NIH Data Management and Sharing Policy

Effective January 25, 2023

Genevieve Milliken, MA, MSLIS - November 30, 2022

Content Overview

- Data Sharing Landscape
- Scope of the Policy
- Elements of the DMSP
- Tools and Resources for writing a DMSP
- Institutional Considerations
- Additional Resources
- Q & A

Data Sharing Landscape

Why does NIH Want Data to be Shared?

Advance rigorous and reproducible research

- Enable validation of research results
- Make high-value datasets accessible
- Accelerate future research directions
- Increase opportunities for reuse, citation and collaboration

Promote public trust in research

- Foster transparency and accountability
- Demonstrate stewardship over taxpayer funds
- Maximize research participants' contributions
- Support appropriate protections of research participants' data

Major NIH Wide Data-Related Policies

Policy	Expectations	Year
NIH Data Sharing Policy	Expects investigators seeking more than $500K in direct support in any given year to submit a data sharing plan with their application or to indicate why data sharing is not possible.	2003
Genomic Data Sharing Policy	Expects sharing of large-scale human and non-human genomic data from NIH-funded studies through a publicly available data repository. All studies with human genomic data should be registered in dbGaP, and the data should be submitted to an NIH-designated data repository. Non-human data may be submitted to any widely used data repository.	2014
Intramural Human Data Sharing Policy	Expects all intramural investigators generating or collecting human research data to develop a data sharing plan describing how data will be shared.	2015

Still Work to Do: Data Accessibility & Availability

"Data sharing practices and data availability upon request differ across scientific disciplines," Tedersoo et al., Scientific Data volume 8, Article number: 192 (2021)

- Evaluated **data availability in 875 Nature and Science papers across nine disciplines** published 2000-2019
- Data initially obtained from authors in **39.4% of requests on average**; ranged **27.9–56.1% among research fields**, improved with repeated follow-up
- **19.4% of data requests were declined on average**

Still Work to Do: Reproducibility

"Reproducibility in Cancer Biology: Challenges for assessing replicability in preclinical cancer biology," Errington et al., eLife 2021;10:e67995 (published Dec. 7, 2021)

- Attempted to **repeat 193 experiments from 53 high-impact cancer biology** papers
- Unable to obtain data for **68%** of experiments
- **39 of 53 papers cited NIH funding** (75%)

Scope of the Policy

NIH Data Management & Sharing Policy

Requires Data Management and Sharing Plans (DMPs) for all research conducted at or funded (in whole or in part) by the NIH that **generates scientific data**.

This includes:

- Competing grant applications that are submitted to NIH
- Contract proposals
- NIH Intramural Research Projects
- Other funding agreements with the NIH (e.g., Other Transactions)

Does not apply to funding that does not generate data

What do we mean by scientific data?

NIH defines **Scientific Data** as:

"The *recorded factual material* commonly accepted in the scientific community as of sufficient quality to *validate and replicate research findings*, regardless of whether the data are used to support scholarly publications."

Excludes:

- Lab notebooks
- Preliminary analyses
- Completed case report forms
- Peer reviews
- Physical objects, such as laboratory specimens

What counts as data for sharing?

- Adequate data to validate and replicate study findings
- Data resulting from the study but not necessarily supporting a publication
- Null findings that do not result in publication

Details of the Policy

Scope: All NIH-supported **research generating scientific data**

"Recorded factual material… of sufficient quality to **validate and replicate** research findings, regardless of whether the data are used to support scholarly publications"—relates to the proposed research questions and findings can include unpublished null results

Timeline: When to share **data**?

No later than publication or end of award (for data underlying findings not published in peer- reviewed journals)

Additional Expectations

Sharing Should be…..

- ***The default practice***
  - Data sharing should be maximized (with justifiable limitations)
  - All data should be managed; **not all must be shared**

- ***Responsibly implemented***
  - Plans should outline protection of privacy, rights, and confidentiality
  - Abide by existing laws, regulations, and policies
  - Prospectively planned for at all stages of the research process

Elements of the DMSP

What is a Data Management and Sharing Plan (DMSP)?

- DMSPs state how scientific data will be **managed and shared**, including key procedures and responsibility over the course of the award
- DMSPs include **descriptions** of what data will be shared and what limitations there are on data sharing
- Submitted DMSPs **must comply** with other relevant policies (e.g., NIH Genomic Data Sharing Policy)
- The policy does not proscribe methods of **data sharing or management**, but does have preferences (e.g., use of NIH-supported repositories). The policy also expects PIs to follow the submitting plans.

