



Data Management and Resource Sharing

Rigor & Reproducibility Workshop

17 May 2023

Melissa Eitzen, MLS(ASCP), MS, RQAP-GLP

Director, Regulatory Operations

UTMB Institutional Office of Regulated Nonclinical Studies

mmeitzen@utmb.edu

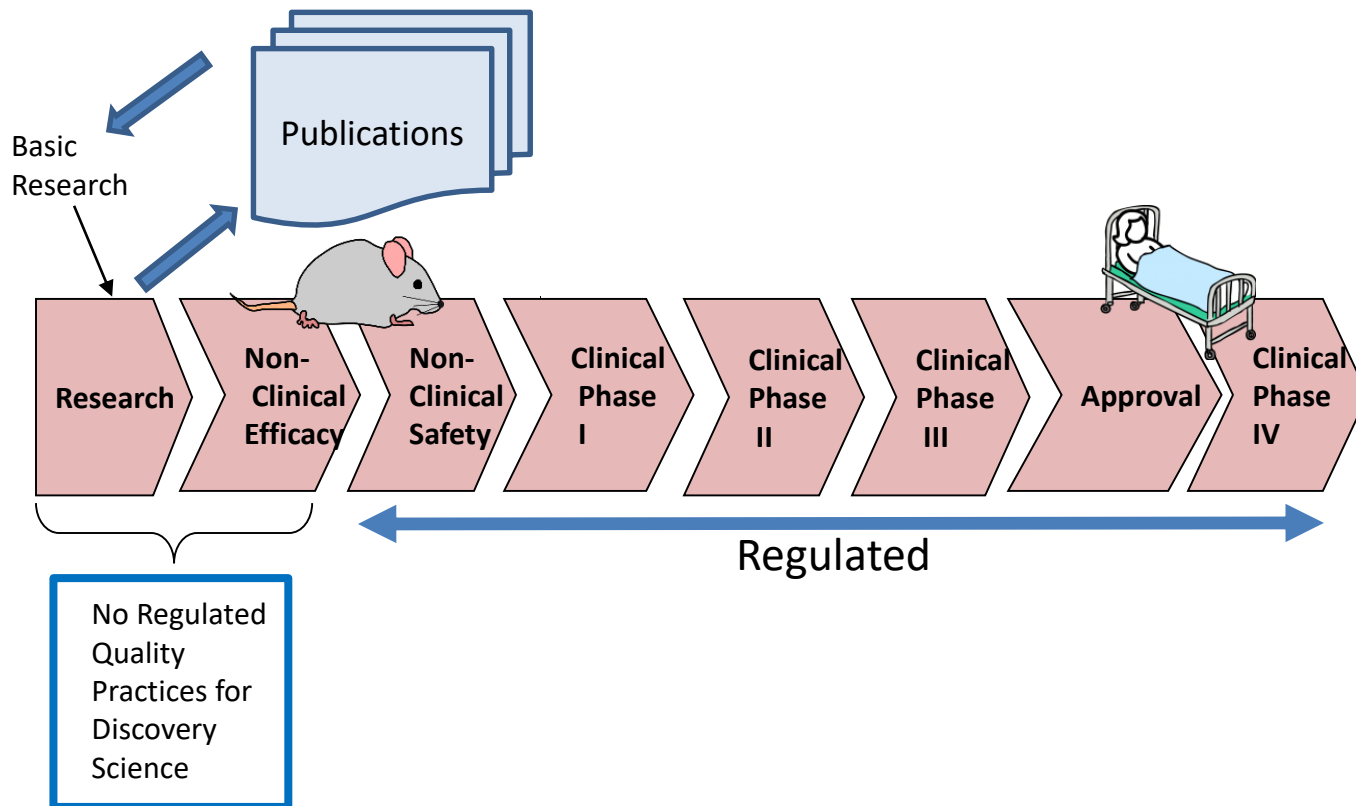
Topics

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



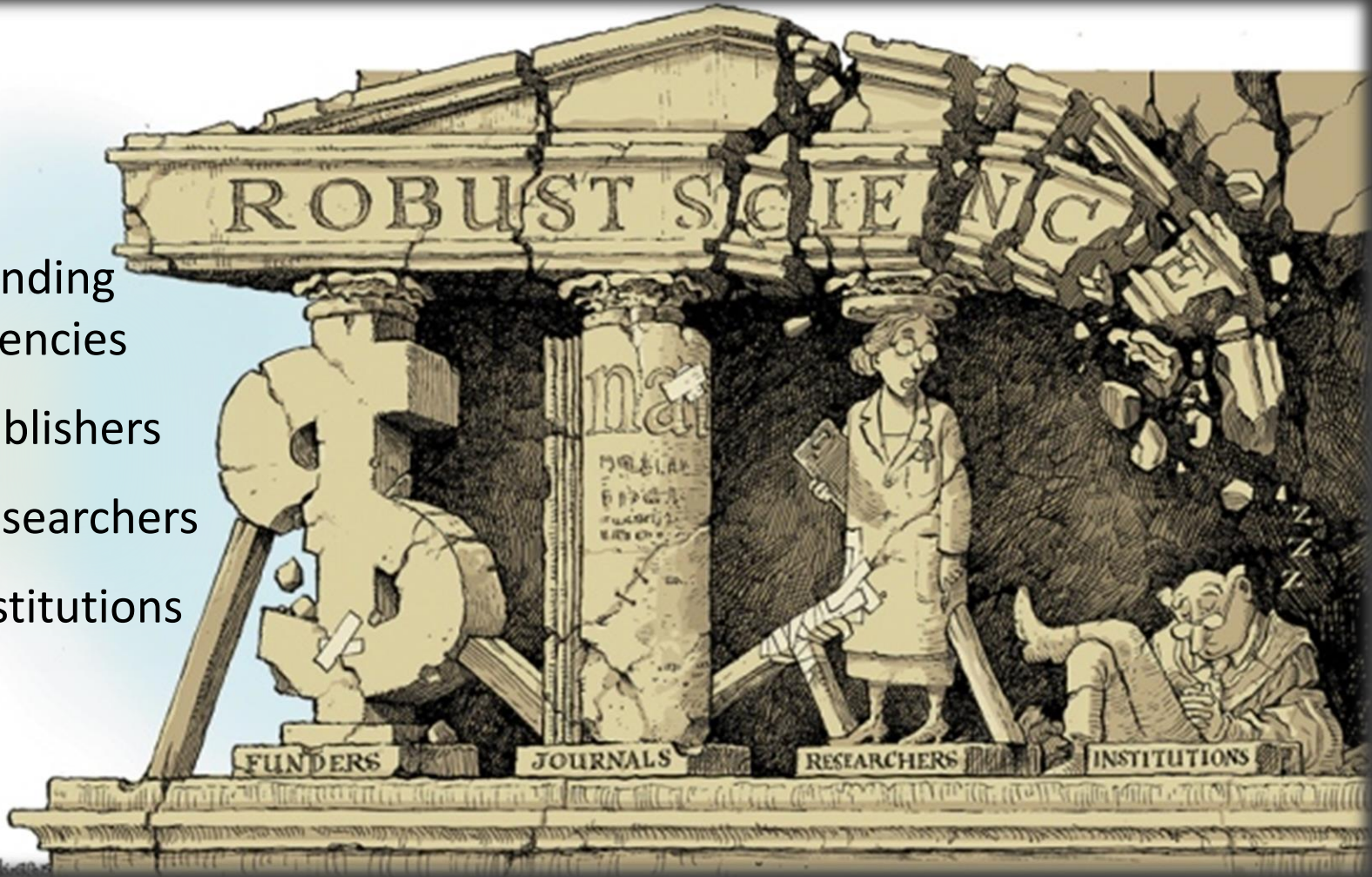
References provided on slides

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>

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>

