Data Management and Resource Sharing

Rigor & Reproducibility Worship
14 February 2018

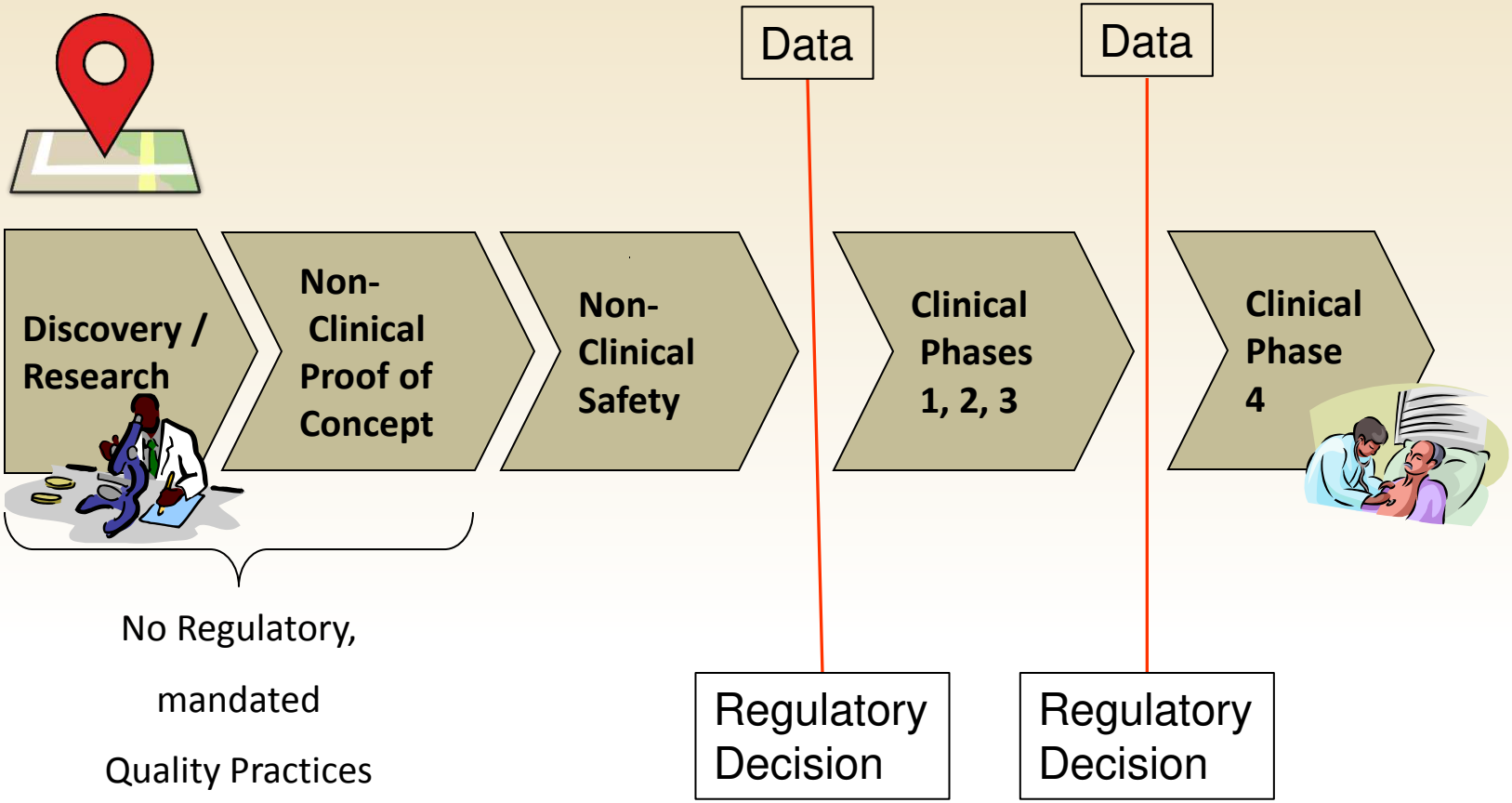
Melissa Eitzen, MT (ASCP), MS, RQAP-GLP
Director, Regulatory Operations
UTMB Institutional Office of Regulated Nonclinical Studies
mmeitzen@utmb.edu

utmb Health

Topics

- Translational Science
- Data Lifecycle and Organization
- Good Documentation Practices

Product Approval Pathway



Translational Science

What is the average cost to develop and gain marketing approval for a new drug?

