NSF'S PUBLIC ACCESS PLAN:

Today's Data, Tomorrow's Discoveries

Increasing Access to the Results of Research Funded by the National Science Foundation

National Science Foundation

March 18, 2015

NSF 15-52

FOREWORD

With the release of the public access plan, *Today's Data, Tomorrow's Discoveries*, the National Science Foundation (NSF) continues its commitment to expand public access to the results of its funded research. Public access is intended to accelerate the dissemination of fundamental research results that will advance the frontiers of knowledge and help ensure the nation's future prosperity. NSF's plan is grounded in the realization that clear and open communication of research results is central to fulfilling NSF's primary mission of promoting the progress of science.

NSF is unique among science agencies for its broad scope, providing support for research and education in all branches of science and engineering. Therefore, the plan necessarily accommodates the diverse array of research results generated by NSF-funded Principal Investigators. The plan takes advantage of existing infrastructure services and seeks to build partnerships with other federal agencies as well as with public and private sector groups, such as institutions of higher education, publishers, libraries, and technology companies.

In this complex world where information technology advances rapidly, the Foundation recognizes the importance of proceeding in a way that enables both extraordinary progress and disruptive change. NSF's public access plan is designed to take advantage of such change when it occurs.

France A. Córdova
Director

TABLE OF CONTENTS

1.0 INTRODUCTION: BACKGROUND AND PURPOSE

The National Science Foundation (NSF) is the primary Federal agency supporting research at the frontiers of knowledge, across all fields of science and engineering (S&E) and all levels of S&E education. On February 22, 2013, the White House Office of Science and Technology Policy (OSTP) issued a memorandum, *Increasing Access to the Results of Federally Funded Scientific Research*,[1] which directs Federal agencies with more than $100 million in research and development (R&D) expenditures per year to develop plans to make publicly available to the "greatest extent and with the fewest constraints possible and consistent with law" the "direct results of federally funded scientific research." This document sets forth NSF's plan for reaching the objectives laid out in the OSTP memorandum through an open, flexible, and incremental approach that ultimately will:

- Integrate publications, data, and other products of NSF funding into a single management system;
- Build on current policies and practices;
- Leverage resources in other Federal agencies, universities and research institutes, and the private sector;
- Provide a platform for innovation; and
- Broaden access to NSF-funded research findings with necessary and appropriate safeguards.

In developing this plan, NSF has considered internal, administrative resources and infrastructure assets; broad patterns of publication by NSF-funded investigators and their sources of funding; and the creation and use of scientific data sets among disciplines; a summary of findings is presented in Exhibit 1. NSF has also taken into consideration studies led by the National Science Board (NSB); international developments, beginning with meetings in Budapest, Hungary, Bethesda, Maryland, and Berlin, Germany in 2002-3 and followed by a series of statements and policies, including policies for sharing of specific data types; and the four days of public comment organized by the National Research Council supported by all of the Federal agencies participating in this initiative in May 2013. Transcripts, videos, and submitted white papers are reported at: http://sites.nationalacademies.org/DBASSE/CurrentProjects/DBASSE_082378. Additional information on public consultation and notice is included in Sections 9.0 and 11.0.

This plan sets forth a framework for increasing access to the results of NSF-funded research and leverages existing NSF policies that provide for data sharing, data management plans, and evaluation, monitoring, and compliance. NSF will continue to identify additional approaches, involving public and private sector entities, and will continue efforts to improve public access to research data. NSF will explore, along with other agencies, how best to achieve improved public access, including data storage and preservation, discoverability, and reuse with a particular focus on data underlying the conclusions of peer-reviewed scientific publications resulting from federally funded scientific research.

Going forward, NSF has identified current and future opportunities for collaboration and public-private partnerships with other Federal agencies, commercial and not-for-profit publishers, universities and libraries, business interests[2], and other potential interested groups. NSF expects to maintain ongoing communication and consultation activities around key and emerging topics of concern to the research community. Because data management is dynamic and practices vary substantially across the broad

---

[1] http://www.whitehouse.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf

[2] For example, search engines supported by Google, Microsoft, and others.

range of scientific disciplines supported by NSF, the Foundation expects to engage in a wide and evolving set of activities appropriate to the needs and concerns of the different research communities. Some of the consultation and community outreach activities are discussed in Sections 3.2, 7.1.1, 7.2.2, 7.3.2, 7.4.2, 7.5, 9.0, and 11.0.

A crosswalk from the organization of this document to the criteria set forth in the OSTP memorandum is presented in Exhibit 2.

1.1 Agency mission

NSF is governed by the National Science Foundation Act of 1950, as amended (Public Law 81-507, the "Organic Act" or the "NSF Act"), which sets forth a mission: "to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense; and for other purposes."[3] As elaborated in the Foundation's strategic plan, *Investing in Science, Engineering, and Education for the Nation's Future; Strategic Plan for 2014-2018 (March 2014)*,[4] NSF meets two fundamental needs: fostering advanced research and developing the skilled knowledge workforce by seamlessly integrating the education of future scientists, engineers, and educators into the broad research portfolio in a strategy that enables ground-breaking discoveries and equips the future S&E workforce with relevant knowledge and experience. NSF believes that implementation of public access activities aligns with this mission and will reduce barriers to communication, foster S&E research, and provide a platform for innovation that will contribute to the economic prosperity and well-being of American citizens.

1.2 NSF strategic plan and public access objectives

Implementing public access requirements is consistent with the three strategic goals set forth in the Foundation's current strategic plan:

- *Transform the frontiers of science and engineering:* This goal emphasizes the seamless integration of fundamental research and education to enhance the ability of the Nation to meet new challenges and to provide new paradigms and capabilities for scientific, technological, and economic leadership in a rapidly changing, competitive global environment. Facilitating access to the results of NSF-funded research fosters advances in research and education and potentially enables translation of these results to applications.

- *Stimulate innovation and address societal needs through research and education:* This goal points to the tight linkage between NSF programs and societal needs where NSF makes a unique contribution through targeted solicitations and core programs, which focus the attention of the S&E research communities on fundamental aspects of high priority national challenges, including research and development on science, technology, engineering, and mathematics (STEM) education and learning to prepare a competent STEM workforce and a STEM-literate citizenry. Public access to NSF research results is threaded through all of these topics.

---

[3] The NSF Act is codified at 42 USC 1861 et seq.

[4] *Investing in Science, Engineering, and Education for the Nation's Future; Strategic Plan for 2014 – 2018* (March 2014), http://www.nsf.gov/about/performance/strategic_plan.jsp.

- *Excel as a Federal science agency:* This goal includes blending strong scientific leadership with robust organizational leadership and supporting the staff with essential information in a culture of continuous improvement to ensure effective, inclusive, and accountable programs and merit review processes. Public access goes to providing high levels of customer service and contributes to developing evaluation systems essential to innovation and improvement.

2.0 SCOPE

NSF will implement its public access requirements in stages. In the first implementation, the following products of NSF-funded research are in scope:

- Articles in peer-reviewed journals in which the research is funded wholly or in part by NSF through new awards resulting from proposals submitted, or due, on or after the effective date of the NSF *Proposal & Award Policies & Procedures Guide* (PAPPG) that will be issued in January 2016 (January 2016 effective date).

- Papers accepted as part of juried conference proceedings in which the research is funded wholly or in part by NSF through new awards resulting from proposals submitted, or due, on or after the January 2016 effective date.

- Articles in peer-reviewed journals and juried papers accepted as part of conference proceedings authored entirely or in part by employees of NSF and by Intergovernmental Personnel Act assignees to NSF on or after the January 2016 effective date.

- Data and associated outcomes that result from NSF-funded research and are subject to the existing Data Management Plan (DMP) requirement, implemented on January 18, 2011, and enumerated in the *Grant Proposal Guide* II.c.2.j and described in Section 3.2.

Definitions:

Final accepted peer-reviewed manuscript

- Author's final manuscript of a peer-reviewed paper accepted for journal publication, including all modifications from the peer-review process.

Final published article or version of record

- Publisher's authoritative copy of the paper, including all modifications from the publishing peer-review process, copyediting, stylistic edits, and formatting changes.

Research data

- As defined in the 2 CFR 200.315(e)(3):[5] "the recorded factual material commonly accepted in the scientific community as necessary to validate research findings." While the focus is on unclassified digital data as defined in the regulation, NSF understands that additional DMP

[5] 2 CFR 200.315(e)(3)—Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards.

guidance at the directorate, division, or program levels may include more nuanced criteria, appropriate to the scientific discipline and proposed activity (see Section 3.2).

In future implementations, NSF expects to expand the scope to include additional products such as technical reports, white papers, instructional materials and other items currently reported as part of the annual and final project reporting requirements.[6] This timing of future implementations will depend upon experience with the initial implementation; planning will begin no sooner than FY 2015. Expansion of scope and related requirements will be announced publicly through the annual process that NSF employs to make revisions to the PAPPG.[7] This revision process ensures that the research and higher education communities have sufficient time to comment and to prepare for changes in requirements (Section 11.0). NSF will also provide updates to OSTP and the Office of Management and Budget (OMB) twice yearly (Sections 12.0 and 13.0).

