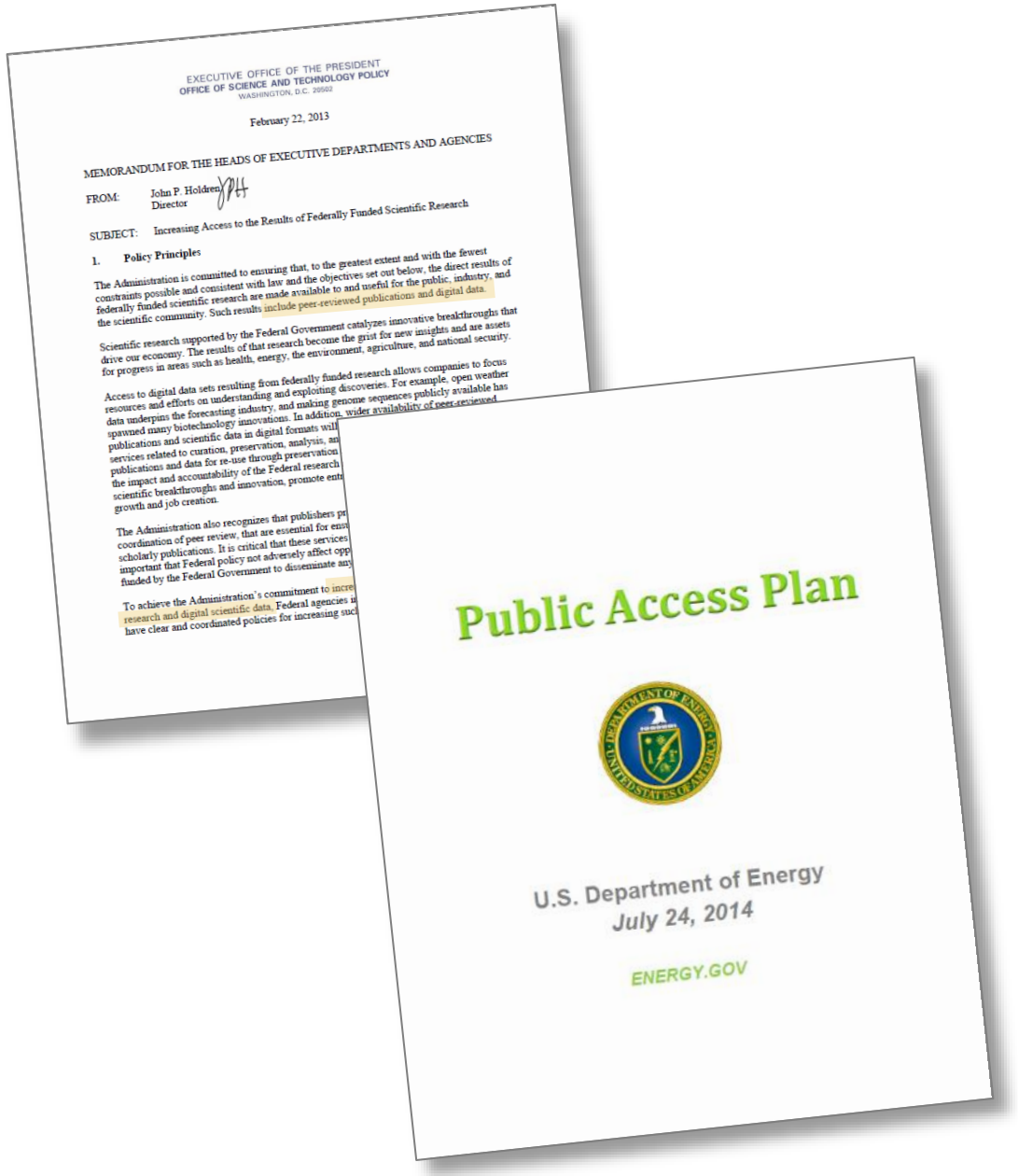
DOE Public Access and Data Management

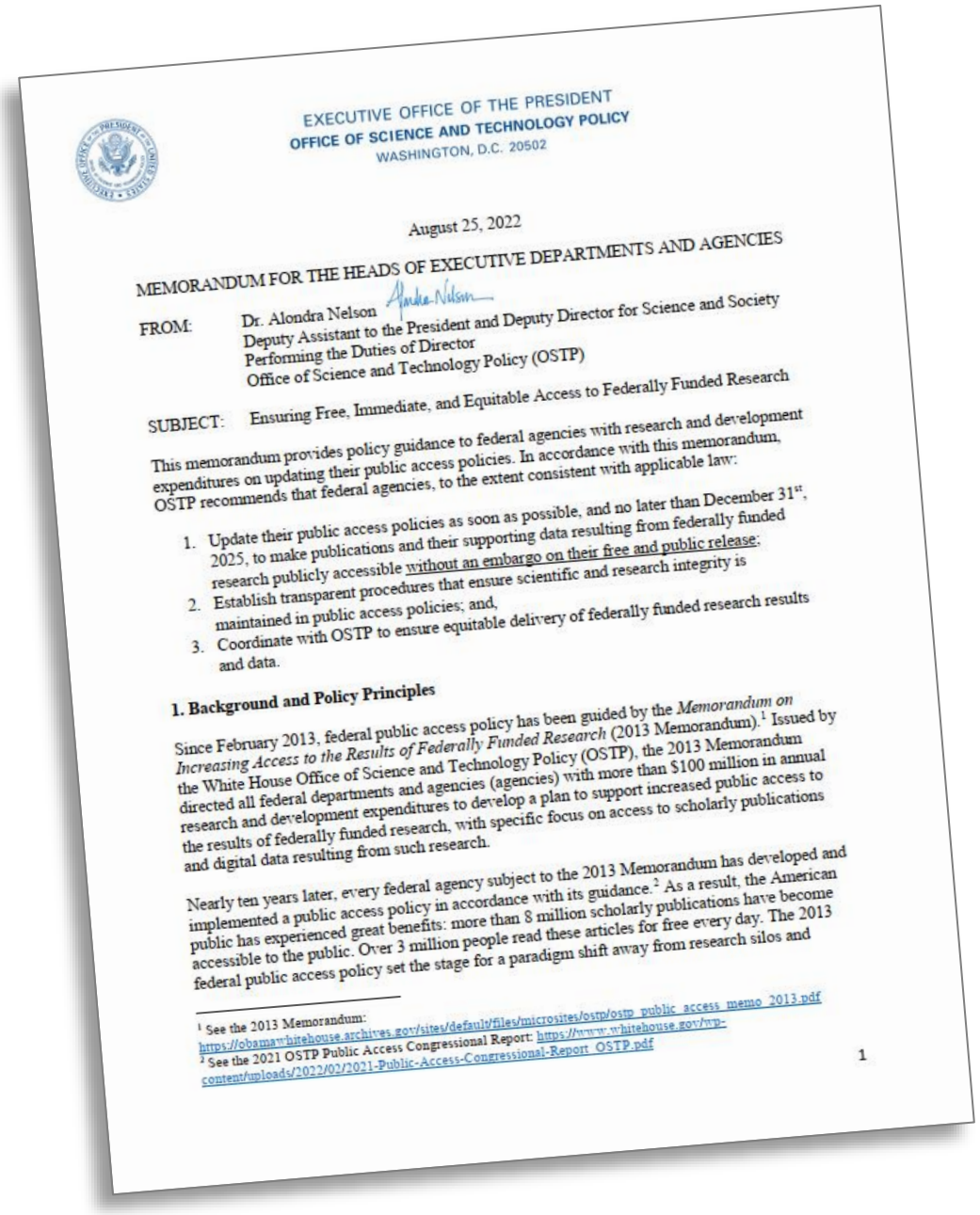
BESAC Meeting
September 2024

Michael Cooke, PhD
Senior Technical Advisor
Office of the Deputy Director for Science Programs



2013 – 2014 Public Access

- 2013 Public Access Memo issued by the White House Office of Science and Technology Policy (OSTP)
  - Results of federally funded scientific research, including peer-reviewed publications and digital research data, should be made publicly available
  - Allowed 1-year embargo of peer-reviewed articles
- 2014 DOE Public Access Plan
  - Requires author submission of accepted manuscript for peer-reviewed publications to DOE within 12 months of publication
  - Government purpose license used to share manuscripts through DOE PAGES® with voluntary participation of publishers
  - Data Management Plan (DMP) requirements for public sharing of digital research data



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

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM: Dr. Alondra Nelson
Deputy Assistant to the President and Deputy Director for Science and Society
Performing the Duties of Director
Office of Science and Technology Policy (OSTP)

SUBJECT: Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

This memorandum provides policy guidance to federal agencies with research and development expenditures on updating their public access policies. In accordance with this memorandum, OSTP recommends that federal agencies, to the extent consistent with applicable law:

1. Update their public access policies as soon as possible, and no later than December 31st, 2025, to make publications and their supporting data resulting from federally funded research publicly accessible without an embargo on their free and public release;
2. Establish transparent procedures that ensure scientific and research integrity is maintained in public access policies; and,
3. Coordinate with OSTP to ensure equitable delivery of federally funded research results and data.

1. Background and Policy Principles

Since February 2013, federal public access policy has been guided by the *Memorandum on Increasing Access to the Results of Federally Funded Research* (2013 Memorandum).[1] Issued by the White House Office of Science and Technology Policy (OSTP), the 2013 Memorandum directed all federal departments and agencies (agencies) with more than $100 million in annual research and development expenditures to develop a plan to support increased public access to the results of federally funded research, with specific focus on access to scholarly publications and digital data resulting from such research.