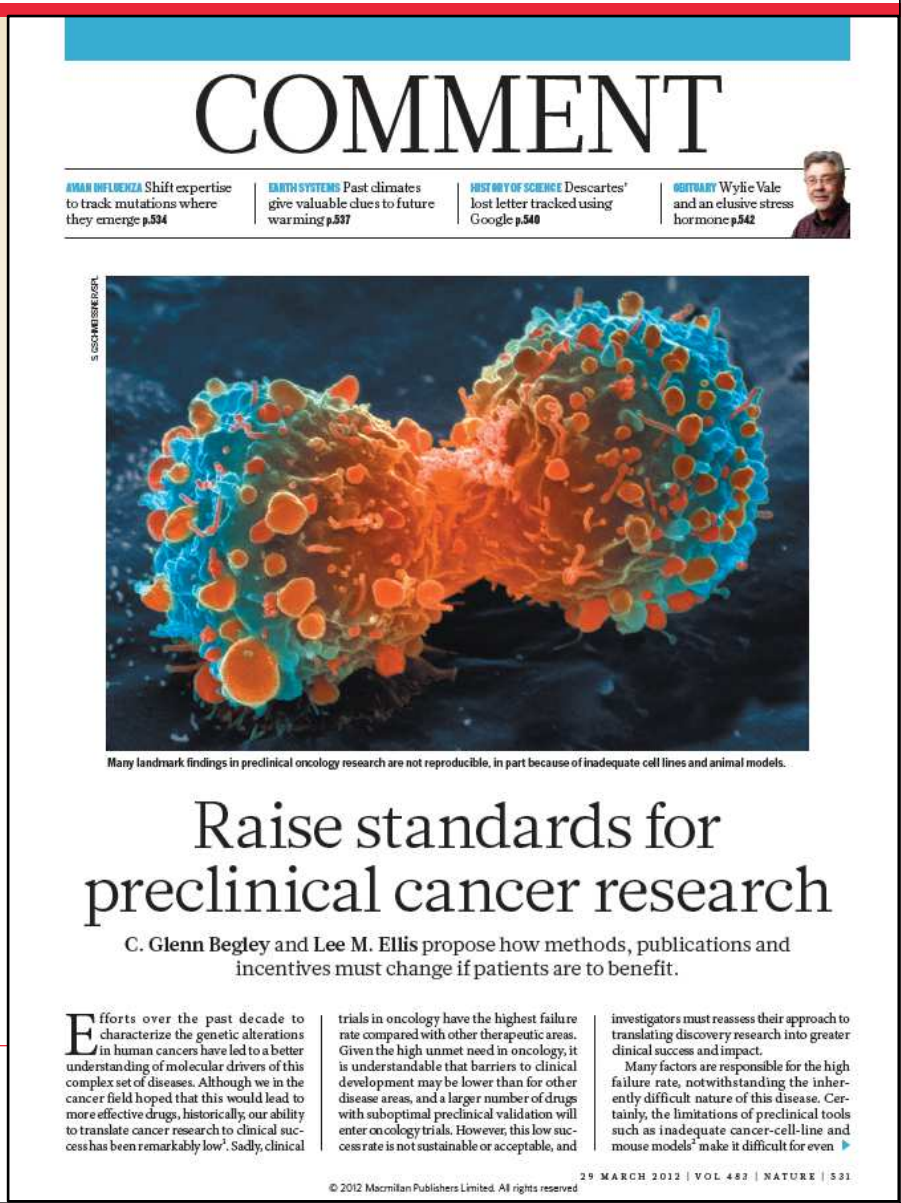
$2.558 billion

$2.870 billion, including post-approval costs

Journal of Health Economics 47 (2016) 20–33

Economics of Reproducibility

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



COMMENT

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

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 ▶

utmb Health

Focus Areas for Reproducibility

- Randomization
- Blinding
- Sample size
- Data handling

http://www.ncbi.nlm.nih.gov/pubmed/23060188

Published in final edited form as:
Nature. 2012 October 11; 490(7419): 187–191. doi:10.1038/nature11556.

A call for transparent reporting to optimize the predictive value of preclinical research

Story C. Landis[1], Susan G. Amara[2], Khusru Asadullah[3], Chris P. Austin[4], Robi Blumenstein[5], Eileen W. Bradley[6], Ronald G. Crystal[7], Robert B. Darnell[8], Robert J. Ferrante[9], Howard Fillit[10], Robert Finkelstein[1], Marc Fisher[11], Howard E. Gendelman[12], Robert M. Golub[13], John L. Goudreau[14], Robert A. Gross[15], Amelie K. Gubitz[1], Sharon E. Hesterlee[16], David W. Howells[17], John Huguenard[18], Katrina Kelner[19], Walter Koroshetz[1], Dimitri Krainc[20], Stanley E. Lazic[21], Michael S. Levine[22], Malcolm R. Macleod[23], John M. McCall[24], Richard T. Moxley III[25], Kalyani Narasimhan[26], Linda J. Noble[27], Steve Perrin[28], John D. Porter[1], Oswald Steward[29], Ellis Unger[30], Ursula Utz[1], and Shai D. Silberberg[1]

[1]National Institute of Neurological Disorders and Stroke, NIH, Bethesda, Maryland 20892, USA [2]Department of Neurobiology, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania 15213, USA [3]Bayer HealthCare, 13342 Berlin, Germany [4]National Center for Advancing Translational Sciences, NIH, Rockville, Maryland 20854, USA [5]CHDI Management/CHDI Foundation, New York, New York 10001, USA [6]Center for Review, NIH, Bethesda, Maryland 20892, USA [7]Department of Genetic Medicine, Weill Cornell Medical College, New York, New York 10021, USA [8]Howard Hughes Medical Institute, The Rockefeller University, New York, New York 10065, USA [9]Department of Neurological Surgery, University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA [10]Alzheimer's Drug Discovery Foundation, New York, New York 10019, USA [11]Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01545, USA [12]Department of Pharmacology and Experimental Neuroscience, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA [13]JAMA, Chicago, Illinois 60654, USA [14]Department of Neurology, Michigan State University, East Lansing, Michigan 48824, USA [15]Department of Neurology, University of Rochester Medical Center, Rochester, New York 14642, USA [16]Parent Project Muscular Dystrophy, Hackensack, New Jersey 07601, USA [17]The Florey Institute of Neuroscience and Mental Health, University of Melbourne, Heidelberg 3081, Australia [18]Neurology and Neurological Sciences and Cellular and Molecular Physiology, Stanford University, Stanford, California 94305, USA [19]Science Translational Medicine, AAAS, Washington DC 22201, USA [20]Department of Neurology, Harvard Medical School, Massachusetts General Hospital, Boston, Massachusetts 02114, USA [21]F. Hoffmann-La Roche, 4070 Basel, Switzerland [22]Department of Psychiatry and Biobehavioral Sciences, University of California Los Angeles, Los Angeles, California 90095, USA [23]Department of Clinical Neurosciences, University of Edinburgh, Western General Hospital, Edinburgh EH4 2XU, UK [24]PharMac LLC, Boca Grande, Florida 33921, USA [25]University of Rochester Medical Center, School of Medicine and Dentistry, Rochester, New York 14642, USA [26]Nature Neuroscience, New York, New York 10013, USA [27]Department of Neurological Surgery, University of California San Francisco, San Francisco, California 94143, USA [28]ALS Therapy Development Institute, Cambridge, Massachusetts 02139, USA [29]Reeve-Irvine Research Center,