Elements of a Data Management and Sharing Plan (2023)

1. **Data type**
   Identifying data to be preserved and shared
2. **Related tools, software, and/or code**
   Tools and software needed to access and manipulate data
3. **Data standards**
   Standards to be applied to scientific data and metadata
4. **Data preservation, access, and associated timelines**
   Repository to be used, persistent unique identifier, and when/ how long data will be available
5. **Access, distribution, or reuse considerations**
   Description of factors for data access, distribution, or reuse
6. **Oversight of data management and sharing**
   Plan compliance will be monitored/ managed and by whom

1. Data Type

- A description of data, associated metadata, and documentation (e.g., data type, data dictionary, study protocol, data collection instruments)
  - Data modality (e.g., imaging, genomic, survey)
  - Level of aggregation (e.g., individual, grouped, summarized)
  - Level of data processing (e.g., raw vs. processed data)

2. Related tools, software, and/or code

- Any additional tools/software are needed to access or manipulate the data (e.g., Python packages/modules)
- Names of specific software tools (e.g. Python, SPSS, etc)
- Availability of tools (e.g. open source vs purchase)
- Expected lifespan of the tools compared to length of data availability

3. Standards for the data/metadata

- Data formats
- Data dictionaries
- Common Data Elements
- Identifiers
- Definitions
- Indicate when no consensus standard exists

4. Data preservation, access, and associated timelines

- Name(s) of repository(ies) where data and metadata will be deposited
- How data will be made identifiable (e.g., Digital Object Identifier)
- When the data will be made available and for how long
    - No later than time of publication or end of performance period
    - How long data is anticipated to be available

5. Access, distribution, or reuse considerations

- Any limitations due to informed consent, privacy and confidentiality protections
- Controlled access protections (e.g., author approval, DUA)
- Restrictions imposed by federal, Tribal, or state laws, regulations, or policies (e.g. HIPAA)

6. Oversight of Data Management and Sharing

- Describe how compliance with the DMSP will be monitored and managed (e.g., frequency, person(s) responsible), including budgeting requirements
- Who will be responsible for oversight
- How often will oversight activities occur

Plan submission, review, and compliance

Submission

With application for funding in Budget Justification section

Assessment

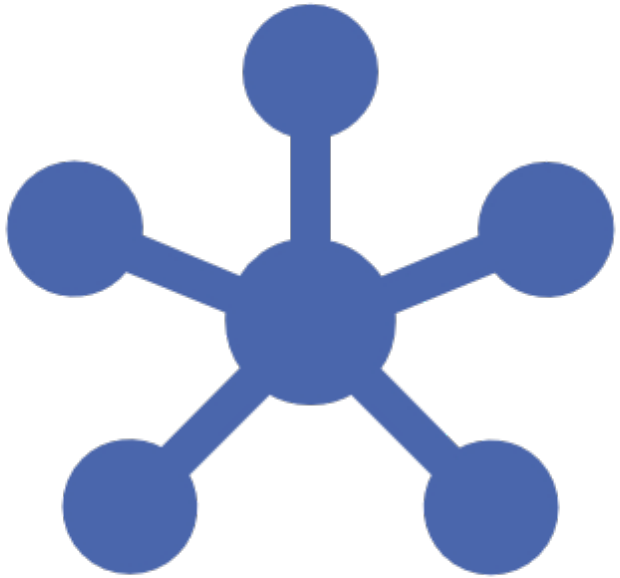
- Peer reviewers may comment on (not score) budget
- NIH program staff assess Plans
- Plans can be updated

Compliance

- Incorporated into Terms and Conditions
- Monitored at regular reporting intervals
- Compliance may factor into future funding decisions

DMSP and the broader policy landscape

- DMSP represents the minimum requirements for all NIH grants
- NIH Institutes, Centers, and Offices (ICOs) may provide more specific guidance
- Other policies may apply (e.g. Genomic Data Sharing policy)



Tools and Resources for Writing a DMSP

Allowable Costs

Reasonable, allowable costs may be included in **NIH budget requests** for:

