



# American Economic Association

Committee on Economic Statistics

<https://www.aeaweb.org/about-aea/committees/economic-statistics>

and

Office of the Data Editor

April 17, 2020

***This comment responds to EPA proposed rulemaking on Strengthening Transparency in Regulatory Science, Docket No. EPA-HQ-OA-2018-0259, Document RIN 2080-AA14***

The proposed EPA rule would fail to assure that the best possible science is available for regulatory decision making. The American Economic Association's Data and Code Availability Policy demonstrates an alternative that assures scientific credibility without restricting good science because it is conducted with data that cannot, for proprietary, statistical, or legislative purposes, be widely distributed.

The importance of sharing data (and computational instructions, "code") for the purpose of transparency and reproducibility of science is paramount to the American Economic Association (AEA) and for science in general. The AEA's Data and Code Availability Policy (<https://www.aeaweb.org/journals/policies/data-code>) emphasizes this importance, but also exhibits flexibility by explicitly recognizing the scientific credibility of research performed with restricted data, and prescribing alternative approaches to replicability in that case. Specifically, it states that

*If data are subject to any access restriction that prevents authors from depositing the files in an openly accessible data repository, additional information is required. Authors shall provide detailed information on how, where, and under what conditions an independent researcher can access the data. This information shall be provided to the editors upon submission... and the AEA Data Editor will verify the information, and may contact the data providers identified by the authors.*

If data contain confidential information - identifiable person or firm/establishment information - de-identification and anonymization are required both for ethical as well as legal reasons (HIPAA). Such data can and regularly are made available at archives that support relevant access controls. A prime example is the government's own [Federal Statistical Research Data Center](#) (FSRDC) system, which currently [houses confidential data from five agencies](#), is used at any time by several hundred researchers (as of the end of 2018, [760 researchers were active](#)), on [hundreds of approved projects](#) (as of the end of 2018, there were [315 active projects](#)). Using this highly restricted, confidential data on businesses, people, their health, their taxes, researchers each year publish in top journals. In 2018, the latest available numbers show at least 82 publications, most of those in peer-reviewed journals, often in the top journals in economics and other social disciplines. Importantly, many of those researchers and those publications use the same underlying data repeatedly, demonstrating that an active scientific discussion can happen in the presence of confidential data, in an ethical and privacy-preserving fashion. While data is restricted from publication, it is available to a very large community of researchers under proper safeguards.



# American Economic Association

*Committee on Economic Statistics*

<https://www.aeaweb.org/about-aea/committees/economic-statistics>

*and*

*Office of the Data Editor*

The FSRDC system is but one of many systematic, objective, and non-discriminatory systems of accessing confidential health, business, and demographic data in the United States. In fact, the Foundations for Evidence-Based Policymaking Act of 2018 ([P.L. 115-435](#)) explicitly calls for making data available “while protecting such assets from inappropriate access and use”, and mandates OMB to “establish a process” by which a large and diverse cadre of stakeholders can access the data under those conditions, with explicit mention of researchers.

At the AEA, if access to the data used is restricted, the Editor needs to be alerted to such situations at the time of submission. In all such cases, the AEA policy still requests provision of the code, and a description of how the data could be accessed or reproduced by another researcher. The Data Editor may test the access protocol prior to acceptance, and may ask for support by third parties in running the provided code. It is likely that this will *not* always be possible, and the ultimate decision about whether to go forward with the submission, and whether to accept an article that uses restricted data rests with the journal's Editor or Co-Editor. However, the AEA has been successful in asking independent third-parties with access to the same data to conduct verification and reproducibility tests, subject to a stringent protocol, and to the satisfaction of the Data Editor.

The AEA is firmly of the opinion that reliable, verifiable, high-quality science can be conducted when using confidential data.

**Bottom line:** Reliable, reproducible, credible science can be done and has long been done, using non-public data. And because non-public data will remain a feature over the future, for ethical and competitive reasons, we must be able to base policy on the conclusions of knowledge arising from its use.

**The proposed EPA policy would fail in assuring that the best possible science is available for regulatory decision making.** It’s rigid and restrictive interpretation of transparency – that anyone in the public requesting the data can obtain it – is unnecessary when restricted access facilities of trusted intermediaries and associated protocols are in place, widely used, and scientifically acceptable. Other means of avoiding disclosure of proprietary or restricted data, such as data masking, are also sometimes available to assure the reproducibility of research for which full data accessibility must be restricted. EPA’s policy needs to recognize and accommodate these contemporary approaches.

Contacts: Katherine Smith Evans, AEA Washington Area Representative [kitty.s.evans@aeapubs.org](mailto:kitty.s.evans@aeapubs.org)  
Lars Vilhuber, AEA Data Editor [dataeditor@aeapubs.org](mailto:dataeditor@aeapubs.org)