larkin Professor of Chemistry, Department of Chemistry, College of Arts and Sciences, University at Buffalo, Buffalo, NY

Dr. John D. Graham, Dean, School of Public and Environmental Affairs, Indiana University, Bloomington, IN

Dr. John Guckenheimer, Professor Emeritus and Interim Director, Center for Applied Mathematics, Cornell University, Ithaca, NY

Dr. Steven P. Hamburg, Chief Scientist, Environmental Defense Fund, Boston, MA

Dr. Robert E. Mace, The Meadows Center for Water and the Environment, Texas State University, San Marcos, TX

Dr. Clyde F. Martin, Horn Professor of Mathematics, Emeritus, Department of Mathematics and Statistics, Texas Tech University, Crofton, MD

Dr. Sue Marty, Senior Toxicology Leader, Toxicology & Environmental Research, The Dow Chemical Company, Midland, MI

Dr. Kristina D. Mena, Associate Professor, Epidemiology, Human Genetics and Environmental Sciences, School of Public Health, University of Texas Health Science Center at Houston, El Paso, TX

Mr. Robert W. Merritt, Independent Consultant, Houston, TX

Dr. Larry Monroe, Independent Consultant, Braselton, GA

Dr. Thomas F. Parkerton, Senior Environmental Scientist, Toxicology & Environmental Science Division, ExxonMobil Biomedical Science, Spring, TX

Dr. Robert Phalen, Professor, Air Pollution Health Effects Laboratory, Department of Medicine, University of California-Irvine, Irvine, CA

Mr. Richard L. Poirot, Independent Consultant, Burlington, VT

Dr. Kenneth M. Portier, Independent Consultant, Athens, GA

Dr. Robert Puls, Owner/Principal, Robert Puls Environmental Consulting, Bluffton, SC

Dr. Kenneth Ramos, Executive Director, Institute of Biosciences and Technology, Texas A&M University, Houston, TX

Dr. Tara L. Sabo-Attwood, Associate Professor and Chair, Department of Environmental and Global Health, College of Public Health and Health Professionals, University of Florida, Gainesville, FL

Dr. Anne Smith, Managing Director, NERA Economic Consulting, Washington, DC

Dr. Richard Smith, Professor, Department of Statistics and Operations Research, University of North Carolina, Chapel Hill, NC

Dr. Jay Turner, Professor and Vice Dean for Education, Department of Energy, Environmental and Chemical Engineering, McKelvey School of Engineering, Washington University, St. Louis, MO

Dr. Brant Ulsh, Principal Health Physicist, M.H. Chew & Associates, Cincinnati, OH

Dr. Donald van der Vaart, Senior Fellow, John Locke Foundation, Raleigh, NC

Dr. Kimberly White, Senior Director, Chemical Products and Technology Division, American Chemistry Council, Washington, DC

Dr. Mark Wiesner, Professor, Department of Civil and Environmental Engineering, Director, Center for the Environmental Implications of NanoTechnology (CEINT), Pratt School of Engineering, Nicholas School of the Environment, Duke University, Durham, NC

Dr. Peter J. Wilcoxen, Laura J. and L. Douglas Meredith Professor for Teaching Excellence, Director, Center for Environmental Policy and Administration, The Maxwell School, Syracuse University, Syracuse, NY

Dr. Richard A. Williams, Retired, U.S. Food and Drug Administration and the Mercatus Center at George Mason University, McLean, VA

Dr. S. Stanley Young, Chief Executive Officer, CGStat, Raleigh, NC

Dr. Matthew Zwiernik, Professor, Department of Animal Science, Institute for Integrative Toxicology, Michigan State University, East Lansing, MI

SCIENCE ADVISORY BOARD STAFF

Dr. Thomas Armitage, Designated Federal Officer, U.S. Environmental Protection Agency, Washington, DC

Enclosure A

The EPA's Charge Questions

Strengthening Transparency in Regulatory Science Proposed Rule Charge Questions for the SAB

Summary

EPA's proposed rule "Strengthening Transparency in Regulatory Science" states:

> "When promulgating significant regulatory actions, the Agency shall ensure that *dose response data and models* underlying *pivotal regulatory science* are publicly available in a manner sufficient for independent validation."
>
> "Information is considered 'publicly available in a manner sufficient for independent validation' when it includes the 'information necessary for the public to understand, assess, and replicate findings.'"
>
> "Where the Agency is making data or models publicly available, it shall do so in a fashion that is consistent with law, protects privacy, confidentiality, confidential business information, and is sensitive to national and homeland security."

Therefore, EPA seeks consultation with its Science Advisory Board on existing mechanisms for secure access to personally identifying information (PII) and confidential business information (CBI) as discussed in the proposed rule consistent with existing laws and policies that protect PII and CBI.

Charge Questions

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

Background

1. <u>Background on existing mechanisms to protect PII</u>

To date, EPA has not issued guidance for de-identifying PII, and intramural researchers either code such datasets as "non-public" or follow the guidance issued by Health and

Human Services (HHS, see https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html#zip). This guidance provides two ways to de-identify information – the "safe harbor" method and the "expert determination" method. [In the "safe harbor" method, a proscribed list of identifiers of the individual who is part of a study or of relatives, employers, or household member of the individual are removed. In the "expert determination" method protected health information is de-identified by "[a] person with appropriate knowledge of and experience with generally accepted statistical and scientific principles and methods for rendering information not individually identifiable:

> (i) Applying such principles and methods, determines that the risk is very small that the information could be used, alone or in combination with other reasonably available information, by an anticipated recipient to identify an individual who is a subject of the information; and
> (ii) [Documenting] the methods and results of the analysis that justify such determination."]

EPA is currently gathering information for managing public access to human subjects research datasets, including the use of a tiered approach with secure data enclave(s)[1] and comparing the benefits of a centralized or distributed approach to protecting PII.

A tiered approach provides access to research data using different strategies based upon disclosure risk. Access to information and data varies by tier. The greatest amount of information is available when access to data are most restricted. Replicating findings requiring PII information (e.g., residence) may not be possible with unrestricted public access. The amount of information available for analysis is dictated by the tier chosen.

2. Background on existing mechanisms to protect CBI

Regulations at 40 CFR § 2 Subpart B govern the use of confidential business information. These regulations establish basic rules governing business confidentiality claims, the handling by EPA of business information which is or may be entitled to confidential treatment, and determinations by EPA of whether information is entitled to confidential treatment for reasons of business confidentiality. Various statutes under which EPA operates contain special provisions concerning the entitlement to confidential treatment of information gathered under such statutes. The regulations prescribe rules for treatment of certain categories of business information obtained under the various statutory provisions. In the event of a conflict between the provisions of the basic rules and those of a special rule which is applicable to the particular information in question, the provision of the special rule shall govern. See https://www.ecfr.gov/cgi-bin/text-idx?SID=505006343d266e51c03f18fc82f41cc1&mc=true&node=sp40.1.2.b&rgn=div6]

[1] A tiering approach was recommended in the recent update to OMB's Information Quality Bulletin [OMB-19-15, April 24, 2019].

Enclosure B

Individual Comments from Members of the EPA Science Advisory Board on Mechanisms for Secure Access to Personally Identifying Information (PII) and Confidential Business Information (CBI) Under the Proposed Rule, *Strengthening Transparency in Regulatory Science.*

(September - 2019)

Dr. Hugh A. Barton B-3

Dr. Barbara Beck B-5

Dr. Deborah Hall Bennett B-7

Dr. Janice Chambers B-10

Dr. John Christy B-11

Dr. Samuel Cohen B-12

Dr. Louis Anthony (Tony) Cox, Jr. B-13

Dr. Joseph A. Gardella, Jr B-14

Dr. John Graham B-16

Dr. John Guckenheimer B-21

Dr. Steven P. Hamburg B-22

Dr. Sue Marty B-23

Mr. Robert W. Merritt B-24

Dr. Thomas F. Parkerton B-26

Dr. Kenneth M. Portier B-28

Dr. Robert Puls B-30

Dr. Tara Sabo-Attwood B-31

Dr. Richard Smith **B-32**

Dr. Donald van der Vaart **B-34**

Dr. Kimberly White **B-35**

Dr. Mark Wiesner **B-36**

Dr. Richard A. Williams **B-37**

Dr. Hugh A. Barton

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

I cannot comment on the approaches used by other agencies as I have no experience with them.