Nearly ten years later, every federal agency subject to the 2013 Memorandum has developed and implemented a public access policy in accordance with its guidance.[2] As a result, the American public has experienced great benefits: more than 8 million scholarly publications have become accessible to the public. Over 3 million people read these articles for free every day. The 2013 federal public access policy set the stage for a paradigm shift away from research silos and

[1] See the 2013 Memorandum: https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/ostp_public_access_memo_2013.pdf
[2] See the 2021 OSTP Public Access Congressional Report: https://www.whitehouse.gov/wp-content/uploads/2022/02/2021-Public-Access-Congressional-Report_OSTP.pdf

1

August 2022 Nelson Memo

All federal science agencies, including DOE, required to develop new Public Access Plans

- Significant updates include:
  - Emphasis on use/re-use; machine readability; equitable access
  - Immediate public access to publications, removing 12-month embargo
  - Immediate public access to data displayed in or underlying publications
  - Expanded use of persistent identifiers (PIDs)

Development of DOE's New Public Access Plan – Released June 2023

Intra-Agency Coordination

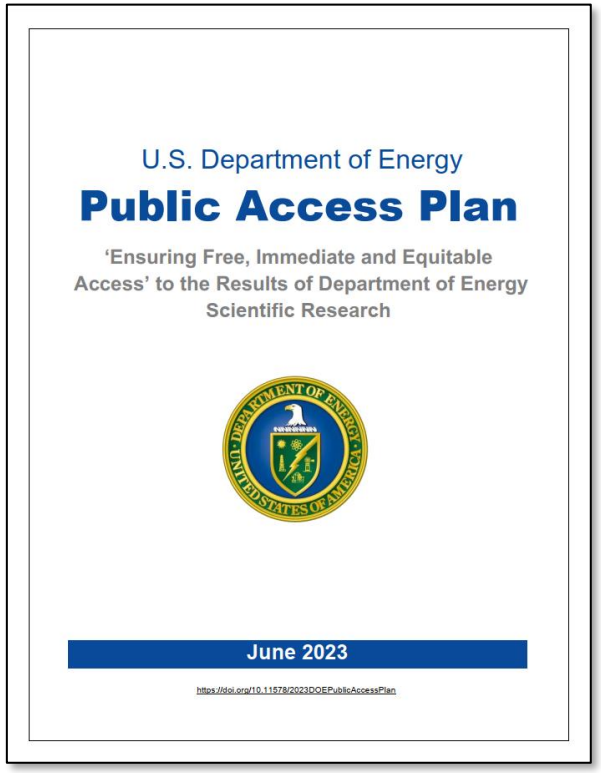
- DOE-wide participation (EERE, FE, NE, OE, ARPA-E, MA, GC, NNSA)
- SC-led author team
- Coordinated with DOE and SC Working Groups on Digital Data
- DOE researcher community input through Labs' STI managers

Interagency Coordination

- OSTP Subcommittee on Open Science (SOS); SC co-chairs three SOS working groups
- Persistent Identifier Services partners from 12 agencies

External Community Engagement

- Professional societies
- Publishers
- Libraries

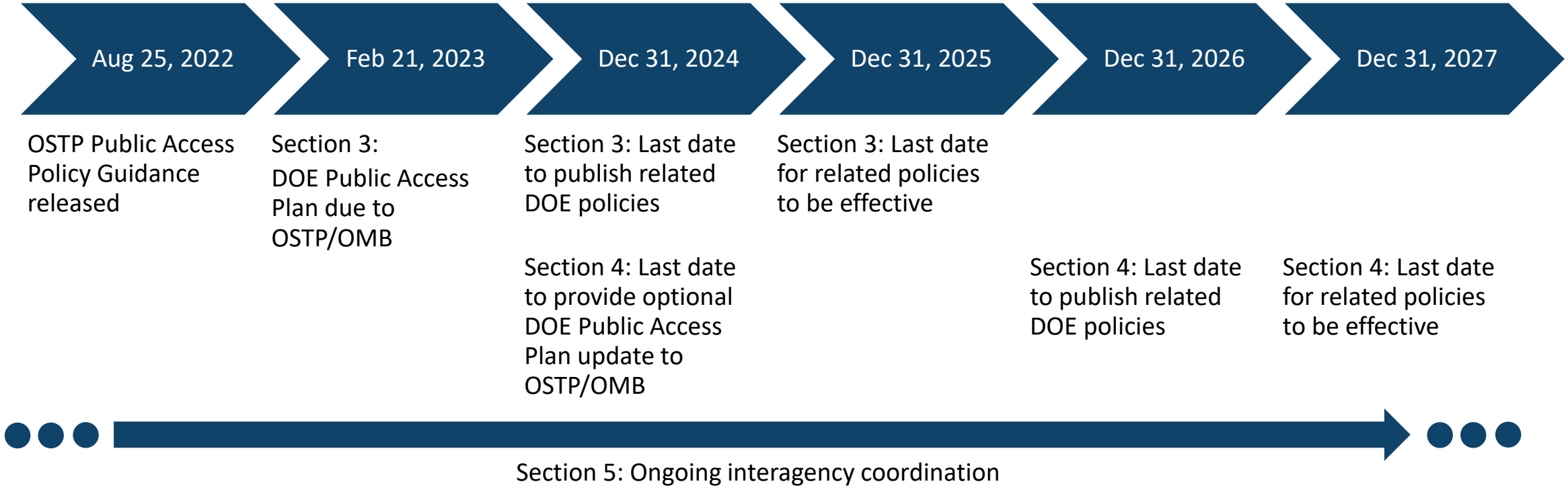


Full plan and FAQ available at:
https://www.energy.gov/doe-public-access-plan
https://doi.org/10.11578/2023DOEPublicAccessPlan

