National Institute on Aging

Developing a Data Management and Sharing Plan for Behavioral and Social Research

June 15, 2023

James McNally, PhD
james.mcnally2@nih.gov
NIA Division of Behavioral and Social Research

Becky Krupenevich, PhD
rebecca.krupenevich@nih.gov
NIA Division of Behavioral and Social Research

NIH Definition of Scientific Data

Scientific data are 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*. Scientific data *do not include* laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens.

What *should* be shared:

- Enough data to validate and replicate findings
- Data resulting from the study even if they does not support a publication
- Null findings that do not result in publication

What *should not* be shared:

- Lab notebooks
- Drafts of papers
- Physical objects, such as biological specimens
- Prelim analyses
- Case reports

Acceptable reasons to not share:

- Existing informed consent limitations
- Privacy or safety of participants would be compromised or risk re-identification
- Federal, state, local, or Tribal law, regulation, or Policy prohibiting disclosure
- *International legal restrictions on data sharing*
- Data cannot practically be digitized

National Institute on Aging

Frequently Asked Questions

FAQ: Research Centers and Pilot Projects

Which DMS policy is applicable to pilot applications for cooperative agreements or center grants that were awarded before Jan 25, 2023?

- The original data sharing plan applies to any pilot studies that may be awarded later

FAQ: Secondary Research and Primary Data

Researchers should share any new, derived data generated from secondary data analysis

Source code should be shared in a code repository (e.g., extraction, recodes, transformation)

Primary **data already shared** should **not** be shared again

Informed Consent for Secondary Research with Data and Biospecimens

FAQ: Where should I archive my data?

NIA/BSR Recommended Repository

- National Archive of Computerized Data on Aging (NACDA)

NIA/NIH-Supported Repositories

- NIA-Supported Repositories
- Filterable List of NIH-Supported Repositories

Other Repository Resources

- Generalist repositories
- Nature's Data Repository Guidance
- Registry of Research Data Repositories

FAQ: What about Institutional Repositories?

- Institutional repositories are an acceptable place to archive data if data/resources are **discoverable and accessible**
- Do they have a DOI or similar recognized identifier?
- Can researchers outside the institution access the data?

Desirable Characteristics for All Data Repositories

- ✓ Unique Persistent Identifiers
- ✓ Long-Term Sustainability
- ✓ Metadata
- ✓ Curation and Quality Assurance
- ✓ Free and Easy Access
- ✓ Broad and Measured Reuse
- ✓ User Support
- ✓ Clear Use Guidance
- ✓ Security and Integrity
- ✓ Confidentiality
- ✓ Common Format
- ✓ Provenance
- ✓ Retention Policy
- ✓ Discoverability

Repository Considerations for Human Data

- ✓ Fidelity to Consent
- ✓ Restricted Use Compliant
- ✓ Privacy
- ✓ Plan for Breach
- ✓ Download Control
- ✓ Violations
- ✓ Request Review

Biorepositories represent a special case, as their "data" have supply and demand issues. Specimens cannot be replicated endlessly like digital content.

FAQ: How to Share Data from Human Participants?

- De-identify to the greatest extent while maintaining scientific utility; Use Common Rule (45 CFR 46) and HIPAA Privacy Rule standards
  - *Consider risks from information even when de-identified*
  - *Share identifiable data only with explicit consent*

