

These slides:

bit.ly/cu-dmsp-townhall



University Research

UNIVERSITY OF COLORADO DENVER | ANSCHUTZ MEDICAL CAMPUS

CRIO COUNCIL DMSP WG TOWN HALL

12.13.2022

<https://research.cuanschutz.edu/rio>



CRIO Council-led Data Management and Sharing Plan Working Group

- WG comprised of members from CRIO Council and the research community (schools, library, OGC, OIT, etc.)
- Identify preparations needed for campus to be ready
- WG activities page, updates, and templates are available on the Research website:

<https://research.cuanschutz.edu/crio/projects/nih-data-management-and-sharing-policy>

- Campus resources and templates for the DMSP
- NIH and community resources and training materials
- Grant submission and RPPR strategies and materials
- List of data repositories for data deposition
- Open to all! Email crio@cuanschutz.edu if you would like to help out!

Entering Slido

Please go to slido.com



Join the meeting at: #DMSPatCU

First poll

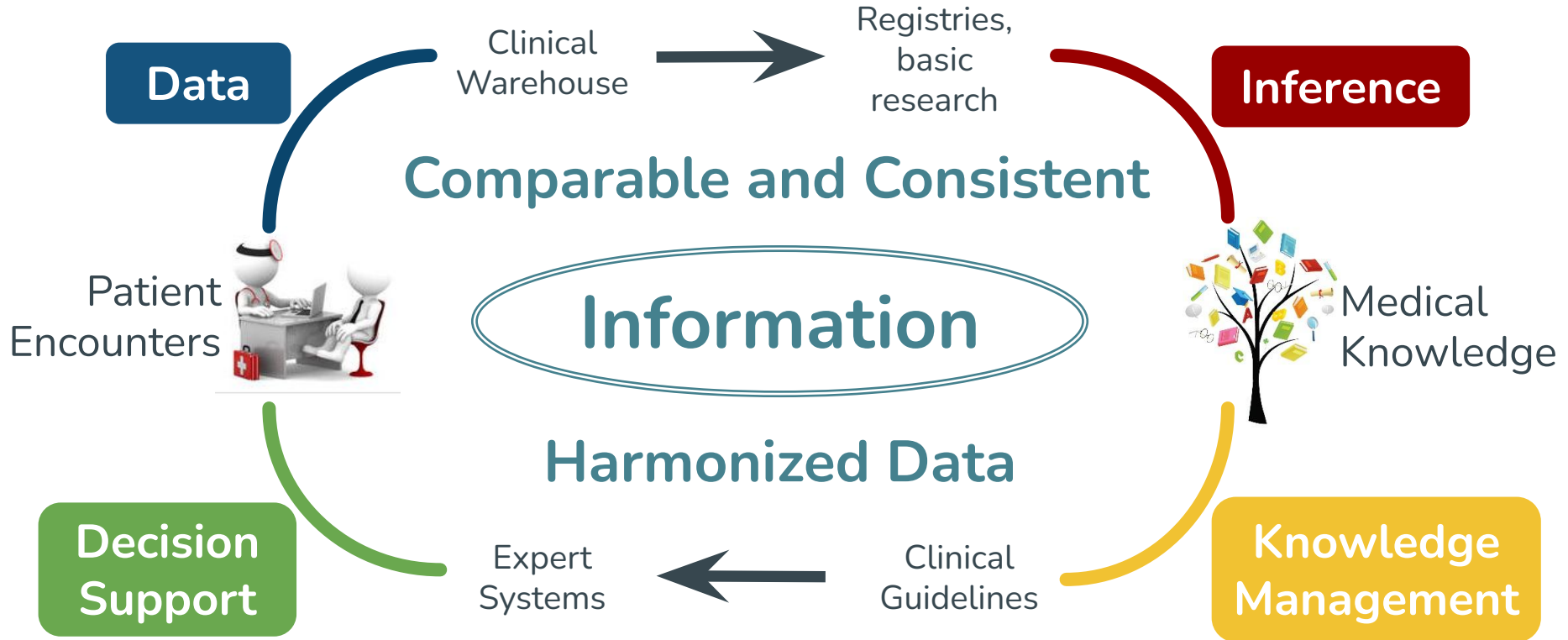
slido



What department or unit are you in?

① Start presenting to display the poll results on this slide.

