Advancing Open Science to Accelerate Research, Development, and Innovation for American Prosperity

A Report by the

Interagency Working Group on Open Science

Committee on Science

National Science and Technology Council

May 2018

About the National Science and Technology Council

The National Science and Technology Council (NSTC) is the principal means by which the Executive Branch coordinates science and technology policy across the diverse entities that make up the Federal research and development enterprise. A primary objective of the NSTC is to ensure science and technology policy decisions and programs are consistent with the President's stated goals. The NSTC prepares research and development strategies that are coordinated across Federal agencies aimed at accomplishing multiple national goals. The work of the NSTC is organized under committees that oversee subcommittees and working groups focused on different aspects of science and technology. More information is available at http://www.whitehouse.gov/ostp/nstc.

About the Office of Science and Technology Policy

The Office of Science and Technology Policy (OSTP) was established by the National Science and Technology Policy, Organization, and Priorities Act of 1976 to provide the President and others within the Executive Office of the President with advice on the scientific, engineering, and technological aspects of the economy, national security, homeland security, health, foreign relations, the environment, and the technological recovery and use of resources, among other topics. OSTP leads interagency science and technology policy coordination efforts, assists the Office of Management and Budget with an annual review and analysis of Federal research and development in budgets, and serves as a source of scientific and technological analysis and judgment for the President with respect to major policies, plans, and programs of the Federal Government. More information is available at http://www.whitehouse.gov/ostp

About the Interagency Working Group on Open Science

The Interagency Working Group on Open Science (IWGOS) advances Federal efforts to support open science by increasing access to and use of the results of federally-funded research and development, including but not limited to scholarly publications and digital data. Among its responsibilities, the IWGOS aims to improve implementation of policies to increase access to the results of federally-funded scientific research and to identify additional steps that Federal departments and agencies can take to enhance the preservation, discoverability, accessibility, quality, and utility of the outputs of federally-funded scientific research.[1]

Copyright Information

This document is a work of the United States Government and is in the public domain (see 17 U.S.C. §105). Subject to the stipulations below, it may be distributed and copied with acknowledgement to OSTP. Copyrights to graphics included in this document are reserved by the original copyright holders or their assignees and are used here under the government's license and by permission. Requests to use any images must be made to the provider identified in the image credits or to OSTP if no provider is identified.

Printed in the United States of America, 2018.

[1] See https://www.whitehouse.gov/sites/whitehouse.gov/files/ostp/IWGOS_Charter.pdf

NATIONAL SCIENCE & TECHNOLOGY COUNCIL

Chair
Ted Wackler, Deputy Chief of Staff and Assistant Director, OSTP

Staff
Chloe Kontos, Executive Director, NSTC

COMMITTEE ON SCIENCE

Co-Chairs

Staff

ED_002389_00007571-00003

INTERAGENCY WORKING GROUP ON OPEN SCIENCE

Co-Chairs *Executive Secretary*

Members

ED_002389_00007571-00004

List of Acronyms

ACL	Administration for Community Living (DHHS)
AHRQ	Agency for Healthcare Research and Quality (DHHS)
ASPR	Assistant Secretary for Preparedness and Response (DHHS)
CDC	Centers for Disease Control and Prevention (DHHS)
DHHS	Department of Health and Human Services
DHS	Department of Homeland Security
DOC	Department of Commerce
DOD	Department of Defense
DOE	Department of Energy
DOT	Department of Transportation
ED	Department of Education
EPA	Environmental Protection Agency
FDA	Food and Drug Administration (DHHS)
IARPA	Intelligence Advanced Research Projects Activity (ODNI)
IWGOS	Interagency Working Group on Open Science
ODNI	Office of the Director of National Intelligence
NASA	National Aeronautics and Space Administration
NIH	National Institutes of Health (DHHS)
NIST	National Institute of Standards and Technology (DOC)
NOAA	National Oceanic and Atmospheric Administration (DOC)
NSF	National Science Foundation
NSTC	National Science and Technology Council
OMB	Office of Management and Budget
OSTP	Office of Science and Technology Policy
PAGES	Public Access Gateway for Energy and Science
PMC	PubMed Central
SI	Smithsonian Institution
USAID	United States Agency for International Development
USDA	United States Department of Agriculture
USGS	United States Geological Survey
VA	Department of Veterans Affairs

ED_002389_00007571-00005

ED_002389_00007571-00006

Executive Summary

Open science is the principle and practice of making the results of scientific research available, accessible, and usable. By removing barriers to the access and use of scientific information, open science can accelerate scientific progress, improve research rigor and reproducibility, spur innovation, and strengthen the economy, increasing the return on investments in research and development.

In recent years, Federal departments and agencies have made considerable progress in advancing open science through policies to increase public access to scientific publications and digital data resulting from federally-funded research. This work has highlighted two key lessons that can apply more generally to efforts to advance open science: First, implementation needs to be tailored to the specific missions, organizations, and fields of science that individual agencies support. While agencies can pursue common objectives, the specific ways they achieve them may differ. Second, interagency cooperation can significantly improve the efficiency and effectiveness of open science implementation, through the sharing of knowledge, infrastructure, best practices, solutions to common challenges, and more.

This report was prepared by the Interagency Working Group on Open Science (IWGOS), which was established to advance Federal efforts to support open science by increasing access to, and use of the results of, federally-funded research and development, including but not limited to scholarly publications and digital data.