Implementation Timeline

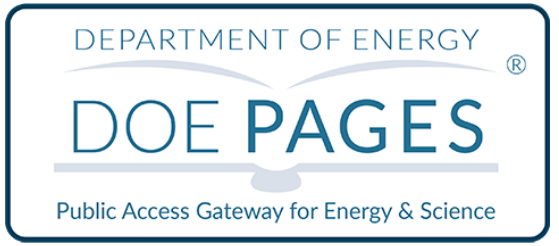
2022 OSTP Public Access Memo Section Descriptions

Section 3: Publications & Data | Section 4: PIDs to Ensure Research & Scientific Integrity | Section 5: Interagency Coordination



Publications – Memo Expectations

- Move from 12-month embargo to immediate access upon publication
- Continue to submit Accepted Manuscripts (AM) via E-Link, but earlier
- Provide access through DOE's designated repository, DOE PAGES®



Author's manuscript is accepted by publisher and completes peer-review process

Author submits the final peer-reviewed accepted manuscript to DOE via E-Link, using established processes

After receipt and processing, accepted manuscript is made available on DOE PAGES®

Getting to Zero-Embargo

Federal Purpose License allows DOE to provide immediate access to the AM

Submission of AM “Upon Acceptance” is encouraged (versus “upon publication”)

Will allow a “transition” period for adjusting workflow process

Pathways to Public Access

DOE allows for
author's choice in publishing

Green Open Access (OA) is encouraged but Gold OA is allowed if fees are "reasonable"

Publications – Implementation

Updating policy and guidance

- Revisions of DOE Directive 241.1B and Contractor Requirements Document (CRD)
- Updates to language in funding announcements and award packages for Financial Assistance Awardees

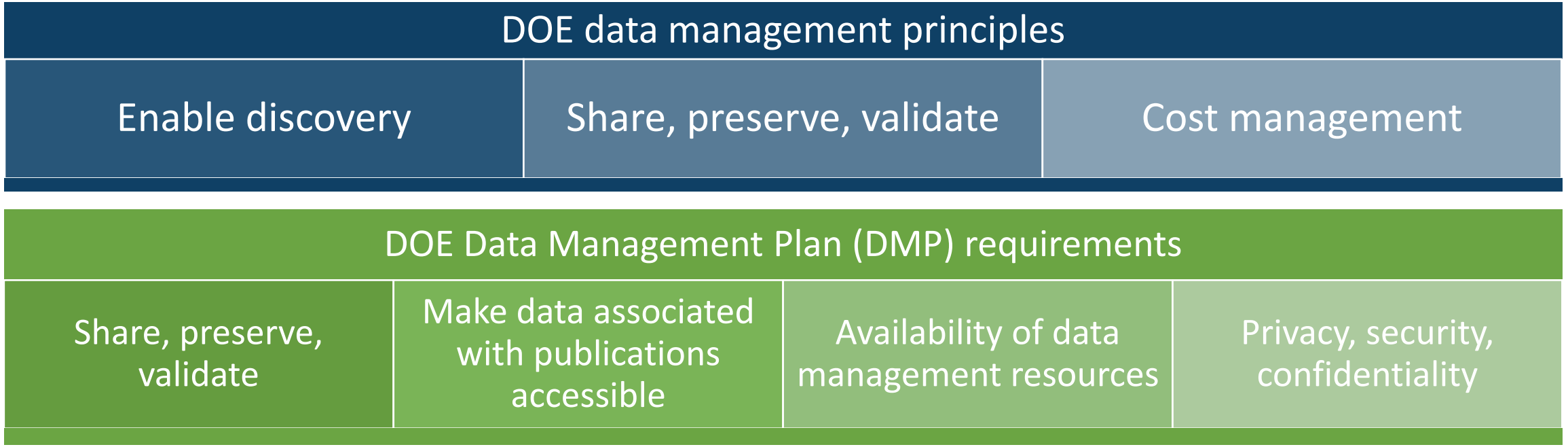
Formation of Scientific and Technical Information Program (STIP) focus group

- Publisher-related topics (APCs, OA fees, Read & Publish agreements)

Coordination with other federal agencies

- OSTP-led Subcommittee on Open Science and participation in various Working Groups

Current DOE Data Management Overview



- DMPs are reviewed, but there is flexibility in the process used for collection and review
- Additional requirements may be identified in a solicitation or invitation for research funding
- Implementation is supported through commensurate budget for the approved DMP scope