From Practice-based Evidence to Evidence-based Practice



New NIH Data Management and Sharing policy

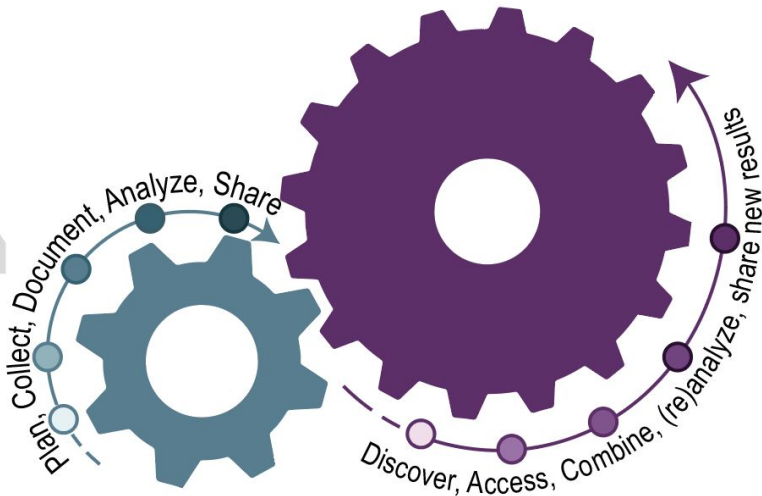
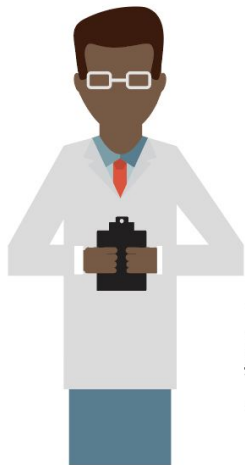
Beginning in January 2023, the Final NIH Policy for Data Management and Sharing (NOT-OD-21-013) will require researchers to include a Data Management and Sharing Plan (DMSP) in all funding applications.

Key Features:

- Applies to all research funded by NIH that results in the generation of scientific data; it does not apply to activities that do not generate scientific data, e.g. training, infrastructure development, etc.
- The NIH recommendation is 2 pages long
- The DMSP is not a scored component of proposals
- The DMSP can contain justified uncertainties, and can be updated annually.
- Annual Notice of Award is dependent on RPPR reporting on compliance to the awardee's DMSP and other NIH data sharing policies.
- Grant budgets may include costs for preserving and sharing data, including personnel and storage costs. These funds must be spent during the funding period.
- Scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of performance period, whichever comes first.

The data re-user's perspective is equally as important as the data provider's

DATA GENERATOR



DATA RE-USER



Whether the data is open or closed, or reusable is often impossible to know until you try.



SCHRÖDINGER'S DATABASE IS
COLOPES

...This trial often comes at significant legal and technical expense since most licenses are missing, vague, or restrictive, and most data are not formatted, documented, or accessible for reuse.

Getting from FA to FAIR requires some TLC



Traceable



Licensed



Connected

**“Traceability, Licensing,
and Connectedness --
OH MY!”**





Traceable

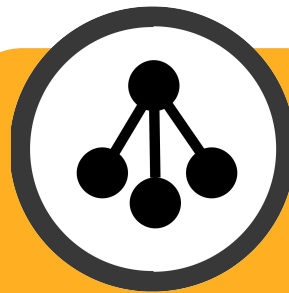
- Evidence
- Provenance
- Attribution

*Traceability and Licensure go hand-in-hand:
Licensing often a (convoluted)
credit hack*



Licensed

- Clearly stated
- Comprehensive and non-negotiated
- Accessible
- Avoid restrictions on kinds of (re)use
- Avoid restrictions on who may (re)use



Connected

- Identifiers
- Data models
- Ontologies
- Concept alignment
- Clinical Data Elements

