INSTITUTE FOR DEFENSE ANALYSES

Public Access to Federally Funded Research: Policy, Evolving Ecosystem, and Agency Efforts

Karen D. Gordon, *Project Leader*

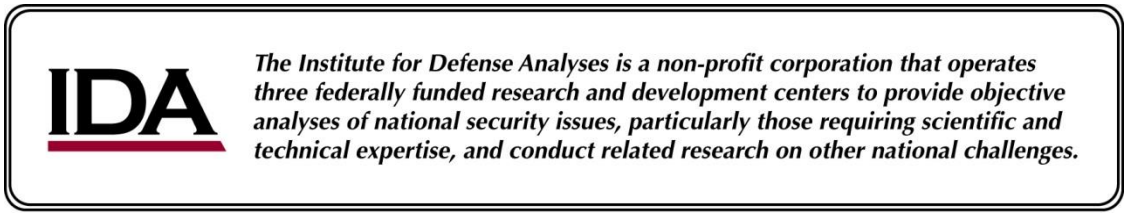
Jonathan R. Agre

Robert F. Leheny

January 2017

IDA Document
D-8320
Log: H 2017-000015
Copy

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882



About This Publication

This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-14-D-0001, Task ET-2-2954.30, "ODNI/IARPA Data Management Plans (DMP)," for Office of the Director of National Intelligence, Intelligence Advanced Research Projects Activity. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments

Arun S. Maiya

For more information:

Karen D. Gordon, Project Leader
kgordon@ida.org, 703-845-2343

Leonard J. Buckley, Director, Science and Technology Division
lbuckley@ida.org @ida.org, 703-578-2800

INSTITUTE FOR DEFENSE ANALYSES

IDA Document D-8320

Public Access to Federally Funded Research: Policy, Evolving Ecosystem, and Agency Efforts

Karen D. Gordon, *Project Leader*

Jonathan R. Agre

Robert F. Leheny

Executive Summary

This document captures the results of preliminary research conducted by the Institute for Defense Analyses (IDA) for a project on public access to the results of Federally funded research. The project was sponsored by the Intelligence Advanced Research Projects Activity (IARPA) within the Office of the Director of National Intelligence (ODNI).

A companion document contains IDA's primary project deliverables – a draft plan for IARPA implementation of ODNI public access policy and draft text for inclusion in IARPA broad agency announcements (BAAs) – as well as other IARPA-specific briefings and memos.

Project Description

Background. The Office of Science and Technology Policy (OSTP) issued a memorandum to all heads of executive departments and agencies in February 2013 that directs agencies with significant research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government. The memorandum details the Policy Principles, Agency Public Access Plans, Objectives for Public Access to Scientific Publications, Objectives for Public Access to Scientific Data in Digital Formats, Implementation of Public Access Plans, and General Provisions.

IARPA, as part of ODNI, needs to develop language and processes to solicit data management plans (DMPs), guide their evaluation, and ensure faithful execution that meets the directions of the OSTP memorandum. Substantial work and experience with DMPs exist within other departments and agencies, notably the National Institutes of Health (NIH) and the National Science Foundation (NSF). This work can be leveraged to establish a basis for ODNI/IARPA DMPs and processes. IARPA would like to incorporate such language into contracts by FY 2016 and have in place the processes, guidance, examples, templates, and infrastructure to implement DMPs to meet the spirit and direction of the OSTP memorandum.

A number of key questions need research and reconciliation. What is the state of the art in DMPs? What is there about the ODNI context and IARPA that may require unique DMP elements or exceptions? How should criteria for making research data public be evaluated? What options does ODNI/IARPA have to post research data publicly?

Objective. The objective of this project is to research DMPs and develop the basis (criteria, language, guidance, standards, examples, processes, strategy, and plans) for IARPA research and development (R&D) contracting and deliverable review.

Expanded Scope. At the project's kickoff meeting, held on 7 October 2014 at IARPA, it was agreed that the scope of the project should encompass not only data but also publications. This expanded scope aligns with the 22 February 2013 OSTP memo, "Increasing Access to the Results of Federally Funded Scientific Research," which directed "each agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support

increased public access to the results of research funded by the Federal Government." The memo stated that "such results include peer-reviewed publications and digital data."

Contents of this Document

This document captures IDA's preliminary research on public access to the results of Federally funded research and, in so doing, lays the foundation for IDA support of IARPA's implementation of public access policy. The document comprises two briefings and four memos. Highlights are given below:

1. **Public Access to Scientific Publications and Data: State of the Practice**. This briefing considers the full scope of the 22 February 2013 OSTP memo, which calls for increased access to the results of Federally funded research, including both scientific publications and scientific data. It points out that requirements for public access to publications are more straightforward than those for public access to data. For publications, requirements are *uniform across disciplines*; for data, requirements are *discipline-specific*.

   Agency public access plans, in keeping with the OSTP memo, take different approaches to publications and data. Namely, the plans are *prescriptive* with respect to publications; they specify the requirements and the process for making peer-reviewed journal articles publicly accessible. However, the plans are *collaborative* with respect to data. That is, agencies, organizational units within agencies, and program managers specify requirements for Data Management Plans (DMPs). Then researchers respond to the requirements by formulating DMPs and submitting them as part of their research proposals or plans. Finally, reviewers and program managers evaluate the DMPs. Sometimes, revisions to DMPs can be negotiated between researchers and funding agencies.

   A public access ecosystem is developing and evolving. It encompasses (1) identifiers for articles, datasets, funders, and researchers; (2) publication archives; (3) data repositories; (4) metadata standards; and (5) open license schemes. This ecosystem is integral to agency public access strategies and plans.

2. **Responding to OSTP Public Access Memo: Processes, Sample Agency Implementations, and Preliminary Analysis of Rights-in-Data.** This briefing starts with a brief update on the status of agency public access plans. It then describes the Department of Energy (DOE) implementation of public access policy, which could be leveraged in IARPA's implementation. The briefing also addresses the copyright issue, summarizing the NIH approach to managing copyright for public access to publications.

   Department of Energy Implementation of Public Access Policy. DOE was the first agency to have its public access plan approved by OSTP and the Office of Management and Budget (OMB), and it has made considerable progress in implementing public access policy as follows:

   - DOE Public Access Plan (24 July 2014): States that DOE is taking a phased approach, with the Office of Science taking the lead with respect to public access to data;

- DOE communication with stakeholders, which includes (1) a link to the DOE Public Access Plan on the DOE Open Government webpage and (2) a press release on DOE public access efforts (4 August 2014);
- DOE Office of Scientific and Technical Information (OSTI) capabilities for facilitating public access: (1) Public Access Gateway for Energy and Science (PAGES) – full-text search of publications, (2) DOE Data ID Service – Digital Object Identifiers (DOIs) for datasets, and (3) DOE Data Explorer – search of metadata describing datasets;
- DOE Office of Science Data Management Policy, which is articulated in the following sources: (1) Statement on Digital Data Management webpage – states principles, specifies requirements, offers additional guidance, and includes extensive Frequently Asked Questions (FAQ) and (2) Suggested elements of Data Management Plan (DMP) – covers data types and sources, content and format, sharing and preservation, protection, and rationale.

National Institutes of Health (NIH) Approach to Managing Copyright. The NIH mandate on submission of peer-reviewed journal articles to PubMed Central dates back to 2008. Over the years, a three-step approach for managing copyright has been developed and could serve as a model for other agencies:

- At award acceptance, the author grants a license to an institutional awardee: The license protects the institutional awardee and ensures NIH rights, because the license stands regardless of the authors' actions with respect to any future publication agreements with a publisher.
- At article submission, the author gives notice to the publisher, informing the publisher of rights that will be retained by author(s), university, and NIH. This preempts the over-reaching copyright transfer provisions of any subsequent publication agreement between author and publisher.
- At article acceptance, the author addendum to the publication agreement reiterates the rights retained by author(s), university, and NIH. This protects the author against publisher accusations of misrepresentation in the case that the publication agreement calls for transfer of all rights to publisher.

3. **Agency Public Access Plans:** This memo summarizes the key publication and data provisions of several early agency public access plans. Key findings with respect to publication provisions include the following: (1) there are two leading implementations of public access to publications – NIH PubMed Central (PMC) and DOE Public Access Gateway for Energy and Science (PAGES); (2) NSF extends public access beyond journal articles to peer-reviewed conference papers; and (3) National Science Foundation (NSF), National Aeronautics and Space Administration (NASA), and U.S. Department of Agriculture (USDA) recognize the importance of supporting bulk downloads of full-text articles for research purposes (e.g., text mining).

   The key finding with respect to data provisions is that agency public access plans have considerable variation on several points, including the following: (1) the scope of the data to be shared and preserved; (2) whether software is considered in scope; (3) the length of data

retention periods; (4) data repositories (often can be institutional or public); (5) metadata repositories (typically are centralized); and (6) whether the research data is integrated into an agency's public data listing (PDL) as part of its Data.gov participation.

4. **Sample Data Retention Periods.** This memo examines research data retention policies. The bottom line is that key policies mandate *minimum* retention periods of modest length. In particular, OMB Circular A-110 and the Council on Governmental Relations (an Association of Research Universities) specify a minimum retention period of 3 years for research data. Furthermore, the top research universities (in terms of Federally financed R&D expenditures) typically have minimum research data retention periods of between 3 and 7 years.

   The longer retention periods at some medical schools and universities seem to be driven in part by 42 CFR Part 93 (Public Health Service Policies on Research Misconduct), which specifies a 6-year time limit on raising allegations of misconduct (§93.105).

5. **National Science Foundation Guidance on Public Access.** This memo describes how NSF is promulgating public access requirements to researchers. It begins by examining a sample NSF program solicitation. The solicitation points to three NSF documents: the NSF Grant Proposal Guide; the NSF Award and Administration Guide; and the NSF Grant General Conditions. These documents in turn point to the NSF "Public Access Policy" webpage and the "Dissemination and Sharing of Research Results" webpage.

   The Grant Proposal Guide specifies detailed NSF-wide Data Management Plan (DMP) requirements. NSF Directorates and individual program solicitations can refine the requirements as necessary.

6. **Department of Energy Guidance on Public Access.** This memo describes how DOE is promulgating public access requirements to researchers. It begins by examining a sample DOE Office of Science funding opportunity announcement (FOA). The FOA points to the DOE "Federal Assistance Reporting Checklist" and the DOE Office of Science "Acknowledgments of Federal Support" webpage for requirements on public access to scientific publications. The FOA states basic DMP requirements and points to the "Statement on Digital Data Management" for more detailed requirements on public access to research data.

   The DOE E-Link tool is used to submit publications (specifically, accepted manuscripts) and associated metadata to DOE. The publications can be searched at the DOE Public Access Gateway for Energy and Science (DOE PAGES). The DOE E-Link tool is also used to submit metadata on research datasets to DOE. The metadata can be searched at DOE Data Explorer.

Contents

Public Access to Scientific Publications and Data: State of the Practice

Karen D. Gordon, Jonathan R. Agre, Robert F. Leheny
5 December 2014

Synopsis

This briefing was prepared early in the course of the project. It puts Data Management Plans (DMPs) into context by considering the full scope of the 22 February 2013 OSTP memo, which calls for increased access to the results of Federally funded research, including both scientific publications and scientific data. DMPs, of course, address the data side of public access.

The briefing begins with a Bottom Line Up Front (BLUF) that serves as an Executive Summary for the briefing and highlights some of IDA's key findings and observations on the state of the practice. The BLUF is structured into six parts:

- Scope of the Office of Science and Technology Policy (OSTP) Memo: Describes what is meant by scientific publications and scientific data;
- Public Access Plans: Notes that the DOE Public Access Plan was the first to be published in response to the OSTP Memo;
- Phased/Tiered Approach: Observes that agencies are taking a phased and/or tiered approach to public access;
- Publications vs. Data: Explains that public access to data is discipline-specific;
- Data Management Plans (DMPs): Lists elements of DMPs and notes that DMPTool is a widely used resource for creating DMPs;
- Public Access Ecosystem: Describes how a public access ecosystem – consisting of identifiers, archives, repositories, standards, and open license schemes – is developing and evolving.

The briefing then provides detailed material in five areas: (1) Background (OSTP Memo and Agency Public Access Plans); (2) Public Access to Scientific Publications; (3) Public Access to Scientific Data; (4) Data Management Plans; and (5) Evolving Public Access Ecosystem.

The briefing captured the state of the practice and proved to be a good source of reference material throughout the course of the project.

IDA

Institute for Defense Analyses

4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Public Access to Scientific Publications and Data: State of the Practice

Dr. Karen D. Gordon

Dr. Jonathan R. Agre

Dr. Robert F. Leheny

5 December 2014

IDA | Outline

- Bottom Line Up Front (BLUF) / Executive Summary
  - Scope of Office of Science and Technology Policy (OSTP) Memo
  - Public Access Plans
  - Phased/Tiered Approach
  - Publications vs. Data
  - Data Management Plans (DMPs)
  - Public Access Ecosystem
- Background
- Public Access to Scientific Publications
- Public Access to Scientific Data
- Data Management Plans (DMPs)
- Evolving Public Access Ecosystem

IDA | BLUF, Part 1—Scope of OSTP Memo

- The Office of Science and Technology Policy (OSTP) memo, "Increasing Access to the Results of Federally Funded Scientific Research," dated 22 Feb 2013, calls for increased public access to the results of unclassified Federally funded research, where results include scientific publications and scientific data
- With respect to "**scientific publications**," the scope, at least initially, seems to be limited to **peer-reviewed journal articles**
  - Does not extend to conference articles or technical reports
- With respect to "**scientific data**," the scope is less clear
  - Includes the data (i.e., the **digital recorded factual material**) "**commonly accepted in the scientific community as necessary to validate research findings**"
    - At a minimum, "research findings" include results published in peer-reviewed journal articles
  - Is constrained by other requirements (e.g., privacy, security, IP rights), as well as by cost

IDA | BLUF, Part 2—Public Access Plans

- OSTP memo requires each Federal agency with over $100 million in annual R&D expenditures to develop a public access plan
  - OMB and OSTP are reviewing draft plans "for compliance with **each requirement in the OSTP memorandum** and for consistency with the **Executive Order** on Making Open and Machine Readable the New Default for Government Information and the accompanying **Open Data Policy issued by OMB**" [Holdren Letter to House and Senate Appropriations Committees, 24 March 2014]
- Department of Energy (DOE) Public Access Plan
  - Released 4 August 2014
  - First agency public access plan to be published
  - Is written in 2 parts:
    - Public Access to Scientific Publications
    - Public Access to Scientific Data in Digital Formats
  - Could serve as a model for IARPA Public Access Plan

IDA | BLUF, Part 3—Phased/Tiered Approach

Agencies are, in some cases, taking a phased and/or tiered approach to implementation of public access

- With respect to emphasis on publications vs. data
  - NIH has emphasized publications; NSF has emphasized data
- With respect to which data is covered, e.g., DOE Public Access Plan considers various classes of data:
  - "[R]esearch data **displayed in publications** [in charts, figures, images, etc.] resulting from the proposed research": DMPs should provide a plan for making this data **"open, machine-readable, and digitally accessible to the public** at the time of publication"
  - "**[U]nderlying digital research data** used to generate the displayed data": Should be made "**as accessible as possible** to the public" in accordance with principles stated in DOE Plan
  - **Other data** "generated in the course of the proposed research": DMPs should describe **"whether and how" this data will be shared and preserved** and, if not shared or preserved, "how results could be validated"
- With respect to which researchers are covered, e.g.,
  - DOD may start with intramural researchers
- With respect to which organizational units of agency are covered, e.g.,
  - DOE/Office of Science is piloting data management policy
  - IARPA could serve as model for Intelligence Community public access efforts

IDA | BLUF, Part 4—Publications vs. Data

- Requirements for public access to publications are more straightforward than those for public access to data
  - For **publications**, requirements are **uniform across disciplines**
  - For **data**, requirements are **discipline-specific**
    - What data? When should data be made available? For how long should data be available? Which repositories? Which metadata standards?
- Agency Public Access Plans, in keeping with OSTP memo, seem to be taking different approaches to publications and data
  - **Prescriptive with respect to publications**
    - Specify requirements and process for making peer-reviewed journal articles publicly accessible
  - **Collaborative with respect to data**
    - Agencies, organizational units within agencies, and program managers specify requirements for Data Management Plans (DMPs)
    - Researchers formulate DMPs as part of research proposals or plans
    - Reviewers and Program Managers evaluate DMPs

IDA | BLUF, Part 5—Data Management Plans (DMPs)

- NIH and NSF policies pre-date OSTP Memo
  - NIH Data Sharing policy dates back to 2003
  - NSF has required DMPs as part of proposals since January 2011
- DOE published "Statement on Digital Data Management" on 28 July 2014
  - **Aligns with OSTP Memo and with DOE Public Access Plan**
    - DOE Plan recognizes importance of **cost/benefit** considerations
    - DOE Plan recognizes **discipline-specific** nature of public access to scientific data; it advocates community-based standards, best practices, and repositories
    - DMPs are reviewed as part of research proposal **merit review process**
  - Suggests the following elements for DMPs: **Data Types and Sources**, **Content and Format**, **Sharing and Preservation**, **Protection**, **Rationale**
- University libraries are playing a central role in facilitating development of DMPs
- **DMPTool**: Widely used resource developed by University of California Curation Center of the California Digital Library
  - **Facilitates development of DMPs**
  - Partners with funding agencies to provide links to agency-specific guidance, templates, and sample DMPs
  - Suggested by many universities for use by their researchers

IDA | BLUF, Part 6—Public Access Ecosystem

Public access ecosystem is developing and evolving

- Identifiers and Metadata Repositories
  - **CrossRef** provides Digital Object Identifiers (DOIs) for **articles**
  - **DataCite** provides DOIs for **datasets**
  - **FundRef** provides DOIs for **funders** and organizes funders hierarchically
  - **ORCID** provides IDs for **researchers**
  - Above organizations provide searchable metadata repositories, e.g., for finding articles written by a specified researcher or supported by a specified funding agency
- Publication Archives: **LOCKSS** and **CLOCKSS**
- Data Repositories:
  - Many repositories, most discipline-specific but some general-purpose
  - **Databib**: 992 repositories organized into 21 subject areas
  - **Registry of Research Data Repositories (re3data.org)**: 1029 repositories tagged with one or more of 152 subject areas
  - Databib and re3data.org are merging under auspices of DataCite
- Metadata Standards: **DataCite** defines a metadata schema for publication and citation of research datasets
- Licenses: **Project Open Data** (OMB and OSTP) identifies **open license principles** and licenses that meet the open criteria

IDA

Background

- OSTP Memo, 22 February 2013
  - Required Elements of Agency Public Access Plans
- Executive Order, 9 May 2013
- OMB Memo, 9 May 2013
- Progress on Agency Public Access Plans
- DOE Public Access Plan

OSTP Memo, "Increasing Access to the Results of Federally Funded Scientific Research," 22 Feb 2013

- **Objectives for Public Access to Scientific Publications:**
  - "To the extent feasible and consistent with law; agency mission; resource constraints; U.S. national, homeland, and economic security; and the objectives listed below, the **results of unclassified research that are published in peer-reviewed publications** directly arising from Federal funding should be stored for **long-term preservation** and **publicly accessible to search, retrieve, and analyze** in ways that maximize the impact and accountability of the Federal research investment."
- **Objectives for Public Access to Scientific Data in Digital Formats:**
  - "To the extent feasible and consistent with applicable law **and policy**; agency mission; **resource constraints**; U.S. national, homeland, and economic security; and the objectives listed below, **digitally formatted scientific data** resulting from unclassified research supported wholly or in part by Federal funding should be stored and publicly accessible to search, retrieve, and analyze."
- **Direction to Federal Agencies to Develop Public Access Plan**
  - "The Office of Science and Technology Policy (OSTP) hereby directs each Federal agency with over $100 million in annual conduct of research and development expenditures to develop a plan to support increased public access to the results of research funded by the Federal Government."
  - "Results of research" means scientific publications and digital scientific data

Required Elements of Agency Public Access Plans [OSTP Memo, 22 Feb 2013]

- **Regarding scientific publications:**
  - Strategy for "leveraging existing archives" and "fostering public-private partnerships with scientific journals"
- **Regarding scientific data**:
  - Strategy for "improving the public's ability to locate and access digital data resulting from federally funded scientific research"
- **Regarding scientific publications and data**:
  - Approach for "optimizing search, archival, and dissemination features … while ensuring long-term stewardship of the results"
  - Plan for notifying awardees and other federally funded scientific researchers of their obligations (e.g., through guidance, conditions of awards, and/or regulatory changes)
  - Strategy for "measuring and …enforcing compliance with its plan"
- **Regarding cost, time, and extenuating circumstances:**
  - "Identification of resources within the existing agency budget to implement the plan"
  - "Timeline for implementation"
  - "Identification of any special circumstances that prevent the agency from meeting any of the objectives set out in this memorandum"

Executive Order—Making Open and Machine Readable the New Default for Government Information

- Issued 9 May 2013
- OMB Open Data Policy:
  - "The Director of the OMB shall issue an Open Data Policy in consultation with the CIO, CTO, and Administrator of the Office of Information and Regulatory Affairs (OIRA), to advance the management of Government information as an asset …"
  - OMB M-13-13, "Open Data Policy—Managing Information as an Asset," was also issued 9 May 2013 (see next slide)
- Agency Implementation:
  - "When implementing the Open Data Policy, agencies shall incorporate **a full analysis of privacy, confidentiality, and security risks** into each stage of the information lifecycle to identify information that should not be released. These review processes should be **overseen by the senior agency official for privacy**. It is vital that agencies **not release information if doing so would violate any law or policy, or jeopardize privacy, confidentiality, or national security**."

OMB M-13-13, "Open Data Policy—Managing Information as an Asset": *Open Data Principles*

- **Public**
  - Subject to privacy, confidentiality, security, or other valid restrictions
- **Accessible**
  - Machine-readable, open, non-proprietary formats
- **Described (strengths, weaknesses, analytical limitations, security requirements, how to process)**
  - Metadata, thorough documentation of data elements, data dictionaries
  - If applicable, purpose of collection, population of interest, characteristics of sample, and method of collection
- **Reusable**
  - Open license that places no restrictions on their use
- **Complete**
  - Primary data (i.e., as collected at the source)
  - Derived or aggregate data, with reference to primary data
- **Timely**
  - Available as quickly as necessary to preserve the value of the data
- **Managed Post-Release**
  - Designated point of contact

Open Data: "Publicly available data … fully discoverable and usable by end users"

IDA | Progress on Agency Public Access Plans

- 21 agencies have submitted draft plans to OSTP (due 22 Aug 2013; done on or before 24 Mar 2014, according to Holdren Letter to House and Senate Appropriations Committees)
- OMB and OSTP have reviewed all agency plans, and all comments have been returned to the agencies. Plans are being reviewed for:
  - Compliance with ***each requirement*** in OSTP memo
  - Consistency with ***9 May 2013 Executive Order***, "Making Open and Machine Readable the New Default for Government Information," and accompanying ***OMB M-13-13***, "Open Data Policy"
- On 4 August 2014, the Department of Energy (DOE) became the first agency to release a Public Access Plan in response to the OSTP memo [DOE press release and American Institute of Physics commentary]

IDA | DOE Public Access Plan: Overview of Contents

Public Access to Scientific Publications

With regard to unclassified and otherwise unrestricted research in scientific publications, the Department proposes a new policy and tool for providing access to peer-reviewed scholarly publications and associated metadata in which publishers retain their rights under copyright to the Version of Record (VoR). Both the policy and tool will be applied to scholarly publications resulting from unclassified and otherwise unrestricted research supported by the Department.

Public Access to Scientific Data in Digital Formats

With regard to unclassified and otherwise unrestricted scientific data in digital formats, the Department proposes a set of principles and requirements to be adopted by all DOE offices supporting open research. Implementing strategies and timelines may differ across the Department depending on the specific communities supported and funding mechanisms used by each office.

IDA

Public Access to Scientific Publications

- Closer Look at “Scientific Publications”
- NIH Approach
- DOE Approach
- DOD Approach
- Limitations to Public Access

IDA | Closer Look at "Scientific Publications"

- NIH Public Access Policy [2008]
  - Applies to any manuscript that is peer-reviewed and **"accepted for publication in a journal"**
- OSTP Memo [22 February 2013]
  - Speaks of "peer-reviewed publications" and "fostering public-private partnerships with **scientific journals**" and ensuring "that **attribution to authors, journals, and original publishers** is maintained"
- DOE Public Access Plan [dated 24 July 2014, released 4 August 2014]
  - "Metadata accompanying the accepted manuscript, e.g., author name, **journal title**, and digital object identifier (DOI) for the VoR [Version of Record], ensures that **attribution to authors, journals, and original publishers** will be maintained."
  - Scholarly publications: "final, peer-reviewed and accepted manuscripts or, for participating publishers, the corresponding **published journal article**"
- DOD Public Access Memo [by USD(AT&L), dated 9 July 2014]
  - Applies to "peer-reviewed scholarly publications" and requires authors to submit "final, peer-reviewed **journal** manuscripts"

So, focus is on peer-reviewed journal articles, not on conference articles or technical reports

NIH Approach to Public Access to Scientific Publications

- Legal Requirement Underlying NIH Public Access Policy: Division F Section 217 of PL 111-8 (Omnibus Appropriations Act, 2009)
  - "The Director of the National Institutes of Health ("NIH") shall require in the current fiscal year and thereafter that all investigators funded by the NIH submit or have submitted for them to the National Library of Medicine's **PubMed Central** an electronic version of their final, peer-reviewed manuscripts upon acceptance for publication, to be made **publicly available no later than 12 months after the official date of publication**: Provided, that the NIH shall implement the public access policy in a manner consistent with copyright law."
- NIH Public Access Policy applies to any manuscript that is **peer-reviewed and that is accepted for publication in a journal** on or after April 7, 2008
- PubMed Central
  - Free full-text archive of biomedical and life sciences journal literature
  - Developed and operated by the National Center for Biotechnology Information (NCBI) of the National Library of Medicine (NLM) at NIH

1. **Agency archive: PubMed Central**
2. **Content of archive: Author's final peer-reviewed manuscripts or, for participating journals, journal's final published article (i.e., the version of record)**
3. **Approach to public access: PubMed Central delivers the content**
4. **Embargo period: Up to 12 months**

DOE Approach to Public Access to Scientific Publications

- DOE Public Access Policy, developed in response to OSTP memo, applies to "**final, peer-reviewed and accepted manuscripts** or, for participating publishers, the corresponding **published journal article**"
- Public Access Gateway for Energy and Science (PAGES) Portal
  - Operated and maintained by Office of Scientific and Technical Information (OSTI)
  - Holds metadata and abstracts that are searchable and available for bulk download
  - Before end of 12-month embargo period, PAGES links to Version of Record (VoR) at publisher's site (typically behind pay wall)
  - After embargo period, PAGES links to "best available version"
    - VoR at publisher's site, if access is open and free
    - Otherwise, accepted manuscript (typically hosted at institutional repository, i.e., national lab or grantee institution; otherwise hosted at OSTI)

1. **Agency portal: PAGES, with centralized metadata and links to decentralized full-text articles**
2. **Dark archive: OSTI repository of accepted manuscripts, accessed only when no other version of a requested article is publicly available**
3. **Approach to public access: "Best available version" delivered from publisher's site (1st choice), institutional repository (2nd choice), or OSTI repository (3rd choice)**
4. **Embargo period: Up to 12 months**

DOD Approach to Public Access to Scientific Publications

- DoD Public Access Policy, developed in response to OSTP memo, was promulgated in USD(AT&L) memo, "Public Access to the Results of Department of Defense-Funded Research," dated 9 July 2014
  - DoD Public Access Plan, following up on USD(AT&L) memo, is in its third draft
- USD(AT&L) memo requires authors to submit "**final, peer-reviewed journal manuscripts**" to the Defense Technical Information Center (DTIC)
  - Requires manuscripts to be freely available to the public not later than 12 months following publication

1. **Agency archive: DTIC**
2. **Content of archive: Author's final peer-reviewed manuscripts, at a minimum**
3. **Approach to public access: TBD**
4. **Embargo period: Up to 12 months**

Limitations to Public Access to Scientific Publications

- NIH restrictions on bulk downloading of articles [PMC FAQ]:

  Crawlers and other automated processes may NOT be used to systematically retrieve batches of articles *from the PMC web site. Bulk downloading of articles from the main PMC web site, in any way, is prohibited because of copyright restrictions.*

  PMC has two auxiliary services that may be used for automated retrieval and downloading of a ***special subset of articles [the PMC Open Access Subset]*** *from the PMC archive.*

- DOE restrictions on bulk downloading of articles [DOE Public Access Plan]:

  The distributed nature of PAGES' full-text content inherently makes unauthorized mass downloading and redistribution more difficult. For the limited full-text content it hosts publicly, ***OSTI will enforce a download limit and post appropriate fair use policies****.*

- Restrictions inhibit analytics based on text and data mining of articles [Society for Scholarly Publishing blog]

  Something that's missing from the DOE's plan appears to be any sort of mechanism for ***text- and data-mining (TDM) of articles****. Bulk download of metadata and abstracts is a good thing, to be certain, but there is increasing demand for full text TDM functionality across scholarly publishing.*

IDA

Public Access to Scientific Data

- OSTP Requirements
- Closer Look at “Scientific Data”
- Discipline-Specific Nature of Public Access to Scientific Data
- Publisher Interest in Public Access to Scientific Data

OSTP Requirements for Agency Public Access Plans — Scientific Data [page 1 of 3]

[Adapted from University of Michigan Inter-university Consortium for Political and Social Research (ICPSR) "Guidelines for OSTP Data Access Plan" at webpage]

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

Increased access is constrained by other requirements (e.g., privacy, security, IP rights), as well as by cost

OSTP Requirements for Agency Public Access Plans — Scientific Data [page 2 of 3]

- Use of data management plans (DMPs)
  - OSTP 4(b): "Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans and, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified"
- Include cost of data management in funding proposals
  - OSTP 4(c): "Allow the inclusion of appropriate costs for data management and access in proposals for Federal funding for scientific research"
- Evaluate data management plans
  - OSTP 4(d): "Ensure appropriate evaluation of the merits of submitted data management plans“
- Ensure researcher compliance with DMPs
  - OSTP 4(e): "Include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies“
- Promote public deposit of data
  - OSTP 4(f): "Promote the deposit of data in publicly accessible databases, where appropriate and available"

Researchers formulate DMPs; Reviewers evaluate DMPs; Agencies track compliance to DMPs

OSTP Requirements for Agency Public Access Plans — Scientific Data [page 3 of 3]

- Private-sector cooperation to improve access
  - OSTP 4(g): "Encourage cooperation with the private sector to improve data access and compatibility, including through the formation of public-private partnerships with foundations and other research funding organizations"
- Mechanisms for identification & attribution of data
  - OSTP 4(h): "Develop approaches for identifying and providing appropriate attribution to scientific data sets that are made available under the plan"
- Data stewardship workforce development
  - OSTP 4(i): "In coordination with other agencies and the private sector, support training, education, and workforce development related to scientific data management, analysis, storage, preservation, and stewardship"
- Long-term support for repository development
  - OSTP 4(j): "Provide for the assessment of long-term needs for the preservation of scientific data in fields that the agency supports and outline options for developing and sustaining repositories for scientific data in digital formats, taking into account the efforts of public and private sector entities“

Agencies participate in development of ecosystem

IDA | Closer Look at "Scientific Data" (aka Research Data)

- OMB circular A-110 [30 Sep 1999]: "**Research data** is defined as
  - the recorded factual material commonly accepted in the scientific community as necessary to validate research findings,
  - but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples).
  - Research data **also do not include**: (A) **Trade secrets, commercial information, materials necessary to be held confidential** by a researcher until they are published, or similar information which is protected under law; and (B) **Personnel and medical information** and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."
- OSTP memo [22 Feb 2013]: "For purposes of this memorandum, **data** is defined, consistent with OMB circular A-110, as
  - the **digital** recorded factual material commonly accepted in the scientific community as necessary to validate research findings
  - including data sets used to **support scholarly publications**,
  - but **does not include laboratory notebooks**, preliminary analyses, drafts of scientific papers, plans for future research, peer review reports, communications with colleagues, or physical objects, such as laboratory specimens."
- OSTP memo calls for maximizing access to digital scientific data, while
  - "protecting confidentiality and personal privacy"
  - "recognizing proprietary interests, business confidential information, and intellectual property rights and avoiding significant negative impact on intellectual property rights, innovation, and U.S. competitiveness"

Key concepts:
1. **Role of scientific community**
2. **Validation of research findings**

OMB addresses security, privacy, and IP through definition of "research data"

OSTP limits "data" to "digital data," and excludes lab notebooks

OSTP explicitly calls out data sets supporting scholarly publications

OSTP addresses security, privacy, and IP in statements of objectives

Discipline-Specific Nature of Public Access to Scientific Data: e.g., NIH Data Sharing Policy

NIH Data Sharing Policy and Implementation Guidance, 5 March 2003

- "Given the breadth and variety of science that NIH supports, neither the precise content for the data documentation, nor the formatting, presentation, or transport mode for data is stipulated."
- "What is sensible in one field or one study may not work at all for others."
- "It would be helpful for members of multiple disciplines and their professional societies to discuss data sharing, determine what standards and best practices should be proposed, and create a social environment that supports data sharing."

