

Subject: David Wojick comment on EPA proposed transparency rule

May 19, 2018

From: David E. Wojick, Ph.D.

*Inside Public Access*

<http://insidepublicaccess.com/>

To: Administrator Scott Pruitt  
Environmental Protection Agency  
1200 Pennsylvania Avenue, NW  
Washington, DC 20460

RE: Docket No. EPA-HQ-OA-2018-0259  
Strengthening Transparency in Regulatory Science

Dear Administrator Pruitt:

EPA's proposed regulation more-or-less banning the Agency's use of so-called "secret science" has received a lot of attention, much of it negative. What has largely been missed is the deep impact that this rule might have on open science generally.

The most common criticism is that the rule rules out the use by EPA of health studies that include data on individuals. This sort of data cannot be shared, due to privacy laws. These sorts of studies can be very important. There is also the issue of proprietary business data, etc.

But in fact the proposed rule allows for these studies, in two different ways. First, it allows for what is called "masking" of data. If the data is properly structured then masking technologies enable the computer to easily remove or replace the sensitive stuff. Second, in extreme cases the EPA Administrator can simply exempt the study from the regulations.

Leaving the regulatory issues aside, consider the positive aspects for open science. The EPA rule is likely to finally establish specific standards for openness. Moreover, these standards will set a potential precedent for other Federal Agencies, possibly even other Governments, or even for scientific journals.

Open science is a major issue throughout the global scientific community. In other words this relatively small action by EPA is potentially a very big pilot project for the whole world.

The Federal Government already has rules about sharing data that is developed in federally funded research. These rules are part of what is called "Public Access," a program which began in 2013. Every federal science agency has a Public Access Plan mandating that research data be shared.

EPA's proposed rules extend Public Access in a big way. It basically extends the access and availability requirements of the Public Access Program from research that is federally funded to research that is federally used. (In fact EPA specifically cites their Public Access Plan as a supporting document for this new regulation.) The researcher is basically required to provide access to everything technical that is involved in getting the research result. In some ways this proposed rule looks like a natural extension of the existing requirement for a Data Management Plan.

But with any big groundbreaking project comes big challenges. The present proposal is pretty vague when it comes to saying what is actually required. It reads as though the concept of replicability were already well defined, which it most certainly is not.

This is a common problem with ground breaking new laws and regulations. They use language which is clear in its way but which has no operational definition. Working out what these new rules mean is then a complex and difficult matter.

I have been studying this messy phenomenon for almost 40 years, beginning with the confusion surrounding the implementation of the 1970 National Environmental Policy Act. NEPA required all Federal Agencies to do Environmental Impact Statements for all physical projects. But it did not say what these documents looked like or how to do them. It took several years of confusion to work these questions out.

I eventually developed a taxonomy of 126 different regulatory confusions, which anyone is free to use.

See "*Engineer Tackles Regulatory Confusion*"

[http://www.stemed.info/engineer\\_tackles\\_confusion.html](http://www.stemed.info/engineer_tackles_confusion.html)

and

"*A Taxonomy of Confusions*"

<https://scholarlykitchen.sspnet.org/2013/02/05/a-taxonomy-of-confusions/>

EPA's open science rule has the same broad impact and the same degree of vagueness as NEPA did. A great deal of work will have to be done before we know what these new rules actually require in practice. Some of this hard work can be done by EPA in formulating its final rules. But much of it is probably going to be done by the scientific community. A second proposal may be necessary.

At some point EPA is going to have to say, on a very specific case by case basis, which research can be considered and which cannot. This is when the rules get very specific.

First they have to figure out what "using" a given research result even means. For example, proposed major rules are accompanied by a voluminous Technical Support Document. It may cite hundreds, or even thousands, of research journal articles. Does each of these have to meet the availability and replicability standard? Or is regulatory usage confined to just a few key studies?

Second, what does availability mean? For example, does the researcher have to document their data, or just provide it? What about the decisions made as the research progressed, which can be numerous. Does each of these have to be explained? How documented does the software have to be, etc. Here the danger is that the availability requirement might become too burdensome.

But assuming that these deep challenges finally get worked out, consider what this regulation does.

In the environmental field a lot of research is done with federal policy in mind so this is potentially a very broad mandate. It in effect creates the new category of "EPA usable research."

A lot of researchers (or their institutions) are likely to want their work to be EPA usable, even if EPA does not fund it. They will then adopt usability practices from the beginning, which may be a new way of doing research. This is exactly what the open science movement is calling for.

So all things considered this regulation is a big extension of Public Access. It is also a big step forward for open science. But it will be a big job for EPA and the research community to work out.

#### EPA usability and use procedures

Not using a given research result looks to be procedurally easy. EPA simply does not cite the research in its justification documents or statements. They might even create a new support document that specifically identifies the research used to justify the regulation.

Determining which research is not usable is the difficult part. This looks very much like what is called a "certification" procedure. That is, EPA will need to certify that each piece of research it wants to use is properly available for attempted replication. Note that neither attempted or successful replication is required under this procedural transparency rule. All that is required for EPA use is that the research be properly available for possible attempted replication.

The best way, perhaps the only way, to do this is for EPA to contact the researchers and determine that the research is properly available. Another possibility is to establish a practice of self certification among those researchers who desire that their research be usable by EPA. Or certification of EPA usability might be done by third parties, such as scholarly societies. Such certification might even become a condition of funding with some funders.

In any case, what EPA needs to do is specify the requirements of proper availability as needed for usability. This might be a simple checklist of questions, or perhaps something more complex. There is already considerable discussion and action within the open science community, as to just what availability for the purpose of replication requires. EPA should draw on this existing work and expertise.

The basic requirements are not that mysterious. The data and codes need to be available and properly documented. The procedures followed and decisions made along the way need to be properly explained. (There is already a relevant literature on this topic in some fields, especially biomedical.) Research that cannot be certified is then excluded.

I shall be very happy to discuss these important issues in greater detail or to provide EPA with additional information.

Respectfully submitted,

David Wojick, Ph.D.  
<davidwojick@insidepublicaccess.com>