


Data Management and Resource Sharing

Rigor & Reproducibility Workshop
17 Oct 2022

Melissa Eitzen, MLS(ASCP), MS, RQAP-GLP
Director, Regulatory Operations
UTMB Institutional Office of Regulated Nonclinical Studies
mmeitzen@utmb.edu

 utmb Health

1



Data Management and Resource Sharing

Topics

- Principles, Guidelines, Policies, Definitions
- Data Lifecycle
 - Data Quality & Integrity
- Case Study—Break out session



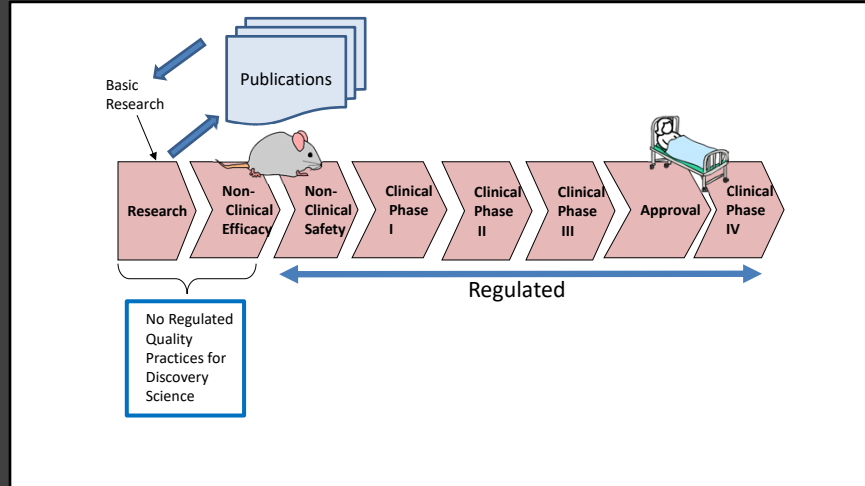
 utmb Health

References provided on slides

2

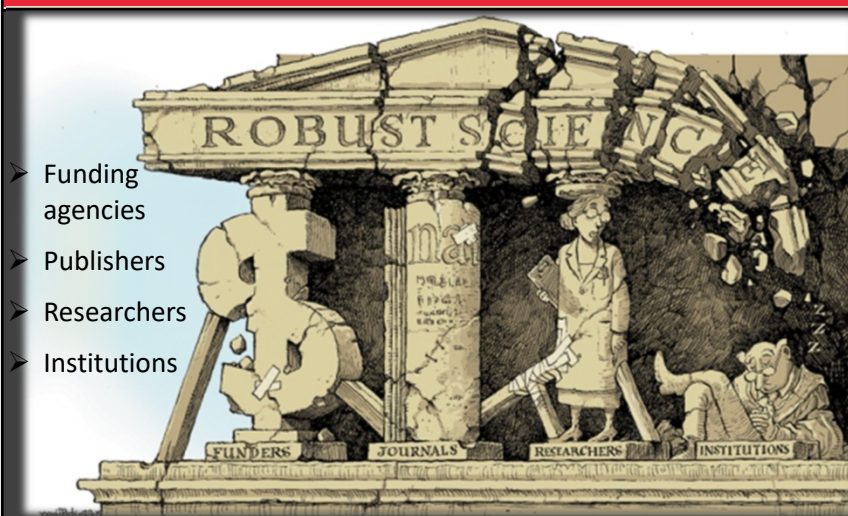
2

Product Development Pathway



3

Stakeholders of Robust Science



- Funding agencies
- Publishers
- Researchers
- Institutions

<https://www.nature.com/news/robust-research-institutions-must-do-their-part-for-reproducibility-1.18259>

4

NIH Public Workshop (2014)