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>

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>

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>

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...*

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

Data and Material Sharing

- Funding agencies
- Publishers
- Researchers
- Institutions

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

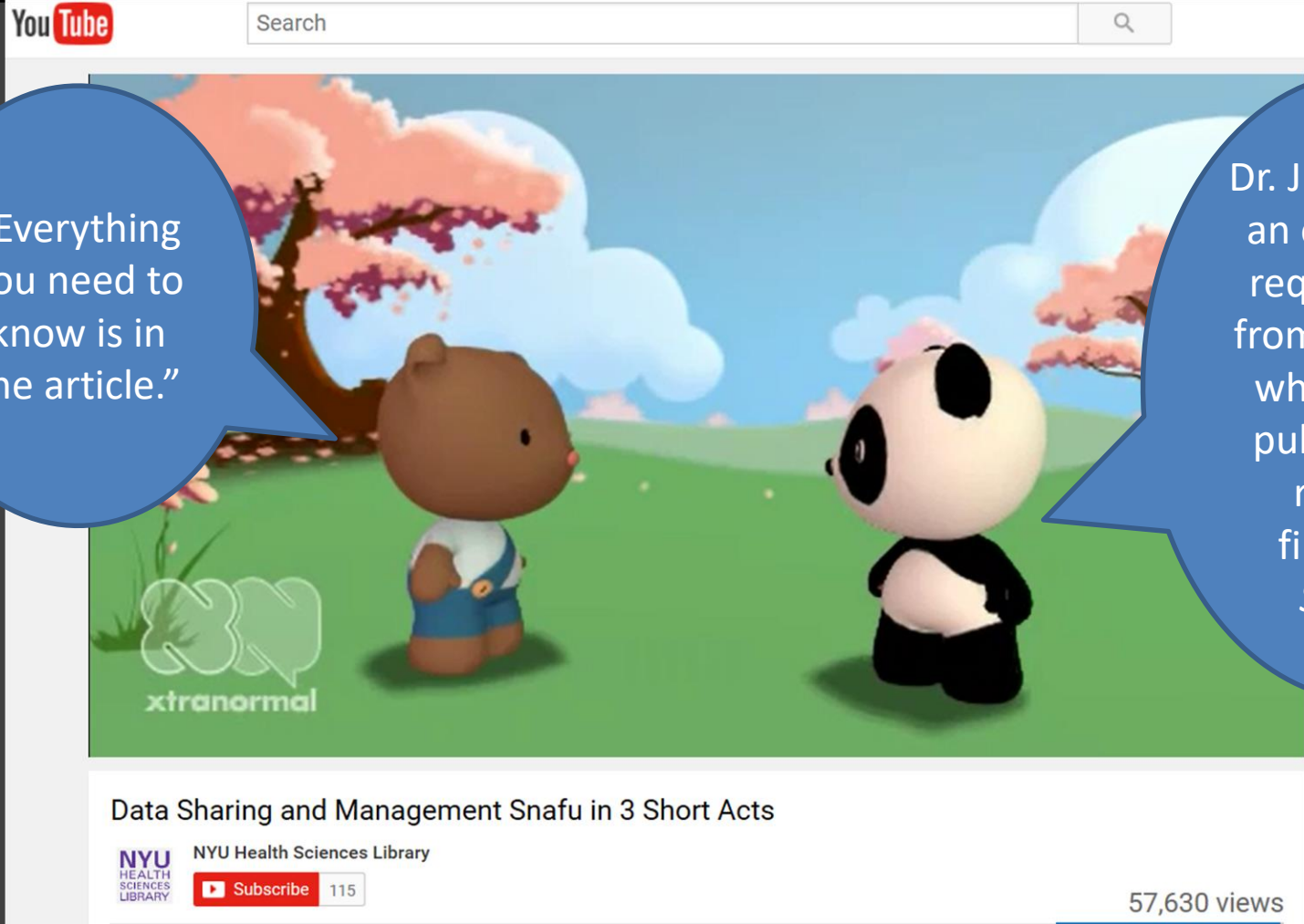
F Findable

A Accessible

I Interoperable

R Re-usable

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*.

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

Resource Sharing—NIH

- Funding agencies
- Publishers
- Researchers
- Institutions

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

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.



Note! Contracts may require retention of additional documents!

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).

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.

Where Do We Begin?