bit.ly/fair-tlc

Entering Slido

Please go to slido.com



Join the meeting at: #DMSPatCU

Second poll

slido

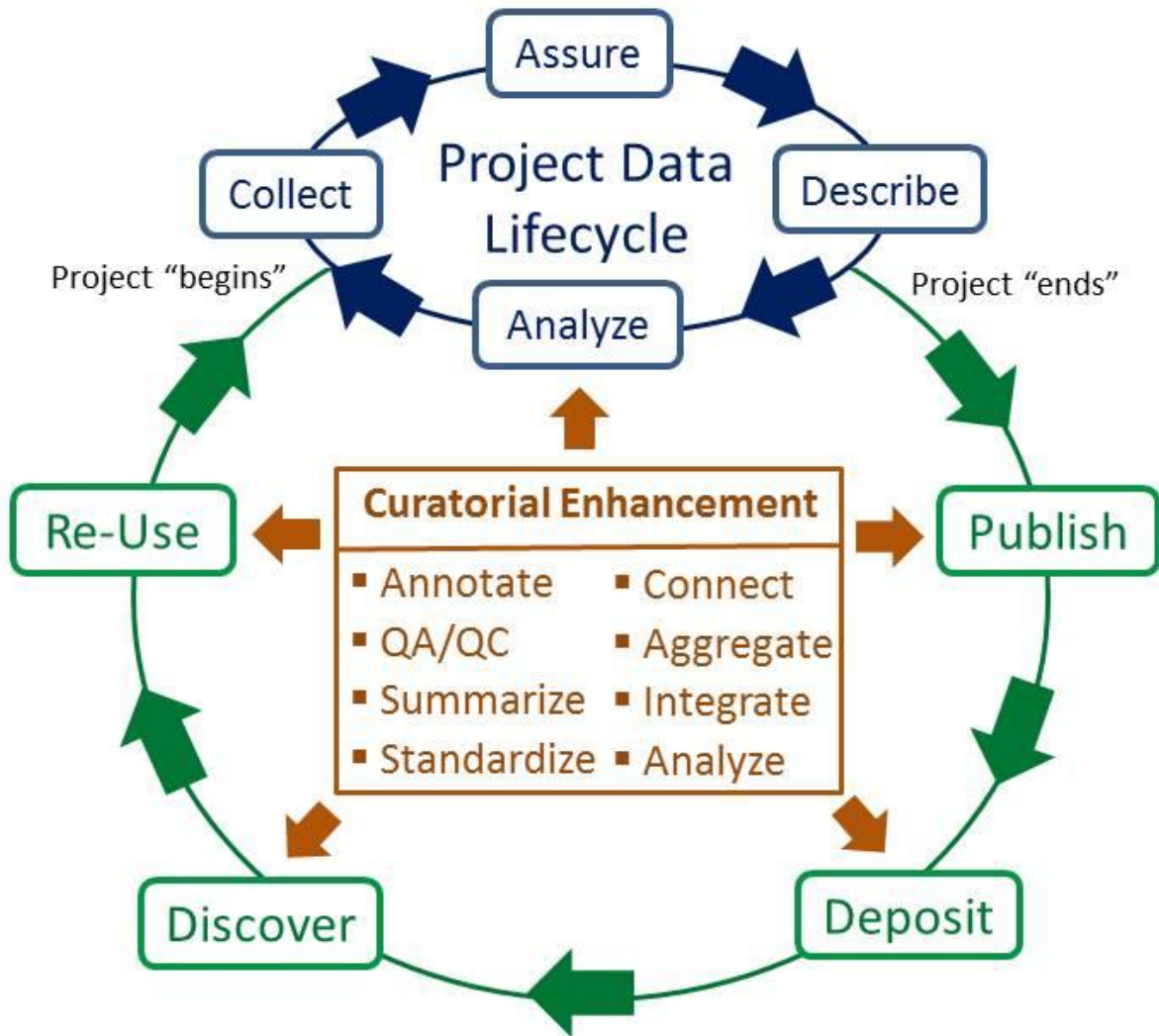


What are your concerns about the new DMSP?

① Start presenting to display the poll results on this slide.

Writing the DMSP:

Think about the
whole data
lifecycle



NIH DMSP Template

NIH has a draft that covers the following categories:

- **Element 1: Data type** (*type, amount, which data will be preserved/shared, metadata and documentation*)
- **Element 2: Tools, software, and code**
- **Element 3: Data standards**
- **Element 4: Data Preservation, access, timelines for sharing and preservation, and repository selection**
- **Element 5: Factors affecting access, distribution, or reuse of scientific data, controlled access and privacy considerations**
- **Element 6: Institutional compliance, monitoring, and roles**

NIH DMSP Template Element 1: Data Type

Summarize the scientific data necessary to validate your findings. List or create a table to describe the datasets that will be created or used as part of the study, including:

- Data type, format, size, and number of files (estimate quantities as necessary).
- Which datasets will be shared.
- The level of aggregation, de-identification, or processing/cleaning that will be done prior to sharing.
- The source of any secondary data, previously collected data reused in this project.
- List the metadata and other documentation (e.g. a README file) that will be shared with your data to facilitate interpretation.