Correspondence and requests for materials should be addressed to S.D.S. (silberbs@ninds.nih.gov).

Author Contributions R.F., A.K.G., S.C.L., J.D.P., S.D.S., U.U. and W.K. organized the workshop. R.B.D., S.E.L., S.C.L., M.R.M. and S.D.S. wrote the manuscript. All authors participated in the workshop and contributed to the editing of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper.

utmb Health

Resource Sharing—NIH

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.

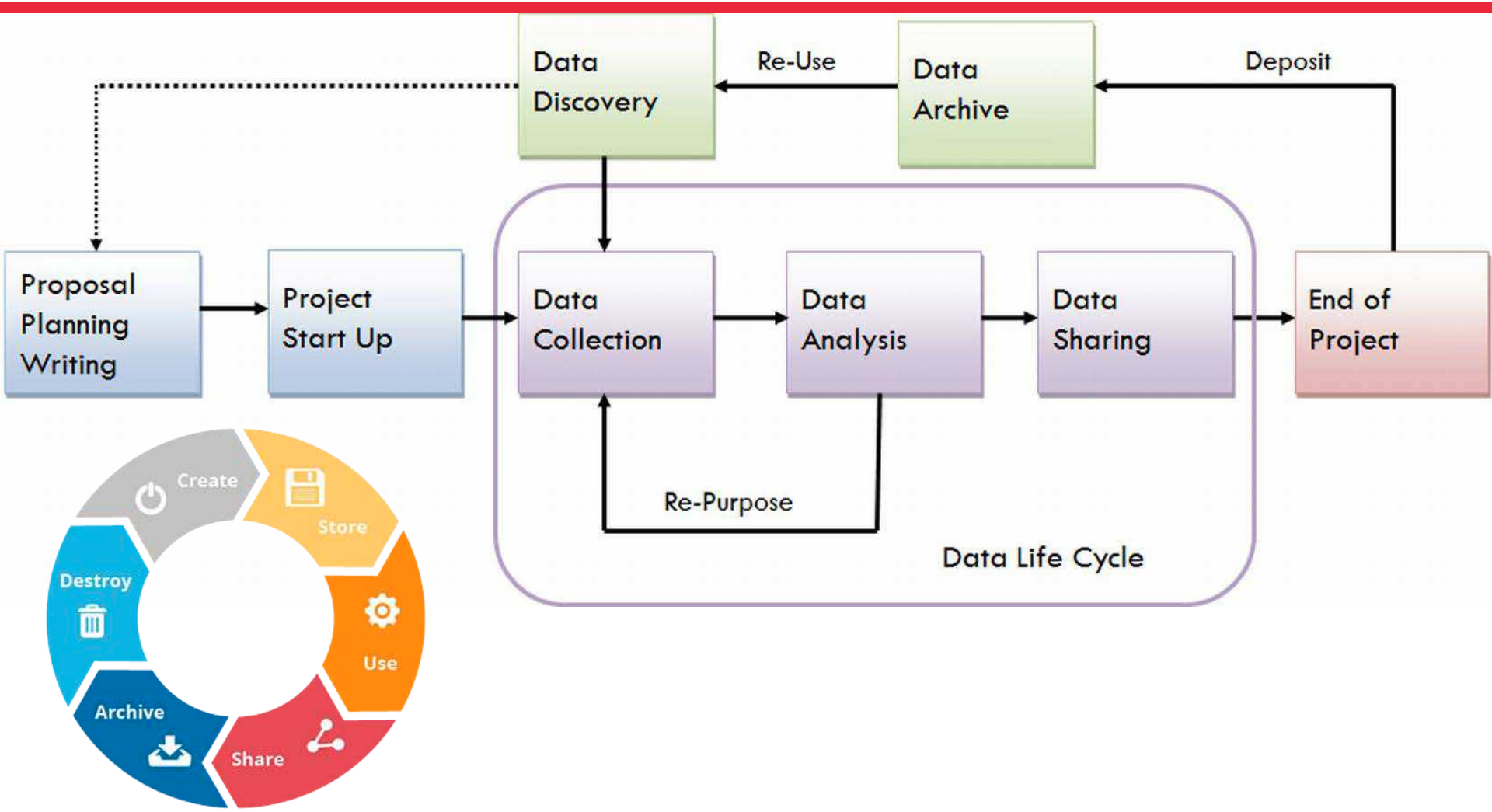
https://grants.nih.gov/grants/peer/guidelines_general/Resource_sharing_plans.pdf

Resource Sharing—NIH

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

Where Do We Start?

Data Lifecycle



What Kind of Data?

