

**Preparing for new Federal guidelines on data publishing:
Evolving Federal initiatives
and the NSF Public Access Repository**

SDSC SAN DIEGO
SUPERCOMPUTER CENTER

August 8, 2022

UC San Diego

Overview

- EAGER project on Public Access and data records (and code)
- Public Access policy and systems in US, and where it's headed
- Discussion – how do we better link science production with low-barrier public access reporting?
 - What are the biggest gaps in practices, systems and services that would deter or ease records submission to Federal Public Access Repositories?
 - AI-readiness (data / community practices)

EAGER: Community Readiness for Increasing Public Access to Data

- Initial work -
 - Gathering expert perspectives from leaders in scholarly communication and data infrastructure ;
 - Semi-structured interviews with domain researchers
 - Focus on Earth & Ag sciences (e.g. Soils; agronomy; other Earth Sci) and Data Science
 - particularly interest in eco / soils using AI / ML methods, or producing data for reuse
 - Am actively recruiting for interviews across the Soil sciences
 - Engagement with a range of research communities on the use of the NSF Public Access Repository.


This work anticipates upcoming policy changes from NSF on the deposit of metadata for datasets resulting from funding (into the PAR).

US Public Access Policy

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

February 22, 2013

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: John P. Holdren 
Director

SUBJECT: Increasing Access to the Results of Federally Funded Scientific Research

1. Policy Principles

Federal Agencies Public Access Policies

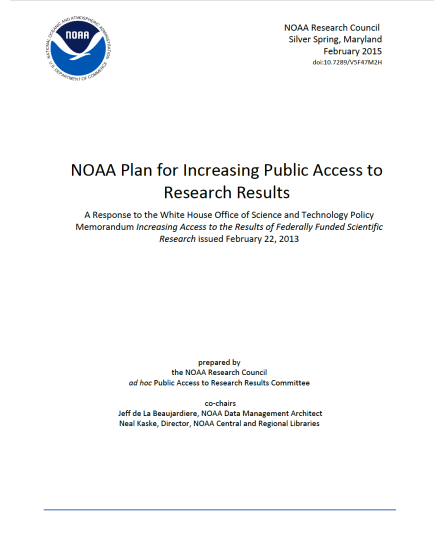
Implementation Plan to Increase Public Access to Results of USDA-funded Scientific Research

November 7, 2014

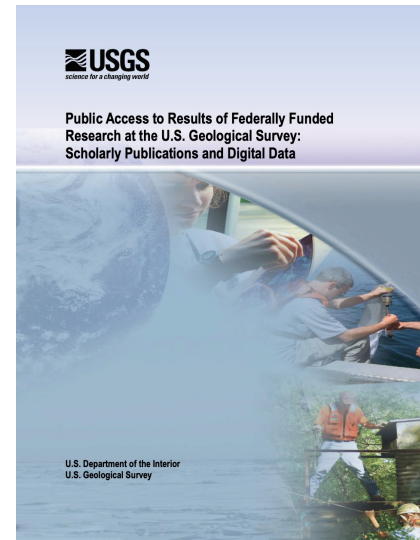


United States Department of Agriculture

2014



2015



2016

NSF'S PUBLIC ACCESS PLAN:

Today's Data, Tomorrow's Discoveries

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

National Science Foundation

March 18, 2015



NSF 15-52

2015

NSF Public Access Repository – Brief History

- Launched in 2015, policy applied starting in 2016
- Roles and responsibilities - (NSF Special Report on Public Access)
 - NSF requires 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.

https://www.nsf.gov/news/special_reports/public_access/



PUBLIC ACCESS To Results of NSF-Funded Research



Public Access Home

[The Plan](#)

[The Executive Summary](#)

[Press Release](#)