- Curating data
- Developing supporting documentation
- Formatting data according to accepted community standards, or for transmission to and storage at a selected repository for long-term preservation and access
- De-identifying data
- Preparing metadata to foster discoverability, interpretation, and reuse
- Local data management considerations, such as unique and specialized information infrastructure necessary to provide local management and preservation (for example, before deposit into an established repository).
- Preserving and sharing data through established repositories, such as data deposit fees.
https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/budgeting-for-data-management-sharing

Unallowable Costs

Budget requests must NOT include:

- Infrastructure costs that are included in institutional overhead (for instance, Facilities and Administrative costs)
- Costs associated with the routine conduct of research, including costs associated with collecting or gaining access to research data.
- Costs that are double charged or inconsistently charged as both direct and indirect costs

https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/budgeting-for-data-management-sharing

Selecting a repository

NIH strongly encourages
subject-specific, open access Data Sharing Repositories
as a first choice.

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Datasets up to **2 gigabytes**

PubMed Central

Stores publication-related supplemental materials and datasets directly associated publications.

Datasets up to **20 gigabytes**

Generalist Repositories

Datasets associated with publications or otherwise and links to PubMed.

High priority datasets, **petabyte-scale**

Cloud Partners
(STRIDES Program)

Store and manage large scale, high priority NIH datasets.

DMPTool

DMPTool
Build your Data Management Plan

- Online application from the UC libraries that helps researchers create data management plans.
- Provides a click-through wizard for creating a DMP that complies with funder requirements.
- Used by many institutions, including Harvard, Johns Hopkins, UPenn, Stanford, and many others.
- Current community focus is on the new NIH DMSP Policy

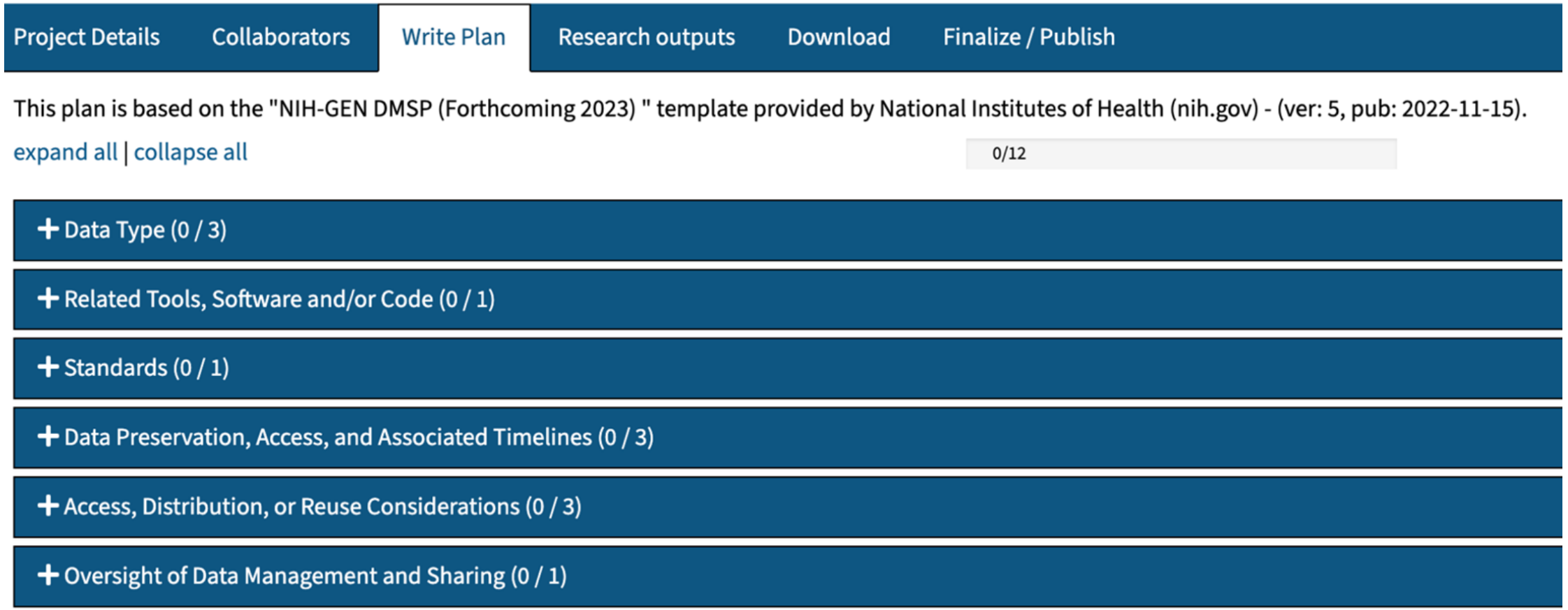
DMPTool. https://dmptool.org/

Functions of DMPTool

- Supports collaborative creation of DMPs across a lab or research project, including tracking changes to the DMP and who makes them
- Template helps ensure you answer necessary questions and makes the policy requirements more comprehensible
- Includes guidance from the grant funder, DMPTool itself, and your institution
- Allows for review and collaboration
- Can be downloaded in various formats
- Provides a repository of existing, public DMPs