🔗 Informed Consent for Secondary Research with Data and Biospecimens

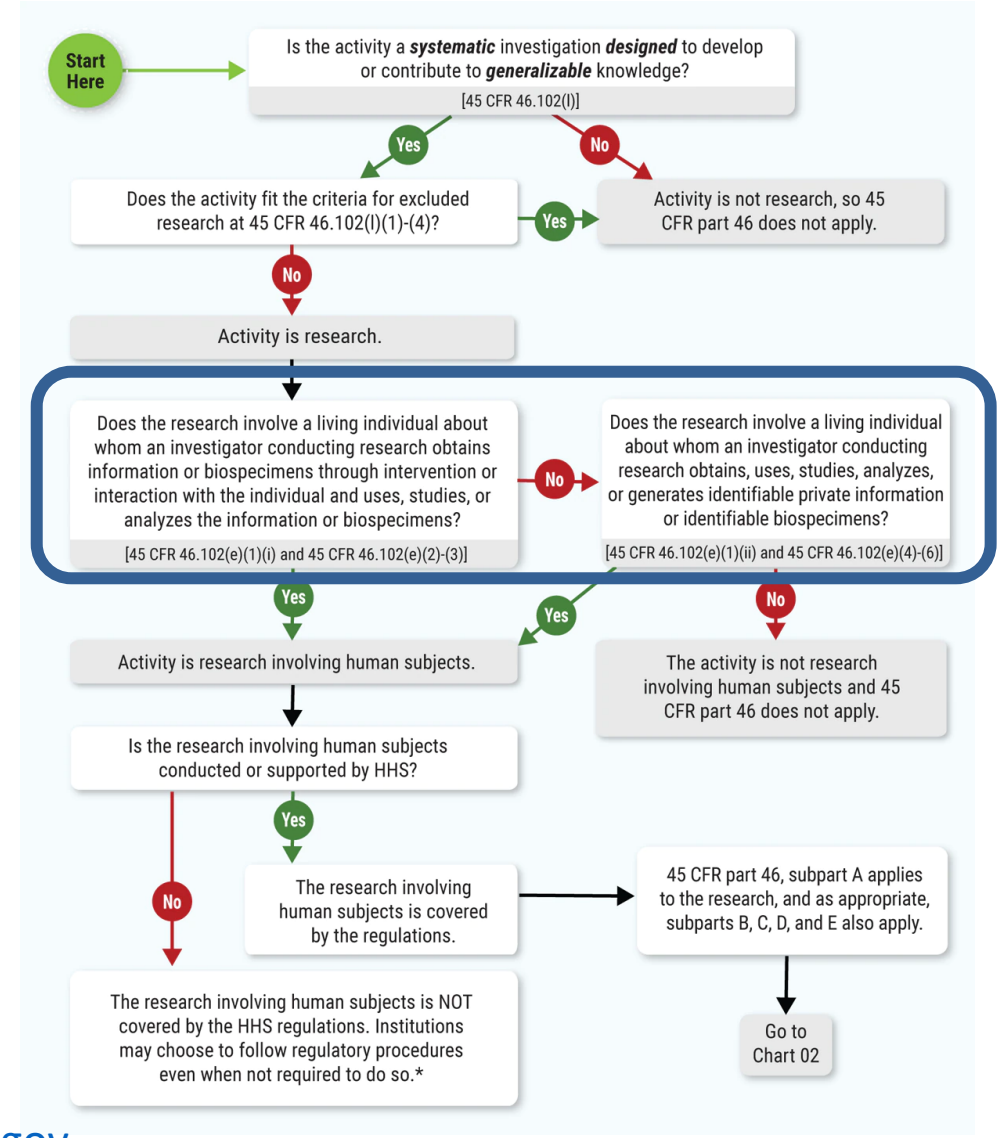
- Understand legal protections and limitations on disclosure

- Just because data on humans is collected does not mean it is "human subjects" data when managing and sharing

FAQ: What is considered human subject data?

THIS IS THE PERTINENT CLAUSE

- Original data collection with your local IRB. Human subjects research applies when interacting with respondents, subjects, or groups
- 45 CFR 46 is no longer applicable when sharing de-identified data resulting from a project IF no attempt is made to contact the original respondent



Human Subject Regulations Decision Charts: 2018 Requirements | HHS.gov

How to Share HIPAA Protected Data

- Under HIPAA's Expert Determination standard, researchers can employ advanced statistical or computational methods to de-identify data and maintain privacy

- Additional resources to assist with de-identification:
  - Guidance on Expert Determination and Safe Harbor (DHHS)
  - Tools for de-identification (NIST)
  - Guidance on images (National Cancer Imaging Archive)
  - Clinical text de-identification tool (NLM-Scrubber)

FAQ: What about grants that Include CMS Data?

