EXECUTIVE OFFICE OF THE PRESIDENT
OFFICE OF SCIENCE AND TECHNOLOGY POLICY
WASHINGTON, D.C. 20502

November 5, 2021

Dear Senate and House Appropriations Committees:

This correspondence is submitted in fulfillment of House Report 116-455 that directs the Office of Science and Technology Policy (OSTP) to report annually on the progress of Federal departments and agencies in implementing their public access plans, including relevant measures of progress, and additional steps being taken to improve access to the results of Federally funded research.

OSTP continues to coordinate Federal public access efforts and identify additional opportunities to enhance access to the results of Federally funded research through the National Science and Technology Council's (NSTC) Subcommittee on Open Science, which meets on a regular basis. This report focuses on progress during Fiscal Year 2021.

Federal public access efforts continue to be guided by the February 2013 OSTP *Memorandum on Increasing Access to the Results of Federally Funded Research*. The Memorandum directs each Federal department and agency with more than $100 million in annual research and development (R&D) expenditures to develop a plan to support increased public access to the results of Federally funded research, specifically access to scholarly publications and digital data resulting from such research.

Of the more than 20 Federal departments and agencies subject to the Memorandum, *all* have developed plans and implemented policies that require Federal scientists and scientists funded by the department or agency to: 1) ensure that the full text of scholarly journal articles reporting on Federally funded research is freely available to the public not more than 12 months after publication; and 2) develop data management plans for new research projects that describe how data generated as part of the research will be made maximally available to other researchers and the public in a timely manner while protecting privacy, security, and other proprietary concerns. Several agencies with annual R&D budgets below the $100 million-dollar threshold are also adopting these principles and practices in their work, resulting in the broader adoption of public access beyond what was mandated in the 2013 OSTP Memorandum.

All of these departments and agencies are now taking steps, individually and with OSTP coordination, to build upon and expand the success of their efforts to improve public access to publications and to other outputs of the research process, such as software, and to advance data management and sharing. The summary below provides an update on these efforts, as well as an overview of key Federal efforts to leverage open science in support of the response to the COVID-19 pandemic.

Mobilizing open science to address COVID-19

The historic coronavirus pandemic demonstrates the importance of timely access to Federally supported research and points to new opportunities for innovation and collaboration in open science.

To support the COVID-19 response, departments and agencies launched and continue to sustain a unique partnership with scholarly publishers, prompted by OSTP and science policy leaders of other nations, to make the full text of scholarly journal articles related to COVID-19 and other coronaviruses freely and immediately available to the public in both human- and machine-readable forms. Launched with 29,000 publications in March 2020, the COVID-19 Open Research Dataset (CORD-19) has since grown to include more than 275,000 full-text journal articles and has been used to motivate several innovation competitions, hosted by public and private organizations, to improve algorithms for extracting meaningful information from biomedical literature and expand public access.[1]

The National Institutes of Health (NIH) launched a Preprint Pilot in June 2020 to accelerate the discoverability of NIH-supported research results that are made publicly available prior to formal publication. NIH has taken steps to help users recognize that the articles have not yet been peer reviewed. The first phase of this pilot focuses on preprints related to NIH-funded COVID-19 research and now includes more than 2,500 preprints.[2]

The Centers for Disease Control and Prevention (CDC) continue to increase the public accessibility of data related to the COVID-19 pandemic. The National Center for Health Statistics continues to release daily provisional counts of U.S. deaths due to COVID-19 by week and state; weekly provisional COVID-19 death counts by select demographic and geographic characteristics, including race and Hispanic origin and age; and estimates of excess deaths related to the COVID-19 pandemic.[3]

Expanding public access to scholarly publications

More broadly, as a result of their public access efforts, Federal departments and agencies now provide free, public access to the full text of more than 2.4 million scholarly journal articles reporting on Federally supported research. Through agreements with scholarly publishers, they provide free public access to an additional 5.7 million articles, mostly in the life and biomedical sciences.[4] Together, these articles are accessed by more than a million Americans per day, driving research and its application to a range of urgent societal needs.

