732 North Washington Street, Alexandria, VA 22314-1943
(703) 684-1221
Fax: (703) 684-2037
Email: asainfo@amstat.org
Web: www.amstat.org

AMERICAN STATISTICAL ASSOCIATION
Promoting the Practice and Profession of Statistics

ASA
FOUNDED 1839
175th ANNIVERSARY
CELEBRATE OUR PAST · ENERGIZE OUR FUTURE

September 5, 2014

The Honorable Kevin McCarthy
Majority Leader
U.S. House of Representatives
Washington, DC 20515

Dear Majority Leader McCarthy,

As president-elect of the American Statistical Association, with 19,000 members, I write regarding H.R. 4012, the "Secret Science Reform Act." We generally applaud the idea that researchers and federal agencies strive to make data available to others—under strict pledges to maintain confidentiality of data provided by individuals and establishments where necessary—and to encourage reproducible research. Access to data and reproducibility of research are crucially important for science to advance.

While H.R. 4012's intent is to make data more widely available, we have several concerns and urge the bill to be revised significantly before further consideration. Our concerns include those voiced by others (especially the American Association for the Advancement of Science) that the bill's statements do not account for the complexities common to the scientific process on research that involves biological materials or physical specimens not easily accessible, combinations of public and private data, longitudinal data collected over many years that are difficult to reproduce, and data from one-time events that cannot be replicated. The bill as written could have far-reaching consequences that would ultimately hamper or undermine the scientific process generally and EPA's work specifically. We also agree with the point that it would be prudent to see the EPA's data access policy—in accordance with the America COMPETES Reauthorization Act of 2010—expected by year's end before further action on H.R. 4012.

Our nation should be striving for transparency in government and, as noted above, data accessibility, but these goals also must be balanced with the necessity to protect individuals' and businesses' privacy. The bill's language of "publicly available" except when "prohibited by law" acknowledges this balance, but that language is vague and may be insufficient to protect individuals and businesses. In particular, some data sets may not fall under "prohibited by law," yet the data are still collected under a pledge to protect the identifiability and confidentiality of the reported values. For example, the government, as well as private and nonprofit sectors, routinely collects data—including private business information and private health information—under strict pledges to protect confidentiality. In some studies, this is backed up with penalties

for violating those pledges. Such data should not be publicly available to every person who might ask for them. Rather, data subjects' confidentiality should be protected, for example by policies and procedures that provide data access to trusted users (i.e., approved users committed to appropriate protections of the confidentiality of study participants) while discouraging breaches of confidentiality and/or by data redaction techniques developed in the statistical and computer science communities. Under the current wording, a choice may have to be made between maintaining data confidentiality and issuing needed regulations.

To emphasize the challenges and importance of confidentiality protection, we note that simple but necessary de-identification methods—like stripping names and other personally identifiable information (PII)—often do not suffice to protect confidentiality. Statisticians and computer scientists have repeatedly shown it can be possible to link individuals to publicly available sources, even with PII removed. Thus, allowing unrestricted public access without appropriate controls could result in unintended disclosures. These could cause significant harm to the advancement of science and the federal government—especially the federal statistical system— as people may be less willing to provide their data if highly publicized breaches occur.

In short, any requirements for making data available should carefully consider the complexities, challenges, and potential ramifications. We hope you will address these concerns, which would require major modifications to the bill. We would be happy to be of any assistance.

Sincerely,
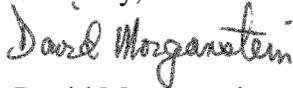
David Morganstein
President-Elect, American Statistical Association

Cc: The Honorable Nancy Pelosi, Democratic Leader
        The Honorable Steve Scalise, Majority Whip
        The Honorable Steny Hoyer, Democratic Whip
        The Honorable Cathy McMorris Rodgers, Republican Conference Chairman
        The Honorable Lamar Smith, Chair, Science, Technology, and Space Committee
        The Honorable Eddie Bernice Johnson, Ranking Member, Science, Technology, and
            Space Committee
        The Honorable Fred Upton, Chair, Energy and Commerce Committee
        The Honorable Henry Waxman, Ranking Member, Energy and Commerce Committee