CODE

Response to RFI from OSTP: Public Access
From the Center for Open Data Enterprise (CODE)

May 6, 2020

From: Joel Gurin, President, the Center for Open Data Enterprise (CODE)
To: Lisa Nichols, Assistant Director for Academic Engagement, OSTP

This document is being submitted in response to "Request for Information: Public Access to Peer-Reviewed Scholarly Publications, Data and Code Resulting From Federally Funded Research".[1] The Request for Information (RFI) is part of an ongoing effort to facilitate implementation of the 2013 memorandum *Increasing Access to the Results of Federally Funded Scientific Research* (the Holdren Memo).

The Holdren Memo set out policy principles aimed at ensuring "the direct results of federally funded scientific research are made available to and useful for the public, industry, and the scientific community."[2] It directed federal agencies that spend more than $100 million annually on scientific research and discovery to develop and implement a plan to support increased public access to the results of federally funded research. Specifically, agencies are required to ensure that the results of research that they fund are available to the public at no charge within 12 months of initial publication.

The Holdren Memo applies to both the publication of research articles in peer-reviewed journals and the publication of digital data. However, while the Holdren Memo provides a specific timetable for making published articles available for free, its guidance for making the underlying data available is more general, simply calling for "maximize[d] access, by the general public and without charge, to digitally formatted scientific data created with Federal funds." The Center for Open Data Enterprise (CODE) believes that, while agencies have broadly achieved the Holdren Memo's vision of increased public access to research results and peer-reviewed publications, much more needs to be done to ensure public access to data produced through federally funded scientific research.

Our organization has studied a range of issues related to scientific data sharing, and the concerns that inhibit data sharing, over the last several years. CODE is a 501(c)3 nonprofit organization dedicated to maximizing the value of open government data for the public good. Since our founding in 2015 we have worked with stakeholders in government, academia, the nonprofit sector, and the private sector to enable new approaches to data publication, sharing,

---

[1] https://www.federalregister.gov/documents/2020/03/31/2020-06622/request-for-information-public-access-to-peer-reviewed-scholarly-publications-data-and-code
[2]
https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

and exchange. We have worked with the White House and numerous federal agencies to help them leverage data more effectively across a range of applications, including applications that require open access and scientific data sharing.

In 2016, CODE partnered with the White House Office of Science and Technology Policy to hold an Interagency Open Data Roundtable series as a rapid, inclusive way to address and help solve challenges that prevent open data from reaching its full potential. Over four months, the Roundtables addressed open data challenges related to privacy, data quality, sharing research data, and public-private collaboration.[3] Our *Roundtable on Applying Research Data* specifically focused on the question of how to best share and apply data from government-funded scientific research.[4] Through that Roundtable, we developed a series of recommendations aimed at increasing public access data produced through federally funded research, building on existing federal initiatives including the Holdren Memo.

Since 2016 we have conducted projects related to health, oceans, and other topic areas that have provided us with additional insight into best practices for data sharing. In particular, through a series of Roundtables with the U.S. Department of Health and Human Services, we have studied the issues of privacy and data security that must be addressed in sharing biomedical research data and strategies for addressing those issues.

CODE would like to share our insights and recommendations as OSTP works to facilitate implementation and compliance with the 2013 memorandum *Increasing Access to the Results of Federally Funded Scientific Research*. We have three main recommendations.

- **Use new incentives to promote research data sharing more widely.** Currently, there are many incentives against sharing research data and few that support it. The challenges include both cultural and pragmatic obstacles. The current scientific culture is not to share data, but for individual researchers to hold datasets for their own use. The academic model does not reward data sharing. Since academics are rewarded for publishing peer-reviewed articles more than for publishing datasets, researchers want to get maximum publication value out of their data before releasing it. They may also worry that other researchers may not cite them as the source of the data, or that others may interpret the data in ways they would not approve of.

    New ways to reward data-sharing through funding, tenure decisions, and other career incentives could significantly increase data-sharing by researchers. A key is to ensure that researchers receive systematic and meaningful credit for sharing their data. Data citation systems, similar to the citations for published papers, could help researchers gain credit for their work, measure the impact of their research, and advance professionally. They could form the basis for "report cards" that researchers can access to see how their data is being used. This would be similar to the way some organizations now support the use of open source software.

---

[3] http://reports.opendataenterprise.org/2016opendataroundtables.pdf

[4]

http://reports.opendataenterprise.org/KeyTakeawaysonOpenDataforSharingandApplyingResearchData.pdf

While focused on researchers, improved citation systems for data could also help federal agencies and research institutions track the use and impact of the data their grantees produce. Several research-focused and academic organizations have developed data citation systems for this purpose.[5] OSTP should support this effort in coordination with the General Services Administration (including data.gov) and Office of Management and Budget, and in partnership with academic institutions and organizations that support open science.