3.0 REQUIREMENTS

3.1 Publications

NSF will require that either the version of record or the final accepted peer-reviewed manuscript in peer-reviewed scholarly journals and papers in juried conference proceedings or transactions described in the scope above (Section 2.0) and resulting from new awards resulting from proposals submitted, or due, on or after the January 2016 effective date must:

- Be deposited in a public access compliant repository designated by NSF;
- Be available for download, reading, and analysis free of charge no later than 12 months after initial publication;
- Possess a minimum set of machine-readable metadata elements in a metadata record to be made available free of charge upon initial publication (Section 7.3.1);
- Be managed to ensure long-term preservation (Section 7.7); and
- Be reported in annual and final reports during the period of the award with a unique persistent identifier[8] that provides links to the full text of the publication as well as other metadata elements.

---

[6] NSF's implementation of the Research Performance Progress Report (RPPR) follows the government-wide standard established by the OSTP/OMB-approved RPPR policy. NSF allows NSF-funded researchers to report:

- Publications, conference papers, and presentations (includes journal publications; books or other non-periodical, one-time publications; and other publications, conference papers, and presentations);
- Website(s) or other Internet site(s);
- Technologies or techniques;
- Inventions, patent applications, and/or licenses; and
- Other products, such as data or databases, physical collections, audio or video products, software or NetWare, models, educational aids, or curricula, instruments, or equipment.

If there is nothing to report under a particular item, the researcher may state "Nothing to Report."

[7] http://www.nsf.gov/publications/pub_summ.jsp?ods_key=papp

[8] The U.S. Geological Survey defines persistent identifiers as "globally unique numeric and/or character strings that reference *a digital object. Persistent identifiers can be actionable in that they* enable a user to access the digital resource via a persistent link" (http://www.usgs.gov/datamanagement/preserve/persistentIDs.php). As noted by

Either the version of record or the final accepted peer-reviewed manuscript will be acceptable (see definitions in Section 2.0).

In the initial implementation, NSF has identified the Department of Energy's PAGES (Public Access Gateway for Energy and Science) system as its designated repository and will require NSF-funded authors to upload a copy of their journal articles or juried conference paper to the DOE PAGES repository in the PDF/A format, an open, non-proprietary standard (ISO 19005-1:2005). Either the final accepted version or the version of record may be submitted. NSF's award terms already require authors to make available copies of publications to the Cognizant Program Officers as part of the current reporting requirements. As described more fully in Sections 7.8 and 8.2, NSF will extend the current reporting system to enable automated compliance.

Future expansions, described in Section 7.3.1, may provide additional repository services. The capabilities offered by the PAGES system may also be augmented by services offered by third parties.

3.2 Data Management Plan

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 NSF funding should be stored and publicly accessible to search, retrieve, and analyze. NSF requires applicants for funding to prepare a DMP. The current requirement, which was a January 18, 2011 implementation of the Foundation's long-standing data-sharing policy, specifies that proposals must include a supplementary document of no more than two pages, labeled "Data Management Plan." This supplement should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results as stated in the *Grant Proposal Guide.* The DMP may address:

- The types of data, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies for access and sharing, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
- Policies and provisions for re-use, re-distribution, and the production of derivatives; and
- Plans for archiving data, samples, and other research products, and for preservation of access to them (*Grant Proposal Guide* II.c.2.j).

Davidson, "persistent identification of digital resources can play a vital role in enabling their accessibility and re-usability over time" (Davidson, J. (2006). "Persistent Identifiers". DCC Briefing Papers: Introduction to Curation. Edinburgh: Digital Curation Centre. Handle: 1842/3368. Available online: http://www.dcc.ac.uk/resources/briefing-papers/introduction-curation). A Digital Object Identifier (DOI) is an example of a persistent identifier. If the investigator does not have a publisher-supplied identifier at the time a document is deposited in the public access repository, the system will provide one at the conclusion of the submission process.

Guidance and data management requirements and plans specific to the directorate, office, division, program, or other NSF unit are available at: http://www.nsf.gov/bfa/dias/policy/dmp.jsp. Individual solicitations may also include specific guidance or requirements. NSF's FastLane system will not permit submission of a proposal that is missing a DMP.

3.2.1 Exceptions to the release of data

Small Business Innovation Research (SBIR)/Small Business Technology Transfer (STTR) proposals and any other proposal may allow for exceptions for proprietary or otherwise restricted data, including but not limited to personally identifiable information, business confidential information, security, among other concerns outlined in section 4.a. of the OSTP memo. Any such data management issues as well as conditions that might affect, delay, or limit data sharing should be discussed in the DMP. Coordination with the Cognizant Program Officer prior to submitting the proposal is also advised.[9] Other potential restrictions are described in Section 7.5.2.

3.2.2 DMP evaluation

DMPs are reviewed as an integral part of the merit review of the proposal, considered under Intellectual Merit or Broader Impacts or both, which are criteria established by the NSB. These criteria are described in the *Grant Proposal Guide III.A.2* and are used to review every proposal submitted to NSF. Results of the merit review process are not made public but are provided to the proposing investigators.

3.2.3 Data deposit and citation

Data that underlie the findings reported in a journal article or conference paper should be deposited in accordance with the policies of the publication and according to the procedures laid out in the DMP included in the proposal that led to the award on which the research is based. In the future, NSF will explore whether all data underlying published findings can be made available at the time of publication. The reference to the data cited should include the NSF award number and appropriate attribution as well as other attributes required by the publisher[10] or the repository where the material is deposited (e.g., metadata, persistent identifier, and so on). Effective January 2013, NSF permitted proposers to include citations to data sets in the Biographical Sketches in proposals, requiring (among other data elements) names of all authors, date of publication or release, and Universal Resource Locator (URL) or other persistent identifier (see *Grant Proposal Guide, Chapter II C.2.f.i (c)*).

---

[9] The current NSF SBIR Phase I solicitation is at http://nsf.gov/pubs/2013/nsf13546/nsf13546.htm with guidance provided in section 9.i.2. Data Management Plan. On exceptions, see guidance provided by the Directorate for Engineering at: http://nsf.gov/eng/general/ENG_DMP_Policy.pdf.

[10] NSF notes that the National Information Standards Organization (NISO) has begun to address the issue of publication and associated data through a new standard, JATS: Journal Article Tag Suite, ANSI/NISO Z39.96-2102, and a set of best practices for managing and representing supplemental journal article materials. See: National Information Standards Organization (NISO), JATS: Journal Article Tag Suite, ANSI/NISO Z39.96-2102 (Baltimore: NISO, 2012), pp. 20, 23 (http://www.niso.org/apps/group_public/download.php/8975/z39.96-2012.pdf); and National Information Standards Organization (NISO), Recommended practices for Online Supplemental Journal Article Materials, NISO RP-15-2103 (Baltimore: NISO, 2013), http://www.niso.org/apps/group_public/download.php/10055/RP-15-2013_Supplemental_Materials.pdf.

All data resulting from the research funded by the award, whether or not the data support a publication, should be deposited at the appropriate repository as explained in the DMP. Metadata associated with the data should conform to community standards and the requirements of the host repository. At a minimum, data elements should include acknowledgement of NSF support as well as the award number and appropriate attribution.

As implementation of NSF's public access initiative unfolds over the next years, NSF expects to explore a series of options to leverage existing data repositories, extend approaches already in use in the development of DMPs, develop standards for repositories and metadata in consultation with the community, and enhance reporting and evaluation procedures (see Sections 7.1.1, 7.3, and 7.4).

3.3 Costs associated with publications and data

NSF's data-sharing policy states: "Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing" (PAPPG's *Award & Administration Guide, Chapter VI.D.4*).

NSF allows proposers to request funds for the "costs of documenting, preparing, publishing or otherwise making available to others the findings and products of the work conducted under the grant. This generally includes the following types of activities: reports, reprints, page charges or other journal costs (except costs for prior or early publication); necessary illustrations; cleanup, documentation, storage and indexing of data and databases; development, documentation and debugging of software; and storage, preservation, documentation, indexing, etc., of physical specimens, collections or fabricated items" (see *Grant Proposal Guide, Chapter II.C.2.g.vi.(b)*). There will be no charge to investigators to deposit their articles (whether final accepted version or version of record).

4.0 APPLICABILITY

These requirements will apply to new awards resulting from proposals submitted, or due, on or after the January 2016 effective date. In addition, these requirements apply to all approved NSF Individual Research and Development (IR/D) plans for NSF employees and Intergovernmental Personnel Act assignees.

NSF requires awardees to adhere to public access requirements as described in this plan for all research that it supports, even when the research underlying the publication and generation of data is supported by multiple funders (including other Federal and state agencies, private funders, and international sponsors).

5.0 AUTHORITY

NSF is an independent Federal agency created by the NSF Act, 42 USC 1861, et seq. The NSF Act authorizes NSF to make discretionary grants to support basic scientific and engineering research and education programs and to impose conditions on those grants. Other relevant legal and policy authorities include:

- The America COMPETES Reauthorization Act of 2010 (P.L. 111-358)
- 2 CFR Part 200
- 45 CFR Part 602
- NSF award terms and conditions, policy guides and manuals, and other policy documents

6.0 ROLES AND RESPONSIBILITIES

National Science Board

- Approves all NSF policies required to implement this plan.

Deputy Director

- Establishes the Public Access Working Group to provide oversight and guidance to the Foundation for on-going activities associated with the implementation of this plan.