- Types (observational, derived, etc.)
- Format (text, numeric, modeling, images, etc.)
- Quantity
- HIPAA
- Proprietary
- Owner

Organization and Storage

- Location
- Accessibility / Security (including encryption)
- Back-up
- Retention / Archival
- Data Migration / Operating System Sustainability

Organization and Storage

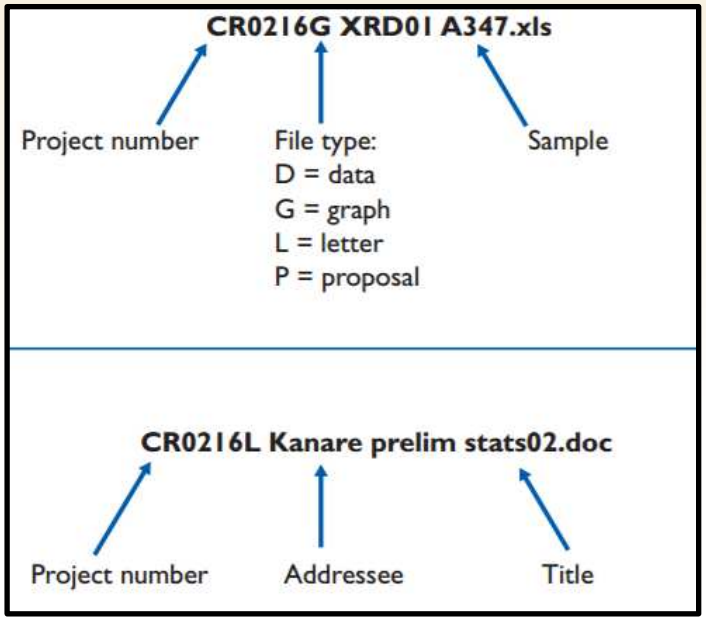
Electronic *Data*

- Audit Trails / Meta Data
- Transposition Errors
- Software Compatibility
- Program Updates
  - Automatic
  - Impact to significant digits

Organization and Storage

Electronic *Documents*

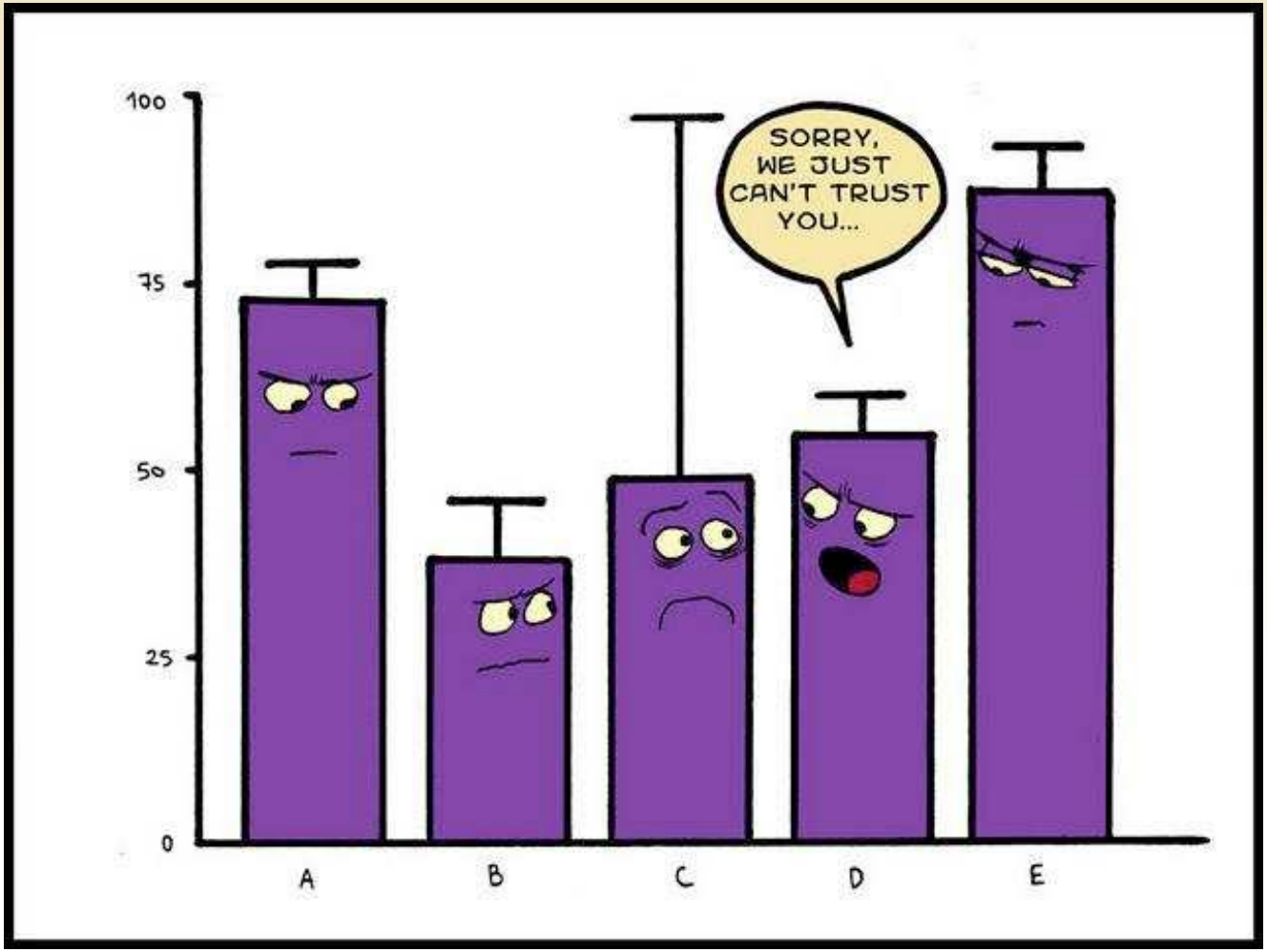
- Standard File Naming System
  - Brief, descriptive, consistent, dated
  - Plans for edits and changes