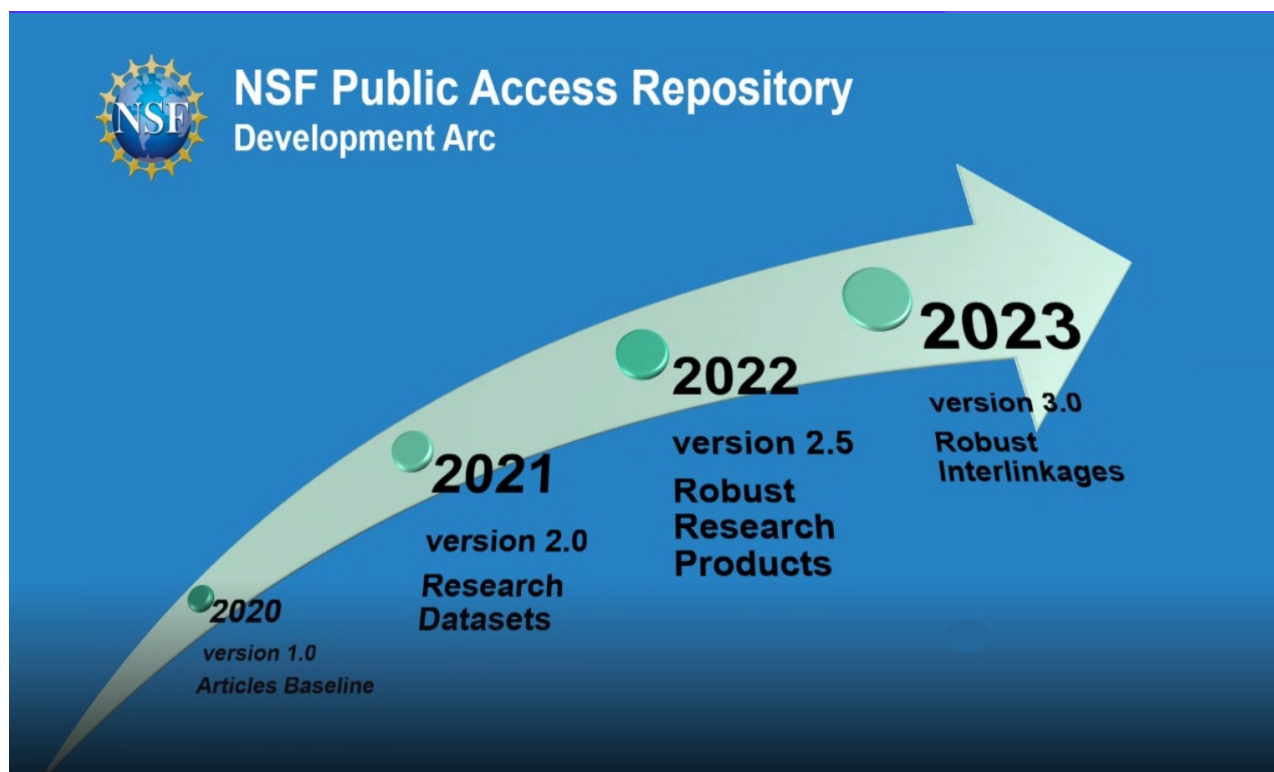
[Frequently Asked Questions \(FAQs\)](#)

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.

Where it's headed

Martin Halbert, NSF

- Opportunity to submit records for data: Voluntary for now
- Principal investigators will be responsible for separately publishing data sets in preferred repositories
- Once a repository has issued a PID for the data set, the PI will be responsible for the PID being entered into the PAR 2.0 system through a submission workflow separate from that of articles



NSF PAR Data Pilot

As of FY22 this new feature does not change reporting requirements.

PIs and co-PIs can add datasets to the NSF-PAR by meeting the following conditions:

Key Requirement: You must have previously deposited your dataset into a repository which associates a DOI with your dataset.

- ✓ Select dataset as your research product
- ✓ Enter your DOI
- ✓ Verify your dataset's metadata information
- ✓ Add your award id to your dataset
- ✓ Acknowledge the NSF Grant Conditions
- ✓ Submit your dataset

- Your dataset can exist in various file formats and in various file sizes in a repository of your choice.
- The NSF-PAR will only store the metadata information from the digital object identifier (DOI) associated to your dataset. Your dataset will not be replicated in NSF-PAR.
- Your repository provider must either register (i.e., mint) a new DOI or upload a pre-existing DOI and associated metadata with [DataCite.org](https://datacite.org).
- Only the metadata that has been added to DataCite.org will be stored in the NSF-PAR and made available to the Public Access search application.