Public Access Working Group

- Provides oversight and guidance on the implementation of this plan; informs senior leadership on progress, resource needs, and other issues that may influence the plan's implementation.

Program Staff

- Serve as the primary interface with Principal Investigators (PI) and assist in overseeing compliance with the plan's requirements.

Budget, Finance, & Award Management (BFA)

- Develops and implements NSF-specific public access policy.

Office of Information & Resource Management (OIRM)

- Ensures IT systems support public access policies and enable compliance and compliance checking.

Awardee Organizations

- Ensure that authors and data generators are aware of and comply with NSF's public access requirements.
- Ensure compliance with all award terms and conditions of awards, including the submission of final peer-reviewed manuscripts that arise directly from the organization's awards and release of any associated data.

Principal Investigators and Authors

- Work with the publisher before any publication rights are transferred to ensure that all NSF's public access requirements can be met.

- Ensure that any agreement signed with a publisher preserves the author's ability to comply with NSF's public access requirements.

7.0 IMPLEMENTATION

NSF proposes to implement its public access plan incrementally and expects to manage data and publications in a coherent, integrated framework that can accommodate products of NSF-funded research, starting with but not limited to data and journal publications. NSF expects to build on current practices and to leverage existing and future investments in many sectors through interagency agreements and public/private partnerships. The potential roles of these different partners are indicated in the following sections.

In the initial implementation, NSF has identified the DOE PAGES system to support managing journal articles and juried conference papers. In the future, NSF may add additional partners and repository services in a federated system.

7.1 Planning and communication

NSF will undertake a series of near- and long-term activities and pilots to develop an operational system with the following objectives:

- A system will be in place that will enable NSF-supported investigators to make their articles (either final accepted manuscript or version of record) available to the public on a voluntary basis by the end of calendar 2015; mandatory deposit in a designated public access repository will become effective for awards resulting from proposals submitted on or after the January 2016 effective date. Either the version of record or final accepted version will be acceptable.
- Communication will be maintained with the research communities and concerned stakeholder groups to identify areas where guidance or adaptations may be required.
- Changes in the system that may result in guidance associated with DMPs will take place incrementally after consultation with the research community and will be implemented no earlier than FY 2016. Several cycles of changes may be required.
- The system will be flexible and extensible to allow for continued evolution, to expand its scope, and to respond to changing technologies and infrastructure capabilities.

For planning purposes, a start date of January 1, 2015, has been assumed.

7.1.1 Maintain ongoing discussions with concerned stakeholder groups and other Federal agencies

NSF will continue to engage stakeholder groups and coordinate with other Federal agencies to identify infrastructure capabilities, resolve outstanding and shared concerns, and develop best practices and standards. NSF has already initiated discussions with the National Research Council to organize the Forum on Open Science, a discussion forum or roundtable -- modeled on the Institute of Medicine (IOM) Research Forum -- at which representatives from industry, the academic community, and possibly others,[11] might meet on a regular basis to discuss future developments in the area of public access to

[11] The range of communities and stakeholders engaged in aspects of public access policies and implementations shifts. NSF recognizes that the current landscape encompasses publishers, higher education, libraries, hard- and

data and publications and related issues. The goal would be to ensure that these stakeholder communities would have regular avenues to communicate, not just between stakeholders and agencies, but between themselves; for example, between the universities and the publishers. The first meeting was held on June 13, 2014.

NSF will also coordinate with other Federal agencies so that emerging standards and guidelines may be harmonized across the research agencies, thus reducing the likelihood of inconsistent requirements and associated confusion and burden for awardees and investigators (see Section 10.0). In parallel, NSF will continue outreach and consultation with the research communities associated with developing guidelines for repositories and other activities (see Sections 7.3 and 7.4).

7.1.2 Implement Memorandum of Understanding (MOU) or similar agreement with DOE for repository services for journal publications (initial implementation)

NSF recognizes that developing the full public access program will require several years and a series of implementation steps. In the initial implementation, NSF has entered into a non-exclusive relationship for repository services with the DOE that will enable authors whose work is subject to the NSF public access requirement to make their articles publicly available in the DOE repository system (PAGES). A MOU was executed in 2014. This agreement sets forth the terms of the cooperation between the two Federal agencies and provides for pilots and testing required to ensure that relevant information from the two systems flows correctly. The system will be available for NSF-funded authors to use on a voluntary basis by the end of calendar 2015.

7.1.3 Implement changes to NSF's guidance documents, procedures, and electronic administrative systems

NSF anticipates changes to current documents, procedures, and systems, including:

- Applicable NSF policy documents, including the PAPPG and *Proposal and Award Manual* (PAM), as needed;
- NSF's Grant General Conditions;
- Independent Research and Development (IR/D) Plans;
- NSF's electronic administrative systems to support new public access requirements; and
- NSF's DMP guidance and requirements, as specified in the PAPPG, based on consultation with the research community.

Consistent with OMB's Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Award (UG), effective December 26, 2014, NSF will hold a Federal purpose license for articles published based on its research awards.

NSF will report regularly to the NSB on the progress of implementation, and proposed changes to the Foundation's policies will be discussed and submitted for the Board's approval. Changes in policy will be announced and made publicly available for comment in accordance with the public notice process described in Section 11.0, and will initially cover items specified in Section 2.0. Coordination with OSTP and OMB is described in Section 12.0 and included in the timeline outlined in Section 13.0.

---

software companies, as well as users of federally funded research results and civil society groups. However, this landscape may change, and NSF expects the membership of this forum to evolve accordingly.

7.1.4 Develop a communications plan and Frequently Asked Questions (FAQs)

The communications plan will comprise an evolving set of activities designed to inform and engage internal NSF audiences (directorate and division leadership, program staff, etc.) and external audiences (research communities and other stakeholder groups) as well as other public agencies. A key element of the plan will be creation and maintenance of FAQs, which clarify details and outline guidance as the implementation of the plan unfolds.

7.1.5 Implement changes to NSF website (nsf.gov)

The NSF webpages will be revised to include, at a minimum, information on the progress of the public access initiative (for example, the final approved plan, FAQs); guidance for awardees and investigators; and access to search capabilities. After the public access requirement goes into effect, the number of NSF-supported journal articles and conference papers submitted to NSF's public access repository will be reported on the website. This information will be updated on an annual basis.

Construction of the website began in FY 2014. The site will be launched no later than April 2015 with updates as the initiative progresses.

7.2 Submission

"Submission" refers to the steps the authors or their designated representatives (acting on behalf of the awardee institutions) take to deposit their journal articles and juried conference papers or digital scientific data into a public access repository. The investigator is then responsible for reporting that deposit back to NSF as a condition of the award. During the period of the award, this information is included in the annual and final reports.

Any author who has had NSF funding or currently has NSF funding that predates the public access requirement (to go into effect in FY 2016) may submit legacy publications to the public access system, which will allow these authors to link the publications to their NSF awards.

7.2.1 Submission: Publications

NSF will require awardee institutions to ensure that authors of articles and papers that fall within the scope of this plan (as defined in Section 2.0) deposit copies of the author's final accepted peer-reviewed manuscript or the version of record[12] in the PDF/A standard in a repository maintained on behalf of NSF by DOE. The system integration, described in more detail in Section 7.8, allows seamless transition from NSF systems to the DOE infrastructure. The DOE system also supports non-text formats (images, video, and supporting digital data).

DOE's system meets applicable Federal requirements of Section 508 of the Rehabilitation Act of 1973. In the event that material is discovered that is not Section 508-compliant, DOE maintains the capability to make the material compliant. Additional discussion is provided in Section 7.7. NSF will provide guidance for authors for converting their documents into the required format in the FAQs on the NSF website (Section 7.1.4).

[12] Ability to submit the version of record will be subject to the policies of the journal.

In future implementations, NSF will explore options that would allow authors to deposit copies in repositories maintained by other Federal agencies or by other public/private third parties that meet all of the criteria set forth in the OSTP memorandum and to report that submission back to NSF. Further explanation is provided in Section 7.3.1.

7.2.2 Submission: Data

NSF will retain its current DMP requirements. Under these guidelines, awardees are required to set forth plans for archiving data, samples, and other research products, and for preservation of access to them, as appropriate (see Section 3.2.1). Investigators report on the progress of these activities in their annual and final reports.

NSF understands that practices surrounding data deposit vary substantially based on the discipline and type of data (simulation, modeling, observational, instrument-driven, laboratory, and so on): "What constitutes reasonable data management and access will be determined by the community of interest through the process of peer review and program management. In many cases, these standards already exist, but are likely to evolve as new technologies and resources become available" (*Data Management and Sharing Frequently Asked Questions*, updated November 30, 2010).

Over the next several years, NSF will consult with the research communities to develop discipline-specific guidance and best practices. Any changes in policy and related guidance would be announced no earlier than March 2016 for implementation in the *PAPPG* no earlier than January 2017.

7.3 Management

"Management" refers to the policies and procedures that the repository has in place to ensure appropriate treatment of the material and to enable the material to become accessible to end users. Specific functions, such as preservation, discovery, search, and access are addressed in Sections 7.3, 7.4, 7.6, and 7.7.