Organization and Storage

- Fixed (No changes)
- Updated with no changes
- Updated with changes—Change Control

Data Manipulation



Mechanisms for Data Sharing

- Email
- Online repositories
- Supplemental to publication
- Sharing agreements
- Mixed
- Conditions / Exclusions

Promoting an Open Research Culture

Data Sharing


SCIENTIFIC STANDARDS

Promoting an open research culture

Author guidelines for journals could help to promote transparency, openness, and reproducibility

Summary of the eight standards and three levels of the TOP guidelines

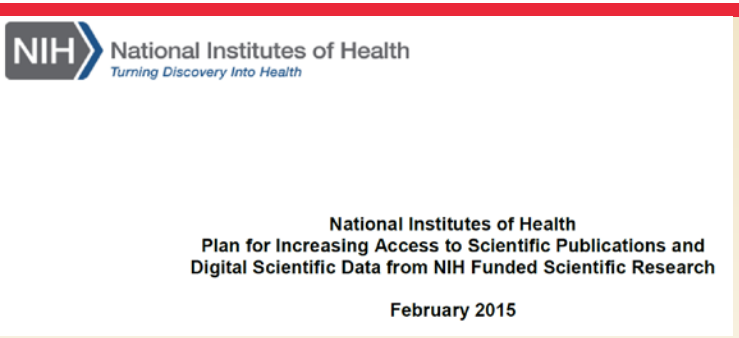
Levels 1 to 3 are increasingly stringent for each standard. Level 0 offers a comparison that does not meet the standard.

	LEVEL 0	LEVEL 1	LEVEL 2	LEVEL 3
Data transparency	Journal encourages data sharing—or says nothing.	Article states whether data are available and, if so, where to access them.	Data must be posted to a trusted repository. Exceptions must be identified at article submission.	Data must be posted to a trusted repository, and reported analyses will be reproduced independently before publication.

1424 26 JUNE 2015 • VOL 348 ISSUE 6242 sciencemag.org SCIENCE

Published by AAAS

Data Management Plan—NIH



Data Management Plans

Ensure that all extramural researchers receiving Federal grants and contracts for scientific research and intramural researchers develop data management plans, as appropriate, describing how they will provide for long-term preservation of, and access to, scientific data in digital formats resulting from federally funded research, or explaining why long-term preservation and access cannot be justified. (OSTP memo, element 4b)

Data Sharing Plan—Template

- What
- Who
- Where
- When
- How

Example Plan addressing Key Elements for a Data Sharing Plan under NIH Extramural Support
(For questions, contact the NIH Office of Extramural Research (OER), Email Sharing@nih.gov)

Example Data Sharing Plan for FOA-XX-XXXX

What data that will be shared:
I will share phenotypic data associated with the collected samples by depositing these data at _____ which is an NIH-funded repository. Genotype data will be shared by depositing these data at _____. Additional data documentation and de-identified data will be deposited for sharing along with phenotypic data, which includes demographics, family history of XXXXXX disease, and diagnosis, consistent with applicable laws and regulations. I will comply with the NIH GWAS Policy and the funding IC's existing policies on sharing data on XXXXXX disease genetics to include secondary analysis of data resulting from a genome wide association study through the repository. Meta-analysis data and associated phenotypic data, along with data content, format, and organization, will be available at _____. Submitted data will confirm with relevant data and terminology standards.

Who will have access to the data:
I agree that data will be deposited and made available through _____ which is an NIH-funded repository, and that these data will be shared with investigators working under an institution with a Federal Wide Assurance (FWA) and could be used for secondary study purposes such as finding genes that contribute to process of XXXXXX. I agree that the names and Institutions of persons either given or denied access to the data, and the bases for such decisions, will be summarized in the annual progress report. Meta-analysis data and associated phenotypic data, along with data content, format, and organization, will be made available to investigators through _____.

Where will the data be available:
I agree to deposit and maintain the phenotypic data, and secondary analysis of data (if any) at _____, which is an NIH-funded repository and that the repository has data access policies and procedures consistent with NIH data sharing policies.

When will the data be shared:
I agree to deposit genetic outcome data into _____ repository as soon as possible but no later than within one year of the completion of the funded project period for the parent award or upon acceptance of the data for publication, or public disclosure of a submitted patent application, whichever is earlier.

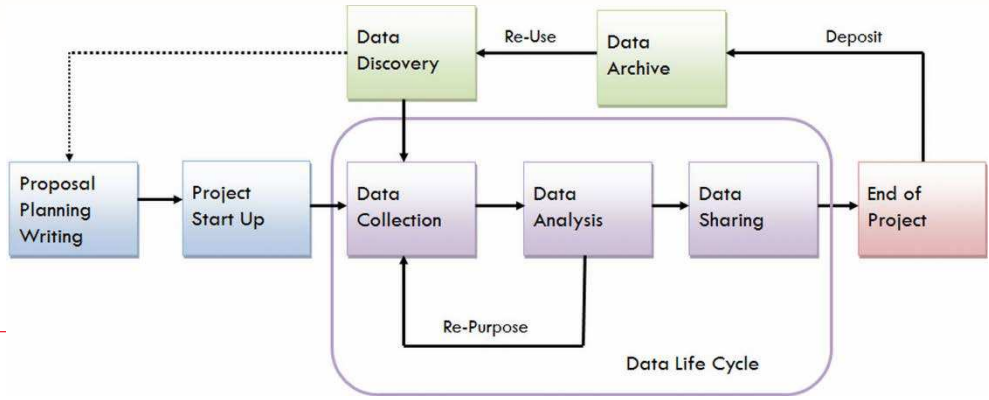
How will researchers locate and access the data:
I agree that I will identify where the data will be available and how to access the data in any publications and presentations that I author or co-author about these data, as well as acknowledge the repository and funding source in any publications and presentations. As I will be using _____, which is an NIH-funded repository, this repository has policies and procedures in place that will provide data access to qualified researchers, fully consistent with NIH data sharing policies and applicable laws and regulations.