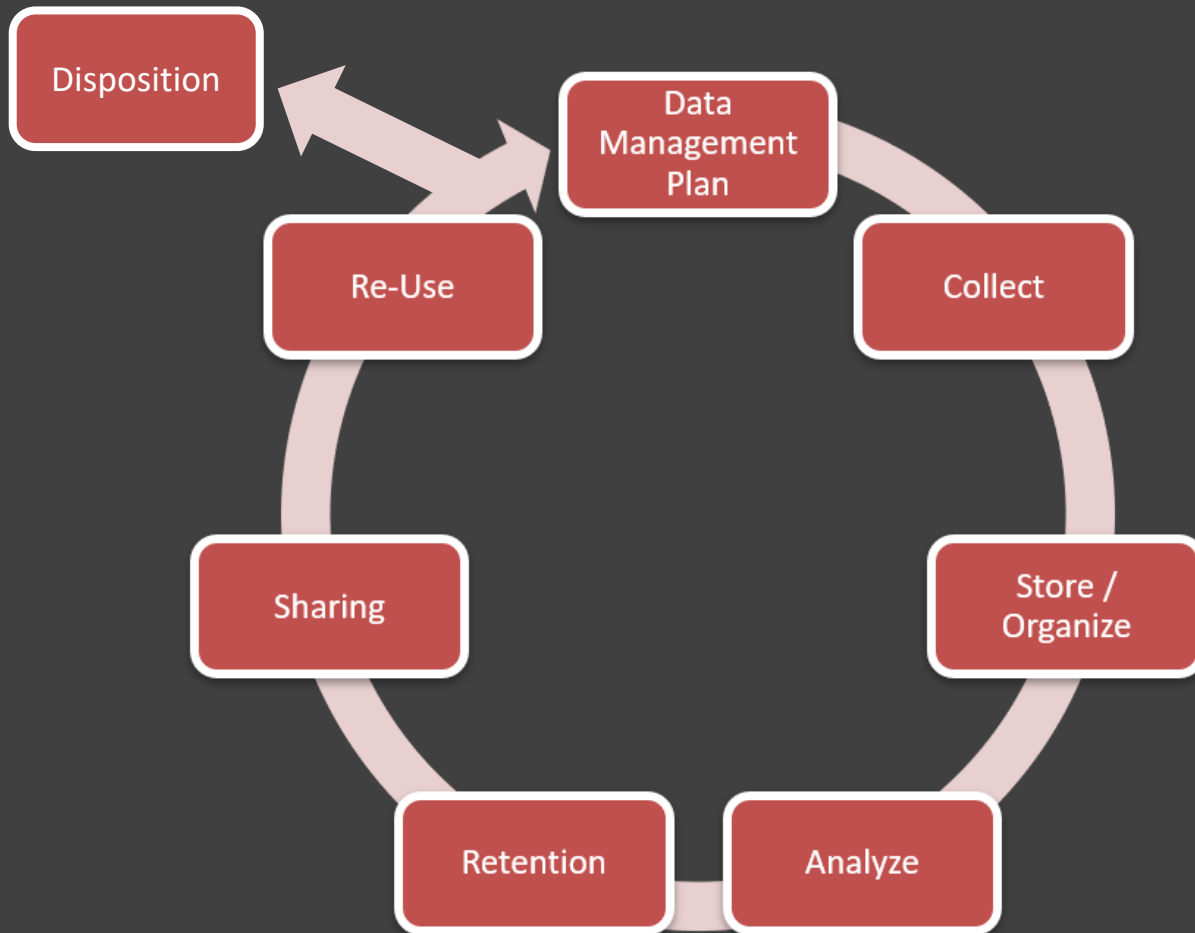
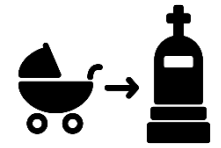
Topics

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

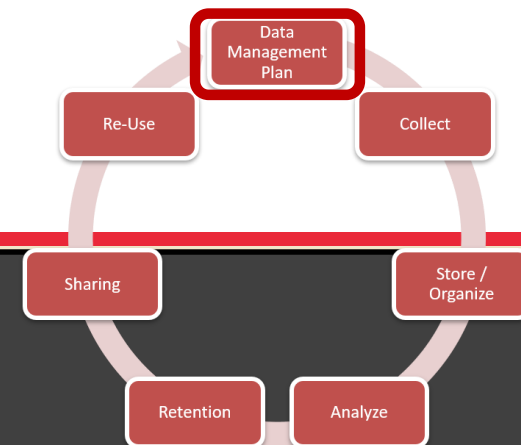


References provided on slides

Data Lifecycle



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

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



Source Data (Original)

First capture of information



ALCOA Principles

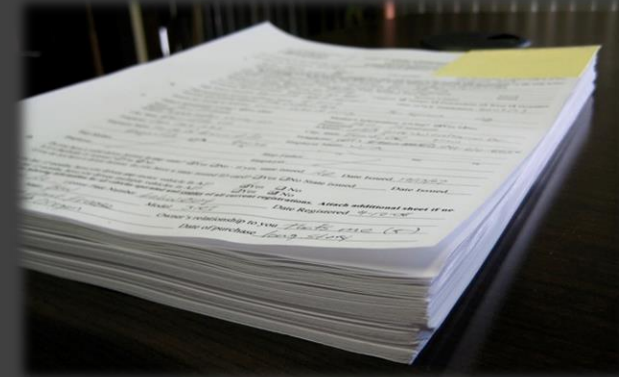
Applies to paper and/or electronic data

Data Quality

- Atributable
- Legible
- Contemporaneous
- Original
- Accurate

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 Draft Advisory Document on GLP Data Integrity; 7 August 2020

Data Risk - Non-enduring



Scenario 1 (Worst case):