DMPTool. https://dmptool.org/

2023 NIH Data Management and Sharing Template



Working Group on NIH DMSP Guidance

- Policy Readiness Checklist for Librarians
- Data Management and Sharing Plan Checklist for Researchers
- Example DMSPs
- Glossaries (e.g., data terms related to the DMS policy, grant glossary for librarians)
- Repository Finder (forthcoming)

All content is under a CC 4.0 License and can be found at:
https://osf.io/uadxr/

Institutional Considerations

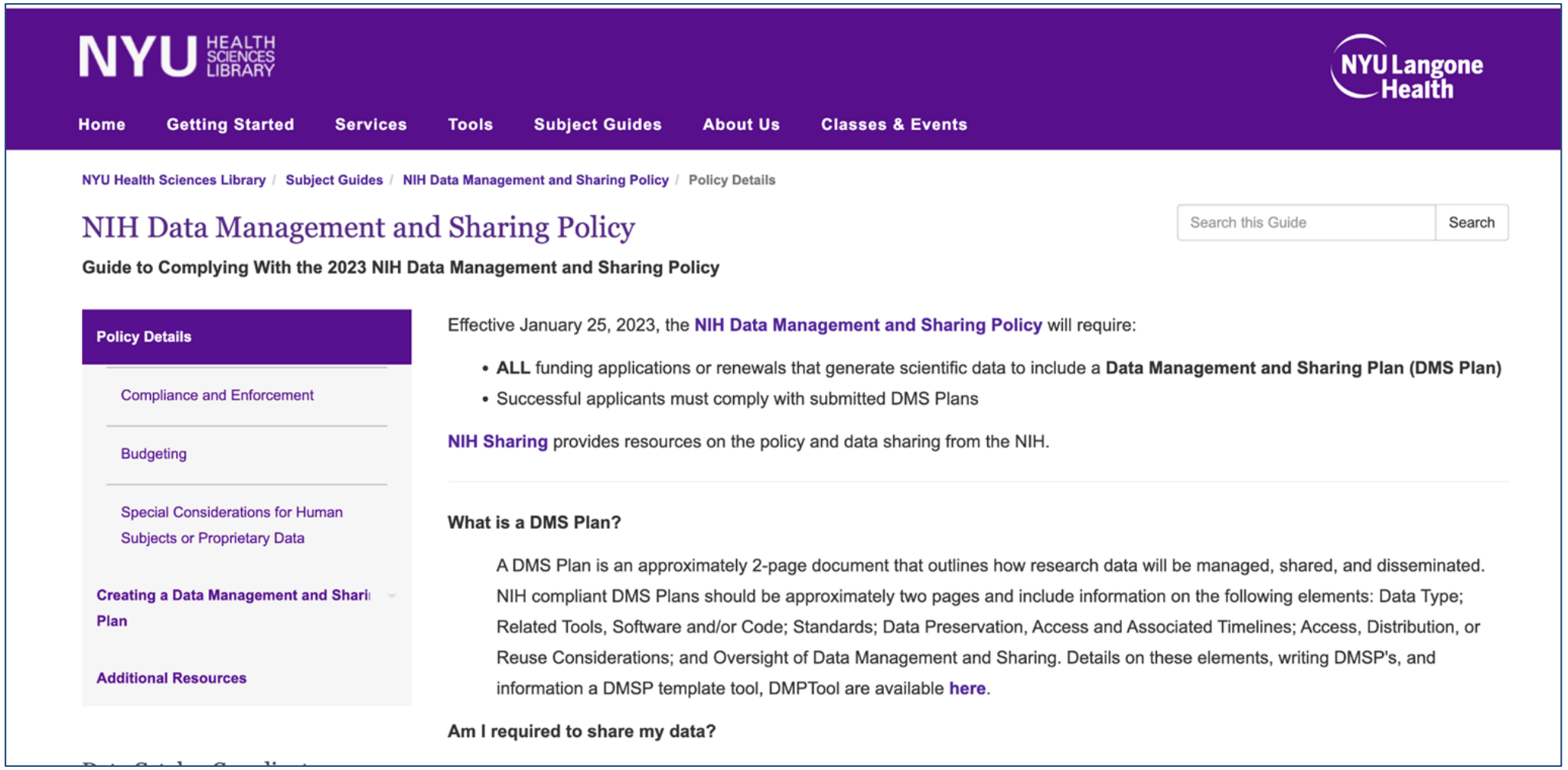
Library Education

- Borrow from existing resources to develop online guides customized for your institution (describing what departments might help, and how they might be available to help)
- Connect with departments and offices to market library services – through faculty meetings, newsletters, social media.
- Hopefully, you can leverage existing relationships to determine the needs of different departments and provide specific education based on needs.
- Keep in mind your capacity for providing education on different levels.
  - Are you able to provide general workshops about the policy for any audience, specific sessions for small groups, one-on-one consultations?
  - Trying a variety of approaches may show you what can be successful at your institution.