- Funding agencies
- Publishers
- Researchers
- Institutions

- **Sponsors:** NIH + Nature Publishing Group + Science
- **Issue:** Reproducibility, Rigor of research findings
- **Attendees:** Journal editors (>30 basic/preclinical science journals where NIH-funded investigators publish)
- **Goals:** Identify common opportunities in the scientific publishing arena to *enhance rigor and further support research that is reproducible, robust, and transparent*
- **Outcome:** set of principles to facilitate these goals, which a considerable number of journals have agreed to endorse

<https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>

5

NIH Principles and Guidelines

- Funding agencies
- Publishers
- Researchers
- Institutions

Principles and Guidelines for Reporting Preclinical Research:

- Rigorous statistical analysis
- Transparency in reporting
- **Data and material sharing**
- Consider establishing best practice guidelines for:
 - Images
 - Biological materials (antibodies, cell lines, etc.)
 - Animals
- Endorsements (journals, associations, societies)
- Adapted Guidelines

<https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>

6

Data and Material Sharing

- Funding agencies
- Publishers
- Researchers
- Institutions

- Require datasets be made available (where ethically appropriate) upon request
 - during manuscript review
 - upon publication
- Recommend datasets in public repositories, where available
- Encourage presentation of all other data values in machine readable format in the paper (or supplementary information)
- Require materials sharing after publication
- Encourage sharing of software
- Require a statement in the manuscript describing if software is available and how it can be obtained

<https://www.nih.gov/research-training/rigor-reproducibility/principles-guidelines-reporting-preclinical-research>

7

Data and Material Sharing

- Funding agencies
- Publishers
- Researchers
- Institutions

NOT-OD-21-013 Final NIH Policy for Data Management and Sharing (DMS)

Release Date: 29 October 2020

Effective Date: 25 January 2023

Purpose: *Data sharing enables researchers to rigorously test the validity of research findings, strengthen analysis through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery....*

<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-22-213.html>

8

Data and Material Sharing

- Funding agencies
- Publishers
- Researchers
- Institutions

NOT-OD-21-013 Final NIH Policy for Data Management and Sharing (DMS)

Release Date: 29 October 2020

Effective Date: 25 January 2023

Purpose: *...In addition, NIH emphasizes the importance of good data management practices, which provide the foundation for effective data sharing and improve the reproducibility and reliability of research findings...*

9

Data and Material Sharing

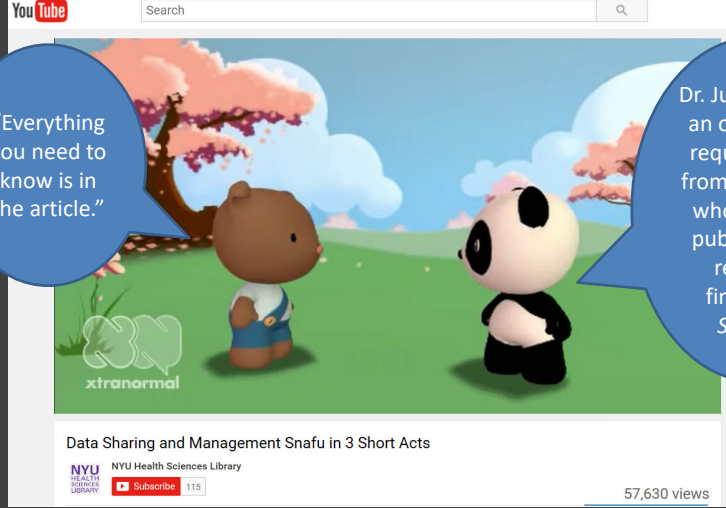
- Funding agencies
- Publishers
- Researchers
- Institutions

NIH encourages data management and data sharing practices consistent with the FAIR data principles.

- F** Findable
- A** Accessible
- I** Interoperable
- R** Re-usable

10

Why is Data Management and Resource Sharing Important?



“Everything you need to know is in the article.”

Dr. Judy Benign, an oncologist, requests data from a scientist who recently published his research findings in *Science*.

