Data Management and Resource Sharing

Rigor & Reproducibility Workshop

02 October 2024

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

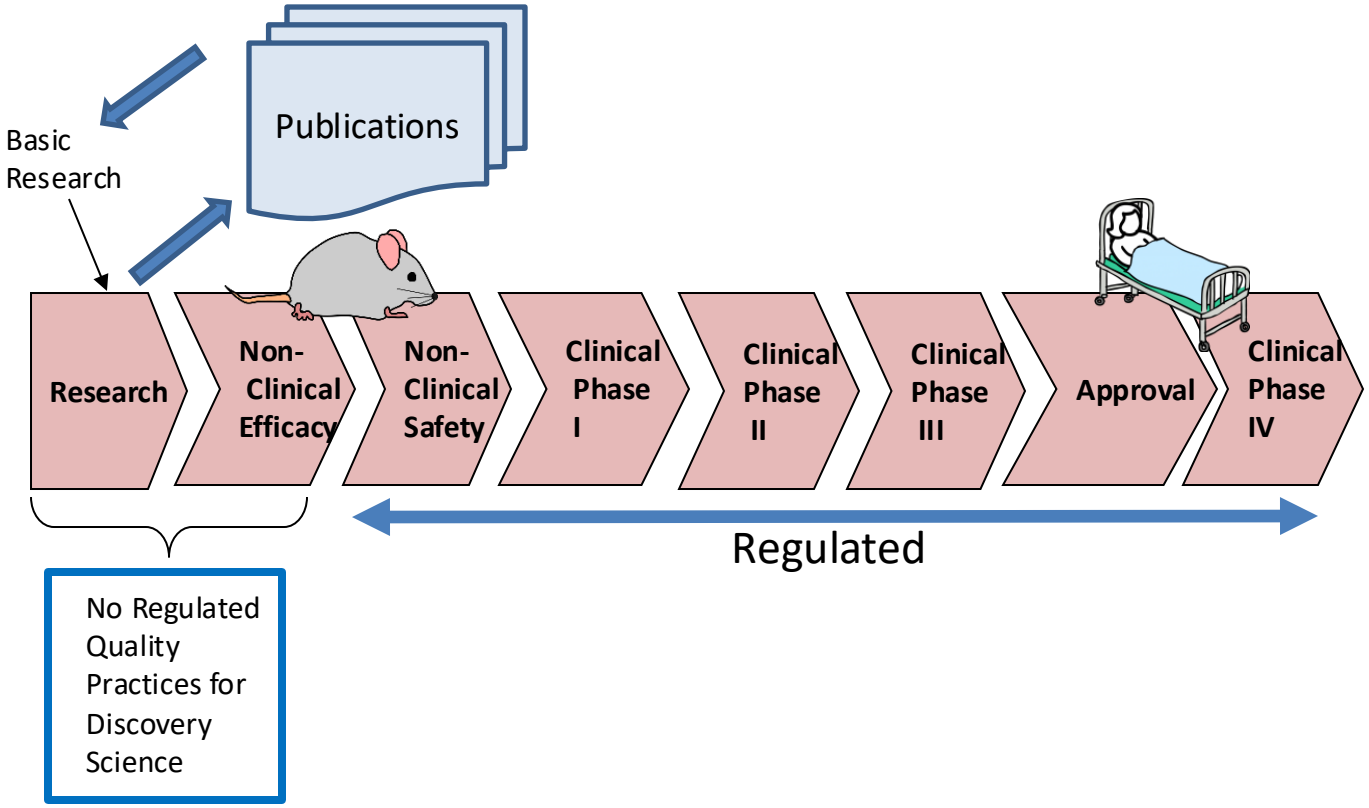
Data Management and Resource Sharing

Topics

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

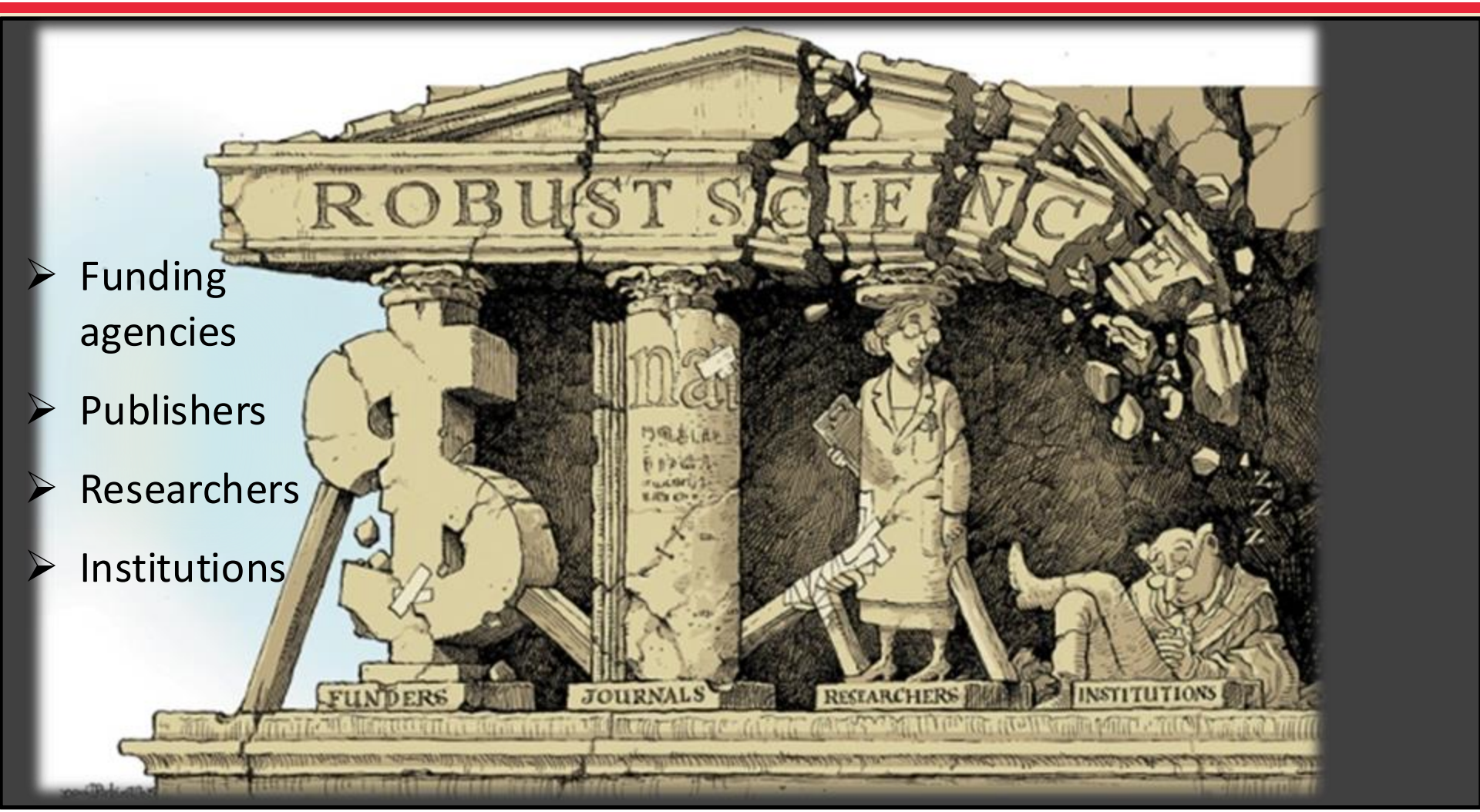
References provided on slides

Product Development Pathway



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

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

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)

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)

Principles and Guidelines for Reporting Preclinical Research | grants.nih.gov

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