It is clear from the numerous public comments to the EPA and the SAB that there are major concerns with whether and how to move forward with the Strengthening Transparency in Regulatory Science proposed rule. The ongoing effort by EPA to work with a Federal Agency partner to explore using their data storage site is a good activity that should provide valuable information and perspectives. However, the breadth of studies with different designs, sizes of populations, reliance on geographic or other data that can make blinding challenging or impossible, the historical consent agreements that researchers signed, and many other factors make it impossible for a single approach to address all the needs to insure EPA access to the best possible science. Statements by some that providing access to blinded data can be simply accomplished do not address the full range of studies of interest. The statement from the International Society for Environmental Epidemiology submitted to SAB by Dr. Gamse[1] as well as submitted to EPA (Comments of the International Society for Environmental Epidemiology on EPA's proposed rule on Strengthening Transparency in Regulatory Science (EPA-HQ-OA-2018-0259-0001) is valuable and deserving of specific consideration and responses by the Agency. An important policy such as this deserves more complete and explicit development of the proposed rule and its implementation by the EPA in written documentation available for public comment and should have complete consensus review by the SAB prior to the EPA taking further actions.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

EPA needs to be able to use the best available scientific data and analyses. Ideally, these would be available either publicly or through access methods such as those referred to the previous question. Having said that, epidemiological studies in the past were done with the consent agreements that were used at the time. I am concerned that "Strengthening Transparency" will unreasonably apply a contemporary standard to studies done in the past and inappropriately make important information unavailable to EPA for regulatory purposes. I am aware of cases where pharmaceutical studies have been substantially impacted by the consent agreements used, leading to changes in agreement wording for future studies. There likely needs to be a similar differentiation of historical studies from future studies for EPA purposes as well.

It is unclear to me why toxicity studies in animals should ever be CBI, except to limit use of their data by other companies. It would seem that it should be possible to have the complete study

[1] Written Statement submitted to the SAB by Roy N. Gamse, https://yosemite.epa.gov/sab/sabproduct.nsf//F997FD0F423473118525845F004A451A/$File/Gamse_written+statement_8_22_19.pdf

publicly available (except perhaps proprietary information about a mixture or product as opposed to a single chemical), but require that submissions by other companies must obtain permission from the owners of the data/studies for which they might have to pay if they wish to use that data to support their regulatory needs.

Dr. Barbara Beck

Comments on Charge Question #1 Regarding "Strengthening Transparency Regulatory Science Proposed Rule"
Barbara D. Beck, PhD, DABT, ATS, ERT September 16, 2019

Charge question #1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to personally identifying information (PII) data. Please comment on whether such an approach would be a good model for US EPA to apply.

The information provided by U.S. EPA is insufficient for me to provide a complete answer. Unanswered questions, for which further information from the agency would help answer this charge question, include (but are not limited to):

- There are important challenges in redacting data to truly deidentify PII for purposes of making such data publicly available. This was demonstrated in several publications cited in comments provided by the International Society for Environmental Epidemiology (ISEE). For example, the ISEE comments describe a 2017 study by Sweeney et al. in which, even using data considered to be adequately deidentified, the investigators were able to identity over 25 % of the participants in an environmental health study. How will the agency demonstrate that PII cannot be inadvertently identified? Perhaps some of the examples where subjects were identified could be used as case studies to investigate this issue.

- Limitations on use data imposed under the Health Insurance Portability and Accountability Act (HIPAA) and other entities could affect the availability of personal data. Does this mean that studies where such restrictions exist, the data could not be used by EPA in regulatory decision making?

In addition, by narrowly focusing on the tiered approach in the charge question, other relevant considerations to the use of PII were not presented:

- How would be the agency address data not available because the studies were not conducted in the U.S. or were conducted sufficiently long ago that it is unrealistic to be able to subject the studies to deidentification approaches?

- Who would pay for the re-analysis of the data?

- What alternatives to making data publicly available has the agency considered? An example of an alternative approach is reflected in the HEI analysis of the Harvard Six-City Study about 20 years ago. The use of expert outside scientists allowed for an independent analysis of the study, while protecting PII. Note that this multi-year and costly re-analysis emphasizes the need for EPA to give more consideration to practical questions of implementation.

- EPA needs to be more specific in its objectives in this exercise. Is it to reproduce the results of a study, using the same data, models and assumptions, or is it to more broadly address whether the findings are supported by the available information?

Dr. Deborah Hall Bennett

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

The tiered approach should work very well for some types of studies, for example, studies involving exposures that were measured using biomarkers. In this case, one would want to be careful to limit the available dataset to include only those outcomes, exposures, and covariates that were used in the model to minimize the chance for de-identification.

One complexity may arise for studies with biomarkers of exposure if the study population was recruited from a very small region, such as a study recruiting woman delivering in a particular hospital. The HIPAA website noted in the background materials references provided by the EPA indicates for example, data should only be reported based on the geographical unit defined by the first 3-digits of a zip code. And, further, if the first three digits of a zip –code define a region less than 20,000 people, the data cannot be reported. In the case of, for example, babies born in a particular hospital, birth month would not provide sufficient de-identification, and quite possibly birth year would not provide enough de-identification. Great care would need to be taken with any study using a small recruiting base. It may be that there is no way to de-identify data in some studies, but this should certainly not be cause for not considering well conducted studies as part of those studies included as critical regulatory science.

However, this question becomes more complicated if the exposures are based on the address, such as air pollution or pesticide exposure through databases such as the California Pesticide Use Registry, or similar state registries. The web-site for the HIPAA Guidance document provided in the background section states: "A tiered approach provides access to research data using different strategies based upon disclosure risk. Access to information and data varies by tier. The greatest amount of information is available when access to data are most restricted. Replicating findings requiring PII information (e.g., residence) may not be possible with unrestricted public access. The amount of information available for analysis is dictated by the tier chosen." This statement recognizes that things like residences, and therefore by extension, residence based pollution measures, since those have been shown to be re-identifiable, would not need to be listed, hence protecting privacy.

But then, going back to the proposed transparency rule, it is stated that "Information is considered 'publicly available in a manner sufficient for independent validation' when it includes the 'information necessary for the public to understand, assess, and replicate findings.' " If spatially derived exposure levels are removed, the public can no longer replicate findings. It is not clear how the EPA would handle these apparently contradictory statements. Ideally, well conducted studies should always be considered when making policy decisions, regardless of whether or not it is feasible to de-identify the data and make them publicly available.

It is important to note there is a real concern regarding de-identification of data. The U.S. National Academy of Sciences, in their report *Improving Access to and Confidentiality of*

Research Data: Report of a Workshop, stated "Since unrestricted access can cause harm to individuals and also conflicts directly with respect for individual autonomy, it is not an appropriate policy (NRC 2000)." This guidance should give the EPA pause when considering if it is appropriate to attempt to provide publicly available data, even in a tiered approach.

Recently, a peer reviewed study examined the identifiability of records from an environmental health study in Northern California. Using data meeting HIPAA requirements to be de-identified, researchers were able to correctly identify over 25% of the participants (Sweeney et al. 2017). Another study searched the Lexis-Nexis database for stories that mentioned hospitalization, and used age, race, sex and zip code from anonymized hospital admissions data base to match 43% of the people named in the news stories to their medical records (Sweeney 2015). This study should give the EPA grave concern as to the ability to truly protect privacy in some studies while making the data publicly available. It may be wiser to not require publicly available data as a criteria for including solid, well-conducted studies as part of the requirements to be considered in regulatory decisions.

Consider Medicare information on deaths. If one were to consider exposure information assigned by zip code, the average number of deaths per year in a zip code is only 23. If one knew the age, race, and sex, age at entry into Medicare (people who work after 65 often enter at a later age), year of death, and zip code of those people, it would not be difficult to identify them. This is why the HIPAA guidelines require information given by zip code to only reflect the first three digits of the zip code. But if exposures are assigned by zip code, even if the zip codes are not given, people are grouped into zip-code based groups. However, it is still possible to determine the zip code from the information that might be used in a study. Consider that you knew region of the country, an often used covariate. There are public maps showing how air pollution concentrations vary across the U.S., improving ones' ability to match a zip code to a location, especially if one considers the trends in year to year levels. Consider other potential covariates such as percent of the zip code that is age 65 or older, is black, is Hispanic, or is living below the poverty level, or items such as median house value in the zip code or the annual average temperature in the zip code. These could all be used to identify the zip code, from which then the individuals can be identified. The reality associated with being able to identify study participants in studies with spatially derived exposures, such as air pollution epidemiology studies, should not be a reason not to include them in regulatory decision making.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the *Strengthening Transparency in Regulatory Science* proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

The scientific community has developed several protocols and guidelines specifically designed for reporting study data, such as CONSORT and STROBE. These protocols define

reporting methods aimed at improving the scientific basis of evaluating studies. It is important to note that public access to all study data is not a required element. This would indicate that there are other methods for evaluating if a study is well conducted rather than accessing study data. For science conducted in the future, requiring studies to comply with one of these protocols might be a good alternative to requiring access to the data.