Full DOE policy: ***https://www.energy.gov/datamanagement/doe-policy-digital-research-data-management***
Full SC policy: ***https://science.osti.gov/Funding-Opportunities/Digital-Data-Management***

2023 Public Access Plan Data Management Overview

2023 PAP: Scientific Data Management Principles		
Increase pace of scientific discovery	Protect integrity, enhance value of science	Maximize appropriate data sharing

2023 PAP: Data Management and Sharing Plan (DMSP) Requirements				
Validation and replication of results	Timely and equitable access	Data repository selection	Data management and sharing resources	Data sharing limitations

- All DOE-funded R&D awards and contracts will be subject to a DOE approved DMSP, data reporting
- Targeting updates to Order 241.1B, Acquisition Letter, Financial Assistance Letter (Award T&C), reporting
- DMSP implementation will be supported through commensurate budget for approved scope
- OSTP Memo sets timeline for implementation by December 31, 2025

Full principles and requirements available in 2023 DOE Public Access Plan (https://doi.org/10.11578/2023DOEPublicAccessPlan)

Data Management Requirements Updates

- Highlighted changes from current data management strategy to 2023 PAP
  - Updating principles to emphasize equity and "maximize appropriate data sharing"
  - Creating opportunities to enhance compliance monitoring and evaluation metrics
  - "Data Management Plans" will become "Data Management and Sharing Plans"

Data Management Plan		Data Management and Sharing Plan
• "Research Data" to validate research findings	→	• "Scientific Data" to validate and replicate research findings
• Data underlying publications should be made as accessible as possible	→	• Data underlying publications should be made available at time of publication
• No explicit requirement regarding other research data	→	• Timeline for sharing other scientific data
• No explicit requirement regarding repository selection	→	• Repository selection should align with NSTC Desirable Characteristics guidance *(next slide)*

Desirable Characteristics of Data Repositories

- Guidance by the National Science and Technology Council (NSTC) Subcommittee on Open Science for federally funded research
  - Improves consistency in instructions to researchers about selecting data repositories
  - Helps ensure research data are findable, accessible, interoperable, and reusable (FAIR) to the greatest extent possible, while integrating privacy, security, and other protections

Organizational Infrastructure	Digital Object Management	Technology	Additional Considerations for Human Data
• Free and Easy Access • Clear Use Guidance • Risk Management • Retention Policy • Long-term Organizational Sustainability	• Unique Persistent Identifiers • Metadata • Curation and Quality Assurance • Broad and Measured Reuse • Common Format • Provenance	• Authentication • Long-term Technical Sustainability • Security and Integrity	• Fidelity to Consent • Security • Limited Use Compliant • Download Control • Request Review • Plan for Breach • Accountability

Desirable Characteristics of Data Repositories for Federally Funded Research, guidance by the NSTC Subcommittee on Open Science, published May 2022

Persistent Identifiers (PIDs) – Memo Expectations

From the White House Memos – ***"A digital identifier that is globally unique, persistent, machine resolvable and processable, and has an associated metadata schema."***

A long-lasting, managed, and registered unique digital reference (often in the form of a URL) to a research object (e.g. person, organization, research output, award) that can be represented or described online.

Collecting Metadata and Associated PIDs	PIDs for Researchers	PIDs for R&D Awards
Need to collect metadata associated with publications and data. Metadata should include: • author names, affiliations, and funding, referencing PIDs, • the date of publication; and, • a unique digital persistent identifier for the research output.	Agencies need to instruct researchers to obtain a PID for themselves. PID must be used in publishing when available and when reporting R&D outputs. PID must meet the common/core standards of a PID service defined in the NSPM-33 Implementation Guidance.	Agencies to assign unique digital persistent identifiers to R&D awards and intramural research protocols.

PID Examples

PIDs for Research Outputs – Publications, Reports, Data, Software		
		https://doi.org/10.1016/j.rinp.2023.106511 https://doi.org/10.11578/dc.20230501.1
PIDs for Awards – Grants, Contracts, Facility Use		
		https://doi.org/10.46936/10.25585/60000014
PIDs for People – PIs, Researchers, Senior/Key Personnel		
	ORCID	https://orcid.org/0000-0002-8523-1478
PIDs for Organizations – Funders, Universities, Publishers, Facilities		
		https://ror.org/01bj3aw27 http://doi.org/10.13039/100000015

PIDs – Implementation, Current Practices

Collecting Metadata and Associated PIDs	PIDs for Researchers	PIDs for R&D Awards
• Use E-Link to collect DOE R&D outputs metadata (including PIDs) associated with records. • Assign and collect R&D output DOIs. • Make R&D output metadata publicly and immediately available with full reuse rights (OSTI.GOV, DOE PAGES).	• Optionally collect PIDs for individuals (ORCID iDs) associated with authors in E-Link. • OSTI.GOV ORCID integration for individuals to claim records and add to their ORCID profile. • Lead the US Government ORCID Consortium.	• DOE offers award DOI assignment service for DOE labs, facilities, or sponsoring research offices.