Principles and Guidelines for Reporting Preclinical Research | grants.nih.gov

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.****"*

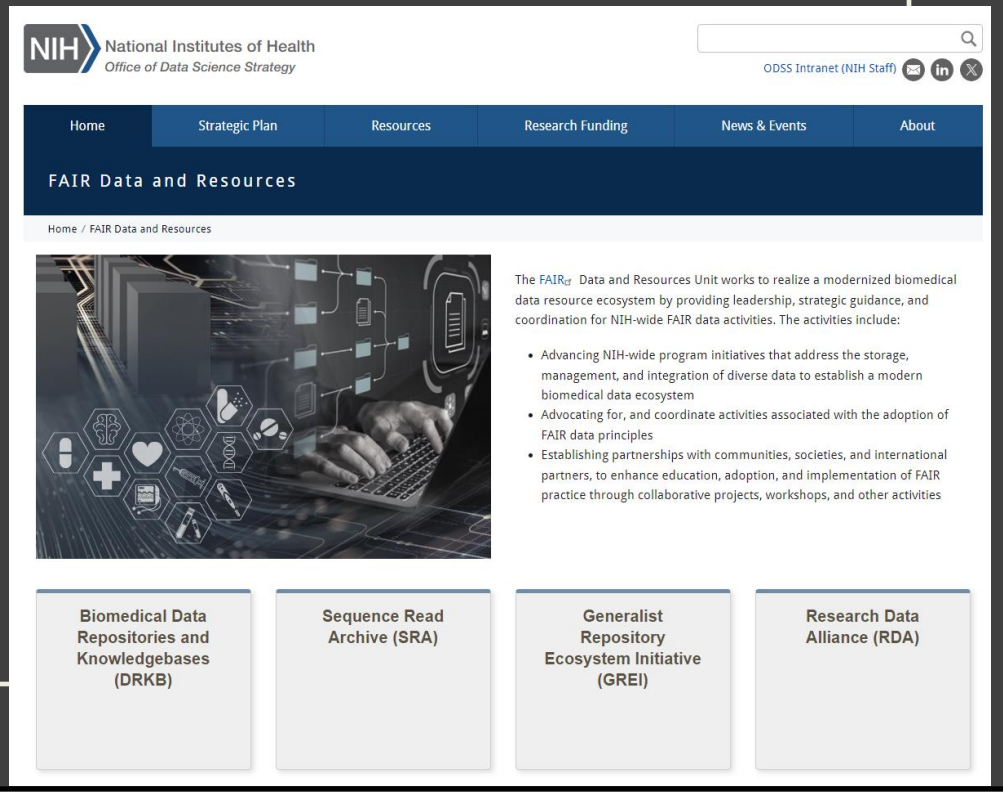
NOT-OD-21-013: Final NIH Policy for Data Management and Sharing

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



FAIR Data and Resources | Data Science at NIH

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

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!

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

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

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).

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

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.