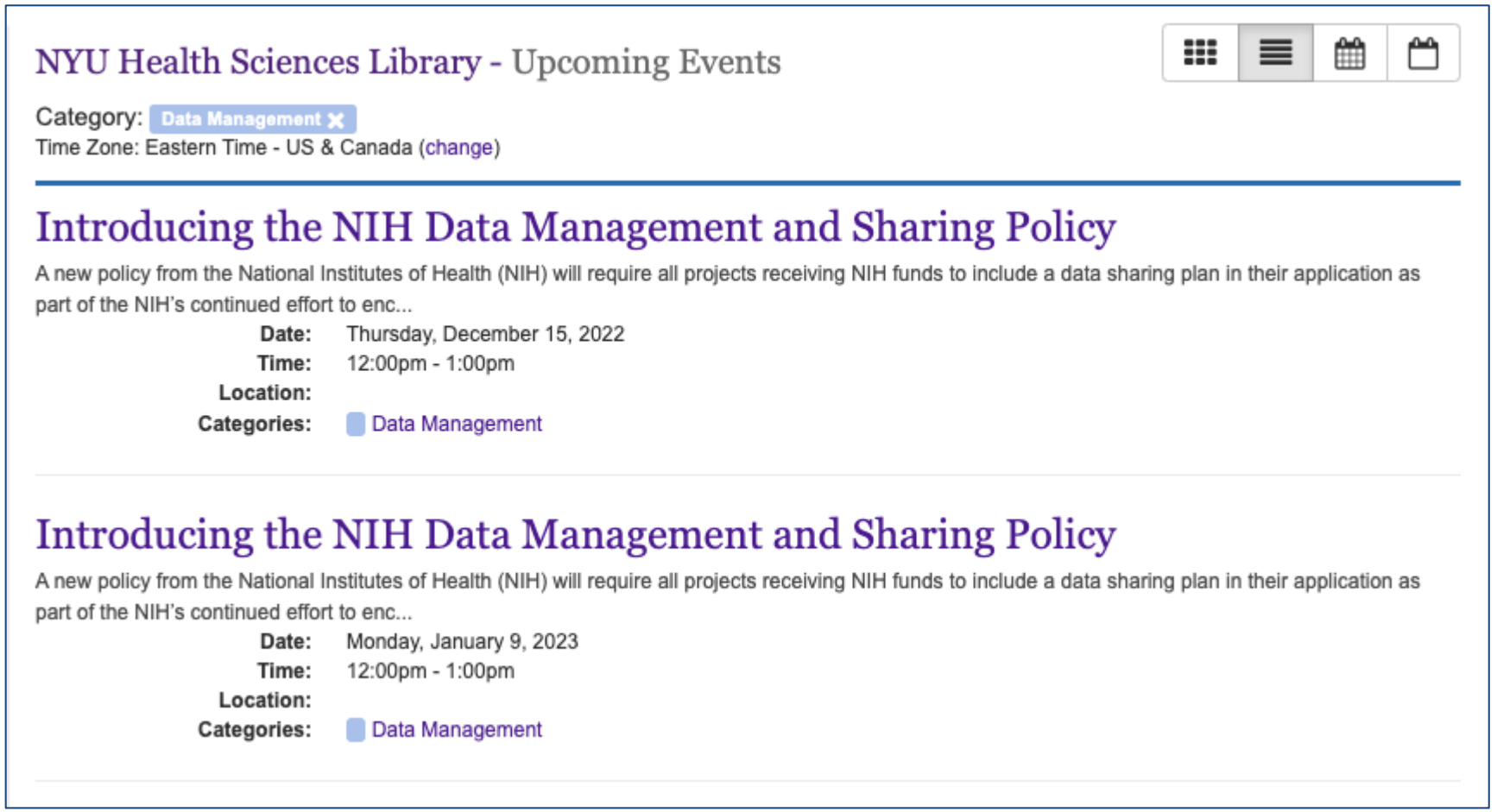
Outreach

- Meet with department head and library leadership to communicate what library services are currently in place, what services are realistic to offer, and how services could grow with more support (more hires, coordination/support from campus).
- Determine if other departments are working on policy compliance:
  - IT? Office of Sponsored Research/Grants? Faculty Development Center? Work with them to figure out exactly how everyone is helping and how you can coordinate and collaborate on services.
- If possible, hold regular meetings with institutional partners to get and provide updates on services.

Creating Library Resources



Hosting Audience-Specific Workshops



NYU Health Sciences Library - Upcoming Events