Beyond these efforts, Federal departments and agencies continue to work individually and in coordination with OSTP to improve access to scholarly publications and other written outputs of the research process. For example:

- Nearly all departments and agencies that support Federal science have developed and promote the use of application programming interfaces to facilitate computer-based discovery of and access to scholarly journal articles resulting from their funded research. A compilation of the interfaces is available on-line for easy reference.[5]
- The National Oceanic and Atmospheric Administration (NOAA) has been working to streamline submission processes for its publication repository and reduce burden for authors

[1] See https://www.kaggle.com/allen-institute-for-ai/CORD-19-research-challenge and https://www.ncbi.nlm.nih.gov/pmc/about/covid-19/.
[2] See https://www.ncbi.nlm.nih.gov/pmc/about/nihpreprints/.
[3] See https://www.cdc.gov/nchs/nvss/covid-19.htm.
[4] These articles are provided through the National Institutes of Health's PubMed Central repository, which contains a total of 7.4 million full text articles, many of which are provided voluntarily by publishers. See https://www.ncbi.nlm.nih.gov/pmc/.
[5] See https://www.science.gov/servicesandtools.html.

and program staff, including intramural researchers, extramural grantees, and NOAA's cooperative institutes.

- The Department of Defense (DOD), Department of Energy (DOE), NOAA, and NIH are cooperating to avoid duplicative submission requirements to their article repositories.

Improving research data management and sharing

Departments and agencies continue to make progress in increasing access to data resulting from Federally funded research. As a first step, several departments and agencies recently developed policies for the preparation of data management plans for new research projects. For example:

- The U.S. Department of Agriculture (USDA) Agricultural Research Service issued data management and public access requirements in September 2020. The USDA National Institute of Food and Agriculture established requirements for data management plans to be included in proposals for competitive grant programs and regularly updated over the life of the grant.
- DOD now requires research award recipients to prepare and submit a data management plan for each new award. It also established a new Research Data Working Group with representatives from the Services and DOD research agencies to coordinate research data management and infrastructure across the department.
- Department of the Interior (DOI) issued an Open Data Directive requiring all data assets to be open by default unless protected by law or regulation. All DOI bureaus have established Associate Chief Data Officers who are responsible for coordinating bureau data assets and implementing the Federal Data Strategy and the Foundations for Evidence-based Policymaking Act.
- Department of Transportation (DOT) has organized three public access and data management workshops since May 2021 to assist researchers at the Federal Aviation Administration in practical skills such as writing and evaluating data management plans.
- Agency for Healthcare Quality and Research (AHRQ) is implementing its Policy on Data Management Plans, which was published in May 2020 and promotes the prompt and broad public dissemination of scientific data from its extramural and intramural research programs.[6]
- NASA rolled out an Open Source Science Initiative in its Science Mission Directorate to support implementation of open science practices, including to make data more findable, accessible, interoperable and reusable and provide access to data ready for machine learning.
- NIH released its *NIH Policy for Data Management and Sharing*[7] in October 2020 and is working with stakeholders to ensure seamless implementation when the policy takes effect on January 25, 2023.

As they continue to implement requirements for data management and sharing, Federal departments and agencies are also enhancing data sharing platforms and increasing the amount of research data they make available to the public through a variety of data platforms. Notably:

- DOE began designating Public Reusable Research (PuRe) Data Resources in April 2021. These data repositories, knowledge bases, analysis platforms, and other activities make curated data from authoritative providers publicly available to further scientific discovery and technical knowledge.