The IWGOS identified several opportunities for advancing open science, taking into account agency missions, operational contexts, and resources. These include:

1. Increasing the availability and timeliness of narrative research products;
2. Improving the availability and timeliness of access to scientific data;
3. Increasing access to other products resulting from federally-funded research, such as software and scientific collections;
4. Improving systems for making research products public and more usable;
5. Lowering barriers to effective use of open science content; and
6. Developing better metrics for measuring advances in open science and its impacts.

To realize the benefits of open science, the IWGOS recommends that Federal agencies take steps to incentivize open science; facilitate interagency development of best practices, guidance, and standards for open science, with a focus on data access; and promote interagency and public-private collaborations to further realize the benefits of open science. Such collaboration would include partnerships with other Federal working groups; public, nonprofit, and academic partners; and international partners, many of which are also pursuing the objective of advancing open science.

ED_002389_00007571-00007

I. Introduction

Open science is the principle and practice that makes the results of scientific research available, accessible, and usable. It accelerates progress in science, improves research rigor and reproducibility, spurs innovation, and strengthens the economy. By making the results of scientific research accessible to the largest possible audience – including other scientists, business innovators, entrepreneurs, students, and American citizens – open science can boost the returns from Federal investments in research and development (R&D). It can drive advances in areas as diverse as agriculture, energy, health, and national security, and catalyze innovation that drives economic growth and prosperity.

The goals of open science are reflected in the public access plans developed by federal agencies.[2] Open science is essential to achieving the compelling R&D priorities outlined by the Director of the Office of Management and Budget, Mick Mulvaney, and Deputy Assistant to the President, Office of Science and Technology Policy, Michael Kratsios, in Memorandum M-17-30 entitled "FY 2019 Administration Research and Development Budget Priorities." Specifically, open science can further administration priorities by advancing American Security, American Prosperity, American Energy Dominance, and American Health.[3]

Accelerating dissemination of research results advances knowledge and leads to broader application of American scientific contributions. Fundamentally, communicating research results is central to scientific progress, allowing others to test the validity of findings and to build upon these results. Open science allows research to be more easily applied to a range of national priorities, including:

- *Protecting American interests.* Making federally-supported science more open and accessible helps to establish American leadership in scientific research and in areas such as international data standards, which can directly impact American scientific and economic interests.

- *Advancing American health.* Making medical research accessible in real time improves patient diagnoses and treatment. Providing access to literature search through clinical decision support systems affords statistically significant improvements in the accuracy of making evidence-based clinical decisions in a physician's everyday clinical practice. In one study, access to online literature was found to boost the accuracy of clinical decision making from 45 percent to 72 percent.[4]

- *Protecting American security.* The Departments of Defense (DOD) and Homeland Security (DHS) maintain robust research portfolios to keep Americans safe, both at home and abroad. Public access to the unclassified results of this research helps to encourage and focus innovation on technologies that, in time, will become integral to our nation's security.

---

[2] Agency public access plans are available online at: https://www.science.gov/publicAccess.html
[3] OMB Memorandum M-17-30, *FY 2019 Administration Research and Development Budget Priorities* (Aug. 17, 2017), available at https://www.whitehouse.gov/sites/whitehouse.gov/files/omb/memoranda/2017/m-17-30.pdf
[4] Tennant, J. P., Waldner, F., Jacques, D. C., Masuzzo, P., Collister, L. B., & Hartgerink, C. H. J. (2016). The academic, economic and societal impacts of Open Access: an evidence-based review. F1000Research, 5, 632. http://doi.org/10.12688/f1000research.8460.3.

ED_002389_00007571-00008

- *Ensuring American energy dominance.* Making the latest scientific evidence available to private sector energy pioneers helps bring Department of Energy (DOE) research to the market faster. For example, through the Gateway for Accelerated Innovation in Nuclear, DOE makes information, data, and R&D infrastructure available to public and private sector stakeholders to speed the development of innovative nuclear energy technologies toward commercial readiness.[5]

- *Increasing government accountability and efficiency.* Making the results of federally-funded research easily available to researchers avoids unnecessary duplication of research, facilitates measurement of the return-on-investment in research, enables needed replication of experiments to improve reliability of results, and permits collected data to be reused for further experimentation and analysis.

- *Maximizing interagency coordination.* By working together, federal agencies have been able to improve the efficiency and effectiveness of their public access policies. For example, eleven agencies use the National Institutes of Health's (NIH) PubMed Central (PMC) to provide public access to scholarly publications, and DOE has partnered with three other agencies further leverage and customize its Public Access Gateway for Energy and Science (PAGES).

- *Supporting Innovative Early Stage Research.* Effective management of early stage research products, including making these products more accessible, has the potential to increase the pace of scientific discovery, accelerate innovation, and promote more efficient and effective use of government and private-sector funding and resources.

Advancing American prosperity through open science. Open science grows the American economy and creates American jobs. Science has more impact, and Federal investments yield better returns, when results are easily accessible and reusable.

- The open flow of ideas, technologies, and people between the U.S. academic sector (funded by Federal investments) and industry have introduced entirely new categories of products and services that have ultimately become billion-dollar industries. In the area of information technology, the "tire tracks diagram"[6] illustrates how companies that create products using the open ideas that result from federally-sponsored research "repay the nation in jobs, taxes, productivity increases and world leadership."

- Sharing data from new technologies in a common format can create a whole new economic sector. Data from the U.S.-developed Global Positioning System (GPS) provide positioning, navigation, and timing information to all parts of the economy, and serve as

[5] For additional information, see https://gain.inl.gov/SitePages/Home.aspx.
[6] Computer Science and Telecommunications Board, National Research Council, *Continuing Innovation in Information Technology*, National Academies Press, 2016.

