**National Institutes of Health
Plan for Increasing Access to Scientific Publications and
Digital Scientific Data from NIH Funded Scientific Research**

February 2015

This document describes NIH's plans to build upon and enhance its longstanding efforts to increase access to scholarly publications and digital data resulting from NIH-funded research.

TABLE OF CONTENTS

Executive Summary

On February 22, 2013, the White House Office of Science and Technology Policy (OSTP) released its memorandum entitled "Increasing Access to the Results of Federally Funded Scientific Research."[1] The memorandum directs federal agencies and offices to develop and submit plans to OSTP that ensure peer-reviewed publications and digital scientific data resulting from federally-funded scientific research are accessible to the public, the scientific community, and industry to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland, and economic security; and, the specific objectives of the memorandum. In response, this plan describes the steps the National Institutes of Health (NIH) will take to meet those directives for scientific publications and digital scientific data. The main elements of the NIH plan are summarized immediately below, and then described in greater detail beginning in Section I of this plan.

Scientific Publications

The NIH Public Access Policy[2] for publications has been in a requirement for all recipients of NIH funds since 2008. It implements Division G, Title II, Section 218 of PL 110-161 (Consolidated Appropriations Act, 2008). The law states:

> *The Director of the National Institutes of Health shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication:* Provided*, That the NIH shall implement the public access policy in a manner consistent with copyright law.*

The NIH Public Access Policy ensures that the public has access to the published results of NIH-funded research. It requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive PubMed Central (PMC)[3] upon acceptance for publication. Scientists can also deposit papers through partnerships NIH has established with publishers. To help advance science and improve human health, the Policy requires that NIH supported papers are accessible to the public on PMC no later than 12 months after publication.

PMC is a public-private partnership to preserve and make public full-text journal articles. It was established in 2000 and is operated by the National Library of Medicine (NLM) of the NIH. The bulk of its papers are voluntarily contributed by publishers. Thousands of

---

[1] http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
[2] http://publicaccess.nih.gov/policy.htm
[3] http://www.ncbi.nlm.nih.gov/pmc/

journals voluntarily submit peer-reviewed author manuscripts to PMC to assist authors in complying with the public access process.[4] Several hundred journal publishers also voluntarily automatically deposit final published versions of articles on behalf of their authors, relieving them of the need to directly submit their manuscripts. Finally, publishers representing about 1,200 journals voluntarily submit the full content of their journals to PMC, regardless of whether the issue contains an article subject to the NIH Public Access Policy.[5]

To the extent allowed by law, this Policy meets all the requirements of the OSTP directive. NIH implemented its Policy in 2008 to:

- Create a stable archive of peer-reviewed research publications resulting from NIH-funded research to ensure the permanent preservation of these vital published research findings;
- Secure a searchable compendium of these peer-reviewed research publications that NIH and its awardees can use to manage more efficiently and to understand better their research portfolios, monitor scientific productivity, and ultimately, help set research priorities;
- Make published results of NIH-funded research more readily accessible to the public, health care providers, educators, and scientists.[6] NIH views this policy as a means to maximize the impact and accountability of the Federal research investment.

The significant usage of PMC underscores the value of public access to the peer-reviewed scientific literature. On a typical weekday, more than 1 million unduplicated users retrieve over 1.65 million articles. The number of articles retrieved has doubled in the past three years, from 17 million retrievals per month in 2009 to 35 million in 2012.

Digital Scientific Data

NIH intends to make public access to digital scientific data the standard for all NIH-funded research. Following adoption of the final plan, NIH will:

- **Explore steps to require data sharing.** NIH will explore the development of policies to require NIH-funded researchers to make the data underlying the conclusions of peer-reviewed scientific research publications freely available in public repositories at the time of initial publication in machine readable formats. NIH will ensure that data management plans include clear plans for sharing research data.

---

[4] These publishers are listed under Method D at http://publicaccess.nih.gov/select_deposit_publishers.htm.
[5] For a full list, see http://publicaccess.nih.gov/submit_process_journals.htm.
[6] See http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html.

- **Ensure that all NIH-funded researchers prepare data management plans and that the plans are evaluated during peer review.** NIH will ensure the development of data management plans for all research activities supported by grants, cooperative agreements, contracts, or intramural funds. The data management plans will include descriptions of items such as the data to be produced in the proposed study, any data standards used, mechanisms for providing access to and sharing of data (including provisions for protection of privacy, confidentiality, security, intellectual property, or other rights), provisions for data reuse and redistribution, and plans for archiving and long-term preservation of the data, as appropriate. The data management plans for all extramural research will be appropriately evaluated during peer review of applications or proposals, and data management plans for all intramural research will be reviewed by the senior official responsible for scientific leadership within each NIH Institute or Center, i.e., the Scientific Director, or his or her designee.

- **Develop additional data management policies to increase public access to designated types of biomedical research data.** NIH will continue to work in consultation with the biomedical research community to identify areas of research for which more specific data management and sharing policies should be developed, as has been done with genomic data, autism research, and other areas of science. In these areas, NIH may develop more specific policies for data management, stipulating, for example, the types of data that should be preserved and made accessible to others; provisions such as the protection of privacy, confidentiality, security, intellectual property, and other rights, as appropriate; designated repositories for archiving such data; timelines for submitting study data and for making it accessible to other researchers; common data elements (CDEs) that should be used to collect the data; and required data formats, consistent with other applicable requirements. The data management plans will be the foundation for more specific policies for data management.

- **Encourage the use of established public repositories and community-based standards.** NIH will encourage supported researchers to deposit data in established public repositories, where applicable, for archiving and preservation. In some cases, NIH data management policies may specify particular standards and repositories to be used by funded researchers (see bullet above); however, in other cases, appropriate, relevant standards and repositories may not yet exist. In addition, given their importance in the subsequent use of data, NIH will encourage all funded researchers to make use of existing data standards relevant to their research community, such as standards for collecting and representing data and information describing the data set (i.e., metadata). NIH will also promote the interoperability of digital data in public repositories. Where needed, NIH will take steps to support the development of selected community-based data repositories and standards.

- **Develop approaches to ensure the discoverability of data sets resulting from NIH-funded research to make them findable, accessible, and citable.** NIH will

explore innovative approaches to improving the discoverability (i.e., the ability to readily locate) of data sets resulting from NIH-funded research. NIH is funding development of a data discovery index to provide a mechanism to enhance discoverability and facilitate appropriate attribution to those responsible for the data set and link the citations to associated publications. Moreover, the NIH will explore ways to advance data as a legitimate form of scholarship through data citation and other means.

- **Promote interoperability and openness of digital scientific data generated or managed by NIH.** Scientific digital data generated by contracts, intramural research, or deposited in an NIH data repository for use and distribution will be required to meet certain standards that support downstream information processing and dissemination. NIH will ensure that new policies for such data, as appropriate, are consistent with overarching White House requirements in its Open Data Policy Memorandum M-13-13. New NIH information systems for providing access to digital scientific data will also take into account requirements for discoverability, interoperability and accessibility.

- **Explore the development of a data commons.** NIH will explore the development of a commons, a shared space for basic and clinical research output including data, software, and narrative, that follows the FAIR principles of Find, Access, Interoperate and Reuse. A particular focus of the effort will be on making the data underlying the conclusions of peer-reviewed scientific publications resulting from federally funded scientific research available for free at the time of publication. In this effort, NIH will welcome opportunities to collaborate with other Departments and Agencies.

Section I: Background and Purpose

On February 22, 2013, the White House Office of Science and Technology Policy (OSTP) released its memorandum entitled "Increasing Access to the Results of Federally Funded Scientific Research."[7] The memorandum directs federal agencies and offices to develop and submit plans to OSTP that ensure peer-reviewed publications and digital scientific data resulting from federally-funded scientific research are accessible to the public, the scientific community, and industry, to the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland, and economic security; and the specific objectives of the memorandum. In response, this plan describes the steps the National Institutes of Health (NIH) will take to meet those directives.

NIH's mission is to seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability. In addition to funding biomedical and behavioral research into the causes, diagnosis, prevention, and cure of human diseases and the training of basic and clinical researchers capable of carrying out such research, NIH also is responsible for expanding the knowledge base in basic, medical, and associated sciences and ensuring a continued high return on the public investment in research.

The goals of the OSTP memorandum are in keeping with NIH's mandate[8] and continued commitment to ensure that, to the fullest extent possible, the results of federally-funded scientific research are made available to and are useful for the general public, industry, and the scientific community. Validation and progress in science are predicated on access to research results, and, to that end, NIH has developed a number of policies to support this effort.[9] Commensurate with the Administration's goal of increasing public access to scientific publications and digital data, this plan outlines current NIH policies, programs, and procedures that support the overall goals set out in the OSTP memorandum, and it identifies further steps that may be taken to fulfill the goals in a more comprehensive manner. New policies, programs, or procedures implemented by the final NIH plan will apply to manuscripts submitted for publication and digital data generated after the effective date of the final plan. In developing this plan, NIH considered the perspectives of advisors and stakeholders, including federally funded researchers, universities, libraries, publishers, users of federally funded research results, and civil and scientific society groups.[10] As the planning process

---

[7] http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
[8] See, for example, the Consolidated Appropriations Act, 2008, P.L. 110-161, division G, section 218.
[9] http://grants.nih.gov/grants/sharing.htm
[10] Including the report and recommendations on data and informatics of the Advisory Committee to the Director, NIH, and the perspectives of participants of a May 2013 meeting at the National Academy of Sciences to gather input on the Administration's goal of increasing public access to federally-funded research and development data and publications.

continues, NIH intends to continue soliciting and accounting for stakeholders' views, as appropriate.

Section II: Scientific Publications

1. Introduction

The NIH Public Access Policy for publications, hereafter referred to as the "Public Access Policy" or "the Policy," implements Division G, Title II, Section 218 of PL 110-161 (Consolidated Appropriations Act, 2008). The law states:

> *The Director of the National Institutes of Health shall require that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's PubMed Central an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made publicly available no later than 12 months after the official date of publication: Provided, That the NIH shall implement the Public Access Policy in a manner consistent with copyright law.*

The NIH Public Access Policy[11] ensures that the public has access to the published results of NIH-funded research. It requires scientists to submit final peer-reviewed journal manuscripts that arise from NIH funds to the digital archive PubMed Central (PMC)[12] upon acceptance for publication. Scientists can also deposit papers through partnerships NIH has established with publishers. To help advance science and improve human health, the Policy requires that NIH supported papers are accessible to the public on PMC no later than 12 months after publication.