Rev. 20100831

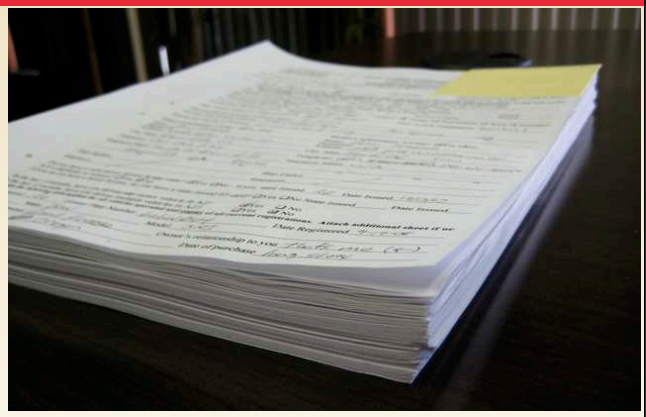
https://www.niaid.nih.gov/research/sample-data-sharing-plan

Data Management and Sharing Plan

- ❑ Data Description / Types
- ❑ Data Standards for Format and Content
- ❑ Mechanisms for Access and Sharing (Provisions, Privacy Protection, confidentiality, Security, Intellectual Property, etc.)
- ❑ Provisions for Data Reuse and Redistribution
- ❑ Archiving / Long-term Preservation and access
- ❑ Other…



Data Quality: GDP Principles



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

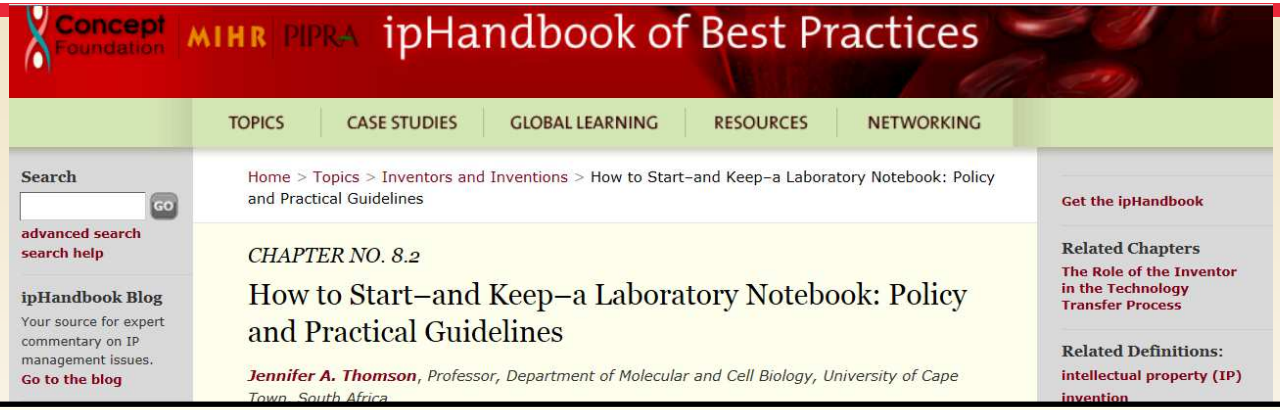
ALCOA+

Complete, Consistent, Enduring, Readily Available

Data Quality: Exercise

- ❑ Month / Day / Year
- ❑ Day / Month / Year

Laboratory Notebooks



1. What is a Laboratory Notebook?

Although you may think you will remember what you did and why you did a certain experiment in a week's time, YOU WILL NOT! And nor will anyone else in your laboratory. Hence the need for laboratory notebooks. In short, a laboratory notebooks is:

- a daily record of every experiment you do, think of doing, or plan to do
- a daily record of your thoughts about each experiment and the results thereof
- the basis of every paper and thesis you write
- the record used by patent offices and, in the case of disputes, courts of law (in the event you file patents on your findings)
- a record that would enable successive scientists, working on the same project, to pick up where you left off or reproduce your results

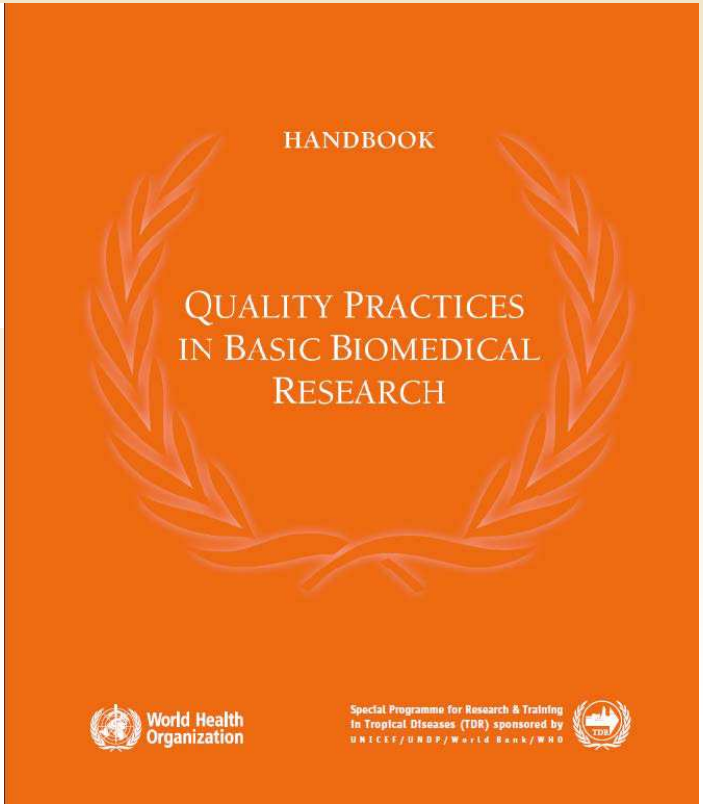
http://www.iphandbook.org/handbook/ch08/p02/

