

Building data services: data management & sharing plans

Shanda Hunt

Research data education & outreach librarian

University of Minnesota

Slides: z.umn.edu/NCDS_DMSP

Where I'm coming from

- Data Repository for the University of Minnesota (DRUM)
 - Human participant & health sciences data curator
 - DRUM in existence for 11 years
- Research Data Services (RDS)
 - Lead educator & outreach coordinator
 - Large suite of services
 - Large network - unofficial at this time
 - RDS in existence for 11 years
- Data Management & Sharing (DMS) Plans
 - 58 DMS plan consultations/reviews
 - 28 presentations to campus groups

Why we're here today

Final NIH Policy for Data Management and Sharing

Notice Number: NOT-OD-21-013

Effective Date: January 25, 2023

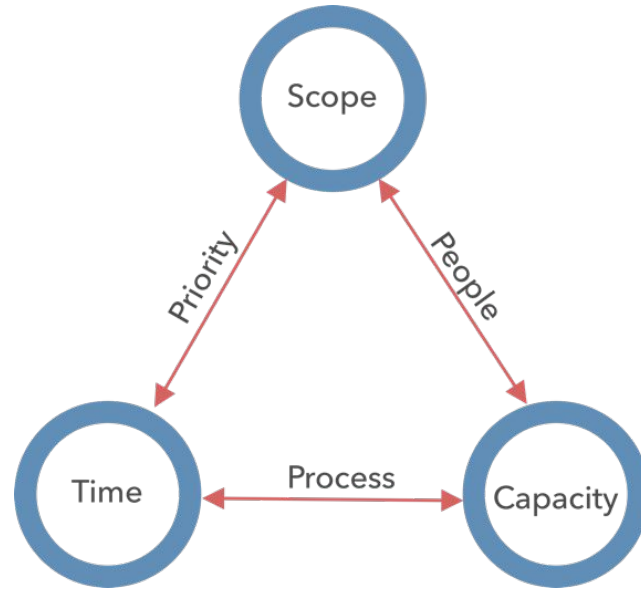
And supplements to the policy:

- Supplemental Information to the NIH Policy for Data Management and Sharing: Responsible Management and Sharing of American Indian/Alaska Native Participant Data
- Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data

And specifically, data management & sharing (DMS) plans!

Service considerations

- Capacity + resources
- Scope
- Sustainability
- Support



Agenda

1. Presentations on DMS plans
2. Review of DMS plans
3. Implementation of DMS plans
4. Marketing & assessing services



Presentations on DMS plans



Present

Elements of a Data Management & Sharing Plan

1. Description of the data plus metadata and documentation
2. Related tools, software, code
3. Standards for the data/metadata
4. Data preservation, access, and associated timelines
5. Access, distribution, and reuse considerations
6. Oversight of data management and sharing

Start with the [NIH DMS plan template](#)

Element 1: Data type

Type and amount of data that will be collected or used

- Data modality (e.g. imaging, genomic, survey)
- Level of aggregation (e.g. individual or aggregated)
- Degree of data processing (e.g. raw vs processed data)
- Anticipated file and/or sample size
- File types (e.g., CSV, BCL sequence, etc.)

Element 2: Related tools, software, and/or code

- Specialized tools that are needed to access the shared data
- Names of specific software tools (e.g. Python, SPSS, etc.)
- Availability of tools (e.g. open source vs proprietary)
- Expected lifespan of the tools compared to length of data availability

Element 3: Standards

- Data formats (e.g., disciplinary, open source file formats)
- Metadata standards (e.g., repository metadata standards)
- Instruments / collection methods (e.g., [NIH Common Data Elements](#))
- Identifiers (e.g., global unique identifiers or accession numbers)
- Indicate when no consensus standard exists

Element 4: Data preservation, access, and associated timelines

- Name of repository where data will be shared
- How the data will be findable and accessible (dependent on the data repository; e.g., assigned DOI or accession number)
- When data will be shared (e.g., “at time of manuscript publication or end of performance period, whichever comes first”)
- How long data is anticipated to be available (dependent on the data repository; look for preservation or deaccessioning policy)