NIH Policy, formulated in 2003, recognizes discipline-specific nature of data sharing policies and practices

Discipline-Specific Nature of Public Access to Scientific Data: e.g., NSF Perspective

NSF Data Management and Sharing FAQ, updated 30 Nov 2010

- What constitutes "data" covered by a DMP?
- Am I required to deposit my data in a public database?
- Does data management and access include supporting documentation and metadata, such as validation protocols, field notebooks, etc.?
- How long should data be archived and made accessible?
- Does this policy mean that I must make my data available immediately, even before publication?
- What are NSF's expectations regarding the release of data that include sensitive information (e.g., information about individuals or locations of endangered species)?

Still today, answers to all these kinds of questions are "determined by the community of interest through the process of peer review and program management"

Discipline-Specific Nature of Scientific Data: e.g., *Science Magazine* Guidance on Repositories

Science Magazine, General Information for Authors: "appropriate data sets … must be deposited in an approved database"

- ***Molecular structure data.*** Approved repositories: Worldwide Protein Data Bank [through the Research Collaboratory for Structural Bioinformatics, Macromolecular Structure Database (MSD EMBL-EBI), or Protein Data Bank Japan], BioMag Res Bank, and Electron Microscopy Data Bank (MSD-EBI), and for synthetic molecules, the Cambridge Crystallographic Data Centre
- ***DNA and protein sequences***. Approved repositories: GenBank of other members of the International Nucleotide Sequence Database Collaboration (EMBL or DDBJ) and SWISS-PROT
- ***Microarray data***. Approved repositories: Gene Expression Omnibus and ArrayExpress
- ***Climate data***. Approved repositories: NOAA climate repository or other public databases
- ***Ecological data***. Recommended repository: Dryad

Science Magazine advocates use of public discipline-specific research data repositories

Discipline-Specific Nature of Scientific Data: e.g., *Nature* Journals Guidance on Repositories

For the following types of data set, submission to a community-endorsed, public repository is mandatory. Accession numbers must be provided in the paper. Examples of appropriate public repositories are listed below.

Mandatory deposition	Suitable repositories
Protein sequences	Uniprot
DNA and RNA sequences	Genbank
	DNA DataBank of Japan (DDBJ)
	EMBL Nucleotide Sequence Database (ENA)
DNA and RNA sequencing data	NCBI Trace Archive
	NCBI Sequence Read Archive (SRA)
Genetic polymorphisms	dbSNP
	dbVar
	European Variation Archive (EVA)
Linked genotype and phenotype data	dbGAP
	The European Genome-phenome Archive (EGA)
Macromolecular structure	Worldwide Protein Data Bank (wwPDB)
	Biological Magnetic Resonance Data Bank (BMRB)
	Electron Microscopy Data Bank (EMDB)
Microarray data (must be MIAME compliant)	Gene Expression Omnibus (GEO)
	ArrayExpress
Crystallographic data for small molecules	Cambridge Structural Database

http://www.nature.com/authors/policies/availability.html

Recommendations for other datasets

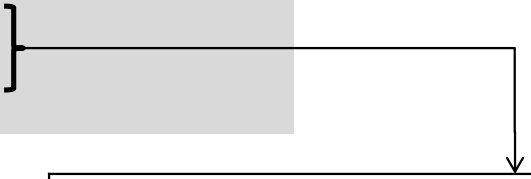
In addition to these mandates, the preferred way to share any data sets is via **public repositories**. Scientific Data, a sister publication to Nature journals, maintains a list of **approved and recommended data repositories organized by discipline**. Please consult this list to identify an appropriate repository for your data sets. When repositories do no exist for a particular data type, authors can deposit and share data via **figshare or Dryad, two general-purpose scientific data repositories**.

Nature Journals Recommended Repositories by Scientific Discipline

http://www.nature.com/sdata/data-policies/repositories

View data repositories

- Life sciences:
  - nucleic acid sequence; protein sequence; molecular & supramolecular structure; neuroscience; omics; taxonomy & species diversity; life-science community resources
- Chemistry & chemical biology
- Environmental & geoscience
- Physics, astrophysics & astronomy
- Social sciences
- General science



Physics, astrophysics & astronomy

Reaction Database Standard Search Interface	view re3data entry
SIMBAD Astronomical Database	view re3data entry
UK Solar System Data Centre	

Social sciences

Harvard Dataverse Network	view re3data entry

Publisher Interest in Public Access to Scientific Data

Science, General Information for Authors.
http://www.sciencemag.org/site/feature/contribinfo/prep/gen_info.xhtml

Data and materials availability All **data** necessary to **understand, assess, and extend** the conclusions of the manuscript must be available to any reader of *Science*. All **computer codes** involved in the creation or analysis of data must also be available to any reader of *Science*….
Science supports the efforts of databases that aggregate published data for the use of the scientific community. Therefore, **appropriate data sets** (including microarray data, protein or DNA sequences, atomic coordinates or electron microscopy maps for macromolecular structures, and climate data) **must be deposited in an approved database** [specified below], and an accession number or a specific access address must be included in the published paper….

Large data sets with no appropriate approved repository must be housed as **supplementary materials** **at *Science***, or only when this is not possible, on an **archived institutional Web site**, provided a copy of the data is held in escrow at *Science* to ensure availability to readers….

Publishers have vested interest in availability of scientific data underlying journal articles.

IDA | Joint Editorial: *Science* and *Nature*, 5 Nov 2014

Published Online November 5 2014
Science 7 November 2014:
Vol. 346 no. 6210 p. 679
DOI: 10.1126/science.aaa1724

EDITORIAL

Journals unite for reproducibility

Marcia McNutt

» Marcia McNutt Editor-in-Chief Science Journals

NATURE | EDITORIAL

Journals unite for reproducibility

Consensus on reporting principles aims to improve quality control in biomedical research and encourage public trust in science.

05 November 2014

First-ever Joint Science/Nature Editorial

"Reproducibility, rigour, transparency and independent verification are cornerstones of the scientific method…. A transparent and rigorous approach … ensures that science moves forward, through **independent verifications** as well as the course corrections that come from refutations and **the objective examination of the resulting data**…."

In June 2014, NIH convened a gathering of editors (representing 30 major journals), funding agency representatives, and scientific leaders to address reproducibility. "The attendees agreed on a common set of **Principles and Guidelines in Reporting Preclinical Research** that list proposed journal policies and author reporting requirements in order to promote transparency and reproducibility…."

Among the guidelines is the statement: **"Journals should recommend deposition of data in public repositories, where available, and link data bidirectionally when the paper is published…."**

IDA

Data Management Plans (DMPs)

- OSTP Approach
- NIH Data Sharing Plans
- NSF Data Management Plans (DMPs)
- DOE DMPs
- DMPTool

OSTP Approach to Discipline-Specific Nature of Data Sharing Policies and Practices

- **Agencies establish requirements for DMPs**
  - Some may apply on agency-wide basis (e.g., NSF)
  - Some may apply to specific sub-agency component (e.g., NSF Directorate)
  - Some, at **Program Manager's discretion**, may apply to specific program or solicitation
- **Researchers, taking into account best practices of their scientific discipline, formulate DMPs**
  - "Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified"
- **Agencies, in part through reviewers engaged in merit review process, evaluate DMPs**
  - "Ensure appropriate evaluation of the merits of submitted data management plans"
- **Agencies, in part through Program Manager's efforts, monitor compliance with DMPs**
  - "Include mechanisms to ensure that intramural and extramural researchers comply with data management plans and policies"

OSTP provides guiding principles and outlines a collaborative process

IDA | NIH Data Sharing Plans

NIH Data Sharing Requirements

- Are documented in "NIH Data Sharing Policy and Implementation Guidance," dated 5 March 2003
- Have not been revised to align with OSTP Memo

Recommended elements of Data Sharing Plan (required for proposals) (quoting)**:**

- Expected schedule for data sharing
- Format of the final dataset
- Documentation to be provided
- Whether or not any analytic tools also will be provided
- Whether or not a data-sharing agreement will be required and, if so, a brief description of such an agreement
- Mode of data sharing (e.g., under their own auspices by mailing a disk or posting data on their institutional or personal website, through a data archive or enclave)

Policy (quoting)

- Investigators seeking $500,000 or more in direct costs in any year should include a description of how **final research data** will be shared, or explain why data sharing is not possible
- [Final research data] does not mean summary statistics or tables; rather, it means the **data on which summary statistics and tables are based**
- It is especially important to share **unique data** that cannot be readily replicated.
- **Reviewers will not factor the proposed data-sharing plan into the determination of scientific merit** or priority score.
- **Program staff will be responsible for overseeing the data sharing policy** and for assessing the appropriateness and adequacy of the proposed data-sharing plan.

IDA | NSF Data Management Plans (DMPs)

NSF DMP Requirements

- Are documented in NSF 15-1, "Proposal and Award Policies and Procedures Guide (PAPPG)," Effective 26 December 2014
- Refer to policy, "Dissemination and Sharing of Research Results," in NSF 11-1, Part II, Award & Administration Guidelines (AAG), dated January 2011
- Cover "**data management and *sharing of the products of research***"

Recommended elements of DMP (supplementary document of up to 2 pages, required for proposals) (quoting)**:**

- 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
- 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
- plans for archiving data, samples, and other research products, and for preservation of access to them

Policy (quoting)

- A valid Data Management Plan may include only the statement that **no detailed plan is needed**, as long as the statement is accompanied by a clear justification.
- The Data Management Plan will be reviewed as an integral part of the proposal, **considered under Intellectual Merit or Broader Impacts or both**, as appropriate for the **scientific community of relevance**.

IDA | DOE Data Management Plans (DMPs)

DOE DMP Requirements

- Align with 22 February 2013 OSTP Memo
- Align with DOE Public Access Plan, dated 24 July 2014 and released 4 August 2014
- Are documented in DOE Office of Science "Statement on Digital Data Management," dated 28 July 2014
- Apply to all Office of Science research solicitations issued on or after 1 October 2014

Suggested elements of DMP (quoting)**:**

- **Data Types and Sources**. A brief, high-level description of the data to be generated or used through the course of the proposed research and which of these are considered digital research data necessary to validate the research findings.
- **Content and Format**. A statement of plans for data and metadata content and format
- **Sharing and Preservation**. A description of the plans for data sharing and preservation. This should include, when appropriate: … cost/benefit considerations to support whether/ where the data will be preserved after direct project funding ends and any plans for the transfer of responsibilities for sharing and preservation …
- **Protection**. A statement of plans, where appropriate and necessary, to protect confidentiality, personal privacy, PII, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; and avoid significant negative impact on innovation, and U.S. competitiveness.
- **Rationale**. A discussion of the rationale or justification for the proposed data management plan including, for example, the potential impact of the data within the immediate field and in other fields, and any broader societal impact.

Policy (quoting):

- The **Principal Investigator** should determine which data should be the subject of the DMP and, in the DMP, propose which data should be shared and/or preserved….
- DMPs should reflect **relevant standards and community best practices for data and metadata**, and make use of **community accepted repositories** whenever practicable.
- DMPs will be reviewed as part of the overall Office of Science **research proposal merit review process**.

IDA | DMPTool: Widely Used Resource for Creating DMPs

- DMPTool is a service of the University of California Curation Center of the California Digital Library
  - **Facilitates development of DMPs**
  - Suggested by many universities for use by their researchers
- DMPTool partners
  - Institutions, profit and nonprofit entities, individuals, other groups:
    - Include DOE, NIH, NSF, USGS, etc.
    - Include 21 of top 25 universities (measured in terrms of Federal R&D funding)
- DMPTool resources
  - Links to funder guidance, templates, and sample plans
    - Covers 23 funders, including DOE, NIH, NSF (generic), NSF Directorates (several), USGS, etc.
    - e.g., for DOE, DMPTool links to the DOE Statement on digital data management
  - Public DMPs, plans created using DMPTool and shared publicly by owners


DMPTool
Home
DMP Requirements
Public DMPs
News
Help
Contact Us
About
Log In
Data Management Planning Tool
Create, review, and share data management plans that meet institutional and funder requirements.
Get Started
PUBLIC DMPS
List of sample data management plans provided by DMPTool users.
A unified approach to preserving cultural software objects and their development histories
Arthropod responses to grassland nutrient limitation
Atmospheric CO2 Concentrations, Mauna Loa Observatory, Hawaii, 2011-2013
DMPTOOL NEWS
Latest information about data management and the DMPTool.
Got an idea for DMPTool? Share it!
DMPTool unavailable, Saturday October 1...
US Dept of Energy data management requi...
We need API use cases!
DMPTool downtime this Saturday 7/12/14
DMPTOOL HELP
Overview of how to use the tool, plus resources and guidance on data management.
Frequently Asked Questions
Create a DMP
Administer the DMPTool
Data management guidance
Community resources

IDA | DMPTool: Funder Guidance, Templates, Sample Plans

Use the A-Z links below to narrow down the list by institution or use the search box to search for specific DMP Templates. Sample plans are not necessarily from the funder.

Search

A - F | G - L | M - S | T - Z | All

Template	Funder	Funder Links	Sample Plans (if available)
Alfred P. Sloan Foundation	Alfred P. Sloan Foundation	Alfred P. Sloan Foundation	
Department of Energy: DOE	Department of Energy (DOE)	DOE Statement on digital data management	
DMP Template from DCC	University of California, Office of the President	Information on DCC website	
⋮			
National Institutes of Health	National Institutes of Health	Guidance	NIH: Sample Plans
⋮			
NSF-ENG: Engineering	National Science Foundation	Guidance	
NSF-GEN: Generic	National Science Foundation	NSF Grant Proposal Guide NSF Data Management FAQ NSF - Dissemination of research results	NSF-GEN Sample 1 NSF-GEN Sample 2
NSF-PHY: Physics	National Science Foundation	Advice to PIs	
NSF-SBE: Social, Behavioral, Economic Sciences	National Science Foundation	Contents	
U.S. Geological Survey DMP Guidance	U.S. Geological Survey	USGS Data Management Plans	

IDA | University Adoption of DMPTool [page 1 of 2]

	Institution (top 25, ordered by Federal R&D)	FY2012 Federal R&D[1]	DMP-Focused Webpage
1	Johns Hopkins U. (including APL)	$1,858M	http://dmp.data.jhu.edu/assistance/guidance-on-writing-data-management-plans/
2	U. WA, Seattle	$910M	http://guides.lib.washington.edu/dmg
3	U. MI, Ann Arbor	$792M	http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/index.html
4	U. PA	$670M	http://guides.library.upenn.edu/dmp
5	U. CA, San Diego	$657M	http://idi.ucsd.edu/data-curation/dmp.html
6	Columbia U. in the City of New York	$646M	http://scholcomm.columbia.edu/data-management/data-management-plan-templates/
7	U. Pittsburgh, Pittsburgh[2]	$638M	http://www.library.pitt.edu/data-management
8	Stanford U.	$636M	http://library.stanford.edu/research/data-management-services/data-management-plans
9	U. NC, Chapel Hill	$606M	http://guides.lib.unc.edu/researchdatatoolkit
10	Harvard U.	$590M	http://isites.harvard.edu/icb/icb.do?keyword=k78759&pageid=icb.page407320
11	Duke U.	$587M	http://library.duke.edu/data/guides/data-management
12	U. WI, Madison	$581M	http://researchdata.wisc.edu/
13	U. CA, San Francisco	$559M	http://hub.ucsf.edu/data-management
14	U. CA, Los Angeles	$539M	http://www.library.ucla.edu/support/publishing-data-management/data-management-curation-services/data-management-plans
15	PA State U., University Park and Hershey Medical Ctr.	$531M	http://www.libraries.psu.edu/psul/pubcur/datamanagement.html

IDA | University Adoption of DMPTool [page 2 of 2]

	Institution (top 25, ordered by Federal R&D)	FY2012 Federal R&D[1]	DMP-Focused Webpage
16	Yale U.	$518M	http://guides.library.yale.edu/dmp
17	MA Institute of Technology[2]	$496M	http://libraries.mit.edu/data-management/
18	U. MN, Twin Cities	$485M	https://www.lib.umn.edu/datamanagement
19	GA Institute of Technology	$484M	http://www.library.gatech.edu/research-data/data-management-plan
20	Cornell U.	$473M	http://data.research.cornell.edu/
21	Vanderbilt U.[2]	$449M	http://researchguides.library.vanderbilt.edu/datamanagement
22	OH State U.	$446M	https://library.osu.edu/researchcommons/data-management-services/
23	U. Southern CA[2]	$444M	https://research.usc.edu/dcg/proposal-preparation/
24	Washington U., St. Louis	$441M	https://research.wustl.edu/PGC/Pages/ProposalDevelopment.aspx
25	Northwestern U.	$393M	http://www.library.northwestern.edu/dmp

1 FY 2012 Federally Financed R&D Expenditures [http://ncsesdata.nsf.gov/herd/2012/html/HERD2012_DST_05.html]
2 Not a contributing member of DMPTool [https://dmptool.org]. The other 21 universities are members. Of the 4 non-members, all but U. Southern CA point to DMPTool.

All but 4 of these top 25 universities are contributing members of DMPTool

IDA

Evolving Public Access Ecosystem

- Identifiers
  - Articles, Datasets, Funders, Researchers
  - Searchable Metadata Repositories
- Publication Archives
- Data Repositories
- Metadata Standards
- Licenses

IDA | Identifiers for Articles and Datasets

- Article Identifiers
  - Digital Object Identifier (DOI) assigned by coordinated effort of CrossRef (DOI Registration Agency for scholarly and professional research content) and publisher
  - CrossRef has over 5,000 Members, including DOE Office of Scientific and Technical Information (OSTI) (DOI prefix for DOE/OSTI is 10.2172)
  - CrossRef provides searchable metadata repository at http://search.crossref.org/
- Dataset Identifiers
  - DOI assigned by coordinated effort of DataCite (DOI Registration Agency for datasets), DataCite member ("allocator"), and data publisher
  - DataCite has 3 members in U.S.: California Digital Library, DOE/OSTI, Purdue University Libraries [https://www.datacite.org/members]
  - DataCite provides searchable metadata repository at http://search.datacite.org/ui

IDA | Identifiers for Funders and Researchers

- Funder Identifiers
  - FundRef has a taxonomy of funders containing 8,180 funders
    - DOI Link for DOE is http://dx.doi.org/10.13039/100000015, where "http://dx.doi.org" is the DOI Directory, "10.13039" is the FundRef DOI Prefix, and "100000015" is the DOI Suffix representing DOE
  - Pilot publishers and funders included DOE/OSTI, AIP Publishing, Nature Publishing Group, NASA, and NSF [see http://www.crossref.org/01company/pr/news052813.html]
  - CrossRef provides searchable metadata repository at http://search.crossref.org/fundref
- Researcher Identifiers
  - ORCID maintains a registry of unique researcher identifiers and provides a method of linking research activities and outputs to these identifiers
  - Federal agencies that are members of ORCID include DOE/OSTI and FDA [see http://orcid.org/about/community/members]
  - ORCID provides searchable metadate repository at http://orcid.org/

IDA | Closer Look at Article IDs (DOIs and DOI Links)

http://dx.doi.org/10.1006/jmbi.1995.0238

DOI Directory | Prefix | Suffix

DOI = DOI Prefix + DOI Suffix

DOI Link = DOI Directory URL (http://dx.doi.org) + DOI

Clicking on DOI Link
http:/dx.doi.org/10.1006/jmbi.1995.0238

Article is at URL:
http://www.sciencedirect.com/science/article/pii/S0022283685702380

URLs in the DOI directory are updated when an item changes location, ensuring persistence. The DOI System and directory is managed by the International DOI Foundation.

CrossRef DOIs
CrossRef members deposit DOIs by submitting XML containing citation metadata and DOIs to the CrossRef system. CrossRef stores the metadata in our lookup service, and registers the DOI and URL in the central DOI directory

Source: http://help.crossref.org/how_do_dois_work

IDA | Closer Look at Dataset IDs

- DataCite leverages DOI infrastructure, but maintains extensibility to other identifier systems
- DataCite DOIs resolve to a public landing page that contains information about the associated dataset and a direct link to the dataset itself

IDA | Closer Look at Funder IDs and FundRef Taxonomy

15 Executive Departments
(DOE DOI is 10.13039/100000015)

- Dept of Defense
  http://dx.doi.org/10.13039/100000005
- Dept of Energy (DOE)
  http://dx.doi.org/10.13039/100000015
- Dept of Health & Human Services
  http://dx.doi.org/10.13039/100000016
- Dept of Justice
  http://dx.doi.org/10.13039/100000074
- Dept of Education
  http://dx.doi.org/10.13039/100000138
- Dept of Transportation
  http://dx.doi.org/10.13039/100000140
- Dept of Homeland Security
  http://dx.doi.org/10.13039/100000180
- …
- Dept of Veterans Affairs
  http://dx.doi.org/10.13039/100000738
- Dept of Labor
  http://dx.doi.org/10.13039/100000778
- Dept of the Treasury
  http://dx.doi.org/10.13039/100006097

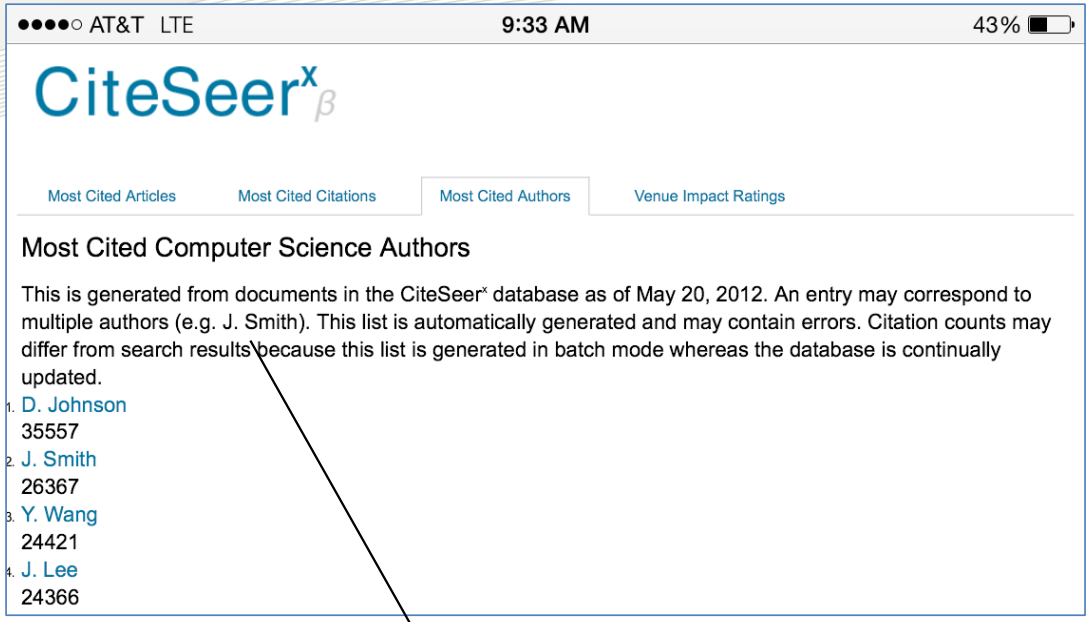
24 DOE Top-level Organizations

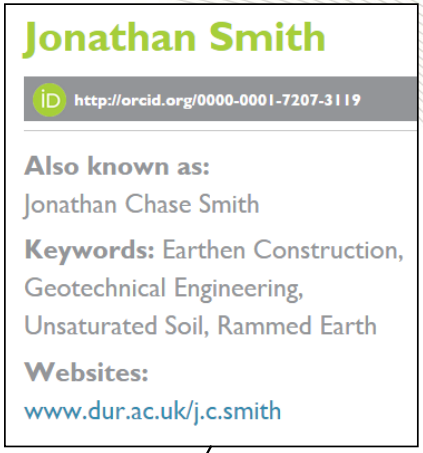
- Energy Information Administration
  http://dx.doi.org/10.13039/100006162
- Office of Electricity Delivery and Energy Reliability
  http://dx.doi.org/10.13039/100006130
- …
- Office of Science
  http://dx.doi.org/10.13039/100006132
- Advanced Research Projects Agency – Energy
  http://dx.doi.org/10.13039/100006133
- …
- Oak Ridge National Laboratory
  http://dx.doi.org/10.13039/100006228
- …

6 DOE Office of Science Organizations

- Workforce Development for Teachers and Scientists
  http://dx.doi.org/10.13039/100006210
- Nuclear Physics
  http://dx.doi.org/10.13039/100006209
- Fusion Energy Sciences
  http://dx.doi.org/10.13039/100006207
- High Energy Physics
  http://dx.doi.org/10.13039/100006208
- Biological and Environmental Research
  http://dx.doi.org/10.13039/100006206
- Basic Energy Sciences
  http://dx.doi.org/10.13039/100006151

IDA | Closer Look at Researcher IDs





Motivation for Unique Researcher IDs:
Search for articles by J. Smith in CiteSeer yields 9,606 hits. 1st hit has author John R. Smith, 2nd hit has author Alan Jay Smith, 3rd hit has author Jonathan M. Smith, etc.