ED_002389_00007571-00009

a time and frequency standard for the world.[7] Without such a standard, companies would have to use multiple systems, and the world's economy would be less efficient.

- For every $1 invested by the Federal government, the Human Genome Project's open research data have resulted in the return of $141 to the U.S. economy. Between 1988 and 2010, human genomics generated an economic output of $796 billion, personal income exceeding $244 billion, and 310,000 jobs. The benefits of the Human Genome Project have been widespread and continue to increase over time. [8]

- Publicly available genomic data has also been used to minimize the economic damage from invasive species and support the rapid identification of invasive species. [9] In one example, locating key genes involved in insecticide resistance, mating behavior, and sterility of the Mediterranean Fruit Fly, a highly invasive, economically disastrous crop pest.[10]

- The use of freely available government satellite imagery from the Landsat satellite missions of the U.S. Geological Survey (USGS) generated annual economic activity more than double the agency's annual budget.[11]

- The National Animal Nutrition Program makes data on more than 1.5 million feed ingredients openly accessible online. Previously isolated information is now routinely used by researchers, extension staff, producers, regulators, and industry professionals in more than 30 countries to improve livestock feeding strategies.[12]

Barriers to scientific information decrease scientific impact and increase costs. Closed science can inhibit creative and entrepreneurial reuse of research results, reducing the return on investment. This is especially true of interdisciplinary research or research that requires knowledge from beyond a single discipline. Sharing scientific advances broadly across disciplines can lead to novel applications. For example, the National Aeronautics and Space Administration (NASA) has a long history of analyzing weak radio signals to identify seismic activity or minor alterations in a satellite's trajectory around another planet. NASA technology was adapted for use in picking out the faint heart beat and breathing of victims trapped under a

---

[7] Leveson, I., "GPS Civilian Economic Value to the U.S.," Report prepared for the National Executive Committee for Space-Based Positioning, Navigation and Timing, 2015.
[8] Simon Tripp and Martin Greuber, "Economic Impact of the Human Genome Project," Battelle Memorial Institute, May 2011.
[9] Lyal, C. Capacity of United States Federal Government and its partners to rapidly and accurately report the identity (taxonomy) of non-native organisms intercepted in early detection programs. https://www.doi.gov/sites/doi.gov/files/uploads/lyal_federal_capicity_taxonomy_draft_8march2018.pdf
[10] Papanicolaou, A., Schetelig, M. F., Arensburger, P., Atkinson, P. W., Benoit, J. B., Bourtzis, K., Handler, A. M. (2016). The whole genome sequence of the Mediterranean fruit fly, Ceratitis capitata (Wiedemann), reveals insights into the biology and adaptive evolution of a highly invasive pest species. Genome Biology. http://doi.org/10.1186/s13059-016-1049-2.
[11] H. Miller, L.Richardson, S. Koontz, J. Loomis, L. Koontz, "Users, Uses, and Value of Landsat Satellite Imagery—Results from the 2012 Survey of Users," U.S. Geological Survey Open-File Report 2013–1269, 2012.
[12] The State Agricultural Experiment Stations, and Hatch Funds provided by the National Institute of Food and Agriculture, U.S. Department of Agriculture, animalnutrition.org.

pile of rubble and spun off into the FINDER device that is being used by the Federal Emergency Management Agency and DHS to locate victims of natural disasters, such as the earthquake in Nepal in 2015.[13]

The United States is particularly well positioned to reap the benefits of open science. While open science is a global phenomenon,[14] the strong connections between American industry, universities, and government enable the United States to be at the forefront of innovation with its ability to translate new scientific knowledge into products and services. Open science strengthens these relationships by providing another channel for industry to access the latest research findings and data stemming from both, U.S. universities and Federal government laboratories.

==========

The IWGOS was charged to prepare a report that recommends "additional objectives for Federal open science policies that further enhance access to a broad range of results of federally-funded scientific research, consistent with agency capabilities and missions" and outlines "effective strategies for improving preservation, discoverability, and accessibility of scientific data, taking into account the capabilities and activities of Federal departments and agencies, as well as the private sector."[15]

The IWGOS views open science as a principle and practice that can help guide Federal investment in science. If Federal research results in publications, those publications should be as easy to access as possible. If the Federal government invests in data production, those data should be shared as broadly as possible. Investments in openness amplifies the value of research investment. The IWGOS recognizes that important practical considerations must inform the degree of openness for any research product. These include, but may not be limited to: law; agency mission; resource constraints; U.S. national, homeland, and economic security; research participant privacy; and proprietary interests.

This report provides appropriate next steps the Federal government can take to further open science, incorporating these important considerations.

The report is structured as follows:

- Section II reviews U.S. Federal government activities taken to date to promote access to the results of federally-funded research and identifies the most important lessons learned.

---

[13] Howard, C. B. "NASA Technology Finds Nepal Survivors by Their Heartbeats," *National Geographic* ,May, 7 (2015).
[14] For example, see OECD (2017), "Business models for sustainable research data repositories", *OECD Science, Technology and Industry Policy Papers*, No. 47, OECD, Paris. http://dx.doi.org/10.1787/302b12bb-en. .
[15] National Science and Technology Council, "Charter of the Interagency Working Group on Open Science," October 2016. See https://www.whitehouse.gov/sites/whitehouse.gov/files/ostp/IWGOS_Charter.pdf.