Element 5: Access, distribution, or reuse considerations

- Researcher access (aka controlled access or restricted access) vs public access (aka open access)
- Informed consent
- Privacy and confidentiality protections
- Any other ethical considerations (e.g., politically charged or endangered animal research)
- Restrictions imposed by federal, Tribal, or state laws, regulations, or policies (e.g. HIPAA)

Element 6: Oversight of data management and sharing

- Name who will monitor and manage compliance with the DMS plan (e.g., PI, data manager, research coordinator)
- Name the individual ultimately responsible for oversight (hint: it's the PI)
- Determine how the plan will be monitored
- Determine frequency of check-ins

Common questions from researchers

- Can I choose to not share my data?
- Where should I share my data?
- How do I share qualitative data?
- How do I budget for data sharing?

Customize!

Common additions to our standard presentation

- Finding and evaluating repositories
- Budgeting for data sharing
- Human participant considerations
- Data ownership
- ICO specifics
- Withholding data

Sample slide: finding and evaluating repositories

Step 1

Check your [Notice of Funding Opportunity](#) or with your [NIH Institute, Center, or Office \(ICO\)](#) to determine whether either encourages or **requires** the deposit of data into a specific data repository.

- NIMH → NDA
- NIAID → “publicly accessible repositories”
- NINDS → refers readers back to NIH
- NIDDK → you choose, NIDDK repo, check your NOFO

Look up your funding institute in [NIH Institute and Center Data Sharing Policies](#)

Go to your funding institute’s website and read their instructions

Follow the more specific guidance, if contradictory

Sample slide: finding and evaluating repositories

Features	DRUM	ICPSR	OpenICPSR	Harvard Dataverse	Dryad	Open Science Framework
Public-access sharing	yes	no	yes	yes	yes	yes
Offers controlled/ restricted access	no	yes	yes	yes	no	yes
Who controls access requests	n/a	repository	repository	depositor	n/a	depositor
Allows custom terms of use	no	no	no	yes	no	no
Fee for data deposit	no	no for UMN	no	no	no for UMN	no
Fee for data access	no	for non-members	no	no	no	no
Allows blind peer review	no	no	no	no	yes	yes

<https://libguides.umn.edu/c.php?g=1164012&p=10159871>

Sample slide: finding and evaluating repositories

Ensure where you share meets **desirable characteristics**

- [NIH's list of desirable characteristics for data repositories](#)
- [Office of Science and Technology Policy list of characteristics](#)

Recommended repositories make data **FAIR**

- Findable
- Accessible
- Interoperable
- Reusable



<https://www.go-fair.org/fair-principles/>

Sample slide: budgeting for data sharing

New study on data sharing costs

Realities of Academic Data Sharing (RADS) Initiative

- 6% of total grant award was spent on DMS, which varied by grant award
 - 15% for smaller awards (<\$200,000)
 - 1% for larger awards (>\$1 million)
- \$30,000 was the average cost incurred for DMS for a research project
- Staff time was the largest portion of this expense

[Making Research Data Publicly Accessible: Estimates of Institutional & Researcher Expenses](#)

Sample slide: human participant considerations

Most repositories have **de-identification requirements**

- **Element 5.C. of the DMS Plan** requires that you describe protection of participants including de-identification of the data.
- Be explicit in the process you will use to address direct and indirect identifiers.
- The same information should be included in the consent form.
- The level of de-identification may depend on the level of accessibility of the data repository.

Sample slide: human participant considerations

Be explicit about data sharing practices in the **consent form**

- **Element 5.A. of the DMS Plan** requires that you describe how informed consent will be obtained for data sharing, and if there will be any access restrictions to the data related to consent.
 - How the data will be processed before it is shared, including de-identification methods
 - What data will be shared (and what data will not be shared)
 - Where the data will be shared (name the specific repository)
 - How the data will be accessed (publicly available or restricted to specific requesters)
 - Who will grant access (the repository or the PI)
- **Certificate of Confidentiality** (issued to all NIH awardees) requires explicit consent for data sharing.