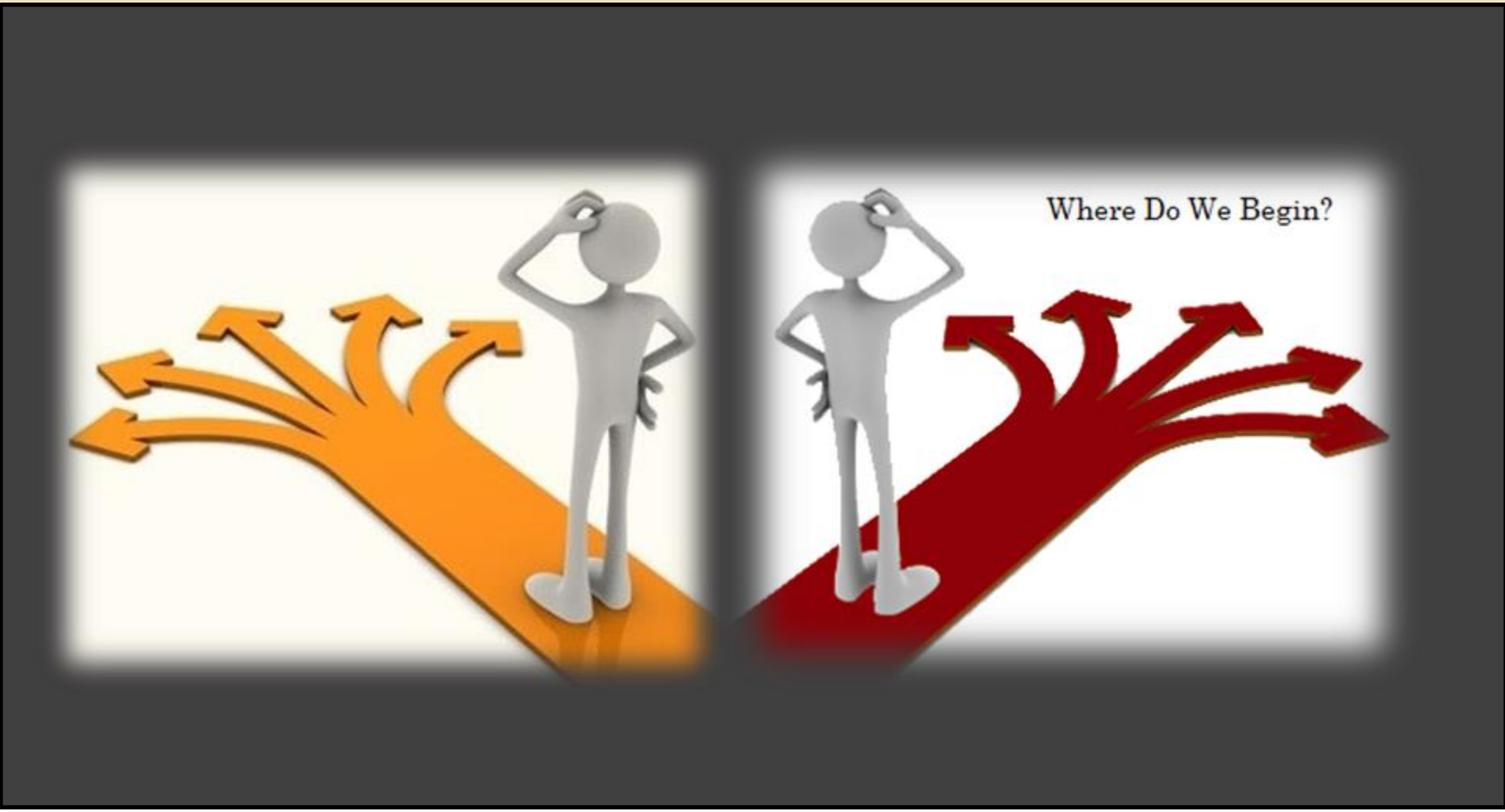
https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html#:~:text=For%20the%20purposes%20of%20the,used%20to%20support%20scholarly%20publications.

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

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

Where Do We Begin?