ED_002389_00007571-00011

- Section III identifies opportunities for further increasing access to publications and data resulting from federally-funded research and options for expanding the reach of open science to include other research products.

- Section IV recommends steps that the Federal Government might consider in pursuit of the identified opportunities, consistent with agency missions, capabilities, and resources.

II. Accomplishments and Lessons Learned

The U.S. government is making considerable progress in increasing public access to scientific publications and data. As of May 2018, twenty-two Federal departments and agencies (accounting for 99% of the Federal R&D budget) had completed development of public access plans for publications and data resulting from funded research: U.S. Department of Agriculture (USDA), DOD, DOE, DHS, Department of Transportation (DOT), Department of Education (ED), Department of Veterans Affairs (VA), Environmental Protection Agency (EPA), NASA, National Institute of Standards and Technology (NIST), National Oceanic and Atmospheric Administration (NOAA), National Science Foundation (NSF), Office of the Director of National Intelligence (ODNI), Smithsonian Institution (SI), USGS, U.S. Agency for International Development (USAID), and five operating divisions within the Department of Health and Human Services (DHHS): Administration for Community Living (ACL), Agency for Healthcare Research and Quality (AHRQ), Office of the Assistant Secretary for Preparedness and Response (ASPR), Centers for Disease Control and Prevention (CDC), Food and Drug Administration (FDA), and National Institutes of Health (NIH).[16]

Agency public access plans describe how the agencies will implement policies to ensure that publications are freely available not more than 12 months after publication and that data management plans are prepared and evaluated for new research projects. The policies aim to maximize the sharing of data while respecting privacy, security, and other legal limits. Development of these plans and implementation of resulting policies was informed by numerous public consultations.

The benefits associated with public access policies are already coming to fruition. As of May 2018, agency systems were making more than 5 million research articles and 250,000 agency data sets publicly available, including research data that are available on Data.gov.[17] Millions of Americans were making use of these resources every day.

Federal agencies continue to improve upon their public access efforts.

- DOE, for example, added scientific software to its public access efforts with the November 2017 launch of DOE CODE, an open source platform that makes it easy for DOE-funded researchers and scientific software developers to share scientific software and for the public to discover DOE-funded code. The alpha release contained roughly 700 open source

[16] See department and agency public access plans at https://www.science.gov/publicAccess.html 2017
[17] Estimates based on information available at designated agency public access repositories and at Data.gov.

software projects, a volume which is expected to increase significantly over the next 12 months.

- USGS put in place a process for certifying USGS Public Access data servers as Trusted Digital Repositories for data, and three USGS repositories have been so certified. Additional USGS systems providing public access to scholarly publications and data have initiated or are exploring the possibility of certification under this process.

- NIH launched in autumn 2017 its Data Commons Pilot Phase, an initiative to test the feasibility of, and develop best practices for, making NIH-funded data sets and computational tools available through communal, collaborative platforms on public clouds. The Data Commons Pilot Phase made nine awards using a special type of funding mechanisms.

Federal agencies continue to gain considerable experience through the development and implementation of their public access plans – lessons that inform future Federal efforts to advance open science. Two of these experiences transcend efforts to increase access to both publications and data:

- **Agencies need flexibility in implementing policy solutions.** Flexibility is necessary because of differences in agency missions, capabilities, and authorities, as well as the mechanisms they use to support R&D, the balance of intramural versus extramural research, and the different types of publications and data that result from their funded research. While they can adhere to common principles, they should also be able to tailor the specific details of implementation in a way that is relevant to their agencies.

- **Interagency cooperation has broadened the understanding of the challenges involved in developing public access policies and continues to improve the efficiency and effectiveness of public access implementations.** Through shared knowledge and services departments and agencies continue to overcome challenges. A number of agencies are using shared infrastructure and common procedures for providing public access to publications and, consistent with the M-17-30 R&D priorities memo, which notes that "agencies should maximize the coordination, promotion, and planning of their R&D programs through the National Science and Technology Council (NSTC)."

Increasing Access to Publications

In terms of increasing public access to scholarly publications resulting from federally-funded research, agency efforts to date demonstrate the following:

- **Interagency collaboration can improve implementation.** Collaboration allows agencies to establish joint solutions (see sidebar "Implementing Public Access to Publications: The Value of Interagency Collaboration"). Although agencies and scientific communities have different needs with regard to the necessary infrastructure, there is enough commonality that a relatively small set of platforms is being used to preserve and provide access to publications, maintaining simplicity and usability for researchers and end users.

ED_002389_00007571-00013

- **Successful implementation of public access policies leverages public–private partnerships.** Private sector publishers are actively supporting public access policies in a number of ways:
  - DOE PAGES complements author and institutional deposits of metadata and accepted manuscripts with publisher-supplied metadata and links to full text publications on publisher websites.
  - NIH's PMC has agreements with more than 2,000 journals to submit all of their final published content, or in some cases only NIH-funded content, directly to PMC. More than 70% of papers collected are submitted with active support from the publisher.
  - SI developed blanket contract addenda with its major publishers for all published journal articles that are authored by one or more SI employees, expediting review and execution of publishing agreements for a large portion of articles.