Unique ORCID ID for Jonathan Chase Smith:
http://orcid.org/0000-0001-7207-3119

Sample Search on CrossRef Metadata Repository using FundRef Taxonomy and Funder ID

Search on http://search.crossref.org/fundref for articles with funder "High Energy Physics," which falls under DOE/Office of Science in FundRef Taxonomy, yields 2 articles

IDA | Publication Archives: LOCKSS and CLOCKSS

- LOCKSS (Lots of Copies Keep Stuff Safe)
  - LOCKSS box: Digital bookshelf with automated preservation referred to as LOCKSS technology
  - Libraries install and maintain their own LOCKSS boxes—running LOCKSS software—to facilitate purchasing content and building permanent library collections, rather than leasing temporary access to materials
  - LOCKSS box content is available to an institution's own readers when the publisher's site is unavailable, thus providing 100% continual access and 100% post-cancellation access
- CLOCKSS (Controlled LOCKSS)
  - CLOCKSS: Global dark archive that preserves content using LOCKSS technology
  - CLOCKSS preserves content in 12 strategically chosen libraries in Asia, Europe, and North America to optimize the content's safety against political and environmental threats
  - Many libraries support CLOCKSS through financial contributions and by participating in archive governance; they do not store any technology locally
  - CLOCKSS preserved content is available free to everyone when it is not available from a publisher or "triggered" from the dark archive

http://www.clockss.org/clockss/FAQ

IDA | Data Repositories: 992 Registered in Databib

Databib: Searchable Registry of 992 Research Data Repositories at databib.org

Agriculture (18)
Area, Ethnic, and Gender Studies (9)
Biological Sciences (174)
Business (2)
Communications and Information Sciences (3)
Ecosystem Sciences (15)
Education (9)
Engineering (8)
Environmental Sciences (87)
Fine and Performing Arts (3)
Geosciences (117)
Health and Medical Sciences(60)
History (2)
Interdisciplinary (9)
Language and Literature (11)
Law and Legal Studies (2)
Mathematical and Physical Sciences (106)
Multidisciplinary (179)
Philosophy and Religion (1)
Social Sciences (92)
Unclassified (85)

Communications and Information Sciences

Cooperative Association for Internet Data Analysis (CAIDA), The
Informatics Research Data Repository (IDR) [NII]
Social Computing Data Repository

Title: Cooperative Association for Internet Data Analysis (CAIDA), The
URL: http://www.caida.org/data/
Other Metadata includes ***Authority, Subjects, Description, Access, Country, Reuse, Deposit, Type***

IDA | Data Repositories: 1029 Registered in re3data.org

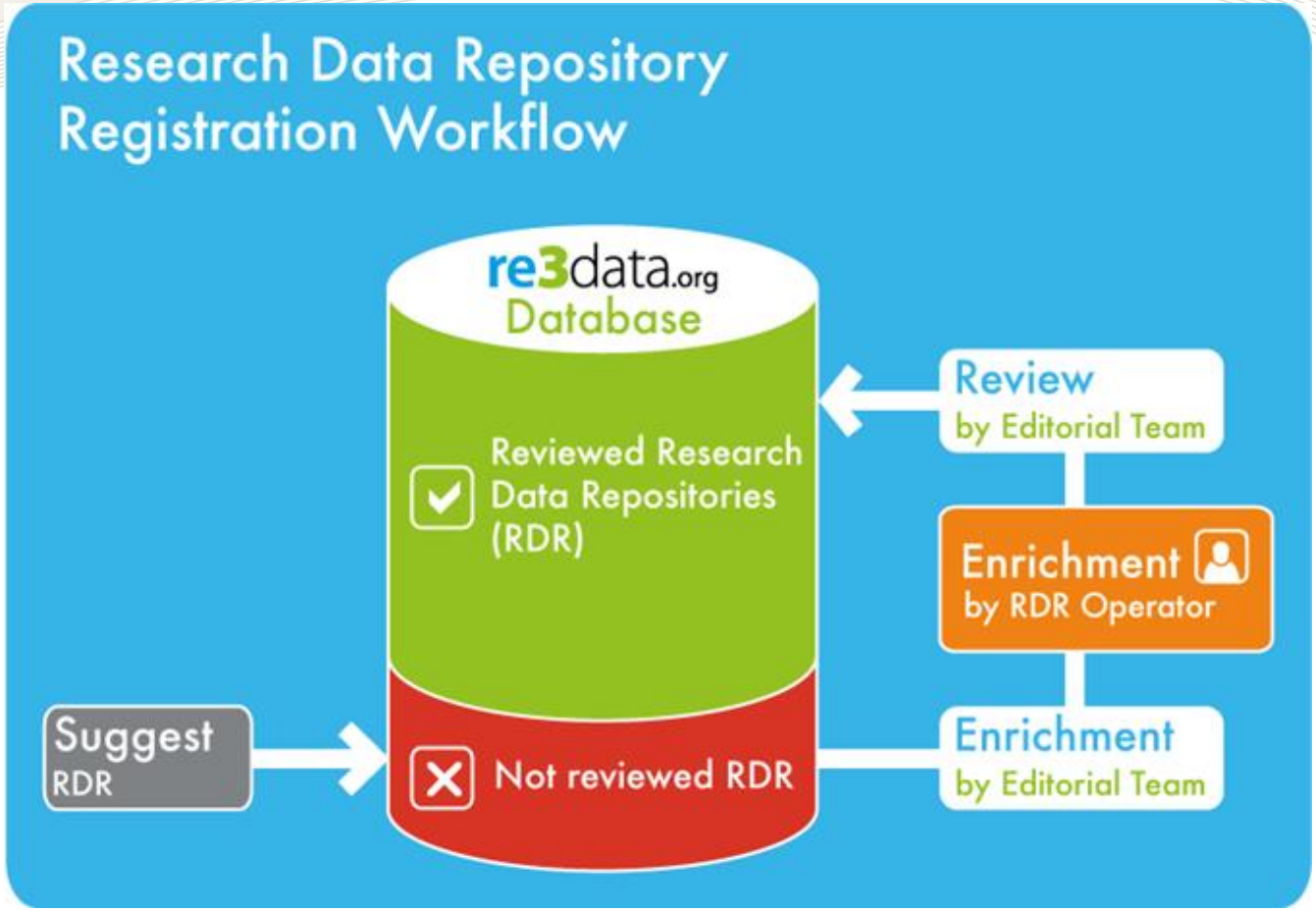
Registry of **Re**search Data **Re**positories (re3data.org)

- 152 subject categories
- Data repository can have multiple subjects, e.g., CAIDA Data has 2 subjects

Name of repository	**CAIDA Data**
Additional name	The Cooperative Association for Internet Data Analysis
Repository URL	http://www.caida.org/data/
Subjects	Computer Science, Electrical and System Engineering; Engineering Sciences
Description	The Cooperative Association for Internet Data Analysis (CAIDA) is a collaborative undertaking among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure.It is an independent analysis and research group with particular focus on: Collection, curation, analysis, visualization, dissemination of sets of the best available Internet data, providing macroscopic insight into the behavior of Internet infrastructure worldwide, improving the integrity of the field of Internet science, improving the integrity of operational Internet measurement and management, informing science, technology, and communications public policies.
Content types	Images; Raw data; Scientific and statistical data formats; Software applications; Structured graphics; Structured text; other
Keywords	internet; analyzing; visualizing; topology; security; traffic analysis; internet economics; internet topology structure; internet infrastructure; UCSD network telescope; DNS Domain name system; internet security; cyber security; internet mapping project; Ark - Archipelago

Other metadata includes ***Repository type, Mission statement for designated community, Research data repository language(s), Data and/or service provider***

IDA | Data Repositories: Re3data Registration Workflow



Source: http://www.re3data.org/faq/

IDA | Data Repositories: Merger of Databib and re3data.org

- Databib, online in April 2012
  - Purdue University Libraries
- Registry of Research Data Repositories (re3data.org), online in August 2012
  - Library and Information Services of GFZ German Research Centre for Geosciences, Computer and Media Service at Humboldt-Universität zu Berlin, and KIT Library at the Karlsruhe Institute of Technology
- DataCite, formed December 2009
  - Not-for-profit organization formed in London on December 1, 2009, with an aim to:
    - Establish easier access to research data on the Internet
    - Increase acceptance of research data as legitimate, citable contributions to the scholarly record
    - Support data archiving that will permit results to be verified and re-purposed for future study
- On 25 March 2014, Databib and re3data.org announced plans to merge into **one service managed under auspices of DataCite** by end of 2015
  - Aim is "to reduce duplication of effort and to better serve the research community with a **single, sustainable registry of research data repositories**…."
  - Joint registry will operate under name "re3data.org – Registry of Research Data Repositories"
  - Editorial board will retain name Databib

IDA | Data Repository: DRYAD

Source: re3data.org

Name of repository	**DRYAD**
Repository URL	http://datadryad.org
Subjects	Agriculture, Forestry, Horticulture and Veterinary Medicine; Basic Biological and Medical Research; Biochemistry and Animal Physiology; Bioinformatics and Theoretical Biology; Biology Evolution, Anthropology; General Genetics; Geology and Palaeontology; Geosciences (including Geography); Humanities and Social Sciences; Life Sciences; Medicine; Microbial Ecology and Applied Microbiology; Microbiology, Virology and Immunology; Natural Sciences; Plant Ecology and Ecosystem Analysis; Plant Sciences; Social and Behavioural Sciences; Virology; Zoology
Description	**DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad is an international repository of data underlying peer-reviewed scientific and medical literature, particularly data for which no specialized repository exists. The content is considered to be integral to the published research. All material in Dryad is associated with a scholarly publication.**
Content types	Plain text, Scientific and statistical data formats, Software applications, Source code, Standard office documents, Structured text, other
Keywords	scientific and medical publications, Biodiversity, interdisciplinary
Repository size	5142 data packages, 15457 data files, 312 journals, 18826 authors
Repository type	other
Mission statement for designated community	http://datadryad.org/pages/organization#community
Research data repository language(s)	eng
Data and/or service provider	dataProvider serviceProvider

IDA | Data Repository: Harvard DataVerse Network

Source: re3data.org

Name of repository	**IQSS Dataverse network**
Additional name	IQSS DVN Harvard Dataverse Network
Repository URL	http://thedata.harvard.edu/dvn/
Subjects	Humanities and Social Sciences; Social Sciences; Social and Behavioural Sciences
Description	**The Harvard Dataverse Network is open to all scientific data from all disciplines worldwide. It includes the world's largest collection of social science research data. It is hosting data for projects and archives.**
Content types	Archived data, Databases, Raw data, Scientific and statistical data formats, Software applications, Source code, Standard office documents
Keywords	human societies, social societies, human behavior, demography, epidemiology, automes research, multidisciplinary
Repository size	746 dataverses,54.529 studies,745.760 files
Repository type	disciplinary institutional
Mission statement for designated community	http://www.iq.harvard.edu/book/mission
Research data repository language(s)	eng
Data and/or service provider	dataProvider serviceProvider

IDA | Data Repository: Data.gov

Source: re3data.org

Name of repository	**Data.gov**
Repository URL	http://www.data.gov/
Subjects	Economics; Engineering Sciences; Geosciences (including Geography); Humanities and Social Sciences; Jurisprudence; Life Sciences; Natural Sciences; Social Sciences; Social and Behavioural Sciences
Description	**Data.gov increases the ability of the public to easily find, download, and use datasets that are generated and held by the Federal Government. Data.gov provides descriptions of the Federal datasets (metadata), information about how to access the datasets, and tools that leverage government datasets**
Repository Contact	https://www.data.gov/contact
Content types	Archived data, Images, Plain text, Raw data, Scientific and statistical data formats, Software applications, Standard office documents, Structured graphics, Structured text other
Keywords	atmosphere, biology, business, climate, communication, culture
Repository size	112.509 datasets
Repository type	institutional other
Mission statement for designated community	http://www.data.gov/about
Research data repository language(s)	eng
Data and/or service provider	dataProvider serviceProvider

DataCite Metadata Schema for Publication and Citation of Research Data: Top Level View

ID	DataCite-Property	Obligation
1	Identifier (with type sub-property)	Mandatory
2	Creator (with name identifier and affiliation sub-properties)	Mandatory
3	Title (with optional type sub-properties)	Mandatory
4	Publisher	Mandatory
5	PublicationYear	Mandatory
6	Subject (with scheme sub-property)	Recommended
7	Contributor (with type, name identifier, and affiliation sub-properties)	Recommended
8	Date (with type sub-property)	Recommended
9	Language	Optional
10	ResourceType (with general type description sub-property, e.g., Dataset, Text)	Highly recommended
11	AlternateIdentifier (with type sub-property)	Optional
12	RelatedIdentifier (with type and relation type sub-properties)	Recommended
13	Size	Optional
14	Format	Optional
15	Version	Optional
16	Rights	Optional
17	Description (with type sub-property, e.g., Abstract)	Highly recommended
18	GeoLocation (with point and box sub-properties)	Recommended

Source: http://schema.datacite.org/meta/kernel-3/doc/DataCite-MetadataKernel_v3.1.pdf

DataCite Metadata Schema: Detailed View of "Identifier" Property

ID	DataCite-Property	Occ	Definition	Allowed values, examples, other constraints
1	Identifier	1	The Identifier is a unique string that identifies a resource.	DOI (Digital Object Identifier) registered by a DataCite member. Format should be "10.1234/foo"
1.1	identifierType	1	The type of the Identifier.	Controlled List Value: DOI

Occ (i.e, Occurrence) indicates cardinality/quantity constraints for the properties as follows:

0-n = optional and repeatable
0-1 = optional, but not repeatable
1-n = required and repeatable
1 = required, but not repeatable

Extensibility of Dataset Identification:
Schema provides for identifierType and Identifier properties, but currently supports only one identifierType, namely, DOI

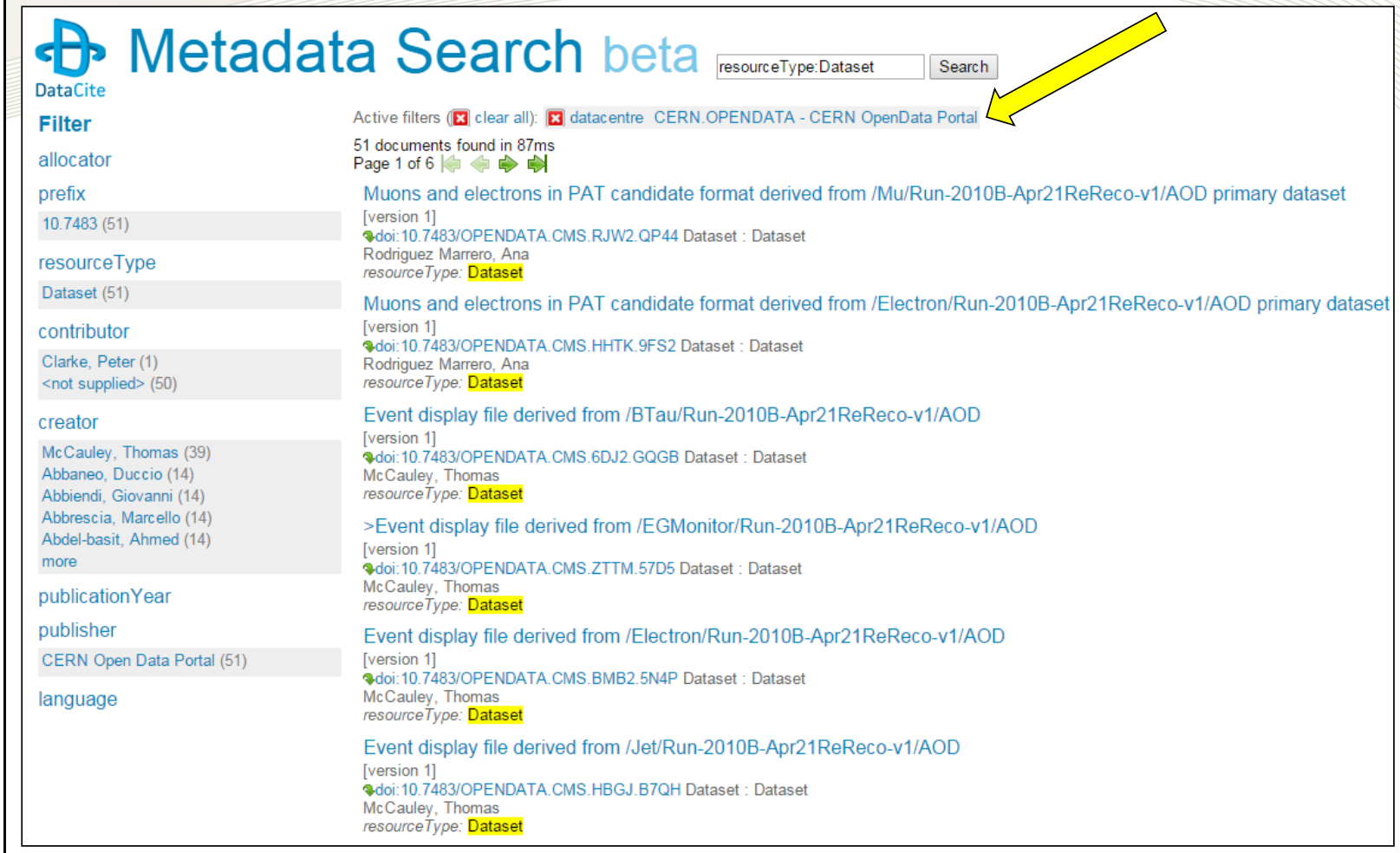
DataCite Metadata Schema: Detailed View of "Creator" Property

ID	*DataCite-Property*	*Occ*	*Definition*	*Allowed values, examples, other constraints*
2	Creator	1-n	Main researchers involved in producing the data, or the authors of the publication, in priority order	May be a corporate/institutional or personal name
2.1	creatorName	1	Name of the creator	Example: Smith, John
2.2	nameIdentifier	0-1	Uniquely identifies an individual or legal entity, according to various schemes	Format is dependent upon scheme
2.2.1	nameIdentifier Scheme	1	Name of the name identifier scheme	Mandatory if nameIdentifier is used. Example: ORCID
2.2.2	schemeURI	0-1	URI of the name identifier scheme	Examples: http://www.isni.org http://orcid.org
2.3	affiliation	0-n	Organizational or institutional affiliation of the creator	Free text

Occ (i.e, Occurrence) indicates cardinality/quantity constraints for the properties as follows:

0-n = optional and repeatable
0-1 = optional, but not repeatable
1-n = required and repeatable
1 = required, but not repeatable

Sample Search for Datasets at a Specified Data Center



IDA | Open License Principles [Project Open Data]

- Reuse
  - Must allow for reproductions, modifications, and derivative works and permit their distribution under the terms of the original work
- Redistribution
  - Shall not restrict any party from selling or giving away the work either on its own or as part of a package made from works from many different sources
  - Shall not require a royalty or other fee for such sale or distribution….
- No Discrimination against Persons, Groups, or Fields of Endeavor
  - Must not discriminate against any person or group of persons
  - Must not restrict anyone from making use of the work in a specific field of endeavor (e.g., it may not restrict the work from being used in a business, or from being used for research)

Open license: "[Grants] permission to access, re-use, and redistribute a work with few or no restrictions"

Creative Commons Licenses that Qualify as "Open": CC0 (public domain), CC BY, CC BY-SA

Attribution
CC BY

This license lets others distribute, remix, tweak, and build upon your work, even commercially, as long as they credit you for the original creation. This is the most accommodating of licenses offered. Recommended for maximum dissemination and use of licensed materials.

View License Deed | View Legal Code

Attribution-ShareAlike
CC BY-SA

This license lets others remix, tweak, and build upon your work even for commercial purposes, as long as they credit you and license their new creations under the identical terms. This license is often compared to "copyleft" free and open source software licenses. All new works based on yours will carry the same license, so any derivatives will also allow commercial use. This is the license used by Wikipedia, and is recommended for materials that would benefit from incorporating content from Wikipedia and similarly licensed projects.

View License Deed | View Legal Code

Attribution-NoDerivs
CC BY-ND

This license allows for redistribution, commercial and non-commercial, as long as it is passed along unchanged and in whole, with credit to you.

View License Deed | View Legal Code

Attribution-NonCommercial
CC BY-NC

This license lets others remix, tweak, and build upon your work non-commercially, and although their new works must also acknowledge you and be non-commercial, they don't have to license their derivative works on the same terms.

View License Deed | View Legal Code

Attribution-NonCommercial-ShareAlike
CC BY-NC-SA

This license lets others remix, tweak, and build upon your work non-commercially, as long as they credit you and license their new creations under the identical terms.

View License Deed | View Legal Code

Attribution-NonCommercial-NoDerivs
CC BY-NC-ND

This license is the most restrictive of our six main licenses, only allowing others to download your works and share them with others as long as they credit you, but they can't change them in any way or use them commercially.

View License Deed | View Legal Code

http://creativecommons.org/licenses/

Responding to OSTP Public Access Memo: Processes, Sample Agency Implementations, and Preliminary Analysis of Rights-in-Data Issues

Karen D. Gordon and Jonathan R. Agre
18 February 2015

Synopsis

Part 1: DOE Implementation of Public Access Policy

The Department of Energy (DOE) was the first agency to have its public access plan approved by OSTP and OMB. The DOE implementation of public access policy includes the following components:

- **DOE Public Access Plan** (24 July 2014): States that DOE is taking a phased approach, with the Office of Science taking the lead with respect to public access to data.
- **DOE communication with stakeholders**:
  - DOE Open Government webpage: Includes a link to the Plan;
  - Press release on DOE public access efforts (4 August 2014).
- **DOE/OSTI capabilities for facilitating public access**:
  - Public Access for Energy and Science (PAGES): Full-text search of publications;
  - DOE Data ID Service: Digital Object Identifiers (DOIs) for datasets;
  - DOE Data Explorer: Search of metadata describing datasets.
- **DOE Office of Science Data Management Policy**:
  - Statement on Digital Data Management webpage: States principles, specifies requirements, and offers additional guidance – includes extensive Frequently Asked Questions (FAQ);
  - Suggested elements of Data Management Plan (DMP): Covers data types and sources, content and format, sharing and preservation, protection, and rationale.

Part 2: NIH Approach to Managing Copyright

The National Institutes of Health (NIH) mandate on submission of peer-reviewed journal articles to PubMed Central dates back to 2008. Over the years, a three-step approach for managing copyright has been developed and could serve as a model for other agencies:

- At award acceptance, the author grants a license to an institutional awardee: The license protects the institutional awardee and ensures NIH rights, because the license stands regardless of the authors' actions with respect to any future publication agreements with a publisher.
- At article submission, the author gives notice to the publisher, informing the publisher of rights that will be retained by author(s), university, and NIH. This preempts the over-reaching copyright transfer provisions of any subsequent publication agreement between author and publisher.
- At article acceptance, the author addendum to publication agreement reiterates the rights retained by author(s), university, and NIH. This protects the author against publisher accusations of misrepresentation in the case that the publication agreement calls for transfer of all rights to publisher.

IDA

Institute for Defense Analyses

4850 Mark Center Drive • Alexandria, Virginia 22311-1882

Responding to OSTP Public Access Memo: Processes, Sample Agency Implementations, and Preliminary Analysis of Rights-in-Data Issues

Dr. Karen D. Gordon

Dr. Jonathan R. Agre

18 February 2015

IDA | Outline

- Background
- Bottom Line Up Front (BLUF) / Executive Summary
- DOE Process for Responding to OSTP Public Access Memo
  - DOE Communication with Stakeholders: Release and Promulgation of *DOE Public Access Plan*
  - DOE Office of Scientific and Technical Information (OSTI) *Public Access Gateway for Energy and Science (PAGES)* and *DOE Data Explorer*
  - DOE Office of Science (SC) *Data Management Policy*
- NIH Process for Managing Copyright with respect to Public Access to Scientific Publications
- Backup: Federal Regulations on Copyright and Government Purpose Rights – Title 2 (Grants and Agreements) and Title 48 (Federal Acquisition Regulation (FAR) and Defense Federal Acquisition Regulation Supplement (DFARS))

IDA | BLUF, Part 1—Status of Agency Public Access Plans

- **24 March 2014**: OSTP report to Congress listed **21 agencies** that had submitted draft public access plans at that time :
  - DHS, DOC/NIST, DOC/NOAA, DoD, DOE, DOI, DOT, ED, EPA, HHS/AHRQ, HHS/ASPR, HHS/CDC, HHS/FDA, HHS/NIH, NASA, NSF, ODNI (IARPA), Smithsonian Institute, USAID, USDA, and VA
- **4 August 2014**: DOE became first agency to release an approved Public Access Plan
  - DOE Public Access Plan is dated 24 July 2014 and was released 4 August 2014
- **13 November 2014:** OSTP report to Congress gave the following status:
  - "We have already provided final clearance on two Department plans, one of which, from the Department of Energy (DOE), is available online. The other will be releasing its final plan shortly…"
- **Between 13 November 2014 and 15 February 2015**: At some point, USDA ***very quietly*** released its Public Access Plan (no readily accessible press releases or notices on USDA website)
  - USDA webpage Open Government at USDA displays title of plan as a hyperlink (Implementation Plan to Increase Public Access to Results of USDA-funded Scientific Research) to the 24-page plan, which is dated 7 November 2014
- **~10 February 2015**: DoD (DTIC) sent a revised draft DoD Public Access Plan to OSTP for review
- **In Progress:** Dr. Steve Thompson, ODNI/AT&F, is working on next revision of the ODNI/IARPA Public Access Plan

BLUF, Part 2—DOE Process for Responding to OSTP Public Access Memo

Several components of DOE's response to OSTP Public Access Memo could serve as a model for IARPA and other agencies:

- **DOE Public Access Plan**: One of only two plans approved by OSTP and OMB to date
- **DOE communication with stakeholders**:
  - DOE Open Government webpage
  - Press release on DOE public access efforts
- **DOE/OSTI capabilities for facilitating public access**:
  - PAGES, DOE/OSTI Data ID Service, DOE Data Explorer
- **DOE Office of Science Data Management Policy**:
  - Statement on Digital Data Management webpage
    - Principles
    - Requirements
    - Additional guidance
  - Suggested elements of Data Management Plan (DMP)
  - Frequently Asked Questions (FAQ)

BLUF, Part 3—NIH Process for Managing Copyright for Public Access to Scientific Publications

NIH mandate on submission of peer-reviewed journal articles to PubMed Central dates back to 2008. Over the years, a 3-step approach for managing copyright has been developed and could serve as a model for IARPA and other agencies:

1. **Author grants license to institutional awardee at award acceptance**
   - Institutional awardee (e.g., university) requires principal investigator to sign an agreement granting the institution certain non-exclusive rights
   - **Protects institutional awardee and ensures NIH rights**. License stands *regardless of authors' actions with respect to any future publication agreements with publisher* (author cannot transfer rights he has already granted to another party)
2. **Author gives notice to publisher at article submission,** informing publisher of rights that will be retained by author(s), university, and NIH
   - **Preempts over-reaching copyright transfer provisions** of any subsequent publication agreement between author and publisher
3. **Author addendum to publication agreement at article acceptance,** reiterating rights retained by author(s), university, and NIH
   - **Protects author against publisher accusations of misrepresentation** in the case that the publication agreement calls for transfer of all rights to publisher (the author cannot transfer all rights if he has already granted non-exclusive, irrevocable rights to his institution)

BLUF, Part 4—Current IARPA Approach to Making Scientific Publications Visible

Strengths of IARPA approach:

- Straightforward – canned Google Scholar queries on contract numbers
- Free and universally available search capability – Google Scholar
- Extensive Coverage – **Over 1200 scholarly publications and over 50 patents/applications** resulting from IARPA-funded research
- Per program results

Challenges:

- Authors must (correctly) include contract number in acknowledgment of IARPA support
- Program Managers (or designees) must keep canned Google Scholar queries up to date
- Google Scholar does not distinguish publications by type
- Many journal articles are behind paywalls

Anomalies:

- False negatives (IARPA-funded articles not found)
  - Contract numbers missing or incorrectly formatted
- False positives (non-IARPA or otherwise inappropriate pubs returned)
  - Article DOIs (sometimes added to canned queries): Results include publications that reference the desired article by DOI
  - BAAs (sometimes added to canned queries): Results include publications that refer to an IARPA program by its BAA number
  - Unfiltered results (e.g., poster presentation vs. journal article)

IDA

DOE Communication with Stakeholders:

Release and Promulgation of DOE Public Access Plan

DOE Homepage

↓

DOE Open Government Webpage

↓

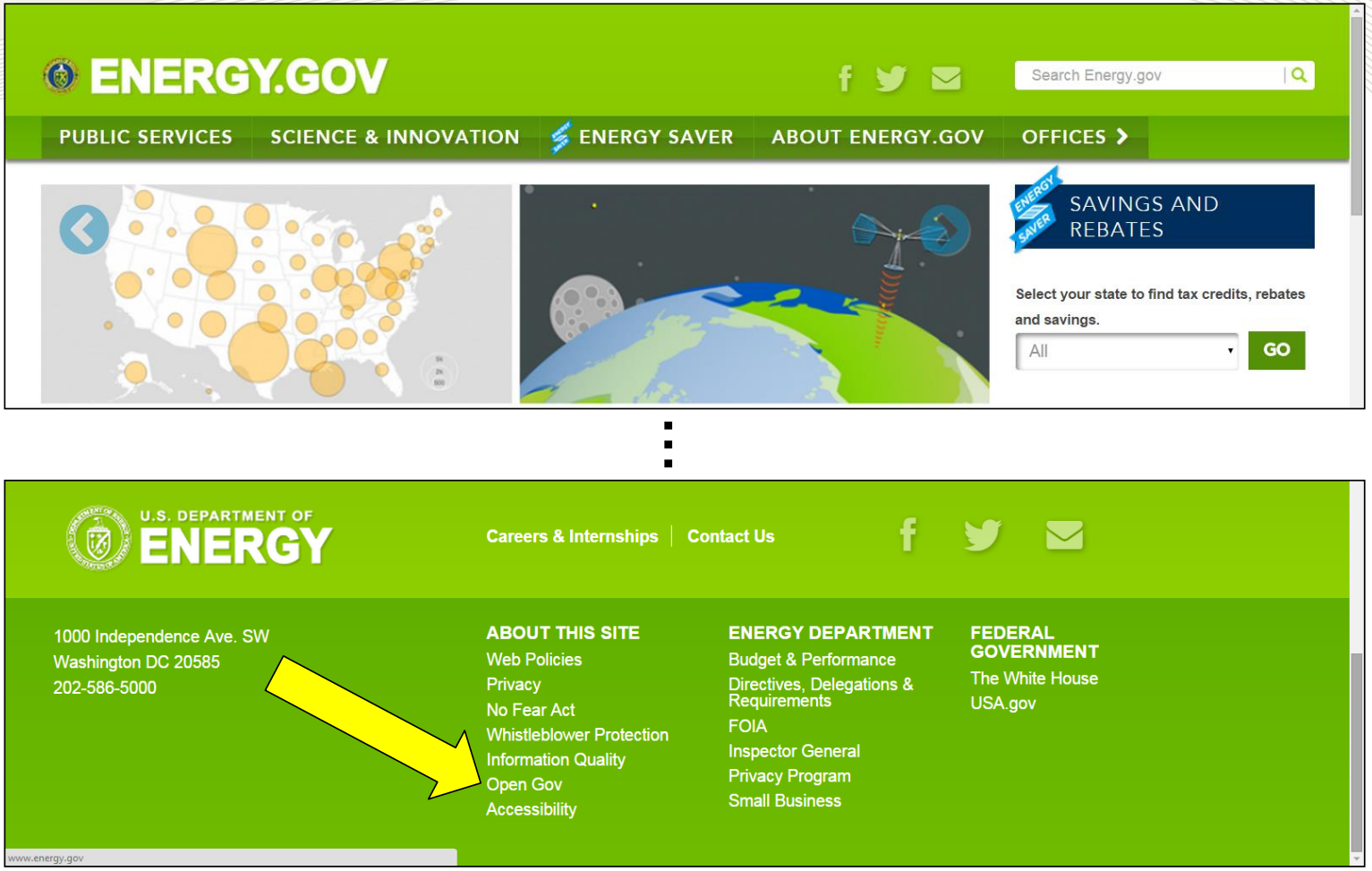
DOE Public Access Plan, 24 July 2014

(states that DOE is taking a phased approach, with DOE Office of Science taking the lead with respect to public access to scientific data)

plus 4 August 2014 DOE Press Release on Public Access Efforts

IDA | DOE: Link to Open Government Webpage

http://www.energy.gov/



IDA | DOE Open Government: Link to Public Access Plan

http://www.energy.gov/open-government

IDA | DOE Public Access Plan, 24 July 2014

http://www.energy.gov/downloads/doe-public-access-plan

ENERGY.GOV

Search Energy.gov

PUBLIC SERVICES | SCIENCE & INNOVATION | ENERGY SAVER | ABOUT ENERGY.GOV | OFFICES

Home » DOE Public Access Plan

DOE PUBLIC ACCESS PLAN

On February 22, 2013, the Director of the White House Office of Science and Technology Policy (OSTP), Dr. John Holdren, issued a memorandum to all agency and department heads entitled, "**Increasing Access to the Results of Federally Funded Scientific Research**." The memo directed federal agencies with more than $100 million in annual conduct of research and development to develop plans for increasing public access to peer-reviewed scientific publications and digital data resulting from federally funded research investments.