It is important to note that while studies conducted in the future could conform with some sort of protocols on reporting study procedures, this is not feasible considering studies conducted in the past. Historically, the peer review process was utilized to evaluate the quality of a study. Many studies conducted in the past were done when exposures were very different than today, and therefore those studies cannot be replicated. For studies conducted in the past, EPA should rely on the standard of peer review that existed at the time the study was published. It is unrealistic for studies conducted long ago to provide all study data, but that does not make them any less valuable.

Death and birth certificate information is publicly available from organizations such as state departments of health or the National Death Index, and hospital admissions data is available from Medicare. To access this data, researchers sign Data Use Agreements prohibiting them from making public anything other than aggregate data summarizing statistics from large numbers of people.

Other researchers can, and have applied to the same organizations to obtain their own copies of the data, after signing their own Data Use Agreements. This approach could be considered an alternative to publicly available data. Although it is not truly publicly available, other scientists can apply to work with the data and test either the same hypothesis or alternative hypothesis.

References

Sweeney L, J.S. Yoo,, L. Perovich, K.E. Boronow, P. Brown. and J.G. Brody. 2017. Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study. *Technology Science*. 2017082801. August 28, 2017. https://techscience.org/a/2017082801.

Sweeney L. 2015. Only You, Your Doctor, and Many Others May Know. Technology Science. 2015092903. September 29, 2015. https://techscience.org/a/2015092903

National Research Council. 2000. *Improving Access to and Confidentiality of Research Data: Report of a Workshop*. Washington, DC: The National Academies Press. https://doi.org/10.17226/9958.

Dr. Janice Chambers

Charge Questions

1. **Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.**

The tiered approach seems to be a reasonable model for EPA to apply. For one thing, it is already used by at least other federal agencies, so it has a precedent for use by agencies that have concerns about access to PII of study participants. This type of approach could use the most restrictive tier initially for the greatest participant protection if this tier was sufficient to allow verification of the data sets or calculations in question. Only when the accessed information was insufficient for verification would there be a need to release more PII and thereby increase the risk for identification of the participants or their personal information.

However, I remain concerned about access to PII when participants in the study have not agreed to their information being shared with people outside the group of investigators who are conducting the study. I would think that many of the older epidemiology studies would not have considered this option in their Informed Consent Forms—we certainly did not in the ICF's that we developed for our epidemiological studies a few years ago. Certainly the option of PII being released to unknown people in the future would have caused some participants to decline participation. It is now unfair to them to put their PII at risk of disclosure without their knowledge and consent. If the rule only applies to studies going forward, then this possibility could be added to the ICF's, but there is no way to make this possibility retroactive to past studies.

2. **Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.**

There are EPA laws protecting both CBI and PII and certainly there are Non-disclosure Agreements (NDA's) that are used in many situations to protect CBI, so there are mechanisms for this protection that have legal standing. EPA may need to obtain the protected information to make regulatory decisions and EPA provides assurance that its staff will not divulge protected information; this should not change in the future. EPA will need to provide the same assurances of protection for any individuals in the public who are granted permission to access the protected information, and these members of the public need to be fully informed about the legality of maintaining the confidentiality and the legal ramifications if the protection is violated.

Dr. John Christy

I agree with the intention of this EPA transparency rule - the underlying data and models utilized by EPA should be made available, as much as feasibly possible, for replication and analysis. I don't think there would be any disagreement about making the models and non-human data available for independent examination. And, I don't believe that any past study is above the need to be re-examined. However, I have no professional experience at dealing with protecting personal information other than here in academia. In my own field, the application of better statistical techniques and the discovery of data errors have demonstrated flawed results in many past studies ... past studies that at the time were used for policy. Also, the EPA should ensure personally identifiable information remains anonymous to prevent unintended and unforeseen reprisals.

Dr. Samuel Cohen

Dr. Cohen provided the following article addressing secure access to PII and CBI.

Perignon, C. K. Gadouche, C. Hurlin, R. Silberman, and E. Debonnel. 2019. Certify reproducibility with confidential data. *Science* 12, Jul 2019. Vol. 365, Issue 6446, pp. 127-128. DOI: 10.1126/science.aaw2825
https://science.sciencemag.org/content/365/6449/127

Dr. Louis Anthony (Tony) Cox, Jr.

Charge Questions

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

Such a tiered approach for access to data could be a good model if it is implemented well. However, EPA could also make data without PII publicly available with few or no restrictions (similar to what has been done previously with NMMAPS, NHANES and other epidemiological data sets), and then impose a tier system to allow restricted access to PII if needed. For many analyses, non-PII data are fully adequate. The final *analysis data sets* used to generate published estimates of concentration-response (C-R) functions often do not contain PII data, but instead consist of numbers of people receiving different exposures and numbers of mortalities or morbidities occurring over some interval of time, as well as values of other causally relevant covariates (e.g., daily temperature and humidity variables). Such data sets can be shared openly if they do not contain high-resolution geographic coordinates or other variables that pose PII risks.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

EPA can use studies involving CBI and/or PII to make regulatory decisions by creating analysis data sets (as just discussed) from them that do not contain CBI and/or PII data (similar to the microaggregation approach that Dr. Stan Young has discussed), but that provide the information needed to construct C-R functions and support regulatory decisions. In addition, in principle, a validated representation of the joint distribution of variable (e.g., a Bayesian network) can also be used to generate data that is exchangeable with the original data and that does not create CBI or PII risks. Finally, differential policy techniques (used by the Census Bureau since at least 2008), such as privacy-loss budgets, may also be used to share data without disclosing PII.

Dr. Joseph A. Gardella, Jr.

The notion of "increased transparency" for studies, data and methods that are used as the basis for rule making and eventual regulations is desirable, and difficult to argue against. A critical factor involves the ***details*** of what constitutes fair access to confidential data.

I would like to endorse the analysis of the Transparency rule submitted by Professor Richard Smith. I have a few additional comments regarding the breadth of impact of the proposed Transparency Rule to data that have been published and used for past regulatory development. It is valuable to consider the efforts by scientific journal editors and professional societies to establish open data policies. EPA's proposed transparency rule acknowledges the need to respect legal requirements and confidentiality, but I am concerned that that the intent is that studies subject to such restrictions will be excluded entirely, not exempted from the requirements. This will result in the exclusion of scientific studies that are potentially significant for rule making and that contain important scientific results.

This could lead to de facto attempts to engineer particular conclusions by arbitrarily eliminating studies from consideration. Because I cannot support any type of ***confirmation bias*** I cannot support the proposed rule as it stands.

I suggest that a great deal of work is needed to define the process of de-identification of data sets to promote both confidentiality and access. University-sponsored studies under human subjects protection rules are generally approved by a university's Institutional Review Board (IRB), who have to respect federal privacy laws such as HIPAA as well as rules related to the confidentiality of individual participants. Access to such datasets is restricted to approved individuals who are expected to sign a data use agreement (DUA) that typically includes additional restrictions on what can be done with the data. In many cases, these DUAs do not permit individual-level data to be distributed in any form, even with personal identifiers removed. The idea that the transparency problem is resolved by just using de-identified data ignores the practical reality of how these datasets are protected from misuse. Geocoded data will not be easily handled by de-identification so that an independent researcher can reconstruct the results of analysis of geospatial data.

Because of the lack of detail on issues like those above, I do not think the transparency rule as it is currently formulated is workable. There are still positive steps that can be taken to improve transparency.

I agree with a strategy that releases computer code written during the course of a study and full examples of calculations that underlie the transformation from original data (aka "raw data") to data that is used in making conclusions.

The third thing I would like to suggest is this: even where direct access to data cannot be obtained, a detailed description of how the author obtained the data could be required. From my observations of the discussions, it is clear that many agree on the difficulty of reviewing "raw" or original data that was collected more than 10 years ago. That data is likely stored on media that may not be readable with current technology. Should those studies be

jettisoned because they cannot be reviewed at the level of original data? I think that it would be sound to consider that the new rule be promulgated for studies that are recent (within the past five years, for example) and for the future.