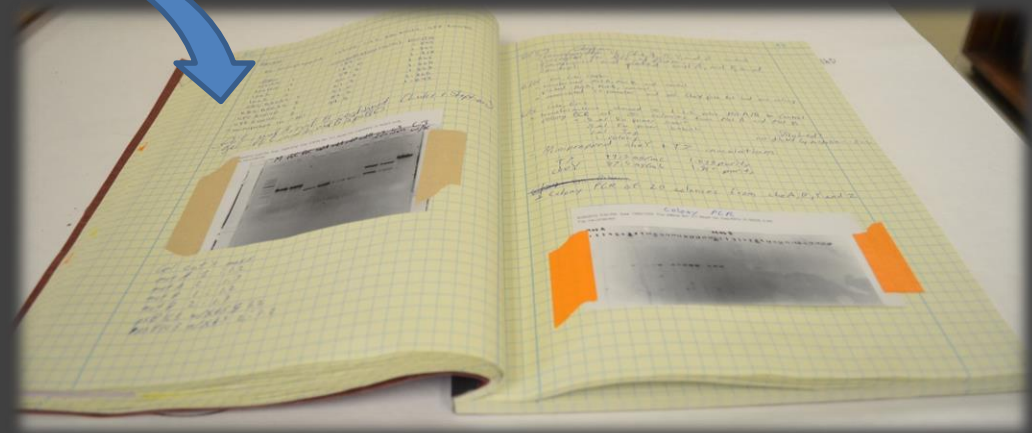
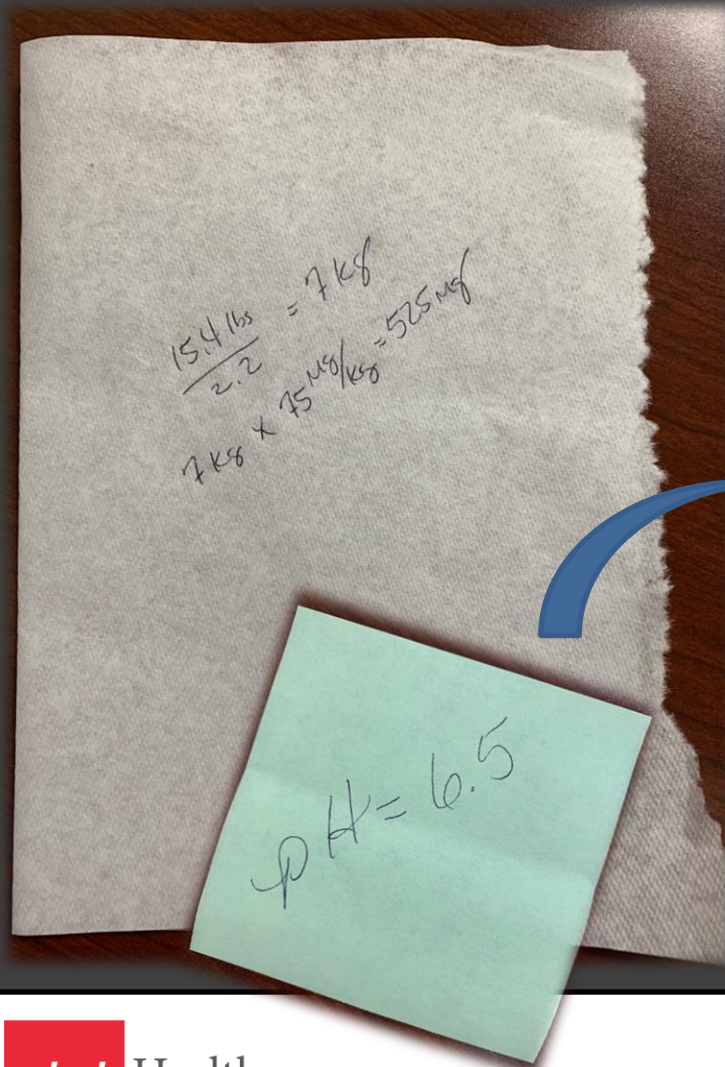


