NYU HEALTH SCIENCES LIBRARY

Perspectives in Research Data Management

MLA Continuing Education
October 10, 2017

Alisa Surkis PhD, MLS
Assistant Director of Research Data and Metrics

Kevin Read MLIS, MAS
Knowledge Management Librarian

Course Schedule

1. **Introduction**
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

Objectives: Knowledge Gained

NYU HEALTH SCIENCES LIBRARY

Understand:

Current roles libraries play in research data management (RDM)

Research process

Differences between clinical and bench science researchers and their RDM needs

Current climate around data management and sharing

Best practices in data documentation and description

Relevance of standards to data management

Issues in storage, preservation, and sharing of data

Objectives: Skills Obtained

NYU HEALTH SCIENCES LIBRARY

Be able to:

Conduct data interviews

Assist researchers with data management best practices

Evaluate data sharing options for researchers

Develop a strategy for implementing a data management service at your institution

First word you think of when you hear the word **data?**

First word you think of when you hear data

NYU HEALTH SCIENCES LIBRARY

Cool

Numbers

Statistics

Science

Information

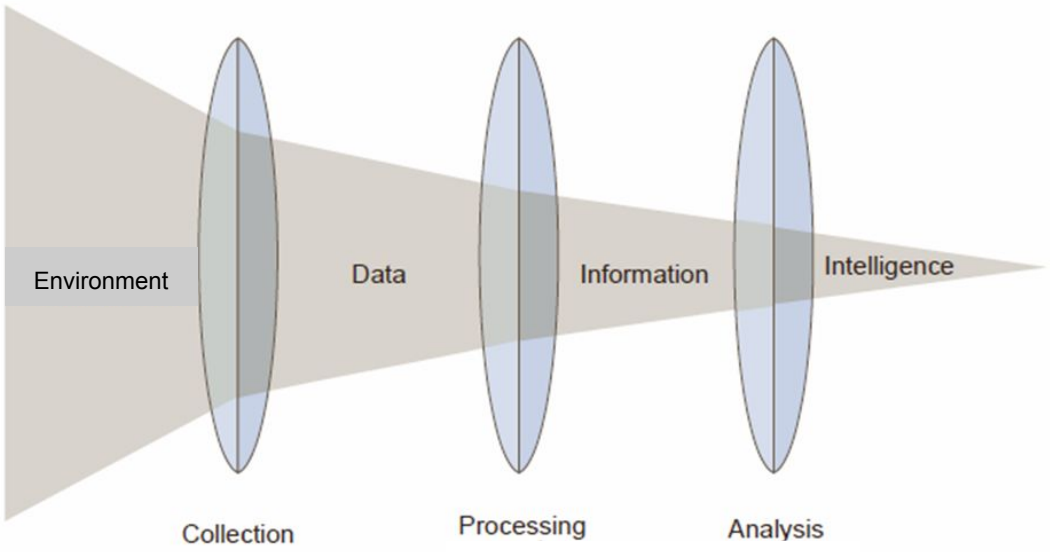
Necessary

Storage

Big

First word you think of when you hear data

Relationship of Data, Information and Intelligence



Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. **Current library roles in RDM**
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

NYU HEALTH SCIENCES LIBRARY

What roles should librarians play in **RDM**?

?

Support **access** to data?

Support **the use** of data?

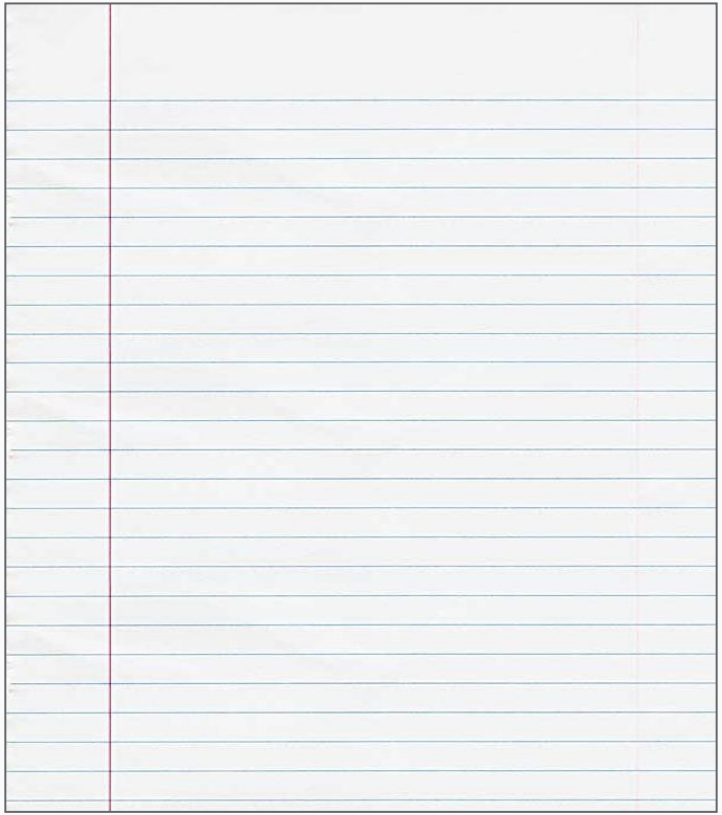
Support **preservation** of data?

Hard to support RDM once the research is over...

"here's all of my data"

“Here’s all of my documentation”

Or worse: “I have no documentation”



How have libraries responded?

Institutional Repositories

NEWS FEATURE DATA SHARING

NATURE|Vol 461|10 September 2009

Empty archives

Most researchers agree that open access to data is the scientific ideal, so what is stopping it happening? **Bryn Nelson** investigates why many researchers choose not to share.

Data management plans

National Science Foundation
WHERE DISCOVERIES BEGIN

QUICK LINKS

SEARCH

HOME | FUNDING | AWARDS | DISCOVERIES | NEWS | PUBLICATIONS | STATISTICS | ABOUT NSF | FASTLANE

Office of Budget, Finance and Award Management (BFA)

- DIAS Home
- CAAR Branch
- Policy Office
- Systems Office
- View DIAS Staff

Search DIAS Staff

Dissemination and Sharing of Research Results

NSF Data Sharing Policy

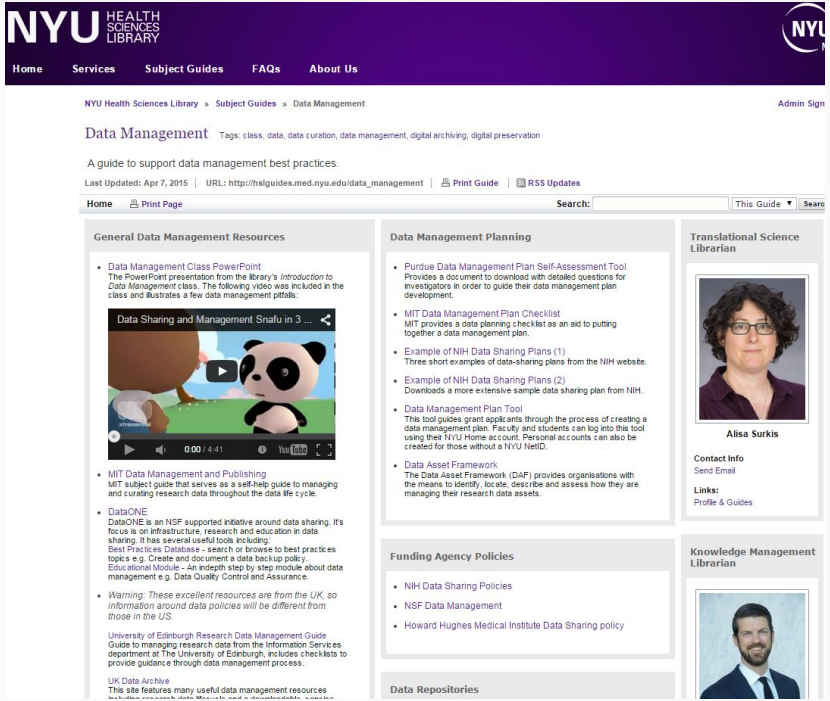
Investigators are expected to share with other researchers, at no more than incremental cost and within a reasonable time, the primary data, samples, physical collections and other supporting materials created or gathered in the course of work under NSF grants. Grantees are expected to encourage and facilitate such sharing. See Award & Administration Guide (AAG) Chapter VI.D.4.

NSF Data Management Plan Requirements

Proposals submitted or due on or after January 18, 2011, must include a supplementary document of no more than two pages labeled "Data Management Plan". This supplementary document should describe how the proposal will conform to NSF policy on the dissemination and sharing of research results. See Grant Proposal Guide (GPG) Chapter II.C.2.j for full policy implementation.

Adapting reference services

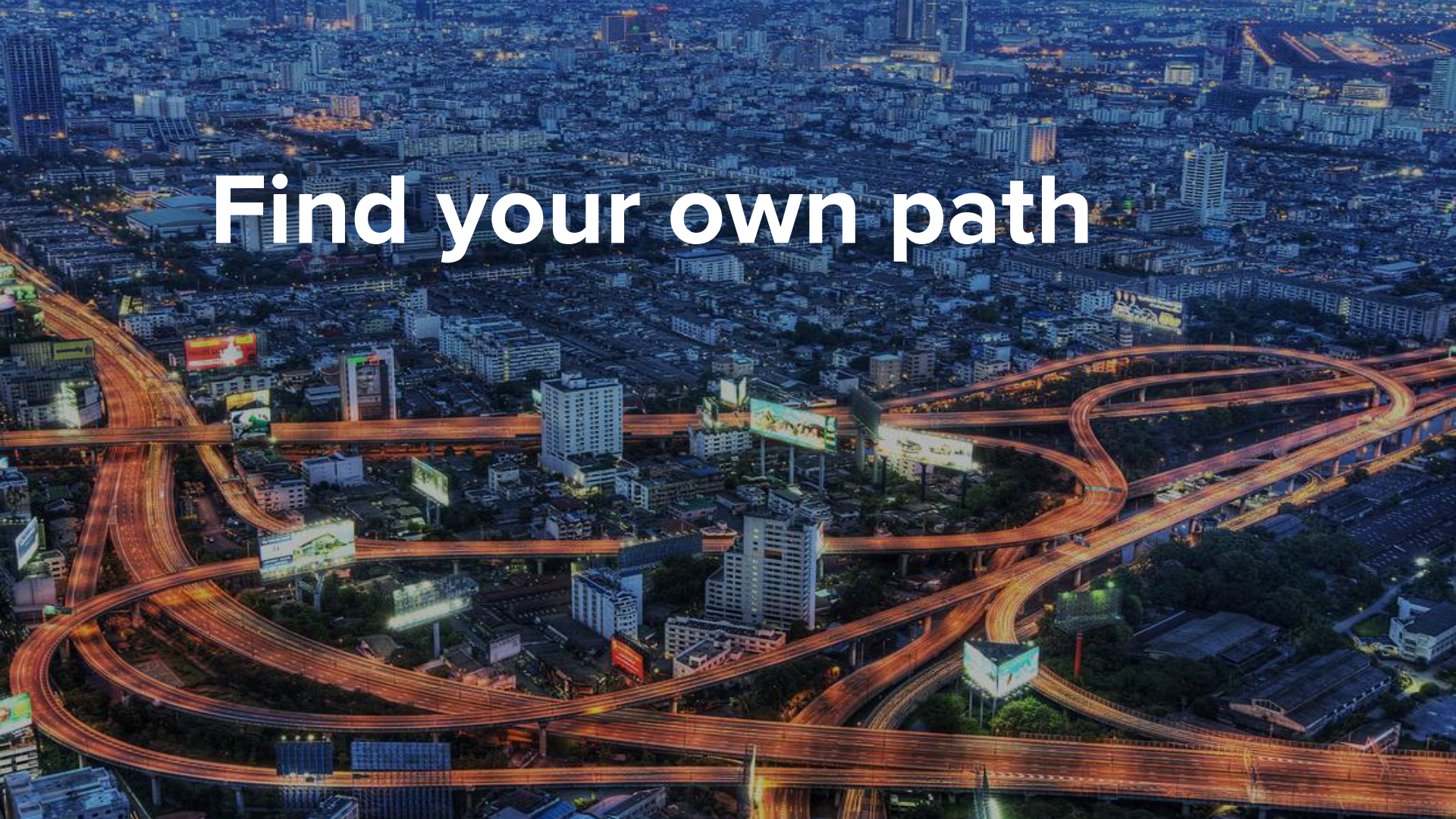
LibGuides


NYU HEALTH SCIENCES LIBRARY
Home
Services
Subject Guides
FAQs
About Us
NYU Health Sciences Library » Subject Guides » Data Management
Admin Sign
Data Management
Tags: class, data, data curation, data management, digital archiving, digital preservation
A guide to support data management best practices.
Last Updated: Apr 7, 2015 | URL: http://hslguides.med.nyu.edu/data_management | Print Guide | RSS Updates
Home
Print Page
Search:
This Guide
General Data Management Resources
Data Management Class PowerPoint
The PowerPoint presentation from the library's Introduction to Data Management class. The following video was included in the class and illustrates a few data management pitfalls:
Data Sharing and Management Snafu in 3 ...
0:00 / 4:41
MIT Data Management and Publishing
MIT subject guide that serves as a self-help guide to managing and curating research data throughout the data life cycle.
DataONE
DataONE is an NSF supported initiative around data sharing. It's focus is on infrastructure, research and education in data sharing. It has several useful tools including:
Best Practices Database - search or browse to best practices topics e.g. Create and document a data backup policy.
Educational Module - An indepth step by step module about data management e.g. Data Quality Control and Assurance.
Warning: These excellent resources are from the UK, so information around data policies will be different from those in the US.
University of Edinburgh Research Data Management Guide
Guide to managing research data from the Information Services department at The University of Edinburgh, includes checklists to provide guidance through data management process.
UK Data Archive
This site features many useful data management resources
Data Management Planning
Purdue Data Management Plan Self-Assessment Tool
Provides a document to download with detailed questions for investigators in order to guide their data management plan development.
MIT Data Management Plan Checklist
MIT provides a data planning checklist as an aid to putting together a data management plan.
Example of NIH Data Sharing Plans (1)
Three short examples of data-sharing plans from the NIH website.
Example of NIH Data Sharing Plans (2)
Downloads a more extensive sample data sharing plan from NIH.
Data Management Plan Tool
This tool guides grant applicants through the process of creating a data management plan. Faculty and students can log into this tool using their NYU Home account. Personal accounts can also be created for those without a NYU NetID.
Data Asset Framework
The Data Asset Framework (DAF) provides organisations with the means to identify, locate, describe and assess how they are managing their research data assets.
Funding Agency Policies
NIH Data Sharing Policies
NSF Data Management
Howard Hughes Medical Institute Data Sharing policy
Data Repositories
Translational Science Librarian
Alisa Surkis
Contact Info
Send Email
Links:
Profile & Guides
Knowledge Management Librarian

Education

Liaison/Subject librarians

Informationist projects


Find your own path

CASE STUDY 1

Maternal smoking during pregnancy and newborn neurobehavior

Case Study: Goals

Understand researchers' study, data practices, and needs

Improve researchers' data collection and organization

Identify avenues for researchers to share their data

Course Schedule

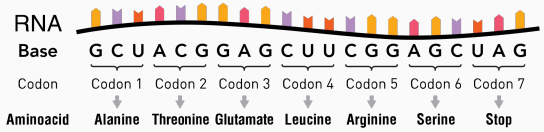
NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. **Story of data**
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

Data?