Data Sharing and Management Snafu in 3 Short Acts

NYU Health Sciences Library

57,630 views

<https://www.youtube.com/watch?v=N2zK3sAtr-4>

utmb Health

11

Resource Sharing—NIH

- Funding agencies
- Publishers
- Researchers
- Institutions

NIH considers the sharing of unique research resources developed through NIH-sponsored research an important means to enhance the value and further the advancement of research.

When resources have been developed with NIH funds and the associated research findings published or provided to NIH, it is important that the results be made readily available for research purposes to qualified individuals within the scientific community.



https://grants.nih.gov/grants/peer/guidelines_general/Resource_sharing_plans.pdf

utmb Health

12

12

Resource Sharing—NIH

- Funding agencies
- Publishers
- Researchers
- Institutions

- Samples
- Reagents
- Model organism (e.g., transgenic mouse strain)
- Scientific Data

13

Scientific Data – Policy Definition

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 (NIH policy definition) *do not* include laboratory notebooks, preliminary analysis, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communication with colleagues, or physical objects, such as laboratory specimens.

14

Metadata – Policy Definition

Data that provide additional information intended to make scientific data interpretable and reusable (e.g., date, independent sample and variable construction and description, methodology, data provenance, data transformations, any intermediate or descriptive observational variables).

15

Policy Definitions

Data Management = The process of validating, organizing, protecting, maintaining, and processing scientific data to ensure the accessibility, reliability, and quality of the scientific data for its users.

Data Sharing = The act of making scientific data available for use by others (e.g., the larger research community, institutions, the broader public), for example via an established repository

Data Management and Sharing Plan (Plan) = A plan describing the data management, preservation, and sharing of scientific data and accompanying metadata.

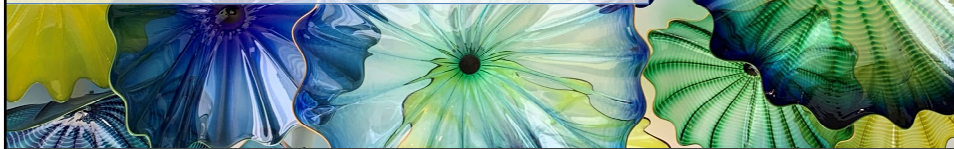
16

Where Do We Begin?