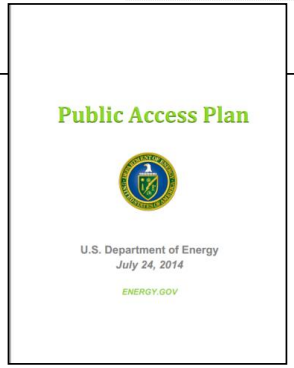
In response to the OSTP memo, DOE has completed a **Public Access Plan**, which commits the Department to deploying a web-based portal, the DOE Public Access Gateway for Energy and Science (**DOE PAGES**), that will make scholarly scientific publications resulting from DOE research funding publicly accessible and searchable at no charge to readers; and to instituting data management principles and requirements that ultimately will apply to proposals for research funding submitted to all DOE program offices.

MORE DOCUMENTS & PUBLICATIONS

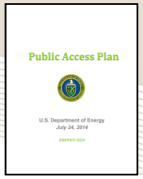
- **Attachment FY2011-58 OPAM**
- **Federal Assistance Reporting Checklist and Instructions for Projects**
- **Open Government Technology Summit**

Department of Energy Public Access Plan

http://www.energy.gov/sites/prod/files/2014/08/f18/DOE_Public_Access%20Plan_FINAL.pdf



DOE Public Access Plan: Overview of Contents



Public Access to Scientific Publications

With regard to unclassified and otherwise unrestricted research in scientific publications, the Department proposes a new policy and tool for providing access to peer-reviewed scholarly publications and associated metadata in which publishers retain their rights under copyright to the Version of Record (VoR). Both the policy and tool will be applied to scholarly publications resulting from unclassified and otherwise unrestricted research supported by the Department.

Public Access to Scientific Data in Digital Formats

With regard to unclassified and otherwise unrestricted scientific data in digital formats, the Department proposes a set of principles and requirements to be adopted by all DOE offices supporting open research. Implementing strategies and timelines may differ across the Department depending on the specific communities supported and funding mechanisms used by each office.

DOE Public Access Plan: Pointer to Office of Science as Lead with respect to Public Access to Scientific Data

Page 9 of DOE Public Access Plan

Public Access to Scientific Data in Digital Formats

The Department affirms that the following principles for the management of digital research data support its mission and align with the objectives of the OSTP memo.

- Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.
- Sharing and preserving data are central to protecting the integrity of science by facilitating validation of results and to advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. To the greatest extent, with the fewest constraints possible, and consistent with the requirements and other principles stated in this document, data sharing should make digital research data available to and useful for the scientific community, industry, and the public.
- Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning.

The Department is taking a phased approach to the implementation of requirements set forth by the OSTP memo. In particular, **the Office of Science, which supports roughly two-thirds of the total R&D for the Department, plans to pilot a data management policy with the requirements described below by July 28, 2014**. Other DOE Offices and elements with over $100 million in annual conduct of research and development expenditures will implement data management plan requirements that satisfy the requirements of the OSTP memo no later than October 1, 2015 in such a way that there is a single DOE policy for data management planning.

The result will be a Department-wide policy. Should it be necessary, additional supplementary guidance and requirements addressing specific needs would be issued by each Office or element and coordinated centrally.

IDA | DOE Press Release on Public Access Efforts, 4 August 2014

ENERGY.GOV

Search Energy.gov

PUBLIC SERVICES | SCIENCE & INNOVATION | ENERGY SAVER | ABOUT ENERGY.GOV | OFFICES

Home » U.S. Department of Energy Increases Access to Results of DOE-funded Scientific Research

U.S. Department of Energy Increases Access to Results of DOE-funded Scientific Research

August 4, 2014 - 10:49am

Tweet +1

NEWS MEDIA CONTACT

- 202-586-4940

WASHINGTON, D.C. – The U.S. Department of Energy is introducing new measures to increase access to scholarly publications and digital data resulting from Department-funded research.

The Energy Department has launched the Public Access Gateway for Energy and Science – **PAGES** – a web-based portal that will provide free public access to accepted peer-reviewed manuscripts or published scientific journal articles within 12 months of publication.

RELATED ARTICLES

Readily Available Data Help to Overcome Geothermal Deployment Barriers

Secretary Abraham Announces

See next slide for full text of press release

IDA | DOE Press Release on Public Access Efforts (full text)

WASHINGTON, D.C. – The U.S. Department of Energy is introducing new measures to increase access to scholarly publications and digital data resulting from Department-funded research.

Regarding Scholarly Publications

The Energy Department has launched the **Public Access Gateway for Energy and Science –** PAGES – a web-based portal that will provide free public access to accepted peer-reviewed manuscripts or published scientific journal articles within 12 months of publication.

"Increasing access to the results of research funded by the Department of Energy will enable researchers and entrepreneurs to capitalize on our substantial research and development investments," said Secretary of Energy Ernest Moniz. "These new policies set the stage for increased innovation, commercial opportunities, and accelerated scientific breakthroughs."

As it grows in content, PAGES will include access to DOE-funded authors' accepted manuscripts hosted primarily by the Energy Department's National Labs and grantee institutions, in addition to the public access offerings of publishers. For publisher-hosted content, the Department is collaborating with the publisher consortium CHORUS -- the Clearinghouse for the Open Research of the United States.

PAGES contains an initial collection of accepted manuscripts and journal articles as a demonstration of its functionality and eventual expanded content. Additional metadata and links to articles and accepted manuscripts will be added as they are submitted, with anticipated growth of 20,000 to 30,000 articles and manuscripts annually.

Regarding Digital Data

The Energy Department's Office of Science also has issued new requirements regarding management of digital research data by Office of Science-supported researchers. All proposals for research funding submitted to the Office of Science will be required to include a Data Management Plan that describes whether and how the digital research data generated in the course of the proposed research will be shared and preserved.

The new requirements regarding management of digital research data will appear in funding solicitations and invitations issued by the Office of Science beginning Oct. 1, 2014. A statement of the new requirements, including **guidance on the development of a Data Management Plan**, can be found on the Office of Science website. Other Energy Department research offices will implement data management plan requirements within the next year.

IDA

DOE Office of Scientific and Technical Information: Capabilities in Support of Public Access

- *For Scientific Publications:*
  Public Access for Energy and Science (PAGES)
- *For Scientific Data:*
  DOE/OSTI Data ID Service
  DOE Data Explorer

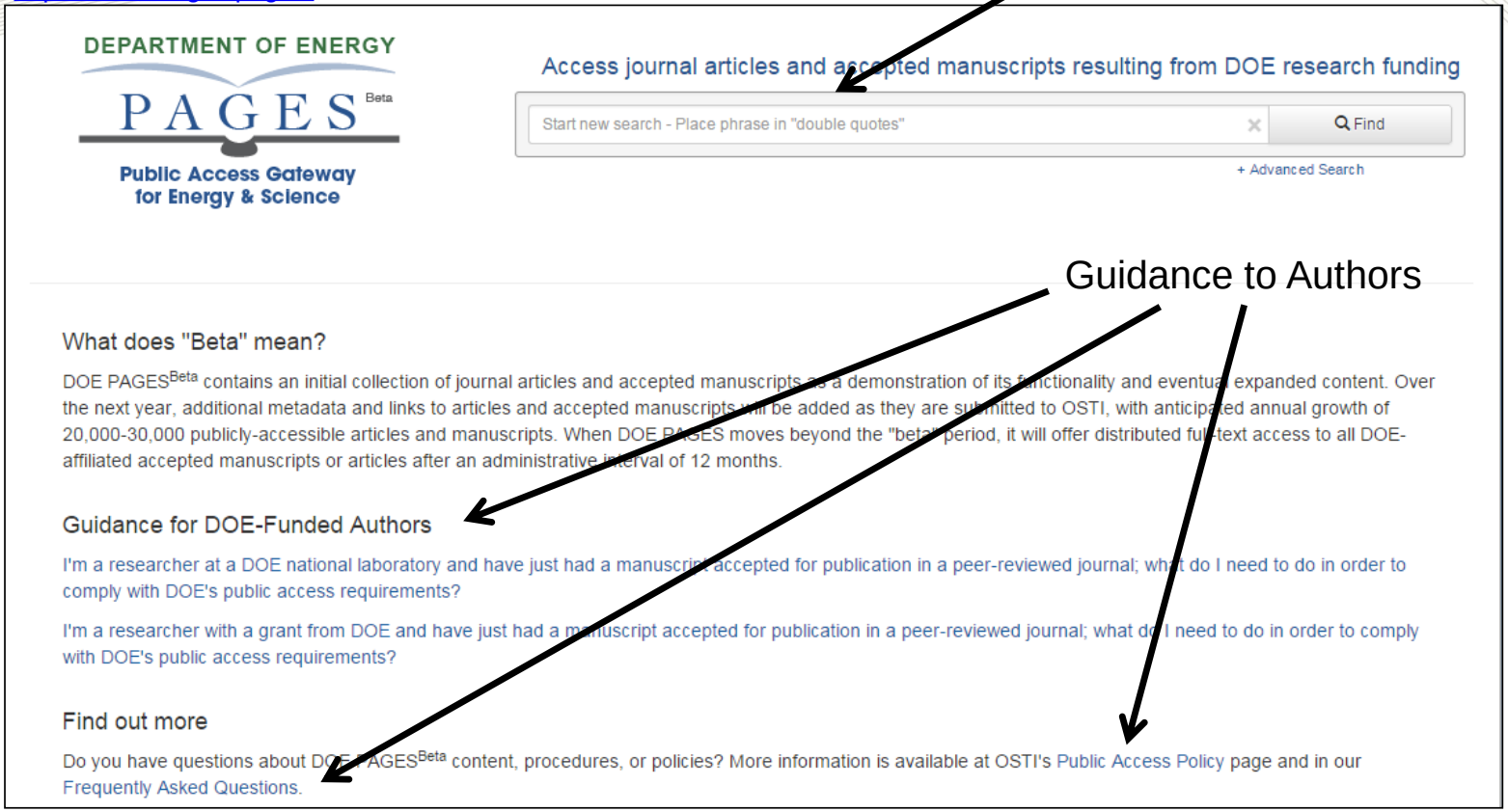
DOE Public Access Gateway for Science and Science (PAGES)

- DOE Public Access Policy, developed in response to OSTP memo, applies to "**final, peer-reviewed and accepted manuscripts** or, for participating publishers, the corresponding **published journal article**"
- Public Access Gateway for Energy and Science (PAGES) Portal
  - Operated and maintained by Office of Scientific and Technical Information (OSTI)
  - Holds metadata and abstracts that are searchable and available for bulk download
  - Before end of 12-month embargo period, PAGES links to Version of Record (VoR) at publisher's site (typically behind pay wall)
  - After embargo period, PAGES links to "best available version"
    - VoR at publisher's site, if access is open and free
    - Otherwise, accepted manuscript (typically hosted at institutional repository, i.e., national lab or grantee institution; otherwise hosted at OSTI)

1. **Agency portal: PAGES, with centralized metadata and links to decentralized full-text articles**
2. **Dark archive: OSTI repository of accepted manuscripts, accessed only when no other version of a requested article is publicly available**
3. **Approach to public access: "Best available version" delivered from publisher's site (1st choice), institutional repository (2nd choice), or OSTI repository (3rd choice)**
4. **Embargo period: Up to 12 months**

IDA | DOE PAGES Portal

http://www.osti.gov/pages/



DOE Guidance for DOE-Funded Authors, Part 1 – Financial Assistance Awardees* (e.g., at Universities)

- **DOE Guidance** (http://www.osti.gov/pages/):
  "Regarding copyright transfer, for Financial Assistance Awardees, the **Government retains nonexclusive and irrevocable rights to use the works published under an award for federal purposes** (2 CFR § 200.315(b) (d)). As per the DOE Terms and Conditions for Grants and Cooperative Agreements, DOE requires an acknowledgement of awarding agency support to be marked on the publication of any material, whether copyrighted or not. **OSTI continues to work closely with procurement points of contact regarding additional guidance related to DOE's Public Access Plan**."

- **2 CFR § 200.315(b) and (d)**:
  (b) **The non-Federal entity may copyright any work that is subject to copyright and was developed, or for which ownership was acquired, under a Federal award**. The Federal awarding agency reserves a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so.
  (d) The Federal Government has the right to:
  (1) Obtain, reproduce, publish, or otherwise use the data produced under a Federal award; and
  (2) Authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes.

* *Financial Assistance Awardees are researchers funded via grants or cooperative agreements, as opposed to researchers funded via contracts.*

DOE Guidance for DOE-Funded Authors, Part 2 – Contractors (e.g., at DOE National Labs)

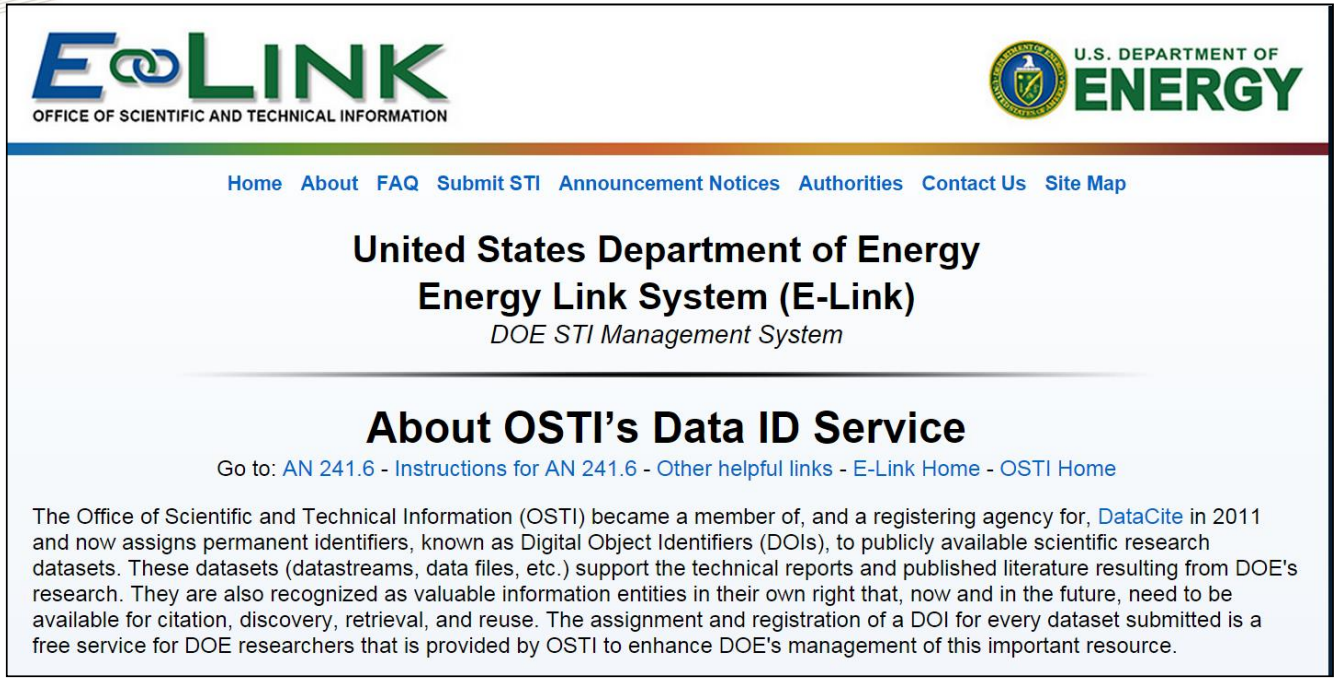
- **DOE Guidance** (http://www.osti.gov/pages/):

  Regarding copyright transfer, … a key point to note [is] that regardless of the specific wording [of any publisher's publication agreement], the Government retains rights to the article. In fact, **within the prime contract for DOE Laboratories** is wording which states: "...(2) **The contractor shall mark each scientific or technical article first produced or composed under this Contract** and submitted for journal publication or similar means of dissemination with a notice, similar in all material respects to the following, on the front **reflecting the Government's non-exclusive, paid-up, irrevocable, world-wide license in the copyright**.

  > *Notice: This manuscript has been authored by [insert the name of the Contractor] under Contract No. [insert the contract number] with the U.S. Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes.*
  >
  > (End of Notice) [Ref. **DEAR970.5227-2 Rights in data-technology transfer**]

IDA | DOE/OSTI Data ID Service

https://www.osti.gov/elink/aboutDataIDService.jsp



E-LINK
OFFICE OF SCIENTIFIC AND TECHNICAL INFORMATION

U.S. DEPARTMENT OF ENERGY

Home About FAQ Submit STI Announcement Notices Authorities Contact Us Site Map

United States Department of Energy
Energy Link System (E-Link)
DOE STI Management System

About OSTI's Data ID Service

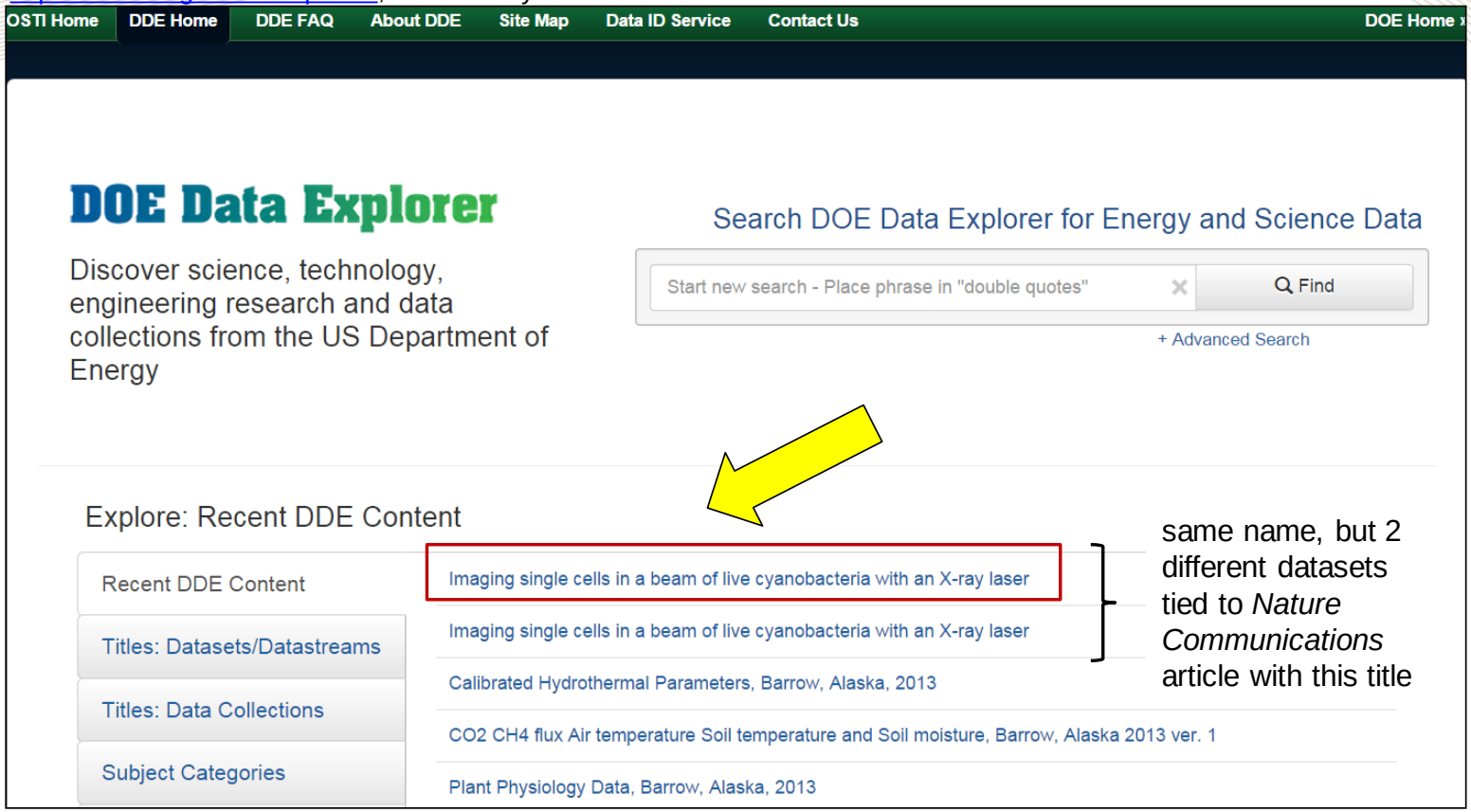
Go to: AN 241.6 - Instructions for AN 241.6 - Other helpful links - E-Link Home - OSTI Home

The Office of Scientific and Technical Information (OSTI) became a member of, and a registering agency for, DataCite in 2011 and now assigns permanent identifiers, known as Digital Object Identifiers (DOIs), to publicly available scientific research datasets. These datasets (datastreams, data files, etc.) support the technical reports and published literature resulting from DOE's research. They are also recognized as valuable information entities in their own right that, now and in the future, need to be available for citation, discovery, retrieval, and reuse. The assignment and registration of a DOI for every dataset submitted is a free service for DOE researchers that is provided by OSTI to enhance DOE's management of this important resource.

DOE OSTI, as registering agency for DataCite, can assign Digital Object Identifiers (DOIs) to Datasets

IDA | Quick Look at DOE Data Explorer (DDE): Recent DDE Content

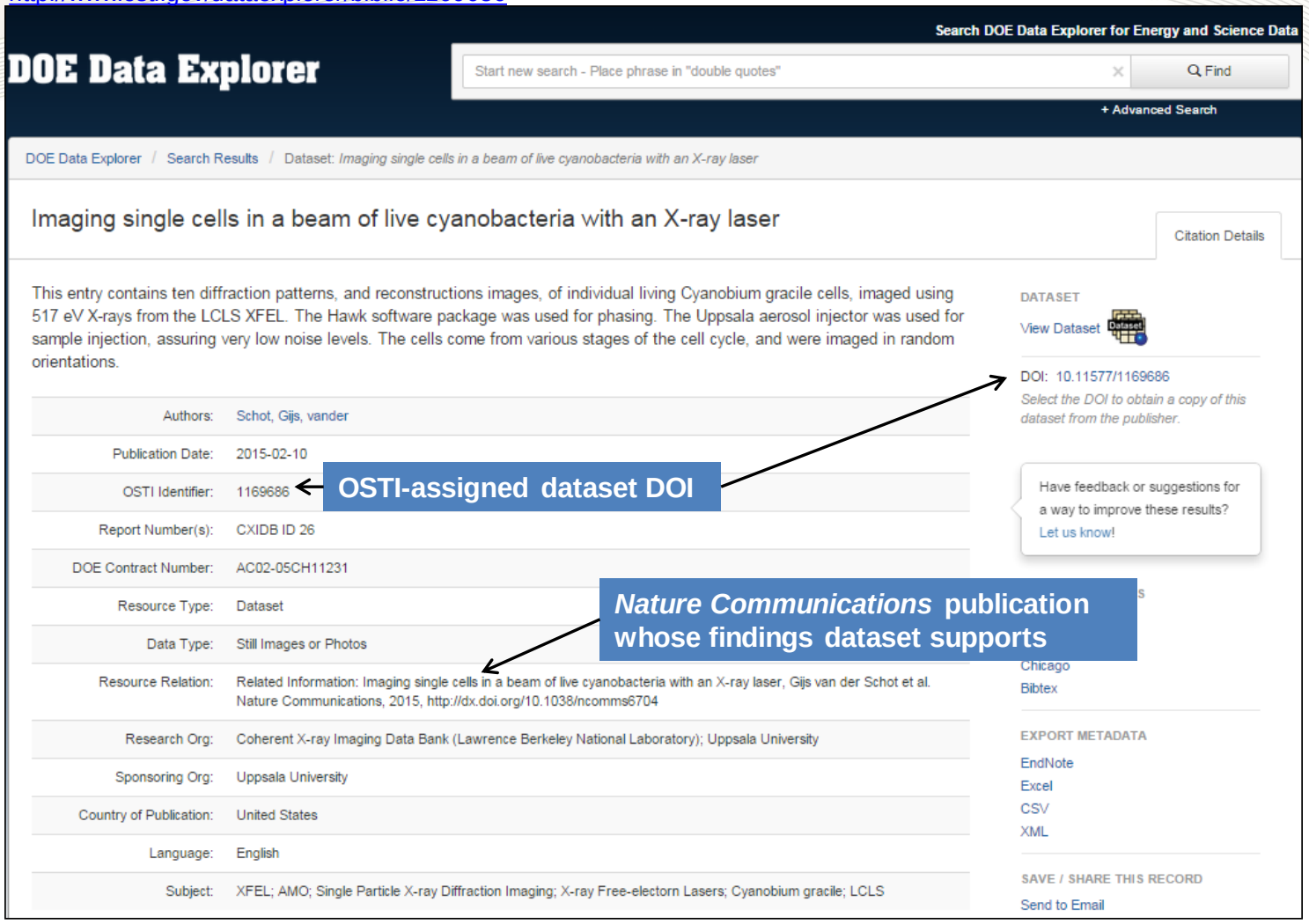
http://www.osti.gov/dataexplorer/, 17 February 2015



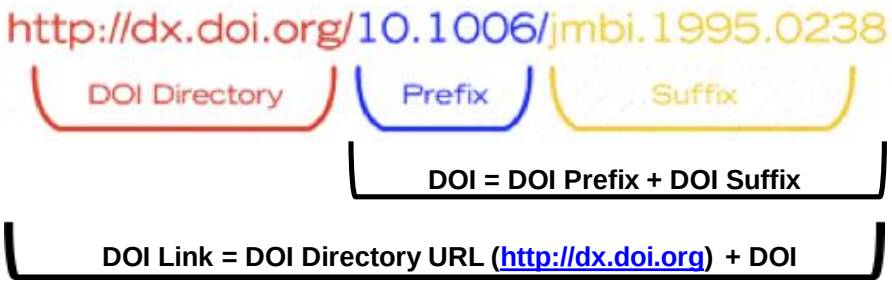
same name, but 2 different datasets tied to *Nature Communications* article with this title

IDA | DOE Data Explorer Metadata on Dataset

http://www.osti.gov/dataexplorer/biblio/1169686



Closer Look at OSTI Identifier for van der Schot Dataset



OSTI Identifier (i.e., DOI Suffix) for van der Schot dataset is 1169686

meaning **Digital Object Identifier (DOI) is 10.11577/1169686**

meaning **DOI Link is dx.doi.org/10.11577/1169686**

which binds to dataset landing page at URL **http://cxidb.org/id-26.html in Coherent X-ray Imaging Data Bank (CXIDB)**
(i.e., clicking on DOI link takes person to landing page)

IDA

DOE Office of Science Data Management Policy

DOE Office of Science (SC) Funding Opportunities Tab on Homepage

↓

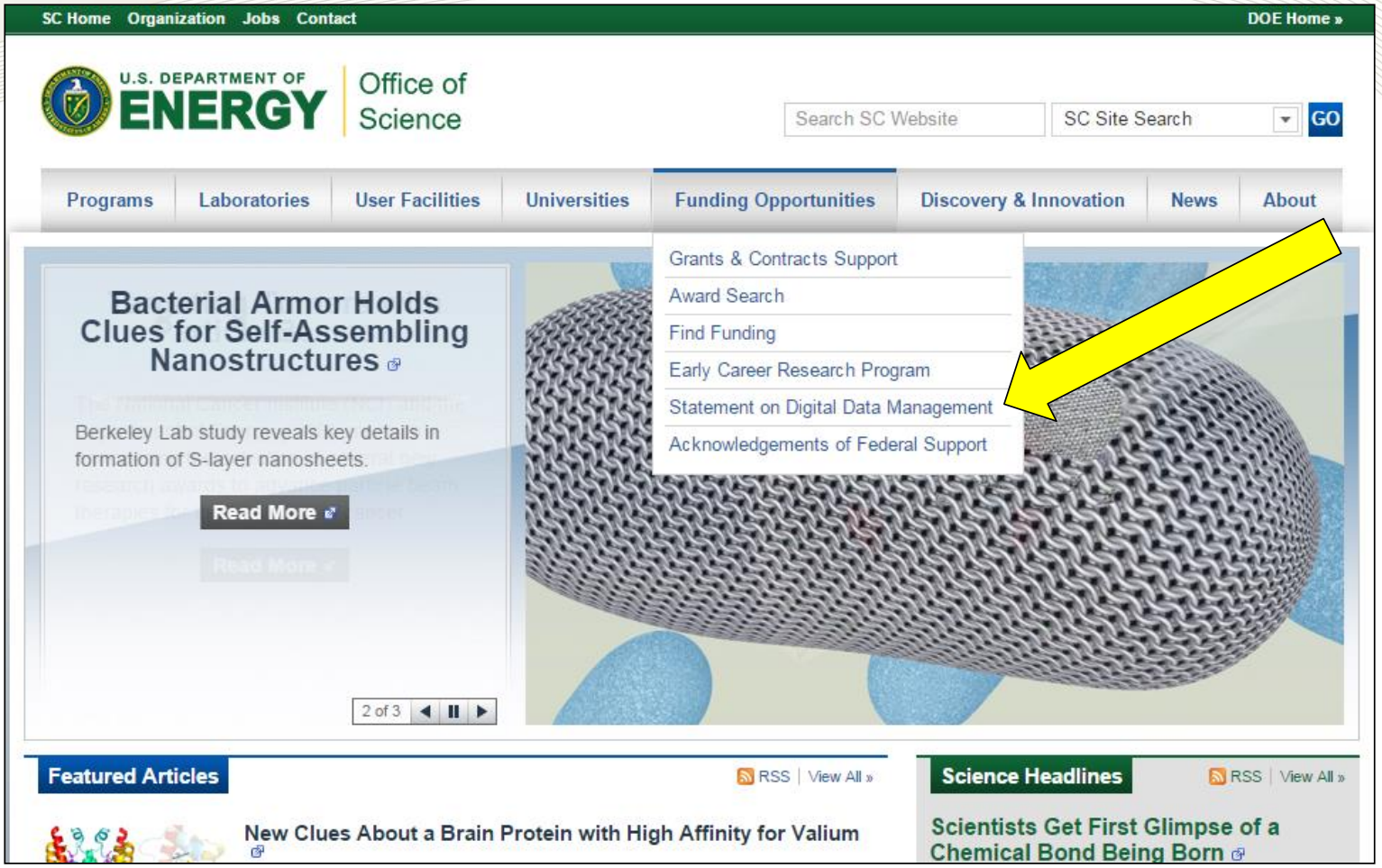
DOE/SC Statement on Digital Data Management Webpage
(*Text on Webpage provides Principles, Requirements, and Additional Guidance)*