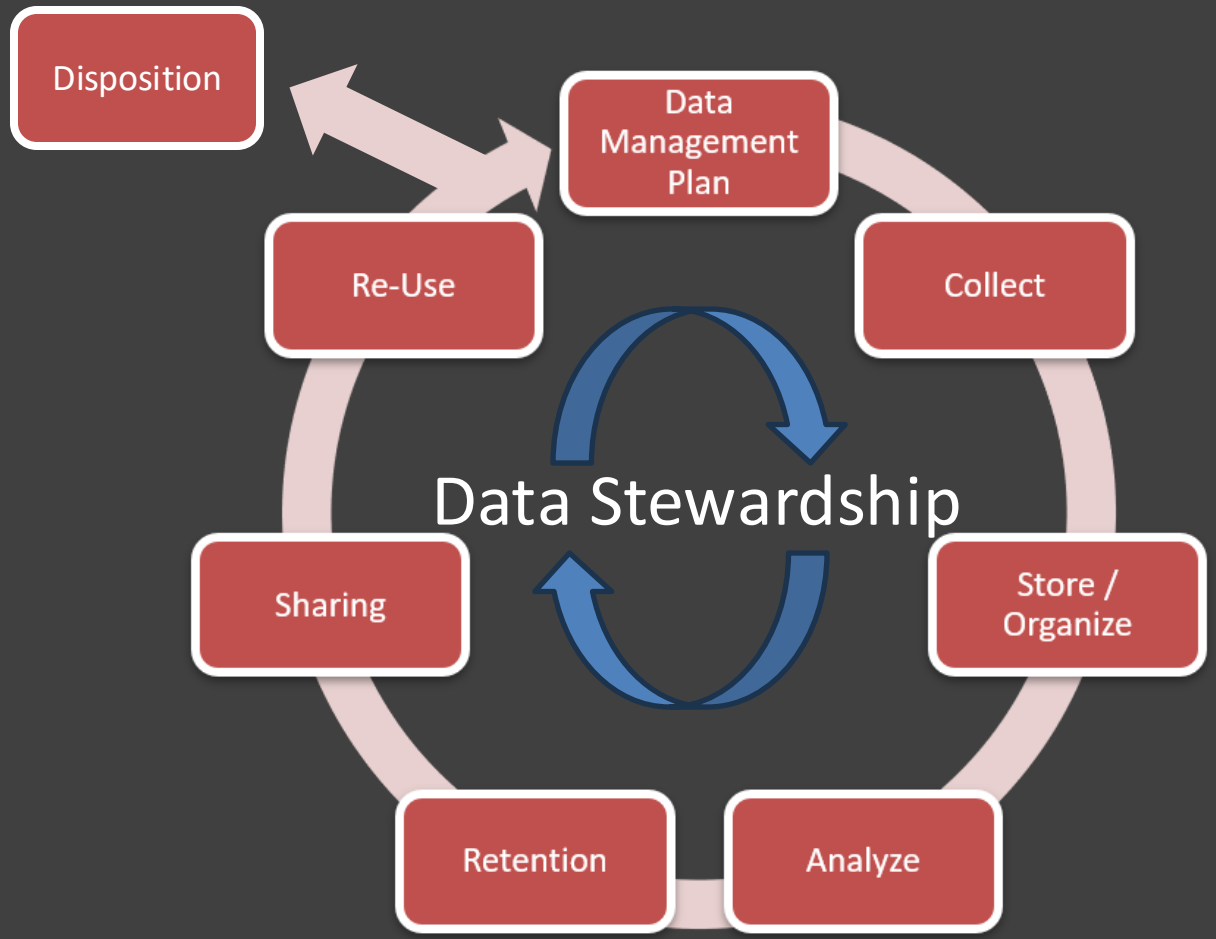
Data Management and Resource Sharing

Topics

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

References provided on slides

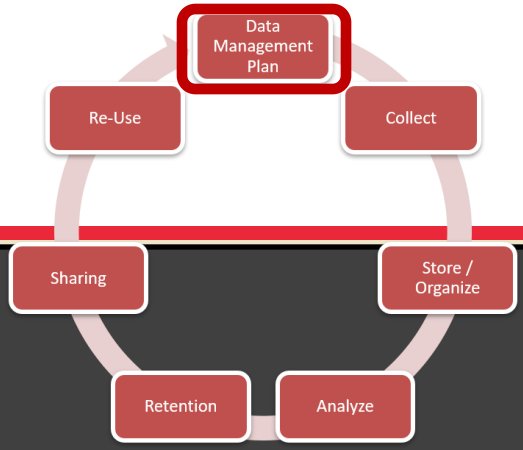
Data Lifecycle



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

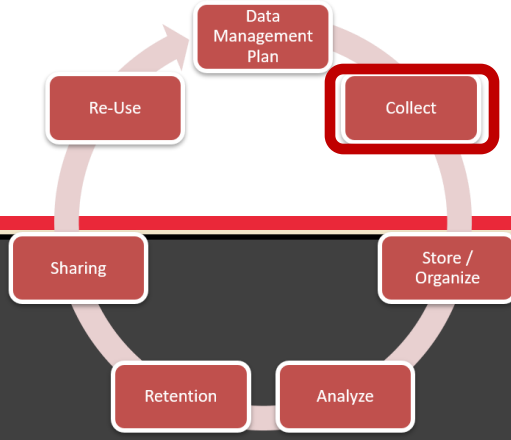
Data Management



- 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

Data Collection



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

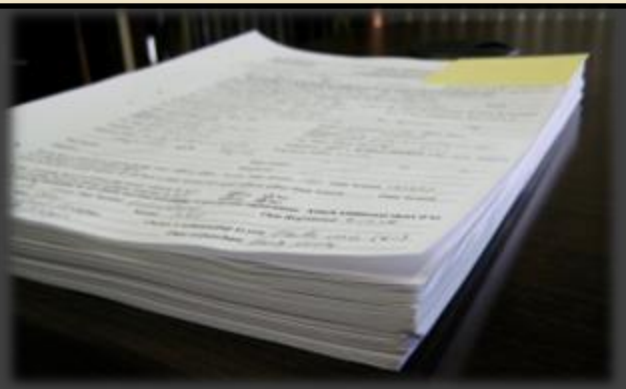
Source Data (Original)

First capture of information

Images courtesy of Bing

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

Data and Data Integrity

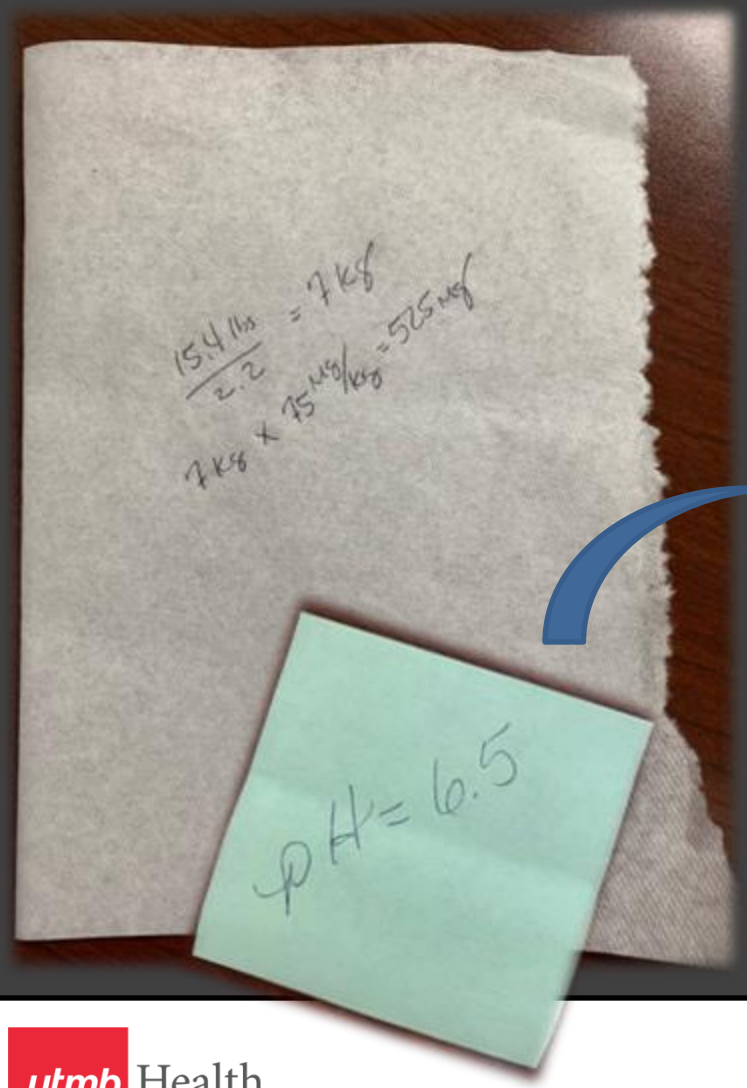
"***Data*** are the foundation on which scientific, engineering, and medical knowledge is built."

~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

Data Risk - Non-enduring



Scenario 1 (Worst case):