7.3.1 Management: Publications

In the initial implementation, NSF will work with DOE to enable NSF-funded investigators to make their articles (either the final accepted version or the version of record) publicly available in a repository that meets the criteria set forth in the OSTP memorandum of February 22, 2013, including public availability of appropriate metadata, Section 508 compliance, preservation (see Section 7.7) and guidance for appropriate use. During the period of the award, investigators will be able to report submissions back to NSF by providing a unique persistent identifier as part of their annual and final reports. Persistent identifiers may be the widely used digital object identifier (DOI) or another unique identifier system created for NSF by the host repository. After the period of the award, investigators can continue to upload copies of their journal articles and conference papers to the repository hosted by DOE and map them to the relevant awards.

The DOE PAGES system offers centralized metadata and indexing together with the flexibility of a distributed system of linking to authoritative copies of the full-text of the material (either the final accepted manuscript or the publisher's version of record). PAGES also accepts manual upload of PDF/A-compliant documents, which will be required of all NSF-funded authors. In the initial implementation, NSF expects to rely on DOE's hosting capability, possibly augmented by other services in future implementations.

In later implementations, NSF expects to add additional partners (discussed in the next paragraphs), leveraging DOE PAGES' capability to maintain centralized metadata records and link to other repository systems. This enables NSF to maintain management control of the information without unnecessary duplication of submission and the associated burden on the awardees and investigators, or the risk of multiple and inconsistent versions. A list of designated public access repository systems will be maintained on the NSF website (Sections 7.1.4 and 7.1.5).

Public/private partnerships. Various groups (including publishers, Federal agencies, and academic libraries) are actively working on initiatives that will maintain consistent metadata (CrossRef[13]); identification of agency funding (FundRef[14]); identification of rights and Open Access status as proposed by NISO/NFAIS[15]; and consistent identification of authors and other contributors (ORCID[16], ResearcherID[17]). The DOEPAGES system relies on the concept of "best available version" and takes advantage of the publishers' consolidated metadata repository, CrossRef, to link metadata records in DOE's system to full-text versions of papers maintained by cooperating publishers.

At the conclusion of the embargo period, end users who search for material through the PAGES interface will obtain a list of search results that will include pointers to the final accepted manuscript or to the version of record, depending on the publisher's public access policies. Thus, end users have access to the most authoritative public access version of a given item. In the event a publisher's site becomes inaccessible or if the version of record is not made available for public access, the item housed in DOE's dark archive[18] is illuminated and becomes available to the end user.

NSF supports these public-private, cross-agency activities and will incorporate new infrastructure capabilities as they mature. The Foundation's incremental approach allows NSF to evolve its systems as new capabilities become available. Over time, NSF expects to expand the range of eligible repositories as follows:

- Systems operated by other Federal agencies. NSF investigators typically have multiple funding sources. Since a given item may be based on funding from more than one agency, NSF expects to allow submissions of articles and papers to public access repositories operated by other Federal agencies that meet the standards of the OSTP February 22, 2013, memorandum and for which the investigator can provide a persistent identifier as an element in annual or final reports. Implementation of this expansion is likely to begin no earlier than FY 2016.

- Systems operated by third parties. A coalition of publishers (Clearinghouse for Open Research of the United States (CHORUS))[19] and a group of institutions of higher education in combination

[13] http://www.crossref.org/
[14] http://www.crossref.org/fundref/
[15] http://www.niso.org/news/pr/view?item_key=d2e5f409bc6af6b7f504a10edf0329203ffec6f9
[16] http://orcid.org/
[17] ResearcherID, http://www.researcherid.com/Home.action
[18] "Dark archive" is a term used for a digital repository to which access is restricted and which is often used for secure storage and back up; see Blue Ribbon Task Force on Sustainable Digital Preservation and Access, *Sustainable Economics for a Digital Planet: Ensuring Long-Term Access to Digital Information* (February 2010), http://brtf.sdsc.edu/biblio/BRTF_Final_Report.pdf, p. 107.
[19] Association of American Publishers, Understanding CHORUS, Clearinghouse for the Open Research of the United States, June 5, 2013, http://www.publishers.org/press/107/

with the research libraries under the leadership of AAU/ARL/APLU (e.g., SHARE)[20] have also proposed potential solutions. NSF will continue discussions with these groups (and others that may come forward) with the expectation that information from them may be incorporated into subsequent implementations of the proposed system.

The ability to expand NSF's system to accommodate multiple repositories in a federated system depends both on the capabilities of potential partner systems and on the level of technological complexity that NSF's internal systems can efficiently support. In collaboration with other Federal agencies and interested parties, NSF will develop criteria for eligible repositories, based on the criteria set forth in the OSTP memorandum, and will provide appropriate guidance for awardees and investigators on the website.

NSF may initiate these discussions as early as FY 2016.

Metadata. To facilitate integration with NSF's internal administrative systems, support simple searches, maintain the connection between the metadata record and the full content of the article, preserve the attribution to the author and to the original publisher, and provide access to a description of the material independently of its embargo status, NSF will require a minimum of eight metadata fields in a record that will become available free of charge upon initial publication of the article:

1. Persistent identifier (e.g., DOI, NSF identifier created by DOE, etc.) , which links the metadata record with the associated content;
2. Author names(s) with associated persistent identifiers (such as the NSF investigator ID or, eventually, ORCID or a similar system);
3. Title of the article;
4. Journal or serial title, preferably with identifiers (e.g., ISSN);
5. Name(s) of agency/agencies and award number(s);
6. Representation of intellectual property rights [21];
7. Link(s) to underlying data including but not limited to the Supplementary Material published with the journal article itself; and
8. Author- or publisher-supplied Abstract[22] .

7.3.2 Management: Data

As described elsewhere, NSF will continue to refine implementation of the current DMP requirement in consultation with the research communities. In general, a DMP must identify relevant standards and data archiving plans, which vary by discipline and research topic, and the sufficiency of the DMPs is evaluated during merit review, as described in Section 3.2.2. Specific NSF guidance for DMPs in the *Grant Proposal Guide* includes:

---

[20] Association of Research Libraries, Shared Access Research Ecosystem (SHARE)Proposed by ARL, AAU, APLU, June 7, 2013, http://www.arl.org/news/arl-news/2773-shared-access-research-ecosystem-proposed-by-aau-aplu-arl

[21] Encoding of rights and Open Access status has been proposed by NISO/NFAIS; http://www.niso.org/news/pr/view?item_key=d2e5f409bc6af6b7f504a10edf0329203ffec6f9. NSF will monitor the progress of this group.

[22] NSF understands that publishers may not release the abstracts that they create. Therefore, this field may include an author-supplied abstract that may not be identical to the abstract that the publisher eventually releases.

- The standards to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
- Policies and provisions for re-use, re-distribution, and the production of derivatives; and
- Plans for archiving data, samples, and other research products, and for preservation of access to them (*Grant Proposal Guide* II.c.2.j).

Metadata. NSF funds a wide range of sciences, and the data resulting from NSF-funded research are highly heterogeneous. NSF does not currently specify a single metadata standard. However, any acceptable minimum set of data elements would include the names of all authors, date of publication or release, and Universal Resource Locator (URL) or other persistent identifier, as required by Biographical Sketches in proposals (Section 3.2.3 and *Grant Proposal Guide, Chapter II C.2.f.i (c)*). NSF notes that metadata and citation standards for data are the subject of continued investigation and discussion through several venues (for example, Research Data Alliance[23], Digital Curation Centre, Force11[24], CENDI, Board on Research Data and Information (BRDI), and others) and will continue to monitor developments in the research communities.

Repositories. Managing data is complex and will require further exploration and development, given their inherent heterogeneity; potentially very substantial size; and the challenges of addressing data resulting from modeling and simulation, and of streaming data generated from a sensor or experiment. Rarely does NSF expect that retention of all data that are streamed from an instrument or created in the course of an experiment or survey will be required.

NSF encourages development of broad guidelines and communities of practice around data description and management, including appraisal, retention, and disposal, which are reflected in repositories' policies and in individual DMPs.[25] NSF believes that management of large investments, such as centers, major facilities, and cross-directorate and interagency initiatives, are consistent with data management objectives, but will review these entities to insure that expectations about data management are communicated clearly. In addition, and based on community consultation, NSF may develop guidance for program officers and panelists to assist them in reviewing DMPs during the merit review process.

---

[23] NSF supports the Research Data Alliance through NSF Award Number 1249473 and NSF Award Number 1349002.

[24] We note that Force11 has taken an active role in data management and data citation; see https://www.force11.org/.

[25] EarthCube (http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504780), a joint effort by the NSF Directorate for Geosciences and the Directorate for Computer and Information Science and Engineering (CISE), is one example of a community-driven activity to develop cyberinfrastructure, including data and data management (http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=504780). The Directorate for Social, Behavioral and Economic (SBE) Sciences provides support (with other Federal agencies) for major databases that are central to SBE research; see General Social Survey (http://www3.norc.org/gss+website/), Panel Study of Income Dynamics (http://psidonline.isr.umich.edu/), and American National Election Studies (http://www.electionstudies.org/). Other examples include: (i) The iDigBio hub (https://www.idigbio.org/), an ADBC program (Advancing Digitization for Biological Collections) that supports activities for the implementation of the Network Integrated Biocollections Alliance (NIBA), which is a coordinated, community-led effort to integrate and make available via the web data from biological specimen collections; and (ii) GoLife, a new cross-directorate program funded by the Directorate for Biological Sciences and the Directorate for Geosciences, designed to develop integrated data and information resources to facilitate comparative biology research across organismal and temporal scales.