Category: Data Management

Time Zone: Eastern Time - US & Canada (change)

Introducing the NIH Data Management and Sharing Policy

A new policy from the National Institutes of Health (NIH) will require all projects receiving NIH funds to include a data sharing plan in their application as part of the NIH's continued effort to enc...

Date: Thursday, December 15, 2022
Time: 12:00pm - 1:00pm
Location:
Categories: Data Management

Introducing the NIH Data Management and Sharing Policy

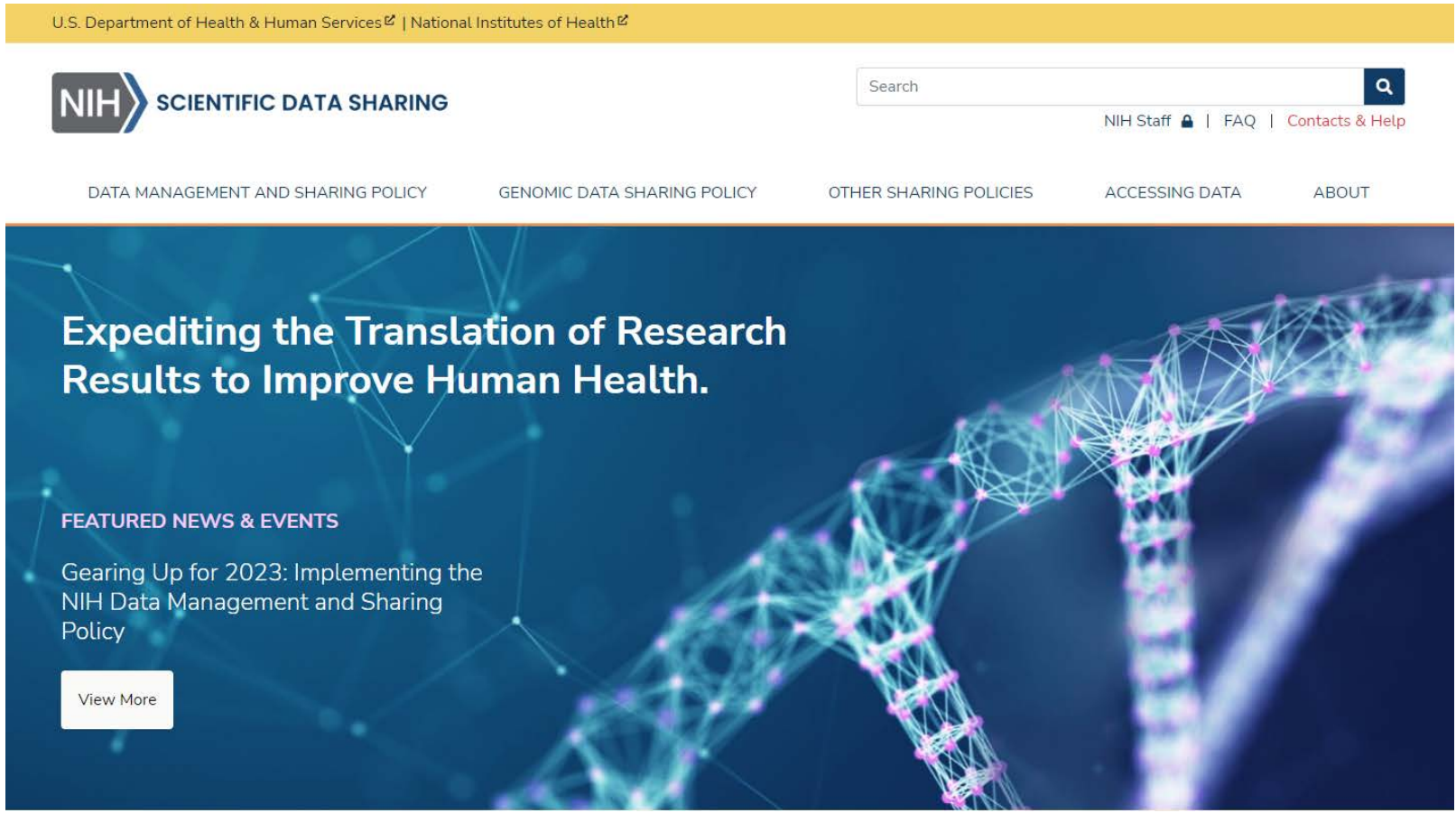
A new policy from the National Institutes of Health (NIH) will require all projects receiving NIH funds to include a data sharing plan in their application as part of the NIH's continued effort to enc...