PIDs – Implementation, New Practices

PIDs for Researchers	PIDs for R&D Awards	Collecting Metadata and Associated PIDs

• DOE O 241.1 and FOA and award clauses, instruct federal and contract employees, and financial assistance recipients conducting R&D to obtain PID for themselves.
• Compliance check will be when submitting STI to OSTI – DOE and DOE-funded authors will need to have an associated PID.
• Implementation approach is aligned with NSPM-33.
	• Working to collect and associate authenticated DOE contract and grant numbers with STI reports in E-Link 2.0. Ensures connection from DOE funding identifiers to PIDs for people, organizations, and outputs.
	• Explore options to assign PIDs to awards.
	• Participating in interagency discussions to talk through options and best practices.
		• New Organization Authority – used to associate organization PIDs with various organization metadata fields in E-Link 2.0.
		• DOE O 241.1C Requirements
		  • PIDs for R&D output records (consistent with current DOI practices).
		  • PIDs (ORCID iDs) for DOE and DOE-funded authors.
		• Financial Assistant Letter Requirements
		  • Recipients to provide PID for themselves when reporting to OSTI.

DOE PID WG – Federal integrated project team working on PID implementation strategies. Coordinating approach across DOE. Focused on PIDs for individuals, ensuring implementation of that aspect of OSTP Memo and NSPM-33.

Public Access Plan Implementation Process

- Coordinated effort across DOE to update our research funding and reporting mechanisms to reflect the 2023 Public Access Plan
  - Requires updating DOE Orders, contractor requirements documents, financial assistance guidance, award terms & conditions, reporting requirements, applicant guidance, etc.
- Targeted updates include:
  - Scientific and Technical Information (STI) reporting requirements and guidance
  - Guidance, requirements and suggested elements for DMSPs
  - Guidance and requirements for researcher PID reporting

Comments welcome via comments@osti.gov

2023: THE YEAR OF **OPEN SCIENCE**

"The principle and practice of making research products and processes available to all, while respecting diverse cultures, maintaining security and privacy, and fostering collaborations, reproducibility, and equity."

https://open.science.gov/

PIDs@ OSTI.GOV

https://www.osti.gov/pids/

PuRe Data Resources

https://science.osti.gov/Initiatives/PuRe-Data

DOE CODE

https://www.osti.gov/doecode/

Award DOI Service

https://www.osti.gov/award-doi-service/

Champions of Open Science

- White House OSTP Year of Open Science Recognition Challenge
  - Recognizing open science stories to benefit society
  - 5 winners announced in March 2024

Project Jupyter: reproducible and collaborative computational science and education

Category: Technical Advancement to Enable Open Science

Federal Support: NSF and DOE

Project leads: Brian Granger, Jason Grout, Fernando Pérez, Ana Ruvalcaba, and Steven Silvester

"Because Jupyter notebooks help to make computational research reproducible and accessible, Jupyter has enabled millions of researchers and thousands of organizations to adopt open science practices in their research and education."

Used by:

KBase PREDICTIVE BIOLOGY
DOE Systems Biology Knowledgebase

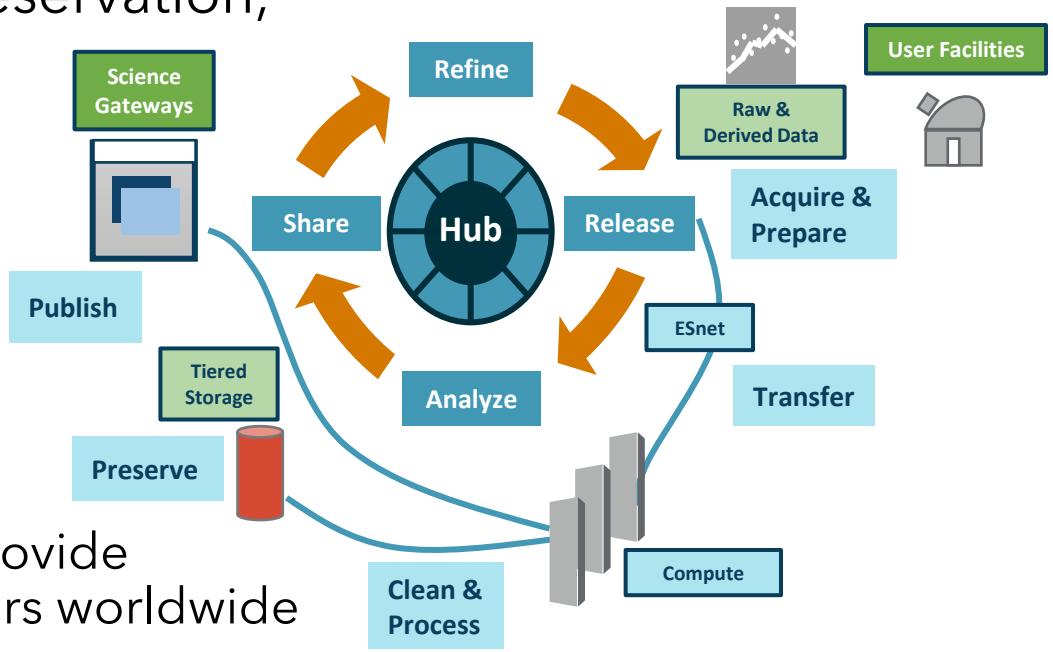
VERA C. RUBIN OBSERVATORY

Integrated Research Infrastructure & HPDF

2023 IRI Architecture Blueprint Activity identified three science patterns:

- *Time-sensitive patterns*
- *Data-integration-intensive patterns*
- *Long-term campaign patterns*