Over the next three years, NSF will consult with the community and with other Federal agencies and facilitate the establishment of standards for metadata and repository systems. NSF understands that issues of description and access as well as cost, use, and preservation are all elements of these discussions. Possible investments include:

- Workshop and pilot activities on data archiving by journals and institutional and disciplinary repositories, including developing best practices and procedures for identifying and providing attribution; preserving a balance between long-term preservation and costs, providing access, and managing and deaccessioning data, and the associated costs;
- Workshops and pilot activities on standards development, networking, and linkage activities associated with data and data repositories;
- Pilots in data infrastructure and sustainable operation; and
- Challenge or prize competition(s) to promote development of persistent identifier approaches and evaluation of their utility.

7.4 Discovery, search, and access

Discovery, search and access are processes by which end users find, display, download, analyze, and use information. NSF seeks to insure that the material is identified, described, and managed to facilitate its creative and appropriate use as outlined by the objectives of the OSTP memorandum.

7.4.1 Discoverability, search, and access: Publications

NSF observes that end users discover information in many ways, frequently by going to a commercial third-party search system such as Bing, Google, or GoogleScholar, or by returning to a site where they are accustomed to finding relevant information.[26] Therefore, NSF expects to make information about NSF awards and related publications available through several public and private mechanisms, taking advantage of users' expectations as well as the widespread use of commercial search systems.

Initial implementation. NSF will employ four mechanisms to enable end users to search NSF awards and related publications, access the material, and download it for local use. Some of these capabilities already exist.

- NSF currently allows public search of active and inactive awards and related publications through its 508-compliant website (www.nsf.gov) and through research.gov. Users can download results in several formats for further analysis.

[26] Ways that people search is a rich vein of research. For an introduction, see Lea Winerman, "Tracking the scent of information," *Monitor on Psychology*, March 2013, Vol. 42, No. 3 (http://www.apa.org/monitor/2012/03/information.aspx). On the importance of search engines, see Simone Kortekaas, Thinking the unthinkable: a library without a catalogue – Reconsidering the future of discovery tools for Utrecht University library, Association of European Research Libraries, Re-inventing the Library for the Future, 4 September 2012 (http://www.libereurope.eu/blog/thinking-the-unthinkable-a-library-without-a-catalogue-reconsidering-the-future-of-discovery-to). Similar findings concerning the broad and heterogeneous character of investigators' search strategies and the importance of search engines are discussed in University of Tennessee and CIBER Research Ltd., Trust and Authority in Scholarly Communications in the Light of the Digital Transition, Final Report, Prepared for the Alfred P. Sloan Foundation, December 2013, http://ciber-research.eu/download/20140115-Trust_Final_Report.pdf.

- NSF will point from the NSF website to a search capability for NSF-funded publications on the NSF public access repository that will ensure users can read, download, and analyze in digital form final author-submitted peer-reviewed manuscripts or final published documents from the publishers' websites, if the version of record is made available free of charge after the period of the embargo. This service will be 508 compliant. DOE builds its search capability on the metadata provided to the system by the author and by indexing the full-text of the material housed locally and on the publishers' websites to which the PAGES metadata points, thus enhancing the results.

- A 508-compliant search capability will be required of any third party service (e.g., CHORUS, SHARE) with which NSF enters into an agreement. These third party services must also ensure that users can read, download, and analyze material in digital form as a condition of the public/private partnership. As these services are brought online, NSF will direct users to them from the NSF website.

- Metadata provide topical descriptions as well as information on identity and attribution, terms and conditions, and other material critical for enabling appropriate use. Metadata records will be made available to third-party crawlers to enable NSF-funded material to be included in these services (e.g., Bing, GoogleScholar, etc.).

Future implementations. NSF's approach recognizes that search is an area in which there is substantial activity and innovation in the research community and in commercial applications and services.[27] These services are increasingly providing platforms for collaboration.[28] In the proposed federated system that NSF envisions, search capabilities would be available at the partner nodes (e.g., DOE, other Federal agencies, third party systems hosted by publishers and universities/academic libraries) as well as through third parties (Google, Bing, and so on).

The strategy of developing a federated multi-organization repository system with public access to NSF-funded publications implies several search-related technical requirements for participating organizations:

- Support for standardized publication metadata, e.g., Dublin Core, MARC, etc. Ideally, a single metadata standard would be adopted, but more realistically it should be possible to support a small number of metadata standards.
- Support for a standard query language, e.g., SQL, Z39.50, Datalog, etc.
- Support for a standard application program interface (API), e.g., REST API, to access various functions on the metadata and/or document collection.
- Support for a standard data exchange protocol, e.g., HTML, FTP, and so on (for the actual metadata and documents).

These technical requirements will be included in the discussions of criteria for repository systems described in Section 7.3.1.

[27] NSF supports research into discovery, search, and retrieval through several programs in the CISE Directorate.
[28] For example, see ResearchGate (http://www.researchgate.net/); Mendeley (http://www.mendeley.com/); and Zotero (https://www.zotero.org/).

NSF is aware that individual publishers and library systems are experimenting with new approaches to presenting information, linking publications to data, and providing pointers to repository systems. NSF proposes to foster these developments and their use by ensuring consistent and predictable access to the underlying information, thus providing a platform for creativity and innovation.

Interoperability. In the initial implementation, NSF will rely on the standards established by the DOE PAGES system, which makes metadata available in widely used formats, including MLA, APA, Chicago, and Bibtex. The full-text of materials stored in the DOE archive will be maintained in nonproprietary formats, such as PDF/A. Individual content holders, such as publishers, can be monitored within the PAGES system for compliance with full-text interoperability standards, download capabilities, and Section 508 compliance. In cases where externally held content does not meet requisite standards, DOE will link to corresponding content held in its dark archive. If NSF establishes an independent relationship with third parties (such as CHORUS or SHARE), methods for monitoring compliance with interoperability requirements for both metadata and the full-text will be conditions of such future partnerships.

7.4.2 Discoverability, search, and access: Data

Current NSF guidance requires applicants for funding to explain in their DMPs "policies for access and sharing including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements" (*Grant Proposal Guide* II.c.2.j). Practices vary across disciplines and directorates, reflecting the heterogeneity of science and scientists that NSF supports. Guidance is issued at different levels of the Foundation and the adequacy of proposed DMPs is reviewed as part of the merit review process of individual awards, as explained in section 3.2.2.

To increase access to data resulting from NSF funding, over the next three years, NSF intends to:

- Modify internal systems and the NSF public web page to support search of data sets as well as publications associated with NSF awards (see Section 7.4.1);

- Consult with the community and with other Federal agencies and facilitate the establishment of best practices and standards for identification, description, and citation of resources to enable discovery and terms of use;

- Consult with the community and with other Federal agencies and facilitate the establishment of best practices and standards for repositories to ensure appropriate access and use;

- Develop guidance concerning management of and access to sensitive information building on expertise already in the Foundation (for example, National Center for Science and Engineering Statistics (NCSES); Directorate for Education and Human Resources; SBE Directorate; CISE Directorate); and

- Monitor ongoing consultation efforts at other Federal agencies. NSF will explore, along with other Agencies, how best to achieve improved public access, including data storage and preservation, discoverability, and reuse with a particular focus on data underlying the conclusions of peer-reviewed scientific publications resulting from federally funded scientific research.

7.5 Embargo and approach to changing its length

7.5.1 Embargo: Publications

NSF will follow the recommended guideline of an embargo of up to 12 months following initial publication. Modification of the 12-month guideline would constitute a change in NSF policies and procedures and would, therefore, be subject to the annual cycle of revision and public notice, described in Section 11.0.

DOE provides a mechanism for providing input into changing the embargo (or administrative interval). For example, users are encouraged to access the NSF PAGES Feedback page, which will provide multiple channels (email, telephone, and U.S. postal service). NSF will also make available an email contact (publicaccess@nsf.gov) through its webpage (www.nsf.gov/news/special_reports/public_access/) and will provide guidance on preparing a petition for extending the period of the embargo based on continuing discussions with concerned groups and Federal agencies. With regard to petitions for changing the embargo, such petitions should be evidence-based, that is, factually and statistically based evidence that a change in NSF's administrative interval will more effectively promote the quality and sustainability of scholarly publications while meeting the objectives of public access. In considering such evidence, NSF will work with other federal science agencies to promote consistent implementation of embargoes for specific scientific fields.

Elements of a technically sufficient petition might include:

- Definition of the stakeholder group that is submitting the petition;
- Evidence that the existing 12-month guideline is inconsistent with the objectives articulated in the OSTP memorandum;
- Proposed new period of embargo/administrative delay and a justification for that period; and
- The scientific field(s) and the titles of the peer-reviewed scholarly journals to be covered by the proposed waiver and a justification for their inclusion.

Additional criteria and considerations, including appropriate evidence, may be added after consultation with the community and included in the guidance.

NSF anticipates that petitions will be reviewed by the NSF Public Access Working Group (as defined in Section 6.0) and consultation with other Federal agencies undertaken as necessary and appropriate. Waiver actions will be reported to OSTP and OMB as part of the Foundation's semi-annual reporting (Section 12.0).