Thus, I am not sanguine that the current version of the rule will achieve transparency. A large number of public comments have been submitted in opposition to the rule, many coming from eminent epidemiologists (including the last two chairs of CASAC) who are well aware of the difficulties the rule will create. I think that avoiding the SAB in promulgating this rule was a serious mistake.

In his comments during our last meeting in Washington, Administrator Wheeler suggested that he wanted to restart appropriate engagement with SAB as a step forward to work together. This rule was promulgated in such a way to leave SAB out of the consideration. I believe that a sound rule to improve transparency could be accomplished if EPA would dedicate to following Administrator Wheeler's offer to rebuild the appropriate relationship with this proposed rule.

Dr. John Graham

Thank you for the opportunity to offer advice on the final rule concerning science and transparency in EPA regulation. I offer these comments based on more than thirty years of experience in the related fields of risk analysis, cost-benefit analysis and regulatory analysis. I served as elected President of the Society for Risk Analysis in 1995 and am a recipient of the Society's highest award for lifetime achievement in the field. I also gained practical experience serving as Administrator of the Office of Information and Regulatory Affairs, U.S. Office of Management and Budget (OMB) (2001-2006). Currently I serve as professor of public affairs at the Paul H. O'Neill School of Public and Environmental Affairs, Indiana University.

The views expressed here are strictly my own. I start by making some broader contextual points about the rulemaking and then sharpen my focus to the privacy and confidentiality issues, which are the most difficult ones to resolve in the final rule.

1. I support the objective of the rulemaking, which is to achieve, with only limited exemptions or waivers, public access to the influential scientific and technical information that forms the basis of EPA rulemaking. Such access is equally important, regardless of whether the action is deregulatory or regulatory in nature. I would support a similar transparency requirement for all federal agencies, but I understand that this action covers only EPA. To make the process less burdensome for EPA and the scientific community, the transparency requirements should apply only to influential scientific and technical information as defined in OMB and EPA information-quality guidelines under the Information Quality Act of 2000.

2. Public access to data and model structure facilitates the process of reanalysis of data by independent investigators. Reanalysis is not as important in science as the process of replication (where a study's findings are investigated with a new experiment or a new population of subjects), but reanalysis is quite important to ensure that a study's findings are robust, especially when there is not enough time to undertake replication. For risk analysis, cost-benefit analysis, and regulatory analysis, it is crucial to know how robust scientific findings are to plausible changes in model structure, covariates, and other factors. Indeed, there is a useful interplay between reanalysis and replication, as robustness issues uncovered through reanalysis can help motivate the need for the more intensive and time-consuming process of study replication. In some cases, the processes of reanalysis and replication may be merged, as new investigators combine some of the original information used in one study with new information that is collected and then analyzed together with the original information.

3. The standard processes of scientific peer review, literature reviews, and expert panel review are extremely useful, but they are not a substitute for access to original information because the standard processes do not typically include in-depth reanalysis of original information.

4. In this rulemaking, EPA is the regulated entity, as scientists are impacted only indirectly and only if they want their methods, data and findings to be relied upon in EPA's

rulemaking processes. I expect that scientists will typically seek such reliance, in part because scientists are often rewarded and promoted by their employers on the basis of the fact that their scientific studies have influenced EPA policy. It is therefore appropriate to work through the practical ramifications concerning how scientists would assist EPA in efforts to comply with a public transparency requirement.

5. Once the final EPA transparency rule is operational, I envision a link on EPA's web site – one for each covered rulemaking action – that contains one or more files of original scientific and technical information that was judged by the agency to be influential. Each file would include analysis data files and analytic models, with guidance as to how a qualified third party could process the information and reproduce the specific results that the agency is relying upon. Under the final rule, EPA should not require posting of files of data that are already publicly available (e.g., government-provided mortality or medical-claims/discharge data that are available for research purposes under data-use agreements). EPA should simply provide a relevant link to guidance on how such government data can be obtained, what the terms of use agreements are, and how the government data can be properly linked with the other data (e.g., pollution measurements and covariates) used and made available to EPA by the original investigators.

6. It is more feasible for EPA and scientists to accomplish transparency with regard to studies initiated after the final transparency rule is adopted than for studies that were initiated before – sometimes years or decades before – the rule is adopted. Thus, the final rule should treat separately studies initiated prior to the effective date of the final rule and studies initiated before the rule becomes effective.

7. A case-by-case process of EPA exemption or waiver from some transparency requirements will be necessary, and the opportunities for exemption or waiver will need to be broader for studies initiated prior to the rule's effective date. For example, the investigators of an old but influential study may have passed away and thus the location of the data is unknown or the data have been discarded. Even for studies initiated after the rule is effective, there will likely be a limited need for case-by-case exemptions or waivers (e.g., to address some privacy and/or confidentiality concerns).

8. I advise EPA to establish a standing panel of the SAB to assist the agency in determining whether requests for exemption or waiver from the rule's requirements are appropriate. In the final rule, it will not be feasible to anticipate all of the unusual circumstances that might lead to a valid request for exemption or waiver. Thus, a credible process needs to be established for considering case-by-case requests for exemption or waiver. Note the request to the standing SAB panel will come from the agency itself, after consultation with the relevant authors.

9. When agency staff are working on a rulemaking, and wish to rely on specific pieces of influential scientific and technical information to support the action, they would typically reach out to the relevant author(s) and seek submission of the relevant underlying information, assuming it is not present in the published article or already available on the author's website. In complex cases, the agency may need to establish a consulting or

contractual arrangement to facilitate the transfer of information to EPA in a format that meets the agency's needs for public transparency. This process should be expected to occur for any author, regardless of whether the author is located in the United States or in other countries.

10. The Agency will not typically need the raw data that investigators collected. Raw data is not yet in a format suitable for analysis to support publication or rulemaking action. What the Agency typically needs is the "analysis data set," which is the final data set used by authors to support the published findings of interest to the agency (e.g., findings that are typically presented in a key table, chart or graph). The analysis data set has typically been cleaned (errors removed, personal identifiers removed, category assignments made, and – where appropriate – missing data imputed) and is usually in a form that could be sent, upon request, to a journal editor, a peer reviewer, or a professional colleague after the study has been published. This type of information exchange is commonplace within the scientific community on a day-to-day basis. In fact, many authors of scientific and technical papers are now posting on their web sites supplementary information that supports a paper that has been published or submitted for publication. Some journals expect such transparency as a condition of publication. The supplementary information typically includes the analysis data set and detailed descriptions of analytic models and computer code. It does not typically include raw data. Concerns about violation of privacy and confidentiality will be much less common with analysis data sets than would be the case if raw data were disclosed publicly because analysis data sets do not typically include personal identifiers or confidential business information.

11. There could be unusual cases where the Agency will need access to some raw data during the course of a rulemaking. Those unusual cases need to be addressed case by case, and public access to raw data should not be necessary (see point 16 below).

12. Some segments of the scientific community will resist EPA's efforts to accomplish public access to the data through this rulemaking. This resistance should not be surprising. The scientific community is not accustomed to mission-oriented federal agencies making detailed requests for access to their work products, especially input data, methods and tables that may have been generated years ago when transparency was not considered as crucial as it is today. Working the kinks out of the process will require that the Agency take some time, meet with scientific organizations, engage in dialogue about key concerns, and do some consensus building. It is well known that many of the key issues that need to be resolved were not resolved (or even addressed) in the proposed rule. In order to ensure that the final rule has sufficient acceptance within the scientific community, it may be wise for the agency to re-propose the revised rule for public comment, especially the provisions related to privacy and confidentiality.

13. Some scientists resist public access to their data, methods, and work products because they fear that impacted interest groups will hire consultants to reanalyze their data and discredit, confuse, or criticize their published findings. While I understand this perspective, I believe the public at large (including affected interest groups) are entitled to access the information used by government to support the regulations (or deregulatory

actions) that impact them. It is quite possible that hired consultants will find important errors or robustness problems in a study, even one that has been authored by investigators at prestigious universities and published in a well-respected, peer-reviewed academic journal. EPA and its advisory bodies cannot be expected to discover all of these problems on their own. Public participation in the rulemaking process can help but only if the public has access to the original information.