What can OSTP do?
- ○ Collaborate with the GSA and OMB to review existing and proposed data citation systems, with input from open science and academic organizations.
- ○ After that review, coordinate with GSA and OMB to issue guidance for agencies requiring their grantees to apply a commonly accepted data citation system.

- **Require data sharing and publication as a condition of research funding and help researchers meet that requirement.** While federal guidelines now include an expectation of data-sharing for federally funded research projects, the guidelines can be made stronger and more specific. Guidelines now require grantees to develop data management plans with an expectation that, at minimum, the data underlying publications will be made accessible and shared. Federal funders could tie grants to clearer, binding requirements to adopt open standards and share data publicly to the greatest extent possible, taking privacy and other concerns into consideration. Additionally, federal guidelines currently allow researchers to include data management costs in their proposed budgets.[6] Guidance should require that these costs be included and equal 5 to 10 percent of total project budgets to align with best practices.[7]

  At the same time, federal grant-makers can provide positive incentives and help researchers meet the data-sharing requirement. They can value open data more highly in funding decisions, giving extra points to grant applicants who are committed to sharing their data. Over time, as open data sharing becomes the norm, agencies could also give points to researchers whose public data from previous research has been widely cited. Funders can also provide sample data management plans for federal grantees. While grant applicants are now required to develop these plans, it would be beneficial to clearly encourage data sharing within the goals and recommendations for data management and data infrastructure development.

What can OSTP do?
- ○ Require agencies to include more stringent standards and data sharing requirements in their data management plans.

---

[5] See: https://www.force11.org/datacitationprinciples and https://www.nature.com/articles/sdata2018259
[6] The 2013 memorandum *Increasing Access to the Results of Federally Funded Scientific Research* as well as a Draft NIH Policy for Data Management and Sharing (https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf) currently being circulated include language to this effect.
[7] https://www.frontiersin.org/articles/10.3389/fmars.2019.00440/full, 9-10

- ○ Require agencies to take planned data release into account when awarding new research funding.
- ○ Develop and provide sample data management plans to federal grantees.
- ○ Require agencies to mandate that research project budgets include 5 to 10 percent for data management.

- **Promote biomedical research data sharing while ensuring the privacy and autonomy of research participants.** Sharing biomedical research data is critical to research that can provide new treatments and improve public health. However, such data sharing has been inhibited by legitimate concerns over patient privacy and data security. A number of strategies and approaches have now been developed to protect individual privacy and comply with relevant laws while still sharing health research data.

  While the Holdren Memo requires agencies to account for confidentiality and personal privacy in their public access plans, more stringent legal and regulatory requirements apply to research involving human participants including biomedical and other health research data. Since at least 2003, The National Institutes of Health (NIH) has been developing strategies to meet these requirements while making the results of federally funded scientific research, and data from that research, available to qualified researchers and to the public when possible.[8] The latest version of the *NIH Policy for Data Management and Sharing* specifically "prioritizes the responsible management and sharing of scientific data derived from human participants," and requires researchers to explicitly outline how they will protect human participants' privacy rights and comply with relevant laws and regulations.[9]

  The NIH is the primary federal funder of medical research, but not the only one. Agencies including the Centers for Disease Control and Prevention, National Science Foundation, U.S. Food and Drug Administration, and others also fund biomedical research. These agencies also need to take relevant privacy laws, regulations, and norms into account when preparing their public access plans and releasing research and data.

  What can OSTP do?
    - ○ Specifically require all agencies that fund medical research to address relevant privacy laws, regulations, and norms in their public access plans, using language from the *NIH Policy for Data Management and Sharing* as a basis.
    - ○ Develop and provide privacy best practices for agencies funding biomedical research.

CODE believes that these and similar actions can help ensure that data produced through scientific research is appropriately shared, whether the research is funded by the government or other organizations. Work like the National Academies of Science's ongoing Roundtable on
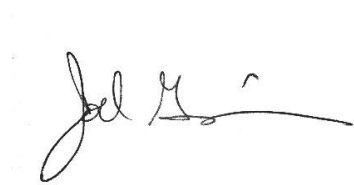
---

[8] https://www.federalregister.gov/documents/2019/11/08/2019-24529/request-for-public-comments-on-a-draft-nih-policy-for-data-management-and-sharing-and-supplemental

[9] https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf

Aligning Incentives for Open Science[10] and SPARC's work on open data[11] show the importance of these issues in academic settings as well as in government. While these recommendations are geared towards OSTP, the issues they tackle should be relevant to anyone interested in sharing scientific data.

CODE applauds OSTP's ongoing work to promote and advance open science and open data across federal research. We thank OSTP for the opportunity to provide our insights and feedback in response to this RFI. We welcome further engagement and look forward to reviewing updates to the 2013 memorandum *Increasing Access to the Results of Federally Funded Scientific Research.*

Sincerely,

Joel Gurin, President, CODE (joel@odenterprise.org)

---

[10] https://www8.nationalacademies.org/pa/projectview.aspx?key=51293
[11] https://sparcopen.org/open-data/