NSF recognizes many scientific societies and publishers are concerned about the impact of a delay on their subscription-based journals and related business models. NSF also recognizes that potential economic harm to publishers must be weighed against other public access objectives and science research goals.

7.5.2 Embargo: Data

Practices governing use of embargos and delayed data release vary widely across the research communities supported by NSF and should be discussed as part of the DMP. For large-scale projects that

are supported primarily to generate data for community use, the timing of release will be part of the award terms and conditions and clearly stated in the public award abstracts.

NSF recognizes that some classes of data, particularly those that relate to human subjects, education, personally identifiable information, national security, or proprietary interests, may be subject to restrictions. Such restrictions must be described in the DMP and changes addressed in annual and final reports.

7.6 Bulk downloads of content

NSF understands that investigators increasingly wish to download entire corpora of articles in bulk to support computational analysis or to develop more sophisticated algorithms. NSF intends to enable such uses while protecting the integrity of the scientific record from unauthorized redistribution of scholarly content. NSF already offers bulk download of information resulting from search of NSF active and inactive awards and related publications in several formats (CSV, EXCEL, HTML). NSF does not permit download of copyrighted or otherwise restricted information.

In the initial implementation based on the DOE's PAGES system, the distributed nature of DOE's PAGES full-text content renders *unauthorized* mass download and redistribution of its material inherently difficult. For the full-text content it hosts, DOE enforces a download limit and posts appropriate fair use policies. Users who search DOE-hosted content are able to follow links to the full-text and may build a collection appropriate for analysis, subject to the fair use policies instituted by the hosting repository and the rights associated with the content (indicated in the metadata or the license attached to the document).[29]

Measures to enable appropriate use of material together with security features necessary to protect the integrity of the system and to discourage unauthorized access and use will be taken into account when identifying partners and developing standards for eligible repositories, as described in Section 7.3.1.

7.7 Preservation

NSF will require NSF-funded authors to deposit a copy of their journal articles and juried conference papers in the DOE-hosted repository in PDF/A (Section 3.1) and will employ DOE's archiving and preservation capabilities to preserve copies of them, which can be made available to end users free of charge.

DOE stores and preserves the information in a dark archive[30] in a climate-controlled, appropriate environment in Oak Ridge, Tennessee, with redundant, backup systems located in geographically distinct locations. DOE accommodates both the widely used non-proprietary PDF and PDF/A formats and can convert material in PDF to PDF/A, should the need arise.

As described in Sections 3.2 and 7.3.2, NSF requires applicants for funding to address archiving and preservation in their DMPs, according to community-based best practices and standards; these plans are evaluated during the merit review process Strategies for providing long-term storage and preservation will be a requirement for any future NSF-designated repository system whether for data or publications

[29] NSF recognizes that guidance for authors about appropriate rights management will be important and will incorporate such guidance in the FAQs on the NSF website.
[30] "Dark archive" is defined in footnote 19.

(Sections 7.3.1 and 7.3.2). There are several technical strategies for achieving long-term preservation including redundancy, dark archives, secure data centers, and so on. In general, good practice calls for duplicating the collection at a geographically distinct location and for regular monitoring and format migration, given exigencies of media degradation and format obsolescence.

Preservation of digital content continues to be an area of active research,[31] and NSF remains open to different approaches to achieving long-term storage and preservation. NSF also encourages the disciplinary communities to develop best practice and other guidance documents to inform policies on data appraisal, retention, and deaccessioning.

7.8 Integration into NSF pre- and post-award systems

As part of the Foundation's ongoing enterprise modernization goals for its information technology (IT) systems, NSF has already begun planning for the internal system configuration that seamlessly integrates internal pre-award proposal submission and post-award reporting and management systems with the infrastructure to be maintained by DOE on behalf of NSF. In the eventual configuration (which will require several years to implement fully), only journal articles and juried conference papers that have been submitted to the public access repository will be included in annual and final reports and therefore, be visible to the Cognizant Program Officers. For the present, reporting other research products (e.g., websites, book chapters, software, and so on) will be unaffected.

The initial conceptual integration architecture was completed by September 30, 2013, and the advanced conceptual architecture based on integration with DOE's systems was completed in March 2014. NSF developed detailed plans, including proof of concept, tests, and pilots, in the spring and summer of 2014. Ongoing changes to the internal systems are cataloged and scheduled for implementation accordingly, along with changes to existing award and report search capabilities. NSF anticipates standing up the system for voluntary deposit by the end of calendar year 2015. Since the public access requirement will not be applied retrospectively, NSF anticipates it will be up to five years before all active awards, which include multiyear awards made prior to the January 2016 effective date, will be subject to the public access requirement.

Over a period of years, the system will also be expanded to include other outputs of NSF-funded research as already laid out in the RPPR, beginning with identifiers and citations to data sets, based on what has been learned from managing journal articles and juried conference papers. NSF notes that DOE's system already handles diverse content. Future changes may also expand template-driven approaches to preparing DMPs, based on consultations with the research communities as laid out in Sections 7.3.2 and 7.4.2, to enable more consistent evaluation as well as post-award management. These templates would be developed in consultation with the research community and will leverage the work that has already been done by several university and library systems.[32]

---

[31] NSF's CISE directorate funds research in digital preservation.

[32] For example, see the California Digital Library DMP Tool at: https://dmp.cdlib.org/; services offered by The Johns Hopkins University at: http://dmp.data.jhu.edu/; and services offered by Indiana University at: http://www.libraries.iub.edu/index.php?pageId=1003264.

8.0 METRICS AND EVALUATION, COMPLIANCE, AND ENFORCEMENT

8.1 Metrics and evaluation

NSF will develop measures of compliance at the agency level, compliance at the award level, and determination of non-compliance. In addition, NSF will set compliance goals, coordinate with the National Archives and Records Administration (NARA) to ensure compliance with applicable public records requirements, and undertake additional relevant activities to ensure effective evaluation. NSF will coordinate with other agency-wide evaluation activities. NSF will post the number of publications submitted to the public access repository annually on the NSF website (see Section 7.1.5).

8.2 Compliance

Deposit of journal articles and juried conference papers in the NSF public access repository will ensure public access, subject to the terms of the embargo (if any). As described in Section 7.8, the system will be configured to require submission of journal articles and juried conference papers as prerequisite for preparing annual and final reports. Only journal articles and juried conference papers that have been deposited in the public access repository can be included in annual and final reports, and Cognizant Program Officers will only see public access compliant journal articles and juried conference papers. Thus, the system ensures compliance with the public access requirement as a function of reporting.

DMPs are required as a component of proposal submission (*Grant Proposal Guide* II.c.2.j). NSF's FastLane system will not permit submission of a proposal that does not include a Data Management Plan. Proposals that do not include a DMP are returned without review. Thus, all active awards have met the DMP requirement.

During the period of the award, investigators currently report the management of data as part of the annual and final reports. Large complex projects that are governed by cooperative agreements (rather than grants) include data release among the award terms and conditions. These are currently monitored by Cognizant Program Officers who also review and approve the annual and final reports.

Based on consultation with the research communities, NSF expects to move toward a more structured system over the next several years that will support automated checking of compliance with DMP requirements through annual and final reports. This may entail the use of identifiers and template-driven approaches to preparing DMPs.

8.3 Enforcement

Enforcement of the public access requirement for journal publications and juried conference papers has been built into the deposit and reporting systems, with minimal additional burden for either investigators or program staff. Similarly, preparing a DMP has been integrated into the proposal submission process and any proposal that does not include the DMP is returned without review.

Consistent with long-standing policies and internal processes, NSF has relied on Cognizant Program Officers to monitor activities and to ensure that all terms and conditions of awards, including deposit of data and verification of publications, are met. NSF supports a range of research communities that have a variety of experience with data intensive science and associated data management activities. In some cases (for example, genomics; oceanography; social, behavioral, and economic sciences), communities of researchers are accustomed to working with and contributing to large, well-organized data sets and

data centers. In others, experience is more individual and the community-shared data resources are not as centralized. In instances of non-compliance, the Foundation can exercise a range of administrative options depending on the specific circumstances, including withholding future funding, if warranted.

9.0 PUBLIC CONSULTATION EXPERIENCE

NSF, together with other Federal agencies, supported an effort by the National Research Council to hold two two-day public consultation sessions on publications (May 14-15, 2013) and data (May 16-17, 2013). Transcripts, videos, and white papers are available at: http://sites.nationalacademies.org/DBASSE/CurrentProjects/DBASSE_082378. The sessions were announced in the *Federal Register* on May 1, 2013 (Vol. 78, No. 84, p. 25484). Approximately 500 individuals participated in each of the sessions, either in person or remotely.

Over several months, NSF also met with individuals and groups representing commercial and scholarly publishers, Scholarly Publishing and Academic Resources Coalition (SPARC), Association of American Universities (AAU), Association for Research Libraries (ARL), Association for Public and Land -grant Universities (APLU), and made presentations at numerous conferences in the summer, fall, and winter of 2013-14.

Opportunities to comment on specific changes are part of the public notice process, described in Section 11.0. As previously noted, DOE offers a comment function on its portal, and NSF will institute an email communication capability (publicaccess@nsf.gov) on its website (www.nsf.gov/news/special_reports/public_access/).