Scenario 2: Transcribe



Scenario 3 (Lower Risk): Affix

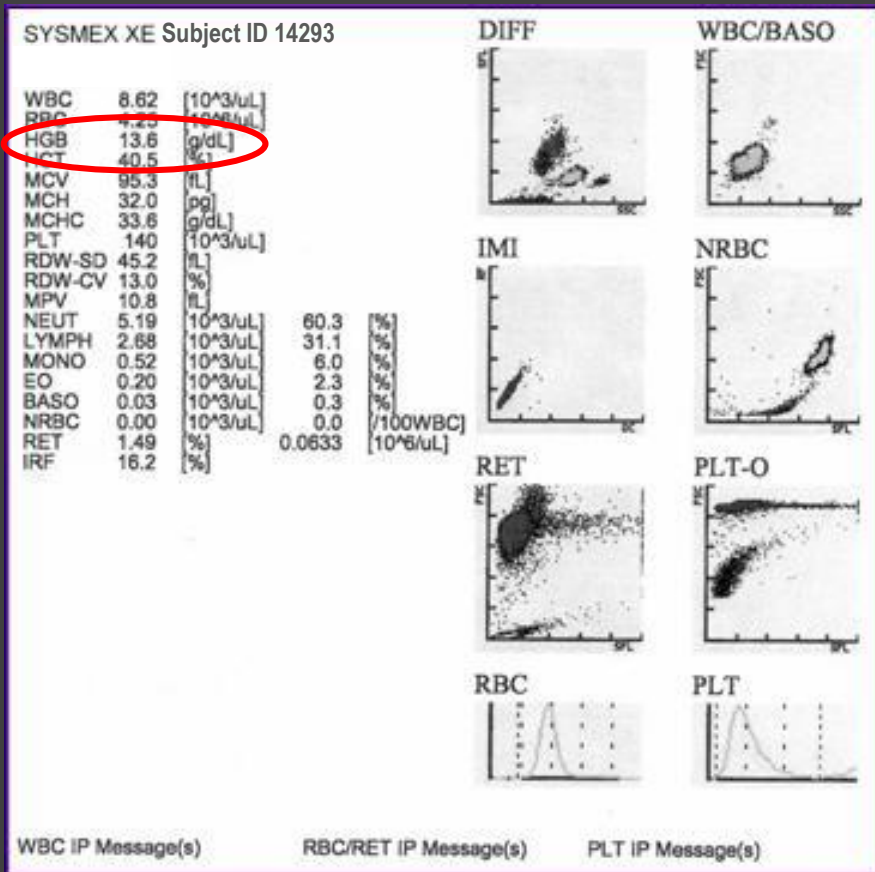
Scenario 4: Eliminate bad practice



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



May 29, 2011

5/3/2022

Data Quality/Reproducibility Exercise

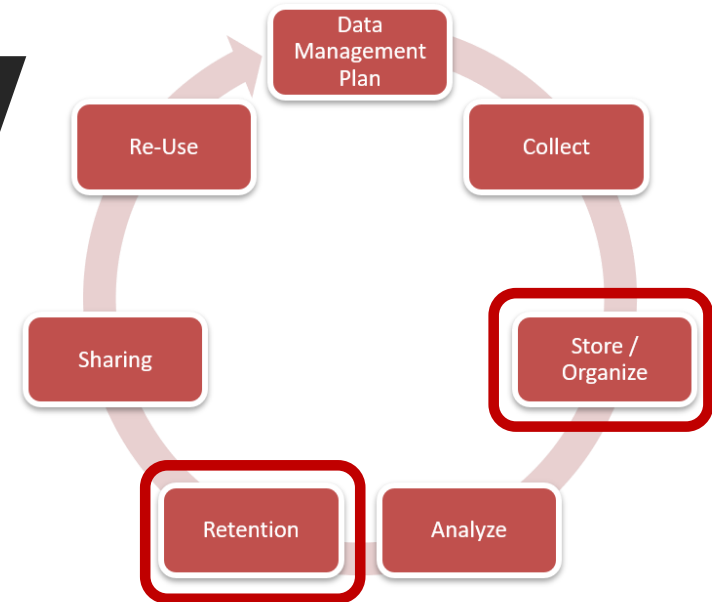
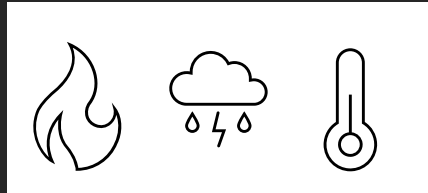


Organization and Storage / Retention

Things to think about prospectively...

Conditions

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




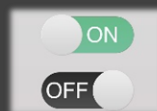
Contents

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



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

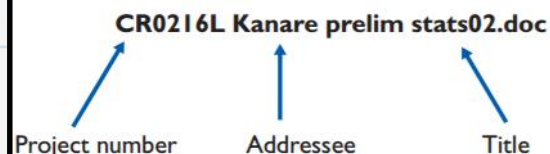
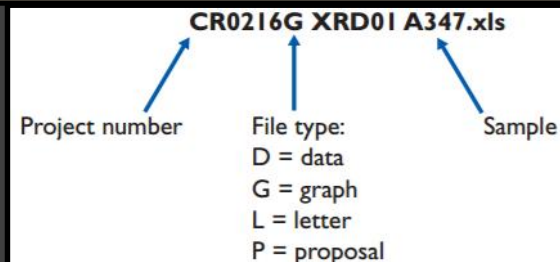