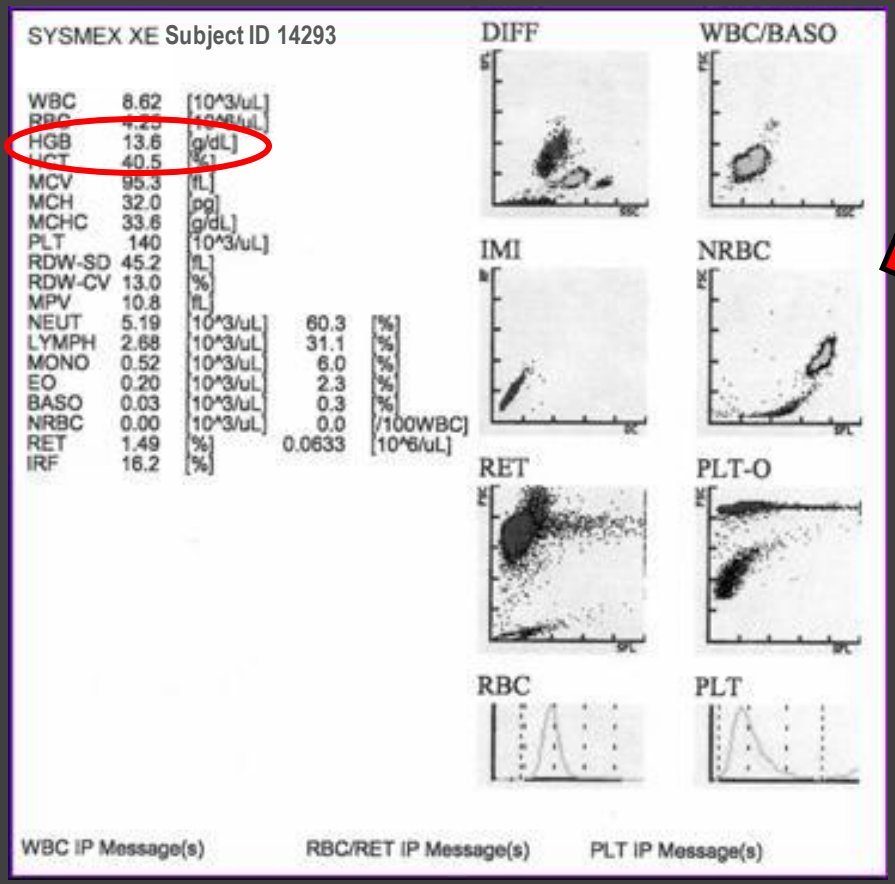
Scenario 2: Transcribe

Scenario 3 (Lower Risk): Affix

Scenario 4: Eliminate bad practice

Data Risk - Transcription Errors

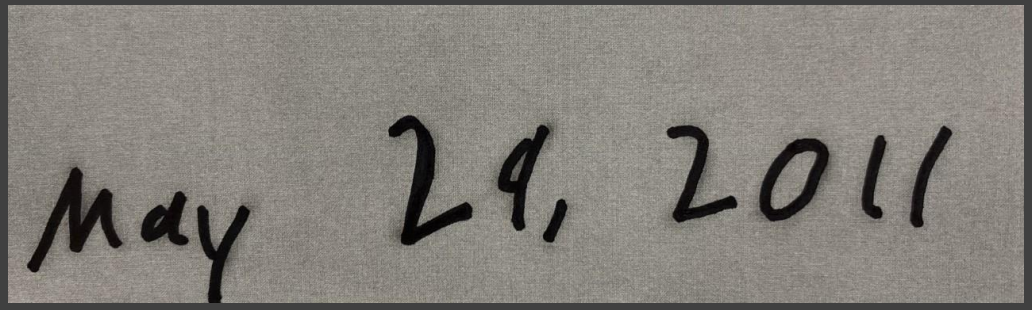
Hemoglobin Value



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

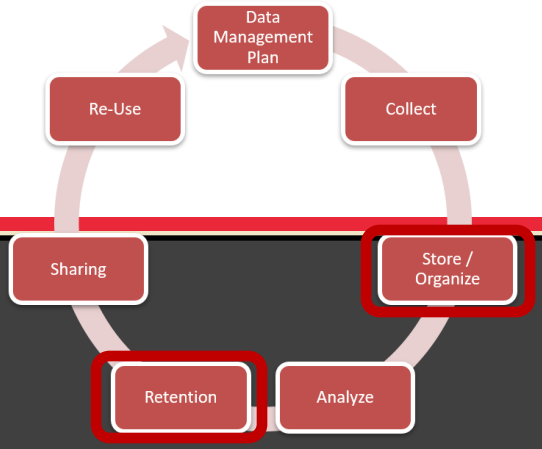
Data Risk - Illegible Data Entries



5/3/2024

Data Quality/Reproducibility Exercise

Organization, Retention

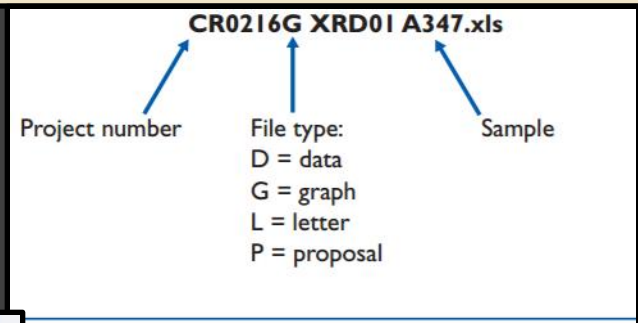


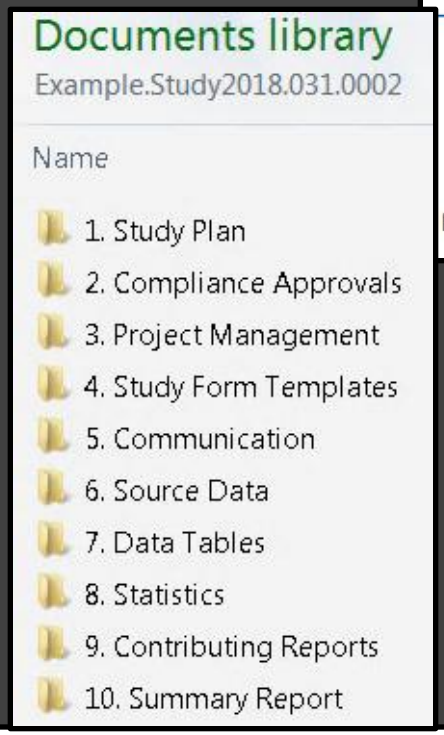
Conditions

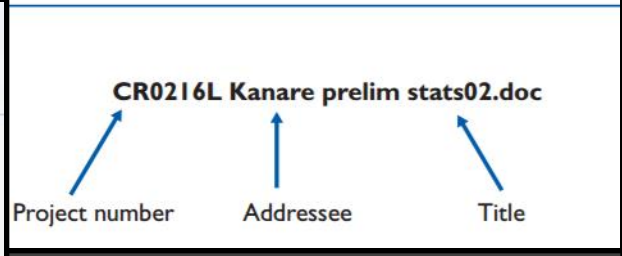
- Location (physical / electronic)
- Transcription of source data
- Accessibility (limited)
- Security
- Change control
- Protection
- Migration

