Data Management and Resource Sharing

Rigor & Reproducibility Workshop

15 October 2025

Melissa Eitzen, MLS(ASCP), MS, RQAP-GLP
Director, Regulatory Operations
UTMB Office of Research, Regulations, and Compliance
UTMB Institutional Office of Regulated Nonclinical Studies

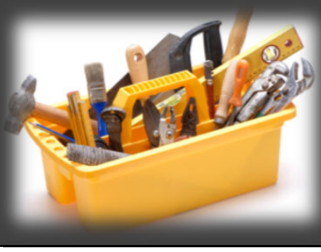
utmb Health

1

Data Management and Resource Sharing

Topics

- History and Policies
- Data Lifecycle (Data Management)
  - Data Quality & Integrity
- Case Study—Break out session



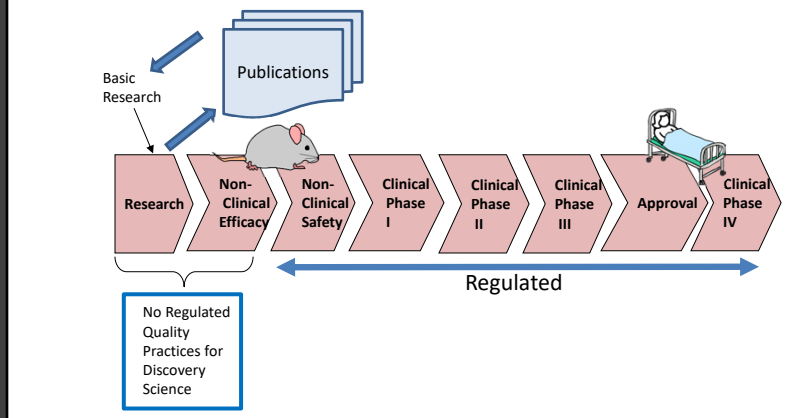
utmb Health

References (links) provided on slides

2

2

Product Development Pathway

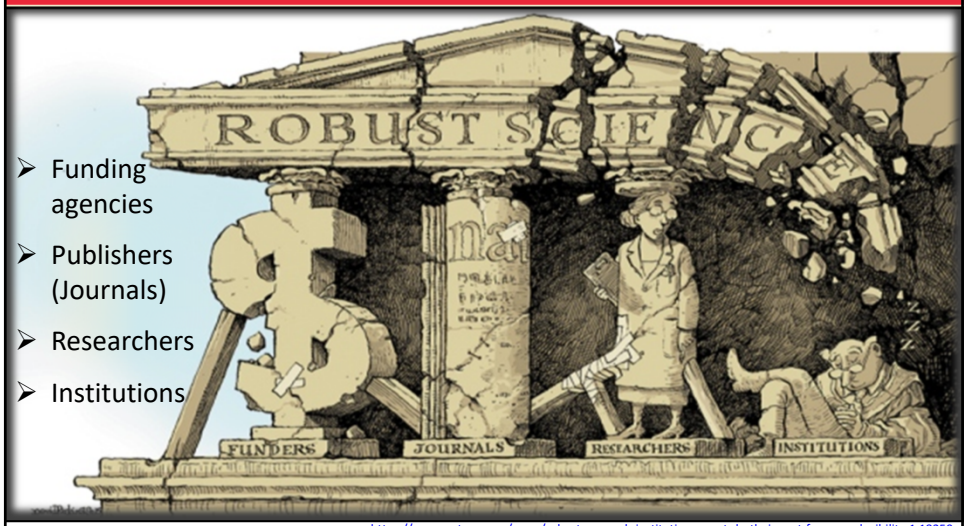


utmb Health 3

3

Stakeholders of Robust Science

- Funding agencies
- Publishers (Journals)
- Researchers
- Institutions



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

utmb Health 4

4

Scientific Advancement

- **Funding agencies**
- Publishers
- Researchers
- Institutions

*"Two of the cornerstones of science advancement are **rigor in designing and performing** scientific research and the ability to **reproduce** biomedical research findings."*