10.0 INTERAGENCY COORDINATION

NSF launched informal discussions with other science agencies in the late fall of 2012 and continued them through February 2013 to explore options and begin to move toward coordinated public access approaches. NSF has participated fully in the interagency working groups on data and publications organized and led by OSTP and will continue to participate in such interagency groups organized through OSTP, the National Science and Technology Council (NSTC), or other venues.

NSF recognizes the need for ongoing coordination and communication across agencies and as well as among the diverse participants in the scholarly communication process (Section 7.1.1) to reduce confusion and burden for the public as well as for awardees and investigators and to minimize waste and unnecessary redundancy while enhancing the value of the public investment in scientific research. At the agency level, NSF participates actively in interagency initiatives surrounding data management.[33]

[33] For example, through the Emerging Frontiers division, the Directorate for Biological Sciences supports U.S. participation in the Global Biodiversity Information Facility (GBIF). Through the Division of Biological Infrastructure, NSF supports the Intergovernmental Platform for Biodiversity and Ecosystem Services (IPBES). NSF also participates in the U.S. Group on Earth Observations (USGEO) and the USGEO Data Management Working Group (DMWG), which will implement a framework for lifecycle data management, stewardship, and preservation of Earth observation data.

11.0 PUBLIC NOTICE

Changes to NSF's *PAPPG* are made on a well-established annual cycle. The anticipated schedule is:

- February 2015: Announce NSF's intent to revise the *PAPPG* in the *Federal Register* (FR); this starts the clock ticking for a period of 60 days.
- April 2015: Publish second notice in the *Federal Register*. This FR Notice will contain a link to NSF's draft revised PAPPG and allow 60 days for public comment. At the same time that NSF publishes the notice, the Foundation also must provide the supporting statement to OMB.
- July-August 2015: NSF resolves comments submitted to the draft PAPPG.
- October 2015: A new PAPPG is published, with an effective date of January 2016.

The first major change, the revision of NSF's policies, will be announced in draft form no later than April 2015 for implementation by the January 2016 effective date, followed by a set of changes in 2016-2018 concerning data management plans. NSF expects that there will be continued revisions to the *PAPPG* and that these will be embedded in the annual cycle of updates to the *PAPPG*.

The NSF webpage will be updated to reflect activities relating to public access and FAQs will be prepared by April 2015.

12.0 UPDATE, RE-EVALUATION OF THE PLAN, AND REPORTING TO OSTP AND OMB

NSF expects to revisit the plan on an annual basis to ascertain progress in developing and deploying systems and in reaching compliance goals. The NSF Public Access Working Group will be responsible for proposing periodic updates of the plan. Any amendments will be made in consultation with OSTP and OMB.

NSF will provide updates to the directors of OSTP and OMB twice yearly, on January 1 and July 1 of each year, for two years after the effective date of the agency's final plan.

13.0 TIMELINE FOR IMPLEMENTATION

The public access initiative is a set of activities with multiple timelines that will unfold incrementally. Assuming a start date of January 1, 2015, NSF expects to:

1. Establish a system that will enable NSF-funded investigators to deposit their work in a public access-compliant repository (DOE PAGES) and to report that information back to NSF through the annual and final project reports in a way that is integrated into NSF internal administrative systems and allows for monitoring, evaluation, and compliance (Sections 7 and 8). This effort began in FY 2014 and will continue for five-to-seven years. NSF expects the system to become available for voluntary submission by the end of calendar year 2015 with deposit to become mandatory for articles (final approved manuscript or version of record) based on awards resulting from proposals submitted on or after the January 2016 effective date.

2. Revise the *PAPPG* and *PAM* to reflect new requirements for publications and data management plans, as they are developed (Sections 3.0, 7.1.3, and 11.0). Discussion began in FY 2014 and initial draft language will be released in FY 2015 to go into effect no earlier than the January 2016 effective date of the *PAPPG* revisions according to the sequence described in Section 11.0. Ongoing revisions will continue, as necessary, on an annual cycle thereafter.

3. Revise the terms and conditions to accommodate public access requirements (Sections 3.0 and 7.1.3). This task began in FY 2014 and will become effective no earlier than the January 2016 effective date with the implementation of the revised *PAPPG*.

4. Establish a search system that will allow members of the public to search NSF awards and associated publications and provide guidance for more sophisticated searches provided by repository or third-party systems (Section 7.4). This capability already exists. Planning for enhancements began in FY 2014, and the next instantiation is anticipated to be concluded by the end of FY 2015 with updates, as appropriate, thereafter.

5. Establish the NSF Public Access website (www.nsf.gov/news/special_reports/public_access/) which will include: a copy of the approved plan; the FAQs, an email address (publicaccess@nsf.gov) for feedback, and information for requesting a waiver to the 12-month embargo. Elements of the website will become active on a rolling basis, as follows:

    - The website with the plan: March 2015
    - Email address for feedback: March 2015
    - FAQs, including guidance for authors of publications and instructions for format conversion to PDF/A: no later than April 2015, depending on initial feedback from users. NSF expects to update the FAQs as needed.
    - Guidance for requesting a waiver: no later than June 2015.
    - Information on numbers of submissions to the public access repository: to be updated annually, starting January 2017.

6. Provide guidance for reviewers and program officers to enable greater consistency in review of DMPs while remaining sensitive to disciplinary considerations and practices (Sections 7.2.2, 7.3.2, and 7.4.2). This process began in FY 2014 with consultations in FY 2015 and may result in guidance in FY 2016 or later, as the communities reach consensus on best practices.

7. Develop a series of community outreach, consultative, and pilot activities to explore issues related to release, identification and use of data sets (Section 7.3.2 and 7.4.2). This process began in FY 2014 with consultations; follow-on activities to take place no sooner than FY 2016.

8. Develop revisions to NSF's online proposal submission capability to enable template-driven responses to the DMP requirement. This will enable NSF to move toward automated DMP compliance. Timing is based on the outcomes of consultations with the community (see items 6 and 7 above) and concurrent internal enterprise modernization activities. NSF expects to begin this process in FY 2016 and to conclude it in FY 2019.

9. Implement expansion of the federated system to include DOE and other partners, repositories, and research products. Planning will begin in FY 2016 with the addition of one partner, which

will be identified by the NSF Public Access Working Group, and implementation of this second partnership may begin in FY 2017.

10. Review data management in the NSF large facilities. This will begin in FY 2016 and conclude in FY 2018.

11. Maintain coordination and reporting to OSTP and OMB. As described in Section 12.0, the NSF Public Access Working Group will review the plan on an annual basis to determine the need for modifications and updates. NSF will provide updates to the directors of OSTP and OMB twice yearly, on January 1 and July 1 of each year for two years after the effective date of the agency's final plan. This process will begin in FY 2015 and conclude in FY 2017.

14.0 RELATIONSHIP OF THE PLAN TO OPEN DATA POLICY (M 13-13)

On May 9, 2013, President Obama signed an Executive Order (EO) entitled *Making Open and Machine Readable the New Default for Government Information*,[34] which requires all Federal agencies to comply with a new Open Data Policy (M 13-13). Based on government-wide guidance on implementation of the Open Data Policy, NSF intends to exercise discretion in determining whether scientific research data resulting from an award will be subject to the requirements of the EO.

In support of agency and Federal transparency initiatives such as the Digital Government Strategy and the Open Data Policy, NSF makes agency data easily accessible to the public in open formats that can easily be shared via a variety of mechanisms (email, Facebook, Twitter, etc.), printed, or downloaded for use with data-mining and extraction tools. As required by the Open Data Policy, NSF's public data inventory includes over 90 agency data sets, including those reflected in Data.gov, agency resources listed on the agency's Open Government page, and other high visibility agency data sets. NSF is committed to continually expanding, enriching, and opening existing agency data. About 30 of the agency's data sets are machine-readable, and NSF has made available three web APIs, including an Award Search Web API that was released in the spring of 2014.

Pursuant to M-13-13, NSF's public data listing is available at http://www.nsf.gov/data.json; the metadata include the common core metadata schema described at https://project-open-data.cio.gov/. NSF does not collect data sets from recipients of NSF awards, and so research data sets are not included in NSF's public data listing.

15.0 TRAINING AND WORKFORCE DEVELOPMENT RELATED TO SCIENTIFIC DATA MANAGEMENT, ANALYSIS, STORAGE, PRESERVATION, AND STEWARDSHIP

NSF already provides for training and workforce development within the framework of its existing policies and solicitations,[35] and is aware of and has supported related programs at graduate schools of information in data and data management.[36]

---

[34] http://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-

[35] For example, the Division of Biological Infrastructure's Advances in Biological Informatics program, Directorate for Biological Sciences has made an award to the Ecological Society of America (ESA) to develop a training program

The Data Science Priority Goal has launched six activities designed to promote data science training and workforce development:

- Posted the solicitation in March 2014 for the National Science Foundation Research Traineeship (NRT) Program with the priority theme: Data-Enabled Science and Engineering (DESE).[37] NRT is an agency-wide program designed to encourage the development of bold, new, potentially transformative, and scalable models for STEM graduate training that ensure that graduate students develop the skills, knowledge, and competencies needed to pursue a range of STEM careers.

- Identified approximately 50 solicitations as potential vehicles for encouraging data science education and training.