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

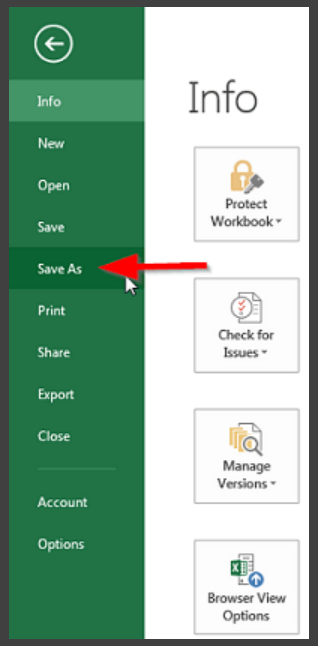




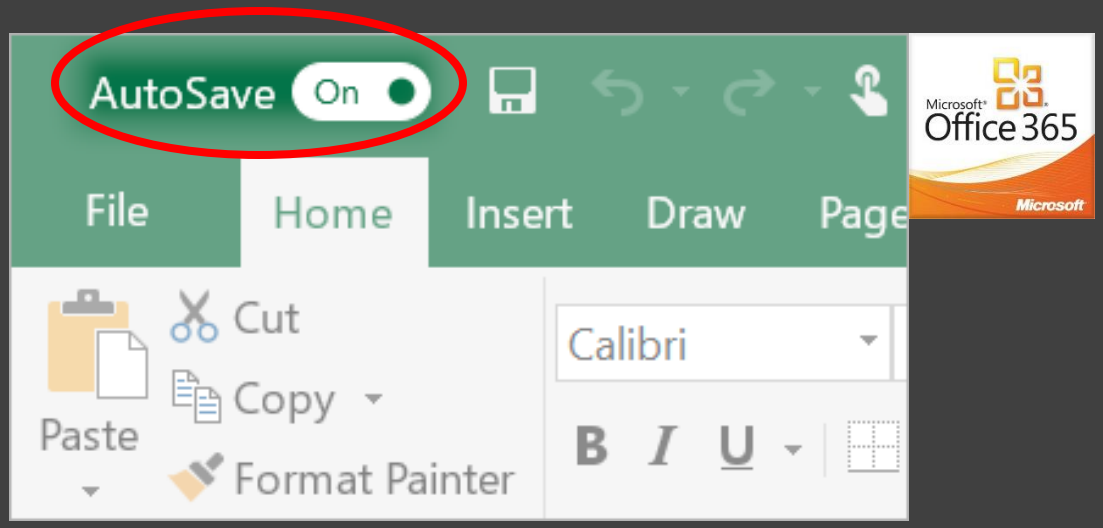


Risks to Electronic Data

Overwriting of information



- Save
- Save As
- AutoSave



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)

utmb Health

Lab Notebooks

Maintaining a laboratory notebook

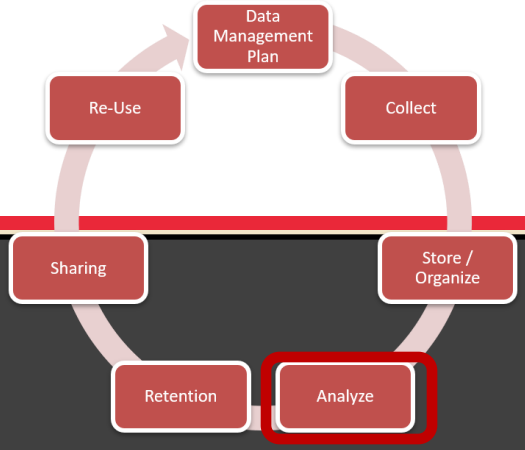
Tips for undergraduates, but perhaps useful for anyone.

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/

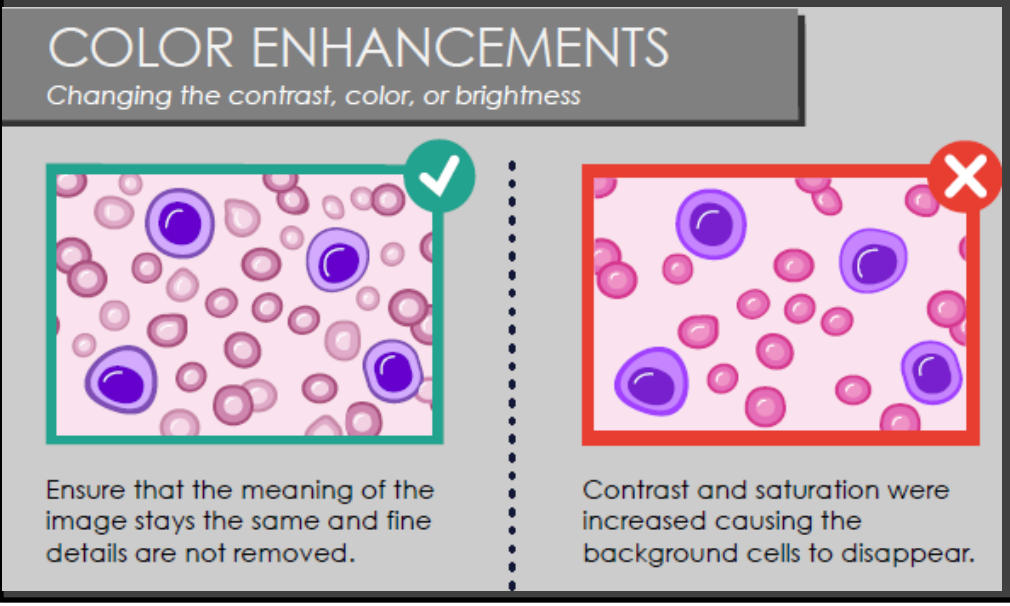
Data Analysis



- Implement methods to reduce transcription errors
- 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

Image Manipulation

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

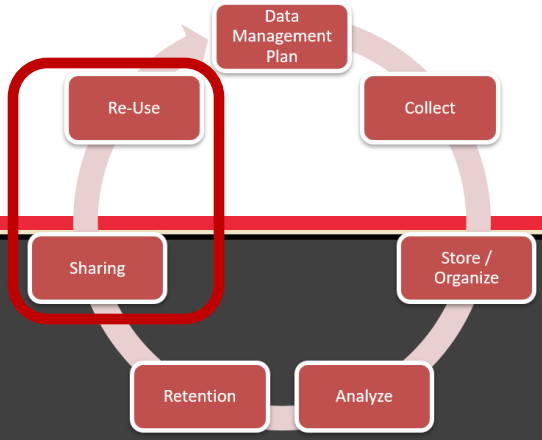


- Western Blot Manipulation
- 1997 – 2023
- 132 Research Papers
- 18,000 Citations
- Prasinezumab in Phase I trials