~ NIH Central Resource for Grants and Funding Information

Enhancing Reproducibility through Rigor and Transparency | grants.nih.gov

utmb Health

5

5

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

Marcia McNutt, Journals unite for reproducibility. Science 346, 679679(2014)

utmb Health

6

6

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:
  - Antibodies
  - Cell lines
  - Animals
- Endorsements (journals, associations, societies)
- Adapted Guidelines (to fit unique need)

7

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)
- Encourage sharing of software

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

Section I, Purpose:

"The National Institutes of Health (NIH) Policy for Data Management and Sharing…reinforces NIH's longstanding commitment to making the results and outputs of NIH-funded research available to the public through effective and efficient data management and data sharing practices. ***Data sharing enables researchers to rigorously test the validity of research findings, strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery.*** *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.* ***NIH encourages data management and data sharing practices consistent with the FAIR data principles.****"*

NOT-OD-21-013: Final NIH Policy for Data Management and Sharing

utmb Health

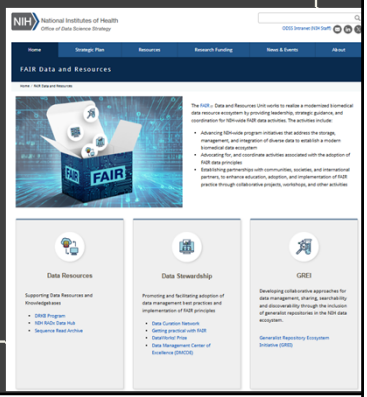
Data and Material Sharing

- **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

Data Engineers!



FAIR Data and Resources | Data Science at NIH

utmb Health

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

Data Engineers!

utmb Health

11

11

Policy Definition—Scientific Data

Scientific Data = 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…

…***does 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.

But wait…

Note! Contracts and/or other applicable regulations may require retention of additional documents!

utmb Health

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html#:~:text=For%20the%20purposes%20of%20the,used%20to%20support%20scholarly%20publications.

12

12

Consortium Written Agreements

2024

*"**For foreign subrecipients, a provision requiring the foreign subrecipient to provide access to copies of all lab notebooks, all data, and all documentation that supports the research outcomes as described in the progress report,** to the primary recipient with a frequency of no less than once per year, in alignment with the timing requirements for Research Performance Progress Report submission. Such access may be entirely electronic."*

Policy: NOT-OD-23-182 *https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-182.html* effective January 1, 2024

Video Resource: https://www.youtube.com/watch?v=mfHIV53-M3A

Webinar On-Demand Video (Broadcast Oct. 17, 2023): https://grants.nih.gov/learning-center/nih-subaward-requirements

13

Policy Definition—Metadata

Metadata = 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).

14

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.

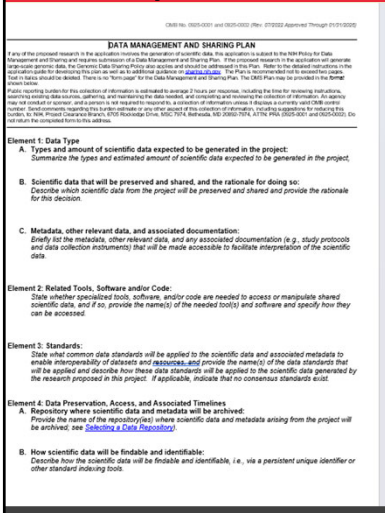
utmb Health

https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html#:~:text=For%20the%20purposes%20of%20the,used%20to%20support%20scholarly%20publications.

15

15

Data Management & Sharing Plan – Template / Examples



DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as for additional guidance on sharing.nih.gov. The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the format below.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

Element 1: Data Type

A. 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.

B. 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.

C. 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.

Element 2: Related Tools, Software and/or Code:
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.

Element 3: Standards:
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.

Element 4: Data Preservation, Access, and Associated Timelines

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).

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.

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.

Element 5: Access, Distribution, or Reuse Considerations

A. Factors affecting subsequent access, distribution, or reuse of scientific data:
NIH expects that in drafting Plans, researchers maximize the appropriate sharing of scientific data. Describe and justify any applicable factors or data use limitations affecting subsequent access, distribution, or reuse of scientific data related to informed consent, privacy and confidentiality protections, and any other considerations that may limit the extent of data sharing. See Frequently Asked Questions for examples of justifiable reasons for limiting sharing of data.