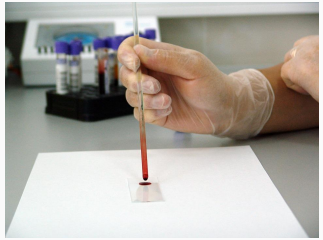
SubjectID	Age	SBP	DBP
001	30	130	70
002	24	145	80
003	28	120	180

Tables of numbers

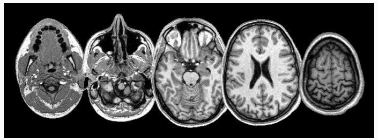


Sequences, base pairs

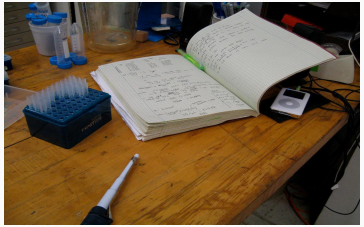
That means...



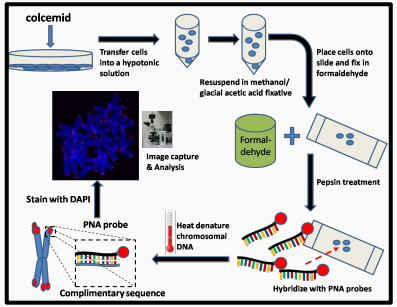
Samples, specimens, slides



Audio, video, imaging

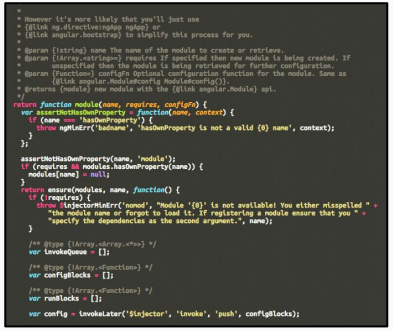


Lab notebooks



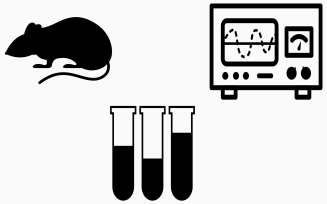
Protocols

Reproducibility



```
 *
 * However it's more likely that you'll just use
 * {@link ng.directive:ngApp ngApp} or
 * {@link angular.bootstrap} to simplify this process for you.
 *
 * @param {!string} name The name of the module to create or retrieve.
 * @param {!Array.<string=>} requires If specified then new module is being created. If
 *        unspecified then the module is being retrieved for further configuration.
 * @param {Function=} configFn Optional configuration function for the module. Same as
 *        {@link angular.Module#config Module#config()}.
 * @returns {module} new module with the {@link angular.Module} api.
 */
return function module(name, requires, configFn) {
  var assertNotHasOwnProperty = function(name, context) {
    if (name === 'hasOwnProperty') {
      throw ngMinErr('badname', 'hasOwnProperty is not a valid {0} name', context);
    }
  };

  assertNotHasOwnProperty(name, 'module');
  if (requires && modules.hasOwnProperty(name)) {
    modules[name] = null;
  }
  return ensure(modules, name, function() {
    if (!requires) {
      throw $injectorMinErr('nomod', "Module '{0}' is not available! You either misspelled " +
         "the module name or forgot to load it. If registering a module ensure that you " +
         "specify the dependencies as the second argument.", name);
    }

    /** @type {!Array.<Array.<*>>} */
    var invokeQueue = [];

    /** @type {!Array.<Function>} */
    var configBlocks = [];

    /** @type {!Array.<Function>} */
    var runBlocks = [];

    var config = invokeLater('$injector', 'invoke', 'push', configBlocks);
```

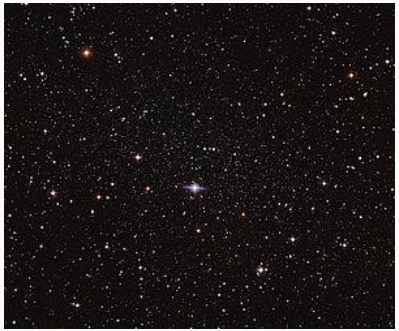
Software



And more...

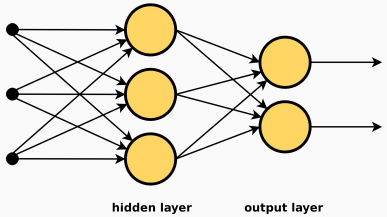
Categories of Data

Observational

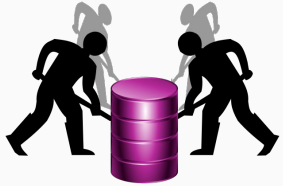


Experimental

Computational



Derived/Compiled



The story of data

The story of data

Identify:

Raw data

Transformed data

Analyzed data

What processes create these data?

The story can be simple

The story can be simple

Create

The story can be simple

Create

The story can be simple

Create

Process

Patient ID	TumorArea PreTreat (mm^2)	TumorArea PostTreat (mm^2)	Number Therapy Sessions
1001	454	317	4
1002	234	82	7

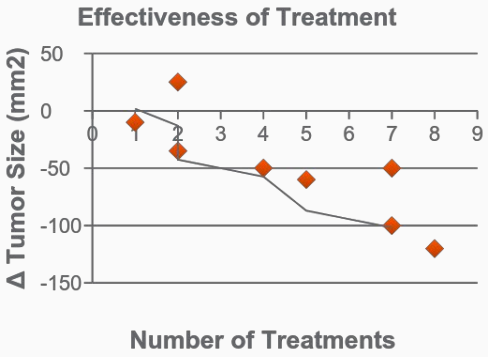
The story can be simple

Create

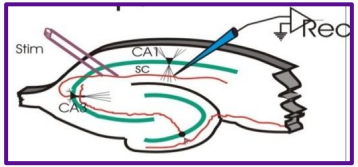
Process

Patient ID	TumorArea PreTreat (mm^2)	TumorArea PostTreat (mm^2)	Number Therapy Sessions
1001	454	317	4
1002	234	82	7

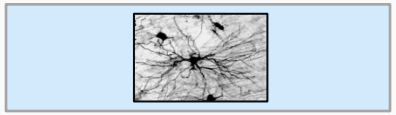
Analyze



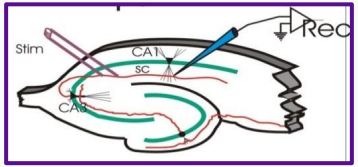
The story can be complex



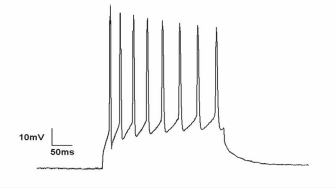
The story can be complex



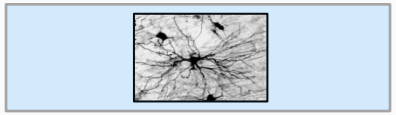
Create



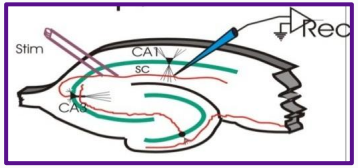
Create



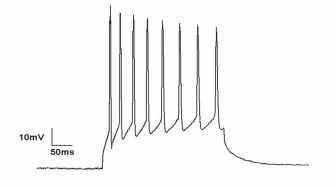
The story can be complex



Create



Create



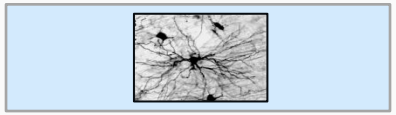
Process

X (µm)	Y (µm)	Z (µm)	Thickness (µm)
34	55	75	1.2
36	58	75	1.3

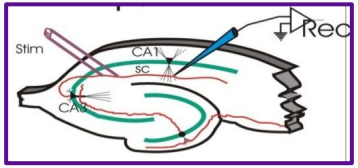
Process

Time (ms)	Signal (mV)
0.1	4.2
0.2	4.3
0.3	4.8

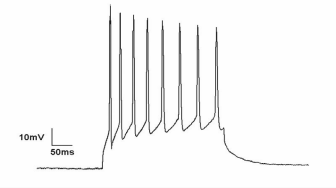
The story can be complex



Create



Create



Process

X (µm)	Y (µm)	Z (µm)	Thickness (µm)
34	55	75	1.2
36	58	75	1.3

Process

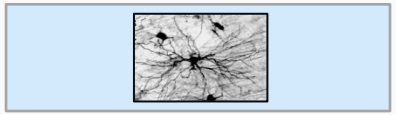
Time (ms)	Signal (mV)
0.1	4.2
0.2	4.3
0.3	4.8

Analyze

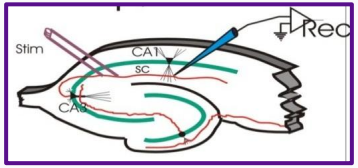
Analyze

Model 1 • • • Model 37 • • • Model 300

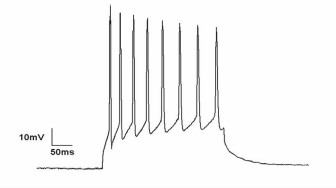
The story can be complex



Create



Create



Process

X (µm)	Y (µm)	Z (µm)	Thickness (µm)
34	55	75	1.2
36	58	75	1.3

Process

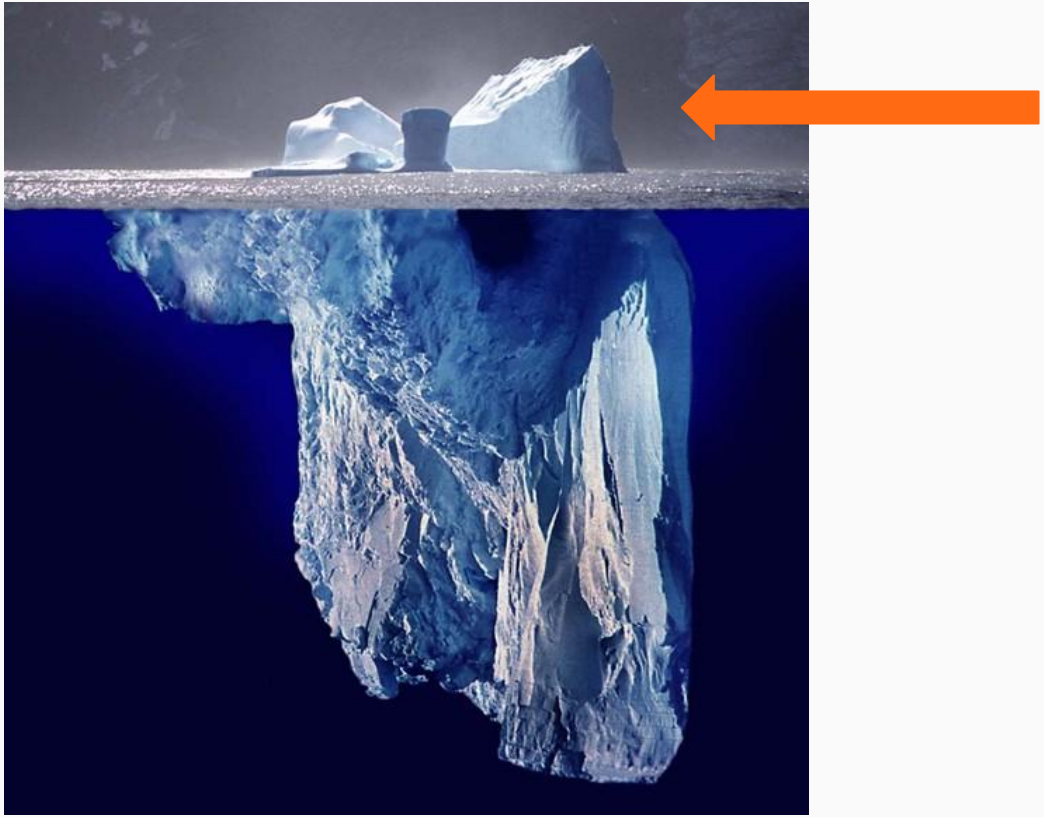
Time (ms)	Signal (mV)
0.1	4.2
0.2	4.3
0.3	4.8

Analyze

Analyze

Model 1 • • • **Model 37** • • • Model 300

Data in an article



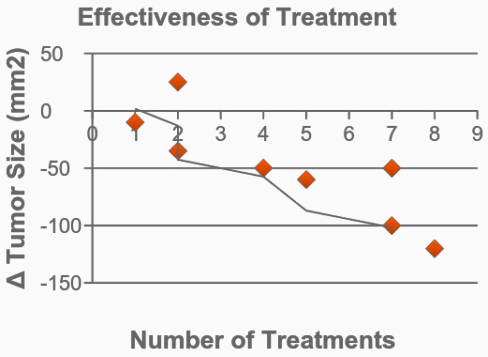
Where does all the data go?