14. Some analysis data sets used in environmental and occupational epidemiology contain detailed information about each individual subject. Even without access to personal identifiers (e.g., name, address, social security number), a creative analyst can ascertain the identity of some subjects with a high probability. For example, there may be very few people in a specific community of a particular gender, race, age, income profile, smoking status, date of death, pollution-exposure profile, clinical diagnosis, and so forth. Under these conditions, public access to the analysis data set could inadvertently compromise the privacy of some subjects and/or the terms for human-subjects agreements that were signed when the study was initiated.

15. There are creative ways to avoid inadvertent violations of privacy and confidentiality that could occur as a result of public access to analysis data sets. For example, the hazard function in a cohort study can be analyzed at the individual level using weekly, monthly, or quarterly mortality data rather than daily mortality data; the wider temporal interval reduces significantly the opportunity to use time of death as a subject identifier. Wider temporal intervals may induce some imprecision in estimates but do not necessarily introduce statistical bias. In cases where multiple cities are studied, it may be necessary to exclude some small towns or neighborhoods from the data set that is made publicly available by EPA. If there are robustness problems with the results for medium- and large-sized cities, then EPA may wish to pursue non-public approaches to reanalyzing the data for small towns and neighborhoods (see point 16 below). The creative solutions will often reduce the amount of data available for reanalysis but the solutions may retain sufficient data to facilitate a meaningful and insightful process of reanalysis to take place. Crafting creative solutions will require cooperation between the original authors, EPA staff, and a standing SAB committee that possesses appropriate expertise in the prevention of inadvertent violations of privacy and confidentiality.

16. Exemptions or waivers should not be granted by EPA unless the standing SAB panel concurs with the agency that there are no creative solutions that offer adequate protection of privacy and confidentiality. If situations arise where the privacy or confidentiality concern is valid and no creative solution can be found, the rule should outline an agency-sponsored procedure for undertaking reanalysis of data through a qualified third party that will respect privacy and confidentiality. In the past, EPA has worked with independent scientific research organizations, such as the Health Effects Institute (HEI), to perform a confidential reanalysis, including verification of raw data, analysis of additional covariates, incorporation of supplemental data, and robustness checks concerning plausible changes in specification. For more information on the origins of HEI, see Chapter 4 (author Thomas Grumbly) of my 1991 book Harnessing Science for Environmental Regulation, Praeger, Westport, Connecticut, 39-62. In the case of HEI's

extensive reanalysis of the pollution-mortality relationships in the American Cancer Society cohort, HEI pursued a confidential reanalysis that was independent of the original authors (who cooperated with the HEI effort). HEI did not find major analytical errors in mortality results but did find robustness issues concerning the inclusion of additional pollution variables and the interaction of level of education with pollution measurements. Thus, the HEI experience with reanalysis underscores the value of the science and transparency rule and provides a pathway for reanalysis when public access cannot be accomplished due to privacy and confidentiality concerns.

Thank you again for the opportunity to offer this advice. I stand ready to provide additional information as appropriate.

Dr. John Guckenheimer

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

Response: Many Federal agencies need to protect confidentiality of PII data. I support efforts to develop common approaches and policies across the Federal government for this purpose. Since confidential personal information is more central to the mission of some other agencies like NIH and HHS, EPA can benefit from following their leadership in this area. In particular, a tiered approach to PII accompanied by penalties for revealing confidential data seems like a good choice for the EPA. Adopting the procedures used by the National Center for Health Statistics would simplify matters for everyone with an interest in this issue.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

Response: EPA regulations should mandate that data about dose responses, toxicity, releases of chemicals into the environment, etc. that underlie its regulations should be deposited in databases maintained by the EPA, alone or in partnership with other Federal agencies. Archiving the data will enable the EPA to reuse it when there are future challenges to its regulations. In the case of CBI and PII, public access to these data should be limited to those parties tasked with peer review, validation or approved proposals for further research with the data. Furthermore, access should occur only to those who agree to maintain the confidentiality of the data.

From a long term perspective, EPA seeks to establish standards for experimental methodology and scientific models that connect the data to regulations. The public interest can be protected by policies that ensure that reasonable questions about data validity and analysis will be investigated by independent experts, similar to scientific peer review. In particular, review of requests for access should prevent nefarious uses of the data insofar as possible. A key aspect of making this process work is that EPA maintain a staff of experts who will manage access to confidential information.

Dr. Steven P. Hamburg

Analysis of the implications of the proposed Science and Transparency rule with respect to PII and CBI is lacking and beyond the scope of this consultation and as such it is impossible to understand the implications of the proposed rule on both protected information and the expectation of privacy by both individuals as well as with respect to commercially sensitive information. The lack of analysis is greatly compounded by the lack of details about the proposed rule itself.

Without specific definitions clarifying a host of the underlying ambiguities including 'publicly available', 'independent validation', 'models' or 'data' it is anyone's guess as to what the implications of the rule would be on PII and CBI. Similarly the lack of contrast to extensive systems already in place with respect to protecting PII and CBI within the federal government and other public agencies eliminates a crucial opportunity to effectively evaluate the strengths and weaknesses of the proposed rule. Given that no information addressing these issues was provided to the SAB by the EPA, despite our requests, one can only assume that the EPA has not yet thought through the design of the system to protect PII or CBI. This in turn means the rule is not complete and requires further work and clarification before it should be promulgated.

Overall the impacts of the rule could be positive or perverse on EPA rule making, but the lack of details regarding the rule makes discerning between these two outcomes impossible beyond providing ungrounded speculation. Given the lack of details about the rule and the absence of a detailed analysis carried out by EPA that helps define the implications of the proposed rule on privacy issues specifically and the quality of any rule making generally there is no way to address the charge questions beyond doing the analytical work required to define the options and the implications. If the EPA wanted the SAB to provide substantive input to help address the current short comings then EPA should have requested a full review with an expert panel and not a consultation. One can only speculate that the Administrator and his representatives deliberately did not complete the required analysis of the implications or request substantive input from the SAB to avoid providing clarity about the rule and its implications.

Dr. Sue Marty

I agree with the general conversation of the SAB that a tiered approach is a good model for EPA to apply. Also, that EPA should be able to use studies with PII and/or CBI for decision making, but that data protections may limit access to the data in some instances (i.e., where blinding of key data elements is not possible). There were several proposals introduced during the discussion. It would be ideal to have some level of access with increasing stringency applied as needed – e.g., blinded data that the public can access or access granted to limited requestors by special permission from EPA or when this is not possible, a small independent group of experts that could review the data and compile a summary report. In the latter case, a balanced group of experts would be key. Again, there are models (HHS, CDC) that EPA may wish to consider for this.

Mr. Robert W. Merritt

For my comments, I have chosen to highlight a specific item from the May 12th, 2018 SAB Workgroup memorandum[1], which I have attached in full with this email. Reviewing this document, I find no reason to modify its conclusions, and would request that all of its elements be represented and amplified in any comments passed forwarded to the EPA on this issue. In addition, I have drawn on my experience related to public data repositories to argue that the proposed rule should only apply to future studies considered for addition to the regulatory process.

From the referenced memorandum, "the proposed rule fails to mention that EPA has mechanisms for vetting science through several expert panels, including the EPA Science Advisory Board, the EPA Clean Air Scientific Advisory Committee, and the EPA FIFRA Scientific Advisory Panel." The Agency has gone to considerable (and clearly stated) effort since 2017 to broaden state and industrial representation on these panels. The motivation behind this rule illustrates a clear (if unspoken) concern about anti-industry bias in study selection related to regulations. If so, the Agency should look to the input of these newly reconstituted expert panels to address the issue before proposing a rule that smacks of overkill.

In addition to this item in the memorandum, I would include a comment as to the enormous cost and extended timeline implied by this rule change, an issue I feel has not been fully addressed, and which argues for limiting this rules applicability to future studies. I am an expert in database design and conversion with four decades of experience in working with state and federal databases, during their transition from internal to public-facing, particularly as a result of the Paperwork Reduction Act of 1980 and as amended in 1995. I can attest that estimation of the cost and completion of these efforts were generally off by a factor of five, at a minimum, and some were never entirely completed. This was primarily for two reasons:

1. The original data repositories for these sources were influenced by the efficiencies and limitations of the idiosyncrasies of the original storage mode (paper files, computer punch cards, ASCII files on magnetic tapes or other media, digital spreadsheets or legacy database systems) and did not envision, or provide for, the need for transfer to a future storage technology. Inevitably there will be some data loss (due either to actual loss of the raw data due to media failure) or an insupportable costs for transferring materials to a new storage mode I can think of no case where these limitations prompted an agency to force a *de novo* repetition or replacement of a study.