Date: Monday, January 9, 2023
Time: 12:00pm - 1:00pm
Location:
Categories: Data Management

Additional Resources

Keeping up-to-date: sharing.nih.gov

- Provides a central source of guidance related to multiple NIH data sharing policies
- Covers Data Management and Sharing, Genomic Data Sharing, Model Organisms, and Research Tools policies
- Content will be updated



Informed Consent for Secondary Research with Data and Biospecimens: Points to Consider and Sample Language for Future Use and/or Sharing

- **Both data and biospecimens can be reused to benefit science and society**
  - Responsible stewardship is essential
  - Effective consent is needed to facilitate data and biospecimen storage and sharing
- **Serves as a resource to:**
  - Uphold individual autonomy, strengthen trust in research
  - Communicate clearly the potential risks, benefits
- **Revised based on community input**
  - “Points to consider" for investigators, IRBs when modifying consent language
  - Sample consent language for data, biospecimen storage, sharing
  - Use is completely voluntary

https://osp.od.nih.gov/wp-content/uploads/Informed-Consent-Resource-for-Secondary-Research-with-Data-and-Biospecimens.pdf

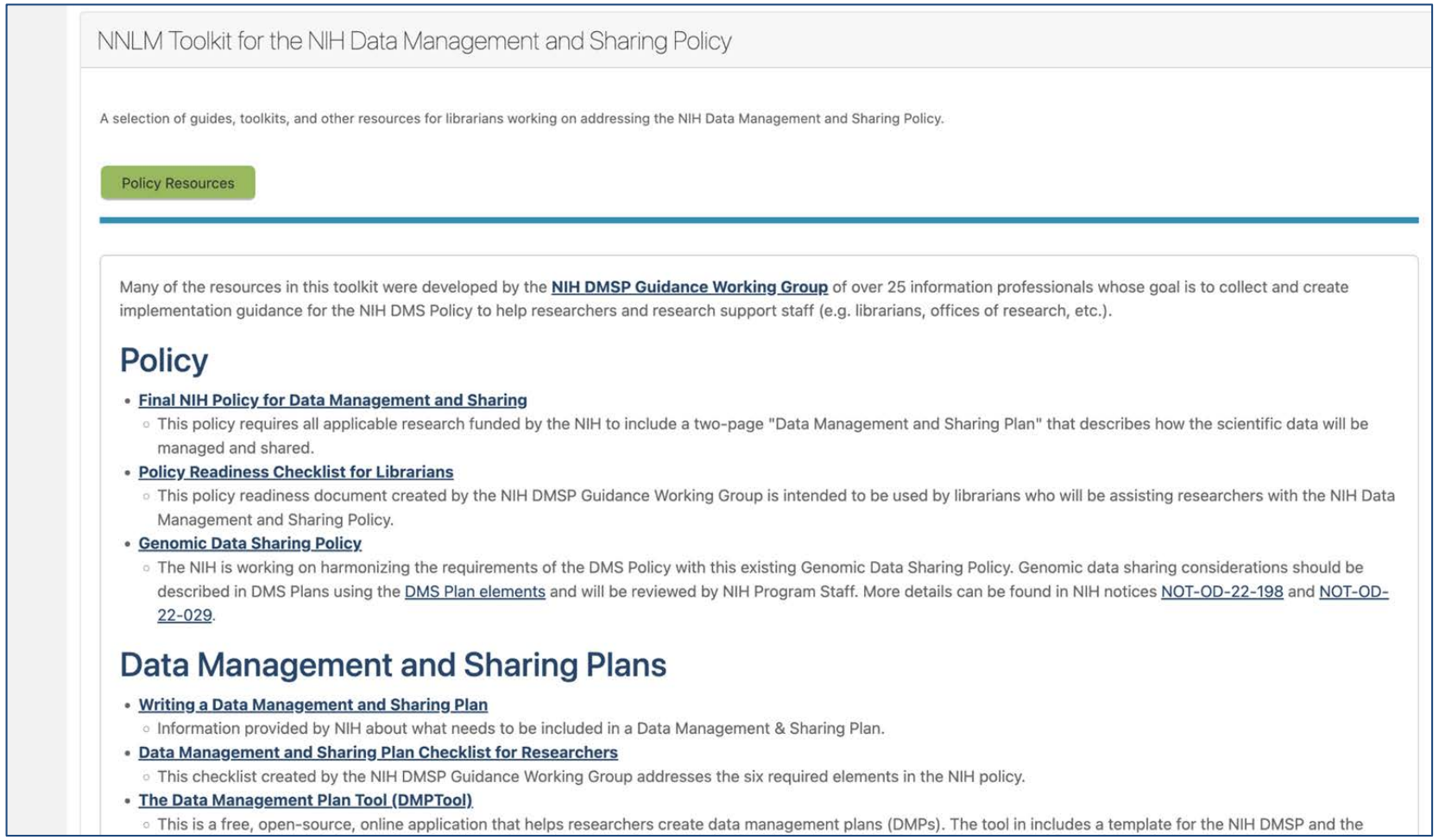
Policy Specific Resources

- NIH Data Sharing Policy
  - Includes guidance on the current and new policies
- NIH Data Sharing Policy FAQs
  - Some answers to frequently asked questions about the 2023 policy
- NIH Data Sharing Resources
- List of Domain-Specific Repositories
- List of Generalist Repositories
- DMPTool
- Registry of Research Data Repositories (RE3Data)
  - An international directory of online data repositories
- FAIRSharing.org
  - Includes information on repositories, data sharing policies, and data standards