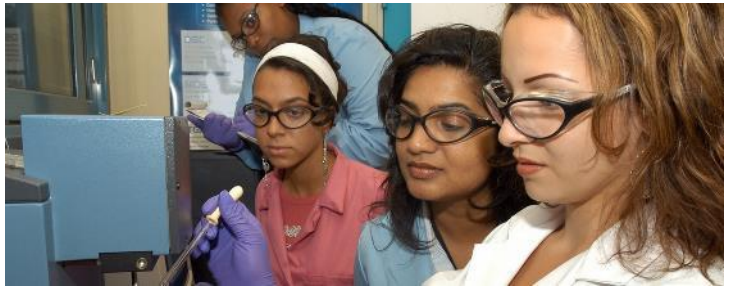
- DOE envisions a revolutionary ecosystem, the **Integrated Research Infrastructure**, to deliver seamless, secure interoperability across National Lab facilities
- The **High Performance Data Facility (HPDF)** will enable analysis, preservation, and accessibility to the staggering amounts of experimental data produced by SC facilities
  - Distributed operations model will be essential to long-term success and required performance levels
  - Hub & spoke architecture will provide seamless, tailored service to users worldwide



Office of Science Initiatives: Broadening Participation

RENEW

Reaching a New Energy Sciences Workforce



FAIR

Funding for Accelerated, Inclusive Research

PIER Plans

Promoting Inclusive and Equitable Research

Vision and Burning Questions

- Office of Science vision for implementing the DOE Public Access Plan
  - Enable SC researchers to develop responsive DMSPs
  - Connect SC scientists to the data (and other) experts needed to craft and implement DMSPs
  - Integrate PIDs into the scientific workflow to connect research outputs to data/code/tools
  - Encourage data citation and reuse to enable replication of results and new modes of science
  - Enable career-impacting recognition for contributing public data and tools of open science

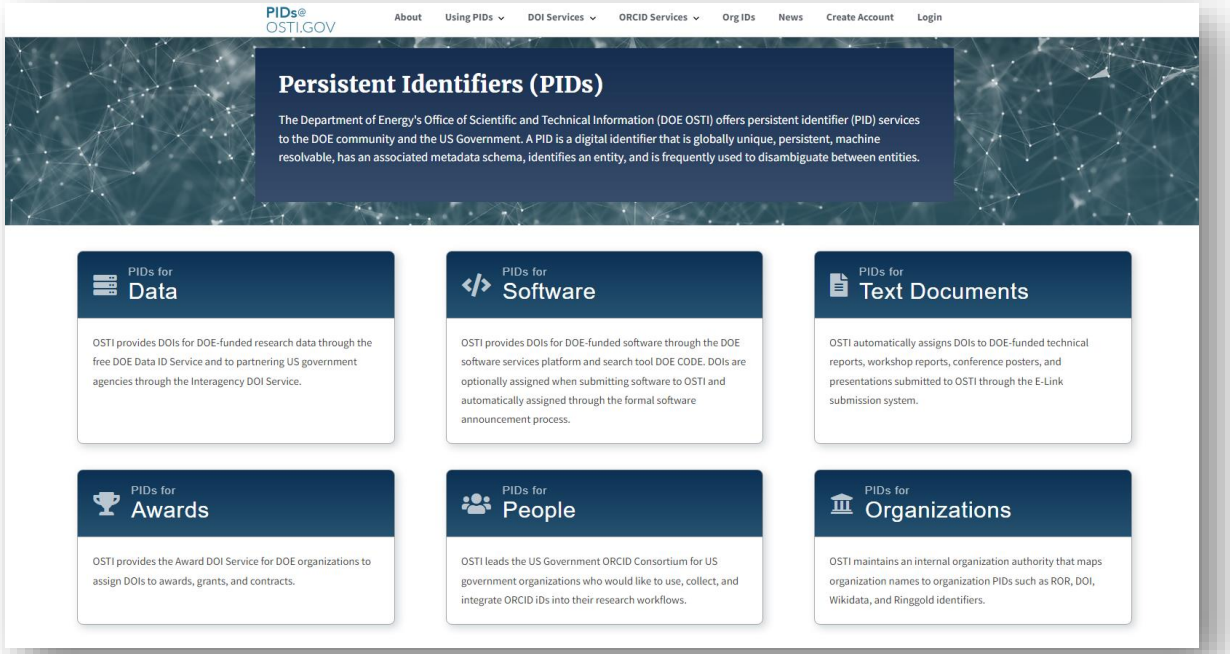
- Burning questions
  - How can we enable researchers to develop and implement responsive DMSPs while letting them focus on their science?
    - Through SC infrastructure? Through helpful guidance? Through community best practices?
  - How can we enable scientists to explore new opportunities made possible by FAIR data?
    - While also recognizing impactful data/code/tools and the people behind them?
  - How can we lower the barrier for broadening participation in science?