Create

Process

Patient ID	TumorArea PreTreat (mm^2)	TumorArea PostTreat (mm^2)	Number Therapy Sessions
1001	454	317	4
1002	234	82	7

Analyze



Where does all the data go?

Create

Images folder

Process

Effectiveness of Treatment

Δ Tumor Size (mm2)

50

0

Figures folder

Analyze

Spreadsheets folder

Patient ID	Number Therapy Sessions
1001	4
1002	7

Data Lifecycle

NYU HEALTH SCIENCES LIBRARY

Idea → Process Methods → Create Collect Assemble → Processing Analysis → Findings → Disseminate

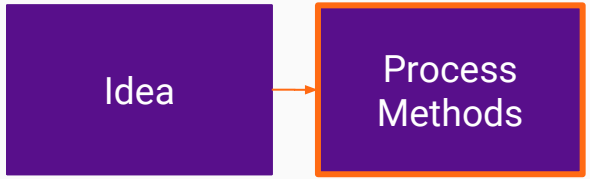
Data Lifecycle: Idea

NYU HEALTH SCIENCES LIBRARY

Idea

- ➔ Hypothesis
- ➔ Methodology
- ➔ Exploration

Data Lifecycle: Process & Methods

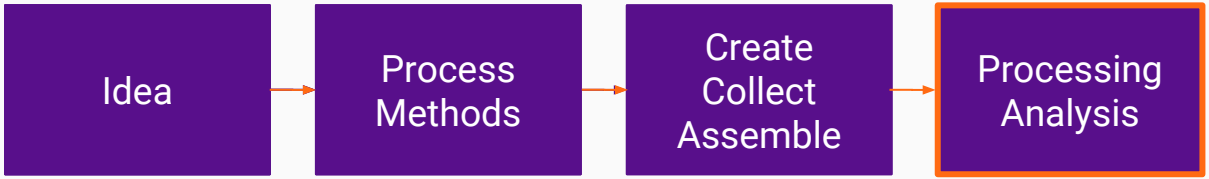


- ➔ Protocol (experimental workflow)
- ➔ Organism/Study population
- ➔ Equipment
- ➔ Software

Data Lifecycle: Gathering data

- ➔ Observations
- ➔ Lab experiments
- ➔ Simulations
- ➔ Aggregation/Extraction

Data Lifecycle: Processing & Analysis

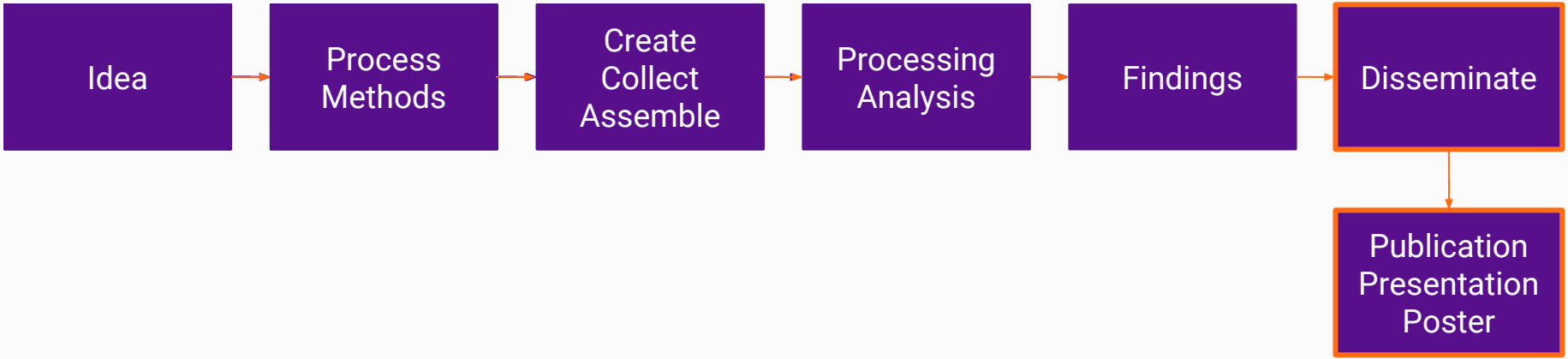


- ➔ Statistical analyses
- ➔ Visualizations
- ➔ Modeling
- ➔ Coding

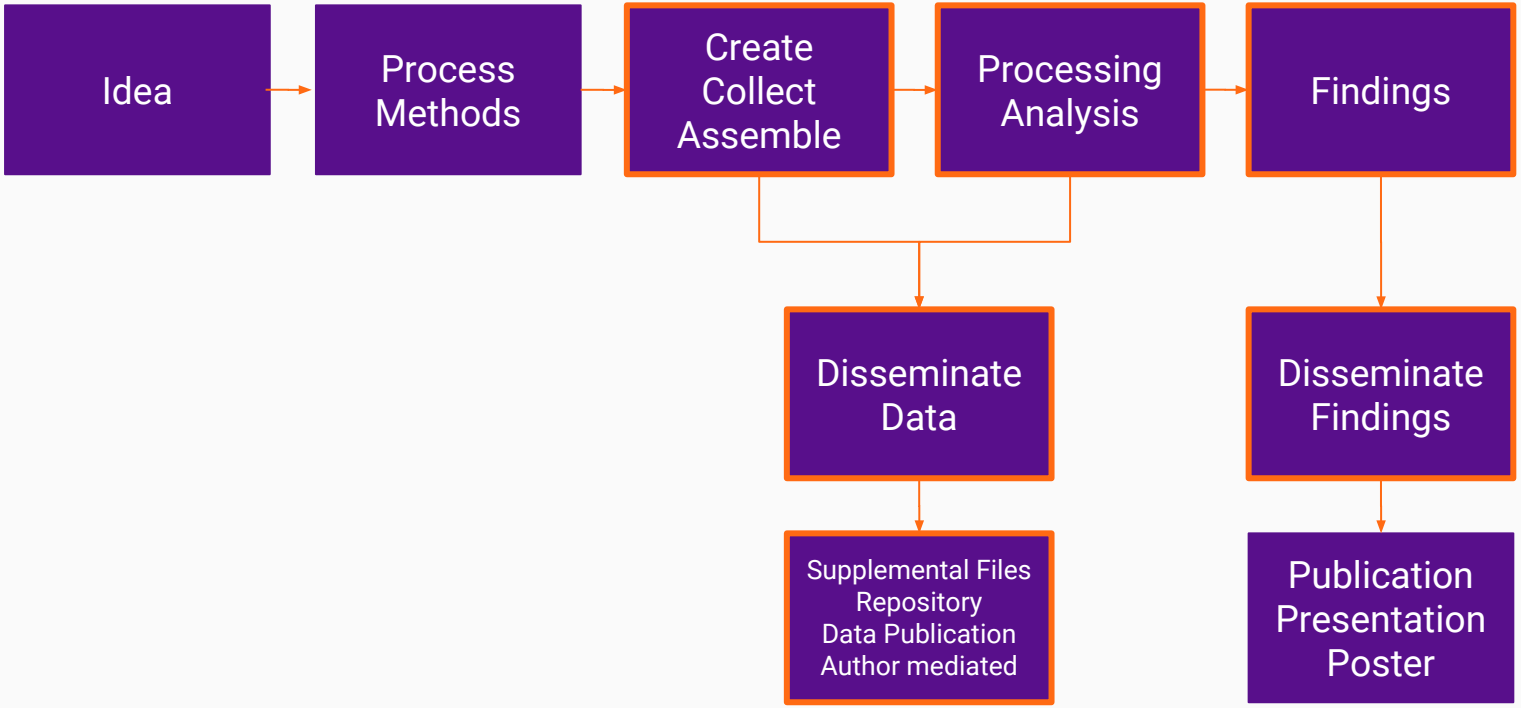
Data Lifecycle: Findings

- ➔ Results
- ➔ Conclusion
- ➔ Discussions

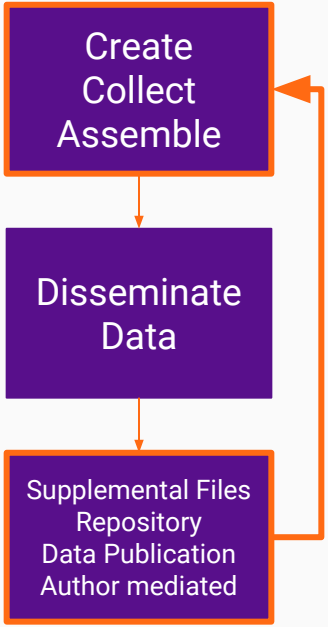
Data Lifecycle: Dissemination



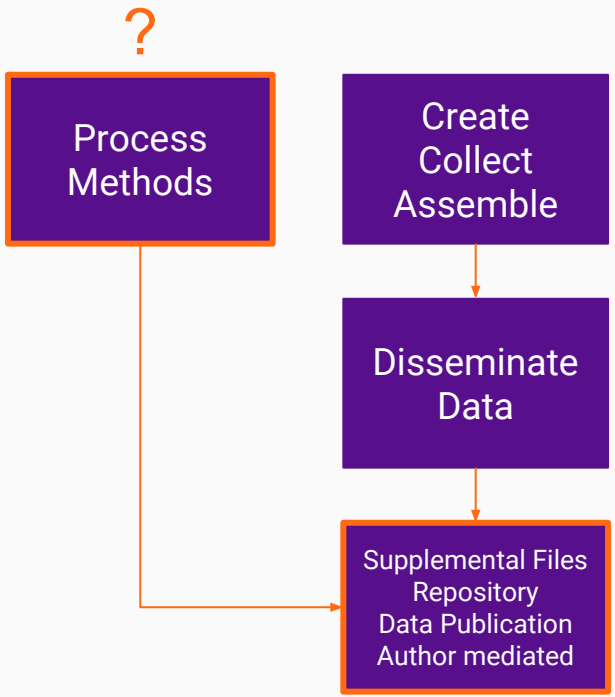
Data Lifecycle: Disseminating data



Data Lifecycle: Data Reuse



Data Lifecycle: Data reuse + reproducibility



Why would anyone need someone else's data?

Cow concept: Dorothea Salo, "Save the Cows", 2009.
http://www.slideshare.net/cavlec/save-the-cows-data-curation-for-the-rest-of-us-1533252

Why would anyone need someone else's data?

Analyze & Process

Cow concept: Dorothea Salo, "Save the Cows", 2009.
http://www.slideshare.net/caylec/save-the-cows-data-curation-for-the-rest-of-us-1533252

Why would anyone need someone else's data?

Analyze & Process

Publish

Cow concept: Dorothea Salo, "Save the Cows", 2009.
http://www.slideshare.net/caylec/save-the-cows-data-curation-for-the-rest-of-us-1533252

Why would anyone need someone else's data?

Analyze & Process

Publish

Cow concept: Dorothea Salo, "Save the Cows", 2009.
http://www.slideshare.net/caylec/save-the-cows-data-curation-for-the-rest-of-us-1533252

No need to “kill the cow”

?