2. A lack of funding from the government agency. Rules such as the one proposed are usually a kind of internally directed "unfunded mandate", where the Agency management listens to their staff when they hear that conversion to a public-facing data repository is technically possible, but pay less attention to the cost portion of the discussion. In my experience, federal and state

[1] Preparations for Chartered Science Advisory Board (SAB) Discussions of Proposed Rule: Strengthening Transparency in Regulatory Science RIN (2080-AA14)
[Available at:
https://yosemite.epa.gov/sab/sabproduct.nsf/E21FFAE956B548258525828C00808BB7/$File/WkGrp_memo_2080-AA14_final_05132018.pdf]

employees have done excellent work in carrying out directives of this type, but they cannot produce man-hours out of thin air, particularly in times of staff reduction and budget cuts. This leads to long delays and the need to limit or stage the scope of the repository conversion, with some projects failing to reach completion before agency priorities change and they face a cancellation of funding.

A further significant consideration is the financial burden on the research agencies. This rule would ask universities and institutes (some outside the U.S.) to find new funding for finished projects, long after the grants or institutional support has expired. This is a classic example of an unfunded mandate, where the burden is placed on a (usually) non-profit organization to provide funding so that the EPA can continue to meet their regulatory responsibilities. Even if the EPA provides a simple and efficient method for embedding raw study data into a public-facing repository, a significant burden will be placed on the research organizations to bring their data up to this threshold due to issues of media recovery and conversion, reformatting, quality control and a need to provide continuous support for the data, once embedded (as surely the repository will require the submitting organization to provide a contact to answer questions about the data).

Absent some (as yet unmentioned) commitment by the EPA to fund such costs, it is ridiculous to assume that most of these organizations will willingly shoulder this significant financial burden. Even if they do, they will have to shift funding from ongoing, vital new research to fund this activity, at the net negative cost to the nation's health. A cynical observer might suggest that, absent EPA funding support for participants, only organizations producing results favorable to deep-pocketed industrial concerns will easily find funding to participate, which could be an underlying *raison d'etre* for the rule.

As you know, I have no expertise in the areas of epidemiology or toxicology, but for the reasons stated above, I would argue that if such a rule as this is required, it should only apply to new studies being adopted for regulatory use, where proper expert panel review, use of a data storage design and method that envisions transfer of raw data into a public repository, and inclusion in research agency funding models, can all properly occur without impeding the research and regulatory function.

Dr. Thomas F. Parkerton

As a preface to my response to the specific charge questions posed on the consultation to the proposed rule Strengthening Transparency in Regulatory Science, I would like to provide some general remarks. As explained in the Chartered SAB's April 25, 2019 memorandum on EPA's Planned Agency Actions and their Supporting Science in the Spring 2018 Regulatory Agenda[1], a consensus view was reached by the SAB in May 2018 concluding that this proposed rule warranted a full SAB review. This position was reiterated in a June 19, 2018 letter[2] from SAB to the Administrator highlighting both the broad implications of the proposed rule on EPA's foundational policies related to the use of science in rulemaking and policy development and the essential need for a formal, deliberate review by relevant experts to logically inform a final regulation. The SAB's views were reinforced by a number of the public comments received on this proposed rule at the last SAB meeting held on August 27, 2019. The present consultation process is both limited in scope and in the relevant expertise that is available via the Charted SAB members for addressing the broad science challenges posed. Therefore, I would urge that the Administrator give serious consideration to a more transparent evaluation of the science challenges that underpin the proposed rule by either constituting a dedicated SAB panel or supporting a focused National Academy of Sciences study. The output obtained from either of these options would in my view provide not only critical input prior to final rule-making but also promote public confidence in the regulatory outcome.

Charge Question 1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply

RESPONSE: A tiered approach to handling data with PII seems like a promising strategy. However, EPA should conduct a more detailed evaluation of different de-identification protocols in use in other agencies and in the private sector and determine if such approaches could be practically implemented in supporting this rule. If de-identification is determined to be inadequate for managing disclosure risk for a given pivotal dataset, an interagency government or third-party data archive could be created to collect and manage data specifically used to support "significant regulatory actions." Researchers could access these data for independent evaluation after preparing a research plan and agreeing to confidentiality agreements subject to penalties for violating the agreement.

Charge Question 2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

---

1 https://yosemite.epa.gov/sab/sabproduct.nsf/91C3DBA65025B525852583FD00530ED4/$File/Work+Group+Memo+Spring+2018+Reg+Review.pdf

2 https://yosemite.epa.gov/sab/sabproduct.nsf/LookupWebProjectsCurrentBOARD/4ECB44CA28936083852582BB004ADE54/$File/EPA-SAB-18-003+Unsigned.pdf

RESPONSE: As an initial step, it may be helpful for EPA to better understand the nature and magnitude of past and expected future PII and CBI challenges. Historically, PM2.5 and ozone epidemiology studies have served as the pivotal regulatory science studies that have driven "significant regulatory actions" based on my understanding. Presumably such studies will continue to be a logical future focus for increasing data transparency. However, limiting application of data access provisions to "significant regulatory actions" may effectively reduce the relevance of CBI related data since most specific chemical rules would unlikely meet this definition.

To gain further insights, EPA could review the major rules over last two decades to determine just how many such rules rely on pivotal studies with non-accessible PPI/CBI data. EPA could also develop a list of key rulemakings expected over the next few years that are likely to meet the "significant regulatory action" definition and rely on regulatory science studies involving PPI/CBI data. This will allow a more realistic assessment of the challenges ahead so that EPA can better focus resources on solutions for the types of studies/data sets that are likely to be of greatest relevance and of highest priority for providing timely public access before major rule-making. Based on this analysis, EPA could engage with investigators of key pivotal studies in upcoming major rule makings and seek their cooperation in de-identifying and/or data sharing using a tiered protocol as highlighted in charge question 1.

EPA may also want to consider developing a data archive system that allows access to pivotal regulatory science studies. For studies involving PPI that cannot be de-identified, researchers should be allowed to request access providing they adhere to a research plan and agree to confidentiality agreements subject to penalties for disclosure. In principle, the data archive provided should not only include the raw data used in analysis but also the research plan, model documentation and results including sensitivity analyses of assumptions/variables. To be practically useful, data supporting significant regulatory actions should be made well in advance (e.g. at least one year) of rule-making.

Looking forward, EPA should include data-sharing provisions in EPA grants to ensure grant recipients make provisions and include budget for sharing data upon study completion. NIH's threshold for data sharing protocol is triggered when an applicant receives $500,000 or more in any one year. EPA could consider establishing an analogous funding threshold.

In the case of past studies in which investigators are unwilling or unable to share data, EPA should consider a preference for relying on reliable studies with available data (if such studies are available) in making significant regulatory decisions.

Dr. Kenneth M. Portier

For question 1. A key difference between EPA's data needs and the typical NIH, HHS data is the utility of geospatial and temporal items in the more useful human data. An EPA tiered approach for access to PII data must address the presence of spatial/temporal information. But spatial and temporal data elements in human data are two items that HIPAA calls out as identifiable elements that need to be stripped out or "grouped/categorized" before data can be considered as de-identified. This will make the creation of truly public data set of this type that ALSO satisfy the informed public's need to validate previously published results very difficult if not impossible. These kinds of data will of necessity be required to be only available via the safe harbor mechanisms as used by NCHS.

For question 2. One issue not discussed is i) WILL owners of independent human data be willing to release some subset of their data to public repositories, even such safe harbor data archives as used by NCHS. Related to this is ii) CAN owners of independent human data legally release these data to public repositories, given IRB and HIPAA requirements for informed consent. For example, a prospective epidemiology study may not have received participant informed consent to release their data in this way. Many key epi studies, such as the ACS second cancer cohort, are not able to go back and get this kind of informed consent since a large fraction of their cohort have either died are no longer are able to provide informed consent due to age-related illness.

While spatial or temporal aggregation may be possible to allow these data to be made acceptably public, who decides which aggregation is used? One aggregation may be best for one type of study and another aggregation will be best for another study. Does this require the original data owner to continue making these aggregations? How many is reasonable? For how many years past the end of the study? The only reasonable solution is to have the highest level resolution data available behind some firewall and allow access only under the most controlled conditions (e.g. formally submitted analysis proposals, peer review, independent analysts running the analysis for the requestor.)