→ Suggested Elements of Data Management Plan (DMP)

→ Frequently Asked Questions (FAQ)

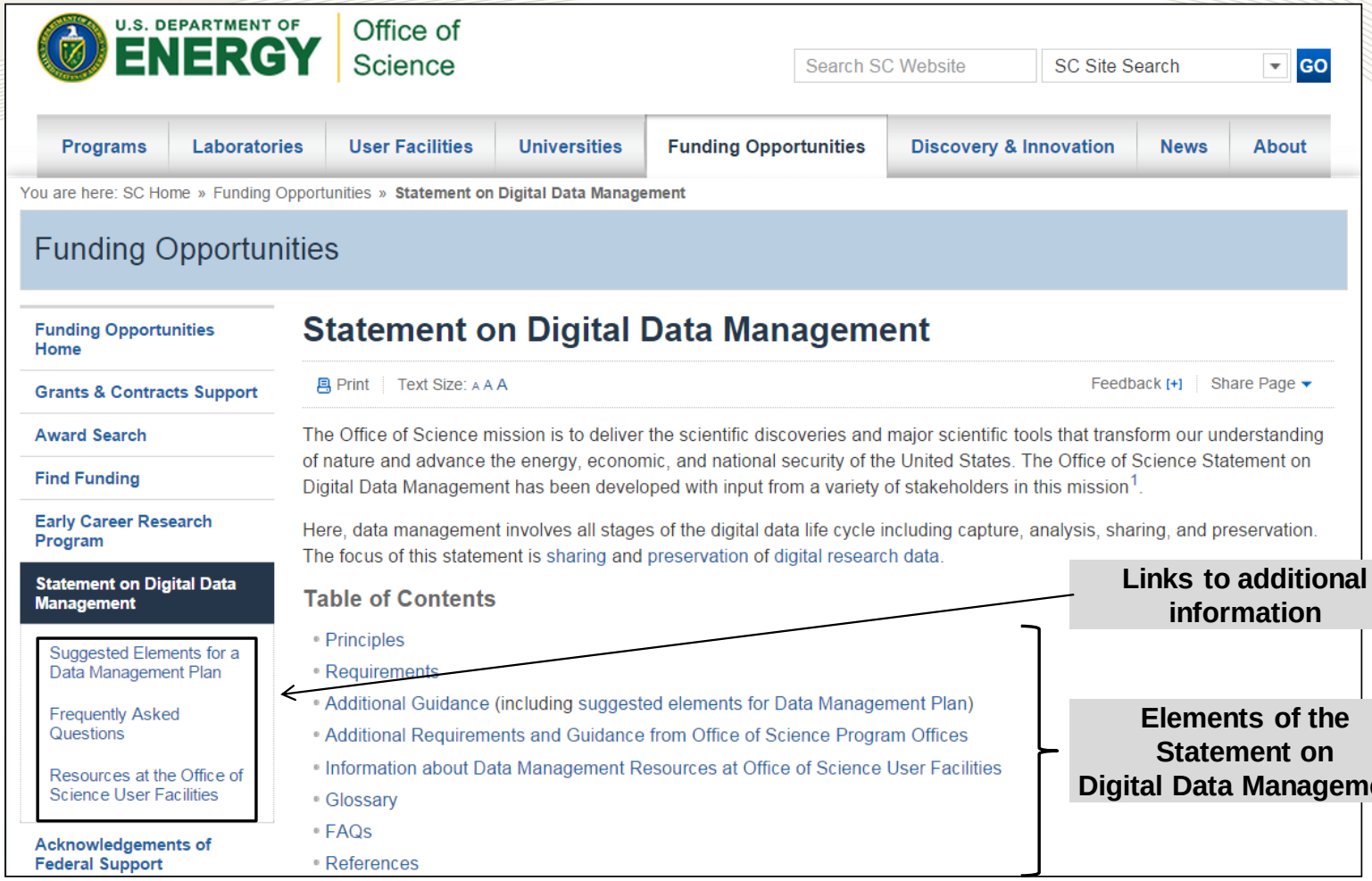
IDA | DOE Office of Science (SC) – Funding Opportunities

http://science.energy.gov/



IDA | DOE/SC – Statement on Digital Data Management (DDM)

http://science.energy.gov/funding-opportunities/digital-data-management/



IDA | DOE/SC – Statement on DDM – Requirements

http://science.energy.gov/funding-opportunities/digital-data-management/

1.DMPs should describe whether and how data generated in the course of the proposed research will be shared and preserved. If the plan is not to share and/or preserve certain data, then the plan must explain the basis of the decision (for example, cost/benefit considerations, other parameters of feasibility, scientific appropriateness, or limitations discussed in #4). **At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved**.

2.DMPs should provide a plan for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. This includes data that are displayed in charts, figures, images, etc. **In addition, the underlying digital research data used to generate the displayed data should be made as accessible as possible** to the public in accordance with the principles stated above. This requirement could be met by including the data as **supplementary information to the published article**, or through other means. The published article should indicate how these data can be accessed.

3.DMPs should consult and reference available information about data management resources to be used in the course of the proposed research….

4.DMPs must protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all applicable laws, regulations, and DOE orders and policies. There is no requirement to share proprietary data….

DMPs will be reviewed as part of the overall Office of Science research proposal merit review process. Additional requirements and review criteria for the DMP may be identified by the sponsoring program or sub-program, or in the solicitation.

IDA | DOE/SC – Statement on DDM – Additional Guidance

http://science.energy.gov/funding-opportunities/digital-data-management/

Additional Guidance

- The Principal Investigator should determine which data should be the subject of the DMP and, in the DMP, **propose which data should be shared and/or preserved** in accordance with the Requirements.
- In determining which data should be shared and preserved, researchers must consider the data needed to **validate research findings** as described in the Requirements, and are encouraged to consider the potential benefits of their data to their own fields of research, fields other than their own, and society at large.
- DMPs should reflect **relevant standards** and **community best practices** for data and metadata, and make use of **community accepted repositories** whenever practicable.
- **Costs associated with the scope of work and resources articulated in a DMP may be included in the proposed research budget as permitted by the applicable cost principles.**
- To improve the discoverability of and attribution for datasets created and used in the course of research, the Office of Science encourages the citation of publicly available datasets within the reference section of publications, and **the identification of datasets with persistent identifiers such as Digital Object Identifiers (DOIs)**. In most cases, the Office of Science can provide DOIs free of charge for data resulting from DOE-funded research through its Office of Scientific and Technical Information **(OSTI) DataID Service**.
- View a list of suggested elements for a DMP.

IDA | DOE/SC – Statement on DDM – Suggested Elements for a DMP

http://science.energy.gov/funding-opportunities/digital-data-management/

Data Types and Sources. A brief, high-level description of the data to be generated or used through the course of the proposed research and **which of these are considered digital research data necessary to validate the research findings**.

Content and Format. A statement of plans for **data and metadata content and format** including, where applicable, a description of documentation plans, annotation of relevant software, and the rationale for the selection of appropriate standards….

Sharing and Preservation. A description of the plans for data sharing and preservation. This should include, when appropriate:

- the anticipated means for sharing and the rationale for any restrictions …;
- a timeline for sharing and preservation that addresses …;
- any special requirements for data sharing, for example, proprietary software …;
- any resources and capabilities (equipment, connections, systems, software, expertise, etc.) …;
- cost/benefit considerations …;
- whether, when, or under what conditions the management responsibility for the research data will be transferred to a third party (e.g. institutional, or community repository);
- any other future decision points regarding the management of the research data….

Protection. A statement of plans, where appropriate and necessary, to protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; and avoid significant negative impact on innovation, and U.S. competitiveness.

Rationale. A discussion of the rationale or justification for the proposed data management plan including, for example, the potential impact of the data within the immediate field and in other fields, and any broader societal impact.

IDA | DOE/SC – Statement on DDM - FAQ

http://science.energy.gov/funding-opportunities/digital-data-management/faqs/

Do I need to submit a Data Management Plan (DMP)? (15 question, 1 shown below)

1. Will a proposal be funded if it does not include a Data Management Plan?
   No. SC reserves the right to reject, without merit review, any proposal that does not include a DMP.

What to include in a DMP (4 questions)

16. The Office of Science Statement on Digital Data Management requires that I submit a Data Management Plan (DMP) with my research proposal. What should I include in this plan?
    A list of suggested elements for a DMP can be found here.

Sharing and Preservation (9 questions)

20. The data or data products from my research will likely be cited by me and/or others. What should I to do ensure that these are cited appropriately and that I receive proper attribution for their use?
    There are no global standards for how to cite data products. Suggestions for what information to include in a citation for your data product and how to format this information can be found here. To facilitate the citation of data products, the Office of Science encourages the use of persistent identifiers such as Digital Object Identifiers (DOIs)…. DOE/SC can provide DOIs free of charge for datasets resulting from DOE-funded research through its Office of Scientific and Technical Information (OSTI) DataID Service.

Evaluation (2 questions)

29. Should my proposed budget specifically address the resources and costs in implementing my DMP?
    Costs associated with the scope of work and resources articulated in a DMP may be included in the proposed research budget as permitted by the appropriate cost principles.
30. What consequences, if any, result from failure to carry out the data management plan of a funded research proposal?
    The DMP is part of the overall research proposal and, as such, it is expected that researchers will follow, to the best of their ability, the proposed research and associated data management plan. **Failure to do so will negatively influence future funding opportunities.**

IDA

NIH Experience on Public Access to Scientific Publications

- NIH Public Access Policy on Publications
- Copyright Issue and Solution
- Sample Guidance to University Researchers
- Public Access Policy Agreement between Researcher and University at Funding Acceptance
- Author Notice to Publisher at Article Submission
- Author Addendum to Publication Agreement

IDA | NIH Public Access Policy on Publications

- Established in response to Consolidated Appropriations Act of 2008 and applies to peer-reviewed journal articles accepted for publication after 7 April 2008 [NOT-OD-08-033]
- Implements Division F Section 217 of PL 111 (Omnibus Appropriations Act of 2009), which states:

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

- Is consistent with Division H Section 527 of PL 113-76 (Consolidated Appropriations Act of 2014), which extends public access requirements beyond NIH to additional Federal agencies:

  > Sec. 527. Each Federal agency, or in the case of an agency with multiple bureaus, each bureau (or operating division) funded under this Act that has research and development expenditures in excess of $100,000,000 per year shall develop a Federal research public access policy that provides for--
  >
  > (1) the submission to the agency, agency bureau, or designated entity acting on behalf of the agency, a machine-readable version of the author's final peer-reviewed manuscripts that have been accepted for publication in peer-reviewed journals describing research supported, in whole or in part, from funding by the Federal Government;
  >
  > (2) free online public access to such final peer-reviewed manuscripts or published versions **not later than 12 months** after the official date of publication; and
  >
  > (3) compliance with all relevant copyright

IDA | Copyright Issue and Solution

- Issue: Historically, publishers have required authors to sign publication agreements that transfer entire copyright to publisher or grant publisher exclusive rights incompatible with legislative mandate and NIH policy
- Solution: Careful management of rights under copyright (NIH chose not to rely on Government Purpose Rights – see upcoming slide on NIH Risk Mitigation)
  - **Step 1: Author grants license to institutional awardee at award acceptance**
    - Institutional awardee (e.g., university) requires principal investigator to sign an agreement granting the institution non-exclusive rights sufficient to enable institution to comply with NIH Policy
    - **Protects institutional awardee and ensures NIH rights**. License stands *regardless of authors' actions with respect to any future publication agreements with publisher* (author cannot transfer rights he has already granted to another party)
  - **Step 2: Author gives notice to publisher at article submission**
    - Author informs publisher of rights that will be retained by author(s), university, and NIH
    - **Preempts over-reaching copyright transfer provisions** of any subsequent publication agreement between author and publisher
  - **Step 3: Author addendum to publication agreement at article acceptance**
    - Author reiterates rights retained by author(s), university, and NIH
    - **Protects author** against publisher accusations of misrepresentation in the case that the publication agreement calls for transfer of all rights to publisher (the author cannot transfer all rights if he has already granted non-exclusive, irrevocable rights to his institution)

IDA | Sample Guidance to University Researchers

HARVARD UNIVERSITY GUIDELINES NIH Public Access Policy

1. All Harvard paid investigators working under NIH grants will be required to sign and return an "NIH Public Access Policy Agreement." Under this agreement, **investigators who author articles subject to the NIH Policy grant Harvard a limited, nonexclusive license to use their articles to comply with the policy**, and to authorize the NIH to use their articles and make them publicly available in accordance with the NIH Policy. This license will help ensure that sufficient rights are reserved to enable compliance with the NIH Policy. The agreement can be found at https://www.countway.harvard.edu/publicaccess/forms and should be submitted to one of the three sponsored programs contacts listed at the bottom of this page.
2. When articles subject to the NIH Policy are submitted to peer-reviewed journals for publication, the submission letter should include an "Article Submission Letter Attachment." The Attachment serves as a **notice to the publisher that the articles are subject to the NIH Policy and that the author or authors and Harvard will have sufficient nonexclusive rights** in the article to enable full compliance with the NIH Policy. The Article Submission Letter Attachment is available at: https://www.countway.harvard.edu/publicaccess/forms.
3. Authors entering into copyright or publishing agreements with publishers must reserve sufficient rights to enable full compliance with the NIH Policy. Some publishing agreements require that authors transfer the entire copyright in their articles to the publisher, or grant exclusive rights that are incompatible with the NIH Policy. **Authors must reserve sufficient rights to avoid breach of the publishing agreement** and ensure compliance with the NIH Policy.
4. All NIH-funded final, peer-reviewed manuscripts, including all graphics and supplemental materials associated with the article, must be submitted electronically to PubMed Central upon acceptance for publication. Information about the submission requirements and submission instructions are available at the Countway Library's HMscholar website http://hmscholar.countway.harvard.edu.

https://legacy.countway.harvard.edu/menuNavigation/libraryServices/nihPublicAccess.html

IDA | Step 1 (at Award Acceptance): Institutional Awardee License

Agreement Regarding NIH Public Access Policy

The National Institutes of Health Public Access Policy (the "NIH Policy") applies to peer-reviewed articles that arise, in whole or in part, from direct costs funded by NIH, or from NIH staff, and that are accepted for publication on or after April 7, 2008.[1]

The NIH Policy requires that Articles be submitted to the National Library of Medicine's PubMed Central and then be made publicly accessible through PubMed Central no later than 12 months after publication. The NIH Policy also requires that any publishing or copyright agreements concerning submitted Articles fully comply with the NIH Policy. Both the President and Fellows of Harvard College ("Harvard") and I wish to comply fully with the NIH Policy.

To ensure compliance with the NIH Policy, and in consideration of my employment by Harvard or opportunities made available to me to perform sponsored research or use funds or facilities administered by Harvard, I agree as follows with respect to all Articles of which I am the author or a co-author:

1. Harvard will have, and I hereby grant to Harvard, the nonexclusive, irrevocable right to use the Articles in order to comply with the NIH Policy, and to authorize the NIH to use the Articles and make them publicly available in accordance with the NIH Policy.

2. Where I am the corresponding author for any Article, I will reserve rights at least as broad as those set forth in paragraph 1 above when granting rights in the Article to the publisher in the publishing or copyright agreement, so that the agreement will comply with the NIH Policy.

3. I will comply with the procedures Harvard specifies for dealing with copyright and for submission of the Articles to PubMed Central, to the extent those procedures apply to me, so that the requirements of the NIH Policy will be met.

Signed: _____

Name (Printed): _____

Date: _____

_____

[1] The NIH has further clarified that the NIH Policy applies to peer-reviewed articles based on work in one or more of the following categories: (i) directly funded by an NIH grant or cooperative agreement active in Fiscal Year 2008 (October 1, 2007- September 30, 2008) or beyond; (ii) directly funded by a contract signed on or after April 7, 2008; (iii) directly funded by the NIH Intramural Program; or (iv) if NIH pays an investigator's salary.

https://legacy.countway.harvard.edu/menuNavigation/libraryServices/nihPublicAccess/forms/AgreementNIHPolicy3_31_08.pdf

IDA | Step 2 (at Article Submission): Notice to Publisher

Dear [Publisher or Editor name],

Enclosed is a manuscript to be considered for publication in _____ [Journal name]. The research reported in this manuscript has been funded through the National Institutes of Health and therefore its publication must comply with the NIH Public Access Policy (http://grants.nih.gov/grants/guide/notice-files/NOT-OD-08-033.html).

In order to ensure compliance with the NIH policy I, as corresponding author on behalf of all the authors, am retaining the rights to:

- Provide a copy of the authors' final manuscript, including all modifications from the publishing and peer review process, to the NLM's PubMed Central (PMC) database at the time the manuscript is accepted for publication; and
- To authorize NIH to make such copy of the manuscript available in digital form for public access in PMC no later than ___ months (indicate 0 to 12 months) after publication.

[Universities may insert any additional terms pertaining to author and institutional rights for reproduction, distribution for academic activities, deposition in institutional archive, etc. Suggested language for this option is:

- To prepare derivative works from the manuscript;
- To authorize others to make any use of the manuscript provided that it is not sold for a profit and that the author receives credit as author and the journal in which the manuscript has been published is cited as the source of first publication; and
- To distribute copies of the manuscript in connection with teaching and research by the author and by the author's employer.]

By accepting this manuscript for review, [publisher name] accepts these terms and agrees that the terms of this agreement are paramount and supersede any provisions in any publication agreement for this article, already signed or to be signed at a later date, that may conflict.

_____
(Signature of corresponding author on behalf of all authors)

Carroll, Michael, Joint SPARC/SC/ARL White Paper, 2008, http://www.sparc.arl.org/resources/papers-guides/nih-copyright

IDA | Step 3 (at Publication Agreement): Author Addendum

ADDENDUM TO PUBLICATION AGREEMENT

1. THIS ADDENDUM hereby modifies and supplements the attached Publication Agreement concerning the following Article:
manuscript title and *journal name*
2. The parties to the Publication Agreement as modified and supplemented by this Addendum are:
Author(s) (if more than one author, collectively, Author) and *Publisher*
3. This Addendum and the Publication Agreement, taken together, allocate all rights under copyright with respect to all versions of the Article. The parties agree that wherever there is any conflict between this Addendum and the Publication Agreement, the provisions of this Addendum are paramount and the Publication Agreement shall be construed accordingly.
4. **Author's Retention of Rights.** Notwithstanding any terms in the Publication Agreement to the contrary, AUTHOR and PUBLISHER agree that in addition to any rights under copyright retained by Author in the Publication Agreement, Author retains: (i) the rights to reproduce, to distribute, to publicly perform, and to publicly display the Article in any medium for non-commercial purposes; (ii) the right to prepare derivative works from the Article; and (iii) the right to authorize others to make any non-commercial use of the Article so long as Author receives credit as author and the journal in which the Article has been published is cited as the source of first publication of the Article. For example, Author may make and distribute copies in the course of teaching and research and may post the Article on personal or institutional Web sites and in other open-access digital repositories.
5. **Publisher's Additional Commitments.** Publisher agrees to provide to Author within 14 days of first publication and at no charge an electronic copy of the published Article in a format, such as the Portable Document Format (.pdf), that preserves final page layout, formatting, and content. No technical restriction, such as security settings, will be imposed to prevent copying or printing of the document.
6. **Acknowledgment of Prior License Grants.** In addition, where applicable and without limiting the retention of rights above, Publisher acknowledges that Author's assignment of copyright or Author's grant of exclusive rights in the Publication Agreement is subject to Author's prior grant of a non-exclusive copyright license to Author's employing institution and/or to a funding entity that financially supported the research reflected in the Article as part of an agreement between Author or Author's employing institution and such funding entity, such as an agency of the United States government.
7. For record keeping purposes, Author requests that Publisher sign a copy of this Addendum and return it to Author. However, if Publisher publishes the Article in the journal or in any other form without signing a copy of this Addendum, such publication manifests Publisher's assent to the terms of this Addendum.

_____ _____

AUTHOR (corresponding author on behalf of all authors) Date PUBLISHER Date

http://www.sparc.arl.org/sites/default/files/Access-Reuse_Addendum.pdf

NIH Risk Mitigation: Avoid Relying Solely on Government Purpose License

HHS Government Purpose Copyright License [45 C.F.R. § 74.36 (2007)]:
"The recipient may copyright any work that is subject to copyright and was developed, or for which ownership was purchased, under an award. The HHS awarding agency reserves a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work **for Federal purposes**, and to authorize others to do so."

NIH Assertion on Scope of Government Purpose License [NOT-OD-05-022, 3 February 2005]:
"Although the NIH, at this time, is not relying on the government purpose license, **it is an available means for NIH to reproduce, publish or otherwise use copyrighted works resulting from NIH funding for Federal purposes, as well as to authorize others to do so**. Arguments put forth and cases cited by the commenter as support for the premise that the government purpose license could not be used as a basis for PMC to post the manuscripts are not persuasive. None of the cases address circumstances where a government agency is acting to fulfill its own statutory purposes with regard to publications resulting from its own research funding. **Creation of a publicly accessible, permanent archive of NIH-funded research publications is squarely within the statutory authorities of the NIH and the NLM and clearly constitutes a Federal purpose**."

NIH Decision to Mitigate Risk by Acquiring Explicit Permission from Author rather than by Relying on Government Purpose License [Carroll 2008]:
"… NIH faced a non-trivial risk that it would have to litigate the issue had it chosen to rely on this license. Consequently, NIH chose as part of the February 3, 2005 version of the policy to require the **person submitting the manuscript to set the embargo period [as part of submittal process] and to specifically grant NIH permission** to make the manuscript publicly accessible after that period."

IDA | Backup: Key Federal Regulations on Rights in Data

IDA | 2 CFR Grants and Agreements (19 December 2014)

Title 2: Grants and Agreements
PART 200—UNIFORM ADMINISTRATIVE REQUIREMENTS, COST PRINCIPLES, AND AUDIT REQUIREMENTS FOR FEDERAL AWARDS
Subpart D—Post Federal Award Requirements

§200.315 Intangible property.

(a) …

(b) The non-Federal entity may copyright any work that is subject to copyright and was developed, or for which ownership was acquired, under a Federal award. **The Federal awarding agency reserves a royalty-free, nonexclusive and irrevocable right to reproduce, publish, or otherwise use the work for Federal purposes, and to authorize others to do so**.

(c) …

(d) The Federal Government has the right to:

(1) Obtain, reproduce, publish, or otherwise use the data produced under a Federal award; and

(2) Authorize others to receive, reproduce, publish, or otherwise use such data for Federal purposes.

OMB Guidance on Federal Purpose Copyright License

Applies to DOE Financial Assistance Awardees
(i.e., those funded via grants and cooperative agreements)

IDA | FAR 52.227-14 Rights in Data – General (May 2014) [FAR 1]

(b) Allocation of rights….

(2) The Contractor shall have the right to—

(i) Assert copyright in data first produced in the performance of this contract to the extent provided in paragraph (c)(1) of this clause; ….

(c) Copyright—

(1) Data first produced in the performance of this contract.

(i) Unless provided otherwise in paragraph (d) of this clause, **the Contractor may, without prior approval of the Contracting Officer, assert copyright in scientific and technical articles based on or containing data first produced in the performance of this contract and published in academic, technical or professional journals, symposia proceedings, or similar works. The prior, express written permission of the Contracting Officer is required to assert copyright in all other data first produced in the performance of this contract.**

(ii) When authorized to assert copyright to the data, **the Contractor shall affix the applicable copyright notices of 17 U.S.C. 401 or 402, and an acknowledgment of Government sponsorship (including contract number).**

(iii) **For data other than computer software,** the Contractor grants to the Government, and others acting on its behalf, a paid-up, nonexclusive, irrevocable, worldwide license in such copyrighted data to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly by or on behalf of the Government. **For computer software**, the Contractor grants to the Government, and others acting on its behalf, a paid-up, nonexclusive, irrevocable, worldwide license in such copyrighted computer software to reproduce, prepare derivative works, and perform publicly and display publicly (but not to distribute copies to the public) by or on behalf of the Government

IDA | FAR 52.227-14 Rights in Data – General (May 2014) [FAR 2]

Alternate IV (Dec 2007). As prescribed in 27.409(b)(5), substitute the following paragraph (c)(1) for paragraph (c)(1) of the basic clause:

(c) Copyright—

(1) Data first produced in the performance of the contract. Except as otherwise specifically provided in this contract, the Contractor may assert copyright in **any data** ***[vs. "scientific and technical articles … published in academic, technical or professional journals, symposia proceedings, or similar works" as in the original paragraph (c)(1) on preceding slide]*** first produced in the performance of this contract. **When asserting copyright**, the Contractor shall affix the applicable copyright notice of 17 U.S.C. 401 or 402, and acknowledgment of Government sponsorship (including contract number), to the data when such data are delivered to the Government, as well as when the data are published or deposited for registration as a published work in the U.S. Copyright Office. **For data other than computer software**, the Contractor grants to the Government, and others acting on its behalf, a paid-up, nonexclusive, irrevocable, worldwide license for all such data to reproduce, prepare derivative works, distribute copies to the public, and perform publicly and display publicly, by or on behalf of the Government. **For computer software**, the Contractor grants to the Government and others acting on its behalf, a paid up, nonexclusive, irrevocable worldwide license for all such computer software to reproduce, prepare derivative works, and perform publicly and display publicly (but not to distribute copies to the public), by or on behalf of the Government.

"**Data**" means recorded information, regardless of form or the media on which it may be recorded. The term **includes technical data and computer software**. The term does not include information incidental to contract administration, such as financial, administrative, cost or pricing, or management information.

"**Technical data**" means recorded information (regardless of the form or method of the recording) of a scientific or technical nature (including computer databases and computer software documentation). This term **does not include computer software** …. The term includes recorded information of a scientific or technical nature that is included in computer databases (See 41 U.S.C. 116)

IDA | FAR 27.404-3 Copyrighted Works (May 2014) [FAR 3]

(3) … Except for contracts for management or operation of Government facilities and contracts and subcontracts in support of programs being conducted at those facilities or where international agreements require otherwise, **Alternate IV shall be used in all contracts for basic or applied research to be performed solely by colleges and universities**. Alternate IV shall not be used in contracts with colleges and universities if a purpose of the contract is for development of computer software for distribution to the public (including use in solicitations) by or on behalf of the Government. In addition, **Alternate IV may be used in other contracts** if an agency determines that it is not necessary for a contractor to request further permission to assert copyright in data first produced in performance of the contract. The **contracting officer may exclude any data, or items or categories of data, from the provisions of Alternate IV** by expressly so providing in the contract or by adding a paragraph (d)(4) to the clause, consistent with 27.404-4(b).

(4) Pursuant to paragraph (c)(1) of the clause at 52.227-14, the **contractor grants the Government a paid-up nonexclusive, irrevocable, worldwide license** to reproduce, prepare derivative works, distribute to the public, perform publicly and display publicly by or on behalf of the Government, **for all data (other than computer software)** first produced in the performance of a contract. **For computer software**, the scope of the Government's license includes all of the above rights **except the right to distribute to the public**…. If an agency obtains a different license, the contractor shall clearly state the scope of that license in a conspicuous place on the medium on which the data is recorded. For example, if the data is delivered as a report, the terms of the license shall be stated on the cover, or first page, of the report.

(5) The clause requires the contractor to affix the applicable copyright notices of 17 U.S.C. 401 or 402, and acknowledgment of Government sponsorship, (including the contract number) to data when it asserts copyright in data. **Failure to do so could result in such data being treated as unlimited rights data** (see 27.404-5(b)).

CMU Guidance with respect to FAR, Rights in Data, Copyrighted Works, and Alternative IV [FAR 4]

Introduction to Intellectual Property using the Federal Acquisitions Regulations (FAR)

Slide 11: **Most agencies (other than DOD, DOE and NASA) use FAR 52.227-14**

Slide 12: Inclusion of **Alternate IV** is important because it provides university with the right to claim copyright without limitation **in any data, including software**

Slide 14: **Why request Alternate IV**?

Differentiates between data and software: Gives government rights to data; Government cannot distribute software

Provides universities with **right to claim copyright without limitation in any data (including technical data and software)**

Required to be used in contracts for **basic or applied research to be performed solely by universities**

Some restrictions apply

Significance of Alternate IV:

1. **Required for all contracts for basic or applied research to be performed solely by colleges and universities**
2. ***May* be used for other contracts**
3. **Applies to any data – not just to scientific and technical publications – thus enabling Contractor to assert copyright, without prior approval, to software**

DFARS 252.227-7013 Rights in Technical Data—Noncommercial Items (February 2014) [DFARS 1]

(a) Definitions. As used in this clause—

(12) "**Government purpose**" means any activity in which the United States Government is a party, including cooperative agreements with international or multi-national defense organizations, or sales or transfers by the United States Government to foreign governments or international organizations. Government purposes include competitive procurement, **but do not include the rights** to use, modify, reproduce, release, perform, display, or disclose technical data **for commercial purposes** or authorize others to do so.

(13) "**Government purpose rights**" means the rights to—

(i) Use, modify, reproduce, release, perform, display, or disclose technical data **within the Government without restriction**; and

(ii) Release or disclose technical data **outside the Government** and authorize persons to whom release or disclosure has been made to use, modify, reproduce, release, perform, display, or disclose that data **for United States government purposes**.