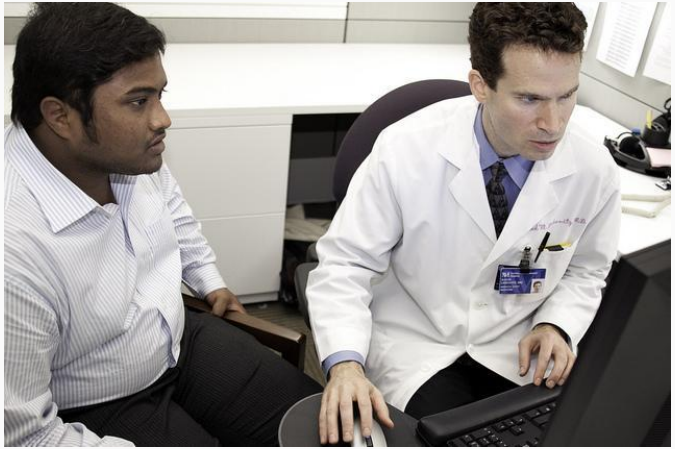
Questions

Course Schedule

1. Introduction
2. Current library roles in RDM
3. Story of data
4. **Understanding your research community**
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

Understand your researchers

Bench Science Researchers

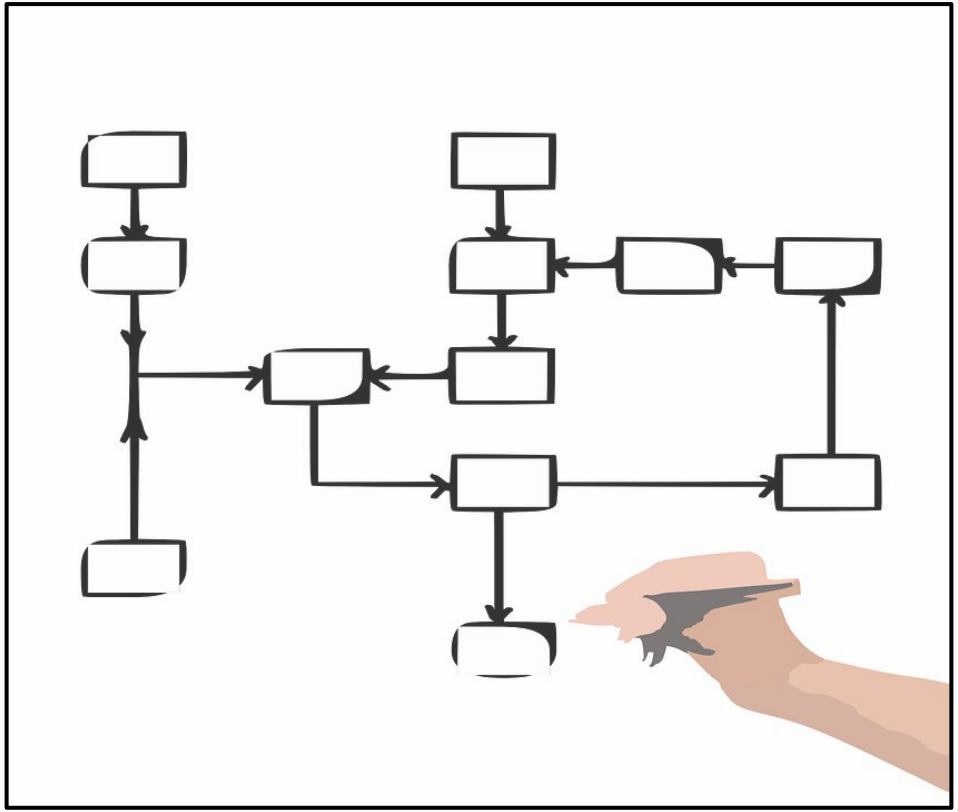


Clinical Researchers

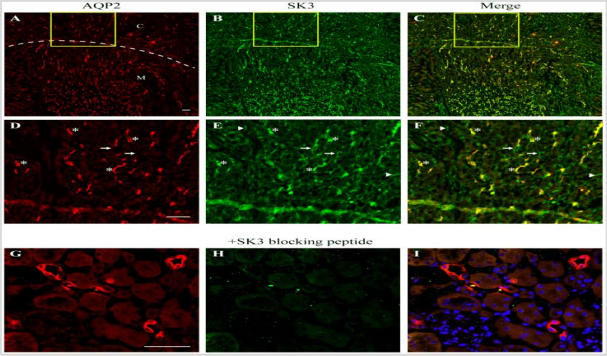
Bench Research: The lab is their home

Bench Research: Many researchers

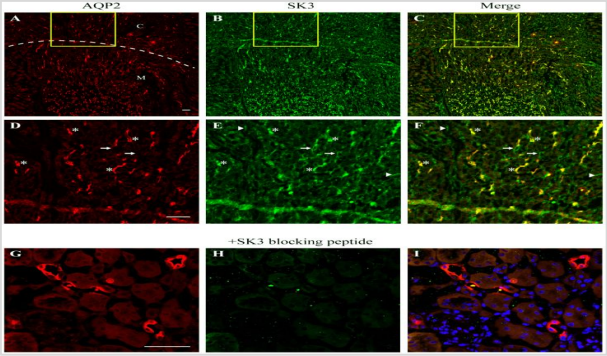
Bench Research: Different workflows

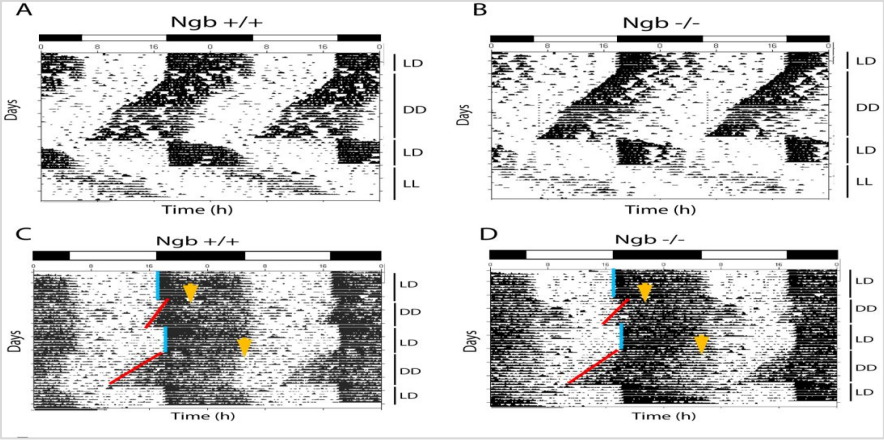


Bench Research: Wide variety of data

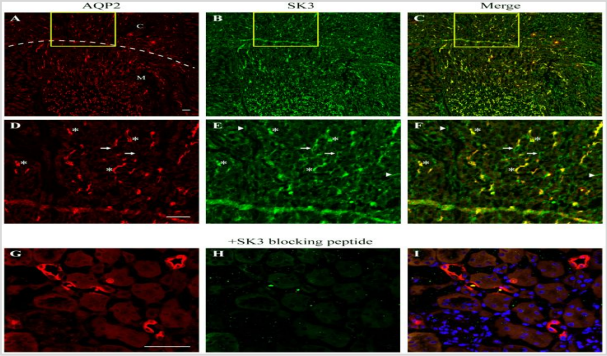


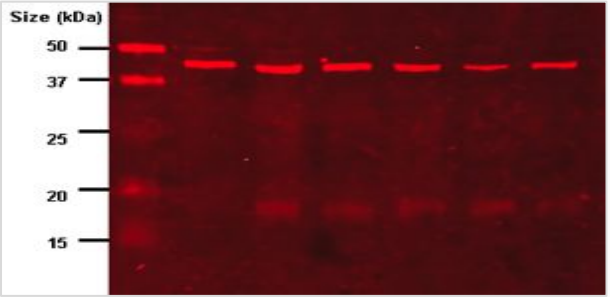
Bench Research: Wide variety of data

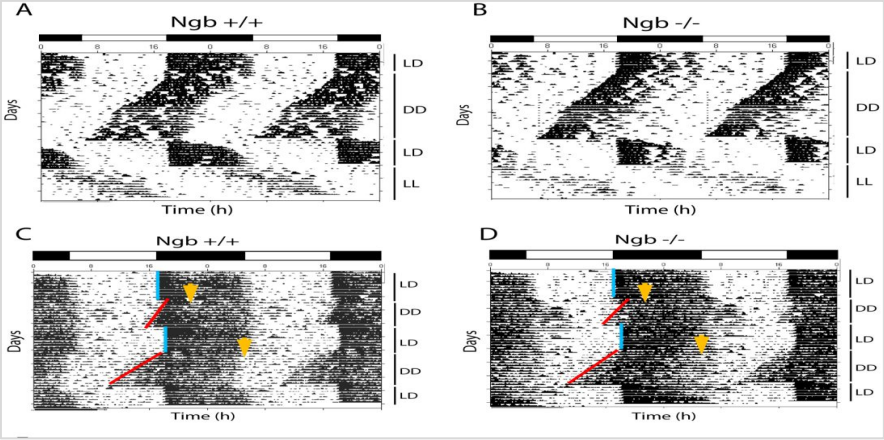




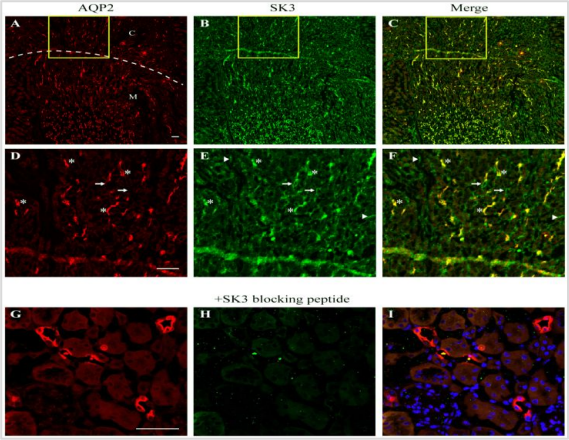
Bench Research: Wide variety of data

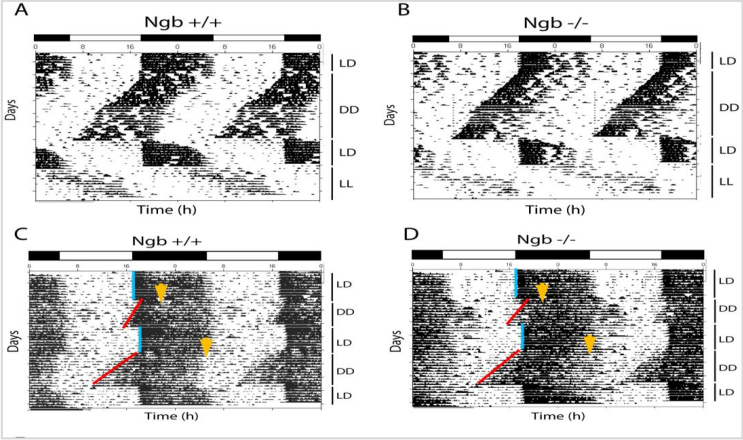


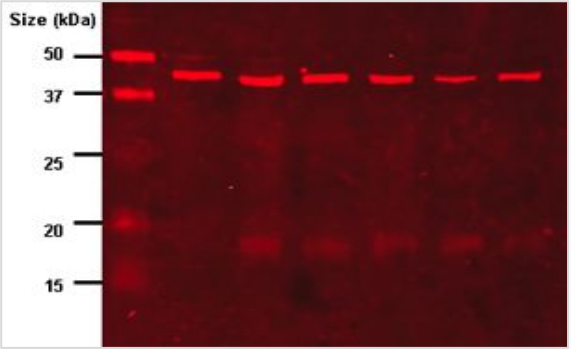


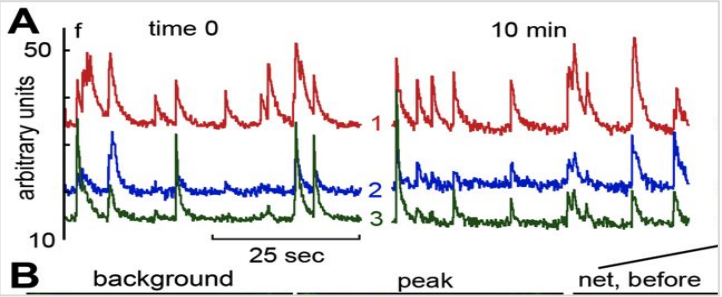


Bench Research: Wide variety of data

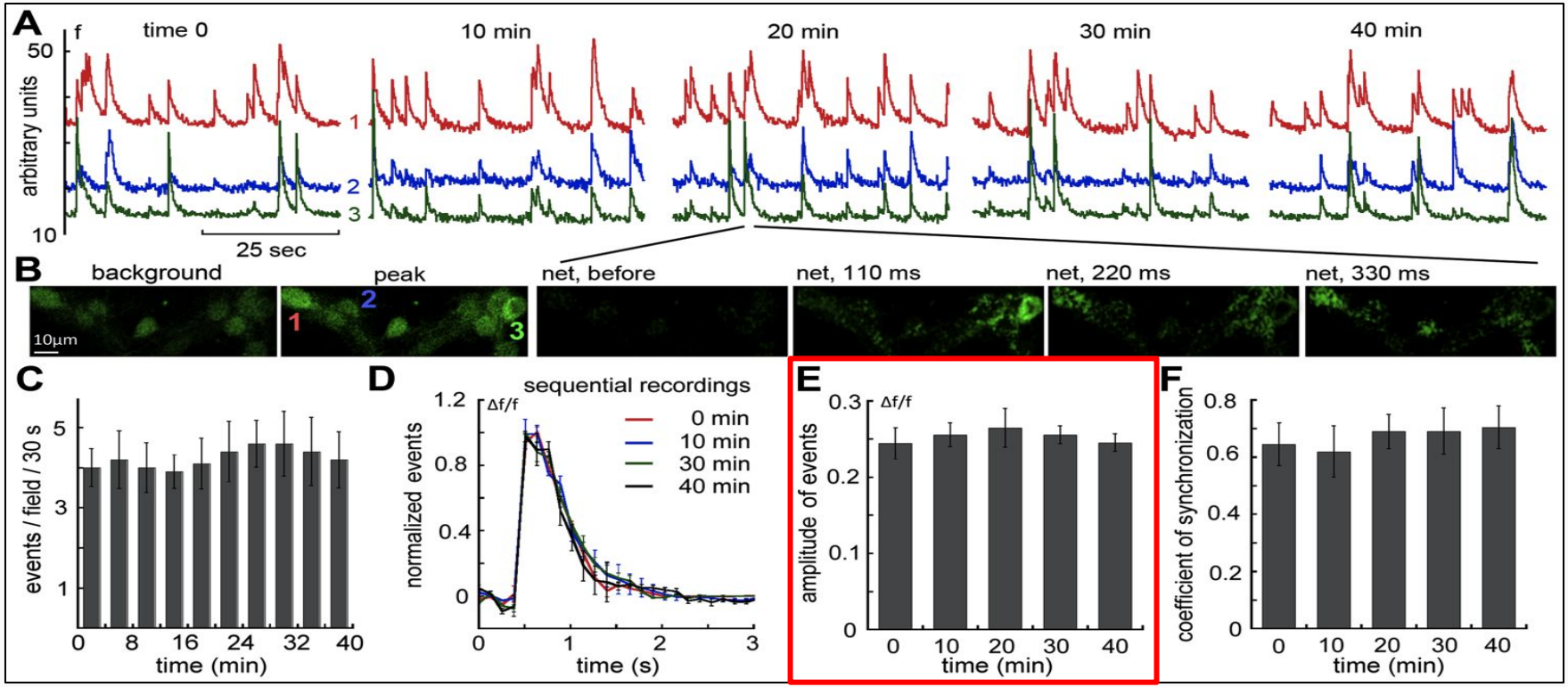




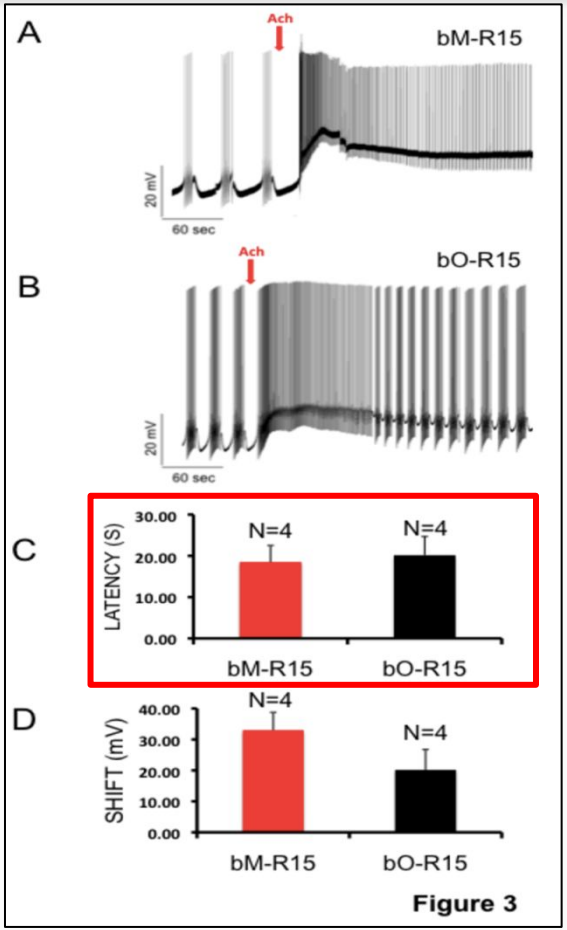