The Policy applies to any manuscript that:

- Is peer-reviewed;
- Is accepted for publication in a journal on or after April 7, 2008; and
- Arises from:
    o Any direct funding from an NIH grant or cooperative agreement active in Fiscal Year 2008 or beyond, or;
    o Any direct funding[13] from an NIH contract signed on or after April 7, 2008, or;
    o Any direct funding from the NIH Intramural Program, or;
    o An NIH employee.

To the extent allowed by law, agency mission, resource constraints, U.S. national, homeland, and economic security, this Policy meets all the requirements of the OSTP directive. NIH implemented its Policy in 2008 to:

---

[11] http://publicaccess.nih.gov/
[12] http://www.ncbi.nlm.nih.gov/pmc/
[13] "Directly" funded means costs that can be specifically identified with a particular project or activity.

- Create a stable archive of peer-reviewed research publications resulting from NIH-funded research to ensure the permanent preservation of these vital published research findings;
- Secure a searchable compendium of these peer-reviewed research publications that NIH and its awardees can use to manage more efficiently and to understand better their research portfolios, monitor scientific productivity, and ultimately, help set research priorities;
- Make published results of NIH-funded research more readily accessible to the public, health care providers, educators, and scientists.[14]

NIH views this Policy as a means to maximize the impact and accountability of the federal research investment (OSTP element 3).

2. Public Comment, Implementation, and Notification

A plan for notifying awardees and other federally funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, and/or regulatory changes); Identification of resources within the existing agency budget to implement the plan; A timeline for implementation. (OSTP memo, elements 2d, 2f, 2g)

The NIH Public Access Policy was initially issued in 2005, after an extensive comment process.[15] The NIH Public Access Policy became a legislative requirement in 2008, and it was subject to public comment that year.[16] It was announced to the extramural community through notices,[17] and policy guidance for NIH employees is posted on the NIH website.[18] Since its implementation, the Policy has been fully operational and funded at approximately $4.0 to 4.5 million per year.[19]

3. Public-Private Partnerships

A strategy for leveraging existing archives, where appropriate, and fostering public-private partnerships with scientific journals relevant to the agency's research; Encourage public-private collaboration; Encourage public-private collaboration to … otherwise assist with implementation of the agency plan; Ensure that publications and metadata are stored in an archival solution that… uses standards, widely available and, to the extent possible, nonproprietary archival formats for text and associated content (e.g., images, video, supporting data). (OSTP memo, elements 2a, 3d, 3d.iv, 3f.ii)

---

[14] http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html
[15] 2005 public comments summarized in http://grants.nih.gov/grants/guide/notice-files/NOT-OD-05-022.html/.
[16] 2008 public comments available at http://publicaccess.nih.gov/comments.htm.
[17] Guide Notice at http://publicaccess.nih.gov/policy.htm.
[18] See http://publicaccess.nih.gov/nih_employee_procedures.htm.
[19] Current information about the policy is available at http://publicaccess.nih.gov.

The NIH Policy is a true public-private partnership in that it codifies the principle that the primary means to disseminate NIH-funded research findings are through privately peer-reviewed and published journals, and not government reports.

Further, the Policy's host archive, PMC, is a public-private partnership to preserve and make public full-text journal articles in a nonproprietary, widely-distributed archival XML format.[20] PMC also archives and makes public any supplementary data associated with journal articles (e.g., images, tables, video, or other documents / files).[21] It was established in 2000 and is operated by the National Library of Medicine (NLM) of the NIH. The bulk of its papers are voluntarily contributed by publishers. Roughly 1,600 journals voluntarily submit the complete content of their journals to PMC, regardless of whether the issue contains an article subject to the NIH Public Access Policy.[22] Many other journals submit NIH-supported articles.

The NIH Public Access Policy also collects papers through public-private partnerships. NIH has established methods in which publishers can partner with NIH and authors to support the Policy and deposit materials in PMC.

There are four methods to ensure that an applicable paper is submitted to PMC in compliance with the NIH Public Access Policy. Authors may use whichever method is most appropriate for them and consistent with their publishing agreement.[23]

Submission Methods A and B

Methods A and B are where the publisher agrees to post the final published article directly to PMC (in XML) around the time of publication. Approximately 1,900 journals have signed agreements with NIH to post content routinely, in compliance with the Policy.[24] Further, more than 20 publishers, representing thousands of journals, have signed agreements with NIH to post the final published article on request of the author (usually for a fee, reimbursable from the supporting NIH grant).[25]

Submission Methods C and D

Methods C and D require final peer-reviewed manuscripts to be deposited to the NIH Manuscript Submission System (NIHMS) upon acceptance for publication. The NIHMS

---

[20] The specific XML format used is the Journal Article Tag Suite, available at http://jats.nlm.nih.gov/. This suite is a National Information Standards Organization standard, widely used in the publishing industry. See also section 5.c.

[21] PMC ensures the supplementary material is archived, public, and to the extent possible, available in non-proprietary formats. See http://www.ncbi.nlm.nih.gov/pmc/pub/filespec/#supp for more details.

[22] See the "Full" participation level journals listed at https://www.ncbi.nlm.nih.gov/pmc/journals/.

[23] More information about the submission process is available at http://publicaccess.nih.gov/submit_process.htm/.

[24] List available at http://publicaccess.nih.gov/submit_process_journals.htm

[25] See the list under Method B at http://publicaccess.nih.gov/select_deposit_publishers.htm.

takes these manuscripts and converts them from their original format to PMC's XML standard. Under Method C, authors initiate the submission process by depositing a final peer-reviewed manuscript. Under Method D, the publisher deposits the manuscript for the author. Method D publishers include most of the major commercial publishers and represent thousands of journals.[26]

4. Making Papers Public

a. Embargo

Identification of any special circumstances that prevent the agency from meeting any of the objectives set out in this memorandum, in whole or in part; Each agency plan…shall use a twelve-month post-publication embargo period as a guideline for making research papers publicly available; however, an agency may tailor its plan as necessary to address the objectives articulated in this memorandum, as well as the challenges and public interests that are unique to each field and mission combination; Provide a mechanism for stakeholders to petition for changing the embargo period for a specific field by presenting evidence demonstrating that the plan would be inconsistent with the objectives articulated in this memorandum; (OSTP memo, elements 2h, 3a.i, 3a.ii)

Per the Consolidated Appropriations Act of 2008, NIH must make papers public on PMC no later than 12 months after the official date of publication. This statute does not authorize NIH to embargo materials longer than 12 months. The rights holder (author or publisher) has the option to set a shorter embargo.

NIH will rely on the Health and Human Services (HHS) petition process for considering requests to shorten the embargo period for publications in a specific field. This process is described in greater detail in the *HHS Guiding Principles and Approach for Enhancing Public Access*.

b. Public use

Ensure that the public can read, download, and analyze in digital form final peer-reviewed manuscripts or final published documents within a timeframe that is appropriate for each type of research conducted or sponsored by the agency. (OSTP memo, element 3a)

i. License

Awardees must ensure that any publishing agreement allows the paper to be posted to PMC in accordance with the NIH Public Access Policy. NIH does not dictate the means

---

[26] See the list under Method d at http://publicaccess.nih.gov/select_deposit_publishers.htm

by which awardees must do so, but does offer guidance in an FAQ[27] on suggested wording for publishing agreements.

NIH is also exploring the possibility of using the government use license specified in 45 CFR 74.36 to help make papers public. Under these terms, the government has 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.

NIH employees must use a publishing agreement cover sheet,[28] which ensures the paper can be posted to PMC in accordance with the policy.[29]

ii. **Terms of use**

Each agency plan…shall use a twelve-month post-publication embargo period as a guideline for making research papers publicly available; however, an agency may tailor its plan as necessary to address the objectives articulated in this memorandum, as well as the challenges and public interests that are unique to each field and mission combination; Ensure that publications and metadata are stored in an archival solution that…provides for long-term preservation and access to the content without charge. (OSTP memo, elements 3a.i, 3f.i)

All of the material available from the PMC site is provided by the respective publishers or authors. Almost all of it is protected by U.S. and/or foreign copyright laws, even though PMC provides free access to it (Public domain material[30] is an exception).

All PMC content is available to anyone to "read, download and analyze in digital form" without charge. Final peer-reviewed manuscripts collected under the Public Access Policy are made public under fair use principles. Publishers submitting final published articles can offer access beyond fair use. The copyright holders retain rights for reproduction, redistribution, and reuse.

Users of PMC are directly and solely responsible for compliance with copyright restrictions and are expected to adhere to the terms and conditions defined by the copyright holder. Transmission, reproduction, or reuse of protected material, beyond that allowed by the fair use principles of the copyright laws, requires the written permission of the copyright owners. U.S. fair use guidelines are available from the U.S. Copyright Office at the Library of Congress.[31,32]

---

[27] http://publicaccess.nih.gov/FAQ.htm#778
[28] Copies available at http://www1.od.nih.gov/oir/sourcebook/oversight/NIHCover%20Sheet.pdf
[29] For more detail on employee procedures, see http://publicaccess.nih.gov/nih_employee_procedures.htm.
[30] http://www.ncbi.nlm.nih.gov/pmc/about/copyright/#public-domain
[31] For fair use guidelines, see http://www.copyright.gov/fls/fl102.html
[32] For more information, see http://www.ncbi.nlm.nih.gov/pmc/about/copyright/.

PMC provides access to NIH-supported papers and over two million other papers, while taking appropriate steps to prevent unauthorized mass redistribution. NIH has established sophisticated monitoring systems to detect and prevent potential misuse that are very similar to the processes used by publishers websites. NIH systems detect and prevent bulk downloading and will immediately cut off any sites, foreign or domestic, that appear to be abusing copyrighted property. Unless publishers have chosen PMC as the sole distributor of the electronic contents of their journals, all the articles that are accessible on PMC are also accessible on publisher websites.

NLM also provides extensive scripting tools that facilitate search and some forms of analysis without using their own computing resources. Please see section II.5b for more details.

iii. Providing bulk downloads for research, and managing the restriction on unauthorized bulk downloads

… an agency may tailor its plan as necessary to address the objectives articulated in this memorandum, as well as the challenges and public interests that are unique to each field and mission combination (OSTP memo, element 3a.i)

PMC has two services that may be used for automated retrieval and downloading of a subset of articles and all the metadata from the PMC archive. The PMC Open Archives Initiative (OAI) service and the PMC File Transfer Protocol (FTP) service are the only services that may be used for automated downloading of articles in PMC. See the PMC Open Access Subset[33] for information about which articles are included in this special subset, and for links to the PMC OAI and FTP services.