Reconstructability

4.3 Documentation

Making a full record of all information is essential not only to permit appropriate scientific interpretation of the results but also to enable complete reconstruction of the study, should this be necessary. Documentation is the only way of demonstrating what actually went on at the time of the experiment. *Without documentation the process is meaningless; essentially there has been no study.*

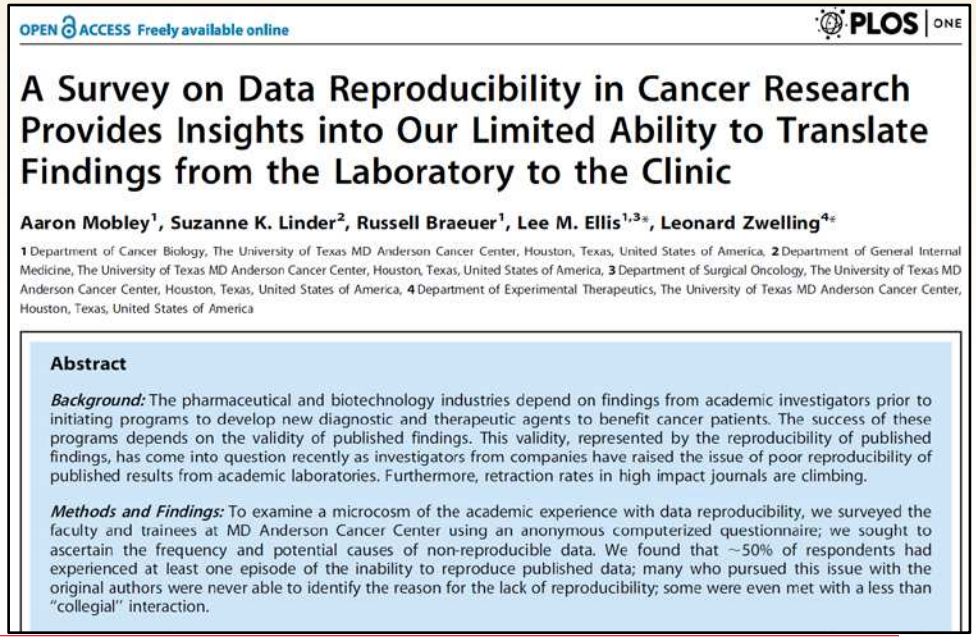
"One group is attempting to reproduce the results of another group. The group is unable to reproduce the same results even after following the methods. After contacting the original author withheld crucial information out of the methods section."



http://www.who.int/tdr/publications/documents/quality_practices.pdf?ua=1

Closing Comment—Data Integrity

"When data are published, the authors must be fully responsible for its integrity, and thus must be prepared to discuss their findings and help investigators troubleshoot experiments when others are unable to reproduce their work…"



OPEN ACCESS Freely available online

PLOS ONE

A Survey on Data Reproducibility in Cancer Research Provides Insights into Our Limited Ability to Translate Findings from the Laboratory to the Clinic

Aaron Mobley[1], Suzanne K. Linder[2], Russell Braeuer[1], Lee M. Ellis[1,3]*, Leonard Zwelling[4]*

1 Department of Cancer Biology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **2** Department of General Internal Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **3** Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America, **4** Department of Experimental Therapeutics, The University of Texas MD Anderson Cancer Center, Houston, Texas, United States of America

Abstract

Background: The pharmaceutical and biotechnology industries depend on findings from academic investigators prior to initiating programs to develop new diagnostic and therapeutic agents to benefit cancer patients. The success of these programs depends on the validity of published findings. This validity, represented by the reproducibility of published findings, has come into question recently as investigators from companies have raised the issue of poor reproducibility of published results from academic laboratories. Furthermore, retraction rates in high impact journals are climbing.

Methods and Findings: To examine a microcosm of the academic experience with data reproducibility, we surveyed the faculty and trainees at MD Anderson Cancer Center using an anonymous computerized questionnaire; we sought to ascertain the frequency and potential causes of non-reproducible data. We found that ~50% of respondents had experienced at least one episode of the inability to reproduce published data; many who pursued this issue with the original authors were never able to identify the reason for the lack of reproducibility; some were even met with a less than "collegial" interaction.

May 2015, Vol. 8, Issue 5, e63221
www.Plosone.org

Case Study 1—Data Sharing

Your research study 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, you will be collecting identifying information.

Even though the final dataset will be stripped of identifiers prior to release for sharing, there remains the possibility of deductive disclosure of subjects with unusual characteristics. Identify options (i.e., conditions) for sharing the data.