NIH DMSP Template Element 1: Data Type example

1. Types and amount of scientific data expected to be generated in the project:

Summarize the types and estimated amount of scientific data expected to be generated in the project.

Our genomic study will be registered with dbGaP, and our raw whole genome sequencing data and derived data will be submitted to the NIMH Data Archive (NDA). Phenotypic and clinical data for all 500 research subjects will be collected and deposited in NDA using the data dictionaries available in NDA (described below).

2. Scientific data that will be preserved and shared, and the rationale for doing so:

Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

All raw and processed genomics files and all clinical and phenotypic data will be shared.

3. Metadata, other relevant data, and associated documentation:

Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

The Institutional Certification will be submitted to NIH during the dbGaP registration process once we have been told that a grant award is likely. Within the first six months following the award, we will submit the Data Submission Agreement to NDA and will create the Data Expected list in our new NDA Collection. A brief study protocol will also be submitted to NDA and will be made freely available.

NIH DMSP Template Element 2: Tools, Software & Code

Identify tools, software, and/or code necessary to access or manipulate the shared data or that will be developed as part of the research

- State whether or not specialized tools are needed. For each tool that is necessary, list:
 - Version number and operating system
 - How the tools can be accessed (i.e., open source and freely available, generally available for a fee in the marketplace, etc.
 - How long they will be available

NIH DMSP Template Element 2: Tools, Software & Code example

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Genotypic data undergo an extensive automated data cleaning process in the laboratory. Our replication plan for observed associations is outlined in the Research Strategy. While all sequencing data from this proposal will be generated using Illumina pipelines, differences in read depth and primer libraries between studies will require joint re-calling of all genotypes from raw read files to yield the highest possible quality calls and a harmonized dataset for future use in follow-up and unrelated studies. Using the Broad Institute's Genome Analysis Toolkit (GATK), we will apply standard Best Practices workflows for single nucleotide variant (SNV) and Indel discovery from whole genome sequence alignment files (SAM/BAM). These steps should ensure that final association results are representative of "true" genotypes rather than miscalls or confounded genotypes that are unlikely to replicate in independent populations.

NIH DMSP Template Element 3: Data Standards

List the standards that will be used for sharing the data and metadata.

State whether or not there are data standards for your field that are applicable to your project.

Typical data standards include:

- Metadata schemas
- Standard Terminologies (Controlled Vocabulary and Ontologies)
- Content/ Encoding Standards
- Common Data Elements
- Identifiers (PIDs)

NIH DMSP Template Element 3: Data Standards example

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and resources and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Example 1: Formal standards for 256 channels EEGs data have not yet been widely adopted. However, our data and other materials will be structured and described according to best practices. Data will be stored in common and open formats, such as CSV and JPEG files for our 256 channels EEGs and fMRI images. Information needed to make use of this data will be recorded in data dictionaries that will be accessible to the research team and will subsequently be shared alongside the final datasets.

Example 2: All genotype-phenotype data and gene function data will be annotated using ontologies such as the Human Phenotype Ontology and the Gene Ontology. Curation rules will be validated using inter-annotator consistency measures higher than 90% across 3 curators for at least a sample of 15 data entries for each annotation type. Finally, we will adhere to standards such as the GA4GH Phenopackets and the GO GAF format for distribution.

Goldilocks approach to standards



Legend

- Indirectly relevant for
- Directly relevant for

Lesson

- Lesson 1. Credit any derived content using its original identifier
- Lesson 2. Help local identifiers travel well: document prefix and patterns
- Lesson 3. Opt for simple, durable web resolution
- Lesson 4. Avoid embedding meaning, or relying on it for uniqueness
- Lesson 5. Design new identifiers for diverse uses by others
- Lesson 6. Implement a version-management policy
- Lesson 7. Do not reassign or delete identifiers
- Lesson 8. Make URIs clear and findable
- Lesson 9. Document the identifiers you issue and use
- Lesson 10. Reference and display responsibly

Types of actors

Designers & creators	Providers & maintainers	Reusers & referencers
●	○	●
●	●	●
●	○	○
●	●	○
●	●	○
●	●	○
●	●	○
●	●	●
●	●	●

bit.ly/id21c-plosbio

NIH DMSP Template Element 4: Data Preservation & Repositories

Provide details and timelines for sharing and preserving data for long term usability.