Documents library

Example.Study2018.031.0002

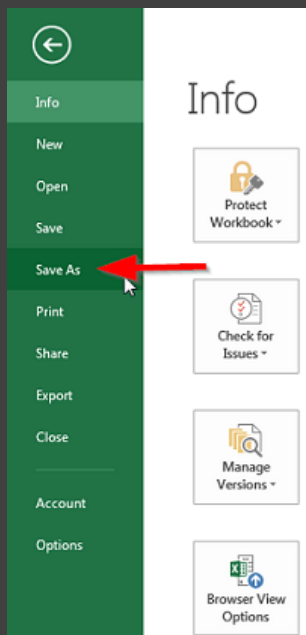
Name

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

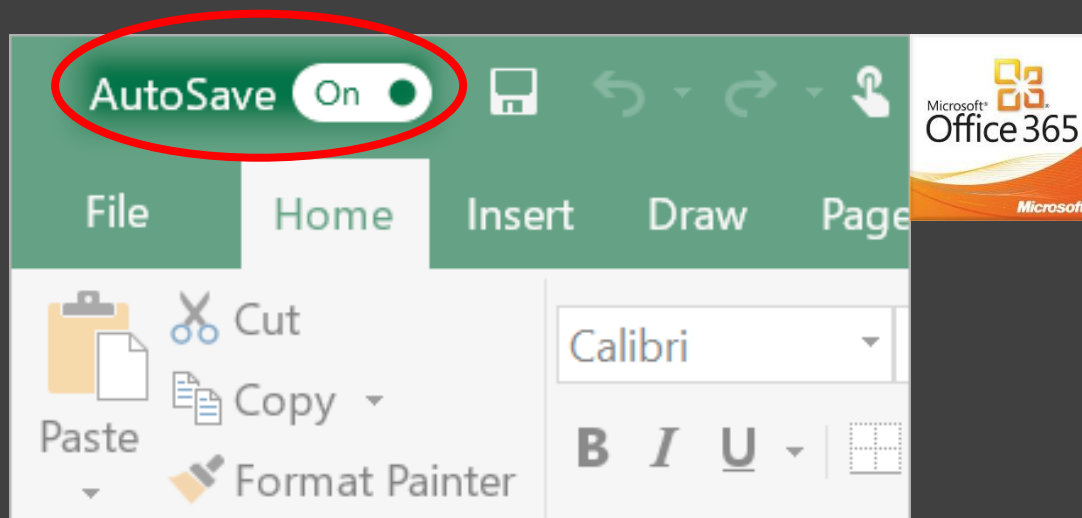


Risks to Electronic Data

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)

Data Analysis (Data Manipulation)

Prospective thinking...

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

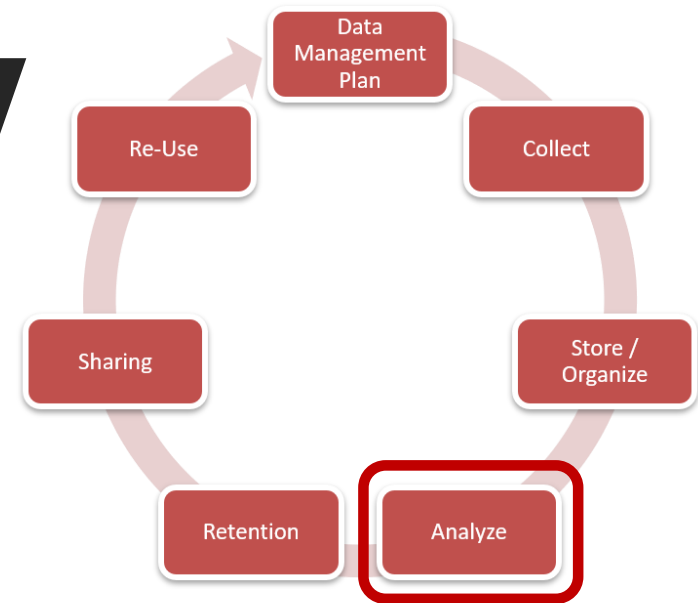
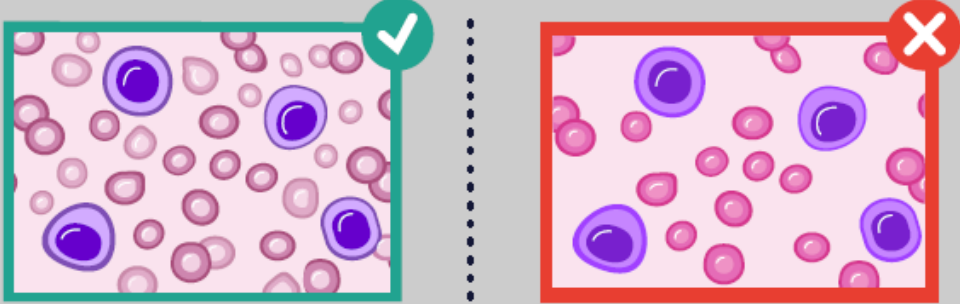


Image Manipulation

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

COLOR ENHANCEMENTS
Changing the contrast, color, or brightness



Ensure that the meaning of the image stays the same and fine details are not removed.