17

Data Management and Resource Sharing

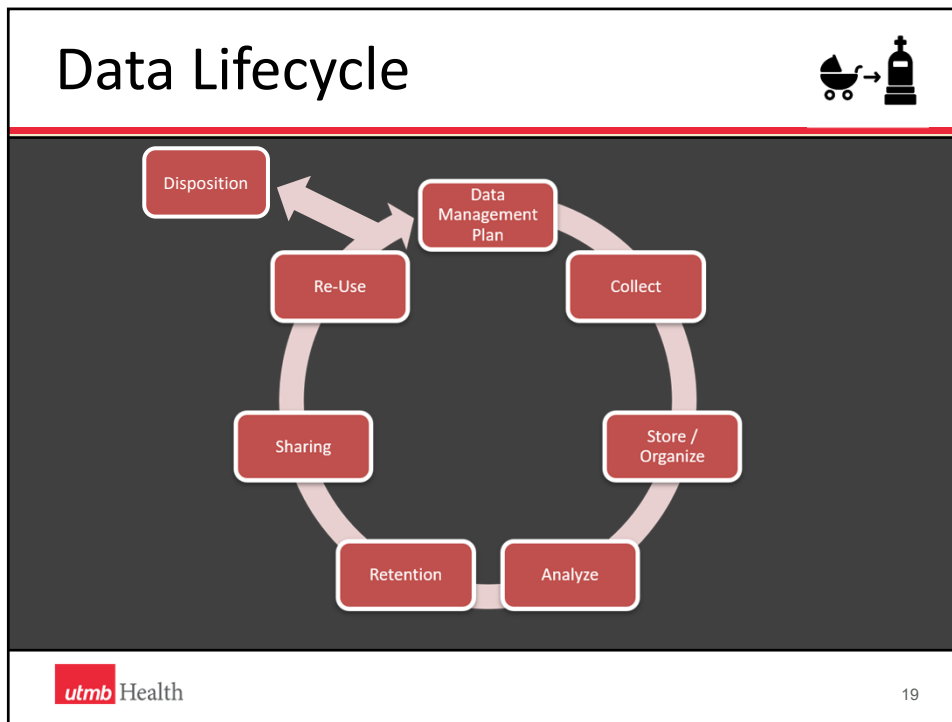


Topics

- Principles, Guidelines, Policies, Definitions
- **Data Lifecycle**
 - Data Quality & Integrity
- Case Study—Break out session



18



19

Data Management

- Data is (are) a scholarly product
- Data are fragile and easily lost
- Growing research data requirements
- Good management helps prevent errors and increases the quality of your analysis
- Well-managed and accessible data allows others to validate and replicate findings
- **Research data management** facilitates sharing of research data and, when shared, data can lead to valuable discoveries by others outside of the original research team

<https://pitt.libguides.com/managedata>

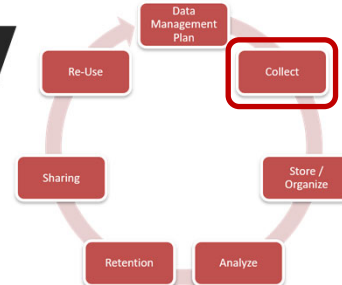
utmb Health 20

20

Data Collection

Perform an inventory...

- Source (Raw) Data
- Types (observational, derived, etc.)
- Format (text, numeric, modeling, images, etc.)
- Quantity
- Standards (e.g., HIPAA)
- Proprietary
- Owner



21

Source Data (Original)

First capture of information



22

Images courtesy of Bing

22

ALCOA Principles

Applies to paper and/or electronic data

Data Quality

- Atributable
- Legible
- Contemporaneous
- Original
- Accurate



Data Integrity

- Complete, Consistent, Enduring, Readily Available

23

Data and Data Integrity

“Data are the foundation on which scientific, engineering, and medical knowledge is built.”

~Ensuring the Integrity, Accessibility, and Stewardship of Research Data in the Digital Age, National Academy of Science, National Academy of Engineering, and Institute of Medicine; Preface, 2009

“Data integrity is the degree to which data are complete, consistent, accurate, trustworthy and reliable and these characteristics of the data are maintained throughout the data life cycle.”

~OECD Draft Advisory Document on GLP Data Integrity; 7 August 2020

24

Data Risk - Non-enduring



Scenario 1 (Worst case): →

Scenario 2: Transcribe →

Scenario 3 (Lower Risk): Affix

Scenario 4: Eliminate bad practice

25

Data Risk - Transcription Errors



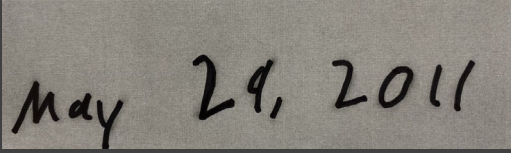
Hemoglobin Value

Animal	HGB
12938	12.2
14039	8.9
14293	3.6
14980	13.8
15209	12.5
15490	9.5
15560	14.0



SYSMEX XE Subject ID 14293			
WBC	8.62	[10 ³ /uL]	
HGB	13.6	[g/dL]	
HCT	40.6	[%]	
MCV	80.3	[fL]	
MCH	32.0	[pg]	
MCHC	33.6	[g/dL]	
PLT	140	[10 ³ /uL]	
RDW-SD	45.2	[fL]	
RDW-CV	13.0	[%]	
MPV	10.8	[fL]	
NEUT	5.19	[10 ³ /uL]	80.3 [%]
LYMPH	2.88	[10 ³ /uL]	31.1 [%]
MONO	0.52	[10 ³ /uL]	6.0 [%]
EO	0.20	[10 ³ /uL]	2.3 [%]
BASO	0.03	[10 ³ /uL]	0.3 [%]
NRBC	0.00	[10 ³ /uL]	0.0 [%]
RET	1.49	[%]	0.0633 [10 ⁶ /uL]
IRF	18.2	[%]	

