

Sharpless, Katherine E.

From: public-access-bounces@nist.gov on behalf of David Wojick
<davidwojick@insidepublicaccess.com>
Sent: Wednesday, August 12, 2015 3:13 PM
To: public-access
Subject: [Public-Access] Comments on NIST's Public Access plan, as requested
Attachments: NIST Public Access comments from Inside Public Access by Wojick.pdf; NIST Public Access comments from Inside Public Access by Wojick.doc; ATT00001.txt

Hello Katherine Sharpless,

Attached are my initial comments on NIST's Public Access plan, in response to your July 7, 2015 Federal Register notice, in two formats -- PDF and Doc. I will be pleased to answer any questions you may have regarding these comments.

Thank you for this opportunity.

David

David E. Wojick, Ph.D.
<http://insidepublicaccess.com/>

To: Katherine Sharpless,
NIST
By email to public-access@nist.gov
August 12, 2015

Inside Public Access comments on the NIST Plan for Providing Public Access to the Results of Federally Funded Research (NIST Public Access Plan).

NIST officially [requested comments](#) on its Public Access plan in the July 7, 2015 Federal Register. *Inside Public Access* is a newsletter and consulting service with a deep interest in the US Public Access program. We are especially interested in making Public Access work as well as possible, while minimizing the burden on the research community.

We are therefore pleased to make the following comments:

The legal background here is that the Paperwork Reduction Act (PRA), which NIST specifically cites, requires that all federal agencies obtain a clearance before they collect information from the public. Requiring authors to submit copies of their published articles is certainly an information collection under PRA.

This request for comment is presumably a step in the clearance process, or at least feeds into it. NIST may even change its Public Access Policy in response. In any case it will then submit an Information Collection Request (ICR) to OIRA. Central to the ICR is a so-called burden estimate, which typically focuses on the hours of public labor required to understand and comply with the collection requirements.

The point is that this is the time for interested parties to raise whatever issues they have with NIST's proposed Public Access Policy. The primary issue areas are the need for the articles, how to maximize their value and how to minimize the burden of providing them. CHORUS obviously plays into the burden issue, so NIST needs to justify not using CHORUS.

Statutory authority for the collection may also be an issue because there is no clear authority given by Congress for the US Public Access program. It was created by an Executive Branch memo. NIST needs to address this issue, as their Authority section provides no actual statutory authority.

Whether the government has a right to demand and publish these copyrighted articles may also be an issue, in view of the potential for damaging the copyright holders. This brings the copyright issue to the fore and NIST needs to address this issue.

Just to elaborate, I have been talking to the Energy Department's Office of Science about this issue. They claim that their right to collect and publish journal articles is based on the

acquisition regulation 2 CFR 200.315(b).

<http://www.gpo.gov/fdsys/pkg/CFR-2014-title2-vol1/xml/CFR-2014-title2-vol1-sec200-315.xml>

Here is the text of that regulation:

“(b) The non-Federal entity may copyright any work that is subject to copyright and was developed, or for which ownership was acquired, under a Federal award. The Federal awarding agency reserves a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.”

As I read this the journal article (the work) has to have been written (developed) under the award. The award contracts all have specific end dates. Journal articles are often written after the contract end date, because one does not write up the results until the research is over. Such articles are not developed under the award.

As I read this the government has no right to these post-contract articles, but NIST is going to collect and publish them under the US Public Access program. I suppose one can argue that the research was part of the development, but then what is being copyrighted is the ideas not the text and that is incorrect. The copyright only begins with the writing.

The basic point is that the NIST Public Access program is a separate information collection. Submitting published articles to a funding agency is clearly an information collection, because the agency gets the information. But so is one researcher sharing data with another, when that activity is mandated by agency rule, as it is under the Public Access program. So the data sharing requirements will also have to be cleared through OIRA.

Every proposal must include what is called a data management plan or DMP. The DMP says how the researcher will make available and share their data, which then becomes a contract commitment if they are awarded a grant.

It is true that preparing a DMP is part of the burden of a proposal, but there is also the potentially much larger burden of preparing and sharing data. It can be very laborious to prepare data for others to use. It can require a lot of refining, formatting, documentation and explanation.

The point is that NIST should be estimating how much sharing will go on and what the burden is. It appears that this is not being done. More broadly, all of the agency Public Access programs that use the data management plan requirement should be clearing the burden of data sharing with OIRA.

How to estimate (and minimize) the burden of mandated data sharing is a research problem, one that NIST should take seriously. There are really just two questions to be answered in estimating burden. First, how often will required sharing happen? Second, what will the average burden likely be?

Note that the DMP data sharing requirement does not end when the research contract ends, far from it. This is also true of the requirement to submit published articles, but data requests can come long after publication. Article usage data suggests that data requests may well come many years after publication. So the estimate of how many mandated data sharings will occur needs to take into account this potential longevity.

As to the average burden per request, here the problem is the extreme vagueness of the requirement. What is emerging as the standard requirement language is something to the effect of "whatever the community normally requires to validate the research" which NIST specifically cites.

As regulatory language this is hopelessly vague, so hopefully standard practices will develop. It does seem to be leading in some cases to discipline specific agency research programs developing their own [specific data sharing rules](#).

The specialization of data sharing rules may well be happening at other agencies as well, or develop in the future as the Public Access program matures. It may make sense scientifically but it also makes burden estimation more complex. It may also make the goal of minimizing burden more difficult to pursue, given the multiplicity of discipline specific requirements.

A related concern is that this vague DMP requirement can be taken to mean that the data to be shared is whatever the requestor wants. The danger here is that imposing burden via endless requests can be used as a political weapon.

We already see this in some cases with Freedom of Information Act (FOIA) requests. But while FOIA puts the burden on the agency, Public Access data requests put the burden on the public, namely the researcher. Data sharing rules may be needed to limit this sort of abuse.

Note too that the fact that a researcher deposits their data in a repository does not mean that their burden ends. People requesting data from the repository may well then ask the researchers for explanations, more documentation, or even more data. The vagueness of the DMP requirements may contribute to this.

It seems likely that the burden of Public Access data sharing mandates may be large, even very large. The agencies and OIRA need to take this potential burden seriously and take steps to minimize it. The first step is to do proper burden estimating. NIST can take the lead in this important research.

Respectfully submitted by
David Wojick, Ph.D.
<http://insidepublicaccess.com>
davidwojick@insidepublicaccess