Case Study 1—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.

http://guides.lib.vt.edu/NIH/examples

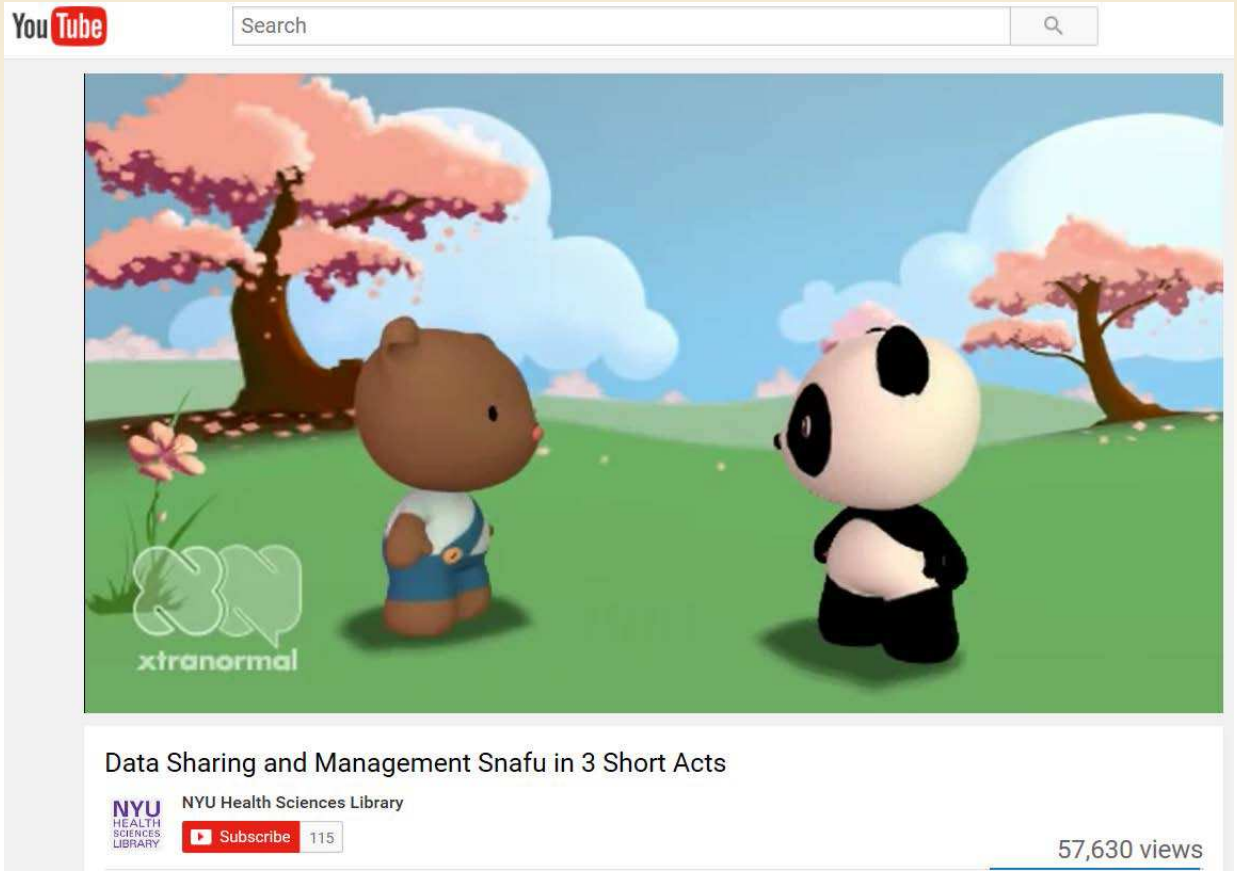
Case Study 2—Data Retention

One of the important elements about rigor and reproducibility is knowing exactly how to "reproduce" conducting an experiment. This in turn depends upon access to primary records of research studies. Sometimes, it is difficult to obtain the primary lab/research records after investigators leave an institution.

What are best practices to insure that primary data and precise methodologies are maintained in an EASILY accessible format for required time periods and maybe even longer?

(Homework): Watch the Data Management and Sharing Video

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



References – Data Management

https://dmptool.org/dm_guidance (General guidance)

http://www.icpsr.umich.edu/files/datamanagement/DataManagementPlans-All.pdf (Guidelines)

http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/plan.html (Sample plan)

http://lgdata.s3-website-us-east-1.amazonaws.com/docs/2784/1033892/DM101Checklist101314.pdf (Checklist)

http://libguides.northwestern.edu/datamanagement/federalfundingagency (Other Federal funding agencies)

http://www.hhmi.org/sites/default/files/Educational%20Materials/Lab%20Management/Making%20the%20Right%20Moves/moves2_ch8.pdf (Data Management and Lab Notebooks)

https://grants.nih.gov/grants/NIH-Public-Access-Plan.pdf (NIH Public Access Plan)

http://www.data-archive.ac.uk/media/2894/managingsharing.pdf (Managing and Sharing Data, UK Data Archive)

http://data.library.virginia.edu/data-management/lifecycle/ (Data lifecycle)

Hyperlinks verified 12 Feb 2018

References—Resource Sharing

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_guidance.htm (Policy and Implementation)

https://grants.nih.gov/grants/policy/data_sharing/data_sharing_faqs.htm (FAQs)

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

https://www.niaid.nih.gov/grants-contracts/resource-sharing-plan (Resource sharing)

Other References:

http://www.iphandbook.org/handbook/ch08/p02/ (Laboratory Notebook)

http://www.who.int/tdr/publications/documents/quality_practices.pdf?ua=1 (WHO Handbook)

Hyperlinks verified 12 February 2018