- Name the repository(ies) where data will be archived:
 - If a particular metadata standard is required, list in the standards section.
 - A specific NIH repository may be required in the funding opportunity announcement.
- Specify which type of unique identifier is used by the repository (DOI, handle, ID number, accession number) (Note- an identifier is not required at time of DMS plan submission).
- Revisit your data list from section 1 and state when the data will be made available (portions of the data may be released at different times). Timelines required by the policy are:
 - Data will be made available when the work is published or the award/support period ends (whichever comes first); OR
 - Data will be made available earlier.
- State the minimum number of years data will be available, based on repository policies.

NIH DMSP Template Element 4: Data Preservation & Repositories example

A. Repository where scientific data and metadata will be archived:

Provide the name of the repository(ies) where scientific data and metadata arising from the project will be archived; see [Selecting a Data Repository](#).

All data will be deposited to NDA (NIMH Data Archive) starting 12 months after the award begins and will be deposited every six months thereafter following the usual NDA data submission dates.

NIH DMSP Template Element 4: Data Preservation & Repositories example

B. How scientific data will be findable and identifiable:

Describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools.

Data will be findable for the research community through the NDA collection that will be established when this application is funded. In addition, the dbGaP study, which will point to NDA, will help researchers find the data. For all publications, an NDA study will be created. Each of those studies is assigned a digital object identifier (DOI). This data DOI will be referenced in the publication to allow the research community easy access to the exact data used in the publication.

NIH DMSP Template Element 4: Data Preservation & Repositories example

C. When and how long the scientific data will be made available:

Describe when the scientific data will be made available to other users (i.e., no later than time of an associated publication or end of the performance period, whichever comes first) and for how long data will be available.

The research community will have access to data at the end of the grant award or when a publication has been submitted. Once the data are submitted to NDA, that archive will control the long-term persistence of the data set. Currently, NDA has no process for deleting or retiring data sets.

Identifying a data repository

NIH has much documentation for selecting a repository to sustainably host your data:

- <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-016.html>
- https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
 - Open NIH-supported **domain-specific repositories** that house data of a specific type or related to a specific discipline;
 - Other NIH-supported **domain-specific resources**, including repositories and knowledge bases, that have limitations on submitting and/or accessing data; and
 - **Generalist repositories** that house data regardless of type, format, content, or subject matter.

NIH DMSP Template Element 5: Access & Distribution

Describe how sharing will be maximized while respecting restrictions.

- Describe any considerations that may affect the extent of data sharing:
 - Legal
 - Technical
 - Ethical
- Consider whether data can be shared with access controls or, if there are intellectual property concerns, an embargo period, rather than refraining from sharing altogether
- If you have human subjects data, describe how you will protect the privacy, rights, and confidentiality of study participants (de-identification, etc.).

NIH DMSP Template Element 5: Access & Distribution example

Example 1: The two existing data sets from NDA used consents that allow broad data sharing. The new dataset to be uploaded to NDA also was collected using informed consent terms that allow broad data sharing. Access to data housed by the NDA requires the completion of a Data Use Certification (<https://nda.nih.gov/faq.html#dac.3>), which prohibits any redistribution or attempts to re-identify research participants.

Example 2: Organizations can operate as data contributors or data users or both; contribution is not required for use. For contributing organizations, the first step is a Data Transfer Agreement (DTA) which is executed between the National Center for Advancing Translational Sciences (NCATS) and the contributing organization (and its affiliates where applicable). For organizations using data, a separate, umbrella/institute-wide Data Use Agreement (DUA) is executed between organizations and NCATS. Interested investigators submit a Data Use Request (DUR) for each project proposal, which is reviewed by a Data Access Committee (DAC). The DUR includes a brief description of how the data will be used, a signed User Code of Conduct (UCoC) that articulates fundamental actions and prohibitions on data user activities, and if requesting access to patient-level data a proof of additional institutional review board (IRB) approval.

NIH DMSP Template Element 6: Institutional Compliance & Monitoring

Oversight of Data Management and Sharing

Identify who will be responsible for plan compliance and oversight.

- List names and titles/roles of everyone who will be responsible for monitoring compliance with the data management plan and updating it as needed.
- State how often compliance with the data management plan will be verified (e.g. every ____ months, on the first of each month, etc.)