Sample slide: data ownership

nih supplement

Responsible Management and Sharing of American Indian/Alaska Native Participant Data

“Tribal Nations have unique rights associated with Tribal sovereignty to control how biomedical research can be performed within Tribal jurisdiction, including how their data can be collected, used, managed, and shared.”

- Acknowledge historical exploitation, data misuse, health disparities, & need for representative data
- Proactively engage Tribes in planning for data management and sharing
- Incorporate AI/AN preferences in DMS plans
- Consider additional protections for future data sharing
- Incorporate DMS plans in the informed consent process

Sample slide: data ownership

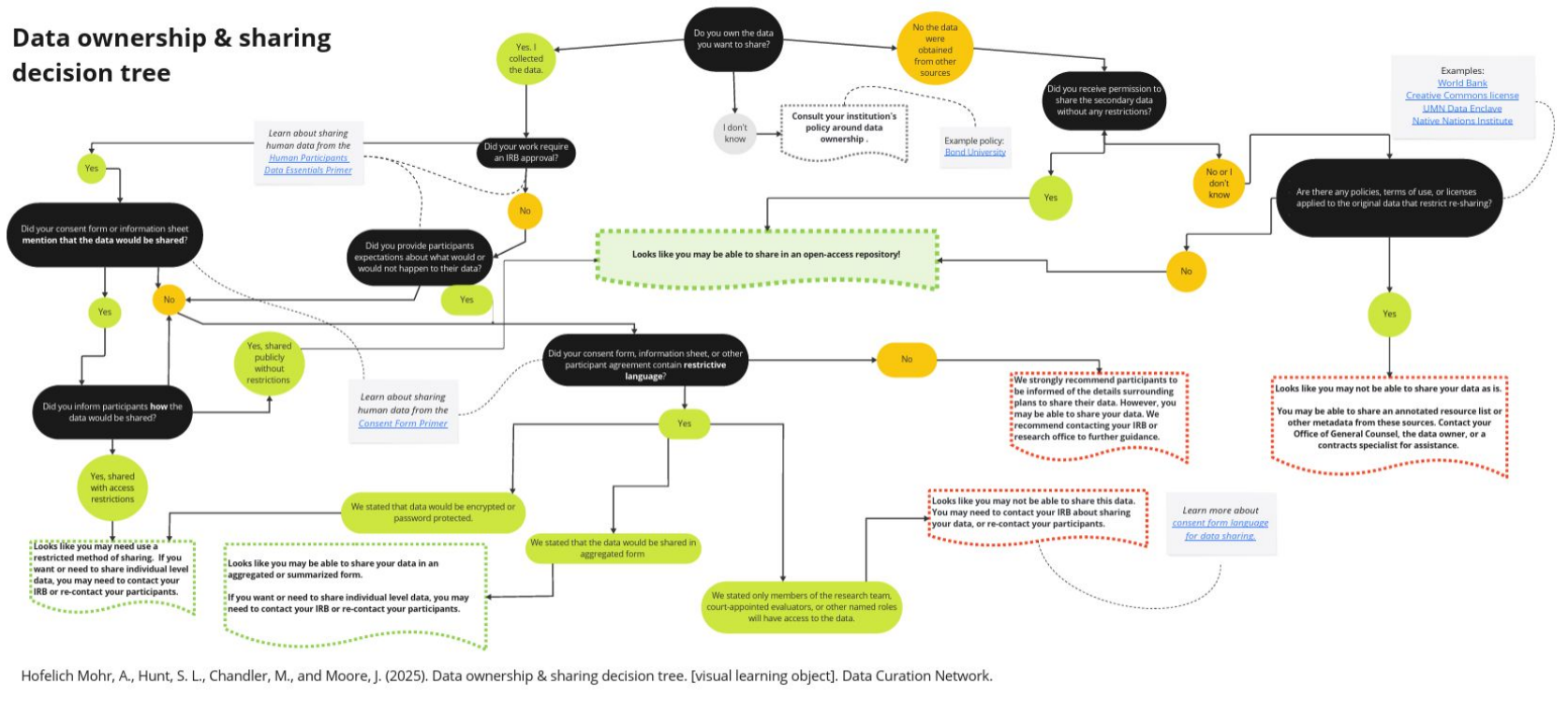
umn policies & guidelines

- [UMN Indigenous research policy](#)
- [UMN guidelines for Indigenous research](#)
- [Research with Indigenous partners library guide](#)