Forcing Agency decisions to be made only on analyses for which supporting data can be made "public," even if such data are within a safe haven archive, could produce biased decisions.

Note that it is not clear that EPA rule making would be considered a public health issue under current HIPAA regulations/interpretation meaning that invoking the public health need may not be a mechanism whereby EPA makes such data public.

For environmental epidemiological studies (or others) that combine environmental data with human health data (e.g., Medicare data), there is the concern that the combined data set that is used in the analysis would need to be made available for others to validate analysis findings. This brings up the issue of whether the computer code used to create the combined dataset needs to be made publicly available. Actually, in some sense, if the original data have already been deemed as "quality", the code used to create the combined dataset is more useful, especially if the public is interested in assessing the code for potential biases that entered at the time of creation. In addition, the code used to analyze these data would also be useful to be made public if the public would be interested in whether biases occurred at the time of analysis.

Regarding CBI, most CBI seems to address aspects of chemical, substance, or product chemical characteristic data and in (copyrighted, non-standard) protocols for generating animal dose response data. There are work around methods (such as consent decrees) for EPA to allow the critical data/information to be released AND to allow enough information on methodology to be made public so as to provide an assessment on the quality of the resulting data. There is often an ongoing need for the original raw data, if available, to be used in assessing the quality of the final data used in analyses which in turn informs the value to be placed on these data in a weight of evidence assessment. This suggests that there may need to be a tiered data access approach to even these CBI protected data.

Dr. Robert Puls

Robert Puls, Response to SAB Charge Questions on CBI and PII

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

I cannot comment on the approaches used by other agencies as I have no experience with them.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

I believe that EPA already has in place rules and regulations which govern the protection of CBI and PII. When I was leading the EPA-ORD Study on Hydraulic Fracturing from 2010-2012, we had rules in place to protect oil and gas industry data collected for the study as well as CBI data related to private wells used in the field portion of the study. These rules worked well and both the public and oil and gas industry were satisfied with the implementation and use of these rules and practices.

Dr. Tara Sabo-Attwood

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

My answer to both charge questions is that I do not feel that I have enough information to be able to formulate a specific response. The consultation asked of the SAB and charge questions seem to be a situation of 'putting the cart before the horse' which became very clear during the public meeting held on August 27, 2019. In addition, I documented several terms that were not well defined yet critical to addressing the issues at hand such as 'validation' and 'replication.' The purpose of a 'consultation' (non-consensus process) with the SAB with limited to no information available from a working group is still not clear, nor is it clear how the responses to these charge questions will be utilized.

This is an important issue with potentially significant consequences regarding how we currently use and protect data in a way that is considered ethical. It is also imperative that we do not move in a direction that eliminates the best research/science to support regulation nor devalues the IRB process. Therefore it is imperative that we get this right. If the SAB could have a more thorough report as is typically done, it would likely lead to answers to these questions that are grounded in science and a foundational information base.

Dr. Richard Smith

Response to EPA Charge Questions on the Transparency Rule

1. *Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.*

Before I give a detailed response to this, I think we need to distinguish among different types of PII data. (Please note that I am answering this question mainly from the perspective of epidemiological studies involving human health outcomes; different considerations ay apply in other types of investigations, such as animal toxicity studies.) Datasets held by federal agencies are, in principle, public datasets, though the actual rules of access may differ substantially from one source to another. Many epidemiological studies have been based on daily death counts, which are maintained by the National Center for Health Statistics (NCHS), part of CDC. Individual-level data are accessible through one of the FSRDCs. As they are described in Wikipedia, "Federal Statistical Research Data Centers are partnerships between U.S. federal government statistical agencies and leading research institutions to provide secure facilities located throughout the United States that provide access to restricted-use microdata for statistical purposes to authorized individuals. There are 29 FSRDCs across the country, primarily located at academic institutions and federal reserve banks." The individual researcher must visit one of these centers in person and perform their analysis under the supervision of a FSRDC employee. The researcher is allowed to take away output data that meets certain requirements as judged by the supervisor. Essentially, aggregate datasets (or tables of results of statistical analyses) are allowed, but any kind of output table that could allow identification of individuals is not. A different method is in place for data from the Center for Medicare and Medicaid Services (CMS) that is the other major source of death data used in epidemiology research. The sponsoring organization must negotiate a contract with CMS that will provide for the data to be maintained on a secure server; again, there is a data supervisor who is responsible for monitoring any output, but the researcher does not have to be present in person at the data center. I have personally used this system and found it perfectly manageable; but the contract with CMS took some months to set up and (as I understand) a considerable sum of money. If this is what EPA means by a tiered approach (and if it's understood that research using such methods would count as public data in EPA's transparency rule) then I think this could be a viable approach; my main caveat is that it only applies to PII data that are held by federal agencies and that's only a small fraction of published epidemiology research. In particular, datasets held by universities or by private organizations such as the American Cancer Society would not seem to be covered under these procedures.

So in summary, if by a "tiered approach" EPA means using the FSRDCs to allow access by researchers to individual-level datasets held by federal agencies, I believe that could be a viable approach. I think it would broadly be helpful if CMS data could be handled in this way as well, if that would avoid the need for the contracting organization to go through a lengthy approval process. It does not solve the problem of PII data held by non-federal entities.

2. *Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule,*

please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

I find this question impossible to answer because it is asking the Science Advisory Board to resolve two incompatible objectives: allowing access to CBI and PII without violating the laws that protect such data. The only answer I can provide is that EPA needs to fundamentally rethink the objectives of the transparency rule.

To address the PII question first, many published papers on environmental epidemiology rely on data collected by universities and other private institutions that are protected by legally binding data use agreements (DUAs). To exclude all such studies from the regulatory process would make a mockery of EPA's commitment to use best scientific evidence in formulating regulations. Making data available in partially redacted form (in effect, removing all personal identifiers) could go a long way towards resolving this problem, but there are difficulties that I don't think have been properly thought through. One question, for example, is what level of data aggregation would be needed to fulfil the requirements of EPA's transparency rule – would it be necessary to provide original raw data down to the level of individual observations or would it suffice to provide an "analysis dataset" that was sufficiently detailed to allow an independent researcher to verify the results of a statistical analysis, but without providing information needed to verify every single data value in such a dataset? The current wording of the rule and EPA's response to SAB's question on this specific point (7/25/2019) seem to leave the issue as ambiguous. I believe that, if it could be clarified that the analysis dataset was what is required, that would help resolve the issue, but not the more fundamental one that even an analysis dataset cannot be publicly released under most DUAs that I am familiar with. So while it sounds attractive to say that data should be released with personal identifiers removed, this would still be subject to the agreement of the universities or other entities that ultimately own the data, and I cannot see that being readily agreed.

A further point here is that even de-identified data might not be sufficient to confirm the analysis. For example, a quite common practice in environmental epidemiology is geocoding: individual participants' addresses are geocoded into latitude-longitude coordinates, which are then integrated with spatially distributed air pollution data (from monitors, air quality models, or satellites) to derive individual estimates of ambient air pollution exposure. However, geocoding is reversible: publishing individual latitude and longitude information would in most cases be sufficient to allow identification of a participant's address, if not necessarily down to the level of an individual address or apartment number, at least with enough accuracy that when combined with additional information in the public domain, there is the possibility that an individual participant could be identified. For that reason, major datasets that include individual participant addresses (for example, the Women's Health Initiative) maintain strict rules on the confidentiality of that information. I don't see how this is going to be reversed.

Since I work for a university and not a private company, I have much less to say about CBI. I don't know what rules have applied in the past, or how or why EPA is proposing to change them. However, I do not believe EPA should treat CBI data in a way fundamentally different from PII data.

Dr. Donald van der Vaart

Strengthening Transparency in Regulatory Science Proposed Rule
Charge Questions for the SAB

My responses are as follows:

Charge Questions

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

The benefits of providing wider access to research for significant rule-making far outweigh the minor risk that a properly implemented protocol may represent. A tiered approach is one reasonable mechanism to provide this critical opportunity.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

As stated above, a thoughtful protocol developed with these laws in mind must be developed if we are to develop rules that potentially harm our country's standard of living. An intermediate or first step may be to limit re-review to an independent stakeholder group.

Dr. Kimberly White

Responses to the Questions

1. Other agencies (e.g., NIH and HHS) use a tiered approach for access to PII data. Please comment on whether such an approach would be a good model for EPA to apply.