Implementing Public Access to Publications: The Value of Interagency Collaboration

While more than 22 Federal agencies have developed public access plans, agencies are sharing infrastructure where possible and collaborating on a handful of systems. For example, 11 agencies make use of the NIH's PubMed Central (PMC).[18] DOE has partnered with the NSF, ODNI's Intelligence Advanced Research Projects Activity (IARPA), and DOD to further leverage and customize its Public Access Gateway for Energy and Science (PAGES) infrastructure for use by those agencies. These agencies are developing a multi-agency portal for submitting articles to their systems. Both NOAA and DOT are making use of CDC's Stacks system. This experience indicates that a small set of platforms can support a large number of agencies with common approaches and needs.

- **Public access can be accomplished without disrupting existing avenues for scientific communication.** For example, in May 2008, NIH's Public Access Policy began to require peer-reviewed scientific papers to be available without charge on the Internet 12 not later than months after publication. Concerns that this policy would undermine scientific publishing proved unfounded – from 2007 to 2011, the number of biological sciences and agriculture journals as well as medicine and health journals grew by 15% and 19%, respectively.[19] Instead, publishers created new services around public access, such as a system that aggregates article metadata from participating publishers. This system is used by several agencies, including DOD, DOE, IARPA, NIST, NSF, SI, USDA, and USGS.

- **Interoperability with existing systems can streamline procedures and minimize burden.** A number of agencies (e.g., NIH, NSF, SI) have successfully tied deposition of publications into a public access repository with their existing award management and other systems. In some cases, integrating a public access repository with existing systems can reduce data entry and tracking burdens for investigators, and oversight burdens for agencies. Many functions can even be automated.

[18] Those making use of PMC include six agencies and offices within DHHS--ACL, AHRQ, ASPR, CDC, FDA, and NIH—and five other departments and agencies--DHS, VA, EPA, NASA, and NIST

[19] National Institutes of Health, "The NIH Public Access Policy," April 2012. Available online at: https://publicaccess.nih.gov/public_access_policy_implications_2012.pdf

ED_002389_00007571-00014

- **Public access to publications can be achieved without major costs.** Agencies are implementing public access within "existing budgetary resources." For example, DOE, USGS, and SI rely on existing infrastructures that collect other forms of scientific and technical information and made the incremental change of collecting final accepted manuscripts. In other cases, (e.g., NSF) public access implementation costs were minimized by leveraging existing infrastructure at DOE (see sidebar). While there are costs involved, for some agencies, those costs have not been large, especially when measured against the investment in research.

- **Researchers can comply with agency policies.** NSF requires the deposit of journal and juried conference publications as part of a Principal Investigator's annual/final report, making compliance follow naturally from existing business processes. The NIH experience with public access to publications, wherein publications are required to be publicly available on the internet within 12 months, has a compliance rate of 88% over its more than 10-year history, suggesting that post-award compliance is a reasonable expectation. One reason for this behavior is that compliance is public, and scientists must cite these works in applications for new NIH funding.

Increasing Access to Data

With respect to data and the preparation of data management plans, agency experience to date demonstrates that:

- **Data management plans are an important tool for agency stewardship of data resulting from Federal research funding**. Agencies that began receiving data management plans for new research projects[20] have found that researchers are starting to submit stronger data management and sharing plans over time. Additionally, those agencies are finding that the plans improve communication between the researcher and the Federal funding agency about the disposition, planning, and support for data arising from a project. A plan that includes components for both management and sharing of data requires investigators to think through issues such as long-term preservation of data, data distribution, and limitations (if any) on data access and use.

- **Data management plans must balance common elements against discipline-specific needs**. Expectations for the specific content of data management plans vary across disciplines and across agencies. The NSF and DOE experience with data management plans, dating back to 2011 and 2015, respectively, demonstrates that practices of the various scientific disciplines differ considerably. Specifically, plan evaluation by scientific domain experts during the proposal review process is critically important. Nevertheless, agencies identify similar categories of content to be included in data management plans and make use of good practices and materials developed by other agencies to improve consistency across agencies. Some, such as the NSF, developed agency-wide guidance that individual directorates and divisions (within directorates) then interpret and implement, while adding their own discipline-specific guidance, to meet the specific needs of their research communities.

[20] For information on the implementation dates of Federal agency policies for data management plans, see https://www.science.gov/publicAccess.html#AwardDates

ED_002389_00007571-00015

- **Sharing interim data is beneficial in some circumstances.** Accessible, quality interim data, in addition to final data, are essential to rapid response in emergencies. Particularly in health crises where time is a valuable resource, the sharing of interim-data can save the lives of Americans and people globally. Two examples from the CDC are the outbreak of *Elizabethkingia anopheles*,[21] and the sharing of provisional of data on the Zika outbreak.[22] Similarly, preliminary, near-real time data from the USGS associated with earthquake intensity, location of volcanic ash plumes, and flooding of rivers and streams are essential to public health and safety when associated natural disasters occur.

- **Privacy and confidentiality can be protected, while allowing for meaningful sharing of data.** Even in situations involving personally identifiable data where privacy and confidentiality must be protected, meaningful sharing of data is still possible. For example, NIH developed effective approaches for allowing researchers to reuse individual level genomic data while protecting patient privacy.[23] Likewise, in 2005, CDC developed an early warning system for potential tuberculosis outbreaks based on research from actual case reports. This system is useful in identifying outbreaks and tracking trends, while maintaining privacy and confidentiality.[24]

- **A combination of public and private-sector repositories can support data preservation and access.** Data preservation and access makes use of a diverse set of publicly- and privately- supported data repositories. Many Federal agencies created or contracted with (or externally funded the creation of) effective data repositories that store data resulting from their funded research. Many private sector solutions (institutional, for-profit and nonprofit; discipline-specific and multidisciplinary) also exist and can provide long-term data preservation, access, and discovery capabilities. Given the diversity of data types generated by federally-funded research, a combination of public and private-sector repositories will likely become part of the solution to data preservation and access needs.