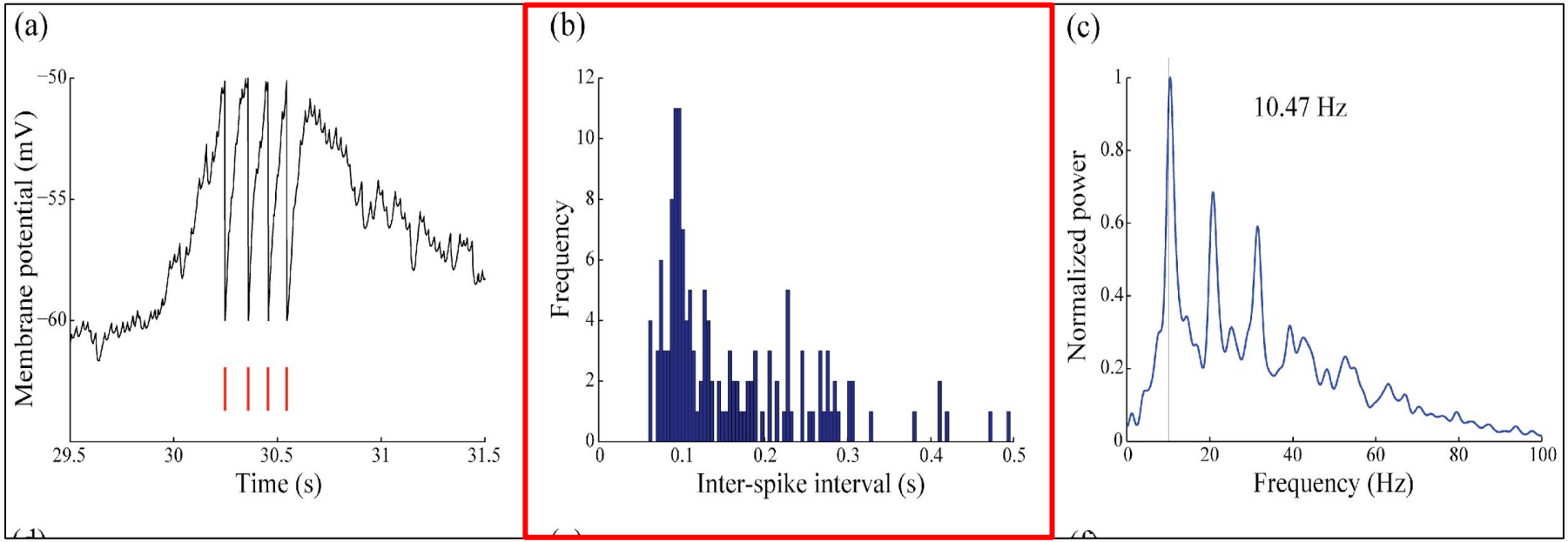
Bench Research: The same but different



Bench Research: The same but different



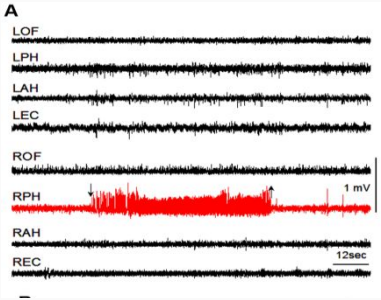
Bench Research: The same but different



Bench Research: Different but the same

NYU HEALTH SCIENCES LIBRARY

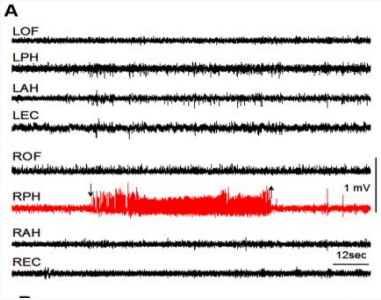
Systems level, functional genomics analysis of chronic epilepsy:

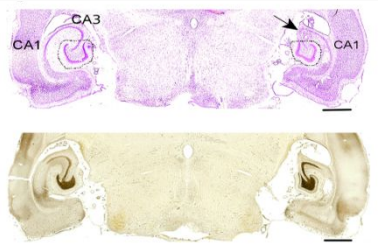


Bench Research: Different but the same

NYU HEALTH SCIENCES LIBRARY

Systems level, functional genomics analysis of chronic epilepsy:

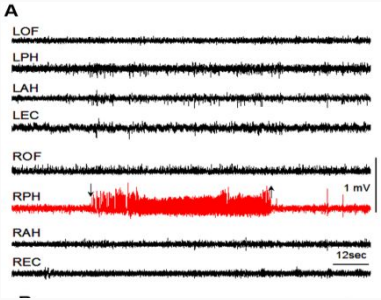


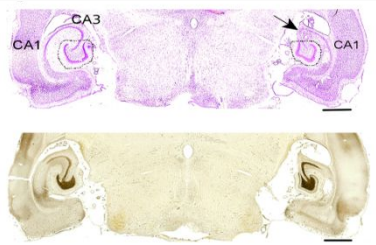


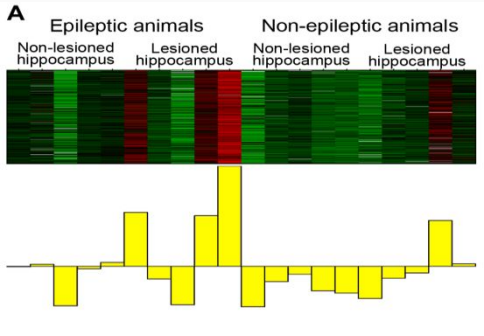
Bench Research: Different but the same

NYU HEALTH SCIENCES LIBRARY

Systems level, functional genomics analysis of chronic epilepsy:



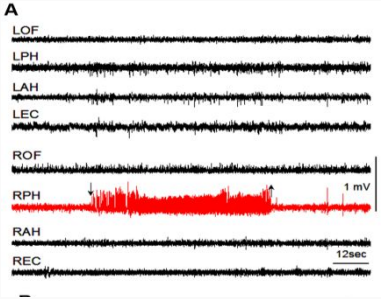


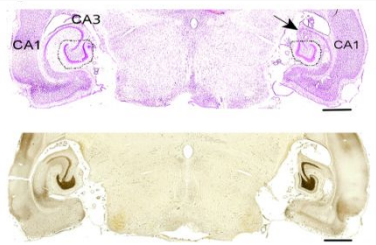


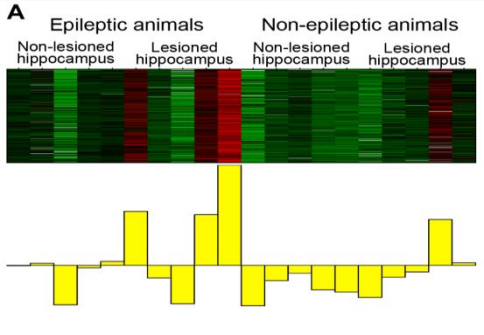
Bench Research: Different but the same

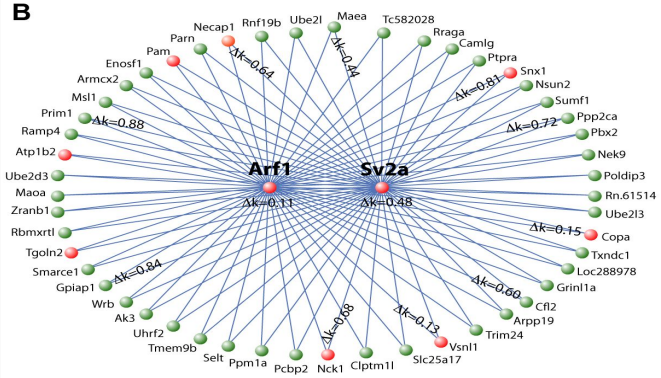
NYU HEALTH SCIENCES LIBRARY

Systems level, functional genomics analysis of chronic epilepsy:

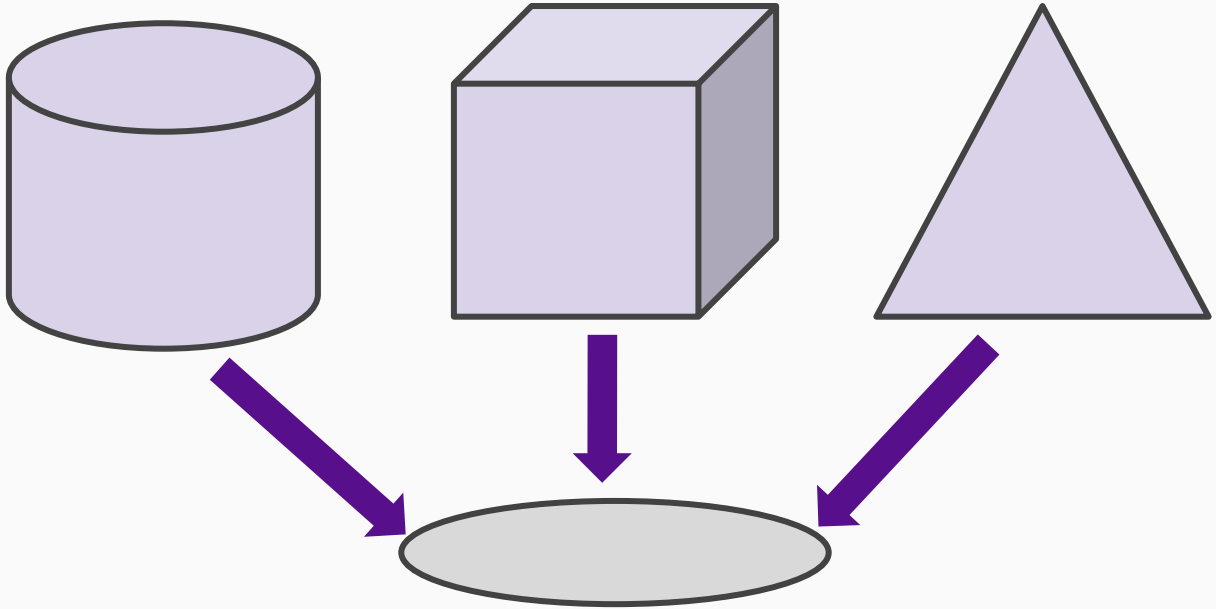






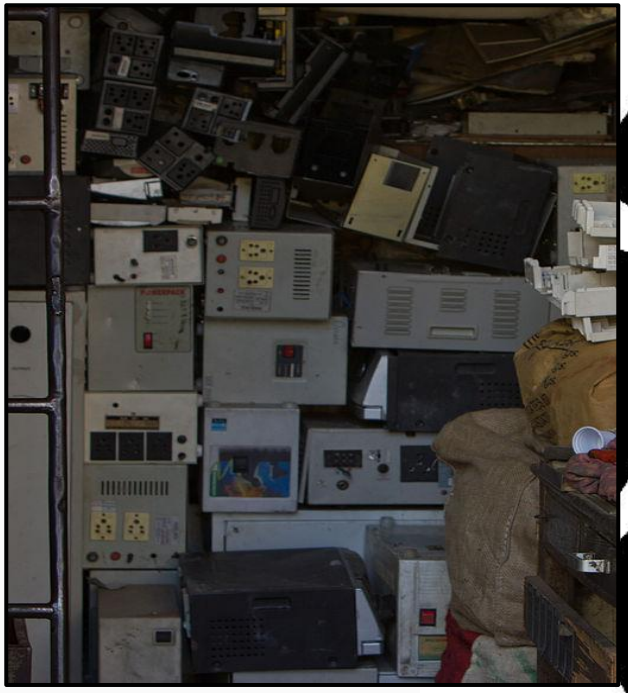


Bench Research: One size DOES NOT fit all

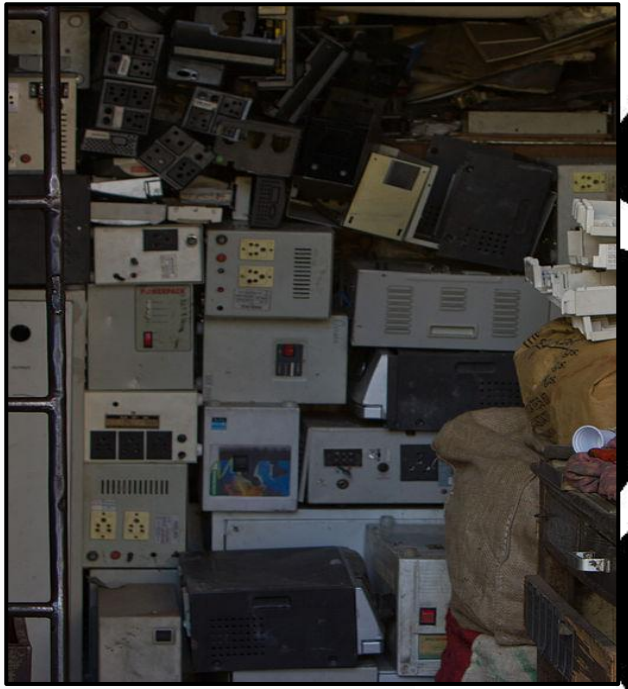


Bench Research: Lab turnover

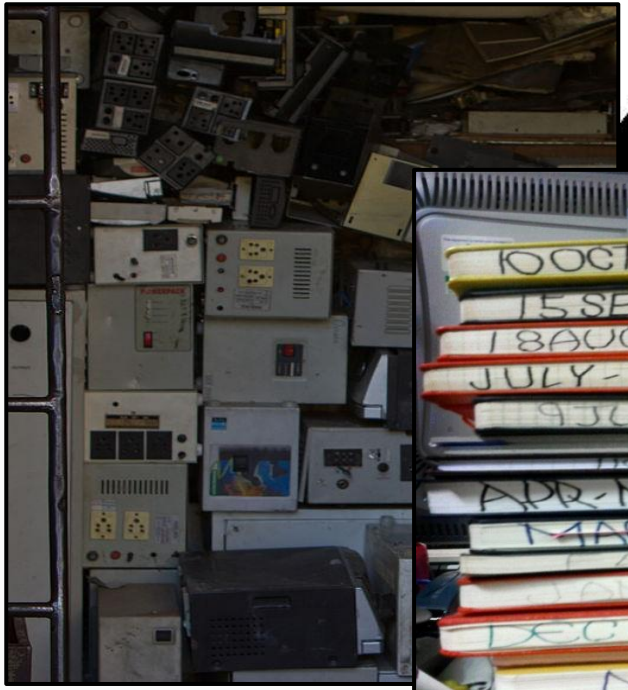
Bench Research: Lab turnover + many experiments

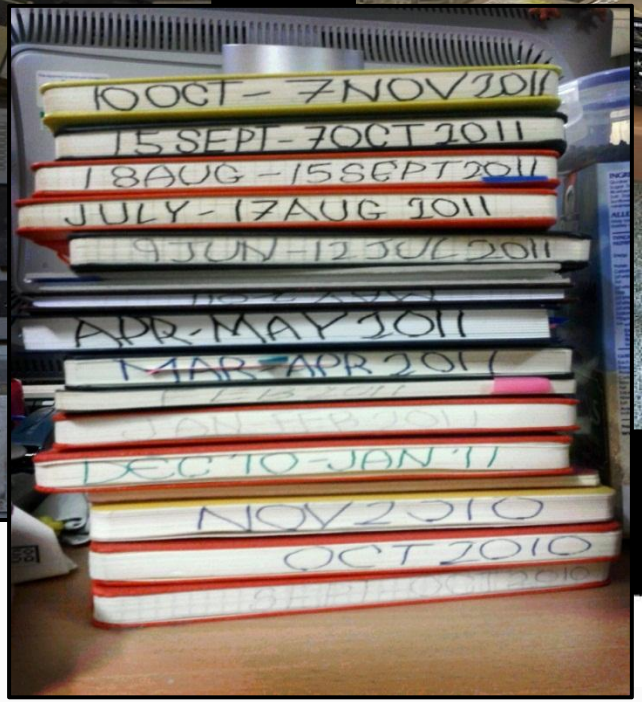


Bench Research: Lab turnover + many experiments



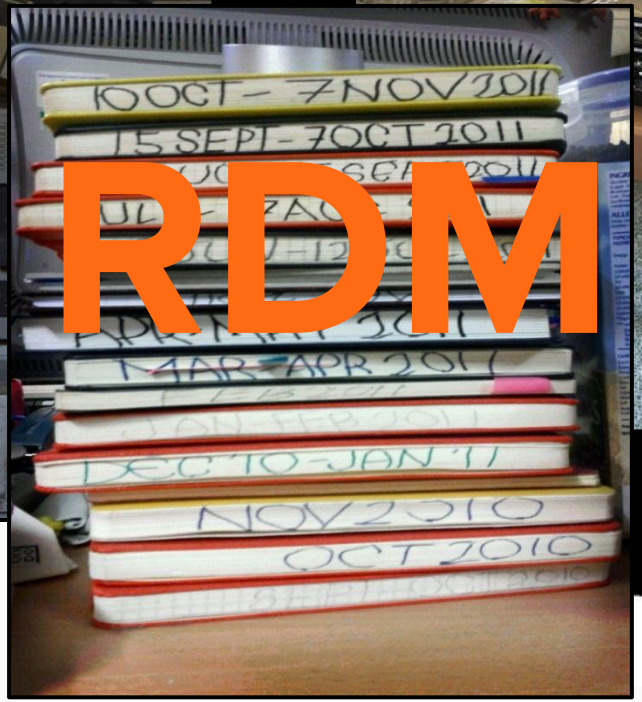
Bench Research: Lab turnover + many experiments





Bench Research: Lab turnover + many experiments

NYU HEALTH SCIENCES LIBRARY



An RDM Mess

But we're **NOT** scientists!

So you're not a scientist...

Work with one researcher at a time

DO YOUR HOMEWORK

- ➔ Review researcher backgrounds
- ➔ Read methods sections of papers
- ➔ Look at figures

Clinical Research:
Live in two worlds

- Primarily trained as clinicians, not researchers

Clinical Research:
Live in two worlds

- ➔ Primarily trained as clinicians, not researchers
- ➔ Many take on research with minimal training

Clinical Research:
Live in two worlds

- ➔ Primarily trained as clinicians, not researchers
- ➔ Many take on research with minimal training
- ➔ Split time between research and clinical responsibilities

Clinical Research: Project personnel

NYU HEALTH SCIENCES LIBRARY

PI

Research Coordinator

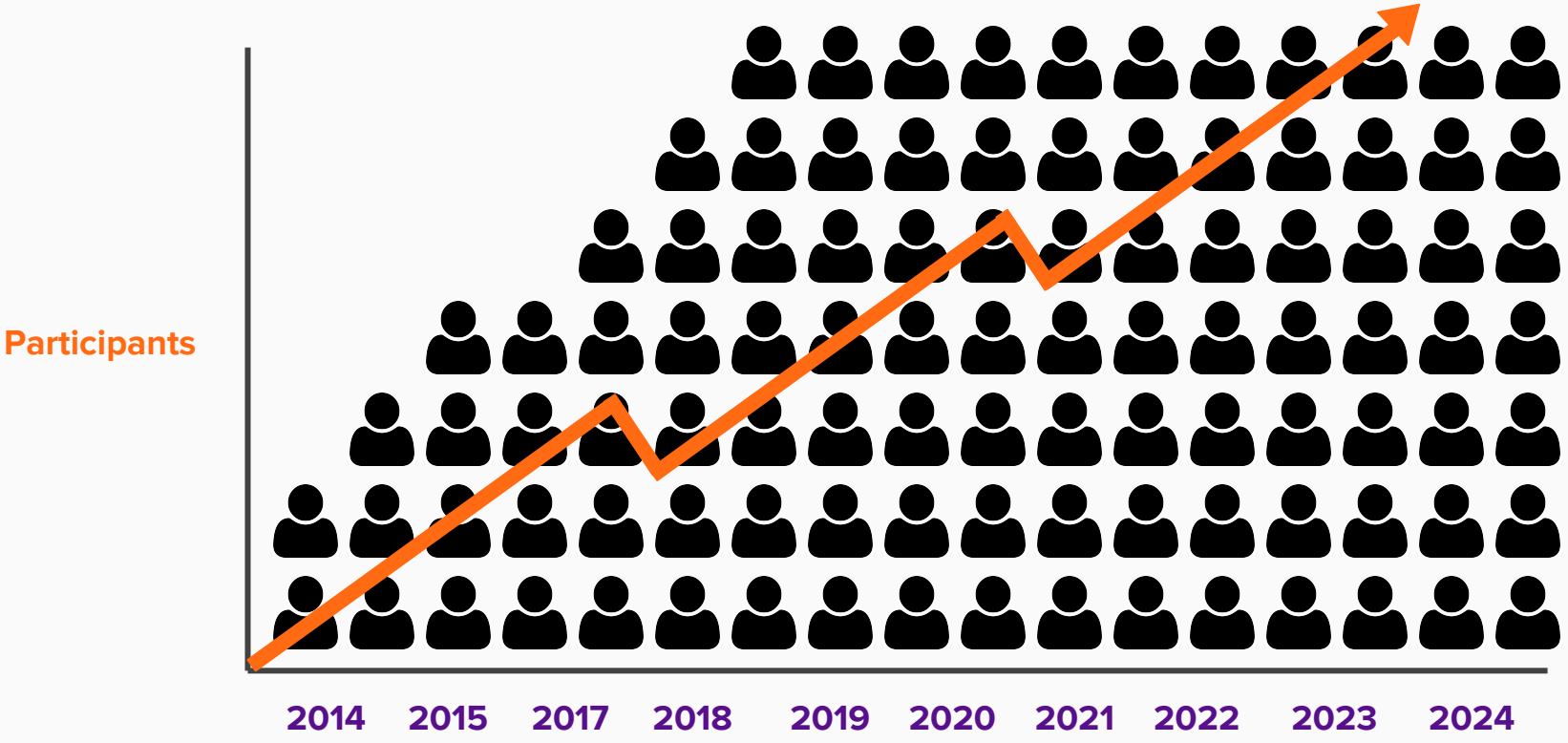
Data Manager

Statistician

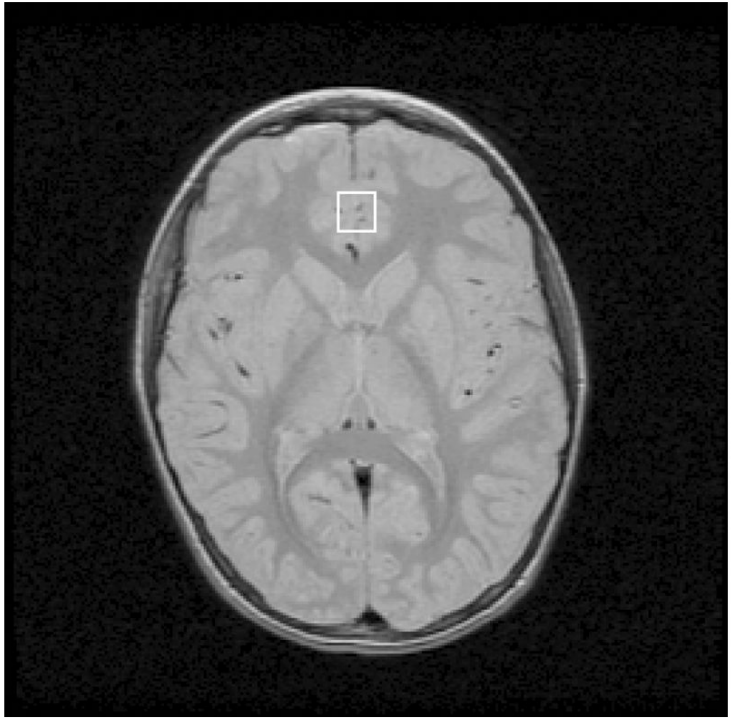
Research Assistant

Research Assistant

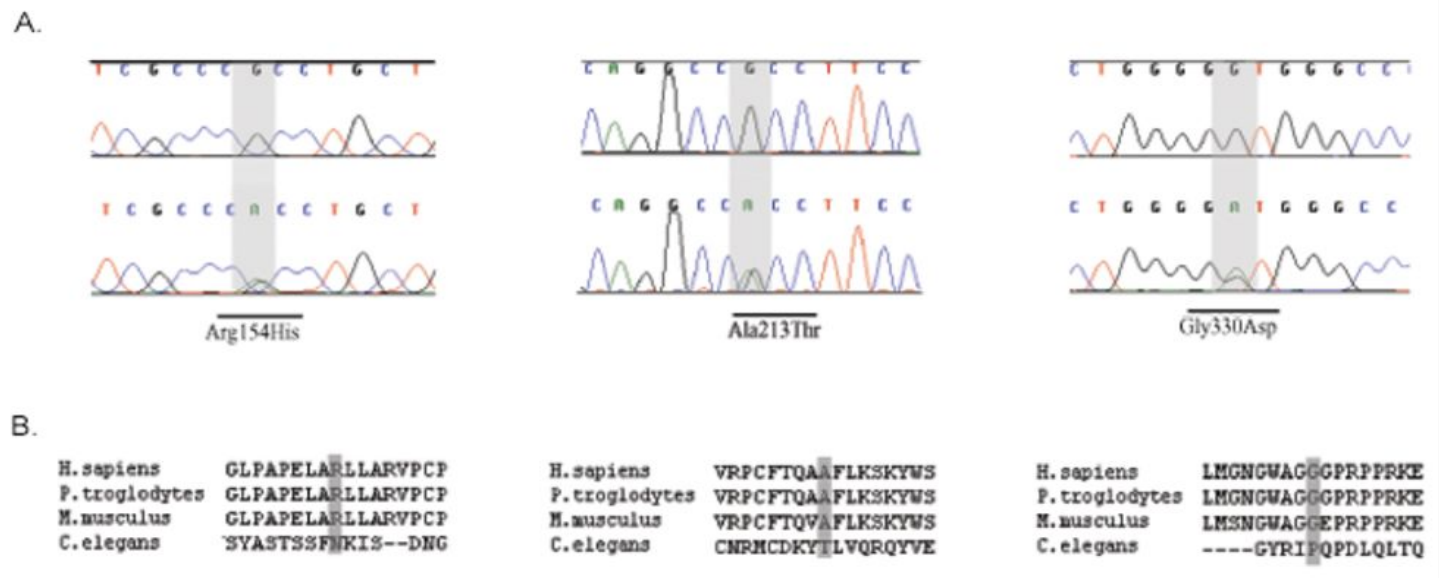
Clinical Research: Long studies, many participants