“Obtain approval from the Office of Native American Affairs before beginning any research by filling out the form [Notification of Indigenous Research](#).”

Sample slide: data ownership

Data ownership & sharing decision tree



Hofelich Mohr, A., Hunt, S. L., Chandler, M., and Moore, J. (2025). Data ownership & sharing decision tree. [visual learning object]. Data Curation Network.

Sample slide: ICO specifics

Additional NIMH DMS Plan guidelines

In addition, the Data Management and Sharing Plan must **propose a schedule to validate the quality of the data being uploaded so these data are compliant with the data dictionary or other standards that are being used**. In cases where a data dictionary has been defined in the NIMH Data Archive, the NDA validation tool (<https://nda.nih.gov/tools/nda-tools.html#vt>) should be used. Compliance with the approved data management and sharing plan will become a term and condition in the Notice of Award and will be monitored by the NIMH throughout the duration of the award.

Sample slide: withholding data

NIH acceptable reasons to withhold data



“[A] compelling rationale for limiting scientific data sharing should be provided and will be assessed by NIH.”

- informed consent or existing consent (e.g., in a secondary data analysis) will not permit or will limit the scope or extent of sharing
- privacy or safety of research participants would be compromised or place them at greater risk of re-identification or suffering harm
- explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure
- restrictions imposed by existing or anticipated agreements
- datasets cannot practically be digitized with reasonable efforts

[NIH FAQ B.5.](#)

Sample slide: withholding data

- Provide a clear justification
- Be clear about what WILL be shared
 - Data derivatives
 - Protocols and documentation
 - Statistical code
 - Survey instruments



Hot tip: create your own teaching tools

[NIH Data Management & Sharing Policy 2023 LibGuide](#)

[Research Data Services Good Practices website](#)

[Selecting a repository comparison chart](#)

[Repository selection tool](#) (evaluation checklist)

Activity #1

In the chat, note whether the consent language matches what is in the Data Management & Sharing (DMS) plan

DMS Plan - Element 4.A. Repository where scientific data and metadata will be archived: All data resulting from this study and related documentation will be shared in the Data Repository for the University of Minnesota (DRUM), an open access, publicly accessible institutional repository.

Consent - What will be done with my data and specimens (if applicable) when this study is over? We will use and may share data and/or specimens for future research. Data collected in this study may be made available for others to use, including for future research studies on similar or different topics, teaching, or other purposes. This could include for profit companies. Our goal is to make more research possible. We will not ask for your consent before using or sharing them. We will remove identifiers from your data and/or specimens, which means that nobody who works with them for future research will know who you are.

Mid-point Q&A



Review of DMS plans



Review

Elements of a Data Management & Sharing Plan

1. Description of the data plus metadata and documentation
2. Related tools, software, code
3. Standards for the data/metadata
4. Data preservation, access, and associated timelines
5. Access, distribution, and reuse considerations
6. Oversight of data management and sharing

Start with the [NIH DMS plan template](#)

NIH feedback (NCI and NICHD)

Element 1. Description of the data plus metadata and documentation

- Describe the # of study subjects, # of technical or biological replicates, # of time points, etc.
- Name file formats
- Name metadata to be generated/shared
- Provide sound justification for why data cannot be shared
- Name the species for genetic/genomic data

NIH feedback (NCI and NICHD)

Element 2. Related tools, software, code

- Describe how tools can be accessed
- Note whether the tools proposed are open source (i.e., free of charge for secondary users)

NIH feedback (NCI and NICHD)