- **Data can be preserved and made accessible over the long-term with time-limited funding.** Many repositories ensure long-term preservation with a one-time deposit fee, which is reimbursable from a Federal research grant. Nevertheless, there is still great diversity among agencies about how researchers should address the need to provide for long-term data preservation and which repositories are most suitable. Some agencies offer explicit guidance about particular repositories to be used, while others provide more general guidance to assist awardees in selecting repositories. Federal agencies should consider offering guidance to awardees for including provisions in data management

---

[21] Figueroa Castro CE, et al, Elizabethkingia anophelis: Clinical Experience of an Academic Health System in Southeastern Wisconsin, Open Forum Infectious Diseases 5(4), April 2018.
[22] Statement on Data Sharing in Public Health Emergencies, *The Lancet.* Available online at http://www.thelancet.com/campaigns/zika/statement
[23] See https://federalregister.gov/a/2014-2038; https://grants.nih.gov/grants/guide/notice-files/NOT-OD-14-124.html
[24] Mindra G, Wortham JM, Haddad MB, and Powell KM, "Tuberculosis outbreaks in the United States, 2009-2015," *Public Health Reports.* 2017:132(2), 157-163.

ED_002389_00007571-00016

plans regarding necessary standards for repositories and addressing compliance challenges associated with data meant to be long-lived.

- **Public access to data is more challenging than for publications.** Digital data sets are much more heterogeneous than publications, and therefore more technically difficult to curate and share. A number of important remaining challenges are noted here:

  - *Appropriate incentives are needed to promote culture change.* Datasets are standalone research products that are valuable in their own right and which researchers are accustomed to retaining. Federal agencies are encouraging a new norm, that this scientific data should be shared as consistent with agency mission. This cultural change requires incentives that help credit and reward investigators for sharing their data. Clear incentives will simplify implementation, improve compliance, and reduce oversight efforts.

  - *Cost-effective approaches are needed to provide long-term preservation of, and access to, data.* Most successful models for data repositories depend directly or indirectly on Federal funding. The ability of agencies to document, clean, preserve, and curate research data is therefore limited by agency budgets. Implementing long-term preservation, discovery, access, and curation for data is likely to be costlier than for publications and will demand cost-effective approaches to identify valuable data and support its preservation and accessibility.

  - *Standards must be developed and widely adopted to improve the consistency of data collection, description, and preservation.* Data sharing could become much easier with the development of consistent common data elements for data collection, standard formats for representing specific data types, standard metadata to improve discoverability of dat. Standards are also needed to assist with quality assurance and quality control, and protect against misuse of published data (i.e., intentional falsification or fabrication).

  - *Data sharing and usage expertise vary among disciplines.* Disciplines vary in their experience with data sharing, and not all disciplines are yet prepared to manage or share their data effectively. Departments and agencies may need to facilitate resource and tool development, training, interdisciplinary collaboration, and other efforts to foster a culture change and bridge the expertise gap.

  - *Data discoverability remains a challenge.* Even as agencies and non-Federal entities establish data repositories, tools for data discovery are lacking in many scientific disciplines. A number of disciplines have capable tools for data discovery, while others are actively engaged in improving resources in this area.

ED_002389_00007571-00017

III. Opportunities to Advance Open Science

Federal agencies are making significant strides with public access. Despite this, open science continues to offer new opportunities to increase the impact of Federal investments in research. While it is essential that Federal agencies work together, individual agencies need to determine which specific objectives to pursue, taking into account their different missions, operational contexts, available resources, and stakeholders' needs. Additionally, there are opportunities for coordinating efforts among Federal agencies and collaborating with the private sector.

1. **Increase the Availability and Timeliness of Narrative Research Products**
   Recent efforts to provide public access to peer-reviewed publications resulting from federally-funded research, not more than 12 months after the date of publication, are making considerable contributions to the scientific community. Agencies can build upon these efforts by accelerating the timetable for making publications available to the public and by enhancing access to a wider range of written, narrative research products resulting from federally-funded research. Opportunities to consider include:

   - Reducing the embargo period for peer-reviewed publications authored solely by Federal employees to be less than 12 months.
   - Identifying scenarios where publications may warrant shorter embargo periods (e.g., to facilitate responsiveness to public health epidemics or other national emergencies).
   - Identifying opportunities to increase public access to peer-reviewed publications that were published before Federal agency public access policies went into effect.
   - Enhancing access to Federal agencies' technical reports and narrative research products other than peer-reviewed publications.
   - Exploring opportunities for the responsible sharing of preprints to speed research dissemination, recognizing that preprints are not been peer reviewed, that their use varies considerably across scientific disciplines, and that they may need internal agency clearance before being made publicly available.[25]

2. **Increase the Availability and Timeliness of Access to Scientific Data**
   Agencies continue to make progress in implementing policies for data management plans. Many opportunities remain to improve the availability and timeliness of access to scientific data, including:

   - Working together and with non-Federal partners (in the public and private sectors) to develop the infrastructure needed to provide long-term preservation of, discoverability of, and access to scientific data.

[25] Preprints are complete drafts of written research reports in the format of articles which have not yet been peer reviewed, increasingly made publicly available through established preprint repositories. See https://www.nature.com/search?q=preprints and http://search.sciencemag.org/?q=preprints for additional information.