Response: Implementing a tiered approach as utilized currently by other federal agencies is a reasonable path-forward for EPA. The National Institutes of Health has extensive experience associated with the identification and handling of sensitive information. They have established guidance and policies to facilitate the process which can be evaluated and implemented, as relevant and appropriate by the EPA.

2. Given the laws protecting CBI and PII, as well as the proposed requirements for data availability in the Strengthening Transparency in Regulatory Science proposed rule, please comment on how EPA could use studies involving CBI and/or PII to make regulatory decisions.

Response: Studies that may include confidential business information (CBI) and/or personally identifiable information (PII) can provide important information about exposure and human health impacts which may be critical for the regulatory process. In compliance with all appropriate laws protecting sensitive information, CBI and PII, EPA should utilize all relevant information at its disposable to make science-based regulatory decisions. Specifically, EPA should develop clear guidance and policies for its use of studies containing this type of information and how, when necessary and feasible, it will seek to make information available.

Dr. Mark Wiesner

To directly reply to the charge questions out before us could be construed as the SAB having been consulted and then providing input on the proposed Science and Transparency rule, when in fact we have not been consulted on the core matters surrounding this proposed rule. Moreover, the process we are being asked to participate in represents a departure from the traditional practice of producing a consensus SAB statement. The proposed rule would limit the scientific information that EPA would be allowed to take into consideration in the regulatory process with potentially harmful implications for public health.

Dr. Richard A. Williams

Comment on : "Strengthening Transparency in Regulatory Science"
Tuesday, September 3, 2019

From the Comments I have heard and read, there appear to be two main issues: the first is whether there is a problem with the use of science (particularly epidemiology) in EPA's regulatory rule making and second is and, if there is a problem, can epidemiological data be made publicly available to allow studies to be replicated without inadvertently disclosing the identities of the participants?

There is a Problem

There have long been charges leveled against EPA's use of science. The most prominent one is interference by politicians in EPA's science. This may take the form of EPA managers (or political appointees) using science inappropriately to cover up decisions,[1] ignoring the science when making decisions[2] or changing it to suit their purposes (like turf building).[3]

Rarely leveled, the opposite is true as well: EPA staff scientists may inappropriately try to drive decisions by slanting the science. The most obvious way of doing so is using reference dose (RfD).[4] Derived from FDA's Acceptable Daily Intake (ADI), the RfD is a safety analysis where risk assessors become, for all practical purposes, risk managers by establishing a level of exposure that is "safe." In fact, research has shown that there are multiple ways to establish an RFD for the same compound.[5] In addition, there is no guarantee of safety above or below any RfD.[6]

Similarly, staff risk assessors may include assumptions that drive risk estimates one way or another and those assumptions may be in line with their preferred decision. Finally, staff scientists may choose like-minded people to do funded research confident that they will arrive at the correct outcome.

These issues play out in different administrations, particularly for science with large measures of uncertainty, which characterize most environmental issues. Thus, it is difficult to separate

[1] Wagner, Wendy E., The Science Charade in Toxic Risk Regulation, Columbia Law Review, November 1995. https://www.law.uh.edu/faculty/jmantel/health-regulatory-process/2014/Wagner95ColumlRev1613.pdf
[2] Union of Concerned Scientists, "The State of Science in the Trump Era (2019), https://www.law.uh.edu/faculty/jmantel/health-regulatory-process/2014/Wagner95ColumlRev1613.pdf
[3] Inhofe, James (Senator), et. al., "Big Government and Bad Science: Ten Case Studies in Regulatory Abuse," November 30, 1999 https://www.law.uh.edu/faculty/jmantel/health-regulatory-process/2014/Wagner95ColumlRev1613.pdf
[4] See, Williams, Richard A. and Kimberly M. Thompson, "Integrated Analysis: Combining Risk and Economic Assessments While Preserving the Separation of Powers," Risk Analysis, 26(6) 2004.
[5] Holman, E., "Part I—Comparing Noncancer Chronic Human Health Reference Values: An Analysis of Science Policy Choices," Risk Analysis, 2016.
[6] Carrington, Clark, "The Science-Policy Shell Game: The Probability of Truth," Amazon Kindle.

science from policy, particularly given two facts about executive branch health and safety agencies:

First, science does not, nor cannot, dictate decisions. Decisions are made by managers, career or political appointees.

Second, EPA is a political agency and the decisions are, of necessity, political. Just as with any political decision, it can be expected that there will be outrage when the decision is not one that is preferred. Different administrations may obscure the politics behind their decisions, or couch their decisions in scientific language, but it does not change the political nature of those decisions.

The point of all of this is that there is a lack of trust in EPA's use of science that goes well beyond the actual quality of the science. The point of the current proposed regulation is to help establish trust by minimizing the potential for using poor science to support decisions.

Some of the comments assert that there is no problem. Although different people may not agree on the extent of the problem, almost everyone seems to agree that there is a problem.[1] Let me mention a few of the issues:

- Funding – it doesn't matter who funds a study, it has the potential to create bias. Whether it is government or industry, scholars want continuous funding and it doesn't help them obtain funding when they reach a result contrary to the agencies or industries interests. For example, a recent study examined 5,675 clinical nutrition, food safety, dietary patterns, and dietary supplement scientific papers for "risk of bias."[2] It came to a surprising conclusion (at least for some): Industry funding "is not consistently associated with producing research results that are considered 'biased' using the standard ROB (risk of bias) criteria" as compared to government-funded research.
- Negative Results – There is a bias against publishing negative results which means that, even if the agency has prepared a balanced review of the science in the preamble or economic benefits analysis, there is missing information.
- Bias – Not only by funding source, but authors may be biased by political or ideological concerns that make them shade their analysis, perhaps in ways that are difficult to detect. One way this manifests itself is using large data sets to find statistical significance (p-hacking) with no theoretical justification.
- Training – many, many scientists are not well trained in statistics resulting in poor models or interpretations of findings.
- Journals are reluctant to investigate or retract poor papers post publication.[3]

[1] See, for example, the survey by Nature that revealed that 52% of scientists thought that there is a significant replicability crisis and only 3% said there was no crisis. https://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.19970

[2] Myers, E.F., et. al., "Using risk of bias domains to identify opportunities for improvement in food- and nutrition-related research: An evaluation of research type and design, year of publication, and source of funding," PLOS ONE, July 5, 2018. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0197425

[3] https://www.nature.com/news/reproducibility-a-tragedy-of-errors-1.19264

- Prepublication peer review is often spotty and some problems are undetectable through peer review. To quote the Center for Open Science "Published and true are not synonyms."[1]

Because of these and other issues, journals are increasingly insisting on open data across all sciences.[2] For example, "funders such as the Bill and Melinda Gates Foundation and the Chan Zuckerberg Initiative have explicit open-data requirements for any research they support, and journals such as *Science*, *Nature*, and many more also now have open-data policies."[3] There is also a Reproducibility Initiative with multiple sponsors including the Science Exchange and PLOS.[4]

The Solution

The primary problem with any solution for epidemiological data is to protect the identifies of the people in the studies. Several comments strongly insist that this is, at least sometimes, impossible. However, EPA is suggesting a tiered approach although it has yet to work out the details.

There are many avenues open beyond simple publication of data that others have mentioned. I will add one. The IRS keeps some of the most sensitive data from being made public but they have still found a way to allow researchers to use their data. [5]

Researchers create a model (in this case, use the existing model), come up with a dummy data set with random numbers to test the model, and the IRS runs the model with actual data. As long as models are made available with a description of the statistics used, EPA or a trusted third party could run the replications so that the underlying private (PII) data would not be made public.

Summary

Clearly EPA is addressing a real problem and is attempting to address it just as many journals and professional organizations are doing. It would be helpful if EPA could provide some of the details on possible solutions to the SAB to provide further assistance.

Richard A. Williams, Ph.D.

[1] https://journals.sagepub.com/doi/full/10.1177/1745691612459058
[2] Nosek Brian A., "Scientific Utopia II. Restructuring Incentives and Practices to Promote Truth over Publishability," Perspectives on Psychological Science, Nov 7, 2012.
[3] Callier, Viviane, "The Open Data Explosion," The Scientist, Jan 1, 2019. https://www.the-scientist.com/careers/the-open-data-explosion-65248
[4] https://www.the-scientist.com/careers/the-open-data-explosion-65248
[5] https://www.sciencemag.org/news/2014/05/how-two-economists-got-direct-access-irs-tax-records