NIH DMSP Template Element 6: Institutional Compliance & Monitoring example

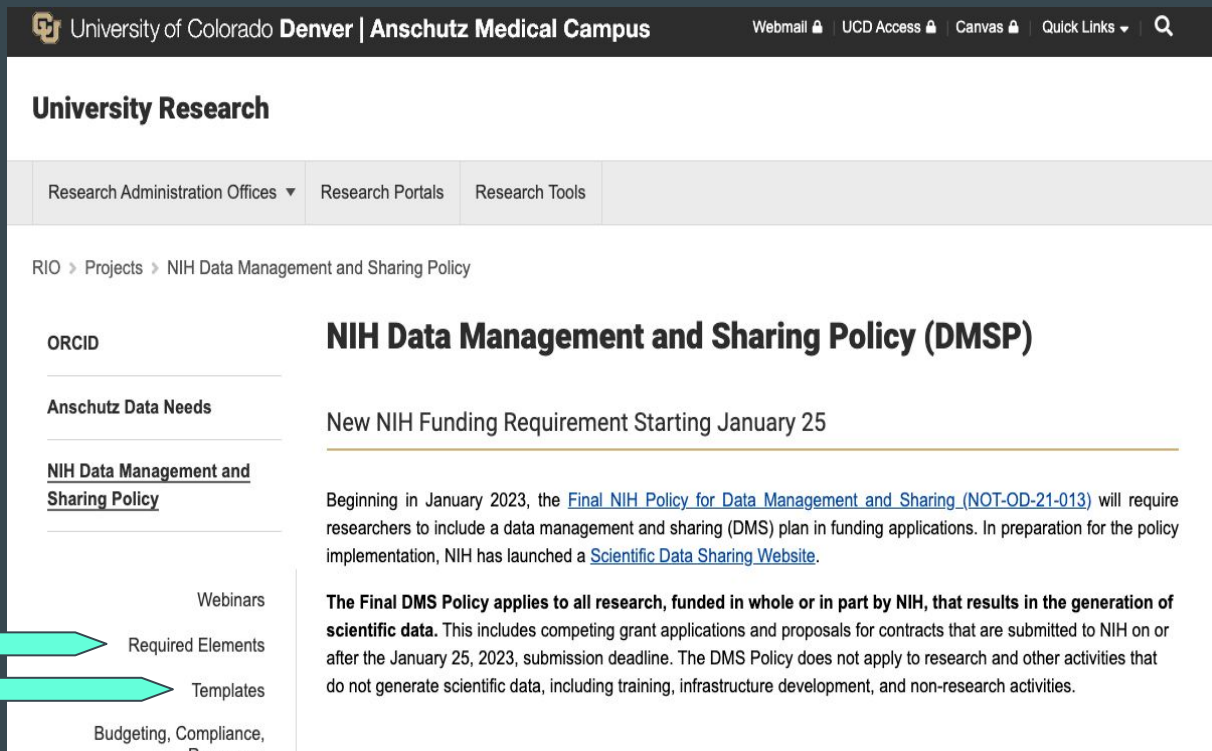
Identify who will be responsible for plan compliance and oversight.

The Research Office has created a data management and sharing plan compliance system as part of their process for submitting the annual NIH progress report. That Office is collecting information related to the number of research participants that are deposited each reporting year. For this award, all of the data will be uploaded in the first year, so the data deposition oversight will end then. The Office of Sponsored Programs will look for the NDA data DOIs when papers are published and will include that information in the annual progress report.

Validation Schedule (this section is required by NIH)

Since this is a secondary data analysis application, validation of newly collected data will not occur. The new data to be deposited to NDA will go through their validation tool when the data are initially uploaded.

Find more information at the CU-AMC RIO Website



University of Colorado Denver | Anschutz Medical Campus

Webmail | UCD Access | Canvas | Quick Links

University Research

Research Administration Offices ▾ | Research Portals | Research Tools

RIO > Projects > NIH Data Management and Sharing Policy

NIH Data Management and Sharing Policy (DMSP)

New NIH Funding Requirement Starting January 25

Beginning in January 2023, the [Final NIH Policy for Data Management and Sharing \(NOT-OD-21-013\)](#) will require researchers to include a data management and sharing (DMS) plan in funding applications. In preparation for the policy implementation, NIH has launched a [Scientific Data Sharing Website](#).