- Developed model language concerning data science training and workforce development for inclusion in future solicitations.

- Modified six solicitations to reflect the new model language.

- Developed data science workshops, initiated in FY 2014 and continuing into FY 2015.

- Is currently developing a mechanism for monitoring graduate students through the NRT and Graduate Research Fellowship (GRF) programs

This approach is consistent with discussions of data management and training activities that increasingly focus on the importance of training that is contextualized in scientific research.

16.0 RESOURCES

The OSTP memorandum directs that resources to implement the plan should be identified within the existing agency budget. NSF public access budget requirements will be developed each fiscal year

---

for PIs on developing sustainable biological infrastructure. ESA will soon hold its first short-course, and expects participation across a range of resource and organization types.

[36] For example, see Award Number 0903629, "IGERT: Open Data: Graduate Training for Data Sharing and Reuse in E-Science" (PI: Margaret Hedstrom, University of Michigan); Award Number 1253980, "CAREER: Enabling High-Throughput Data Management in Scientific Domains," (PI: Yicheng Tu, University of South Florida); Award Number 0848296, "NARA Transcontinental Persistent Archive Prototype" (PI: Richard Marciano, University of North Carolina at Chapel Hill). The Long-Term Ecological Research Network (LTER) Office provides training services, including Information Management Training and a Training and Workshop Computer Lab (http://lno.lternet.edu/services/types/Training). Similarly, the George E. Brown, Jr. Network for Earthquake Engineering Simulation (NEES) supports education and training for students, teachers (at all levels), professionals, and citizen involvement. NEES' continuing education program offers training in data management and data upload and visualization, among other topics. These offerings are both free and paid, and can have professional credits associated with them (see https://nees.org/education/for-professionals/continuing-education).

[37] http://www.nsf.gov/funding/pgm_summ.jsp?pims_id=505015

informed by results in the previous year(s). Overall funding priorities will be reflected in the Foundation's budget submission.

17.0 ADDITIONAL MATERIAL

America COMPETES Reauthorization Act of 2010, Public Law 111-358-January 4, 2011 (http://www.gpo.gov/fdsys/pkg/PLAW-111publ358/pdf/PLAW-111publ358.pdf)

National Science Board, Committee on Strategy and Budget*, Task Force on Data Policies, Digital Research Data Sharing and Management*, NSB-11-79, December 14, 2011 (http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf)

National Science Foundation, Investing in Science, Engineering, and Education for the Nation's Future: NSF Strategic Plan for 2014-2018, March 10, 2014 (http://www.nsf.gov/about/performance/strategic_plan.jsp)

National Science and Technology Council, Executive Office of the President, Interagency Public Access Coordination; A Report to Congress on the Coordination of Policies Related to the Dissemination and Long-term Stewardship of the Results of Federally Funded Scientific Research, March 2012 (http://www.whitehouse.gov/sites/default/files/microsites/ostp/public_access-final.pdf)

National Science Foundation Open Government Plan 2.0, April 2012 (http://www.nsf.gov/pubs/2012/nsf12066/nsf12066.pdf)

Scholarly Publishing Roundtable, *Report and Recommendations from the Scholarly Publishing Roundtable*, January 2010 (http://www.aau.edu/policy/scholarly_publishing_roundtable.aspx?id=6894)

Working Group on Expanding Access to Published Research Findings, Accessibility, Sustainability, Excellence: How to Expand Access to Research Publications, July 2013 (http://www.researchinfonet.org/wp-content/uploads/2012/06/Finch-Group-report-FINAL-VERSION.pdf)

LIST OF EXHIBITS

Exhibit 1: Summary of findings from internal studies

In developing this plan, NSF has considered internal, administrative resources and infrastructure assets; broad patterns of publication by NSF-funded investigators and their sources of funding; and the creation and use of scientific data sets among disciplines, and observed the following:

- NSF investigators publish in a very wide variety of journals.
- NSF investigators typically have funding from more than one agency; patterns in agency sponsorship vary across directorates.
- NSF investigators report a range of products in their annual and final reports; the significance of these products varies by discipline.
- NSF can leverage internal reporting (Research Performance Progress Report (RPPR)), pre-and post-award, and administrative systems.
- NSF has a DMP requirement, which is implemented through the proposal submission process and managed as part of its post-award management procedures.
- NSF supports public search of its active and inactive awards through its website (nsf.gov) as well as through Research.gov. Results include descriptions of the award, the names of the Awardee(s) and Principal Investigator(s), and links to publications that have been reported.
- Data resulting from NSF support are managed through a highly distributed set of external repositories reflective of communities and disciplines.
- NSF does not currently support its own publicly accessible repository for reports, papers, or articles resulting from research the agency funds.
- NSF has determined that creating a separate, internal repository for NSF-supported material is cost-prohibitive.
- NSF will seek one or more partners to establish a permanent repository for publications and other products of NSF-funded research.

Principles:

These findings led to the following principles, which guided the planning process:

- Implement a flexible, incremental, integrated approach to data and publications that can be extended to other products (for example, white papers, technical reports, and so on);
- Minimize burden to awardee institutions, Principal Investigators and NSF staff; and
- Collaborate with other agencies and public/private organizations to minimize cost and burden.

Exhibit 2: Crosswalk from the criteria in the February 22, 2013, OSTP memo to the organization of the NSF plan

Criterion in OSTP memo		Addressed in section:
1	Policy Principles, including consultation with stakeholders	1.0; 1.1; 1.2; 7.0; 9.0; 10.0
2	Agency Public Access Plan	2.0; 3.0; 4.0; 5.0; 6.0; 7.0
2.a	Strategy for leveraging existing archives and fostering public/private partnerships with scientific journals	7.1.1; 7.1.2; 7.2.1; 7.3.1; 7.4.1; 10.0
2.b	Strategy for improving the public's ability to locate and access digital data	7.3.1; 7.4.2
2.c	Approach for optimizing search, archival, and dissemination features that encourages innovation in accessibility and interoperability	7.4.1
2.d	Plan for notifying awardees and other federally funded researchers of their obligations	11.0
2.e	Strategy for measuring and enforcing compliance	7.8; 8.0
2.f	Agency resources	16.0
2.g	Timeline	13.0
2.h	Circumstances that might prohibit meeting objectives set forth in this plan	No impediment (other than availability of resources) is foreseen.
3	Access to publications	7.4
3.a	Read, download, and analyze in digital form	7.6
3.a.i	12-month post publication embargo as a guideline	7.5.1
3.a.ii	Mechanism to petition to change embargo	7.1.1; 7.5.1, 9, 11
3.b	Facilitate easy search	7.4.1
3.c	Ensure full access to metadata without charge on first publication	3.0; 7.4.1
3.d	Encourage public/private collaboration to:	7.1.1; 7.3.1; 7.4.1
3.d.i	Maximize potential for interoperability between public/ private platforms and creative reuse	7.3.1; 7.4.1; 7.7
3.d.ii	Avoid unnecessary duplication	7.2.1; 7.3.1
3.d.iii	Maximize the impact of Federal research investment	1.0; 1.2; 7.3.1; 10.0
3.d.iv	Assist in implementation of Agency plan	7.3.1
3.e	Ensure attribution to authors, journals, and original publishers is maintained	7.3.1
3.f	Ensure that publications and metadata are stored in an archival solution that:	7.3.1; 7.7
3.f.i	Provides for long-term preservation and access without charge	3.3; 7.7
3.f.ii	Uses standards and widely available formats	7.3.1; 7.4.1; 7.7
3.f.iii	Provides access for persons with disabilities consistent with Section 508 of the Rehabilitation Act of 1973	7.2.1; 7.3.1; 7.4.1; 7.7
3.f.iv	Enables integration and interoperability with other Federal public access archival solutions	4.0; 7.1.1; 7.3.1; 10.0

4	Digitally formatted data resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze; shall not include laboratory notebooks, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects such as laboratory specimens	1.0; 2.0; 3.0; 3.2.1;7.5.2
4.a	Maximize access while:	3.0
4.a.i	Protecting confidentiality and personal privacy	3.0; 3.2.1; 7.5.2
4.a.ii	Recognizing proprietary interests and avoiding significant negative impact on intellectual property rights, innovation, and US competitiveness	3.0; 3.2.1; 7.3.2; 7.5.2
4.a.iii	Preserving the balance between the relative value of long-term preservation and access and the associated cost and administrative burden	3.3; 7.0; 7.3.2
4.b	Develop Data Management Plans (DMPs)	3.0
4.c	Allow inclusion of costs for data management and access	3.3
4.d	Allow for appropriate evaluation of DMPs	3.2.1
4.e	Include mechanisms to ensure compliance	8.1.3; 8.1.4
4.f	Promote deposit of data in publicly accessible databases, when appropriate and available	7.2.2; 7.3.2; 7.5.2
4.g	Encourage cooperation with the private sector to improve data access and compatibility	3.2.3; 7.1.1; 7.2.2; 7.3.2. 7.4.1
4.h	Approaches for identifying and providing appropriate attribution	3.2.3; 7.3.2
4.i	Support training, education, and workforce related to scientific data management, analysis, storage, preservation, and stewardship	15.0
4.j	Provide for assessment of long-term needs for the preservation of scientific data	7.3.2
5	Implementation of Public Access Plan	7.0; 12.0; 13.0