The FTP service facilitates bulk download of papers and all PMC metadata. Final peer-reviewed manuscripts submitted by authors are not currently available through FTP. Publishers can set licenses allowing bulk download when they send final published articles directly to PMC. Articles that are available through the PMC OAI and FTP services are still protected by copyright but are distributed under a Creative Commons or similar license that generally allows more liberal use than a traditional copyrighted work.

The PMC web service[34] allows users to discover downloadable resources from the PMC Open Access Subset. It provides a query interface that can be used to find source files (either PDFs or zipped archive files) that meet specified criteria. For example, it could be used to find the PDFs of all articles that have been updated since a specified

---

[33] http://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/

[34] For details about the API and how to use it, see the OA Service documentation page at (http://www.ncbi.nlm.nih.gov/pmc/tools/oa-service/).

date. This could facilitate implementing tools that reuse the Open Access (OA) subset content, such as mirror sites, text mining processes.

5. Access and Discoverability

a. Metadata Management

Ensure full public access to publications' metadata without charge upon first publication in a data format that ensures interoperability with current and future search technology. Where possible, the metadata should provide a link to the location where the full text and associated supplemental materials will be made available after the embargo period; Ensure that attribution to authors, journals, and original publishers is maintained (OSTP memo, elements 3c, 3e)

PMC makes the abstract and citation information available to NIH's abstract and metadata service, PubMed, where it can be accessed without charge. PubMed makes metadata available upon first publication and combines them with any indexing terms or key words (for example Medical Subject Headings assigned by the Medline indexing service for certain journals).

For articles subject to the NIH Public Access Policy, PubMed abstracts include a link to the PMC version and to the version of the paper on the publisher site, if provided. PubMed metadata, like the full text of PMC, are permanently archived. Full text displays of papers on PMC always include publisher-supplied links to the publisher site, authors, journal name and the other information required for a complete academic citation.

b. Search

A strategy for improving the public's ability to locate and access digital data resulting from federally funded scientific research; Facilitate easy public search, analysis of, and access to peer-reviewed scholarly publications directly arising from research funded by the Federal Government; (OSTP memo, elements 2b, 3b)

NLM has set the standard for searching the biomedical literature. The full text of PMC papers can be searched on PMC. In addition, the metadata (abstract, citation and any index terms) can be searched in PubMed.

Both databases support search functions beyond basic look up in a variety of ways:

- links[35] to publisher sites, when links are provided by publishers;
- support for Digital Object Identifiers, when provided by publishers;

---

[35] http://www.ncbi.nlm.nih.gov/projects/linkout/

- a scripting language (E-Utilities, described below);
- the bibliography management service My NCBI, which offers RSS feeds, HTML code, and other methods of bibliography and search sharing, among other features;
- links to cited papers, related genes, chemicals and other elements in other NLM databases (the Entrez[36] system);
- thumbnail sketches of figures in the abstract view;
- highlights similar papers and review articles; some of the highlighted papers can be context sensitive (e.g., methodological papers are highlighted in the margins when displaying the methods section of a paper on PMC).
- PubMed Commons for user generated commentary about individual papers.

Modern search functions do more than help find the specific items; they provide a context. The NCBI system, of which PMC is a part, includes more than 50 databases covering a variety of biomedical data, including nucleotide and protein sequences, gene records, three-dimensional molecular structures, and the biomedical literature. It provides researchers with the capability to integrate their individual discoveries with other publications and scientific data that lead to new areas for exploration.

The NCBI family of databases provide context with a broad array of scientific data that are interconnected through XML and a common archival framework. PMC links to related papers, as well as to papers that were actually cited. It also links to related chemical structures, proteins, viruses, and other data. This level of context is not possible unless papers are housed on PMC.

NCBI databases expose data through the Entrez interface. E-Utilities allow anyone to script searches, or even embed fixed searches (e.g., all papers on diabetes published in the last 7 days) or general searches within their own webpages and applications. The details of using the E-utilities API are documented in the Entrez Programming Utilities Help book.[37]

As with all Entrez databases, PMC defines a set of search fields, filters, and links to enable discovery of information. Search fields enable targeted full-text searches, filters enable finding PMC articles based on specific criteria, and links enable finding records in other Entrez databases that relate in some way to PMC articles (or vice-versa). A list of search fields is provided in the PMC Help Manual.[38]

Finally, since PMC also contains award information, papers can be searched by grant number in PubMed and are discoverable by exploring awards through NIH's Research Portfolio Online Reporting Tools.[39]

---

[36] http://www.ncbi.nlm.nih.gov/About/tools/restable_mol.html
[37] http://www.ncbi.nlm.nih.gov/books/NBK25501/
[38] http://www.ncbi.nlm.nih.gov/books/NBK3825/#pmchelp.Search_field_descriptions_and_ta
[39] http://report.nih.gov/index.aspx

c. **Interoperability and integration into other systems**

A strategy for leveraging existing archives, where appropriate, and fostering public-private partnerships with scientific journals relevant to the agency's research; Encourage public-private collaboration to maximize the potential for interoperability between public and private platforms and creative reuse to enhance value to all stakeholders; Encourage public-private collaboration to avoid unnecessary duplication of existing mechanism; Ensure that publications and metadata are stored in an archival solution that enables integration and interoperability with other Federal public access archival solutions and other appropriate archives. (OSTP memo, elements 2a, 3d.i, 3d.ii, 3f.iv)

NIH has already integrated PMC with a variety of scientific resources, literature archives supported by other research funders, and grants management and accountability systems. NIH stands ready to work with other federal agencies and Health and Human Services (HHS) operating divisions in their efforts to make federally supported research publicly accessible.

Journal Article Tag Suite (JATS)

PMC's XML standard, the Journal Article Tag Suite (JATS), is an American National Standards Institute (ANSI)/National Information Standards Organization (NISO) standard. As stated in the press release from NISO,[40] "JATS provides a common XML format in which publishers and archives can exchange journal content by preserving the intellectual content of journals independent of the form in which that content was originally delivered."

Exposure to third party services

PMC is currently crawled by numerous search services, such as Google and Bing, to make content more discoverable.

PMC International

PubMed Central International (PMCI) is a collaborative effort between NIH and NLM, the publishers whose journal content makes up the PMC archive, and organizations in other countries that share NIH's and NLM's interest in archiving life sciences literature. This effort allows for integration of content and reciprocity in public access policies across funders.

---

[40] http://www.niso.org/news/pr/view?item_key=d92a2bc93b43db6831e68914e134c731d83cbdd1

Over the years, PMC and its partner archives have become host archives for dozens of research funders, and have developed arrangements to mirror content. This means that if a funder requires a paper to be posted to PMC or its international partners within 12 months of publication, compliance with the NIH policy can also count as compliance with other funder policies.

PMC is the host archive for several public access policies, including NIH, the Howard Hughes Medical Institution (HHMI), and the Autism Speaks foundation. PMC Europe is a host archive for a score of funders.[41] PMC Canada[42] is the host archive for the Canadian Institutes of Health Research.

The long-term goal of PMCI is to create a network of digital archives that can share all of their respective locally deposited content with others in the network. There are three primary reasons for doing this:

- The probability of an archive surviving over the long term is greater if there are working copies of the archive in regular use at multiple sites around the world.
- A producer or funder of research literature often will be more inclined to make the primary deposit of its material to a locally or regionally affiliated archive, rather than to one operated elsewhere in the world.
- Each site can integrate the journal articles in the archive with related material, such as national or regional practice guidelines, that has particular significance to its users.

Interoperability with grants management systems

NIH also maintains the grants management system eRA Commons,[43] used by several federal agencies and HHS operating divisions.[44] NIH has integrated eRA Commons with public access compliance information using My NCBI.[45]

Other funders could use eRA Commons for their own public access policies. Alternatively, agencies and HHS operating divisions could establish similar oversight and reporting mechanisms for different funding systems. They would need to authenticate users for My NCBI and the NIHMS, as well as provide NIH systems with a data feed of funding codes.

---

[41] For a list of funders, see http://europepmc.org/Funders/
[42] http://pubmedcentralcanada.ca/
[43] http://era.nih.gov/about_era/index.cfm
[44] For a current list, see http://era.nih.gov/about_era/overview_era_partnerships.cfm
[45] See Managing Compliance with the NIH Public Access Policy: video overview (http://www.youtube.com/watch?v=JYODIOD_YYE) and instructions (http://www.ncbi.nlm.nih.gov/books/NBK53595/#mybibliography.Managing_Compliance_to_th) and Using My Bibliography to generate the publications section of form PHS 2590 (http://www.nlm.nih.gov/pubs/techbull/nd12/nd12_myncbi_pdf.html).

E-Utilities

Finally, the E-Utilities scripting language (described under II.5b) provides another way to access and integrate PMC content with other services. For example, it can be used to create custom search interfaces within mobile apps and webpages.

6. Preservation

Ensure that publications and metadata are stored in an archival solution that: provides for long-term preservation and access to the content without charge; uses standards, widely available and, to the extent possible, nonproprietary archival formats for text and associated content (e.g., images, video, supporting data); provides access for persons with disabilities consistent with Section 508 of the Rehabilitation Act of 197. (OSTP memo, elements 3f, 3f.i, 3f.ii, 3f.iii)

a. **Preservation methods**

Preservation is one of the Public Access Policy's primary objectives. NLM maintains multiple backup copies of the PMC database, both onsite and at a remote, secure location. In addition, most of the content in PMC, including NIH-funded author manuscripts, are redistributed to PMC Canada and PMC Europe. They, in turn, share their content with PMC.

Further, content on PMC is actively curated in that all article views are generated dynamically from the XML record of an article. In this way, every use of an article validates that the archival record is still usable and viable.

Finally, PMC is future proof, in that XML is technology independent and can be easily and reliably migrated as technology evolves. For example, PMC has introduced new display features to better support tablets and other devices for old content as well as new.

b. **508 compliance**

PMC uses the common standard XML to ensure content is 508 compliant. It dynamically renders XML as PDF, HTML, and in a PubReader format.[46]

7. Metrics, Compliance and Impacts

An agency strategy for measuring and, as necessary, enforcing compliance with its plan (OSTP memo, element 2e)

---

[46] http://www.ncbi.nlm.nih.gov/pmc/about/pubreader/

a. **Oversight methods**

NIH staff manually checks applications, proposals, or reports for compliance with the Public Access Policy. They confirm that a citation includes the appropriate reference number that indicates compliance, such as the PubMed Central Identifier (PMCID).[47]

NIH has begun to automate this reporting process by integrating the NLM's Bibliography tool, My NCBI[48], with our electronic grants management system. My NCBI tracks the public access compliance status of every paper associated with an award, generates a citation with the current identifier and clearly indicates its compliance status. It also allows awardees to collaborate[49] with their colleagues to associate publications with NIH awards.