NSF Public Access Repository (NSF-PAR)

A partnership with the Department of Energy,
Office of Scientific and Technical Information



Add Research Products

Manage Research Products

Contacts

FAQs

Welcome, Melissa Cragin

Add Research Products

*** Required**

*** Add a research product into NSF-PAR:**

- Publication
- Dataset

Next >

[Website Policies](#) | [Budget and Performance](#) | [Inspector General](#) | [Privacy](#) | [FOIA](#) | [No FEAR Act](#) | [USA.gov](#) | [Accessibility](#) | [Plain Language](#) | [Contact](#) | [Help](#)

The National Science Foundation, 2415 Eisenhower Avenue, Alexandria, Virginia 22314, USA Tel: (703) 292-5111, FIRS: (800) 877-8339 | TDD: (800) 281-8749

Manage Research Products

Use NSF-PAR to deposit your publications by entering a Digital Object Identifier (DOI) and associating them to specific NSF awards. You can also search for and manage publications that have previously been entered for your awards.

Research Products Previously Submitted

Type:	Award ID:	Title:	Author/Creator:	DOI Number:	PAR ID:
<input type="text" value="--Select Type--"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>	<input type="text"/>

[Filter](#)[Clear](#)

Show

Legend

Publications  Datasets 

Showing 0 to 0 of 0

<< < Prev 1 Next > >>

Type	Award ID(s)	Title	Author(s)/Creator(s)	DOI Number	PAR ID	Actions
------	-------------	-------	----------------------	------------	--------	---------

No results found to match your query. Please refine your search and try again.

Show

Showing 0 to 0 of 0

<< < Prev 1 Next > >>



How to Remove Datasets from the NSF-PAR

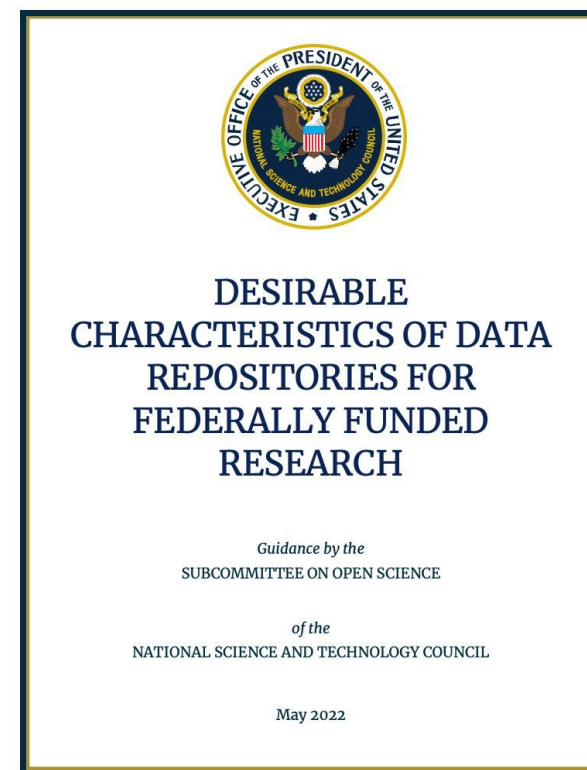
Principal Investigators (PIs) and co-PIs can remove added datasets from the NSF Public Access Repository (NSF-PAR) without assistance from the NSF Help Desk.

- Only the PI and co-PIs that initially added or associated an NSF award to a dataset can disassociate the NSF award from the added dataset in NSF-PAR.
- Award IDs that are removed will immediately stop displaying on the Manage Research Products page and will not be available in the publicly-facing [NSF-PAR search](#) results.
- If the dataset has only one associated award ID, that dataset is immediately removed from the Manage Research Products page and the NSF-PAR search results.
- Removing a dataset that has been added in the NSF-PAR does not remove it automatically from an in-progress project report that you have added the dataset to manually.

Steps to Remove a Dataset from the NSF-PAR

Desirable Characteristics of Data Repositories for Federally-funded Research

- Subcommittee on Open Science
 - National Science and Technology Council
 - Organizational Infrastructure
 - Policies that indicate oversight and sustainability (of the organization)
 - Digital Object Management
 - Research Data management and curation
 - Technology
 - Tech refresh and security
- ** Add'l considerations for Repos storing human data
- Fidelity to consent; security; limited use compliant; download control; request review; plan for breach; accountability



<https://www.whitehouse.gov/wp-content/uploads/2022/05/05-2022-Desirable-Characteristics-of-Data-Repositories.pdf>

Questions / discussion

- What kinds of activities and community-level efforts are working to regularize data publication?
 - What are the biggest challenges at the community level?
 - What are the pain points for Data Managers, and repositories?
- Are the roles or use of Data Management Plans changing?
- What kinds of information on the federal requirements and policy changes would be helpful to have for your community?

Thank you

mcragin@sdsc.edu

This work supported in part by NSF award # 2032705.