The Final DMS Policy applies to all research, funded in whole or in part by NIH, that results in the generation of scientific data. This includes competing grant applications and proposals for contracts that are submitted to NIH on or after the January 25, 2023, submission deadline. The DMS Policy does not apply to research and other activities that do not generate scientific data, including training, infrastructure development, and non-research activities.

Webinars

Required Elements

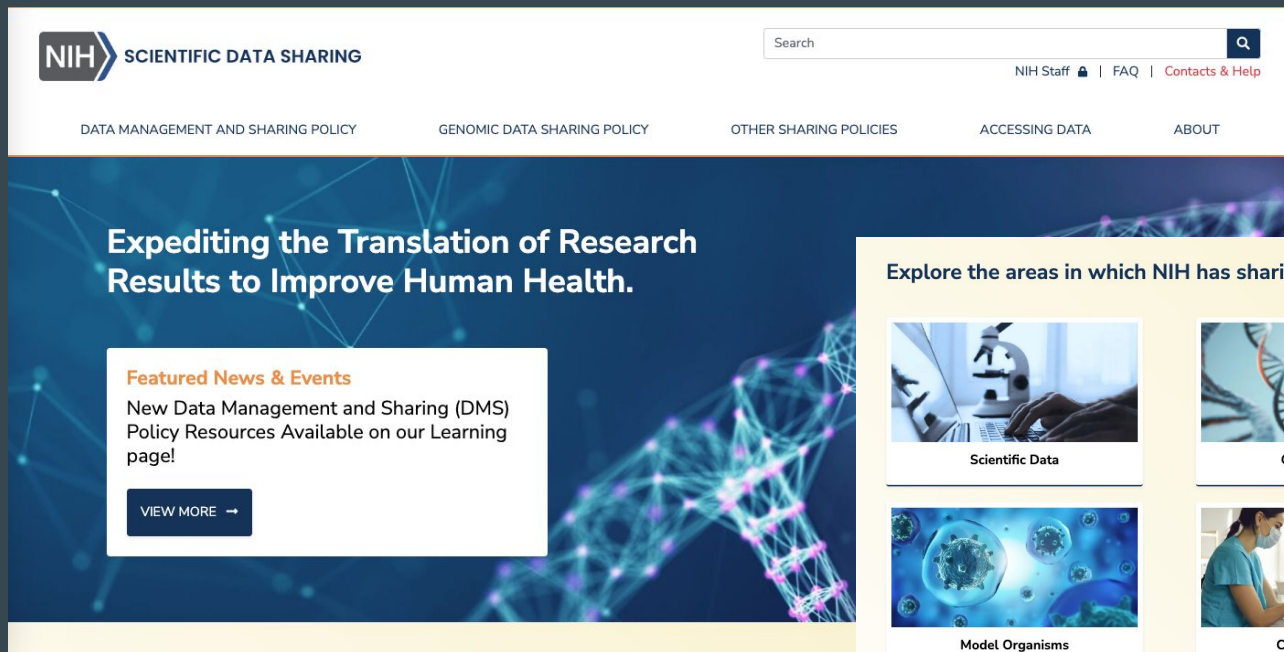
Templates

Budgeting, Compliance, Resources

- ✓ Webinars
- ✓ Examples and templates
- ✓ Data Management best practices
- ✓ Repository selection
- ✓ Budgeting information

<https://research.cuanschutz.edu/rio/projects/ni-h-data-management-and-sharing-policy>

Find more information at NIH (linked from CU-AMC site)



The screenshot shows the NIH Scientific Data Sharing website. The header includes the NIH logo, the text "SCIENTIFIC DATA SHARING", a search bar, and links for "NIH Staff", "FAQ", and "Contacts & Help". Below the header is a navigation bar with links for "DATA MANAGEMENT AND SHARING POLICY", "GENOMIC DATA SHARING POLICY", "OTHER SHARING POLICIES", "ACCESSING DATA", and "ABOUT". The main content area features a large blue banner with the text "Expediting the Translation of Research Results to Improve Human Health." and a "Featured News & Events" section with the text "New Data Management and Sharing (DMS) Policy Resources Available on our Learning page!" and a "VIEW MORE" button.