B. Whether access to scientific data will be controlled:
State whether access to the scientific data will be controlled (i.e., made available by a data repository only after approval).)

C. Protections for privacy, rights, and confidentiality of human research participants:
If generating scientific data derived from humans, describe how the privacy, rights, and confidentiality of human research participants will be protected (e.g., through de-identification, Certificates of Confidentiality, and other protective measures).

Element 6: Oversight of Data Management and Sharing:
Describe how compliance with this Plan will be monitored and managed, frequency of oversight, and by whom at your institution (e.g., titles, roles).

1. Data Type
2. Related Tools, Software and/or Code
3. Standards
4. Data Preservation, Access, and Associated Timelines
5. Access, Distribution, or Reuse Considerations
6. Oversight of Data Management and Sharing

utmb Health

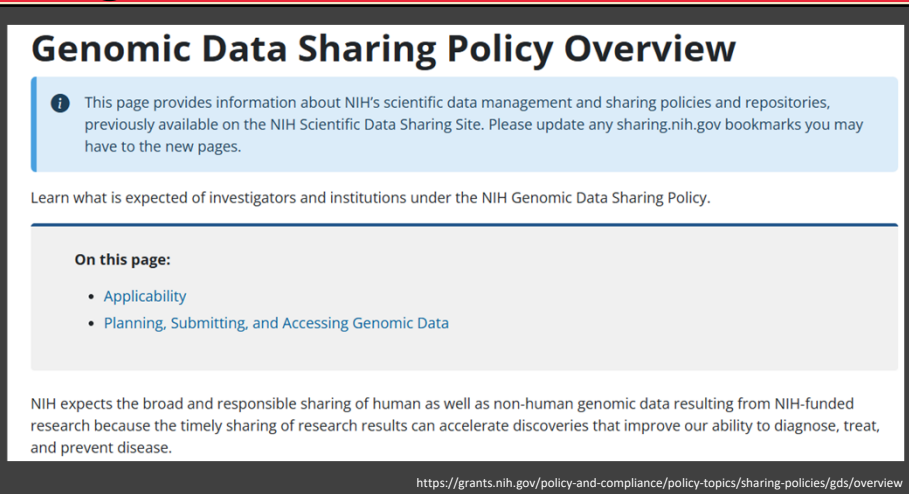
https://grants.nih.gov/grants-process/write-application/forms-directory/data-management-and-sharing-plan-format-page

https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan#sample-plans

16

16

Data Management & Sharing Plan – Large Scale Genomic Data



Genomic Data Sharing Policy Overview

This page provides information about NIH's scientific data management and sharing policies and repositories, previously available on the NIH Scientific Data Sharing Site. Please update any sharing.nih.gov bookmarks you may have to the new pages.

Learn what is expected of investigators and institutions under the NIH Genomic Data Sharing Policy.

On this page:

- Applicability
- Planning, Submitting, and Accessing Genomic Data

NIH expects the broad and responsible sharing of human as well as non-human genomic data resulting from NIH-funded research because the timely sharing of research results can accelerate discoveries that improve our ability to diagnose, treat, and prevent disease.

https://grants.nih.gov/policy-and-compliance/policy-topics/sharing-policies/gds/overview

utmb Health 17

17

Data Management & Sharing Plan – Progress Reports

Reporting Data Management and Sharing (DMS) Plan Activities in the Research Performance Progress Report (RPPR)

Notice Number:
NOT-OD-24-123

Key Dates

Release Date: May 9, 2024

Related Announcements

- **September 23, 2024** - Reminder: Reporting Data Management and Sharing (DMS) Plan Activities in Research Performance Progress Reports (RPPRs) Submitted on or After October 1, 2024. See Notice NOT-OD-24-175.
- **September 23, 2024** - Updated Processes for Requesting Revisions to an Approved Data Management and Sharing (DMS) Plan. See Notice NOT-OD-24-176.
- **October 29, 2020** - Final NIH Policy for Data Management and Sharing. See Notice NOT-OD-21-013.