In November 2012, NIH announced[50] that it will delay processing of all non-competing continuation awards (commonly called progress reports) with a start date of July 1, 2013, and beyond that are not compliant with the Public Access Policy. To simplify reporting and tracking, awardees now use My NCBI to report public access compliance on these awards. This process is fully electronic for our Research Performance Progress Reports (RPPR). My NCBI supports paper based progress reports by generating PDF reports of the publication section of paper forms.

b. **Measures**

Papers are compliant when they are posted to PMC. This forms the numerator of our NIH-wide compliance measure.

Our denominator for overall compliance is determined by the number of papers NIH estimates that fall under the Policy. We generate this number from papers linked to specific NIH awards by:

- Authors in the NIH Manuscript Submission system,
- Authors and principal investigators in the My NCBI bibliography management system,
- Principal investigators and institutions in electronic progress reports, and
- Authors and publishers in the acknowledgment section of the text of papers, as indexed by Medline.

---

[47] See http://publicaccess.nih.gov/citation_methods.htm for details.
[48] http://www.ncbi.nlm.nih.gov/myncbi/
[49] See Collaborating with PIs and Co-Authors to associate papers with NIH grants, monitor compliance, and simplify reporting at http://www.nlm.nih.gov/pubs/techbull/ja12/ja12_myncbi_new_features.html.
[50] http://grants.nih.gov/grants/guide/notice-files/NOT-OD-12-160.html

Combined, these papers serve as the denominator of the Public Access Policy. As of May 2, 2014, NIH funds resulted in approximately 512,000 papers published between July 2008 and October 2013, and 424,000 (83 percent) of these papers have been posted to PMC. As of July 1, 2013, NIH delays processing of all non-competing continuation awards (annual progress reports) that are not in compliance with the Public Access Policy.

c. **Policy Impacts**

i. **Benefits**

An approach for optimizing search, archival, and dissemination features that encourage innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research; Encourage public-private collaboration to maximize the impact of the Federal research investment (OSTP memo, elements 2c, 3d.iii)

The significant usage of PMC underscores the value of public access to the peer-reviewed scientific literature. On a typical weekday, more than 1 million unduplicated users retrieve over 1.65 million articles. The number of articles retrieved has more than doubled in the past four years, from 17 million retrievals per month in 2009 to more than 40 million in 2013.

When the literature is accessible, the NIH investment is more effective. For example, better access to medical research makes our health care system more efficient. More than six out of 10 physicians report changing an initial diagnosis based on new information accessed via online resources/support tools. Nearly nine in 10 physicians report that improved access to online medical information and resources has improved the quality of care at their practice.[51]

In addition, we have found that the NLM databases, with their XML data accessible through a variety of web services, can support additional economic activity. NLM has found more than 300 private sector firms obtain data from Medline for a variety of applications: to supplement literature database products and search tools; to build literature analysis tools for end users of research; to conduct research in a variety biomedical areas; and to gain intelligence on experts and trends. Many others that we do not have the means to track use NLM's other resources.

ii. **Publisher Impacts**

---

[51] www.wolterskluwerhealth.com/News/Documents/White Papers/Wolters Kluwer Health Survey Executive Summary-Media.pdf

The NIH Public Access Policy was initially viewed by some as a possible threat to scientific publishers. The Policy was designed with a delay period and multiple ways for awardees to comply with the Policy that were compatible with all publishing models. This design appears to have not harmed the publishing industry.

Even since the NIH Public Access Policy became a requirement in 2008, available data suggests that publishing industry continues to grow. Even while the U.S. economy suffered a significant downturn, the scientific, technical, and medical (STM) publishing industry appears strong, with increases in both the number and price of STM journals. For example, from 2007 to 2012, the number of biological sciences and agriculture journals and medicine and health journals grew by 17 percent and 21 percent, respectively.[52] From 2007 to 2013, the average prices of biology journals and health sciences journals have increased 54 percent and 24 percent, respectively.[53]

An emerging discussion about the Public Access Policy is whether posting papers to PMC after an embargo might affect the number of downloads of these papers from publisher websites. The evidence is mixed and scant, and may vary by the version posted to PMC. The NIH Public Access Policy is designed to be flexible; it requires the final peer-reviewed manuscript to be posted, but allows publishers to post the final published version at their discretion.

Davis[54] compared the download rates of final published articles from 14 journals that were either publicly accessible exclusively on the publisher site or publicly accessible on both the publisher site and PMC. He found relatively fewer downloads from the publisher site for papers that were available on both the publisher site and PMC. In contrast, the PEER study[55] focused on the final peer-reviewed manuscript. They randomly assign papers from 135 journals to be publicly available exclusively on the publisher's website, or publicly available on a publisher site with final peer-reviewed manuscripts also publicly available from a central repository. This last condition most closely mirrors the requirements of the NIH Public Access Policy. They found papers were downloaded more often from the publisher site if manuscripts were publicly available on central repositories. Neither study demonstrated an association between downloads and financial health of scholarly publishers.

---

[52] http://ulrichsweb.serialssolutions.com/. Most recent data on peer-reviewed journals available. Data accessed 6/13/13.
[53] Library Journal Periodicals Price Surveys, 2009-2011. http://www.libraryjournal.com/article/CA6651248.html ; http://www.libraryjournal.com/article/CA6725256.html; http://www.libraryjournal.com/lj/ljinprintcurrentissue/890009-403/periodicals_price_survey_2011_.html.csp; ISrice data, table 4, http://lj.libraryjournal.com/2013/04/publishing/the-winds-of-change-periodicals-price-survey-2013/.
[54] Davis PM. Public accessibility of biomedical articles from PubMed Central reduces journal readership-- retrospective cohort analysis. *FASEB J.* 2013 Apr 4. http://www.fasebj.org/content/early/2013/04/02/fj.13-229922.full.pdf.
[55] Rowlands I, Clarck D, Nicholas D. *D5.3 PEER Usage Study Randomised controlled trial results.* http://www.peerproject.eu/fileadmin/media/reports/20120618_D5_3_PEER_Usage_Study_RCT.pdf

The final indicator of publisher impact is the strong and active support of the Public Access Policy volunteered by the publishing community. The Policy asks NIH-funded authors to submit their final peer-reviewed manuscripts directly to the NIH. Hundreds of publishers have volunteered to assist their authors in supporting the policy. Thousands of journals voluntarily submit peer-reviewed author manuscripts to PMC to assist authors in complying with the Public Access process.[56] Several hundred journal publishers also voluntarily automatically deposit final published versions of articles on behalf of their authors, relieving them of the need to directly submit their manuscripts. Finally, publishers representing about 1,200 journals voluntarily submit the full content of their journals to PMC, regardless of whether the issue contains an article subject to the NIH Public Access Policy.[57]

Section III: Digital Scientific Data

1. Preamble

NIH-funded research generates digital scientific data across the entire spectrum of behavioral and biomedical research (e.g., basic research, clinical research, epidemiologic research). To the extent feasible and consistent with applicable law and policy; agency mission; resource constraints; U.S. national, homeland, and economic security digitally formatted scientific data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze. To advance this goal and NIH's mission to seek fundamental knowledge about the nature and behavior of living systems and apply that knowledge to enhance health, lengthen life, and reduce illness and disability, NIH intends to make public access to digital scientific data the standard for all NIH-funded research. This section provides a plan for meeting this objective. Underpinning this plan for increasing access to digital scientific data are the following principles:

- The sharing and preservation of data advances science by broadening the value of research data across disciplines and to society at large, protecting the integrity of science by facilitating the validation of results, and increasing the return on investment of scientific research.
- Protecting confidentiality and personal privacy are paramount, and no changes will be made to existing policies that would reduce current protections.
- Proprietary interests, business confidential information, intellectual property rights, and other relevant rights will continue to be appropriately protected.

---

[56] These publishers are listed under Method D at http://publicaccess.nih.gov/select_deposit_publishers.htm.
[57] For a full list see http://publicaccess.nih.gov/submit_process_journals.htm.

- The costs and benefits of data management should be considered in policy development and implementation. It is important to note that not all digital scientific data need to be shared and preserved. A fundamental determinant of the need to share and preserve data is the further insight to be gained from increased public access.
- Data management planning should be an integral part of research planning, and planning for data collection or creation and management should take into account downstream data processing and dissemination.
- Coordination and communication with other Federal agencies and HHS operating divisions is essential to minimize burdens, prevent unnecessary duplication, and conserve resources and funding in achieving the goal of increased access to digital scientific data.
- This plan is intended to address only unclassified research. NIH does not fund classified research.

This plan outlines the current NIH policies, programs, and procedures that support the overall goals of the OSTP memorandum and identifies further steps that may be taken to fulfill these goals in a more comprehensive manner to ensure public access to digital scientific data. NIH programs and initiatives that intend to develop policies in areas addressed by the OSTP memorandum will be informed by the development of this plan. Many of the final details will be determined during policy development and implementation, and in consultation with the scientific community and the public.

2. Definitions

a. **Digital Scientific Data:** Consistent with the OSTP memorandum and OMB Circular A-110,[58] digital scientific data are defined for the purpose of this plan as "the digital recorded factual material commonly accepted in the scientific community as necessary to validate research findings including data sets used to support scholarly publications, but does not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens."[59] NIH considers "digital" data to be data that has been recorded in any electronic format that can be accessed using a computer.

The definition of digital scientific data includes data that are used to support a scientific publication as well as data from completed studies that might never be published. It may include data that support or refute a hypothesis, but does not include draft or preliminary data sets.

---

[58] http://www.whitehouse.gov/omb/circulars_a110

[59] NIH has policies governing the appropriate sharing of NIH-funded research resources other than data (e.g., model organisms, cell lines, research reagents). Those policies can be found at http://sharing.nih.gov.

For the purpose of this plan, the definition of digital scientific data does not include software *per se.* NIH recognizes that software and tools such as interview protocols, coding guides, data collection instruments, or manuals may be necessary to access and interpret data. In such cases, and also when the research involves the collection and analysis of administrative data (e.g., data from the Social Security Administration) the data management plan will be expected to address how information about such items will be made available.

b. **Public Access:** For the purpose of this plan, "Public Access" is defined as the availability of data for public use. It may include either data that are openly accessible and available for any use or data that are accessed in a controlled manner to protect appropriately certain interests, for example, the privacy of research participants, intellectual property, or security.

c. **Individual-level Data:** Individual-level data refers to data that include details about individual organisms, whether non-human or human.

d. **Summary or Aggregate-level Data:** Summary or aggregate-level data refers to data that have been aggregated to describe a group of organisms as a whole.