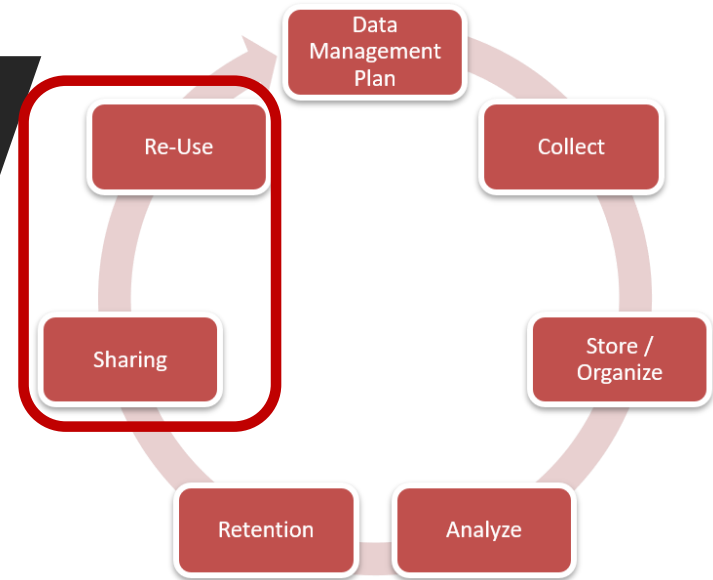
Contrast and saturation were increased causing the background cells to disappear.

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

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
["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://import.niaid.nih.gov/home>

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

Supplemental Information to the NIH Policy for Data Management and Sharing: Allowable Costs for Data Management and Sharing

Notice Number:
NOT-OD-21-015

Key Dates

Release Date: October 29, 2020

Related Announcements

[NOT-OD-23-053](#) - Reminder: NIH Policy for Data Management and Sharing effective on January 25, 2023.

[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-21-014](#) – Supplemental Information to the NIH Policy for Data Management and Sharing: Elements of an NIH Data Management and Sharing Plan

[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

Closing Thoughts...

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



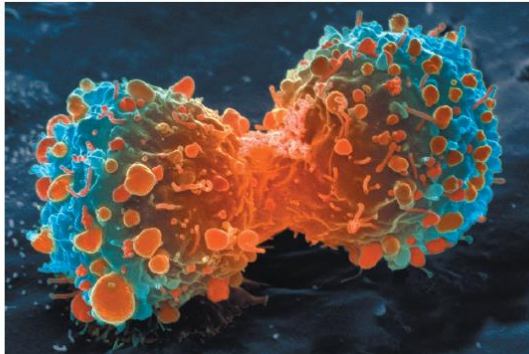
COMMENT

ANAR INFLUENZA Shift expertise to track mutations where they emerge **p.524**

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**



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.

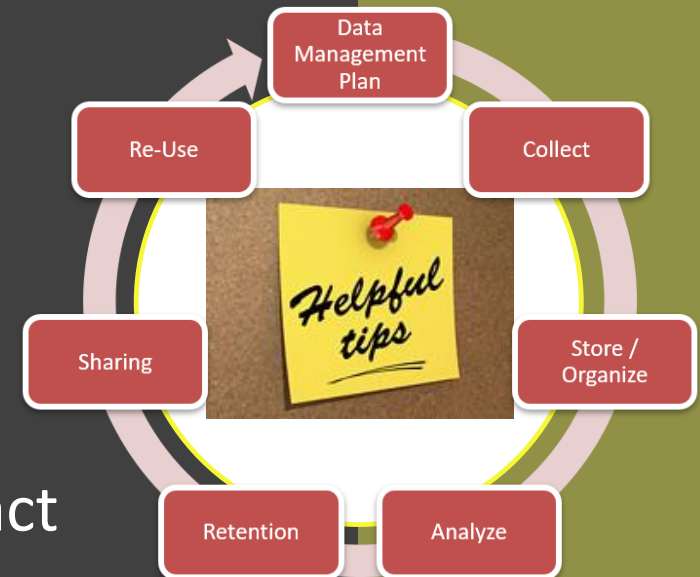
Efforts over the past decade to characterize the genetic alterations in human cancers have led to a better understanding of molecular drivers of this complex set of diseases. Although we in the cancer field hoped that this would lead to more effective drugs, historically, our ability to translate cancer research to clinical success has been remarkably low¹. 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 models² make it difficult for even

© 2012 Macmillan Publishers Limited. All rights reserved. 29 MARCH 2012 | VOL 483 | NATURE | 531

Data Management and Resource Sharing

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



Topics

- Principles, Guidelines, Policies, Definitions
- 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.

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



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