ED_002389_00007571-00018

- Strengthening incentives--and removing disincentives--for effective data management practices to help instill a culture that rewards and recognizes broader use of scientific data. Agency mandates to researchers regarding data management and sharing are a necessary, but not sufficient, condition for success.

- Developing and promulgating methods for mitigating risks associated with privacy, security, and confidentiality when sharing scientific data.

- Assisting researchers in identifying suitable repositories for storing data arising from their work (e.g., by identifying best practices for repositories that store data resulting from federally-funded research).

- Continuing to improve access to Federal data sets that can be used in research (in addition to those resulting from research), such as data from Federal statistical agencies, scientific collections, and monitoring and instrument programs.

- Promoting the release of interim data during national emergencies to spur collaborations and innovative solutions (e.g., releasing interim public health data to advance American health).

3. **Increase Access to Other Research Products**
   Open science efforts to date have focused largely on increasing access to scholarly publications and scientific data. Other research outputs could also be made more accessible, such as:

   - Providing access to research protocols (i.e., documents that describe the methods to be used in specific research studies, such as clinical trials) before studies are completed can reduce unintended duplication of research and increase rigor. Access to protocols after the studies are completed can enable further interpretation of the results, provide contexts for secondary use of the scientific data resulting from the studies, and enable systematic replication.

   - Increasing access to software and code developed through federally-funded research, and/or other outputs that are necessary to use, analyze, or interpret shared data.

4. **Improve the Systems for Making Research Products Public**
   Federal agencies have developed a number of systems to provide access to publications and data resulting from its funded research. By improving those systems, agencies could simplify procedures for researchers to make their publications, data, and other research outputs more accessible, while also allowing users to more easily find, access, and use those materials. Efforts could include:

   - Identifying and implementing technical and administrative approaches to simplify procedures for researchers to provide public access to manuscripts/articles and to scientific data in different systems and under different Federal agency policies.

ED_002389_00007571-00019

- Improving user interfaces and providing documentation that will enable better understanding and use/re-use of research products.

- Using effective approaches to link narrative research products and associated data sets, as well as different data repositories, in order to: facilitate search, access, and citation of data; improve research reproducibility; and assess the success and value of data openness.

- Improving both machine- and human-driven discovery across agency-designated publication and data repositories via application programing interfaces (APIs), federated search tools (e.g., Data.gov, Science.gov), and other innovative approaches.

- Creating metadata more efficiently to enable faster throughput in processing materials and better discovery. Robust metadata are essential for discovery and usability of the product, especially for non-textual materials, such as data sets, software, and artifacts.

- Offering guidance to researchers on how make their research products easier to use through techniques such as documentation standards, repository selection criteria, and use of persistent identifiers.

5. **Lower Barriers to Effective Use of Open Science Content**
Federal agencies can take steps, individually, collectively, and in collaboration with other stakeholders, to make open science content more usable and to unlock its full potential to contribute to economic and societal objectives. Agencies could consider:

- Identifying innovative strategies to fuel open science initiatives, including approaches like crowdsourcing, citizen science, and challenge and prize mechanisms, as well as partnerships with the scientific community, publishers, libraries, and other organizations.

- Supporting community-based efforts to develop standards for data and metadata, adopting community-based standards in Federal systems, and promoting their use by funded researchers.

- Encouraging use of unique persistent identifiers for publications, data, scientific collections, specialized equipment, labs, and facilities, and other resources utilized for research projects to enhance metadata, enable linking, and provide history of studies and analyses.

- Converting narrative research products into machine-readable formats so that computers can help facilitate dissemination of science and increase the impact of federally-funded science.

- Broadening access to peer-reviewed publications through different licensing approaches, investigator incentives, and/or partnership with publishers.

ED_002389_00007571-00020

- Promoting approaches to improve data quality and data documentation (e.g., community standards, open metadata, training, and workforce development).

- Encouraging the private sector to leverage open science content and their downstream application in science and other sectors of the economy. Federal agencies could open opportunities for the private sector to apply new technologies to research outputs and develop value-added services to build on existing Federal resources.

6. **Develop Better Metrics and Measures**
   Agencies are working to improve ways to measure the impacts of Federal public access efforts. Opportunities to enhance these efforts include:

   - Identifying approaches to measure progress in increasing access to narrative research products, scientific data, and other research products.

   - Identifying approaches to enable, encourage, and characterize compliance with agency public access requirements for peer-reviewed publications and data.

   - Supporting research to improve evaluation of open science across Federal agencies and the institutions and infrastructure that they fund, as well as on the research enterprise, entrepreneurship, and the economy.

   - Identifying approaches to use persistent identifiers to enable measurement of the use and impact of research outputs.

IV. Recommendations

Many opportunities exist to advance open science in ways that support the missions of Federal agencies and national R&D priorities. The IWGOS recommends that priority be given to three areas: (1) creating incentives to foster a culture change that recognizes and rewards open science practice; (2) developing interagency best practices, guidance and standards for open science, with a particular focus on data access; and (3) facilitating interagency and public-private collaborations to further realize the benefits of open science. More specific approaches for addressing these priorities are outlined below.