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

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

NIH ImmPort Data Upload Templates

Table Of Contents

1. adverseEvents.txt
2. assessments.txt
3. basic_study_design.txt
4. bioSamples.txt
5. controlSamples.txt
6. CyTOF_Derived_data.txt
7. ELISA_Results.txt
8. ELISPOT_Results.txt
9. experiments.txt
10. experimentSamples.CYTOF.txt
11. experimentSamples.ELISA.txt
12. experimentSamples.ELISPOT.txt
13. experimentSamples.Flow_Cytometry.txt
14. experimentSamples.Gene_Expression_Array.txt
15. experimentSamples.Genotyping_Array.txt
16. experimentSamples.HAI.txt
17. experimentSamples.HLA.txt
18. experimentSamples.Image_Histology.txt
19. experimentSamples.KIR.txt
20. experimentSamples.Mass_Spectrometry_Metabolomics.txt
21. experimentSamples.Mass_Spectrometry_Proteomics.txt
22. experimentSamples.MBAA.txt
23. experimentSamples.Neutralizing_Antibody_Titer.txt
24. experimentSamples.Other.txt
25. experimentSamples.QRT-PCR.txt
26. experimentSamples.RNA_Sequencing.txt
27. experimentSamples.Virus_Neutralization.txt
28. FCM_Derived_data.txt
29. HAI_Results.txt
30. HLA_Typing.txt
31. immuneExposure.txt
32. interventions.txt
33. KIR_Typing.txt
34. labTest_Results.txt
35. labTestPanels.txt
36. labTests.txt
37. Mass_Spectrometry_Metabolomic_Results.txt
38. Mass_Spectrometry_Proteomic_Results.txt
39. MBAA_Results.txt
40. PCR_Results.txt
41. protocols.txt
42. publicRepositories.txt
43. Reagent_Sets.txt
44. reagents.Array.txt
45. reagents.CyTOF.txt
46. reagents.ELISA.txt
47. reagents.ELISPOT.txt
48. reagents.Flow_Cytometry.txt
49. reagents.HAI.txt
50. reagents.HLA_Typing.txt
51. reagents.KIR_Typing.txt
52. reagents.MBAA.txt
53. reagents.Neutralizing_Antibody_Titer.txt
54. reagents.Other.txt
55. reagents.PCR.txt
56. reagents.Sequencing.txt
57. reagents.Virus_Neutralization.txt
58. RNA_SEQ_Results.txt
59. standardCurves.txt
60. study_design_edit.txt
61. subjectAnimals.txt
62. subjectHumans.txt
63. treatments.txt
64. Virus_Neutralization_Results.txt

- Study Design
- Protocols (procedures)
- Public Repositories
- BioSamples
- Control Samples
- Experiment samples
- Lab Tests
- PCR Results
- Reagent sets
- Reagent Sequencing
- Standard Curves
- Treatments

IMMPORT
Private Data
Your site for managing ImmPort data uploads

https://immport.niaid.nih.gov/home

DMS Costs

Planning & Budgeting for Data Management and Sharing

Prospectively planning for how scientific data will be managed and ultimately shared is a crucial first step in optimizing the reach of data generated from NIH-funded research.

- **Determine** if proposed research is subject to the DMS policy.
- **Identify** appropriate methods/approaches and repositories for managing and sharing scientific data.
- **Develop a Plan** for managing and sharing scientific data and include in application or proposal. If subject to Genomic Data Sharing Policy, submit a single Plan that addresses genomic data considerations.
- **Estimate and request funds** for data management and sharing activities (if not already covered by institution or other sources.)

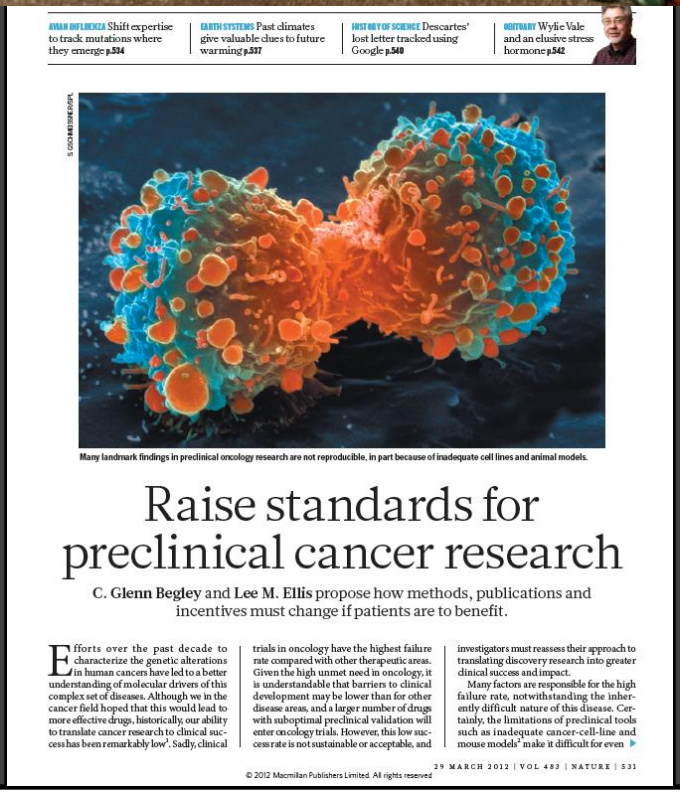
The NIH Data Management & Sharing (DMS) Policy, effective January 25, 2023, applies to all research, funded or conducted in whole or in part by NIH, that results in the generation of **scientific data**

DMS_flyer.pdf (nih.gov)

Data Management and Resource Sharing

Closing Thoughts…

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



AVIAN INFLUENZA Shift expertise to track mutations where they emerge p.534 | EARTH SYSTEMS Past climates give valuable clues to future warming p.537 | HISTORY OF SCIENCE Descartes' lost letter tracked using Google p.540 | OBITUARY Wylie Vale and an elusive stress hormone p.542

S. GSCHMEISSNER/SPL

Many landmark findings in preclinical oncology research are not reproducible, in part because of inadequate cell lines and animal models.