NOT-OD-24-123: Reporting Data Management and Sharing (DMS) Plan Activities in the Research Performance Progress Report (RPPR)

utmb Health 18

18

Where Do We Begin?

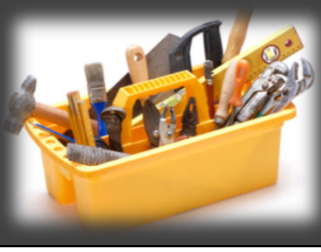
utmb Health

19

Data Management and Resource Sharing

Topics

- History and Policies
- **Data Lifecycle (Data Management)**
  - **Data Quality & Integrity**
- Case Study—Break out session

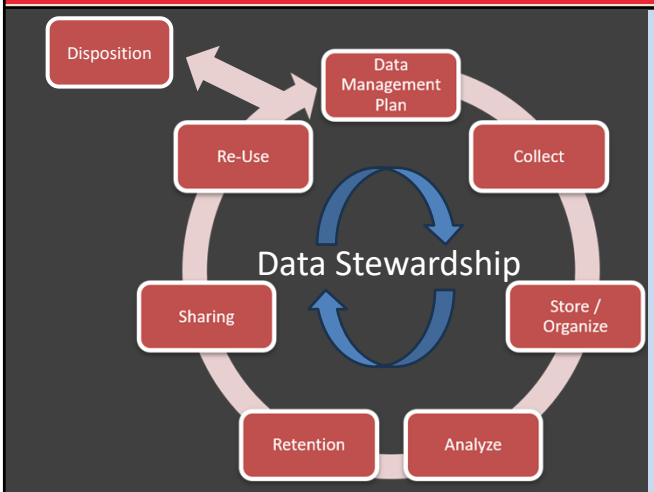


References provided on slides

utmb Health

20

Data Lifecycle



Data Sharing Plan Elements

1. Data Type
2. Related Tools, Software and/or Code
3. Standards
4. Data Preservation, Access, and Associated Timelines
5. Access, Distribution, or Reuse Considerations
6. Oversight of Data Management and Sharing

utmb Health 21

21

Data Management

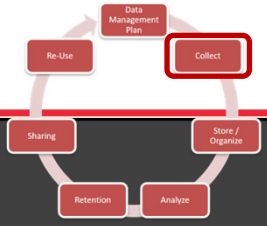
- Data is (are) a scholarly product
- Data are fragile and easily lost
- Growing research data requirements
- Good management helps prevent errors and increases the quality of your analysis
- Well-managed and accessible data allows others to validate and replicate findings
- ***Research data management***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

University of Pittsburgh Library System

utmb Health 22

22

Data Collection



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

23

Source Data (Original)

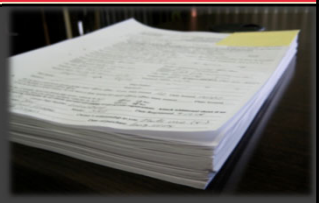
First capture of information

Images courtesy of Bing

24

ALCOA Principles

Applies to paper and/or electronic data



Data Quality

- **A**ttributable
- **L**egible
- **C**ontemporaneous
- **O**riginal
- **A**ccurate

Data Integrity

- Complete, Consistent, Enduring, Readily Available

utmb Health

25

25

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..."

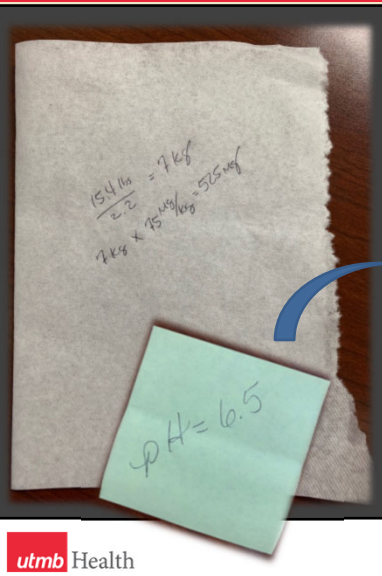
~OECD Advisory Document on GLP Data Integrity; 20 Sept. 2021

utmb Health

26

26

Data Risk - Non-enduring



Scenario 1 (Worst case):