1) **Applicant/Grantee requests CMS data and enters into a data-use agreement**
   **Applicant/Grantee MAY NOT re-share CMS data or any derived data

2) **Applicant/Grantee should plan to:**

- Share enough information about which CMS datasets were requested that another investigator would be able to request the same data
- Share analysis codes

National Institute on Aging

Developing a Data Sharing Plan

Elements of a Data Sharing Plan

NOT-OD-21-014

Description of data

Related tools, software, and codes

Standards

Data preservation, access, and timelines

Access, distribution, and reuse considerations

Oversight of data management and sharing

Element 1

Data Description

A. What type and amount of data will be collected?

- Data modality, estimated number of participants/trials

B. What data will be shared?

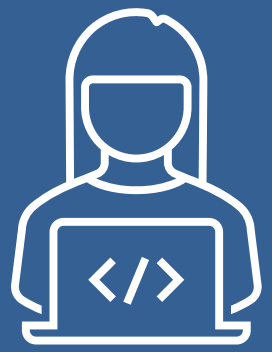
- Level of aggregation, degree of processing
- Include rationale for data that are not shared

C. What metadata and documentation are included?

- Study protocols, codebooks, data collection instruments

Element 2

Related Tools, Software and/or Codes



Are specialized tools or software needed to access shared data?

What are the tools or software?

How can they be accessed?

Element 3

Standards

What standards will be applied to data and metadata?

(or indicate that standards do not exist)

Additional Information about Metadata Standards

- Metadata is defined as "Data about Data"
- One of the challenges when referring to metadata in a DMS plan is that there is no one metadata standard.
- There are dozens of standards currently in use, normally geared to specific kinds of data or disciplines.
- Most of these standards do not "talk" to each other so cross-referencing and creating interoperability standards will be something to keep in mind when developing a DMS plan.

Name	Name	Name	Name
DDI	EML	METS	ISO/IEC 19506
EBUCore	IEEE LOM	MODS	ISO 23081
EBU CCDM	CSDGM	MADS	MoReq2010
FOAF	ISO 19115	XOBIS	DIF
EAD	e-GMS	PBCore	RAD
CDWA	GILS	MPEG-7	RDF
VRA Core	TEI	MEI	MDDL
Darwin Core	NISO MIX	Dublin Core	NIEM
ONIX	<indecs>	DOI	SAML
CWM	MARC	ISO/IEC 11179	

This table lists almost 40 standards routinely used by data curators in the US and internationally.

Non-exhaustive list of metadata standards and their applications

Element 4

Data Preservation, Access, and Timelines

A. Where will data be archived?

- Provide the name of the repository

B. How will data be findable?

- Provide DOI or other standard indexing tools

C. When and how long will data be available?

- Provide appropriate timeline

(NIH Guidance: no later than the time of an associated publication or end of performance period, whichever comes first)

Element 5

Access, Distribution, or Reuse Considerations

What factors may affect data access?

Will data access be controlled?

How will data privacy rights be protected?

Element 6

Oversight of DMS Plan

How will compliance with DMS plan be monitored and managed?

- Frequency of oversight
- Who is responsible for oversight

Links to Policy and Supplements

- Final NIH Policy for Data Management and Sharing
- Responsible Management and Sharing of AI/AN Participants
- Protecting Privacy When Sharing Human Research Participant Data
- Elements of an NIH Data Management and Sharing Plan
- Allowable Costs for Data Management and Sharing

NIH National Institute on Aging

Questions?

Contact:
James McNally, PhD
james.mcnally2@nih.gov

Extra Slides

Activities Subject to the DMS Policy

APPLIES TO…

all research generating scientific data, including but not limited to:

Research Projects

Certain Career Development Awards (Ks)

Small Business SBIR/STTR

Research Centers

DOES NOT APPLY TO…

research projects <u>not</u> generating scientific data or non-research projects, including but not limited to:

- Training (Ts, R25s)
- Fellowships (Fs)
- Certain non-research Career Awards (e.g., KM1)
- Construction (C06)
- Conference Grants (R13)
- Resources (Gs)
- Research-Related Infrastructure Programs (e.g., S06)

Management and Sharing of AI/AN Participant Data

UNDERSTAND	Understand Tribal sovereignty and laws, regulations, policies, and preferences
ENGAGE	Engage early with Tribes when developing a data management and sharing plan, before research begins, and continue throughout research
ESTABLISH	Establish mutually beneficial partnerships
AGREE	Agree who will manage data (e.g., Tribe, researcher, trusted 3rd party)
CONSIDER	Consider additional protections, as necessary

