

# Congress of the United States

## House of Representatives

COMMITTEE ON SCIENCE, SPACE, AND TECHNOLOGY

2321 RAYBURN HOUSE OFFICE BUILDING

WASHINGTON, DC 20515-6301

(202) 225-6371  
[www.science.house.gov](http://www.science.house.gov)

August 8, 2013

The Honorable Eddie Bernice Johnson  
Ranking Member  
Committee on Science, Space,  
and Technology  
2468 Rayburn House Office Building  
Washington, DC 20515

Dear Ms. Johnson,

Thank you for your letter about recent Committee action requiring EPA to make public certain tax-payer funded research data that it uses to support its Clean Air Act regulations.

Currently, only a select few are allowed to access and analyze the information in question. This ignores the need for Congressional oversight of taxpayer-funded research. It also prohibits independent verification of EPA's claims.

For nearly two years EPA has failed to live up to its promise to provide Congress with the basic data used to support its regulations. EPA's newly confirmed Administrator Gina McCarthy promised in September 2011 that she would make the data sets available to the Committee. Even though she now heads the agency, she has yet to provide us with the promised information. At the same time, EPA is moving forward with numerous economically harmful regulations, based on the professed findings from this data. In fact, one of these regulations is expected to be the most costly in U.S. history – costing American taxpayers and businesses \$90 billion annually.

The request of the Committee, and the more recently issued subpoena, is based on the principle of transparency, which requires that the information used to justify major, costly regulations be open and available to the public. This principle is supportive of an open government and consistent with the President's March 2009 commitment that "there should be transparency in the preparation, identification, and use of scientific and technological information in policymaking." It is also vital to the integrity of the scientific process, which requires study results to be independently verifiable.

EPA's policy of partial disclosure, where study results are provided but the underlying data remain hidden, does not support this principle. Indeed, the information the agency has made publically available is useless for verification and oversight purposes. EPA itself admitted in April of this year that the information it has provided is "not sufficient" for independent replication or validation.

I have made clear that any personal health information that may be in the subpoenaed data will be protected and removed before the data are made public. However, I have also made clear that the American taxpayers have a right to see this de-identified information and determine whether the EPA is basing its regulations on sound science.

The Committee is not seeking personal medical records, but the files and analytical data sets that have been used by taxpayer-funded EPA grantees. To this end, the subpoena allows the agency to take the precautions necessary to protect private information before it is released.

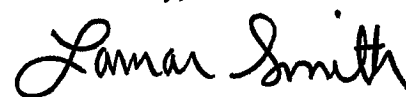
But this is no excuse for further delay or for keeping the data secret. The National Academy of Sciences confirmed in 2005 that de-identification is relatively simple, stating that "The confidentiality of individual data may be protected whether by restricting researcher access to such data (restricted access) or by various alterations that limit the identifiability of the data and hence permit them to be made publicly available (restricted data). Statistical agencies and their contractors currently use both methods." They further explained that "Statistical agencies, working closely with scholars, have for more than 40 years simultaneously improved the technologies that protect confidentiality and the modalities that provide appropriate access to microdata."

In regard to claims that de-identified data sets will not allow independent analysis of EPA's studies, Dr. Stanley Young, Assistant Director for Bioinformatics at the National Institute of Statistical Sciences, has said "It is almost always possible to de-identify data so individual identity is protected and analysis is possible." As has been noted in past communications with EPA, the Health Effects Institute's 2000 re-analysis confirmed that the electronic input and output files used in the actual analysis for these studies do not contain confidential information. Further, if this information cannot be made public in a manner sufficient for validation and re-analysis while protecting confidential information, EPA should not be using it to justify major regulations.

I hope and expect that EPA will provide this information in a manner sufficient for independent validation and replication by the deadline included in the subpoena. It makes no difference whether this information resides at EPA or with another custodian - the Agency has the right under existing policy to obtain, reproduce, publish, or otherwise use data that are funded in whole or part by taxpayers.

Ensuring public access to taxpayer funded-data that are used in regulations supports good science and good government. Consistent with this principle, once the Committee receives the data sets, I intend to make them publically available. Certainly, the principle of an open and transparent government is not supported by policies that allow certain groups access to the information, but prevent access to others. This is precisely what is now occurring and should be corrected.

Sincerely,



Lamar Smith  
Chairman