3. **Planning and Policy Development**

a. To develop a plan to increase access to scientific digital data and coordinate policy development, NIH has established an internal Steering Group on Public Access to Digital Scientific Data, composed of senior representatives from NIH Institutes, Centers, and Offices (ICs) with relevant expertise and policymaking experience in both intramural and extramural programs. The Steering Group is co-chaired by the Office of Science Policy (OSP) and the Office of Extramural Research (OER), which have leading roles in policy development for data sharing. The Steering Group consulted with the NIH ICs and other NIH offices to ensure that the proposed plan is informed by IC-specific and trans-NIH perspectives.

b. The Deputy Director for Extramural Research and the Associate Director for Science Policy, in coordination with the Deputy Director for Intramural Research, also will assist in development of the final plan and, as appropriate, any policy development required by the plan.

c. NIH has established the position of NIH Associate Director for Data Science (ADDS). The NIH ADDS serves as Chair of the newly created trans-NIH Scientific Data Council (SDC), whose responsibilities include:
(a) programmatic oversight of the Big Data to Knowledge (BD2K) initiative;
(b) trans-NIH programmatic leadership and coordination of data science activities;
(c) coordination with data science activities beyond NIH;
(d) significant involvement in the establishment and management of NIH data-

sharing policies;
(e) NIH-wide, long-term strategic planning in data science. The ADDS and the SDC will be responsible for overseeing the implementation of this plan, in coordination with OSP, OER, the Office of Management (OM), the Office of Intramural Research (OIR), and the NIH Chief Information Officer (CIO).

4. Data Management Plans

Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified. (OSTP memo, element 4b)

a. **Current approaches:** Many current NIH policies establish expectations for researchers to share data or develop data sharing plans.

1. Statute-based Policies: Under Title VIII of Food and Drug Administration Amendments Act of 2007 (FDAAA), "applicable clinical trials" are required to be registered in the publicly accessible registry, ClinicalTrials.gov, which is managed by NIH. The law also requires summary results of certain applicable clinical trials to be submitted to the data bank. NIH currently encourages the registration of all NIH-funded clinical trials regardless of whether or not they are subject to FDAAA.

2. Selected Trans-NIH Policies:

    a) The *2002 NIH Intramural Policy on Large Database Sharing* applies the general principles of the *2002 NIH Draft Statement on Sharing Research Data*[60] to intramural researchers involved in the collection and analysis of large-scale databases. These principles include promoting open scientific inquiry, facilitating the creation of new data sets when data from multiple sources are combined, encouraging diversity of analysis and opinion, avoiding duplication of expensive data collections, and expediting the translation of research results into knowledge and products to improve human health.

    b) The *2003 NIH Data Sharing Policy*[61] encourages NIH-funded researchers to share their final research data for use by other researchers in a timely way (i.e., no later than the acceptance for publication of the main findings

---

[60] http://grants.nih.gov/grants/guide/notice-files/NOT-OD-02-035.html
[61] http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html

from the final data set). The Policy expects applicants requesting $500,000 or more in direct costs in funding from NIH for research for any one year to include a data sharing plan or state why data sharing is not possible. Supplemental guidance materials[62] suggest that plans should describe (1) whether and how data will be made available to others, including provisions for protection of privacy, confidentiality, security, intellectual property, or other rights as appropriate; (2) items such as the data to be shared (e.g., genomic, clinical, or images), the expected timeline for when the data will be available, data formats, the format of the final data set, any query and/or analytic tools that will be provided, and the mode of data sharing (e.g., through a data archive or enclave or under the researcher's own auspices by mailing a disk or posting data on an institutional or personal website); and (3) procedures to request the data and any required data sharing agreements including the criteria for accessing data and any limitations placed on the use of data.

c) The *NIH Grants Policy Statement* requires a final progress report in which researchers are expected to describe any data, protocols, software, or other information or materials resulting from the research that is available to be shared with other researchers and how it may be accessed. According to OMB Circular A-110, research data must be retained for three years from the date of submitting a Federal Financial Report.[63]

d) NIH 2014 *Genomic Data Sharing (GDS) Policy* expects researchers receiving NIH funding for genomic research to develop data sharing plans and submit their data to a central repository to facilitate sharing.[64]

3. IC-Specific Policies: The National Institute on Aging (NIA) has had a specific policy in place since 2002 to promote the sharing of data generated through research on the genetic factors that contribute to Alzheimer's Disease.[65] Some ICs such as the National Heart, Lung, and Blood Institute (NHLBI) and the National Institute on Drug Abuse (NIDA) expand on the 2003 policy in a number of ways (e.g., broadening the applicability by adding a number of other criteria for data sharing plans, such as number of participants enrolled in a study).[66]

---

[62] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm and http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm
[63] http://grants.nih.gov/grants/policy/nihgps_2012/
[64] http://gds.nih.gov
[65] http://www.nia.nih.gov/research/dn/sharing-policy-and-guidance-research-genetics-alzheimers-disease
[66] http://www.nhlbi.nih.gov/funding/datasharing.htm; http://www.drugabuse.gov/researchers/research-resources/genetics-research-resources/nih-nida-data-sharing-policies

4. Programs Requiring Data Sharing: Data sharing is also expected or required by numerous programs and program policies currently in place at NIH, such as the Environmental Determinants of Diabetes in the Young Sample and Data Sharing Policy,[67] the Alzheimer's Disease Neuroimaging Initiative (ADNI) Data Sharing and Publication Policy,[68] the data-sharing policy of the National Database for Autism Research (NDAR),[69] the ENCyclopedia Of DNA Elements (ENCODE) Consortia Data Release Policy,[70] the NIA-funded Health and Retirement Study,[71] the Human Microbiome Project's Data Release and Resource Sharing Guidelines,[72] and the Brain Research through Advancing Innovative Neurotechnologies (BRAIN) initiative's requirement for an "exceptional level of coordination and sharing."[73]

b. **Further steps under consideration:** In the area of clinical data, NIH is proposing steps to expand transparency of and public access to data resulting from NIH-funded clinical trials. Public dissemination of information about NIH-funded clinical trials is particularly important because the findings are derived from human volunteers who may assume risk in order to contribute to generalizable medical knowledge and, ultimately, to public health. Sharing of clinical trial information thereby helps investigators, and NIH as an agency, fulfill an ethical obligation to participants. Ready access to clinical trials registration information and results informs future research and improves study design, prevents duplication of unsafe trials, and enhances transparency for increased public trust in clinical research. To that end, NIH is engaged in regulatory and policy development to promote the registration of all NIH-funded clinical trials and the submission of results from those trials to ClinicalTrials.gov. To explore optimal approaches to clinical data sharing, NIH is supporting an Institute of Medicine study of clinical trial data sharing, including participant level data in addition to summary results data. In an interim report on this topic, the IOM noted that a cultural change has occurred in discussions about clinical data sharing. Rather than exploring whether it should occur, the focus is on how it should be accomplished.[74] Further evidence of the shift is provided by the steps that have been taken by the European Medicines Agency to make summary-level, and eventually participant-level clinical trial data accessible for research. [75]

---

[67] https://www.niddkrepository.org/niddkdocs/TEDDY/TEDDYSampleandDataSharingPolicy.pdf
[68] http://adni.loni.ucla.edu/wp-content/uploads/how_to_apply/ADNI_DSP_Policy.pdf
[69] http://ndar.nih.gov/ndarpublicweb/Documents/NDAR_Policy.pdf
[70] http://www.genome.gov/Pages/Research/ENCODE/Mod-ENCODE_Consortia_Data_Release_Policy_revised_11-22-09.pdf
[71] http://hrsonline.isr.umich.edu/index.php?p=data
[72] http://commonfund.nih.gov/hmp/datareleaseguidelines
[73] See for example http://grants.nih.gov/grants/guide/rfa-files/RFA-NS-14-007.html
[74] *Discussion Framework for Clinical Trial Data Sharing: Guiding Principles, Elements, and Activities*, http://www.iom.edu/Activities/Research/SharingClinicalTrialData.aspx
[75] http://www.ema.europa.eu/ema/index.jsp?curl=pages/news_and_events/news/2014/10/news_detail_002181.jsp&mid=WC0b01ac058004d5c1

NIH will explore the development of policies to require NIH-funded researchers to make the data underlying the conclusions of peer-reviewed scientific research publications freely available in public repositories at the time of publication in machine readable formats. NIH will ensure that data management plans include clear plans for sharing research data. NIH is taking steps to ensure all NIH-funded researchers develop data management plans whether they are funded by a grant, cooperative agreement, contract, or intramural funds, regardless of funding level. As a first step, the 2003 NIH Data Sharing Policy will be modified to require that all NIH-funded researchers develop data management plans that will (1) express the investigator's commitment to sharing their data, which will at a minimum consist of the data underlying any publications, [76] and (2) include, as appropriate, descriptions of the data to be produced in the proposed study; any standards to be used for collected data and metadata; mechanisms for providing access to and sharing of the data; provisions for protection of privacy, confidentiality, security, intellectual property, or other rights; provisions for reuse and redistribution of the data; milestones and timelines for making the data publicly accessible; and plans for archiving and long-term preservation of the data. The data management plans may also be expected to describe tools, including software, which may be needed to access and interpret the data. Certain funding mechanisms that are unlikely to generate scientific data, such as training grants, may be exempted from the requirement to develop a data management plan. NIH intends to develop guidance describing the key elements to be included in data management plans, as is currently available for data sharing plans.[77]

The timing of the sharing of data should take into account the important role that publication of findings plays in the advancement of scientific careers, and NIH will explore appropriate timelines and expectations to ensure that researchers retain the ability to receive appropriate credit for their work and protect intellectual property interests (see Plan Element 5b).

Most digital scientific data that are produced with federal funding are also subject to expectations of the M-13-13 Open Data Policy, including data generated through contracts, intramural research, or deposited in an NIH data repository for use and distribution. These data will be provided to the public in a way that supports downstream information processing and dissemination activities (e.g., are machine-readable, use data standards, utilize open use licenses to the extent possible, and use common core and extensible metadata). (See also Plan

---

[76] These steps are also responsive to recent Congressional interest in NIH efforts to increase access to scientific data. See https://www.congress.gov/congressional-record/2014/12/11/house-section/article/H9307-1
[77] See for example http://grants.nih.gov/grants/sharing_key_elements_data_sharing_plan.pdf

Elements 10b and 11b). Additional information regarding appropriate compliance with these expectations may also be included in data management plans.