[6] See https://grants.nih.gov/grants/guide/notice-files/NOT-HS-20-011.html.
[7] See https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html

- NIH made 16 million genomic sequence records from its genomic Sequence Read Archive available for analysis on two commercial cloud providers. Since January 2021, NIH has added more than 4,000 new results summaries to its ClinicalTrials.gov registry and results database.
- National Institute of Standards and Technology (NIST) makes research data accessible through data.gov and through a dedicated NIST Data Discovery Portal.[8] The number of data files in NIST's public repository grew by an order of magnitude over the last year, from 2,000 in January 2020 to nearly 20,000 in January 2021.
- US Agency for International Development (USAID) has seen steady growth in the number of published data assets in its Development Data Library, which nearly doubled in size between July 2019 and July 2021 to 466 public and restricted-access data assets.
- USGS Science Data Catalog increased the number of records describing publicly accessible scientific data by 39% since 2016 to 21,748 results in FY 2021.

Improving access to other outputs of Federally funded research

Beyond publications and data, several departments and agencies are taking steps to improve access to software, code, and models generated as part of funded research. For example:

- DOE makes scientific software discoverable and accessible through its software services and search platform, DOE CODE. Over the last year, DOE CODE grew to include 3,537 software projects that are accessible to the public. Most of the code is open source.[9]
- NIST ensures software produced for research purposes is inventoried and made accessible to the public through its new NIST Opensource Portal.[10]
- USGS designed and launched a prototype Model Catalog to encourage discovery and reuse of scientific models.[11] The catalog currently contains 115 scientific models, connecting each to their related publications, software code, and authors.

Improving the ability to track Federally funded research

Departments and agencies are also making progress to increase the use of digital persistent identifiers for research-related publications, datasets, federal awardees, and awards. Digital persistent identifiers are a critical enabler of open science, providing unambiguous identification of resources and enabling interlinking of related entities. Among recent examples:

- DOE launched a pilot Award Digital Object Identifier Service, which assigns persistent identifiers to awards made by its user facilities. DOE also leads the US Government ORCID Consortium, which aims to provide unique identifiers for researchers that can allow linking to their research outputs and funding sources. DOE also offers services for registering datasets with standard identifiers and has assigned identifiers to 151,000 datasets.
- NASA also expanded its persistent identifier services, upgrading its service contract for agency-wide coverage and testing an identifier management system to support agency offices wishing to integrate persistent identifiers into data and information repositories and archives.
- NIST instituted a requirement that its researchers and associates obtain and use a persistent author identifier in their publications.

[8] See https://data.nist.gov/sdp/#/.
[9] See https://www.osti.gov/doecode/.
[10] See https://code.nist.gov/.
[11] See https://data.usgs.gov/modelcatalog/.

Sustaining Federal Coordination

OSTP continues its coordination of public access and open science efforts through the NSTC Subcommittee on Open Science. This work led to consensus around key elements of data management plans and desired characteristics of repositories for data from Federally funded research. Departments and agencies can use this work to provide more consistent information to the research community about data management. The Subcommittee on Open Science is completing a toolkit to share good practices for implementing public access policies, including communication with affected stakeholders, simplification of article-submission procedures, and effective monitoring of compliance. Departments and agencies are also developing toolkits for data management and adoption of persistent identifiers to facilitate interlinking of data.

In sum, Federal departments and agencies continue to make steady progress in improving public access to the results of Federally funded research. The COVID-19 pandemic highlighted the importance of open science to society and urgent national priorities and demonstrated new paths for innovation. OSTP aims to continue its coordination efforts to further improve the efficiency and effectiveness of agency initiatives, as well as to reduce the burden on the research community. In addition, future efforts will aim to align Federal-wide public access efforts with emerging priorities related to equity, research security, and infrastructure for open science. OSTP continues to meet with stakeholders from the research, scholarly publishing, library, and other communities to gather information and perspectives to guide next steps in improving public access to the results of Federally funded research and maximize the value of the government's investment in research.

Sincerely,

Eric S. Lander
Director, Office of Science and Technology Policy
and
Assistant to the President for Science and Technology