Clinical Research: Consistent data



Clinical Research: Consistent data



Clinical Research: Consistent data

NYU HEALTH SCIENCES LIBRARY

My Projects
Project Home
Project Setup
Project status: Production

Data Collection

Scheduling
Data Entry

Patient ID 02
Event: Preoperative evaluation (Arm 1: RV Pacing)
Data Collection Instruments:

- Baseline Data
- Preoperative Diagnostic Tests
- Echocardiogram before PM implantation
- Heart Failure Biomarkers
- SF-36 Questionnaire
- Minnesota Living With HF Questionnaire
- Aquarel Questionnaire

Lock all forms

Applications

Calendar
Data Export Tool
Data Import Tool
Data Comparison Tool
Logging
File Repository
User Rights
Record Locking Customization
E-signature and Locking Mgmt
Graphical Data View & Stats
Data Quality
API

Editing existing Patient ID 02

Event Name: Preoperative evaluation (Arm 1: RV Pacing)

Patient ID	02
Baseline Measurements	
Date of baseline evaluation	2012-06-25 (YYYY-MM-DD) Today Y-M-D
Functional Class (NYHA)	I
Clinical presentation	☐ Syncope ☐ Pre-syncope ☑ Dizziness ☐ Asymptomatic bradycardia ☐ Unstable bradycardia ☐ Dyspnea at rest ☐ Dyspnea during exercise ☐ Paroxysmal nocturnal dyspnea ☐ Angine ☐ Palpitations ☐ Fatigue ☐ No symptoms
Presentation to health care facility	Emergency admission for documented or suspected arrhythmia
Etiology	Degenerative
Underlying heart disease	None
Any comorbid condition?	◉ Yes ○ No

reset

r tachycardia)

urgically corrected)
ically corrected)

Basic User Rights

Calendar	☑
Data Export Tool	○ No Access ○ De-Identified ◉ Full Data Set
Data Import Tool	☑
Data Comparison Tool	☑
Logging	☑
File Repository	☑
User Rights	☑
Data Access Groups	☑
Graphical Data View & Stats	☑
Data Quality (What is Data Quality?)	☑ Create & edit rules ☑ Execute rules
Reports & Report Builder	☑
Project Design and Setup	☑
API (What is the REDCap API?)	☐ API Export ☐ API Import

Settings pertaining to record locking and E-signatures:

Record Locking Customization	☐
Lock/Unlock Records	◉ Disabled

Data Entry Rights

	No Access	Read Only	View & Edit
Subject Screening and Recruitment	○	○	◉
Demographics	○	○	◉
Baseline Data	○	○	◉
Preoperative Diagnostic Tests	○	○	◉
Echocardiogram	○	○	◉
Pacemaker Implantation	○	○	◉
Hospital Discharge	○	○	◉
Clinical Evaluation	○	○	◉
Pacemaker Interrogation and Programming	○	○	◉
Heart Failure Biomarkers	○	○	◉
Six Minute Walk Distance Test	○	○	◉
SF-36 Questionnaire	○	○	◉
Minnesota Living With HF Questionnaire	○	○	◉
Completion Data	○	○	◉
Adverse Events	○	○	◉

Clinical Research: Data collection forms

NYU HEALTH SCIENCES LIBRARY

Serum Prealbumin (mg/dL)	
Creatinine (mg/dL)	
Normalized Protein Catabolic Rate (g/kg/d)	
Cholesterol (mg/dL)	
Transferrin (mg/dL)	
Blood draw shift?	◯ 0. AM ◯ PM reset
Blood draw by	◯ RN ◯ LPN ◯ nurse assistant ◯ doctor reset
Level of patient anxiety	
Patient scheduled for future draws?	

Clinical Research: Data collection forms

NYU HEALTH SCIENCES LIBRARY

Serum Prealbumin (mg/dL)	
Creatinine (mg/dL)	
Normalized Protein Catabolic Rate (g/kg/d)	
Cholesterol (mg/dL)	
Transferrin (mg/dL)	
Blood draw shift?	◯ 0. AM ◯ PM reset
Blood draw by	◯ RN ◯ LPN ◯ nurse assistant ◯ doctor reset
Level of patient anxiety	
Patient scheduled for future draws?	

Total Cholesterol = LDL + HDL Cholesterol + Triglycerides/5

Clinical Research: Data quality

NYU HEALTH SCIENCES LIBRARY

PatientID	Date of Birth	Weight	Smoker
1001	1983-01-09	180	Y
1002	1974-04-10	55	
1003	1991-05-02	135	2 packs/day
1005	1972-04-24	80	No

Clinical Research: Regulations

NYU HEALTH SCIENCES LIBRARY

De-Identification Options (optional)

The options below allow you to limit the amount of sensitive information that you are exporting out of the project. Check all that apply.

Known Identifiers:

- ☐ Remove all known Identifier fields *(tagged in Data Dictionary)*
- ☐ Hash the Study ID *(converts record name to an unrecognizable value)*

Free-form text:

- ☐ Remove unvalidated Text fields *(i.e. Text fields other than dates, numbers, etc.)*
- ☐ Remove Notes/Essay box fields

Date and datetime fields:

- ☐ Remove all date and datetime fields

— OR —

- ☐ Shift all dates by value between 0 and 364 days *(shifted amount determined by algorithm for each record)*

  What is date shifting?

Deselect all options

Clinical Research: Audit trails

NYU HEALTH SCIENCES LIBRARY

HIPAA Compliant

Time / Date	Username	Action	List of Data Changes OR Fields Exported
03/24/2016 1:13pm	surkia01	Manage/Design	Download data dictionary
10/29/2015 4:00pm	lapolf01	Manage/Design	Download data dictionary
10/26/2015 10:58am	lapolf01	Manage/Design	Download data dictionary
08/06/2015 9:42am	readk01	Manage/Design	Download data dictionary
08/06/2015 9:33am	lapolf01	Manage/Design	Delete project field
08/04/2015 3:24pm	readk01	Manage/Design	Download data dictionary
08/04/2015 3:22pm	readk01	Manage/Design	Download data import template
08/04/2015 10:43am	lapolf01	Manage/Design	Edit project field
08/04/2015 10:43am	lapolf01	Manage/Design	Edit project field
08/04/2015 10:42am	lapolf01	Manage/Design	Add/edit branching logic
08/04/2015 10:42am	lapolf01	Manage/Design	Reorder project fields
08/04/2015 10:42am	lapolf01	Manage/Design	Edit project field
08/04/2015 10:42am	lapolf01	Manage/Design	Edit project field
08/04/2015 10:35am	lapolf01	Manage/Design	Add/edit branching logic
08/04/2015 10:34am	lapolf01	Created Record 001	lastname = 'Smith', firstname = 'John', address = '123 Whatever Boulevard, Whereversburg, MA, 12345', phonenumber = '(212) 263-8535', email = 'fred.lapolla@med.nyu.edu', age = '29',

Clinical Research: Project personnel

NYU HEALTH SCIENCES LIBRARY

PI

Research Coordinator

Data Manager

Statistician

Research Assistant

Research Assistant

Data Management Nightmare

https://www.youtube.com/watch?v=nNBiCcBIwRA

Different languages

- ➔ Data capture vs data collection
- ➔ Data quality (clinical term)

Tower of Babel

NYU HEALTH SCIENCES LIBRARY

Clinical & Bench

Common issues

- → Documentation
- → Workflow

EXERCISE 1

The research paper

Published in final edited form as:
J Pediatr. 2009 January ; 154(1): 10–16. doi:10.1016/j.jpeds.2008.07.048.

Maternal smoking during pregnancy and newborn neurobehavior: A pilot study of effects at 10–27 days

Laura R. Stroud, Ph.D.[1], **Rachel L. Paster, B.A.**[1], **George D. Papandonatos**[2], **Raymond Niaura, Ph.D.**[1], **Amy L. Salisbury, Ph.D.**[3], **Cynthia Battle, Ph.D.**[1], **Linda L. Lagasse, Ph.D.**[3], and **Barry Lester, Ph.D**[3]

[1] Department of Psychiatry and Human Behavior, Warren Alpert Medical School, Brown University

[2] Center for Statistical Sciences, Brown University

[3] Brown Center for the Study of Children at Risk, Warren Alpert Medical School and Women and Infants' Hospital

Abstract

Objective—To examine effects of maternal smoking during pregnancy on newborn neurobehavior at 10–27 days.

Study design—Participants were 56 healthy infants (28 smoking-exposed, 28 unexposed) matched on maternal social class, age, and alcohol use. Maternal smoking during pregnancy was determined by maternal interview and maternal saliva cotinine. Postnatal smoke exposure was quantified by infant saliva cotinine. Infant neurobehavior was assessed through the NICU Network Neurobehavioral Scale.

Results—Smoking-exposed infants showed greater need for handling and worse self-regulation ($p < .05$) and trended toward greater excitability and arousal ($p < .10$) relative to matched, unexposed infants (all moderate effect sizes). In contrast to prior studies of days 0–5, no effects of smoking-exposure on signs of stress/abstinence or muscle tone emerged. In stratified, adjusted analyses, only effects on need for handling remained significant ($p < .05$, large effect size).

Conclusions—Effects of maternal smoking during pregnancy at 10–27 days are subtle and consistent with increased need for external intervention and poorer self-regulation. Along with parenting deficits, these effects may represent early precursors for long-term adverse outcomes from maternal smoking during pregnancy. That signs of abstinence shown in prior studies of 0–5 day-old newborns did not emerge in older newborns provides further evidence for the possibility of a withdrawal process in exposed infants.

The research paper: Tasks

What are/is the **funding source** for this research paper?

What **types of data** are being collected?

What **method(s)** did the research use to collect the data?

What **data collection methods** can you identify?

The research paper: Collection methods

a) Questionnaires

- TLFB
- NNNS
- Medical history questionnaire
- Hollingshead
- CES-D

b) Bioassay

- Salivary continine
- Infant saliva sample

c) Observation

d) Interviews

What **data types** can you identify?

The research paper: Types of data

Physiological/Clinical measures

Bioassay, saliva samples, interviews, medical history questionnaire

Behavioral

NNNS, Interview, Questionnaire, TFLB

Neurological

NNNS, Observation

Socioeconomic

Hollingshead

Psychological

CES-D

Why is this important?

Data Interviews

Data Interviews: How to reach out

1. Review literature to inform questions
2. Identify researchers with active grant funding
3. Separate researchers into bench and clinical

Data Interviews: Why they are useful

Opportunity to meet with researchers

Learn about researchers' data practices

Establish relationships with research community

Identify data service gaps

Associates the library with data

Data Interviews: Quick tips

Do your homework (review papers, researcher background)

Make meetings **about the researcher**, not about the library

Citation:

Read KB, Surkis A, Larson C, et al. Starting the data conversation: informing data services at an academic health sciences library. *Journal of the Medical Library Association : JMLA*. 2015;103(3):131-135. doi:10.3163/1536-5050.103.3.005.

?