NIH-supported researchers may also be expected to provide information to be used in registration and indexing of data to ensure that data sets resulting from their research are discoverable and can be included in an inventory of data sets; see Plan Element 9b for a description of the data discovery index that is under development through BD2K. In addition to those elements, NIH, through the ADDS and the SDC, will continue to work in consultation with the biomedical research community to identify areas of research for which more specific data management and sharing policies should be developed; see Plan Element 4a for examples of such existing policies.

5. Policy Considerations Related to Data Management

Maximize access, by the general public and without charge, to digitally formatted scientific data created with Federal funds, while: i) protecting confidentiality and personal privacy, ii) recognizing proprietary interests, business confidential information, and intellectual property rights and avoiding significant negative impact on intellectual property rights, innovation, and U.S. competitiveness, and iii) preserving the balance between the relative value of long-term preservation and access and the associated cost and administrative burden. (OSTP memo, elements 4a.i-iii)

a. **Current approaches:**

1. **Maximize access by the general public without charge:** Existing NIH policies to promote public access to digital data are designed to provide free access to applicable data. See element 4a for existing data sharing policies that promote public access.

2. **Protecting confidentiality and personal privacy:** NIH researchers are subject to applicable rules governing the privacy of research participants, including in some cases the HIPAA Privacy Rule,[78] and in all cases HHS regulations governing the protection of human subjects.[79] Personally identifiable information (PII) held by Federal agencies and maintained in a system of records is governed by the 1974 Privacy Act, which prohibits the release of such information with several limited exceptions.[80] The 2003 Data Sharing Policy expects that the rights and privacy of participants in NIH-funded research will be protected. Many NIH policies establish expectations

---

[78] http://www.hhs.gov/ocr/privacy/hipaa/administrative/privacyrule/index.html
[79] http://www.hhs.gov/ohrp/humansubjects/commonrule/
[80] http://www.justice.gov/opcl/1974privacyact-overview.htm

for protecting the interests of research participants that go beyond the scope of protections required currently by law or Federal regulations. For example, the GDS Policy's "controlled access" mechanism allows NIH to control access to human genomic data by ensuring that the research proposed in requests for data is consistent with limitations placed on the use of the data by the submitting institution based on the original informed consent. The GDS Policy also encourages researchers to apply for Certificates of Confidentiality to protect identifiable research information from forced disclosure, and NIH has obtained such a certificate for the database of Genotypes and Phenotypes (dbGaP).[81] The NIH Grants Policy Statement[82] also includes provisions for the protection of sensitive data and information used in research. Submission of summary level, rather than individual level, results to data repositories, such as the submission of clinical trials results to ClinicalTrials.gov, is another way to share data while protecting participant confidentiality.

Data security is an important aspect of protecting privacy, and as needed, NIH policies include provisions related to data security, and/or refer to regulatory requirements, as applicable, such as the HIPAA Security Rule.[83] Additionally, data that belong to the Federal government are subject to the Federal Information Security Management Act (FISMA).[84] The 2003 Data Sharing Policy does not explicitly require that data security be included in data sharing plans, but it is understood that such plans must also be consistent with applicable laws, regulations, rules, and policies. Other policies, such as the GDS Policy, have established standards for data security for researchers working with data from NIH data repositories.[85]

3. **Intellectual property protection:** The 2003 NIH Data Sharing Policy and program policies recognize that there may be times when the sharing of intellectual property and other forms of proprietary information may need to be restricted in order to foster innovation and investment. Currently, several policies, including the GDS Policy, explicitly expect appropriate attribution of data sets. The 2003 NIH Data Sharing Policy recognizes intellectual property rights, and guidance documents[86] state that a delay in disclosure of research findings of 30 to 60 days after data collection may be justified.[87] In addition, applicants or offerors for Small Business Innovation Research (SBIR) grants and contracts are expected to submit data sharing plans but, under the Small

---

[81] http://grants.nih.gov/grants/policy/coc/
[82] http://grants.nih.gov/grants/policy/
[83] http://www.hhs.gov/ocr/privacy/hipaa/administrative/securityrule/
[84] http://csrc.nist.gov/groups/SMA/fisma/overview.html
[85] http://www.ncbi.nlm.nih.gov/projects/gap/pdf/dbgap_2b_security_procedures.pdf
[86] http://grants.nih.gov/grants/policy/data_sharing/
[87] FAQ #10, http://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm

Business Act, may withhold their data for four years after the end of the award to protect proprietary information. Many individual program data-sharing policies at NIH also have established expectations for sharing or not sharing intellectual property, as appropriate, depending upon specific data types. For example, the GDS Policy discourages premature claims on pre-competitive information that may impede research, though it encourages patenting of technology suitable for subsequent private investment that may lead to the development of products that address public needs.

4. **Preserving the balance between cost and value:** The 2003 Data Sharing Policy specifies that researchers are responsible for finding a balance between the value of the final data and the costs of archiving and sharing, but the policy does not include a method for assessing the value of long-term preservation and sharing of data. NIH has recognized in some cases that the expense of archiving and sharing data may outweigh its broader utility. However, if the researcher publishes results based on those data, NIH expects that the data will be shared at the time of acceptance for publication.

5. **Other policy considerations:** In accordance with the 2012 *United States Government Policy for Oversight of Life Sciences Dual Use Research of Concern*, NIH ensures that the risks associated with the communication of research results are considered for NIH-supported research projects that are subject to the policy.[88]

b. **Further steps under consideration:**

1. **Maximize access by the general public without charge:** NIH will continue ensure that NIH repositories and policies for digital data maximize access by the general public without charge. See elements 4b for plans to modify data sharing policies to maximize public access.

2. **Protecting confidentiality and personal privacy:** NIH is working to enhance the policy framework for enabling the ethical sharing of data in less restrictive ways (e.g., through participant consent for open access and broad use of individual-level data). NIH will maintain its commitment to the protection of the rights and welfare of human research subjects. NIH will continue, upon request, to issue Certificates of Confidentiality to applicable databases to protect identifiable research information from forced disclosure and provide guidance to researchers on the issues to consider when using this form of privacy protection. Any changes made to NIH policies will continue to be consistent with applicable laws, regulations, and Executive Orders governing the privacy and confidentiality of individual human data, as

---

[88] http://www.phe.gov/s3/dualuse/Documents/oversight-durc.pdf

well as consistent with the relevant expectations of the M-13-13 Open Data Policy. Program policies that have established greater levels of protection for data sharing activities than in current law will not be required to reduce the levels of protection afforded by those policies. NIH is also obligated to respect research participants' wishes, including those who desire to share their de-identified data openly. NIH will continue to implement and develop methods to ensure privacy and confidentiality, and maintain data security.

3. **Intellectual property protection:** NIH will continue to use existing policies for the appropriate attribution of scientific data sets and plans to develop new methods to ensure attribution of scientific data generated by NIH-funded research (see Plan Element 12b). NIH will continue to recognize intellectual property rights as appropriate, consistent with regulations and program policies, including considerations for intellectual property based on the type of data subject to those policies (e.g., varied embargo dates, conditions for delaying data release). For the purpose of this plan, proprietary interests include receiving appropriate credit for scientific work. If the outcomes of the research result in inventions, the provisions of the Bayh-Dole Act of 1980, as implemented in 37 CFR Part 401, apply.

4. **Preserving the balance between cost and value:** In order to preserve the balance between the relative benefits of long-term preservation and access and the associated cost and administrative burden, NIH will continue to expect researchers to consider the benefits of long-term preservation of data against the costs of maintaining and sharing the data.

5. **Other policy considerations:** NIH will continue to ensure that increased public access to digital scientific data is consistent with the *United States Government Policy for Oversight of Life Sciences Dual Use Research of Concern*.

6. Evaluation of Data Management Plans

Ensure appropriate evaluation of the merits of submitted data management plans. (OSTP memo, element 4d)

a. **Current approaches:** Under the 2003 NIH Data Sharing Policy, NIH evaluates the appropriateness and adequacy of proposed data sharing plans for grantees through an administrative review, although those evaluations are not factored into the scoring of the application. For contracts, depending on the requirements of the contract, program staff ensures that data sharing plans are considered during the process of evaluating proposals. The NIH Guidelines for the Conduct of Intramural

Research Programs state that data management is a responsibility of the Principal Investigator and that data supporting published analyses should be shared.[89] Under the Intramural Policy on Large Database Sharing,[90] the appropriateness of data sharing and the mechanisms involved in sharing and archiving data are considered when projects are reviewed by the Board of Scientific Counselors (BSC) and other external advisory bodies at the NIH IC-level. Some policies have broader expectations, such as the GDS Policy, which expects that intramural researchers develop data-sharing plans consistent with that Policy.

b. **Further steps under consideration:** NIH will determine the additional steps needed to ensure that the merits of digital data management plans are considered during the peer review process for extramural research grants and contracts. NIH also plans to ensure that data management plans for all intramural research are reviewed by the senior official responsible for scientific leadership within each NIH Institute or Center, i.e., the Scientific Directors. In addition, NIH will assess whether the appropriate balance has been achieved in data management plans between the relative benefits of long-term preservation and access and the associated cost and administrative burden. NIH plans to develop guidance for determining which data should be prioritized for long-term preservation and access, in consultation with the scientific community.

7. Ensuring Compliance with Plans and Policies

An agency strategy for measuring and, as necessary, ensuring compliance with its plan; Include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies. (OSTP memo elements 2e and 4e)

a. **Current approaches:** NIH's current expectations for data sharing are communicated to intramural and extramural researchers through numerous guidance documents[91] and outreach activities, and these are reiterated to funded researchers through the Notice of Award or the Contract Award. NIH staff is responsible for monitoring and taking steps to ensure compliance with the terms and conditions of the award. Data-sharing plans that are approved and made a term and condition in the Notice of Award or the Contract Award can be enforced. Accordingly, failure to comply with the terms and conditions of the funding agreement could lead to enforcement actions, including the withholding of funding, consistent with 45 CFR 74.62 and/or other authorities as appropriate. Failure to comply with statutory requirements for submitting the results of

---

[89] http://sourcebook.od.nih.gov/ethic-conduct/Conduct%20Research%206-11-07.pdf
[90] http://sourcebook.od.nih.gov/ethic-conduct/large-db-sharing.htm
[91] http://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm

applicable clinical trials to ClinicalTrials.gov can also result in penalties specified in section 402(j)(5) of the Public Health Service Act.

b. **Further steps under consideration:** To improve monitoring of grantee and contractor compliance, program staff overseeing extramural grants and contracts will review progress reports carefully for adherence with data management plans and policies. For contracts, this will include adherence with the requirements of the M-13-13 Open Data Policy. Development of procedures for enhancing the discoverability of data sets, as part of the BD2K initiative, will provide an additional mechanism for monitoring compliance with data management plans by including a citable unique identifier that can be used to verify that novel data sets are registered in accordance with applicable NIH policy. Under the intramural program, IC Scientific Directors will be responsible for carrying out a prospective review of intramural researchers' compliance with data management policies and plans. OIR plans to use its annual Management Survey of Scientific Directors to monitor overall compliance with data management and data-sharing expectations and policies developed to implement this plan, including the requirements of the M-13-13 Open Data Policy.