1. **Agencies should help incentivize open science practices.** Open science is a relatively new concept for the Federal government and many of the scientists it funds and supports. Federal agencies should collaborate with other stakeholders to identify and strengthen incentives for practicing open science and making the results of federally-funded scientific research more easily accessible and usable. The IWGOS recommends that Federal agencies:

   - *Identify ways to incentivize culture change for open science among federally-funded researchers (e.g., requiring data management and sharing plans, incentivizing the research community to understand the benefits of open science and receive credit/attribution for open science activities).* Incentivizing a culture change that promotes open science practices may include recognizing data sets as

ED_002389_00007571-00021

important, standalone scientific products and as evidence of scientific excellence; rewarding openness in the evaluation of individual scientists' performance; and identifying metrics to measure the impact of shared data on work by other scientists.

- *Identify ways to incentivize the private sector to develop tools that support open science.* Such tools could improve data discoverability and sharing, facilitate text mining, and provide information infrastructure for data storage and access. The private sector is making tremendous strides in data science and agencies can help focus some of that effort on scientific research data. Incentives like public-private partnerships, prizes, and challenges may help publishers, innovators in machine learning and artificial intelligence, and research funders to contribute toward more effective and open science. These incentives could promote the use of, for example, common APIs to publications, data, and software collections. These interoperable APIs could then be used to create virtual, compatible, aggregated collections amenable to large-scale text and data mining applications.

2. **Agencies should facilitate the interagency development of best practices, guidance and standards on open science, with a particular focus on data access.** Agencies are making substantial progress in advancing open science by enabling public access to both publications and data. There are still numerous opportunities ahead to improve public access to publications and other types of research outputs (e.g., software and protocols), but the primary near- and medium-term challenges lie in sharing data.

   Agencies have many opportunities to work together to increase the benefits and reduce the burden of data preservation and access. Agencies should continue to work together to: 1) refine expectations and guidance for data management and the selection of data repositories; 2) improve and refine approaches to protecting privacy; and 3) improve consistency of policies and practices across Federal funding agencies, philanthropic organizations, and international funders. In addition, agencies should encourage continued improvement in the quality, documentation, versioning, and provenance of shared data. The IWGOS recommends that Federal agencies:

   - *Accelerate efforts to create joint solutions for common data sharing challenges.* By connecting agencies working on same or similar research areas or disciplines, seeking consistent approaches to the development of data repositories, discovery tools, and compliance monitoring public access becomes more uniform and thereby easier to navigate and use by both investigators and the end user.

   - *Coordinate their efforts to provide guidance to researchers for selecting suitable repositories for federally-funded research data.*

   - *Recognize the importance of data standards and coordinate efforts to work with relevant communities to identify, develop, promulgate, and encourage the use of standards for data and metadata that facilitate data connectivity, interoperability, and reuse across agencies.*

ED_002389_00007571-00022

- *Improve efforts to make digital data more discoverable and accessible via the use of persistent identifiers for research objects, researchers, and research organizations.* Such identifiers are key to interlinking related research objects and to supporting open science and reproducibility.

3. **Agencies should facilitate interagency and public-private collaborations to further realize the benefits of open science.**

- *Federal agencies should continue to formally meet to discuss open science.* To bolster efforts toward open science, Federal agencies must address issues of research culture, incentives, infrastructure, policy, and more. IWGOS agencies are addressing these challenges through their open science policies, while pursuing common objectives tailored to their specific missions and research communities. Some challenges are common to all disciplines and shared approaches can reduce training and administrative burden. IWGOS recommends Federal agencies continue to meet regularly to address issues such as policy harmonization, burden reduction, implementation, measurement, infrastructure, and emerging opportunities to promote open science. IWGOS also recommends stronger collaboration with other Federal working groups addressing related topics. These include working groups that coordinate Federal-wide research initiatives in specific domains of science (e.g., materials science, microbiome research, and arctic research) and working groups on cross-disciplinary topics such as open data, data science, scientific collections, and technology transfer.

- *Engage all stakeholders to fully realize economic and social benefits of open science.* Successful implementation of open science policies and practices will require close collaboration with stakeholders outside the Federal government. Important partners include the academic and non-profit sectors, businesses and entrepreneurs, and international partners.

  - *Public, nonprofit, and academic partners.* Federal agencies should work closely with the academic community to ensure that open science policies and practices are understood by the research community and practical to implement. Universities also have ambitions related to open science and are developing supporting infrastructure and policies. Such efforts should be coordinated as much as possible with Federal efforts. Federal agencies will need to partner with universities, scientific societies, and other stakeholders in the scientific community to help establish incentives for open science that recognize and reward the practice of open science.

  - *Businesses and entrepreneurs.* Realizing the full economic and societal potential of open science requires participation beyond the scientific research community. Other sectors of the economy have made great strides in pushing knowledge to targeted audiences, data mining, and collaboration and networking tools. Commercial investment in these areas often dwarfs Federal investment. Federal science agencies could enhance their impact by engaging knowledge management solutions, and their associated private investment, to focus on

ED_002389_00007571-00023

research products. Agencies could also work on the standards, infrastructure, and other technical details to ensure federally-funded open science products are attractive resources for the private sector.

- *International partners.* Because scientific communities span national borders, international engagement can advance open science. Open science in other countries can help U.S.-based researchers gain better access to global knowledge. Many countries are seeking innovative solutions for data preservation and access, incentive systems that reward open science practices, and approaches for developing needed skills in the scientific workforce. Federal agencies can promote the international adoption of open science practices that support American interests. Federal agencies should continue to participate in international fora to help identify good practices, standards, and opportunities to align initiatives around American policies, and leverage the solutions (e.g., technical, operational) that emerge from these efforts.

Steps such as those outlined above can help achieve the promise open science offers to accelerate research, development, and innovation for American prosperity.

ED_002389_00007571-00024