Scenario 2: Transcribe

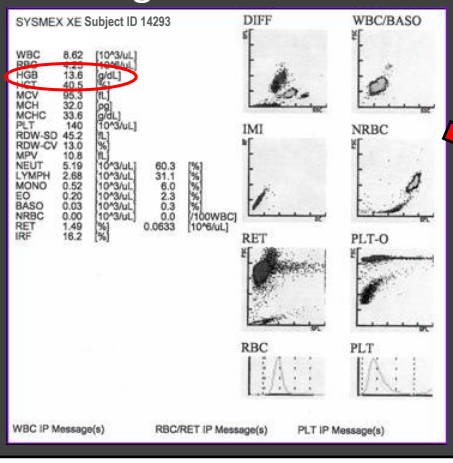
Scenario 3 (Lower Risk): Affix

Scenario 4: Eliminate bad practice

utmb Health

Data Risk - Transcription Errors

Hemoglobin Value

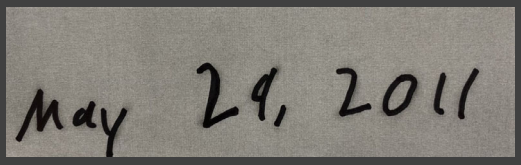


Animal	HGB
12938	12.2
14039	8.9
14293	3.6
14980	13.8
15209	12.5
15490	9.5
15560	14.0

Source: Google Images

utmb Health

Data Risk - Illegible Data Entries



5/3/2024

utmb Health 29

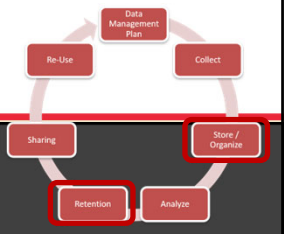
29

Data Quality/Reproducibility Exercise

utmb Health 30

30

Organization, Retention



Conditions

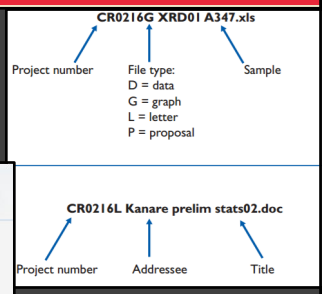
- Location (physical / electronic)
- Accessibility / Security (limited)
- Protection from Loss / Unauthorized Changes
- Change Control
- Move / Migration

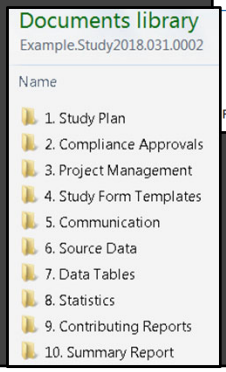
utmb Health 31

31

Managing Electronic Data

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



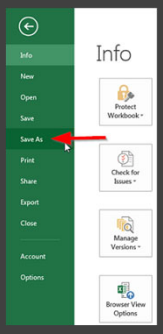


utmb Health 32

32

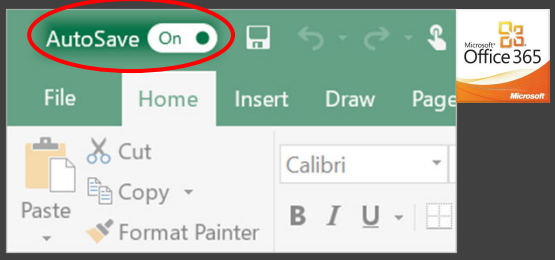
Risks to Electronic Data

Overwriting of information



- Save
- Save As
- AutoSave

- Copy / Paste Errors



33

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)

34

Lab Notebooks

Maintaining a laboratory notebook

Tips for undergraduates, but perhaps useful for anyone.