Element 3. Standards for the data/metadata

- Describe standards to use for *each* data type for the *entire project*
- If no standards exist for a specific data type, indicate as such

[National Institute of Child Health & Human Development website dedicated to the issue](#)

NIH feedback (NCI and NICHD)

Element 4. Data preservation, access, and associated timelines

- DMS policy specifies “no later than time of an associated publication or end of the performance period, whichever comes first”
- Note that different policies may have different timeline requirements for different levels of data (e.g. genomic data sharing policy [GDS])
- Planning to share in “publications” and “conferences” is unacceptable
- Utilize a discipline specific repository first (problematic: GitHub, PubMed, journals, and generalist non-curated repositories)
- If using multiple repositories, note which data goes to which and describe how they will be linked (the latter is my advice)

NIH feedback (NCI and NICHD)

Element 5. Access, distribution, and reuse considerations

- Address factors that may affect access to data including human subjects protections, legal or policy considerations, licensing agreements, etc.

[National Institute of Child Health & Human Development website dedicated to the issue](#)

NIH feedback (NCI and NICHD)

Element 6. Oversight of data management and sharing

- Find out if your institution has requirements or recommended language for this section

NIH feedback (NCI and NICHD)

Overall

- All data proposed in the Research Plan of the grant applications should be discussed in the DMS plan
- There is confusion between Resource Sharing Plan and Data Sharing Plan
- There is contradiction across Elements
- When multiple data types are proposed, certain elements for each data type were often missing
- Most grant submissions who were subject to the GDS policy did not know and did not mention it
- Weak justifications for not sharing and delaying timelines

[National Institute of Child Health & Human Development website dedicated to the issue](#)

Workflow

Feedback medium

- 2-3 business days
- [DMPTool](#)
 - Funder templates
 - Customizable templates
- Google Drive (and my preference)
 - Interactive
 - Archive of completed DMS plans



Hot tip: choose a file naming convention for plans

If you use Google Drive in your DMS plan review workflow, implement a useful file naming convention early on

- Author last name
- Funder / ICO
- General type of data
- Repository selected

Hunt_NIH_NIAAA_human_NDA

Workflow

Research performed as part of the service

- Data types / expected file formats
- Standards
 - Data formats (e.g. [BAM / FASTQ for genomic data](#))
 - Data dictionaries (e.g. [OSF dictionary elements](#))
 - Collection tools & elements (e.g. NIH [Common Data Elements](#))
 - Identifiers (e.g. [NDA global identifiers](#))
 - Repository standards (e.g. [QDA documentation requirements](#))
- ICO requirements
- Appropriate repositories

Workflow

Just in time (JIT)

- Prioritize
- Shortened timeline
- Responding to NIH feedback

Scope

IMHO: Non-researchers cannot, nor should they, write the DMS plan

- Requires deep expertise & personal knowledge
- Part of a legal contract



Hot tip: remind researchers that DMS plans are living documents

- NIH program staff will review the DMS Plan for acceptability and may request modifications prior to award as appropriate
- Provide updates on data management and sharing activities in annual [progress reports](#)
- If plans change over the course of the project, work proactively with NIH Program Officer to obtain review and approval of modifications

Activity #2

In the chat: what feedback would you provide?

Element 1: Data Type - types and amount

Types and amount of scientific data expected to be generated in the project: The study data will be saved as CSVs. This study will provide quantitative data on a target sample of:

- 500 caregivers of a child ages 3-8. Quantitative surveys pre- and post-program participation and coded audio-recorded responses to vignettes.
- 200 primary care clinicians serving children ages 3-8. Clinician-level data: Quantitative surveys pre- and post-pilot year and coded audio-recorded responses to vignettes.