8. **Leveraging Existing Archives and Repositories**

A strategy for leveraging existing archives, where appropriate, and fostering public-private partnerships with scientific journals relevant to the agency's research; Promote the deposit of data in publicly accessible databases, where appropriate and available. (OSTP memo elements 2a and 4f)

a. **Current approaches:** NIH currently supports numerous data repositories for a range of research fields and data types, and many policies at NIH such as the GDS Policy designate a central repository for researchers to deposit their data (i.e., dbGaP). Many NIH program-level data sharing policies expect deposition of data in existing repositories.

b. **Further steps under consideration:** NIH will expect funded researchers to deposit data in appropriate, existing, publicly accessible repositories before considering other means of making data available. Researchers will be expected to describe any use of data repositories in their data management plans. NIH will work to ensure that NIH repositories are designed to minimize the burden of submitting data and will consider whether enhancements to existing archives are needed to accommodate the deposition of additional data. In addition, NIH will ensure that new archives support interoperability and information accessibility to support downstream information processing and dissemination activities (e.g., by using open data standards, open licenses as appropriate and common core and extensible metadata) consistent with the requirements of the M-13-13 Open Data Policy. If a researcher's data management plan proposes to use a repository or other resources supported by an organization other than NIH, the researcher will be expected to notify the organization of their intent.

To help researchers find an appropriate repository to accept their data, NIH will expand its database of existing repositories.[92] NIH plans to develop guidance and criteria to aid researchers in identifying acceptable repositories not funded by NIH. In addition, NIH intends to make use of the BD2K data discovery index initiative to identify the types of data for which public repositories are lacking and may be needed.

9. **Improving Public Access to Scientific Data**

A strategy for improving the public's ability to locate and access digital data resulting from federally funded scientific research. (OSTP memo, element 2b)

a. **Current approaches:** NIH has invested significant time and effort to facilitate the public's ability to locate and access digital scientific data. For example, OER,[93] the National Library of Medicine (NLM)[94] and its National Center for Biotechnology Information (NCBI),[95] and the Trans-NIH BioMedical Informatics Coordinating Committee (BMIC)[96] maintain lists of databases and data repositories, and display them on their websites. NCBI enables searches of the contents of multiple NCBI databases (including PubMed, dbGaP, and PubChem) with one query. ClinicalTrials.gov provides a mechanism for allowing the public to search for summary results of clinical trials.

b. **Further steps under consideration:** NIH will explore the development of a commons, a shared space for basic and clinical research output including data, software, and narrative, that follows the FAIR principles of Find, Access, Interoperate and Reuse. A particular focus of the effort will be on making the data underlying the conclusions of peer-reviewed scientific publications resulting from federally funded scientific research available for free at the time of publication. In this effort, NIH will welcome opportunities to collaborate with other Departments and Agencies. NIH is taking steps to facilitate the discovery of data sets generated via NIH-funded research in ways that will enable researchers to locate and cite data sets, provide proper attribution for efforts related to the generation of data, and to link those data sets to the biomedical literature (see also Plan Element 12b). In 2014, NIH issued a Funding Opportunity Announcement to establish a consortium to develop the data discovery index that will provide these capabilities.[97] NLM is also exploring approaches for indexing and making separately searchable data

[92] http://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
[93] http://grants.nih.gov/grants/sharing.htm
[94] http://www.nlm.nih.gov/
[95] http://www.ncbi.nlm.nih.gov/
[96] http://www.nlm.nih.gov/NIHbmic/
[97] http://grants.nih.gov/grants/guide/rfa-files/RFA-HL-14-031.html

that are published in a number of new journals dedicated to describing datasets, such as *Scientific Data.* NIH recognizes the benefit of collaborating with other federal agencies and public and private stakeholders to develop methods for federated searching across multiple indices and will continue to work toward this goal.

Consistent with requirement of the M-13-13 Open Data Policy to establish a public data listing, NIH is working with HHS to continue to make information publicly available about NIH scientific (and administrative) datasets associated with large information technology (IT) investments, such as those currently made available through HealthData.gov. [98] These initiatives will also provide the capability to view trends in data collection, sharing, and other related activities. NIH will continue to update its data sharing webpages such as http://sharing.nih.gov. NIH also plans to continue efforts to improve access to data through public-private collaborations (see Plan Element 11b).

10. Optimizing Accessibility, Interoperability, and Long-Term Stewardship

An approach for optimizing search, archival, and dissemination features that encourages innovation in accessibility and interoperability, while ensuring long-term stewardship of the results of federally funded research. (OSTP memo, element 2c)

a. **Current approaches:** NIH is supporting initiatives to establish Common Data Elements (CDEs) for relevant areas of research to improve the interoperability and comparability of data collected in research studies. Initiatives such as the consensus measures for Phenotypes and Exposures (PhenX), NIH Toolbox for Assessment of Neurological and Behavioral Function, and Patient Reported Outcomes Measurement Information System (PROMIS), are intended for use by NIH and the biomedical research community.[99] NIH also works collaboratively with other HHS offices and operating divisions, including the Office of the National Coordinator for Health IT (ONC) to enable meaningful research use of electronic health record data, for example via standardized value sets and vocabulary standards supported by NLM. In addition, several programs at NIH support innovative services to archive and disseminate data, and NIH has begun to address issues relating to the adoption of cloud computing services for optimizing searching, archiving, and disseminating digital scientific data. NIH has developed a number of application programming interfaces (APIs) to make data more easily accessible to users in commonly used formats. [100] For research that is subject to the M-13-13 Open Data Policy, the metadata for scientific data will

---

[98] The Office of Management and Budget Open Data Policy requires agencies to develop public data catalogs, and guidance has been published on the Project Open Data website: http://project-open-data.github.io/catalog/
[99] http://cde.nih.gov
[100] http://www.nlm.nih.gov/api/

include, at a minimum, the common core metadata schema in use by the Federal government, found at https://project-open-data.cio.gov/.

b. **Further steps under consideration:** NIH will continue to identify and encourage the use of CDEs[101] and to coordinate such efforts among internal and external partners. BD2K activities will continue to support development of community-based efforts for the development of data and metadata standards to address issues of data interoperability and stewardship.[102] Data produced through intramural research or research supported by contracts, or data that are deposited in an NIH repository for use and distribution, must be provided in a way that supports downstream information processing and dissemination activities (e.g., are machine-readable, use data standards, utilize open-use licenses, and use common core and extensible metadata, as appropriate). (See also 11b.)

To ensure that data are accessible, interoperable, and useful now and into the future, NIH recognizes that it will be important to cite any software and analytical tools that are needed to access, interpret, and use the data. The data and software indexing efforts under BD2K (See Plan Element 12b) will make data and software more discoverable and accessible for the general public and scientific researchers.

As part of BD2K, NIH will also continue to explore and fund innovative tools and services that improve search, archiving, and disseminating of data, while ensuring long-term stewardship and usability. For example, through BD2K, NIH will continue to evaluate cloud computing options, and as appropriate, the NIH will develop policy, guidance, and points to consider for the use of cloud computing by NIH and NIH-funded researchers. NIH will also take part in collaborative efforts toward improving the ability to locate and obtain access to data globally.

11. **Cooperation with the Private Sector**

Encourage cooperation with the private sector to improve data access and compatibility, including through the formation of public-private partnerships with foundations and other research funding organizations. (OSTP memo, element 4g)

a. **Current approaches:** NIH has a long-standing commitment to public-private partnerships to increase public access to scientific data. For example, the Alzheimer's Disease Neuroimaging Initiative (ADNI) is a partnership among NIH, 23 pharmaceutical, imaging, and biotechnology companies, two foundations, and

---

[101] http://www.nlm.nih.gov/cde/
[102] See http://grants.nih.gov/grants/guide/rfa-files/RFA-HG-14-020.html

the Canadian Institutes of Health Research (CIHR). Another example is the pilot program on Discovering New Therapeutic Uses for Existing Molecules,[103] which is designed to develop partnerships between pharmaceutical companies and the biomedical research community by establishing a library of agents originally developed and provided by the private sector. In addition, NLM works with a number of private as well as public sector entities to support the development, maintenance, and dissemination of standards to improve the interoperability of electronic health records and other health information technology systems. (See also Plan Element 10a for other efforts related to interoperability.) NIH is also working with HHS and other agencies in the development of a data infrastructure for a comprehensive, interoperable, and sustainable network that will enable high quality, observational, patient-centered outcomes research to be performed by linking access to private and public claims, electronic health records, biobanks, registries, and other sources of data. NIH also joined a number of public and private sector organizations in commissioning the Institute of Medicine to study the sharing of clinical trial data.[104] Public-private partnerships could highlight challenges to data sharing associated with proprietary information and intellectual property that may need to be resolved. (See also Plan Element 5a and 5b for the protection of privacy and intellectual property.)

b. **Further steps under consideration:** NIH will explore expanding cooperation with the private sector through public-private partnerships and coordination with other agencies and HHS operating divisions. One such effort, which is being planned under the BD2K initiative, is to facilitate and support standards-related activities that are community-based. Many such communities and their stakeholders are in the private sector.

12. Developing Attribution for Scientific Data

Develop approaches for identifying and providing appropriate attribution to scientific data sets that are made available under the plan. (OSTP memo, element 4h)

a. **Current approaches:** Currently, there is no NIH-wide system for providing consistent identification and attribution for NIH-funded data that are made publicly available. While most NIH-supported data repositories provide unique identifers for submitted data sets, practices for attribution and citation vary from one system to another. For example, data sets submitted to NCBI's dbGaP under the GDS Policy are assigned a unique accession number, and the terms of use for the data expect that any publications resulting from secondary analysis of that data include an acknowledgement with the accession number.