26

Data Risk - Illegible Data Entries



5/3/2022



27

27

Data Quality/Reproducibility Exercise



28

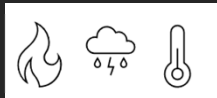
28

Organization and Storage / Retention

Things to think about prospectively...

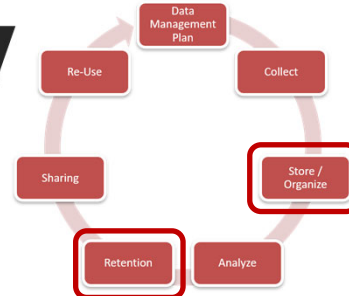
Conditions

- Location (physical / electronic)
- Transcription of source data
- Accessibility (limited)
- Security
- Change control
- Protection
- Migration



Contents

- Retain data and *methods* to allow for study reconstruction
- Critical communication?



29

Managing Electronic Data

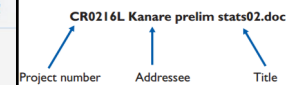
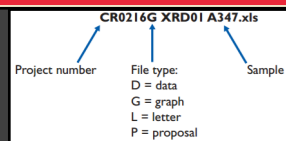
- Audit Trails / Meta Data
- Security / Encryption
- Software Compatibility
- Back-up
- Program Updates
 - Automatic
 - Impact to significant digits
- Data Migration
- Windows PC vs. MAC
- Checksums



Documents library

Example.Study2018.031.0002


- Name
1. Study Plan
 2. Compliance Approvals
 3. Project Management
 4. Study Form Templates
 5. Communication
 6. Source Data
 7. Data Tables
 8. Statistics
 9. Contributing Reports
 10. Summary Report

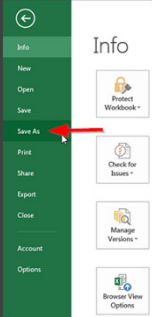


30

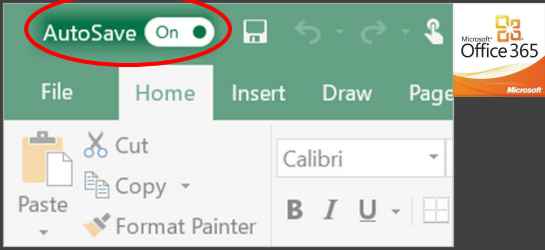
Risks to Electronic Data

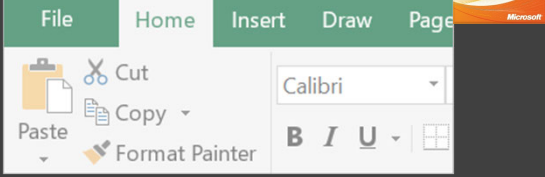
Overwriting of information






- Save
- SaveAs
- AutoSave







31

31

Electronic Laboratory Notebooks


Pros

- Project organization
- Collaboration
- Custom forms/fields to assure all data are captured
- Procedure Checklists
- Time standardization
- Auto reminders
- Searchable
- Audit trail
- Data exportable



Cons

- Cost
- Sustainability (\$)
- System administration
- Compatibility with other systems
- Software updates/data migration verification
- Discontinued (or support discontinued)

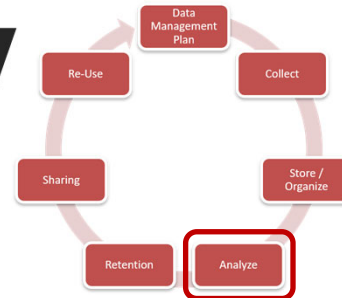

32

32

Data Analysis (Data Manipulation)

