

SDSC SAN DIEGO SUPERCOMPUTER CENTER

Overview

- EAGER project on Public Access and data records (and code)
- US Public Access policy and where it's headed
- NSF Public Access Repository
- 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?
 - Getting beyond the F and the A (in FAIR)...
 - Al-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 projects using AI / ML methods, or producing data for reuse
 - Am actively recruiting for interviews interest in early career, and "Rapid" awards
 - 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

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

1. Policy Principles

SDSC SAN DIEGO

NSF 15-52 Plan

- NSF Public Access Initiative (PAI) was created by the NSF 15-52 Plan as the agency response to the 2013 U.S. Holdren Memorandum
- The Public Access Initiative is an NSF agency wide commitment, implemented through
 - policy,
 - procedures
 - infrastructure
- NSF holds a uniquely broad and influential role among organizational leaders of the commitment to public access in the United States





Adapted from M. Halbert, NSF - Used with permission

NSF Public Access Repository Functions and Characteristics – 1.0

PAR 1.0 is focused on Peer-Reviewed Articles

- Enables researchers to enter metadata for peer-reviewed articles or auto-populate by means of Digital Object Identifiers (DOI)
- Metadata recorded in PAR is also transmitted and synchronized with Award Search database
- The metadata for articles can be searched and displayed
- Researchers may deposit (or retrieve) public access articles as PDF/A files or point to download locations through DOI
- Note that PAR is not a single system, but rather an infrastructure of various interacting software systems in several different parts of the NSF infrastructure, also making use of modified modules from the DOE Office of Scientific and Technical Information (OSTI) system

→ C û û â	https://par.nsf.gov/search/term:COVID	◎ … ♡☆ IN □ ◎			
NSF PAR Home Contact	FAQ				
NSF Public Acces A partnership with the Departme Office of Scientific and Technica	s Repository (NSF-PAR)	National Science Foundation WHERE DISCOVERIES REGIN			
Explore scholarly pu	blications in the NSF Public Access Repos	sitory			
COVID		× 🔎 Find			
		+ Advanced Search			
Home / Search Results / Pag	pe 1 of 6				
Search for: COVII	Sort	by Relevance + « Prev + Next »			
Total Results eo	Note: When clicking on a Digital Object Identifier (DOI) number.	nou will be taken to an external site			
Publicly.Available.Eull.Text 1	maintained by the publisher. Some full text articles may not yet be the embargo (administrative interval). •	available without a charge during			
Citation Only 59	Some links on this page may take you to non-federal websites. Their policies may differ from this site.				
Filtered Results	 Telemedicine and Healthcare Disparities: A cohort study in a large healthcare system in New York City during COVID-19 				
Filter Results	doi: 10.1093/jamia/ocaa217 🗭 Chunzer Rumi - Zhao Yuan - Chen - 8 - Levenece Ketherine - Testa Paul & - Nov Oderl - Menn Davin M				
Filter by Author	(August 2020, Journal of the American Medical Informatics Association)				
Save Results	Abstract Objective Through the coronavirus disease 2019 (COVID-1)	9) pandemic, telemedicine and treatment. Bacial and attnic			
Excel CSV	ith respect to risk of infection and				
XIII.	in-hospital outcomes once admitted, and here we assess disparities in those who access healthcare				
	via telemedicine for COVID-19. Materials and Methods Electronic he	with record data of patients at			
Have feedback or	descriptive and multilevel regression analyses with respect to visit by	pe (telemedicine or more »			
suggestions for a way to improve these	Free, publicly-accessible full text available August 31, 2021				
results? ⊠ <u>Let us know !</u>	 Short-term forecasts and long-term mitigation evaluations for the COVID doi: https://doi.org/10.1016/j.idm.2020.08.001 2* 	9-19 epidemic in Hubel Province, China			
	Yang, Q.; Yi, C.; Vajdi, A.; Cohnstaedt, L.; Wu, H.; Guo, X.; Scoglio, C (A modelling)	lugust 2020, Infectious disease			
	As an emerging infectious disease, the 2019 coronavirus disease (Cl	OVID-19) has developed into a			





SDSC SAN DIEGO SUPERCOMPUTER CENTER

M. Halbert, NSF - Used with permission

			O Select journal na	Magil Sen		
ISSN: 😧				ine / issiv		
Volume:		ssue:	Ρ	age Range or	eLocation-ID	:
Enter full date	e or year:				(e.g., 28 - 32)	
● *Publicatio	n / Issue Full Date:	-OR-	Publication (Y)	/ Issue Year: YYY)		Time Period: Select One V
*Authors: 🕑	Separate multiple authors with a se Last name, First intial or name (mid	emicolon ; Idle name optional) e.g., Smith, J	ohn A; Anderson, S.			
		colon -		11		
Editors: 🕑	Last name, First intial or name (midd	e name optional) e.g., Smith, Joh	n A; Anderson, S.			
Editors: 🕑	Separate multiple editors with a sem Last name, First intial or name (midd	le name optional) e.g., Smith, Joh	n A; Anderson, S.			
Editors: 🕑	Separate multiple editors with a sem Last name, First intial or name (midd Abstract:	le name optional) e.g., Smith, Joh	in A; Anderson, S.			

SF Public Access Repo rtnership with the Department of Energy te of Scientific and Technical Information	sitory (NSF-PAR)		Research. C
Manage Publications	Contact	FAQ	Welcome, Cra
eposit Publication			
1. Retrieve Publication Info	2. Deposit Final Accepted Version	3. Select Award & Acknowledge	4. Review
nter the DOI number and click 'Submit' to Required Fields Enter Digital Object Identifier (DOI) Nu	pretrieve the publication information from the pu	blisher. Once you have reviewed the informat	ion, click 'Next' to proceed.
	Clear O Where do I find my DOI r	umber?	
ave you previously deposited your pu O Yes No Submit	blication in another federal repository?	More Information	
			Previous Next >





What's coming for Data?



SDSC SAN DIEGO SUPERCOMPUTER CENTER

US Data Policy, Open Science







The National Academies of SCIENCES • ENGINEERING • MEDICINE

CONSENSUS STUDY REPORT

OPEN SCIENCE BY DESIGN

Realizing a Vision for 21st Century Research





NSF Public Access Repository Development Arc



2021

version 2.0

Research Datasets version 2.5

Robust Research Products 2023

version 3.0 Robust Interlinkages

M. Halbert, NSF - Used with permission

2020

version 1.0 Articles Baseline



The NSF Public Access Initiative (PAI) strives to make the outputs of scientific research funded by the National Science Foundation publicly available to the greatest extent and with the fewest constraints possible and consistent with law.

- NSF Public Access Repository (NSF PAR)
- See a list of what the PAI has funded
- > FAQ for Public Access

SDSC SAN DIEGO SUPERCOMPUTER CENTER

M. Halbert, NSF - Used with permission

Questions / discussion

NSF PAR

- What kinds of information (or services) are needed to help researchers
 - deposit data in "preferred" repositories?
 - Have a better sense of the Public Access reporting process (submit records of these datasets)?

Adapting practices for publishing and Open Science

- Where (process) are the biggest barriers to researchers generating metadata records for datasets / code?
- Are publishers starting to monetize curation and stewardship?
- Where are we with AI-readiness?
 - machine-readable, AI-ready data this is only part of the challenge
- What is the most pressing challenge that would benefit from closer collaboration with international efforts?







mcragin@sdsc.edu

This work supported in part by NSF award 2032705.

SDSC SAN DIEGO SUPERCOMPUTER CENTER