---

[103] http://www.ncats.nih.gov/research/reengineering/rescue-repurpose/therapeutic-uses/therapeutic-uses.html
[104] http://www.iom.edu/Activities/Research/SharingClinicalTrialData.aspx

b. **Further steps under consideration:** NIH aims to ensure that data generated by NIH-funded research can be cited and attribution can be provided in a consistent manner. NIH will explore ways to advance data as a legitimate form of scholarship through data citation and other means. As part of the data discovery index, a system for unique identifiers for data sets generated by NIH-funded research will be developed, analogous to the PubMed Central identification number (PMCID) that is assigned to all submitted publications resulting from NIH-funded research. The identifier also would provide a means of linking the data with the biomedical literature via associated PubMed records. Such indexing of data sets would also facilitate generation of full citations to allow for differential recognition of scientific contributions (authorship on the scientific paper) and data-related contributions (authorship on the NIH data index entry). Approaches to coordinate with existing NIH data repositories will also be considered. NIH held a public workshop to obtain community input that would inform development of an NIH data discovery index in a way that supports appropriate citation and attribution.[105] NIH recognizes the benefit of collaborating with other federal agencies and public and private stakeholders to adopt consistent practices for citation of data sets across scientific communities and other data set attribution systems and will work toward this goal.

13. Training and Workforce Development

In coordination with other agencies and the private sector, support training, education, and workforce development related to scientific data management, analysis, storage, preservation, and stewardship. (OSTP memo, element 4i)

a. **Current approaches:** NIH awards training grants and has outreach programs designed, among other research-related objectives, to familiarize researchers and reference librarians with NLM databases. NLM supports university-based research training programs in biomedical informatics, short courses in biomedical informatics for health professionals, and training for medical librarians. The 2012 Report of the Data and Informatics Working Group (DIWG) of the Advisory Committee to the Director (ACD) found that training and workforce concerns related to the utilization of big data were a priority among NIH employees, extramural researchers, and the public.[106] Training for big data includes skills for the management, storage, and preservation of scientific data, in addition to analysis. Accordingly, a component of the BD2K initiative is dedicated to enhancing training for big data science, for which the BD2K is assessing training needs in this area, including both the augmentation of existing training mechanisms and the need for new training mechanisms to address researchers at all career stages. In 2014, through BD2K, NIH issued several awards to

---

[105] http://bd2k.nih.gov/pdf/DDI_Workshop_summary.pdfl
[106] http://acd.od.nih.gov/Data%20and%20Informatics%20Working%20Group%20Report.pdf

develop approaches for training, education, and workforce development for analysis and management of scientific data.[107]

b. **Further steps under consideration:** Through BD2K and other NIH initiatives, NIH will build upon its existing efforts to develop approaches for training, education, and workforce development for analysis and management of scientific data. In addition, NIH intends to develop approaches for training program staff and peer reviewers to evaluate data management plans and training NIH staff to evaluate and enforce compliance with data management plans and policies.

14. Notifying Researchers of Obligations

A plan for notifying awardees and other federally-funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, and/or regulatory changes). (OSTP memo, element 2d)

a. **Current approaches:** Grantees and contractors are notified and reminded about NIH's data sharing expectations through many means of public outreach and engagement, including Notices in the NIH Guide for Grants and Contracts,[108] NIH Grants Policy Statement, terms and conditions of each Notice of Award, and through changes to the NIH Request for Proposals and the NIH Contract documents.[109] Intramural researchers are notified via updates to the NIH Manual Chapters and by memoranda posted in the NIH Intramural Research Sourcebook.[110] The Deputy Director for Intramural Research (DDIR) assists in making the intramural community aware of changes. For contracts, notifications of changes are also published in outlets such as the NIH Guide to Grants and Contracts and FedBizOpps.gov.[111]

b. **Further steps under consideration:** Depending on the nature and scope of any change(s) to current policies and practices, NIH will consider public outreach and input through the resources referenced above and/or additional means as appropriate. As the recipient of the grant award and the party responsible for fulfilling the terms and conditions of the award, the funded party (e.g., the institution) should establish approaches to achieve institutional consistency and economies of scale for its researchers.

15. Support for Data Management

---

[107] http://bd2k.nih.gov/funding_opportunities.html
[108] http://grants.nih.gov/grants/guide/
[109] http://oamp.od.nih.gov/DGS/WKF/DGS_contract_workform_12182012.pdf
[110] http://oma.od.nih.gov/manualchapters/ and http://sourcebook.od.nih.gov/
[111] https://www.fbo.gov/

Allow the inclusion of appropriate costs for data management and access in proposals for federal funding for scientific research. (OSTP memo, element 4c)

a. **Current approaches:** NIH recognizes that support for data management should be considered a cost of research. The 2003 Data Sharing Policy allows applicants and offerors to request funds for data sharing and archiving in their applications and proposals, while expecting researchers to find a balance between the value of sharing and preserving the data with the associated costs (see Plan Element 5a.3). Some program policies may also make allowances for researchers to deposit certain types of data into NIH-designated data repositories. The 2003 Data Sharing Policy recognizes that costs will vary and that some researchers will have more experience estimating the costs of data sharing, but it does not provide specific guidance for estimating these costs.

b. **Further steps under consideration:** NIH will continue to allow for the inclusion of appropriate costs for data management and access in applications and proposals for NIH funding, and will highlight the acceptability of including this cost. The requested funding levels for data management plans will be appropriately assessed by NIH, and NIH will develop guidance to assist researchers and reviewers of data management plans in finding a balance between the value of archiving and sharing the data with the associated costs (see Plan Element 5b.3). Comparable assessments will need to be explored by the NIH intramural research program for intramural researchers. NIH also plans to conduct an analysis to estimate the costs of current data management activities in order to determine how best to support future data management (See also Plan Element 16).

16. **Cost of Implementation**

Identify resources within the existing agency budget to implement the plan. (OSTP memo, element 2f)

NIH has identified resources to support some relevant activities described in this plan, such as those under the BD2K initiative. However, since increasing public access to digital scientific data will require resources from within the existing NIH budget, NIH will need to assess the full range of costs associated with data management planning and efforts to support greater data sharing. It will be important to establish priorities to ensure that these activities increase the efficiency of the research enterprise and realize all the benefits of increased data access while maintaining support for research and the conduct of research and generation of new findings.

NIH will continue to review the costs and benefits of supporting repositories for data generated by NIH-funded research and to evaluate various economic models for supporting data repositories over the long-term. The ADDS, through strategic and resource planning and priority setting informed by the NIH Scientific

Data Council and the NIH Administrative Data Council, will coordinate plans to sustain the appropriate level of support for public access to and long-term preservation of digital scientific data.

17. Assessing Long-Term Preservation Needs

Provide for the assessment of long-term needs for the preservation of scientific data in fields that the agency supports and outline options for developing and sustaining repositories for scientific data in digital formats, taking into account the efforts of public and private sector entities. (OSTP memo, element 4j)

a. **Current approaches:** By not including a limit to the period of time for which data shared under the policy must remain archived, the 2003 Data Sharing Policy and other program policies implicitly expect long-term preservation of data. However, the 2003 Data Sharing Policy and many program policies do not provide explicit instructions for the long-term preservation of scientific data.

b. **Further steps under consideration:** Long-term preservation and sustainability will be included in data management plans (see also Plan Element 6b). Development of criteria for and conduct of periodic reviews to identify gaps in preservation coverage and response to changing needs arising from new data types may become responsibilities of the ADDS. In order to determine how best to develop and sustain repositories for digital scientific data, NIH intends to collaborate with HHS operating divisions and other agencies that support research in related areas.

Section IV: Implementation

1. **Timeline for implementation of the plan.** (OSTP memo, element 2g)

   NIH has already implemented policies that meet all of the objectives of the OSTP memorandum for publications and, for the most part, the objectives for digital scientific data. Over the next year, NIH will explore whether additional steps are needed to achieve the fullest measure of public access to results of NIH-funded research. Should any policy changes be required, OSTP has set the end of calendar year 2015 as the date by which policies to fulfill the plan should be issued. NIH will, as necessary, seek public comment on any policy changes. NIH will also continue to work toward achieving the goals of the M-13-13 Open Data Policy during this period and after.

2. **Obstacles to Meeting the Objectives:** *Identify any special circumstances that prevent the agency from meeting any of the objectives set out in this memorandum, in whole or in part.* (OSTP memo, element 2h)

   With respect to human data, the nature of informed consent obtained from participants may prevent some human data from being made publicly accessible or

from being used for some purposes. In addition, some types of data that meet the regulatory definitions for de-identified data may be re-identifiable. NIH has established mechanisms for facilitating broad access to human data while protecting participant privacy and confidentiality, which necessarily limit some access to or uses of the data. For example, in order to access data in NDAR and dbGaP, data access committees evaluate requests to access the data to ensure that proposed research is consistent with limitations on data use. NIH also is actively exploring technical measures to protect the confidentiality of such data and will continue to work to ensure that a robust regime of privacy protections is in place for data that are potentially re-identifiable.

NIH will assess the costs and benefits and may determine that, given current fiscal or other constraints, NIH must delay the implementation of some aspects of this plan (see also Plan Section III, Element 16). The time and expense of establishing suitable data repositories for data generated by NIH-funded research may also slow the pace with which such data will become broadly accessible.

NIH must also evaluate the implementation costs of data management within the constraints of existing budgets and resources. Decisions will need to be made by NIH to prioritize and balance the funding of data management plans within the existing budget for funding research projects. Accordingly, the NIH plan will be implemented consistent with available resources.

Additional challenges to implementing the plan in the near term include any statutory and regulatory limitations, such as for requiring grantees to develop data management plans. Regulatory changes may require a significant amount of time to accomplish. Given different types of data and the varying levels of data sharing currently practiced in different scientific fields, it may be challenging and time consuming to develop a consensus on criteria to determine which data should be preserved and shared. Any changes to the information requested from researchers will be subject to Paperwork Reduction Act and other legal and programmatic considerations, which may further increase the amount of time required to implement the final plan.

Section V: Conclusion

The OSTP memorandum was issued at an opportune time. The growth in the volume of scientific publications and the generation of increasing volumes of digital scientific data have occurred alongside a growing expectation of openness in government and federally-funded research within the scientific community and by the public. Meeting the objectives of the OSTP memorandum is consistent with and in support of NIH's ongoing efforts to respond to this growing demand and further advances NIH's mission to improve human health through biomedical and behavioral research.