(14) "**Limited rights**" means the rights to use, modify, reproduce, release, perform, display, or disclose technical data, in whole or in part, **within the Government**….

(15) "**Technical data**" means recorded information, regardless of the form or method of the recording, of a scientific or technical nature (including computer software documentation). The term **does not include computer software** or data incidental to contract administration, such as financial and/or management information.

(16) "**Unlimited rights**" means rights to use, modify, reproduce, perform, display, release, or disclose technical data in whole or in part, in any manner, and **for any purpose whatsoever**, and to have or **authorize others to do so**.

252.227-7013 Rights in Technical Data – Noncommercial Items (February 2014) [DFARS 2]

(b) Rights in technical data. The Contractor grants or shall obtain for the Government the following royalty free, world-wide, nonexclusive, irrevocable license rights in technical data other than computer software documentation: ….

(1) **Unlimited rights.** The Government shall have unlimited rights in **technical data** that are—

(i) Data pertaining to an item, component, or process which has been or will be developed exclusively with Government funds; …

(2) **Government purpose rights.**

(i) The Government shall have government purpose rights for a five-year period, or such other period as may be negotiated, in **technical data**—

(A) That pertain to items, components, or processes developed with **mixed funding** except when the Government is entitled to unlimited rights in such data as provided in paragraphs (b)(1)(ii) and (b)(1)(iv) through (b)(1)(ix) of this clause; or

(B) Created with **mixed funding** in the performance of a contract that does not require the development, manufacture, construction, or production of items, components, or processes….

(3) **Limited rights.**

(i) Except as provided in paragraphs (b)(1)(ii) and (b)(1)(iv) through (b)(1)(ix) of this clause, the Government shall have limited rights in **technical data**—

(A) Pertaining to items, components, or processes developed exclusively at private expense and marked with the limited rights legend prescribed in paragraph (f) of this clause; or

(B) Created exclusively at private expense in the performance of a contract that does not require the development, manufacture, construction, or production of items, components, or processes….

IDA | DFARS 227.7103-9 Copyright [DFARS 3]

Note: Subparts 227.71 (Rights in Technical Data) and 227.72 (Rights in Computer Software and Computer Software Documentation) prescribe policies and procedures for the acquisition of technical data and computer software and computer software documentation and rights for the Government to use, modify, reproduce, release, perform, display or disclose the technical data and computer software and computer software documentation. [DISA Data Rights webpage]

227.7103-9 Copyright.

(a) Copyright license.

(1) **The clause at 252.227-7013, Rights in Technical Data–Noncommercial Items, requires a contractor to grant or obtain for the Government license rights which permit the Government to reproduce data, distribute copies of the data, publicly perform or display the data or, through the right to modify data, prepare derivative works.** The extent to which the Government, and others acting on its behalf, may exercise these rights varies for each of the **standard data rights licenses** obtained under the clause. When non-standard license rights in technical data will be negotiated, negotiate the extent of the copyright license concurrent with negotiations for the data rights license. Do not negotiate a copyright license that provides less rights than the standard limited rights license in technical data.

(2) The clause at 252.227-7013 does not permit a contractor to incorporate a third party's copyrighted data into a deliverable data item unless the contractor has obtained an appropriate license for the Government and, when applicable, others acting on the Government's behalf, or has obtained the contracting officer's written approval to do so. Grant approval to use third party copyrighted data in which the Government will not receive a copyright license only when the Government's requirements cannot be satisfied without the third party material or when the use of the third party material will result in cost savings to the Government which outweigh the lack of a copyright license.

IDA | Better Buying Power Summary (January 2013) [DFARS 4]

Rights Category	Applies to These Types of TD or CS	Rights Criteria	Permitted Uses Within the Government	Permitted Uses by Third Parties Outside the Government[1]
Unlimited Rights (UR)	Noncommercial TD and CS	Developed exclusively at Government expense, and certain types of data (e.g., FFF, OMIT, CSD)	All uses; no restrictions	All uses; no restrictions
Government Purpose Rights (GPR)	Noncommercial TD and CS	Developed with mixed funding	All uses; no restrictions	For "Government Purposes" only; no commercial use
Limited Rights (LR)	Noncommercial TD only	Developed exclusively at private expense	Unlimited; except may not be used for manufacture	Emergency repair or overhaul[2]
Restricted Rights (RR)	Noncommercial CS only	Developed exclusively at private expense	Only one computer at a time; minimum backup copies; modification[3]	Emergency repair/overhaul; certain service/maintenance contracts[2]
Negotiated License Rights	Any/all TD and CS– including commercial TD and CS	Mutual agreement of the parties; use whenever the standard categories do not meet both parties' needs	As negotiated by the parties; however, must not be less than LR in TD and must not be less than RR in noncommercial CS (consult with legal counsel as other limits apply)	
SBIR Data Rights	Noncommercial TD and CS	All TD or CS generated under an SBIR contract	All uses; no restrictions	Cannot release or disclose except to Government support contractors
Commercial TD License Rights	Commercial TD only	TD related to commercial items (developed at private expense)	Unlimited in FFF and OMIT; other rights as negotiated	
Commercial CS Licenses	Commercial CS only	Any commercial CS or CS documentation	As specified in the commercial license customarily offered to the public[4]	

http://www.disa.mil/about/legal-and-regulatory/datarights-ip/~/media/Files/DISA/About/GC/BuyingPowertrifold.pdf

Agency Public Access Plans: Publication and Data Provisions

Karen D. Gordon
25 November 2015 (updated 16 December 2016)

Synopsis

Key features of several early agency public access plans (those released between July 2014 and April 2015) are summarized on the following pages. Table 1 describes the publication provisions of the plans, and Table 2 covers the data provisions.

Table 1 indicates that there are two leading implementations of public access to publications. The **NIH PubMed Central** (PMC) capability is being used by several Health and Human Services (HHS) organizations, as well as the Department of Veterans' Affairs (VA), NASA, and the National Institute of Standards and Technology (NIST). The Department of Energy **(DOE) Public Access Gateway for Energy and Science** (PAGES) technology (via separate implementations) is being used by the DoD, the National Science Foundation (NSF), and IARPA (via the DoD implementation).

Table 1 also reveals some policy differences among agencies with respect to scope and bulk downloads. NSF gives **peer-reviewed conference papers** the same treatment as peer-reviewed journal papers, and the National Aeronautics and Space Administration (NASA) indicates an inclination to do so as well. Also, NSF, NASA, and the U.S. Department of Agriculture (USDA) recognize the importance of supporting **bulk downloads of full-text articles** for research purposes (e.g., text mining).

Table 2 shows that agency public access plans have considerable variation with respect to data on several points, including:

- The **scope of the data** to be shared and preserved (e.g., data is displayed in peer-reviewed journal articles, data underlying the results is published in peer-reviewed journal articles, all data created or collected during the course of a research project, etc.).
- Whether **software** is considered to be data.
- The length of the **data retention periods.**
- The **Data repositories** where data is stored (agency repositories, discipline-specific repositories, institutional repositories, publisher repositories (as supplementary material), etc.).
- The **Metadata repositories** where metadata on research datasets is stored.

Table 1. Agency Public Access Plans: Publication Provisions

Agency	Public Access Plan	Publication Solution	Scope of Publication Solution	Bulk Downloads of Full-Text Publications
Department of Health and Human Services (HHS) National Institutes of Health	NIH Public Access Plan February 2015	PubMed Central (PMC)	Peer-reviewed journal articles	No, except for "PMC Open Access Subset" of articles. "NIH systems detect and prevent bulk [unauthorized] downloading and will immediately cut off any sites, foreign or domestic, that appear to be abusing copyrighted property."
HHS Agency for Healthcare and Quality	AHRQ Public Access Plan February 2015	PMC	Peer-reviewed journal articles	No, except for "PMC Open Access Subset" of articles.
HHS Office of the Assistance Secretary for Preparedness and Response	ASPR Public Access Plan February 2015	PMC	Peer-reviewed journal articles	No, except for "PMC Open Access Subset" of articles.
HHS Food and Drug Administration	FDA Public Access Plan February 2015	PMC	Peer-reviewed journal articles	No, except for "PMC Open Access Subset" of articles.
Dept. of Veterans Affairs	VA Public Access Plan March 2015	PMC	Peer-reviewed journal articles	No, except for "PMC Open Access Subset" of articles.
National Institute of Standards and Technology	NIST Public Access Plan April 2015	NIST-branded PMC portal	Peer-reviewed journal articles	No, except for "PMC Open Access Subset" of articles.

Agency	Public Access Plan	Publication Solution	Scope of Publication Solution	Bulk Downloads of Full-Text Publications
National Aeronautics and Space Administration	NASA Public Access Plan November 2014	NASA-branded PMC portal	Peer-reviewed journal articles, at least initially **"Applicability to peer-reviewed conference abstracts and proceedings … will be determined"**	No, except for "PMC Open Access Subset" of articles. NASA plans states, "Acceptable use policies will reflect the provisions of PMC's public license or repository Terms of Service." At the same time, NASA plan states, **"Bulk downloads for research purposes should be permitted as an acceptable use."**
HHS Centers for Disease Control and Prevention	CDC Public Access Plan January 2015	PMC and CDC Stacks (peer-reviewed journal articles will be dual-hosted)	Peer-reviewed journal articles, although scope of CDC Stacks is much broader	For PMC: No, except for "PMC Open Access Subset" of articles. For CDC Stacks: No, there is "no automated system for downloading publications."
National Oceanic and Atmospheric Administration	NOAA Public Access Plan February 2015	NOAA Institutional Repository, using CDC Stacks technology	Peer-reviewed journal articles, although scope of NOAA Repository is much broader	No, there is "no automated system for downloading publications in CDC Stacks."
Dept. of Energy	DOE Public Access Plan July 2014	Public Access Gateway for Energy and Science (PAGES)	Peer-reviewed journal articles	No, "[t]he distributed nature of PAGES' full-text content inherently makes unauthorized mass downloading … difficult. For the limited full-text content it hosts publicly, OSTI will enforce a download limit and post appropriate fair use policies."
National Science Foundation	NSF Public Access Plan March 2015	PAGES	Peer-reviewed journal articles and **juried conference papers**	Not initially, but **"NSF intends to enable such uses while protecting the integrity of the scientific record from unauthorized redistribution of scholarly content."**

Agency	Public Access Plan	Publication Solution	Scope of Publication Solution	Bulk Downloads of Full-Text Publications
Dept. of Defense	DoD Public Access Plan February 2015	Defense Technical Information Center (DTIC), using PAGES technology	Peer-reviewed journal articles, although scope of DTIC is much broader	No, except in special cases. "Bulk downloads will be permitted only as authorized and by special arrangement. DoD web usage monitoring protocols will alert system administrators of potentially improper practices. These will be investigated as they occur."
U.S. Dept. of Agriculture	USDA Public Access Plan November 2014	PubAg: Public Access to Agricultural Scholarly Publication System	Peer-reviewed journal articles	Not initially, but **planned enhancements include support for text mining.**
U.S. Agency for International Development[1]	Standard Provisions for U.S. NGOs December 2014	Development Experience Clearinghouse (DEC) (See USAID Development Experience Info)	"Intellectual Work," which includes peer-reviewed journal articles, conference papers, technical reports, and much more	Unclear. There appear to be no mechanisms to facilitate bulk downloads; at the same time, there are no statements indicating that bulk downloads are prohibited.
Other released plans: Smithsonian Institute (August 2015, Plan); DOT (November 2015, Plan); U.S. Geological Survey (February 2016, Plan); Dept of Education (October 2016, Plan); USAID (October 2016, Plan); Office of the Director of National Intelligence (October 2016, Plan); Environmental Protection Agency (November 2016, Plan); HHS Administration for Community Living (February 2016) **In-progress plans**: Dept of Homeland Security				

[1] Note that the USAID publication policy has recently been supplemented with a formal USAID Public Access Plan, approved in October 2016. The plan addresses publications and data, and it builds on the Development Experience Clearinghouse (DEC) as well as on the Development Data Library (DDL).

Table 2. Agency Public Access Plans: Data Provisions

Agency	Link to Public Access Plan	Highlights of Public Access Plan with respect to Data
HHS/ NIH	NIH Public Access Plan February 2015	"[D]igital scientific data includes data that are used to support a scientific publication as well as **data from completed studies that might never be published**…. [but] **does not include software** per se." NIH will require "**data management plans** that will express the investigator's commitment to sharing their data, which will at a minimum consist of the **data underlying any publications**…" "NIH will encourage supported researchers to deposit data in established **public repositories**, where applicable, for archiving and preservation. In some cases, NIH data management **policies may specify particular standards and repositories** to be used by funded researchers…." NIH intends to "[d]evelop additional data management policies to increase public access to **designated types** of biomedical research data … as has been done with **genomic data, autism research**, and other areas of science." Such policies would stipulate, among other things, **designated repositories**.
HHS/ AHRQ	AHRQ Public Access Plan February 2015	"All AHRQ-funded researchers will be required to include a **data management plan** for sharing **final research data** in digital format, or state why data sharing is not possible." "AHRQ will promote the deposit of data in **publicly accessible databases**, where appropriate and available…. [However, at the same,] to ensure long-term preservation and full access to the public, **AHRQ will contract with a commercial repository** to accept and manage data submitted by … researchers…. If the data are made available via another mechanism, AHRQ will publicize the **location of the data** to the public on its Web site and provide a link to the data."
HHS/ ASPR	ASPR Public Access Plan February 2015	ASPR research "must have a reviewed and approved **data management plan**…. All ASPR-funded researchers will be required to make the **data underlying the conclusions of peer-reviewed scientific research publications** freely available in public repositories at the time of initial publication…."

Agency	Link to Public Access Plan	Highlights of Public Access Plan with respect to Data
HHS/ FDA	FDA Public Access Plan February 2015	FDA policies will require researchers to operate under approved **data management plans**. The FDA "will look to the **practices and norms of the scientific disciplines and communities** engaged in the research, as well as applicable restrictions on disclosure, when evaluating data management plans." "[T]o the extent a researcher believes that **long-term preservation and/or public access to data is not justified** or appropriate, researchers must provide an explanation based upon balancing the relative value of long-term preservation and/or access and the associated cost and administrative burden." "Researchers … will be expected to commit to sharing **digital data underlying their research findings upon publication** of the findings in a peer-reviewed article." "… FDA expects that, in accordance with their data management plans, researchers would make datasets publicly accessible in **discipline specific data repositories**, wherever available."
VA	VA Public Access Plan March 2015	VA will research proposals to include "a **data management plan** that describes how, where, and the extent to which they will make the data and results of their research available to the public and specifying what data will be available in machine readable formats." The plan must "specifically include how the **final research datasets underlying scientific publications** will be made available for discovery, retrieval, and analysis…." "VA will begin sharing digital data from VA-funded research through **controlled public access mechanisms** … and move as expeditiously as possible toward open public access mechanisms. All VA-funded researchers will be required to share all **digital data underlying the published results from all VA funded research** at least under controlled public access mechanisms…." "The development of archiving, data platforms, and sharing resources for **VA research data** will necessarily unfold within the larger context of [VA efforts] to promote interoperability and openness in accordance with OMB Memorandum M-13-13." Until **newly proposed VA infrastructure** is in place, "investigators for VA-funded research will be required to complete a data inventory and submit this inventory at the time of application to indicate exactly **where their data will be stored**."

Agency	Link to Public Access Plan	Highlights of Public Access Plan with respect to Data
NIST	NIST Public Access Plan April 2015	"NIST's plan for providing public access to data consists of three components: **data management plans (DMPs)**, an Enterprise Data Inventory (EDI), and a Common Access Platform providing a public access infrastructure." **DMPs** must include a "plan describing whether and how data generated will be reviewed and made available to the public… and explicitly describe how they will make the **data that underlies scientific publications** available for discovery, retrieval, and analysis." The **EDI is a catalog of the datasets** that are generated via NIST-sponsored research to enable researchers to link those datasets to the scientific literature, other datasets, etc. The **CAP** "will use the information gained in the first two phases to put in place production-level infrastructure and populate it with persistent identifiers and metadata for all publicly available NIST data."
NASA	NASA Public Access Plan November 2014	NASA will require proposals to include a "**Data Management Plan (DMP)** that describes whether and how data generated through the course of the proposed research will be shared and preserved (including timeframe), or explains why data sharing and/or preservation are not possible or scientifically appropriate. "DMPs must provide a plan for making **research data that underlie the results and findings in peer-reviewed publications** digitally accessible at the time of publication or within a reasonable time period after publication. This includes data (or how to access data) that are **displayed in charts and figures**…. This requirement could be met by including the data as **supplementary information to the published article**, through NASA archives, or other means."
HHS/ CDC	CDC Public Access Plan January 2015	"**Data management plans** [*one per dataset*] will require both intramural and extramural scientists seeking funding to describe how and where they will make their data available to the public and explicitly describe how they will make the **data that underlies scientific publications** available for discovery, retrieval, and analysis." Plan covers **public health research data** – "those collected or generated systematically to increase the stock of knowledge, including but not limited to epidemiology, laboratory, and environmental studies. Included are **microdata and aggregated data**, **whether or not they lead to publication**, as long as they are determined to be of use or value to the scientific community…. As a general rule, "**[d]ata underlying research papers** will be published coincident with the paper's publication…. At a minimum, the dataset release will consist of a machine-readable version of the **aggregated data used in the paper's data tables**. At most, the dataset release can consist of individual-level (micro) data [possibly in a follow-up dataset release]."

Agency	Link to Public Access Plan	Highlights of Public Access Plan with respect to Data
NOAA	NOAA Public Access Plan February 2015	"For NOAA, **'scientific data' specifically means environmental data**…. The majority of this plan refers to **requirements or activities that already exist**, especially in the context of environmental data…. NOAA will continue to require intramural [and contract] data producers to develop comprehensive **data management plans** [and] extramural grantees to include a **data sharing plan** in their proposals." NOAA will require "scientists seeking funding to describe how and where they will make their data available to the public and explicitly require the **data that underlie the conclusions of peer-reviewed scientific publications be made available** for discovery, retrieval, and analysis for free at the time of publication." "NOAA maintains three world-class **National Data Centers** specializing in the long-term preservation and stewardship of environmental data…. NOAA will continue efforts to ensure intramural datasets are submitted for long-term preservation… **NOAA will endeavor to archive extramural data of long-term relevance but likely cannot archive everything**." If not suitable for archiving in a NOAA National Data Center, extramural datasets can be deposited in "appropriate repositories." "**NOAA Big Data Partnership** will start pilot project(s) in 2015 to copy a subset of datasets to **commercial Cloud** alongside computation resources, to enable the creation of value-added products and services by the private sector."
DOE	DOE Public Access Plan July 2014	DOE will require **data management plans** describing "whether and how data generated in the course of the proposed research will be shared and preserved and, at a minimum, …how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved." "DMPs should provide a plan for making all **research data displayed in publications [in charts, figures, images, etc.]** resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. In addition, the **underlying digital research data** used to generate the displayed data should be made as accessible as possible to the public…." "Individual research offices will encourage researchers to deposit data in existing **community or institutional repositories** or to submit these data to the **article publisher as supplemental information**." "**DOE currently supports a number of publicly accessible repositories of research data** and … is active in developing new repositories to meet mission goals." "The Office of Energy Efficiency and Renewable Energy (EERE) will include detailed requirements to ensure specific research data are submitted to the **Open Energy Information Platform (OpenEI)**, a centralized and secure resource for publicly accessible energy data managed by the National Renewable Energy Laboratory (NREL). All publicly accessible data on OpenEI will be integrated into the Department of Energy's **Enterprise Data Inventory and its Public Data Listing**, which can be found on energy.gov/data." DOE Office of Science Advanced Scientific Computing Research (ASCR) (webpage): "ASCR considers **software** to be a data artifact that is covered by the Office of Science Statement on Digital Data Management."

Agency	Link to Public Access Plan	Highlights of Public Access Plan with respect to Data
NSF	NSF Public Access Plan March 2015	Still calls for DMPs to address "[t]he types of **data, samples, physical collections, software, curriculum materials, and other materials** to be produced in the course of the project." "**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." "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." "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**…. 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." "NSF intends to exercise discretion in determining whether scientific research data resulting from an award will be subject to the requirements of the [Executive Order and Open Data Policy (OMB M-13-13]…. 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**."
DoD	DoD Public Access Plan February 2015	DoD will require DMPs that describe whether, how, and where the digital scientific data created or gathered during the course of a research project will be made available to the public; and "explicitly describe how the **data that underlies scientific publications** will be available for discovery, retrieval, and analysis." DoD plans "a decentralized approach for storing **data in public repositories**, with a **centralized data catalog/locator at DTIC** to consolidate the metadata." "DoD will develop requirements for the submission of metadata to DTIC. The metadata for scientific data will include, at a minimum, the **common core metadata** schema in use by the federal government, found at https://project-open-data.cio.gov/."

Agency	Link to Public Access Plan	Highlights of Public Access Plan with respect to Data
USDA	USDA Public Access Plan November 2014	Extramural and intramural research studies will require **data management plans**. "These data management plans will, at a minimum, describe how the researcher(s) will provide for long-term preservation of, and access to, the digital scientific data created by the proposed study. Alternatively, researchers can explain in their data management plans why long term preservation and access cannot be justified, if applicable." "All USDA-funded researchers will be required to comply with USDA's policy for making the **digital data underlying the conclusions of peer-reviewed scientific research publications** freely available in public repositories in machine readable formats." USDA will "[e]ncourage development of **discipline-based data management standards and data repositories** by scientists who are undertaking pilot program activities."
USAID[2]	USAID Open Data Policy March 2015	The USAID Open Data Policy **does not require data management plans**. Instead, it mandates that data be submitted to the USAID Development Data Library (DDL): "The DDL is the Agency's repository of USAID-funded, machine readable data created or collected by the Agency and its implementing partners. **Datasets and supporting documentation created or collected directly by USAID Operating Units or under USAID-funded awards must be submitted for inclusion in the DDL.**" Research data is among the types of data to be submitted to the DDL. However, "USAID recognizes the value of research data, not only to the general public, but to the academic and scientific communities…. [It supports] the inclusion of USAID-funded research in databases commonly accessed by the academic and scientific communities…. [Therefore,] Should USAID staff or implementing partners submit a Dataset to a **publicly accessible research database**, they are not required to submit the data to the DDL. However, they must submit a **notice to the DDL**, providing details on where and how to access the data…."
Other released plans: Smithsonian Institute (August 2015, Plan); DOT (November 2015, Plan); U.S. Geological Survey (February 2016, Plan); Dept of Education (October 2016, Plan); USAID (October 2016, Plan); Office of the Director of National Intelligence (October 2016, Plan); Environmental Protection Agency (November 2016, Plan); **In-progress plans:** Dept. of Homeland Security		

[2] Note that the USAID data policy has recently been supplemented with a formal USAID Public Access Plan, approved in October 2016. It builds on the Development Data Library, and it requires researchers to submit DMPs.

Sample Research Data Retention Periods

Karen D. Gordon
15 February 2016

Synopsis

IDA investigated retention periods for research data, since there is some debate on how long research data should be preserved. As a part of its investigation, IDA reviewed the data retention policies outlined in OMB Circular A-110[1] and in a March 2012 Council on Government Relations (COGR) publication.[2] IDA also examined the research data retention policies at the top 25 research universities in terms of Federal research and development (R&D), according to FY 2012 Federally Financed R&D Expenditures.[3]

Key findings included the following:

- **OMB Circular A-110** and the **Council on Governmental Relations (COGR)**, which is an Association of Research Universities, specify a **minimum retention period of 3 years** for research data.
- The top **research universities** typically have minimum research data retention periods of **between 3 and 7 years**.

IDA also found that 42 CFR Part 93 (**Public Health Service Policies on Research Misconduct**) specifies a **6-year time limit on raising allegations of misconduct** (§93.105). Exceptions to the 6-year limitation are: (1) subsequent use exception; (2) health or safety of the public exception (i.e., the alleged misconduct, if it occurred, would possibly have a substantial adverse effect on the health or safety of the public); and (3) "grandfather" exception (i.e., the allegation of research misconduct was received before this time limit went into effect.). Thus, it is not surprising that some universities and medical schools have research data retention periods that exceed 3 years but are not unbounded.

Details are provided in the table on the following pages.

1 OMB Circular A-110 (amended 30 September 1999) addresses the subject "Uniform Administrative Requirements for Grants and Agreements With Institutions of Higher Education, Hospitals, and Other Non-Profit Organizations." Subpart C (Post-Award Requirements) addresses "Retention and access requirements for records" in Section 53 (one of four sections under "Reports and Records").

2 COGR is an association of research universities. Its 1 March 2012 publication, "Access to, Sharing and Retention of Research Data: Rights & Responsibilities," addresses research data retention periods.

3 https://ncsesdata.nsf.gov/herd/2012/html/HERD2012_DST_05.html.

Sample Research Data Retention Periods

	Minimum Retention Period	Reference
OMB Circular A-110 (Section _.53), 30 Sep 1999	3 years	https://www.whitehouse.gov/omb/circulars_a110/#53 Section C. _.53, Retention and access requirements for records: "Financial records, supporting documents, statistical records, and all other records pertinent to an award shall [with a few stated exceptions] be **retained for a period of three years** from the date of submission of the final expenditure report."
Council on Government Relations (COGR): An Association of Research Universities, "Access to, Sharing and Retention of Research Data: Rights & Responsibilities," 1 Mar 2012	3 years	http://www.cogr.edu/sites/default/files/access_to_sharing_and_retention_of_research_data-_rights_%26_responsibilities.pdf "Section C._.53 of Circular A-110 requires that all records – financial records and the supporting documentation, scientific data including notebooks, etc. – be maintained for three years or, in the case of litigation started before the end of the original three year period, until any claim or audit is resolved and final action taken. **Thus, a three-year period is the minimum amount of time that research data should be kept by the grantee**."
Johns Hopkins Univ.†	3 years	http://jhuresearch.jhu.edu/DMP_AppendixFive.pdf
Univ. of Washington	3 years	http://www.washington.edu/research/osp/gim/gim37.html
Univ. of Michigan	3 years	http://research-compliance.umich.edu/operations-manual-laws-regulations-and-standards
Univ. of Pennsylvania	7 years	http://www.archives.upenn.edu/urc/recrdret/researchadmin.html
Univ. of California, San Diego	6 years	http://www.ucop.edu/research-policy-analysis-coordination/resources-tools/contract-and-grant-manual/chapter17/chapter-17-300.html
Columbia University	3 years	http://scholcomm.columbia.edu/data-management/frequently-asked-questions/
Univ. of Pitt	7 years	http://www.provost.pitt.edu/documents/RDM_Guidelines.pdf
Stanford Univ.	3 years	https://doresearch.stanford.edu/policies/research-policy-handbook/conduct-research/retention-and-access-research-data
Univ. of North Carolina	5 years	http://gradschool.unc.edu/academics/resources/ethics.html
Harvard Univ.	7 years	http://files.vpr.harvard.edu/files/vpr-documents/files/research_records_and_data_retention_and_maintenance_faq_guidance_7_31_12.pdf
Duke Univ.	5 years	http://provost.duke.edu/wp-content/uploads/FHB_App_P.pdf
Univ. of Wisconsin	7 years	http://www.irb.wisc.edu/documents/PolicyDataStewardship.pdf

Univ. of California, San Francisco	6 years	http://www.ucop.edu/research-policy-analysis-coordination/resources-tools/contract-and-grant-manual/chapter17/chapter-17-300.html
Univ. of California, Los Angeles	6 years	http://www.ucop.edu/research-policy-analysis-coordination/resources-tools/contract-and-grant-manual/chapter17/chapter-17-300.html
Pennsylvania State Univ.	5 years	http://guru.psu.edu/policies/rpg01.html
Yale Univ.	3 years	http://www.yale.edu/hrpp/resources/docs/handbook.pdf
Massachusetts Institute of Technology	3 years	http://osp.mit.edu/help-and-training (points to COGR)
Univ. of Minnesota	3 years	http://policy.umn.edu/research/researchdata-faq
Georgia Institute of Technology	3 years	http://d7.library.gatech.edu/research-data/archiving#retention
Cornell Univ.	3 years 6 years (Medical School)	http://www.dfa.cornell.edu/cms/treasurer/policyoffice/policies/volumes/governance/upload/vol4_7.pdf http://weill.cornell.edu/research_compliance/pdf/data_retention_procedures_final.doc
Vanderbilt Univ.	3 years? 7 years (Medical School)	Assume OMB Circular A-110 http://www.mc.vanderbilt.edu/documents/CoreManagers/files/130922_Core%20Managers%20Meeting.pdf
Ohio State Univ.	5 years	http://orc.osu.edu/files/2011/01/ResearchDataPolicy.pdf
Univ. of Southern California	3 years	https://oprs.usc.edu/files/2013/04/Data_Management_Acquisition-4.5.13.pdf
Washington U, St. Louis	6 years (by inference)	http://research.wustl.edu/PoliciesGuidelines/Documents/ResearchIntegrityPolicy2010Revisions.pdf *(statute of limitations on allegations of research misconduct is 6 years)*
Northwestern Univ.	3 years	http://www.research.northwestern.edu/policies/documents/research_data.pdf
United Kingdom	3 to 10 years (varies by Research Council)	http://www.dcc.ac.uk/sites/default/files/documents/reports/DCC_Curation_Policies_Report.pdf AHRC (Arts & Humanities): 3 years EPSRC (Engineering & Physical Sciences): 10 years
Canada	5 years	http://www.science.gc.ca/default.asp?lang=en&n=2BBD98C5-1
Australia	5 years	http://www.nhmrc.gov.au/_files_nhmrc/publications/attachments/r39_australian_code_responsible_conduct_research_150107.pdf

† The universities listed are the top 25 in terms of Federal R&D, according to FY 2012 Federally Financed R&D Expenditures [http://ncsesdata.nsf.gov/herd/2012/html/HERD2012_DST_05.html].

National Science Foundation (NSF) Guidance on Public Access

Karen D. Gordon
18 March 2016

Synopsis

The NSF public access plan, "Today's Data, Tomorrow's Discoveries: Increasing Access to the Results of Research Funded by the National Science Foundation," was published on 18 March 2015. A link to the plan is posted on the NSF Open Government webpage. The publication of the plan was announced in NSF Press Release 15-021, "National Science Foundation announces plan for comprehensive public access to research results."