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.

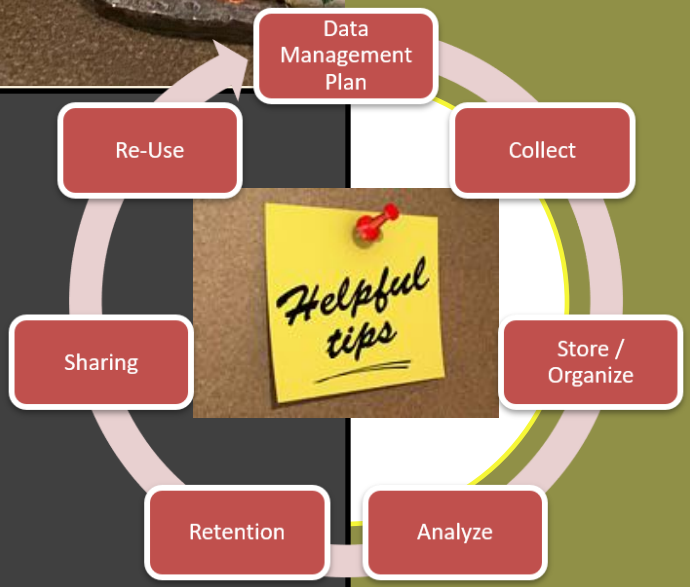
Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low[1]. Sadly, clinical trials in oncology have the highest failure rate compared with other therapeutic areas. Given the high unmet need in oncology, it is understandable that barriers to clinical development may be lower than for other disease areas, and a larger number of drugs with suboptimal preclinical validation will enter oncology trials. However, this low success rate is not sustainable or acceptable, and investigators must reassess their approach to translating discovery research into greater clinical success and impact.

Many factors are responsible for the high failure rate, notwithstanding the inherently difficult nature of this disease. Certainly, the limitations of preclinical tools such as inadequate cancer-cell-line and mouse models[2] make it difficult for even ▶

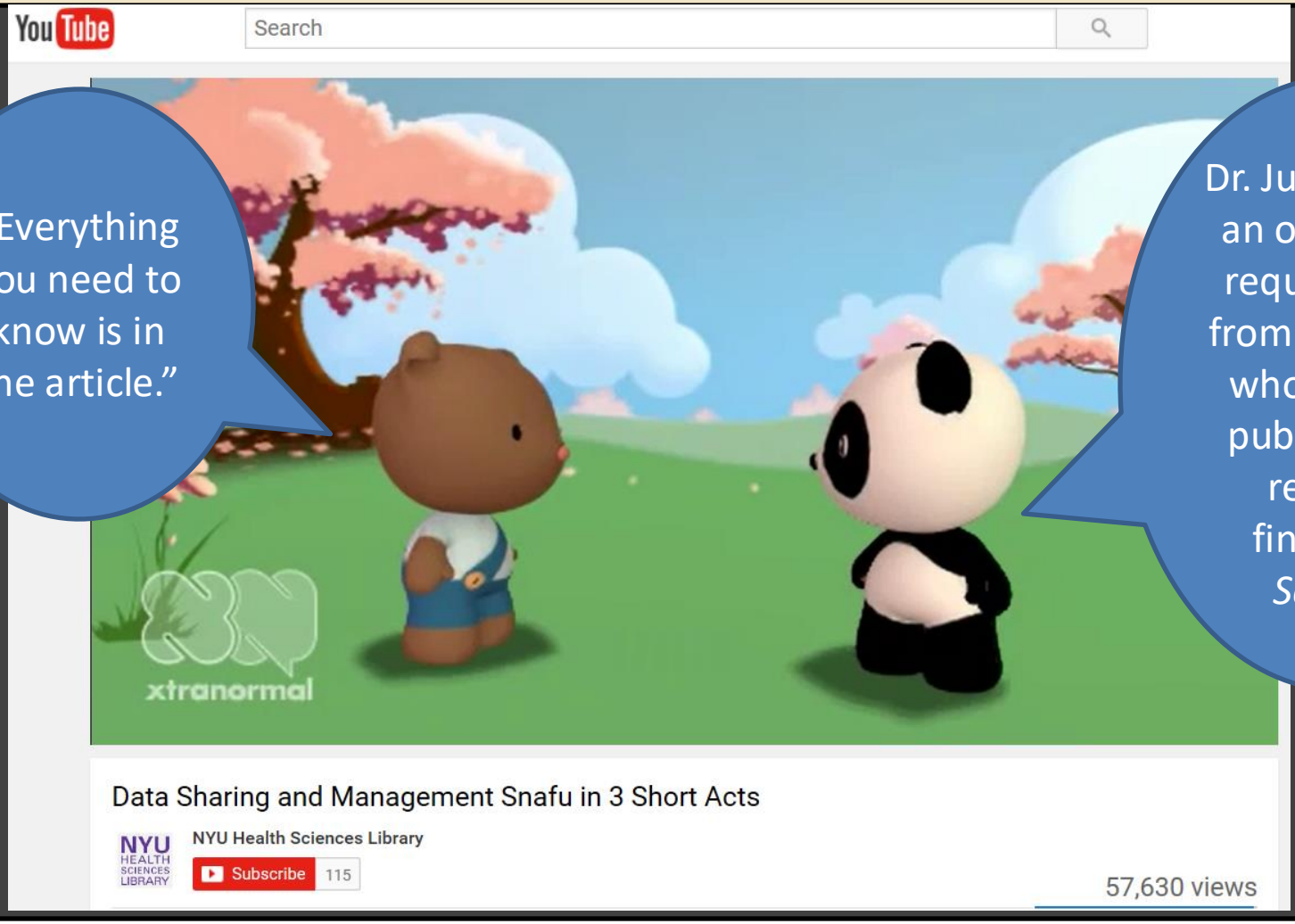
29 MARCH 2012 | VOL 483 | NATURE | 531

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)



Why is Data Management and Resource Sharing Important?



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

Data Management and Resource Sharing

Topics

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

References provided on slides

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.

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.

thank you!

utmb Health

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