Reasons to keep a laboratory notebook

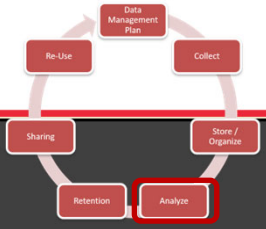
1. To provide yourself with a complete record of why experiments were initiated and how they were performed. You'll forget if you don't. Seriously: even in your youth your brain cells are senescing.
2. To give yourself a central, physical place to record your data, to note statistical outcomes, and to paste graphs that show results. Researchers who keep these items in separate places are unlikely to be productive scientists.
3. To encourage sound thinking. Keeping a notebook gives you a forum to talk to yourself — to ask questions, to record important thoughts about the experimental design, and to speculate on how your results might eventually be interpreted.
4. To provide information to a person who is interested in continuing your research project, even if you deem that possibility hilariously unlikely. And if you're doing important research and die an early, gruesome death, your colleagues might want to pick it up.
5. To get rich. Not everyone sets out with the goal of patenting a process or contraption, but you might stumble onto something actually important, and in such an event you must have a notebook that supports your claims.

https://colinpurrington.com/tips/lab-notebooks/

35

35

Data Analysis



- Implement methods to reduce transcription errors
- Prospectively define inclusion / exclusion criteria
- Develop prospective statistical plan (within the study plan) and analyze data in accordance with the plan
- Retain meta data and methods (protocols) that allow for study reconstruction
- Retain critical communication

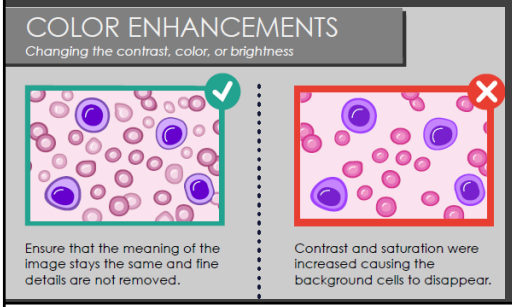
utmb Health

36

36

Image Manipulation

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



- Western Blot Manipulation
- 1997 – 2023
- 132 Research Papers
- 18,000 Citations
- Prasinezumab in Phase I trials

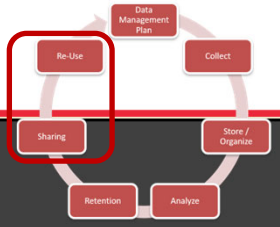
utmb Health

https://ori.hhs.gov/sites/default/files/2017-12/6_Image_Manipulation_scalable.pdf
https://www.science.org/content/article/research-misconduct-finding-neuroscientist-eliezer-masliah-papers-under-suspicion

37

37

Data Sharing



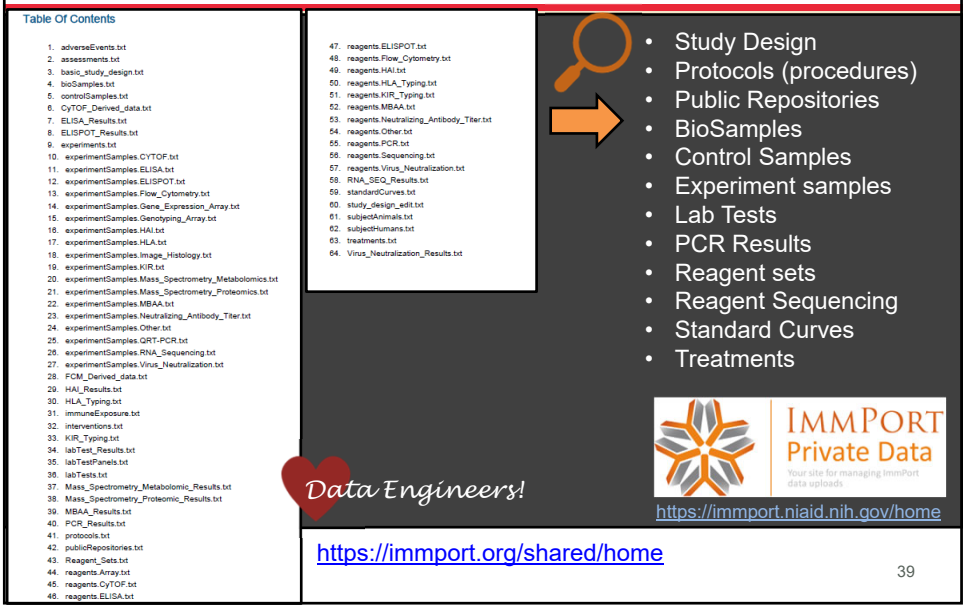
- **Mechanisms & Format**
  - SharePoint / OneDrive
  - Online repositories
  - Coding
  - Mixed media
- **Restrictions (e.g., HIPAA), Conditions / Exclusions**
- **Sharing Plan Timelines**
  "no later than the time of publication or the end of the award or support period, whichever comes first."
- **Acknowledgements of source data**
- **Instructions**

utmb Health

38

38

NIH ImmPort Data Upload Templates



https://immport.org/shared/home

39

39

DMS Costs

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. If the Data Management & Sharing (DMS) plan proposes deposition to multiple repositories, costs associated with each proposed repository may be included.