The NSF public access requirements – as they relate to publications and data – are laid out in a series of documents and webpages:

- NSF Proposal and Award Policies and Procedures Guide (PAPPG), Part I – Grant Proposal Guide (GPG), January 25, 2016;
- NSF Proposal and Award Policies and Procedures Guide (PAPPG), Part II – Award & Administration Guide (AAG), January 25, 2016;
- National Science Foundation (NSF) Grant General Conditions (GC-1), Effective January 25, 2016;
- NSF Public Access Policy: Public Access To Results of NSF-funded Research (webpage);
- Dissemination and Sharing of Research Results (webpage).

NSF program solicitations point to the Grant Proposal Guide (GPG), the Award & Administration Guide (AAG), and the Grant General Conditions (GC-1). These documents in turn point to the NSF Public Access Policy and the Dissemination and Sharing of Research Results webpages, as illustrated in Figure 1 on the next page.

NSF-wide Data Management Plan (DMP) requirements are detailed in the Grant Proposal Guide (GPG) in Section II.C.2.j, "Special Information and Supplementary Documentation." Directorate-specific details are available via links on the Dissemination and Sharing of Research Results webpage. Program solicitations may further refine the requirements as necessary.

The remaining pages of this section provide key excerpts from the above NSF documents/webpages and from a sample NSF Program Solicitation – NSF 16-533, Cybersecurity Innovation for Cyberinfrastructure (CICI) – issued on 20 January 2016.

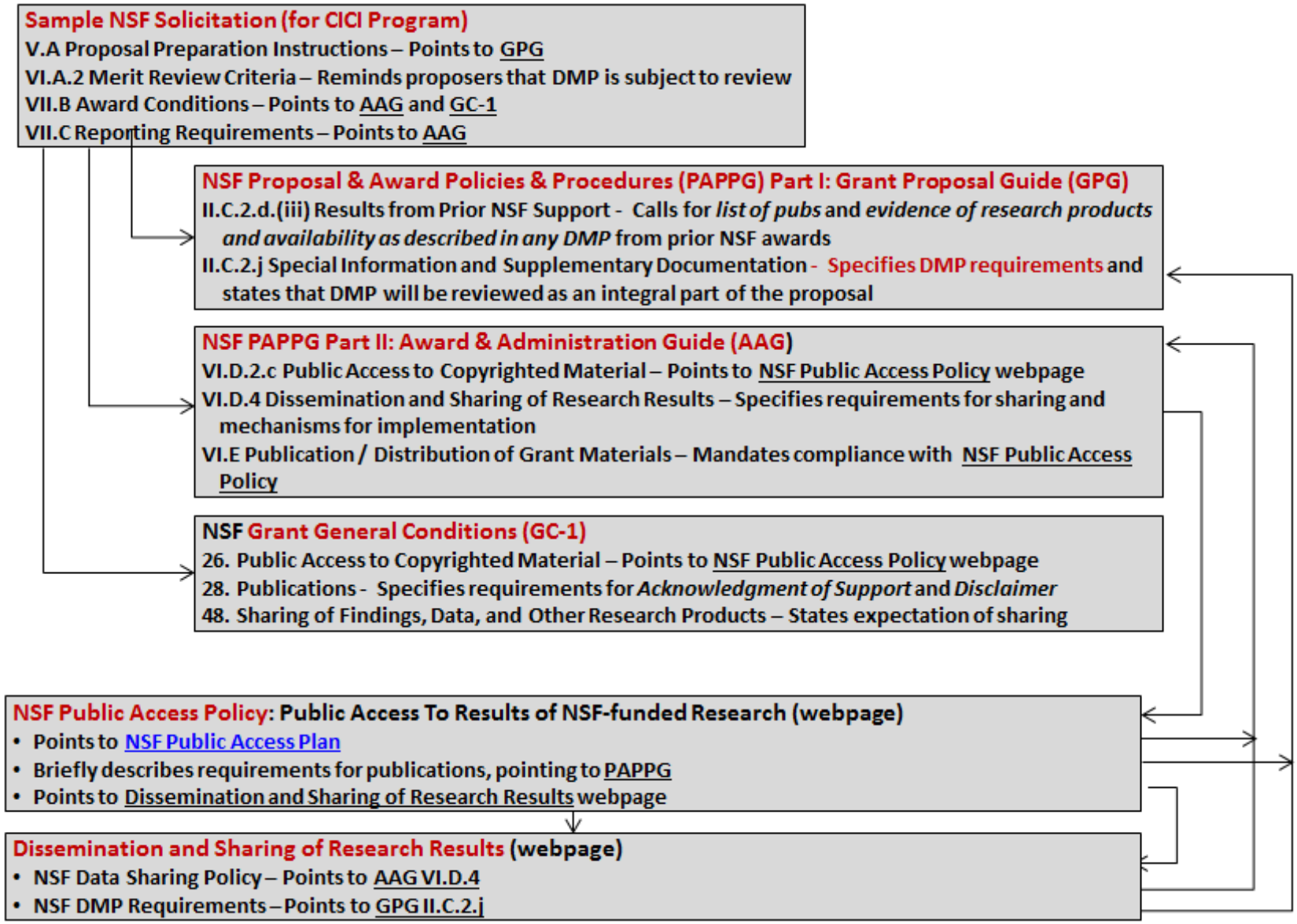


Figure 1. Organization of NSF Guidance on Public Access

Sample NSF Program Solicitation [Excerpts]
January 20, 2016

Cybersecurity Innovation for Cyberinfrastructure (CICI)
PROGRAM SOLICITATION
NSF 16-533

IMPORTANT INFORMATION AND REVISION NOTES
… Any proposal submitted in response to this solicitation should be submitted in accordance with the revised **NSF Proposal & Award Policies & Procedures Guide (PAPPG) (NSF 16-1)**, which is effective for proposals submitted, or due, on or after January 25, 2016.

V. PROPOSAL PREPARATION AND SUBMISSION INSTRUCTIONS
V.A. Proposal Preparation Instructions
Full Proposal Preparation Instructions: Proposers may opt to submit proposals in response to this Program Solicitation via Grants.gov or via the NSF FastLane system.

- Full proposals submitted via FastLane: Proposals submitted in response to this program solicitation should be prepared and submitted in accordance with the general guidelines contained in the **NSF Grant Proposal Guide (GPG)**. The complete text of the GPG is available electronically on the NSF website at: http://www.nsf.gov/publications/pub_summ.jsp ?ods_key=gpg....

VI. NSF PROPOSAL PROCESSING AND REVIEW PROCEDURES
VI.A. Merit Review Principles and Criteria
VI.A.2. Merit Review Criteria
… Proposers are reminded that reviewers will also be asked to review the **Data Management Plan** and the Postdoctoral Researcher Mentoring Plan, as appropriate.

VII. AWARD ADMINISTRATION INFORMATION
VII.B. Award Conditions
An NSF award consists of: (1) the award notice, which includes any special provisions applicable to the award and any numbered amendments thereto; (2) the budget, which indicates the amounts, by categories of expense, on which NSF has based its support (or otherwise communicates any specific approvals or disapprovals of proposed expenditures); (3) the proposal referenced in the award notice; (4) the applicable award conditions, such as **Grant General Conditions (GC-1)***; or Research Terms and Conditions* and (5) any announcement or other NSF issuance that may be incorporated by reference in the award notice. Cooperative agreements also are administered in accordance with NSF Cooperative Agreement Financial and Administrative Terms and Conditions (CA-FATC) and the applicable Programmatic Terms and Conditions. NSF awards are electronically signed by an NSF Grants and Agreements Officer and transmitted electronically to the organization via e-mail.

… More comprehensive information on NSF Award Conditions and other important information on the administration of NSF awards is contained in the **NSF Award & Administration Guide**

(AAG) Chapter II, available electronically on the NSF Website at http://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag.

VII.C. Reporting Requirements

… More comprehensive information on NSF Reporting Requirements and other important information on the administration of NSF awards is contained in the **NSF *Award & Administration Guide* (AAG)** Chapter II, available electronically on the NSF Website at http://www.nsf.gov/publications/pub_summ.jsp?ods_key=aag.

NSF Proposal and Award Policies and Procedures Guide
Part I – Grant Proposal Guide (GPG) [Excerpts]
January 25, 2016

II. Proposal Preparation Instructions
C. Proposal Contents
2. Sections of the Proposal
II.C.2.d. Project Description (including Results from Prior NSF Support)

II.C.2.d.(iii) Results from Prior NSF Support

If any PI or co-PI identified on the proposal has received NSF funding with a start date in the past five years (including any current funding and no cost extensions), information on the award is required for each PI and co-PI, regardless of whether the support was directly related to the proposal or not….

The following information must be provided:

a) the NSF award number, amount and period of support;
b) the title of the project;
c) a summary of the results of the completed work, including accomplishments, supported by the award. The results must be separately described under two distinct headings: Intellectual Merit and Broader Impacts;
d) a listing of the **publications resulting from the NSF award** (a complete bibliographic citation for each publication must be provided either in this section or in the References Cited section of the proposal); if none, state "No publications were produced under this award."
e) **evidence of research products and their availability**, including, but not limited to: **data, publications, samples, physical collections, software, and models**, **as described in any Data Management Plan**; and
f) if the proposal is for renewed support, a description of the relation of the completed work to the proposed work.

II.C.2.j. Special Information and Supplementary Documentation

Except as specified below, special information and supplementary documentation must be included as **part of the Project Description** (or part of the budget justification), if it is relevant to determining the quality of the proposed work. Information submitted in the following areas is **not considered part of the 15-page Project Description limitation**. This Special Information and Supplementary Documentation section also is not considered an appendix. Specific guidance on the need for additional documentation may be obtained from the organization's Sponsored Projects Office or in the references cited below….

- **Plans for data management and sharing of the products of research**. Proposals must include a supplementary document of no more than two pages labeled "**Data Management Plan**". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results (see AAG Chapter VI.D.4), and may include:

1. the **types of data**, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;
2. 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);
3. **policies for access and sharing** including provisions for appropriate **protection** of privacy, confidentiality, security, intellectual property, or other rights or requirements;
4. policies and provisions for **re-use, re-distribution, and the production of derivatives**; and
5. plans for **archiving data**, samples, and other research products, and for preservation of access to them.

Data management requirements and plans specific to the Directorate, Office, Division, Program, or other NSF unit, relevant to a proposal are available at: https://www.nsf.gov/bfa/dias/policy/dmp.jsp. If guidance specific to the program is not available, then the requirements established in this section apply.

Simultaneously submitted collaborative proposals and proposals that include subawards are a single unified project and should include only one supplemental combined Data Management Plan, regardless of the number of non-lead collaborative proposals or subawards included.

A valid Data Management Plan may include only the statement that no detailed plan is needed, as long as the statement is accompanied by a clear justification. Proposers who feel that the plan cannot fit within the limit of two pages may use part of the 15-page Project Description for additional data management information. Proposers are advised that the Data Management Plan must not be used to circumvent the 15-page Project Description limitation. The **Data Management Plan will be reviewed as an integral part of the proposal**, considered under Intellectual Merit or Broader Impacts or both, as appropriate for the scientific community of relevance.

NSF Proposal and Award Policies and Procedures Guide
Part II – Award & Administration Guide (AAG) [Excerpts]
January 25, 2016

VI. Other Post Award Requirements and Consideration
VI.D. Intellectual Property
VI.D.2. Copyright

VI.D.2.c. Public Access to Copyrighted Material

NSF's policy on public access to copyrighted material (Public Access Policy) reflects the Foundation's commitment to making certain that, to the extent possible, the American public, industry and the scientific community have access to the results of federally funded scientific research. Pursuant to this policy, awardees must ensure that **articles in peer-reviewed scholarly journals and papers in juried conference proceedings**:

- are deposited in a public access compliant repository (as identified in the Public Access Policy);
- are available for download, reading, and analysis within 12 months of publication;
- possess a minimum set of machine-readable metadata elements as described in the Public Access Policy; and
- are reported in annual and final reports with a persistent identifier.

Either the **final printed version or the final peer-reviewed manuscript** is acceptable for deposit. NSF's Public Access Policy applies to awards, funded in whole or in part, as a result of proposals submitted, or due, on or after January 25, 2016. NSF's Public Access Policy may be viewed at http://www.nsf.gov/news/special_reports/public_access/index.jsp.

Each NSF grant contains as part of the grant terms and conditions, an article implementing the public access requirements.

VI.D.4. Dissemination and Sharing of Research Results

a. Investigators are expected to promptly prepare and submit for publication, with authorship that accurately reflects the contributions of those involved, all significant findings from work conducted under NSF grants. Grantees are expected to permit and encourage such publication by those actually performing that work, unless a grantee intends to publish or disseminate such findings itself.

b. **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. Privileged or confidential information should be released only in a form that protects the privacy of individuals and subjects involved. General adjustments and, where essential, exceptions to this sharing expectation may be specified by the funding NSF Program or Division/Office for a particular field or discipline to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate the legitimate interest of investigators.

A grantee or investigator also may request a particular adjustment or exception from the cognizant NSF Program Officer.

c. Investigators and grantees are encouraged to share **software** and inventions created under the grant or otherwise make them or their products widely available and usable.

d. **NSF normally allows grantees to retain principal legal rights to intellectual property developed under NSF grants** to provide incentives for development and dissemination of inventions, software and publications that can enhance their usefulness, accessibility and upkeep. Such incentives do not, however, reduce the responsibility that investigators and organizations have as members of the scientific and engineering community, to make results, data and collections available to other researchers.

e. **NSF program management will implement these policies for dissemination and sharing of research results, in a way appropriate to field and circumstances, through the proposal review process; through award negotiations and conditions; and through appropriate support and incentives for data cleanup, documentation, dissemination, storage and the like.**

VI.E. Publication / Distribution of Grant Materials

1. NSF Policy

NSF advocates and encourages open scientific and engineering communication. NSF expects significant findings from research it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. **Copyrighted material published in peer-reviewed scholarly journals and papers included in juried conference proceedings must comply with NSF's Public Access Policy** as implemented in the grant terms and conditions.

2. Costs

Cost of documenting, preparing, publishing, disseminating and sharing research findings and supporting material are allowable charges against the grant. (See AAG Chapter V.A.2.c.)

3. Responsibilities

Unless otherwise provided in the grant, preparation, content, editing, identification of authorship and submission for publication of significant research findings are the responsibility of the investigators, consistent with such policies and procedures as the grantee may prescribe.

4. Grantee Obligations

a. Acknowledgement of Support. Unless otherwise provided in the grant, the grantee is responsible for assuring that an acknowledgment of NSF support is made:

(i) in any publication (including Web pages) of any material based on or developed under this project, in the following terms: "**This material is based upon work supported by the National Science Foundation under Grant No. (NSF grant number).**"

(ii) NSF support also must be orally acknowledged during all news media interviews, including popular media such as radio, television and news magazines.

b. Disclaimer. The awardee is responsible for assuring that every publication of material (including World Wide Web pages) based on or developed under this award, except scientific articles or papers appearing in scientific, technical or professional journals, contains the following disclaimer: "Any opinions, findings, and conclusions or recommendations

expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

c. Copies for NSF. The grantee is responsible for assuring that the cognizant NSF Program Officer is provided access to, either electronically or in paper form, a copy of every publication of material based on or developed under this award, clearly labeled with the award number and other appropriate identifying information, promptly after publication.

d. Compliance with NSF Public Access Policy. The grantee is responsible for ensuring that **copyrighted material published in peer-reviewed scholarly journals** and **papers in juried conference proceedings** are accessible to the public in accordance with the grant terms and conditions.

e. Grantees also should note their obligations in regard to **copyrights (see AAG Chapter VI.D.2)** and their responsibilities as members of the scientific and engineering community to **disseminate and share research results (see AAG Chapter VI.D.4)**.

NSF Grant General Conditions (GC-1) [Excerpts]
Effective January 25, 2016

26. Public Access to Copyrighted Material

NSF's Public Access Policy applies to awards, funded in whole or in part, as a result of proposals submitted, or due, on or after January 25, 2016. NSF's Public Access Policy may be viewed at http://www.nsf.gov/news/special_reports/public_access/.

… Pursuant to this policy, awardees must ensure that **all articles in peer-reviewed scholarly journals and papers in juried conference proceedings**:

- are deposited in a public access compliant repository…;
- are available for download, reading, and analysis within 12 months of publication;
- possess a minimum set of machine-readable metadata elements…;
- are reported in annual and final reports with a persistent identifier.

Either the final **printed version or the final peer-reviewed manuscript** is acceptable for deposit.

28. Publications

a. Acknowledgment of Support

The grantee is responsible for assuring that an acknowledgment of NSF support:

1. is made in any publication (including World Wide Web pages) of any material based on or developed under this project, in the following terms: "**This material is based upon work supported by the National Science Foundation under Grant No. (NSF grant number).**"

b. Disclaimer

The grantee is responsible for assuring that every publication of material (including World Wide Web pages) based on or developed under this grant, except scientific articles or papers appearing in scientific, technical or professional journals, contains the following disclaimer: "Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation."

48. Sharing of Findings, Data, and Other Research Products

a. NSF expects significant findings from research and education activities it supports to be promptly submitted for publication, with authorship that accurately reflects the contributions of those involved. **It expects investigators to share with other researchers, at no more than incremental cost and within a reasonable time, the data, samples, physical collections and other supporting materials created or gathered in the course of the work.** It also encourages grantees to share **software** and inventions or otherwise act to make the innovations they embody widely useful and usable.

b. Adjustments and, where essential, exceptions may be allowed to safeguard the rights of individuals and subjects, the validity of results, or the integrity of collections or to accommodate legitimate interests of investigators.

NSF Public Access Policy:
Public Access To Results of NSF-funded Research (webpage) [Excerpts]

The National Science Foundation (NSF or Foundation) has developed a plan outlining a framework for activities to increase public access to scientific publications and digital scientific data resulting from research the foundation funds. **The plan, entitled "Today's Data, Tomorrow's Discoveries,"** is consistent with the objectives set forth in the Office of Science and Technology Policy's Feb. 22, 2013, memorandum, "Increasing Access to the Results of Federally Funded Research," and with long-standing policies encouraging data sharing and communication of research results.

As outlined in section 3.1 of the plan, NSF will require that either the **version of record or the final accepted manuscript** in **peer-reviewed scholarly journals and papers in juried conference proceedings or transactions** 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;
- Be managed to ensure long-term preservation; and
- Be reported in annual and final reports during the period of the award with a persistent identifier that provides links to the full text of the publication as well as other metadata elements.

This NSF requirement will apply to new awards resulting from proposals submitted, or due, on or after the effective date of the **Proposal & Award Policies & Procedures Guide (PAPPG)** that will be issued in January 2016….

NSF's current data management plan requirement and policies on costs of publication and data citation in biographical sketches will remain unchanged for the present while the Foundation undertakes activities to engage the research communities around data management in support of public access goals. Additional guidance at the Foundation, directorate, division, office or program levels may become available in the future….

- The Plan
- The Executive Summary
- Press Release
- Frequently Asked Questions (FAQs)
- Search NSF Awards
- NSF Public Access Feedback
- NSF Public Access Repository (NSF-PAR) Beta

See also NSF's Open Government website.

See also **Dissemination and Sharing of Research Results**.

Dissemination and Sharing of Research Results (webpage) [Excerpts]

NSF Data Sharing Policy

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. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF Data Management Plan Requirements

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

Requirements by Directorate, Office, Division, Program, or other NSF Unit

Links to data management requirements and plans relevant to specific Directorates, Offices, Divisions, Programs, or other NSF units, are provided below. If guidance specific to the program is not provided, then the requirements established in Grant Proposal Guide, Chapter II.C.2.j apply.

Please note that **if a specific program solicitation provides guidance on preparation of data management plans, such guidance must be followed**.

- Biological Sciences Directorate (BIO)
  - Directorate-wide Guidance
- Computer & Information Sciences & Engineering (CISE)
  - Directorate-wide Guidance
- Education & Human Resources Directorate (EHR)
  - Directorate-wide Guidance
- Engineering Directorate (ENG)
  - Directorate-wide Guidance
- Geosciences Directorate (GEO)
  - Directorate-wide Guidance
- Mathematical and Physical Sciences Directorate (MPS)
  - Division of Astronomical Sciences
  - Division of Chemistry
  - Division of Materials Research
  - Division of Mathematical Sciences
  - Division of Physics
- Social, Behavioral and Economic Sciences Directorate (SBE)
  - Directorate-wide Guidance

Data Management & Sharing Frequently Asked Questions (FAQs)

Department of Energy (DOE) Guidance on Public Access

Karen D. Gordon
22 August 2016

Synopsis

The DOE Public Access Plan was published on 24 July 2014. A link to the plan is posted on the DOE Open Government webpage. The publication of the plan was announced in a 4 August 2014 press release, "U.S. Department of Energy Increases Access to Results of DOE-funded Scientific Research."

The DOE public access requirements – as they relate to publications and data – are laid out in a series of documents and webpages:

- DOE Office of Science Statement on Digital Data Management (webpage):
  - DOE Office of Science Suggested Elements for a Data Management Plan (webpage),
  - DOE Office of Science Statement on Digital Data Management FAQ (which points to DOE OSTI DataID Service that provides Digital Object Identifiers (DOIs) for datasets);
- DOE Federal Assistance Reporting Checklist and Instructions for RD&D Projects, October 2014 (document);
- DOE Office of Science Acknowledgments of Federal Support (webpage);
- DOE AN 241.6 Instructions for using E-Link tool to submit metadata on a dataset;
- DOE AN 241.3 Instructions for using E-Link tool to submit an accepted manuscript and associated metadata;

DOE Office of Science funding opportunity announcements (FOAs) state the basic DMP requirements and point to the Statement on Digital Data Management for further information on DOE requirements for public access to research data.

DOE Office of Science FOAs point to the Federal Assistance Reporting Checklist and the Acknowledgments of Federal Support webpage for DOE requirements for public access to scientific publications.

The DOE E-Link tool is used to submit publications (specifically, accepted manuscripts) and associated metadata to DOE. The publications can be searched at the DOE Public Access Gateway for Energy and Science (DOE PAGES).

The DOE E-Link tool is also used to submit metadata on research datasets to DOE. The metadata can be searched at DOE Data Explorer.

The remaining pages of this document provide key excerpts from the above DOE documents/webpages and from a sample DOE Office of Science Funding Opportunity Announcement – DE-FOA-0001528, Computational Materials Sciences – issued on 9 February 2016.

Figure 1. Organization of DOE Guidance on Public Access

Sample DOE Funding Opportunity Announcement [Excerpts]
January 20, 2016

DOE Office of Science Funding Opportunity Announcement
Computational Materials Sciences
Funding Opportunity Number: DE-FOA-0001528

UPDATES AND REMINDERS
REGULATIONS
This FOA and any awards made under it are controlled by 2 CFR 200, the Uniform Administrative Requirements, Cost Principles, and Audit Requirements for Federal Awards, as modified by 2 CFR 910, the Department of Energy Financial Assistance Rules, and 10 CFR 605, the Office of Science Financial Assistance Program.

DATA MANAGEMENT PLAN
The Office of Science has published a new **Statement on Digital Data Management**, published at http://science.energy.gov/funding-opportunities/digital-data-management/, which governs applications submitted under this FOA, and is detailed in Part IV of this FOA.

ACKNOWLEDGMENT OF FEDERAL SUPPORT
The Office of Science published guidance about how its support should be acknowledged at http://science.energy.gov/funding-opportunities/acknowledgements/.

REPORTING
The Office of Science has implemented the federal-wide Research Performance Progress Report (RPPR) through the Portfolio Analysis and Management System (PAMS). The common RPPR format is described at http://www.nsf.gov/bfa/dias/policy/rppr/. Progress Reports are generally due 90 days before the end of each budget period. The Principal Investigator (PI) will receive an automated email from PAMS (<PAMS.Autoreply@science.doe.gov>) thirty days prior to the progress report due date. Some information will be prepopulated. Additional details and changes will be contained in the **Reporting Requirements Checklist** attached to the Assistance Agreement.

III.D. Other Eligibility Requirements

DATA MANAGEMENT PLAN

A Data Management Plan as described in Section IV C. “Application Contents and Forms” - Appendix 6, is required. **Applications that do not have a Data Management Plan will be deemed nonresponsive** and will be not reviewed…..

IV.C Content and Application Forms

IV.C.APPENDIX 6: DATA MANAGEMENT PLAN

Provide a **Data Management Plan (DMP)** that addresses the following **requirements:**

1. DMPs should describe whether and how data generated in the course of the proposed research will be shared and preserved. If the plan is not to share and/or preserve certain data, then the plan must explain the basis of the decision (for example, cost/benefit considerations, other parameters of feasibility, scientific appropriateness, or limitations discussed in #4). At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.
2. DMPs should provide a plan for making all research data displayed in publications resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. This includes data that are displayed in charts, figures, images, etc. In addition, the underlying digital research data used to generate the displayed data should be made as accessible as possible to the public **in accordance with the principles stated in the Office of Science Statement on Digital Data Management** (http://science.energy.gov/funding-opportunities/digital-data-management/). This requirement could be met by including the data as supplementary information to the published article, or through other means. The published article should indicate how these data can be accessed.
3. DMPs should consult and reference available information about data management resources to be used in the course of the proposed research. In particular, DMPs that explicitly or implicitly commit data management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written approval from that facility. In determining the resources available for data management at Office of Science User Facilities, researchers should consult the published description of data management resources and practices at that facility and reference it in the DMP. Information about other Office of Science facilities can be found in the additional guidance from the sponsoring program.
4. DMPs must protect confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all applicable laws, and regulations. There is no requirement to share proprietary data.

DMPs will be reviewed as part of the overall Office of Science research proposal merit review process. Applicants are encouraged to consult the Office of Science website for further information and suggestions for how to structure a DMP: http://science.energy.gov/funding-opportunities/digital-data-management/

- This appendix will not count in the project narrative page limitation.

V.A.2 Merit Review Criteria

SCIENTIFIC AND/OR TECHNICAL MERIT OF THE PROPOSED RESEARCH

- Is the **Data Management Plan** suitable for the proposed research; to what extent does it make the data available and useful to the scientific community?

VI.B. Administrative and National Policy Requirements

VI.B.2. Terms and Conditions

The DOE Special Terms and Conditions for Use in Most Grants and Cooperative Agreements are located at http://energy.gov/management/office-management/operational-management/financial-assistance/financial-assistance-forms under Award Terms.

The standard DOE financial assistance intellectual property provisions applicable to various types of recipients are located at: http://energy.gov/gc/standard-intellectual-property-ip-provisions-financial-assistance-awards ...

VI.B.5. Additional Conditions

PUBLICATIONS

The recipient is expected to publish or otherwise make publicly available the results of the work conducted under any award resulting from this Funding Opportunity Announcement. Publications and other methods of public communication describing any work based on or developed under an award resulting from this Funding Opportunity Announcement must contain an acknowledgment of DOE Office of Science support. The **format for such acknowledgments** is provided at http://science.energy.gov/funding-opportunities/acknowledgements/. The author's copy of any peer-reviewed manuscript accepted for funding must be announced to DOE's Office of Scientific and Technical Information and made publicly available in accordance with the instructions contained in the **Reporting Requirements Checklist incorporated in all Assistance Agreements**.

VIII. Other Information

VIII.F. INTELLECTUAL PROPERTY DEVELOPED UNDER THIS PROGRAM

Patent Rights: The government will have certain statutory rights in an invention that is conceived or first actually reduced to practice under a DOE award. 42 USC 5908 provides that title to such inventions vests in the United States, except where 35 USC 202 provides otherwise for nonprofit organizations or small business firms. However, the Secretary of Energy may waive all or any part of the rights of the United States subject to certain conditions. (See "Notice of Right to Request Patent Waiver" in paragraph G below.)

Rights in Technical Data: Normally, the government has unlimited rights in technical data created under a DOE agreement. Delivery or third party licensing of proprietary software or data developed solely at private expense will not normally be required except as specifically negotiated in a particular agreement to satisfy DOE's own needs or to insure the commercialization of technology developed under a DOE agreement.

DOE Office of Science Statement on Digital Data Management (webpage)

[Excerpts]

July 28, 2014

The Office of Science mission is to deliver the scientific discoveries and major scientific tools that transform our understanding of nature and advance the energy, economic, and national security of the United States. The Office of Science Statement on Digital Data Management has been developed with input from a variety of stakeholders in this mission.

Here, data management involves all stages of the digital data life cycle including capture, analysis, sharing, and preservation. The focus of this statement is sharing and preservation of digital research data.

Table of Contents

Principles

The Office of Science affirms that the following principles related to the management of digital research data directly support fulfillment of its mission.

- Effective data management has the potential to increase the pace of scientific discovery and promote more efficient and effective use of government funding and resources. Data management planning should be an integral part of research planning.
- Sharing and preserving data are central to protecting the integrity of science by facilitating validation of results and to advancing science by broadening the value of research data to disciplines other than the originating one and to society at large. To the greatest extent and with the fewest constraints possible, and consistent with the requirements and other principles of this Statement, data sharing should make digital research data available to and useful for the scientific community, industry, and the public.
- **Not all data need to be shared or preserved. The costs and benefits of doing so should be considered in data management planning**.

Requirements

To integrate data management planning into the overall research plan, the following requirements will apply to all Office of Science research solicitations and invitations for new, renewal, and some supplemental funding issued on or after October 1, 2014. These requirements apply to proposals from all organizations including academic institutions,

DOE National Laboratories, and others. These requirements do not apply to applications to use Office of Science user facilities.

All proposals submitted to the Office of Science for research funding must include a Data Management Plan (DMP) that addresses the following requirements:

1. DMPs should describe whether and how data generated in the course of the proposed research will be shared and preserved. If the plan is not to share and/or preserve certain data, then the plan must explain the basis of the decision (for example, cost/benefit considerations, other parameters of feasibility, scientific appropriateness, or limitations discussed in #4). **At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.**
2. DMPs should provide a plan for making **all research data displayed in publications** resulting from the proposed research open, machine-readable, and digitally accessible to the public at the time of publication. This includes data that are displayed in charts, figures, images, etc. In addition, the **underlying digital research data used to generate the displayed data should be made as accessible as possible to the public** in accordance with the principles stated above. This requirement could be met by including the data as supplementary information to the published article, or through other means. The published article should indicate how these data can be accessed.
3. DMPs should consult and reference available information about data management resources to be used in the course of the proposed research. In particular, DMPs that explicitly or implicitly commit data management resources at a facility beyond what is conventionally made available to approved users should be accompanied by written approval from that facility. In determining the resources available for data management at Office of Science User Facilities, researchers should consult the published description of data management resources and practices at that facility and reference it in the DMP. Information about other Office of Science facilities can be found in the additional guidance from the sponsoring program.
4. **DMPs must protect** confidentiality, personal privacy, Personally Identifiable Information, and U.S. national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; avoid significant negative impact on innovation, and U.S. competitiveness; and otherwise be consistent with all applicable laws, regulations, and DOE orders and policies. There is no requirement to share proprietary data.