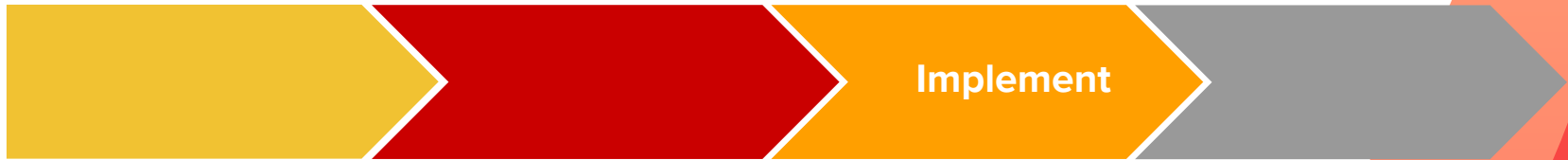
Activity #2

In the chat: what feedback would you provide?

Element 1: Data Type - preserved and shared

Scientific data that will be preserved and shared, and the rationale for doing so: The raw audio files will be stored within a secure computing environment. All direct respondent identifiers (e.g., names and addresses) associated with surveys will be removed and maintained in a secure file for future contact purposes. All other scientific data (de-identified transcribed responses to vignettes and their macro codes, and de-identified raw survey data and scale composites) will be both preserved and shared. Respondent identifiers will not be shared. Demographic data may only be shared at an aggregated level as needed to maintain confidentiality.

Implementation of DMS plans



What does this mean?

- Repositories
 - Curate data for deposit to DRUM
 - Consult on other appropriate repositories - no matter what they wrote in their DMS plan
 - Curate data for deposit to other repositories

What does this mean?

- Defunded and cancelled grants
 - Researchers are obligated to fulfill the grant contract
 - Data must be shared, in full or part
 - Consults around repository selection are crucial and sensitive



Hot tip: know your institutional support structure

- Office of the Vice President for Research (UMN rebranded as the Research Innovation Office)
- Sponsored Projects Administration (grants)
- Office of General Counsel
- Office of Technology & Commercialization
- Storage Council
- Institutional Review Board

Example outreach email

Hello!

I hope your 2025 is off to a great start! Research Data Services is adding a list of our campus data partners to our [homepage](#) as well as our [Data Management & Sharing Plans](#) webpage. Our goal is to continue to de-silo our efforts and alert users to our collaborations, and we see this simple step as one way to accomplish this.

We are curious to learn more about your unit and the services you provide to UMN researchers as we frequently make referrals when researchers approach us with their data questions. And hopefully, we will add you to our list of collaborators.

Are you open to an introductory meeting to learn more about each other?

Thanks for considering.

Shanda Hunt (she/her)
Research Data Education & Outreach Librarian
[Research Data Services](#)

Activity #3

In the Google doc: who will you reach out to?

z.umn.edu/NCDsactivity3

UMN partnerships

Do you have similar offices? What are they named?

Who are the major players on your campus?

Names / acronyms?

What do they do, exactly?

Marketing & assessing services



Marketing

- Rely on your data partners list
 - Websites
 - Newsletters
- Take advantage of existing campaigns
 - Love Data Week - February
 - Research Ethics Week - March (UMN)
 - Data Management Camp - fall (UMN)

Marketing

- Customize messaging to every outlet
 - Audience
 - Current events / Critical information

The federal open research policy landscape is quickly shifting. Did you know that you must share your data in compliance with a data sharing plan even if your funding agency cancels your grant?

Join the UMN Libraries and LATIS for a series of virtual workshops to build and enhance your data management strategies. This year, three of the six workshops will focus specifically on federal mandates and data sharing.

Learn more: z.umn.edu/dmbootcamp.

Assessment

- Tracking log
 - Researcher
 - Department / group
 - Data services staff
 - Date
 - Funding agency
 - Link to Google doc
 - JIT or no

Assessment

- Scalability
 - What you can handle vs when to bring others in
 - Requests
 - Activities
 - Support system
- “Success” rate
 - Grants funded
 - Funder feedback on DMS plans

Assessment

- Contingency plan
 - What could happen?
 - How will you respond?
 - Who will support you?

Final Q&A



Thank you

Shanda Hunt

hunt0081@umn.edu

data@umn.edu

Slides: z.umn.edu/NCDS_DMSP