Budgeting for Data Management & Sharing | Grants & Funding

utmb Health

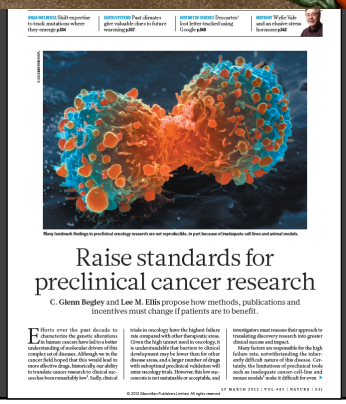
40

40

Data Management and Resource Sharing

Closing Thoughts…

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



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.

utmb Health

41

41

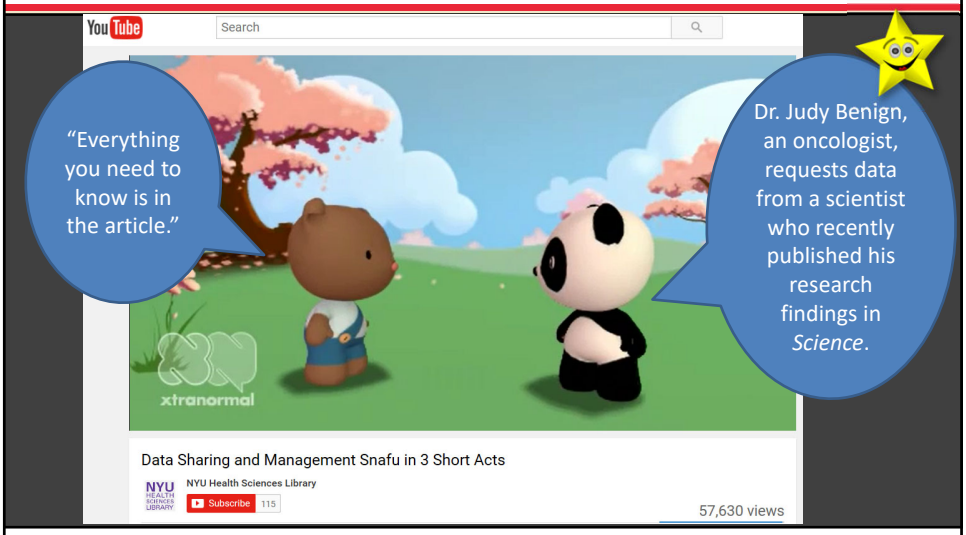
Data Management and Resource Sharing

- Be organized!
- Advocate data stewardship throughout the data lifecycle
- Implement the ALCOA principles
- Verify requirements in RFP / Contract
- Understand that a Data Management and Sharing Plan is a Term and Condition of the Notice of Award (NIH)

42

42

Why is Data Management and Resource Sharing Important?



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

utmb Health

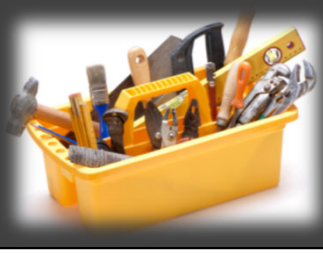
43

43

Data Management and Resource Sharing

Topics

- History and Policies
- Data Lifecycle (Data Management)
  - Data Quality & Integrity
- **Case Study—Break out session**



References provided on slides

utmb Health

44

44

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

45

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.

utmb Health

46

thank you!

Melissa Eitzen, MLS(ASCP), MS, RQAP-GLP
Director, Regulatory Operations
Office of Research, Regulations, and Compliance
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

47

47