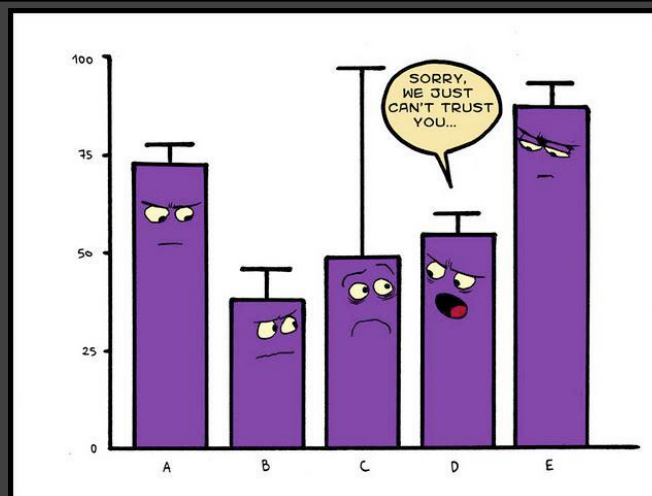
Prospective thinking...

- Methods to reduce transcription errors
- Define inclusion / exclusion criteria
- Develop statistical plan (study plan)
- Retain *methods* to allow for study reconstruction



33

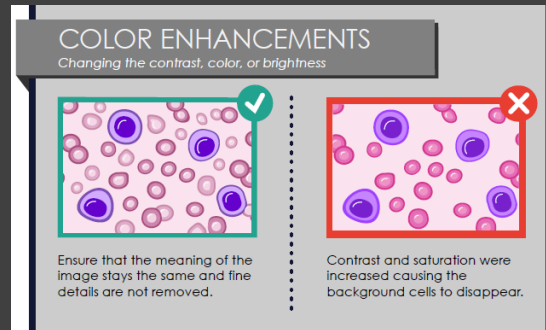
Data Manipulation



34

Image Manipulation

- Document all changes
- Retain unprocessed image
- Follow journal guidelines for permissible processing



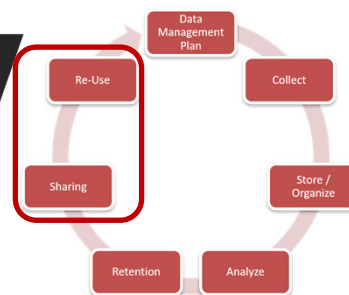
https://ori.hhs.gov/sites/default/files/2017-12/6_Image_Manipulation_scalable.pdf

35

Mechanisms / Conditions for Sharing

Define conditions...

- Mechanisms & Format
 - SharePoint
 - Online repositories
 - Coding
 - Mixed media
- Restrictions (e.g., HIPAA), Conditions / Exclusions
- Sharing agreements / plans
- Schedule/timeline
- Acknowledgements of source data
- Instructions



36

DMS Costs

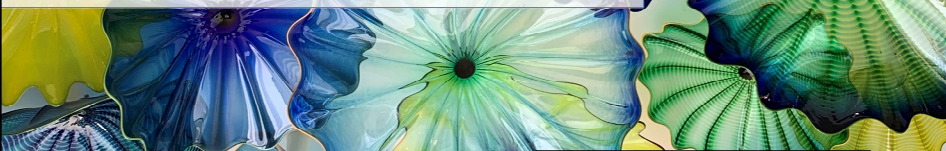
NOT-OD-21-015

- Reasonable, allowable costs may be included in budget
- Allowable costs submitted in budget requests must be incurred during the performance period



37

Data Management and Resource Sharing



Closing Thoughts...