- Forecasting Costs Related to Preserving Biomedical Data
  - Provides guidance on costs related to preserving and sharing data

NNLM Toolkit for the NIH Data Management and Sharing Policy

- About the policy
- Resources for writing a DMS Plan
- Templates and sample plans
- Repository lists
- Resources for cost
- Classes (on demand)
- Glossaries

https://www.nnlm.gov/guides/nnlm-toolkit-nih-data-management-and-sharing-policy



Conclusions

Takeaways

- Recognize what's in the policy and what's not in the policy.
- Consider what you can do, but also be realistic and acknowledge limitations.
- Communicate with leadership and express your needs and limitations.
- Seek out existing resources wherever possible.
- Connect with others doing this work (at your institution and elsewhere!).

Preparing for 2023—No Time like the Present

Steps you can take to begin preparing for the Policy:

- Try **drafting a data management and sharing plan** for your work based on the recommended elements (NOT-OD-21-014)
- **Identify existing resources within your institution** that may be able to assist you, such as data librarians
- **Review your past data sharing practices** to meet other funder or publisher expectations and consider what you may need to update for the new Policy in the future

Credit and Thanks to:

Genevieve Milliken

Justin de la Cruz

Lisa Federer

Sara Samuels

Heather Coates

Nicole Contaxis

Nina Exner

Ariel Deardorff

Michelle Yee

Katie Pierce Farrier

Taunton Paine

Cindy Danielson

Ummea Urmi

Thanks! Questions?

Contact the NCDS and keep connected:

- Email: nnlm-ncds@nyulangone.org
- Twitter: @nnlmncds
- Website: https://nnlm.gov/about/centers/ncds
- Blog & Newsletter: https://news.nnlm.gov/ncds/