Responsible Management and Sharing of AI/AN Participant Data: NOT-OD-22-214

Quick Overview: HIPAA
(Health Insurance Portability and Accountability Act)

- HIPAA is Federal law that required the creation of national standards to protect sensitive patient health information

- The **HIPAA Privacy Rule** was issued to implement the requirements of HIPAA
  - Addresses the use and disclosure of individual's health information by entities subject to the Privacy Rule (see next slide)

- The **HIPAA Security Rule** protects a subset of information covered by the Privacy Rule
  - All individually identifiable information a covered entity creates, maintains, or transmits in electronic form (EHRs)

HIPAA IS ABOUT PLACE NOT PERSON

HIPAA Cont'd: Non-Covered Entities and Repositories

- HIPAA regulations are tied to the entity not to the data itself
- Including Data Repositories in the data sharing plan can simplify the sharing of data that may be collected within a HIPAA covered entity
- Data should be de-identified to address human subjects and confidentiality issues
- HIPAA privacy rules can be incorporated in data sharing plan
- HIPAA Limited Data Set presents major barriers to aging research
- HIPAA should not be used as a rational for not sharing data

HIPAA Cont'd: Covered and Non-Covered Entities

- Covered Entity – HIPAA rules apply
- Non-Covered Entity – not subject to the requirements of HIPAA
  - Individual, business, or agency that is NOT a health care provider, health plan , or health care clearinghouse (e.g., Fitbit, health social media apps)
  - Data repositories that are not directly affiliated with medical centers or covered entities are typically free of HIPAA oversite

A Covered Entity is one of the following:

A Health Care Provider	A Health Plan	A Health Care Clearinghouse
This includes providers such as: • Doctors • Clinics • Psychologists • Dentists • Chiropractors • Nursing Homes • Pharmacies ...but only if they transmit any information in an electronic form in connection with a transaction for which HHS has adopted a standard.	This includes: • Health insurance companies • HMOs • Company health plans • Government programs that pay for health care, such as Medicare, Medicaid, and the military and veterans health care programs	This includes entities that process nonstandard health information they receive from another entity into a standard (i.e., standard electronic format or data content), or vice versa.

https://www.hhs.gov/hipaa/for-professionals/covered-entities/index.html

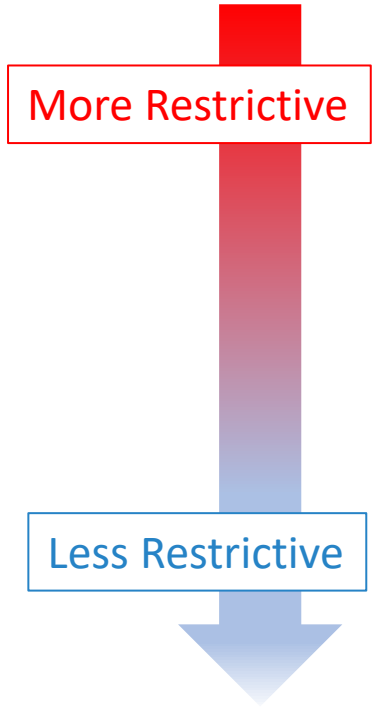
Determining Controlled Access Designation

Researchers should consider a controlled-access repository if data:

- Have specific limitations on subsequent use
- Could be considered sensitive *(such as information about potentially stigmatizing traits, illegal behaviors, or information that could be perceived as causing group harm or used for discriminatory purposes)*
- Cannot be de-identified to established standards or for which the possibility of re-identification cannot be sufficiently reduced

Controlled Access: Data Sharing Agreements

Acceptable Data Sharing agreements come in a variety of forms:



- Material Transfer Agreements (MTA)
- Memorandums of Understanding (MOU)
- Data Use Contracts
- Data Use Agreements
  - One sided agreements
  - Two sided agreements
- Click Through Agreements
- Implied Use Agreements

(Communicate limitations on use, include prohibitions on re-identification or recontact)