THANK YOU!

PIDs@OSTI.GOV

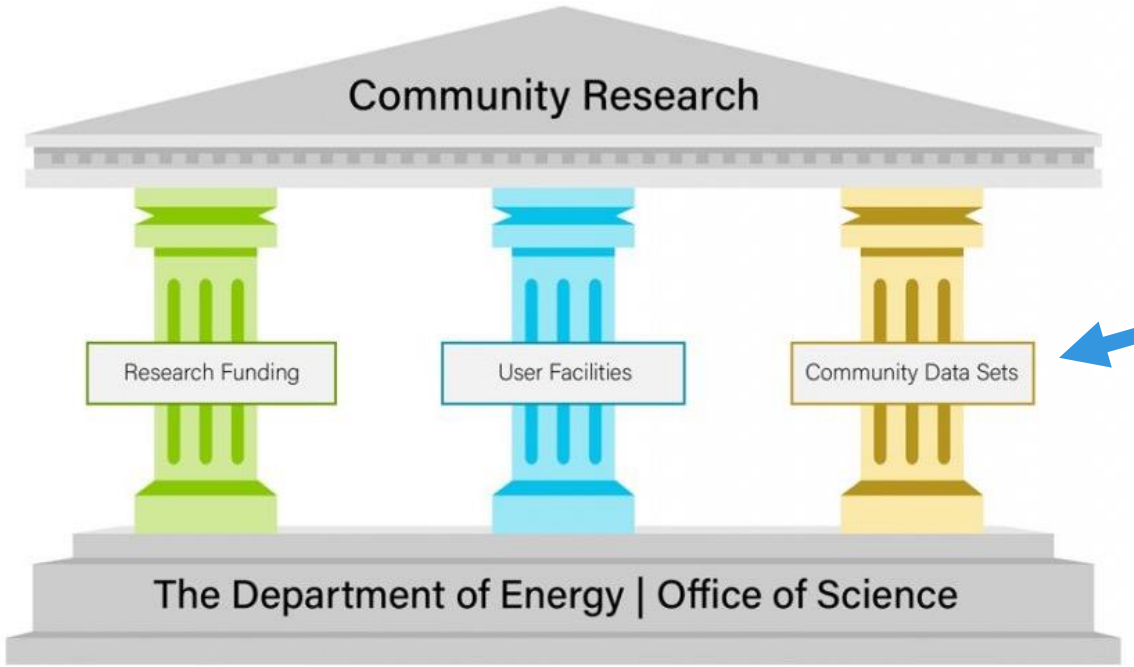
PIDs@OSTI.GOV brings together information about persistent identifiers (PIDs) and the services DOE's Office of Scientific and Technical Information (OSTI) provides for the DOE community and more broadly for U.S. government agencies

- PIDs deliver value to the broader research community by:
  - Enabling greater discovery and reuse
  - Providing appropriate credit
- PIDs@OSTI.GOV provides:
  - General information about PIDs
  - Details about OSTI's PID services
  - Community resources
  - Visualizations of the power of PIDs



https://www.osti.gov/pids/

DOE PuRe Data Resources



PuRe Data Resources

Public Reusable Research (PuRe) Data Resources aim to make data publicly available in order to advance scientific or technical knowledge

https://science.osti.gov/Initiatives/PuRe-Data

ATcT JGI JOINT GENOME INSTITUTE NNDC PDG particle data group The Materials Project ARM KBase PREDICTIVE BIOLOGY DOE Systems Biology Knowledgebase

FAIR Data Principles

The FAIR Guiding Principles for scientific data management and stewardship:
https://doi.org/10.1038/sdata.2016.18

Findable

- F1. (Meta)data are assigned a globally unique and persistent identifier
- F2. Data are described with rich metadata (defined by R1 below)
- F3. Metadata clearly and explicitly include the identifier of the data they describe
- F4. (Meta)data are registered or indexed in a searchable resource

Accessible

- A1. (Meta)data are retrievable by their identifier using a standardised communications protocol
  - A1.1 The protocol is open, free, and universally implementable
  - A1.2 The protocol allows for an authentication and authorization procedure, where necessary
- A2. Metadata are accessible, even when the data are no longer available

Interoperable

- I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.
- I2. (Meta)data use vocabularies that follow FAIR principles
- I3. (Meta)data include qualified references to other (meta)data

Reusable

- R1. Meta(data) are richly described with a plurality of accurate and relevant attributes
  - R1.1. (Meta)data are released with a clear and accessible data usage license
  - R1.2. (Meta)data are associated with detailed provenance
  - R1.3. (Meta)data meet domain-relevant community standards