- 53 landmark studies
- 6 confirmed (11%)
 - Controls
 - Reagents
 - Investigator bias
 - **Described complete data set**

COMMENT

Small text: [Investigator bias in preclinical cancer research](#), [How to improve preclinical cancer research](#), [The impact of preclinical cancer research on clinical trials](#), [The role of preclinical cancer research in drug development](#)

Raise standards for preclinical cancer research

C. Glenn Begley and Lee M. Ellis propose how methods, publications and incentives must change if patients are to benefit.

In the past year, there has been a renewed focus on the quality of preclinical cancer research. This is a welcome development, as preclinical research is the foundation of cancer drug development. However, the current state of preclinical cancer research is far from ideal. In this commentary, we discuss the challenges facing preclinical cancer research and propose ways to improve it. We argue that the current system of preclinical cancer research is broken and that a new system is needed. We propose that the standards for preclinical cancer research be raised, that the methods used be made more rigorous, and that the incentives for preclinical cancer research be changed. We believe that these changes are necessary if patients are to benefit from the results of preclinical cancer research.

© 2022 The Author(s). Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction in any medium, provided the original work is properly cited.

38

Data Management and Resource Sharing

- Get organized!
- Data stewardship throughout the data lifecycle
- DMS plan
- Implement the ALCOA principles
- Verify requirements in RFP



utmb Health

39

39

NIH Policies / Announcements

Related Announcements

[NOT-OD-22-214](#) - Supplemental Information to the NIH Policy for Data Management and Sharing: Responsible Management and Sharing of American Indian/Alaska Native Participant Data.

[NOT-OD-21-013](#) - Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data.

[NOT-OD-22-189](#) - Implementation Details for the NIH Data Management and Sharing Policy

[NOT-OD-22-064](#) - Request for Public Comments on DRAFT Supplemental Information to the NIH Policy for Data Management and Sharing: Responsible Management and Sharing of American Indian/Alaska Native Participant Data

[NOT-OD-21-013](#) - Final NIH Policy for Data Management and Sharing

[NOT-OD-21-015](#) - Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing

[NOT-OD-21-016](#) - Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research

[NOT-OD-22-131](#) -Request for Public Comments on DRAFT Supplemental Information to the NIH Policy for Data Management and Sharing: Protecting Privacy When Sharing Human Research Participant Data

[NOT-OD-22-198](#) - Implementation Changes for Genomic Data Sharing Plans Included with Applications Due on or after January 25, 2023

utmb Health

40

40

Data Management and Resource Sharing



Topics

- Principles, Guidelines, Policies, Definitions
- Data Lifecycle
 - Data Quality & Integrity
- **Case Study—Break out session**



References provided on slides

utmb Health 41

41

Case Study—Data Sharing

Identify options (i.e., conditions) for sharing data from a study with 500 human subjects being screened for sexually transmitted diseases.

utmb Health 42

42

Case Study—Data Sharing

The proposed research will include data from approximately 500 subjects being screened for three bacterial sexually transmitted diseases (STDs) at an inner-city STD clinic. The final dataset will include self-reported demographic and behavioral data from interviews with the subjects and laboratory data from urine specimens provided. Because the STDs being studied are reportable diseases, we will be collecting identifying information. Even though the final dataset will be stripped of identifiers prior to release for sharing, we believe that there remains the possibility of deductive disclosure of subjects with unusual characteristics.

Thus, we will make the data and associated documentation available to users only under a *data-sharing agreement* that provides for:

- (1) a commitment to using the data only for research purposes and not to identify any individual participant;
- (2) a commitment to securing the data using appropriate computer technology; and
- (3) a commitment to destroying or returning the data after analyses are completed.

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm#ex

utmb Health

43

43



Thank you!

Melissa Eitzen, MLS(ASCP), MS, RQAP-GLP
 Director, Regulatory Operations
 Institutional Office of Regulated Nonclinical Studies (ORNCS)
 University of Texas Medical Branch at Galveston
 2.810 Rebecca Sealy, Mail Stop 0184
 301 University Boulevard
 Galveston, TX 77555-0184
 Email: mmeitzen@utmb.edu

utmb Health

44

44