NIH SCIENTIFIC DATA SHARING

Search

NIH Staff | FAQ | Contacts & Help

DATA MANAGEMENT AND SHARING POLICY GENOMIC DATA SHARING POLICY OTHER SHARING POLICIES ACCESSING DATA ABOUT

Expediting the Translation of Research Results to Improve Human Health.

Featured News & Events

New Data Management and Sharing (DMS) Policy Resources Available on our Learning page!

VIEW MORE →

<https://sharing.nih.gov/>

Explore the areas in which NIH has sharing policies.



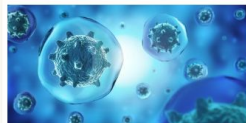
Scientific Data



Genomic Data



Research Tools



Model Organisms



Clinical Trials ↗



Research Publications ↗

Not sure where to start?

Find which policies apply to you

<https://research.cuanschutz.edu/rio/projects/nih-data-management-and-sharing-policy>

Steps being taken to ready our campus

- Best practices in selecting public data repositories for data dissemination, either generalist or specific data types, such as those on the [NIH website](#)
- A library of templates and examples for different data types and institutional language blurbs to put in the specific sections of the DMSP
- An inventory of campus locations where data can be managed
 - Cost structures to support large and small data management needs
 - Budget documentation for infrastructure and personnel
- Compliance monitoring and support for grant applications and RPPR compliance declaration
- Concierge help desk to support researchers and pre-award staff to write quality DMSPs
- New researcher navigational resources

Coming soon: a new front door for Anschutz Research

To help researchers navigate all things research data and IT



 **Research Concierge Portal**

 Help



 Search Resources

Browse Resources 



LEARN



PROPOSE



CONDUCT



CONNECT

Find Health Research Resources and Support

To browse content, select the circle on the left that best fits your current project goals. You can also search content or request a consult by using buttons in upper right.

Centralized ticketing & concierge services to support researchers' data management and sharing needs

Entering Slido

Please go to slido.com



Join the meeting at: #DMSPatCU

Third poll

slido



What would you like to see from Campus to help you in preparing for the new DMSP and sharing your data?

① Start presenting to display the poll results on this slide.



DMSP checklist:
<https://osf.io/awpyt>

DMSP Working Group Members

College of Nursing

Teri Hernandez

Library

Nina McHale

Hannah Pollard

Office of Grants & Contracts

Amy Gannon

Thomas Keith Lii

Garrett Steed

RIO

Steven Andrews

Carlos Goncalves

RIO + Health Data Compass

Ian M Brooks

School of Pharmacy

Heather Anderson

School of Public Health

Katerina Kechris

School of Medicine

Christine Childs

James Costello

Kaitlin A Gil

Jennifer Kemp

Tinalyn Kupfer

Amy Nickson

Sara Rotz

Jana Smilanich-Rose

**Join
Us!**

Connect via email:
crio@cuanschutz.edu

CRIO Council Members

Melissa Haendel *Chief Research Informatics Officer*

Alison Lakin *Associate Vice Chancellor for Regulatory Compliance*

Bruce Dye *Director Dental School*

Dallas Martin *Associate Counsel in the CU Innovations Office*

Heather Anderson *Associate Professor School of Pharmacy*

Jameel Mallory *Program Director, Office of Diversity, Equity, Inclusion*

James Costello *Associate Professor SOM*

Sara Deakyne Davies *Director of Research Informatics and Data Science, CHC*

Teri Hernandez *Professor and Associate Dean for Research*

Tom Campbell *Professor, Medicine-Infectious Disease*

Katerina Kechris *Associate Director of Data Science*

Laura Morris *Associate Vice Chancellor and Chief Information Officer*

Michael Miller *Director of Information Services*

Tom Yager *Director CCTSI*

Our institution has three finalists for the DataWorks! Open Data Prize, which is a testament to what can be achieved when data is shared well.

Please vote for one of our programs:

The Monarch Initiative, which aims to harmonize genotype-phenotype data to aid rare disease diagnostics

The National Covid Cohort Collaborative, which provisions harmonized EHR data from across the nation for public health research

Removing metadata barriers to promote data reuse, which aims to address unstructured/missing metadata that are barriers to reusing omics data

herox.com/dataworks





Additional Questions?



University Research

UNIVERSITY OF COLORADO DENVER | ANSCHUTZ MEDICAL CAMPUS

RIO | Research Informatics Office

Reach out to us! We can help.
CRIO@cuanschutz.edu

<https://research.cuanschutz.edu/rio>