DMPs will be reviewed as part of the overall Office of Science research proposal merit review process. **Additional requirements and review criteria for the DMP may be identified by the sponsoring program or sub-program, or in the solicitation.**

Additional Guidance

- **The Principal Investigator should determine which data should be the subject of the DMP and, in the DMP, propose which data should be shared and/or preserved** in accordance with the Requirements.

- In determining which data should be shared and preserved, researchers must consider **the data needed to validate research findings** as described in the Requirements, and are encouraged to consider the potential benefits of their data to their own fields of research, fields other than their own, and society at large.
- DMPs should reflect **relevant standards and community best practices** for data and metadata, and make use of **community accepted repositories** whenever practicable.
- **Costs** associated with the scope of work and resources articulated in a DMP may be included in the proposed research budget as permitted by the applicable cost principles.
- To improve the discoverability of and attribution for datasets created and used in the course of research, the Office of Science encourages the citation of publicly available datasets within the reference section of publications, and the identification of datasets with **persistent identifiers such as Digital Object Identifiers (DOIs)**. In most cases, the Office of Science can provide DOIs free of charge for data resulting from DOE-funded research through its Office of Scientific and Technical Information (OSTI) DataID Service.
- View a list of suggested elements for a DMP.

Additional Requirements and Guidance from Office of Science Program Offices

- Advanced Scientific Computing Research
- Basic Energy Sciences
- Biological and Environmental Research
- Fusion Energy Sciences
- High Energy Physics
- Nuclear Physics
- Small Business Innovation Research (SBIR) and Small Business Technology Transfer (STTR)

Information about Data Management Resources at Office of Science User Facilities

View information about the data management resources available at the Office of Science User Facilities.

Glossary

Data Preservation:

Data preservation means providing for the usability of data beyond the lifetime of the research activity that generated them.

Data Sharing:

Data sharing means making data available to people other than those who have generated them. Examples of data sharing range from bilateral communications with colleagues, to providing free, unrestricted access to the public through, for example, a web-based platform.

Digital Research Data:

The term *digital data* encompasses a wide variety of information stored in digital form including: experimental, observational, and simulation data; codes, software and algorithms; text; numeric information; images; video; audio; and associated metadata. It

also encompasses information in a variety of different forms including raw, processed, and analyzed data, published and archived data.

This statement focuses on *digital research data,* which are *research data* that can be stored digitally and accessed electronically. *Research data* are defined in regulation (2 CFR 200.315 (e), continuing the definition from 2 CFR 215 (OMB Circular A-110)) as follows:

"Research data is defined as the recorded factual material commonly accepted in the scientific community as necessary to validate research findings, but not any of the following: preliminary analyses, drafts of scientific papers, plans for future research, peer reviews, or communications with colleagues. This 'recorded' material excludes physical objects (e.g., laboratory samples). Research data also do not include:
(A) Trade secrets, commercial information, materials necessary to be held confidential by a researcher until they are published, or similar information which is protected under law; and
(B) Personnel and medical information and similar information the disclosure of which would constitute a clearly unwarranted invasion of personal privacy, such as information that could be used to identify a particular person in a research study."

Validate:

In the context of this statement, *validate* means to support, corroborate, verify, or otherwise determine the legitimacy of the research findings. Validation of research findings could be accomplished by reproducing the original experiment or analyses; comparing and contrasting the results against those of a new experiment or analyses; or by some other means.

FAQs

View Digital Data Management Frequently Asked Questions.

References …

DOE Office of Science Suggested Elements for a Data Management Plan

(webpage) [Excerpts]

The Principal Investigator should determine which data should be the subject of the DMP and, in the DMP, propose which data should be shared and/or preserved in accordance with the Requirements of the Office of Science Statement on Digital Data Management.

The following list of elements for a DMP provides suggestions regarding the data management planning process and the structure of the DMP:

- **Data Types and Sources**. A brief, high-level description of the data to be generated or used through the course of the proposed research and which of these are considered digital research data necessary to validate the research findings.
- **Content and Format**. A statement of plans for data and metadata content and format including, where applicable, a description of documentation plans, annotation of relevant software, and the rationale for the selection of appropriate standards. (Existing, accepted community standards should be used where possible. Where community standards are missing or inadequate, the DMP could propose alternate strategies that facilitate sharing, and should advise the sponsoring program of any need to develop or generalize standards.)
- **Sharing and Preservation**. A description of the plans for data sharing and preservation. This should include, when appropriate:
  - the anticipated means for sharing and the rationale for any restrictions on who may access the data and under what conditions;
  - a timeline for sharing and preservation that addresses both the minimum length of time the data will be available and any anticipated delay to data access after research findings are published;
  - any special requirements for data sharing, for example, proprietary software needed to access or interpret data, applicable policies, provisions, and licenses for re-use and re-distribution, and for the production of derivatives, including guidance for how data and data products should be cited;
  - any resources and capabilities (equipment, connections, systems, software, expertise, etc.) requested in the research proposal that are needed to meet the stated goals for sharing and preservation. (This could reference the relevant section of the associated research proposal and budget request);
  - cost/benefit considerations to support whether/where the data will be preserved after direct project funding ends and any plans for the transfer of responsibilities for sharing and preservation;
  - whether, when, or under what conditions the management responsibility for the research data will be transferred to a third party (e.g. institutional, or community repository);
  - any other future decision points regarding the management of the research data including plans to reevaluate the costs and benefits of data sharing and preservation.
- **Protection**. A statement of plans, where appropriate and necessary, to protect confidentiality, personal privacy, Personally Identifiable Information, and U.S.

national, homeland, and economic security; recognize proprietary interests, business confidential information, and intellectual property rights; and avoid significant negative impact on innovation, and U.S. competitiveness.

- **Rationale**. A discussion of the rationale or justification for the proposed data management plan including, for example, the potential impact of the data within the immediate field and in other fields, and any broader societal impact.

DOE Office of Science Statement on Digital Data Management FAQ

(webpage) [Excerpts]

Do I need to submit a Data Management Plan (DMP)? (15 question, 1 shown below)

1. Will a proposal be funded if it does not include a Data Management Plan?
   No. The Office of Science reserves the right to reject, without merit review, any proposal that does not include a DMP.

What to include in a DMP (4 questions)

16. The Office of Science Statement on Digital Data Management requires that I submit a Data Management Plan (DMP) with my research proposal. What should I include in this plan?
    A list of suggested elements for a DMP can be found here.

Sharing and Preservation (9 questions)

20. The data or data products from my research will likely be cited by me and/or others. What should I to do ensure that these are cited appropriately and that I receive proper attribution for their use in others' research?
    There are no global standards for how to cite data products. Suggestions for what information to include in a citation for your data product and how to format this information can be found here. To facilitate the citation of data products, the Office of Science encourages the use of persistent identifiers such as Digital Object Identifiers (DOIs)…. In most cases, the DOE Office of Science can provide DOIs free of charge for datasets resulting from DOE-funded research through its **Office of Scientific and Technical Information (OSTI) DataID Service.**

Evaluation (2 questions)

29. Should my proposed budget specifically address the resources and costs in implementing my DMP?
    Costs associated with the scope of work and resources articulated in a DMP may be included in the proposed research budget as permitted by the appropriate cost principles.
30. What consequences, if any, result from failure to carry out the data management plan of a funded research proposal?
    The DMP is part of the overall research proposal and, as such, it is expected that researchers will follow, to the best of their ability, the proposed research and associated data management plan. **Failure to do so will negatively influence future funding opportunities.**

DOE Federal Assistance Reporting Checklist and Instructions for RD&D Projects

(doc) [Excerpts]

B. SCIENTIFIC/TECHNICAL REPORTING

NOTE: SCIENTIFIC/TECHNICAL PRODUCTS INTENDED FOR PUBLIC RELEASE **MUST NOT CONTAIN *PROTECTED* PERSONALLY IDENTIFIABLE INFORMATION (PII).** PII is defined as any information about an individual which can be used to distinguish or trace an individual's identity. Some information that is considered to be PII is available in *public* sources such as telephone books, public websites, university listings, etc. This type of information is considered to be **Public PII** and includes, for example, first and last name, address, work telephone number, e-mail address, home telephone number, and general educational credentials. In contrast, ***Protected* PII** is defined as an individual's first name or first initial and last name in combination with any one or more of the following types of information: social security number, passport number, credit card numbers, clearances, bank numbers, biometrics, date and place of birth, mother's maiden name, criminal, medical and financial records, educational transcripts, etc., which could be mis-used if made publicly available.

Journal Article-Accepted Manuscript

Recipients are encouraged to publish their work in scholarly journals. When/if a recipient has an article accepted for publication in a peer-reviewed journal they are **required** to announce the publication to OSTI as detailed below. This Reporting Requirement will be denoted with the Frequency "O – Other" on the Checklist.

Public access to peer-reviewed scholarly publications can be achieved by following these instructions. If the Recipient has a journal article accepted for publication which contains information/data produced under the award, then the Recipient must submit an AN 241.3 for the **author's full-text version of the accepted manuscript**, as described below, **at the time the article meets the status of being "accepted" for publication**. The Federal Government's right to use the data produced under a Federal award is established in 2 CFR 200.315(d), U.S. Government's retained license to published results of federally funded research.

Content. The Recipient is to announce to DOE the final peer-reviewed accepted manuscript (AM), i.e., the version of the journal article content that has been peer reviewed and accepted for publication in a journal, **by providing a persistent link to the accepted manuscript on the recipient's publicly accessible institutional repository or submitting the full text** (see Electronic Submission Process below). The Recipient should NOT submit the journal's published version of the article, i.e., the Recipient should NOT submit a copyrighted reprint. The Recipient should not submit the content of peer reviews or a commitment to publish. The Recipient should provide only the accepted manuscript content intended to be the published article.

DOE will make no additional review of the content of an AM because the AM is a version of the journal article with the content to be published (i.e., publicly released) by the journal publisher. The Recipient is responsible for ensuring the suitability of the content for public release. **The terms and conditions of award provide that PII, proprietary,**

export control or classified information shall be protected. DOE may choose to defer providing public access until an administrative interval period has passed.

The Recipient must self-certify at the time of submission to DOE via E-Link that the content is appropriate and that it is not a copyrighted reprint, i.e., the final version of the published article. Recipients are reminded that the article is to include an acknowledgement of Federal support and a disclaimer.

Electronic Submission Process. The Journal Article-Accepted Manuscript must be announced via the **DOE Energy Link System (E-Link)** by submitting a completed **DOE Announcement Notice (AN) 241.3** (https://www.osti.gov/elink-2413).

Within the AN 241.3, **provide relevant journal information (article title, journal name, volume, issue, and any other pertinent publication information**). Also provide a **persistent link** to the repository location of the accepted manuscript. An example of an acceptable persistent link is a URL to the specific location of the Journal Article-Accepted Manuscript hosted on a public, openly accessible university research publications website. If a persistent link is not available or if the website has access restrictions (preventing public access), then the Recipient must upload the **full-text of the Accepted Manuscript** using the AN 241.3 and E-Link instructions.

Full-text of accepted manuscripts must be in **Adobe Portable Document Format (PDF)** and be **one integrated PDF file** that contains all text, tables, diagrams, photographs, schematics, graphs, and charts. Please refer to http://www.osti.gov/stip/best-practices-portable-document-format-pdf-creation for PDF document creation**.**

Scientific/Technical Conference Paper/Presentation or Proceedings

Recipients are **encouraged to announce Scientific and Technical Conference Papers/Presentations** if they are the primary means by which certain research results are disseminated or if they contain research results not already announced to DOE by the Recipient in technical reports, accepted journal articles, or other STI. This Reporting Requirement will be denoted with the Frequency "O – Other" on the Checklist. Instructions for how to announce such STI can be found below. In cases where the Recipient is required to create and submit a Conference Proceedings, the Frequency will be "F – Final."

Content. The content should include: (1) Name of conference; (2) Location of conference; (3) Date of conference; and (4) Conference sponsor. Also include an acknowledgement of Federal support and a disclaimer.

Electronic Submission Process. Scientific/technical conference papers/presentations or proceedings must be submitted via the DOE Energy Link System (E-Link) with a completed DOE Announcement Notice (AN) 241.3 (https://www.osti.gov/elink-2413).

DOE will not review conference papers or presentations prior to making publicly available via OSTI since they were already presented in a public setting during a conference. The Recipient is responsible for ensuring the suitability of the content for public release. The terms and conditions of award provide that PII, proprietary, export control or classified

information shall be protected. The Recipient must self-certify at the time of submission to DOE via E-Link that the content is appropriate for and has been publicly released.

Scientific/technical conference papers or proceedings that are textual documents must be submitted in Adobe Portable Document Format (PDF) and be one integrated PDF file that contains all text, tables, diagrams, photographs, schematics, graphs, and charts. Please refer to http://www.osti.gov/stip/best-practices-portable-document-format-pdf-creation for PDF document creation. Audiovisual formats, such as PowerPoint (PPT) or video presentations, may be submitted as a Microsoft PPT file or audiovisual file by selecting the appropriate format on the AN 241.3 for the file to be uploaded or, in the case of videos posted on a publicly available website, by providing a link to the specific video. Format options and other instructions can be found at http://www.osti.gov/stip/audiovisualsti.

Scientific/Technical Software & Manual …

DOE Office of Science Acknowledgements of Federal Support (webpage)
[Excerpts]

Peer Reviewed Articles and Technical Papers
For peer reviewed and technical papers, the following acknowledgment of support is required:
For work directly supported by DOE Office of Science Financial Assistance (i.e., Grants and Cooperative Agreements):
Acknowledgment: "This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of [insert the sponsoring SC Program Office, e.g., Basic Energy Sciences], [*Add any additional acknowledgements or information requested by the sponsoring SC Program Office*] under Award Number(s) [Enter the award number(s)]."

> example: "**This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of Basic Energy Sciences Energy Frontier Research Centers program under Award Number DESC0001234**."

For work supported by DOE Office of Science funding at a National Laboratory:
Acknowledgment: "This material is based upon work supported by the U.S. Department of Energy, Office of Science, Office of [insert the sponsoring SC Program Office, e.g., Basic Energy Sciences], [*Add any additional acknowledgements or information requested by the sponsoring SC Program Office*] [optional: under contract number XXXXXX]."

For work done at an Office of Science User Facility:
The acknowledgments should include the name of the user facility and should identify the facility as "a DOE Office of Science User Facility."

> **example**: "This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility."

DOE Energy Link System (E-Link) (webpage):
DOE Scientific and Technical Information (STI) Management System
[Excerpts]

Important News

DOE's Public Access Plan calls for submission of final, peer-reviewed accepted manuscripts. Starting October 1, 2014, the Department began to include requirements for the submission of accepted manuscripts and publication metadata in award agreements. Read more...

A tool, called the "Wizard," is now available to researchers funded by the U. S. Department of Energy (DOE). It offers a set of web-fillable screens in a simple step-by-step format to facilitate the submission of accepted manuscripts.

https://www.osti.gov/elink/forms.jsp

If you have a Financial Assistance Agreement (e.g., a grant) or contract with the Department of Energy and have been asked to provide a final scientific and technical report, the accepted manuscript of a journal article, or other STI product, these are your submission options.

For awards made on or after October 1, 2014, submission to DOE of accepted manuscripts is being required (see DOE STIP Public Access FAQs). Even if not specifically required under your award, **you may submit an accepted manuscript** of a journal article for work published as a result of DOE funding via **AN 241.3** or **the new AM Wizard** for inclusion in DOE PAGES^{Beta}.

IDA Notes:

1. *Announcement Notice (AN) 241.3* ***Instructions for publications*** *are available at https://www.osti.gov/elink/F2413instructC.jsp*
2. *Traditional AN 241.3 Submission Mechanism, which can be used for accepted manuscripts as well as for technical reports and conference papers, is available at https://www.osti.gov/elink/241-3.jsp*
3. *New Accepted Manuscript (AM) Submission Wizard is available at https://www.osti.gov/elink/241-3_Wizard.jsp*
4. *AN 241.6* ***Instructions for datasets*** *are available at https://www.osti.gov/elink/F2416instruct.jsp*
5. *See following page for excerpts from AN 241.3 Instructions on submitting journal article accepted manuscripts and AN 241.6 Instructions on submitting datasets*

Instructions for Journal Article: Accepted Manuscript (webpage) [Excerpts]

Part I: STI PRODUCT DESCRIPTION

DOE Award/Contract Number …

Recipient/Contractor Organization …

STI Product Type

Select "**Journal Article-Accepted Manuscript**" from the choice of four product types. Note: Published versions of journal articles should not be submitted; **you must provide the final peer-reviewed accepted manuscript and certify this is true before submission.**

Additional subfields will appear on the AN 241.3 based on selection of Journal Article: Accepted Manuscript as the STI product type.

- **Journal name** - …
- **Digital Object Identifier (DOI)** - Digital Object Identifier (DOI) - Provide if available….
- Volume
- Issue
- Serial identifier ? such as ISSN
- Page range
- Journal-associated conference information

STI Product Title …

Digital Object Identifier (DOI)

A digital object identifier (DOI) is a **unique persistent identifier** that references a digital object….

For journal articles, publishers will assign a DOI, usually when the article is accepted for publication. **If you have an assigned DOI for your submission, you will be able to auto-populate some of the required metadata, including title, author, and publication date.** DOIs in cited references promote reference linking practices and the reuse of information.

For data sets that are part of a larger repository, a DOI may be assigned by the repository as part of its referencing schema.

For data that are not assigned a DOI by a repository or other authority, **authors may request a DOI from OSTI through the DataID service**….

Publication/Issue Date …

Author(s)

After entering last name, first name, and middle name or initial in the specified blanks, you may also enter the primary author/creator's email address, **his or her ORCID number**, and his or her affiliation in the remaining three fields of the author section….

Contributor Organization(s)

The name of a Research/Project Collaboration, if applicable, should be entered in this field, not in the author field. Contributor organizations are not authors' affiliations, the originating research organization, nor the funding/sponsor organization. They are, instead, any company, institution, or organization to which the submitter wishes to

provide recognition and which clearly does not fit into any of the other organization fields. Examples of possible contributor organizations that a submitter may want to list (in addition to listing a collaboration name) include:

- An external organization that provided significant review of the research product.
- An organization that provided site management but was not directly involved in the research/experiment itself.
- An organization that collected data to provide to the originating research organization.
- A data center or repository that is not listed as the originating research organization.

This is a free text field and can hold multiple organizations names. Each different organization should be separated by a semi-colon followed by a space.

Report/Product Number

This field is **optional for AMs** but can be used to provide a unique identifier that research organization or author placed on the AM itself.

Sponsoring DOE Program Office

Select the name of the **specific DOE program office** that provided the project funding, from the list provided (e.g., USDOE Office of Science (SC) Basic Energy Sciences (BES)). For projects funded by more than one program office, indicate all sources of the DOE funding in descending order of dollar amount of funding appropriated. You may also input additional funding agencies in the textbox. Separate multiple program offices or funding agencies with a semicolon and a space. **If you do not know the sponsoring DOE program office, select "USDOE".**

Description/Abstract

You must provide a clear, concise, and publicly releasable English language executive summary of the information contained in the accepted manuscript, written in terms understandable by an educated layperson. The length should be no more than 5,000 characters. To fill this field, you can cut and paste from any word processing file.

Subject Categories

Select one or more categories from the list provided. Select them in order of relevance to the project described in the product. You can review the list of subject category descriptions on the Authorities page.

Keywords

These are words or phrases that describe the project as summarized in the article. Keywords aide in the online search and discovery of information about the project. More than one keyword may be entered; separate multiple keywords with a semicolon and a space.

Related Document Information

Please provide **citations for supplementary datasets or other items** published with the journal article you are submitting. You may also provide a citation for any other directly-supporting item for technical reports, conference literature, etc. **Or, see the Related Identifiers/DOIs field below**.

Related Identifiers/DOIs

Referencing **other documents, other datasets, or software applications** that relate to the document you are submitting allows the user of your information to follow these vital links and to better understand the scope of your research. **You may add DOIs to cross-reference these other items**. If you do not have a DOI for one of the items you wish to relate to your document, please put the citation information for that item in the Related Document field. Note that you may find assigned DOIs for DOE technical reports in SciTech Connect.

This Related Identifier/DOIs field also allows you, for each DOI you reference, to choose a **controlled vocabulary term that "explains" the type of relationship between the DOI you enter and the document you are submitting.** For additional information about the [How Related] controlled vocabulary see the DataCite Metadata Schema 3.1 Documentation in Table 9 on pages 31 - 34.

Intellectual Property/Distribution Limitations

Since by definition an accepted manuscript will appear in a publicly available scholarly journal, the appropriate category to select should be **"Unlimited Announcement"** - The unrestricted, unlimited distribution of the product will be made publicly available. The government assumes no liability for disclosure of such data. "UNL" is therefore the only value that can be selected. **You will also be required to certify that only appropriate information is contained in the product** (see "Certifications").

Recipient/Contractor Point of Contact …

Part II: STI PRODUCT MEDIA/FORMAT AND TRANSMISSION

Media/Format Information

For accepted manuscripts, the medium will be an "Electronic Document".

Certifications

For Journal Article-Accepted Manuscript, there are **three certifications** that you are required to indicate.

1. The award recipient is responsible for ensuring that content of the accepted manuscript is **suitable for public release**. Therefore, you must **self-certify** that it does not contain limited rights data (proprietary data), classified information, protected PII, information subject to export control classification, or other information not subject to release.
2. You must also **certify that the accepted manuscript being provided is not the published version of the article** (i.e., that it is not a copyrighted reprint of the published article in the journal's format) and that you understand that DOE Policy calls for the manuscript to be submitted after peer-review and acceptance, but before final formatting by the publisher.
3. You must also certify that, in any Publication and Copyright Agreement made with the journal publisher, **rights have been retained to deposit the accepted manuscript with DOE**.

Transmission Information

For Accepted Manuscripts, you have **two options** as described below. Indicate in the appropriate fields which way you are providing the accepted manuscript.

1. Provide a **persistent link** (e.g., a URL or PURL) that is **publicly accessible for the full text of the accepted manuscript.** (NOTE: If the general public cannot access the manuscript at this location, then you must use option 2.)
2. **Upload the full text of the accepted manuscript**.

Provide additional information indicated for selected option. For option 1, provide the URL or PURL that will link directly to the specific document. Note: it must be accessible to a member of the general public. For option 2, upload the file by selecting the "Browse" button to locate the file on your computer. **Valid file formats are: Adobe Portable Document Format (.PDF) or MS Word (.DOC) only**. Before you upload the document, you must certify (in "Certifications" above) that the document being uploaded has been reviewed for, and does not contain, any information not subject to release, such as Protected PII.

IDA Notes [DOE public access policy requires PDF format]

1. *But the DOE Federal Assistance Reporting Checklist and Instructions for RD&D Projects has a stricter requirement: "**Full-text of accepted manuscripts must be in Adobe Portable Document Format (PDF) and be one integrated PDF file** that contains all text, tables, diagrams, photographs, schematics, graphs, and charts."*
2. *The DOE Scientific and Technical Information Program (STPI) Public Access FAQ confirms the requirement for PDF format here:.*
3. *The DOE Scientific and Technical Information Program (STPI) Public Access FAQ clarifies the PDF requirement here: "**Accepted manuscripts should be submitted as a PDF** (portable document format) that is not encrypted, password protected, or corrupted. It is **preferred that the PDF is compliant with one of four standards and with extractable text**. The standards are PDF/A-1a, PDF/A-2a, PDF/A-3a, or PDF/UA….*

Instructions for Announcement of U.S. DOE Publicly Available Scientific Research Datasets (webpage) [Excerpts]

AN 241.6 Metadata Details and Requirements
Part I: STI Product Description

Dataset Type

Note that "Dataset" is automatically defaulted into the record as the product type whenever you use the AN 241.6. Dataset Type, however, allows you to be more specific about the dataset. Select one choice from the drop-down list that best describes the dataset's main or most important content….

Dataset Title

Enter the title exactly as given on the product itself, including part, version, and similar information.

Author(s) [Creator(s)/Principal Investigator(s)]

Enter the name of each person primarily responsible for the dataset, i.e. the person(s) who should be credited with the content of the dataset. After entering first, middle, and last name of each person, you may also enter the person's email address, **ORCID**, and affiliation. Email, ORCID number, and affiliation are optional parts of this required author block….

Related Resource

Enter citation information (title, author, identifier) for key publication, technical report, or software that the dataset being submitted directly supports. This is very important if your dataset is being provided to a journal publisher as supplementary material. If you have several items related to this dataset, you may use the Related Identifiers/DOIs below.

Related Identifiers/DOIs

Referencing other documents, other datasets, or software applications that relate to the dataset you are submitting allows the user of your information to follow these vital links and to better understand the scope of your research. **You may add DOIs to cross-reference these other items**…

Dataset Product Number(s)

An identifying number that has been assigned to the dataset by either the originating/submitting organization or by the organization currently hosting the data….

DOE Contract/Award Number(s)

Enter the DOE contract number under which the work was funded….

Originating Research Organization

Select the name of the organization that performed the research or issued the dataset from the drop-down list. More than one organization may be selected. You may also type in the name of the Originating Research Organization, if you do not see it in the picklist. Select or list the primary organization first and separate multiple entries with a semicolon and a space. *(See also the Contributor Organization(s) field.)*

Sponsoring Organization(s)

Select the DOE Program Office and sub-Program Office (e.g. Office of Science (SC), Office of Basic Energy Sciences (BES, DOE office of Nuclear Energy (NE), Fuel Cycle Research and Development Program) that funded the work described in the STI Product. For projects funded by more than one Program Office, select each source of the DOE funding in descending order of dollar amount of funding. The names of funding offices for work for non-DOE organizations may be typed into this field; separate multiple program offices with a semicolon and a space.

Publication/Issue Date

Provide the **date when the information product was published or issued**, either in format mm/dd/yyyy (example: 04/17/2011), or in format yyyy (example: 1995). If you use the yyyy format, you may also select a Time Period from the drop-down list, if known.

Language….

Country of Publication….

Subject Categories

Select **one or more categories from the drop-down list**. Select the primary one first. A list of subject categories and their descriptions is available at www.osti.gov/stip/subjectcategories. If no subject category is provided, OSTI may generate the appropriate categories.

Keywords….

Description/Abstract

Provide a clear, concise summary of the content of the dataset, as well as specialized parameters that describe the data. Specialized parameters may include a date range during which information was taken (such as May, 01 2002 - December 31, 2002), geographic information (such as a specific state, region, country, latitude and longitude, etc.), information such as well depth ranges, temperature ranges, etc. The abstract length should be no more than 5,000 characters.

Part II. Dataset Location/Technical Specifications

URL where dataset is posted for access

Provide the URL that leads to an HTML "landing page" (information page) that provides context and usage information for the dataset. The landing page must include a direct link to the dataset and/or to its component files….

Digital Object Identifier (if already assigned)

Provide the DOI only if an organization other than OSTI has assigned it. **If the dataset does not already have a DOI, one will be assigned to it by the DOE Data ID Service**….

DOI Infix

A DOI will be assigned to your dataset by OSTI. If you wish to include an alphanumeric string, enter it….

Dataset File Extension
Please provide the file extension of the dataset. The content of the dataset will not be indexed by OSTI but knowing the type of file posted will be important to the users that search our databases. Some common file extensions are .txt, .csv, .ps, etc.

Software needed to utilize dataset (if applicable)
Specialized software tools are often developed to allow a user to manipulate data in various ways. If these tools are available for the user but do not have to be used with the data, they do not need to be listed. However, **if there is a piece of software without which a user cannot open, see, or use the dataset, that software should be noted in this field**.

Dataset Size
Indicate **how many individual data files** are included in the dataset being announced, or if the dataset consists primarily of images, note the approximate **number of images**. You may also indicate **size in megabytes**, and you may indicate whether the dataset is complete or will continue to have files added to it.

Part III. Contact Information….

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE (DD-MM-YY)	2. REPORT TYPE	3. DATES COVERED (From – To)
00-01-2017	Final	

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Public Access to Federally Funded Research: Policy, Evolving Ecosystem, and Agency Efforts	HQ0034-14-D-0001
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBERS

6. AUTHOR(S)	5d. PROJECT NUMBER
Karen D. Gordon, Jonathan R. Agre, Robert F. Leheny	ET-2-2954.30
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESSES	8. PERFORMING ORGANIZATION REPORT NUMBER
Institute for Defense Analyses 4850 Mark Center Drive Alexandria, VA 22311-1882	D-8320 H 2017-000015

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR'S / MONITOR'S ACRONYM
Manfai Fong, Executive Officer Office of the Director of National Intelligence Intelligence Advanced Research Projects Activity Washington, DC 20511	ODNI/IARPA
	11. SPONSOR'S / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT.

Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

Project Leader: Karen D. Gordon

14. ABSTRACT

This document, prepared in support of an Intelligence Advanced Research Projects Activity (IARPA) project, captures preliminary research on public access to the results of Federally funded research. It is a compilation of six briefings and memos: (1) Public Access to Scientific Publications and Data: State of the Practice; (2) Responding to OSTP Public Access Memo: Processes, Sample Agency Implementations, and Preliminary Analysis of Rights-in-Data; (3) Agency Public Access Plans: Key Publication and Data Provisions; (4) Sample Data Retention Periods; (5) National Science Foundation Guidance on Public Access; and (6) Department of Energy Guidance on Public Access. A companion document contains a draft plan for IARPA implementation of public access.

15. SUBJECT TERMS

public access, public access plan, data management plan, federally funded research, research data, retention period, copyright, public access ecosystem

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Manfai Fong, Executive Officer
Unclassified	Unclassified	Unclassified	Unlimited	180	19b. TELEPHONE NUMBER (Include Area Code) 301-851-7586