Questions

Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. **RDM climate**
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

NYU HEALTH SCIENCES LIBRARY

RDM
Carrots

RDM
Sticks

RDM

Carrots

Better Science

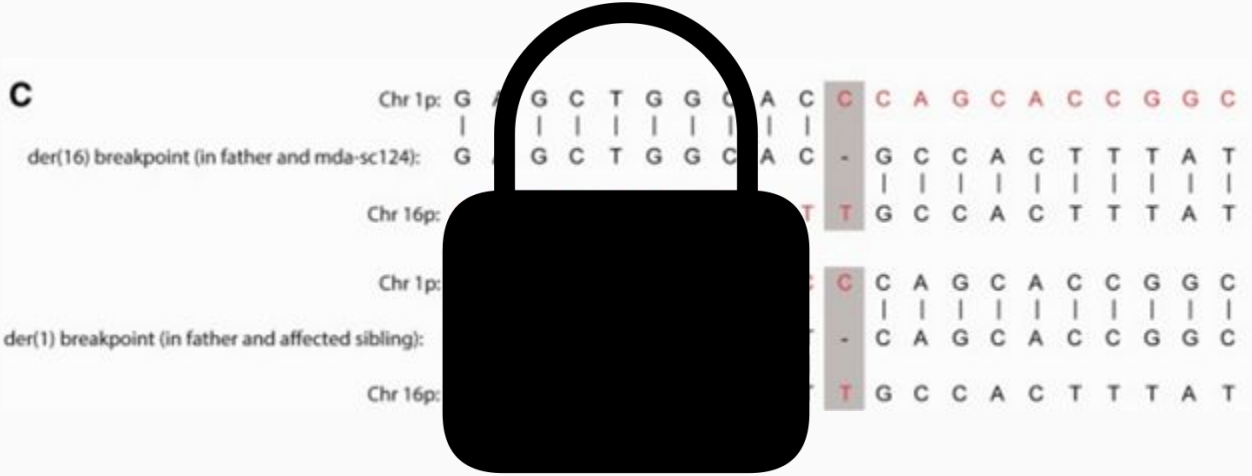
Sharing data across research communities:

- ➔ Human Genome Project
- ➔ Alzheimer's Disease Neuroimaging Initiative

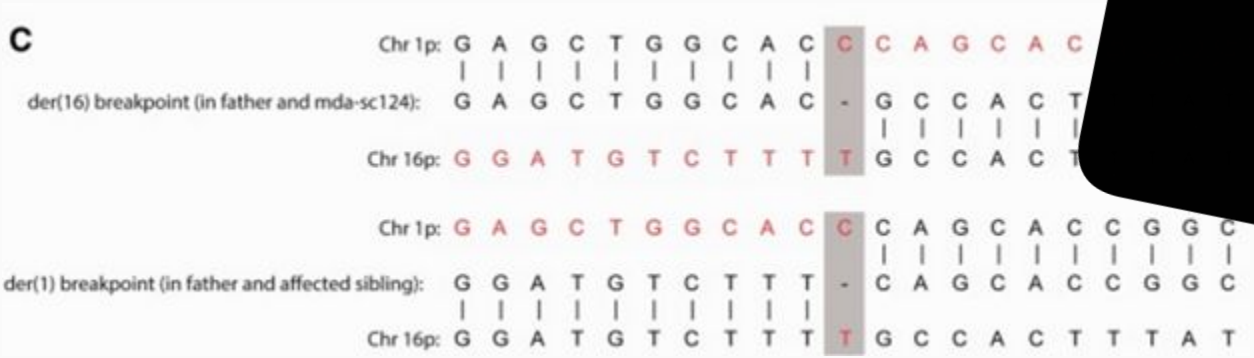
Collaborating to address new diseases (e.g. Zika, Ebola)

Human Genome Project

Human Genome Project: Competition



Human Genome Project: Competition



2 years ahead of schedule

NYU HEALTH SCIENCES LIBRARY

RDM Carrots

Getting credit

RDM carrots: Data repositories

Share with research communities

Receive **proof** for credit (DOI)

NIH Data Sharing Repositories

This table lists NIH-supported data repositories that accept submissions of appropriate data from NIH-funded investigators (and others). Also included are resources that aggregate information about biomedical data and information sharing systems. The table can be sorted according by name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites of the individual repositories.

Show 50 entries Search:

IC	Repository Name	Repository Description	Data Submission Policy	Access to Data
NCI	The Cancer Imaging Archive (TCIA)	The Cancer Imaging Archive (TCIA) is a large archive of medical images of cancer accessible for public download. All images are stored in DICOM file format. The images are organized as "Collections", typically patients related by a common disease (e.g. lung cancer), image modality (MRI, CT, etc) or research focus.	How to Submit Data to TCIA	How to Access TCIA Data
NCI (NHGRI, NIGMS)	PeptideAtlas	PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences.	How to Submit Data to PeptideAtlas	How to Access PeptideAtlas Data
NHGRI	FlyBase: A Drosophila Genomic and Genetic Database	Drosophila Genomic and Genetic database that includes proteomics data, microarrays and Tiling BAC's.	How to Submit Data to FlyBase	How to Access FlyBase Data
NHGRI	The Zebrafish Model Organism Database (ZFIN)	ZFIN serves as the zebrafish model organism database. It aims to: a) be the community database resource for the laboratory use of zebrafish, b) develop and support integrated zebrafish genetic, genomic and developmental information, c) maintain the definitive reference data sets of zebrafish research information, d) to link this information extensively to corresponding data in other model organism and human databases, e) facilitate the use of zebrafish as a model for human biology, and f) serve the needs of the research community.	How to Submit Data to ZFIN	How to Access ZFIN Data
NHGRI	WormBase	WormBase is an international consortium of biologists and computer scientists dedicated to providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of C. elegans and related nematodes.	How to Submit Data to WormBase	How to Access WormBase Data

RDM carrots: Data journals

SCIENTIFIC DATA

Home | About | For Authors | For Referees | Advisory and Editorial Board | Open Access | FAQ

Sign up for Scientific Date e-alert Facebook Twitter

Submit to *Scientific Data* in three simple steps:

1. DESCRIBE
Write a detailed description of your dataset. We have templates to help you and a detailed guide to authors.

2. DEPOSIT
See our list of recommended repositories. We will help you find the right place for your data.

3. SUBMIT
Submit online and get the credit you deserve for your data!

Get credit where credit's due and share your data.

Sample Data Descriptors

Proteomic profiles of human embryonic stem cells, induced-pluripotent stem cells and

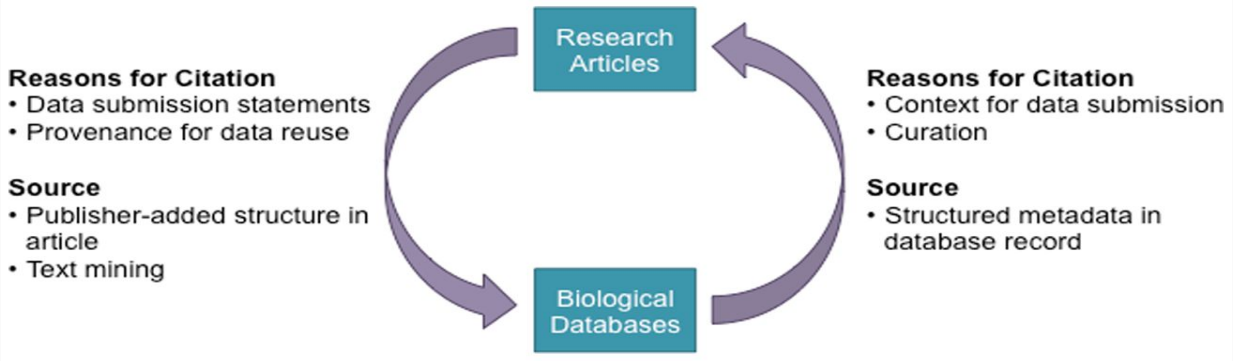
Sequencing of genomes, transcriptomes and methylomes of wild *Arabidopsis thaliana*

RDM carrots: Data citation

Share data = **more citations** (Piwowar, 2013)

Cite data within publications and visa versa

Get a **unique ID**



RDM carrots: NIH Biosketch

C. Contribution to Science [Edit section]

Description

Provided a methodology for librarians to conduct interviews with basic science and clinical researchers to learn about their data management challenges, needs and workflows in the context of an academic medical center.

Citations

a. Read KB, Surkis A, Larson C, McCrillis A, Graff A, Nicholson J, Xu J. Starting the data conversation: informing data services at an academic health sciences library. J Med Libr Assoc. 2015 Jul;103(3):131-5. PubMed PMID: 26213504; PubMed Central PMCID: PMC4511052.

Description

Conducted a study to develop a better understanding of the research datasets that are created as a part of NIH-funded research but not currently documented or deposited in a known repository. This work served to inform the initial stages of development for a NIH Data Discovery Index designed to describe NIH-funded research data.

Citations

a. Read KB, Sheehan JR, Huerta MF, Knecht LS, Mork JG, Humphreys BL. Sizing the Problem of Improving Discovery and Access to NIH-Funded Data: A Preliminary Study. PLoS One. 2015;10(7):e0132735. PubMed PMID: 26207759; PubMed Central PMCID: PMC4514623.

RDM sticks

NIH Data Sharing Policy
Research >$500,000
No enforcement

2003

Timeline of funder data regulations

NIH Data Sharing Policy
Research >$500,000
No enforcement

NIH Public Access Policy
No enforcement

2003

2008

Timeline of funder data regulations

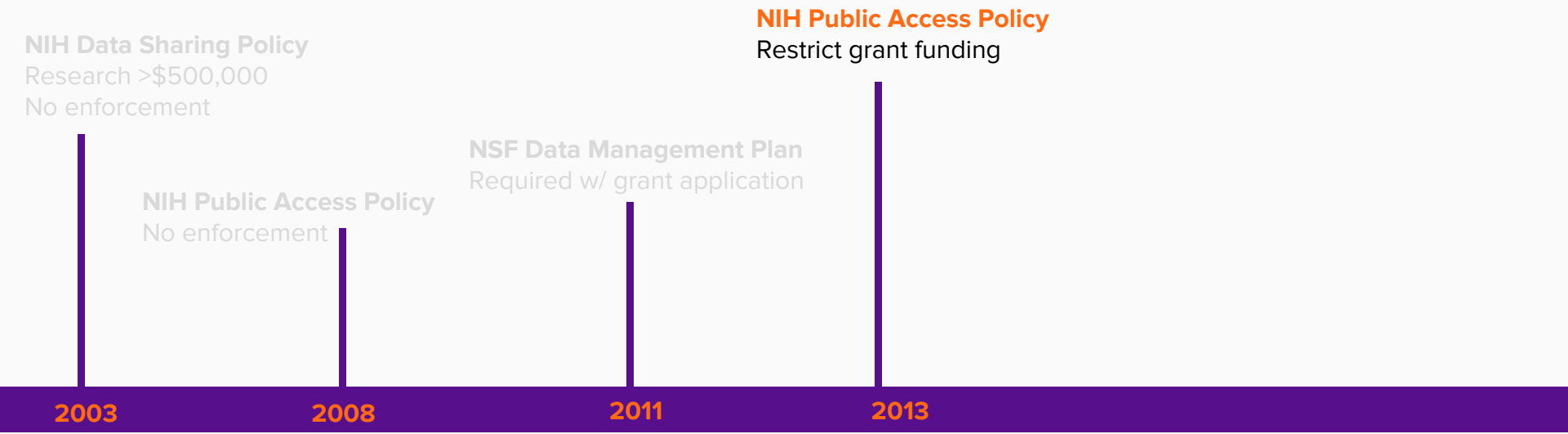
NIH Data Sharing Policy
Research >$500,000
No enforcement

NIH Public Access Policy
No enforcement

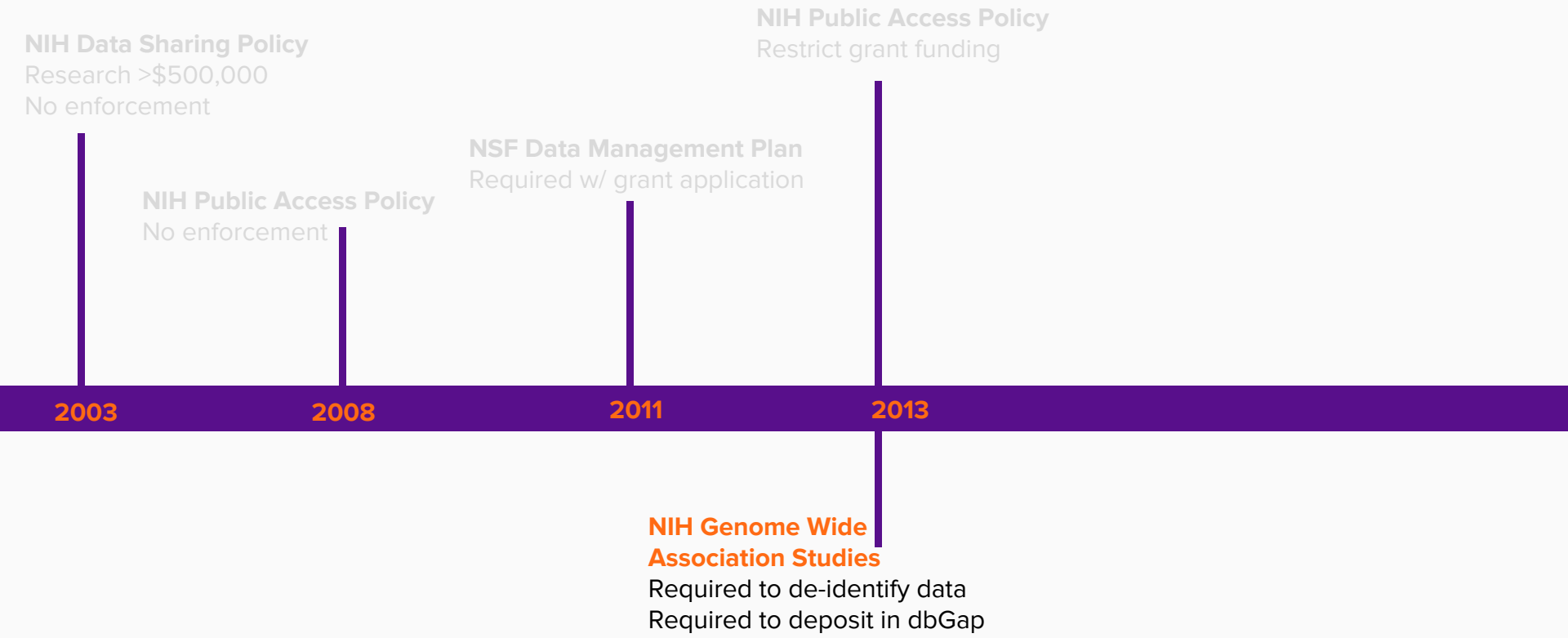
NSF Data Management Plan
Required w/ grant application

2003 **2008** **2011**

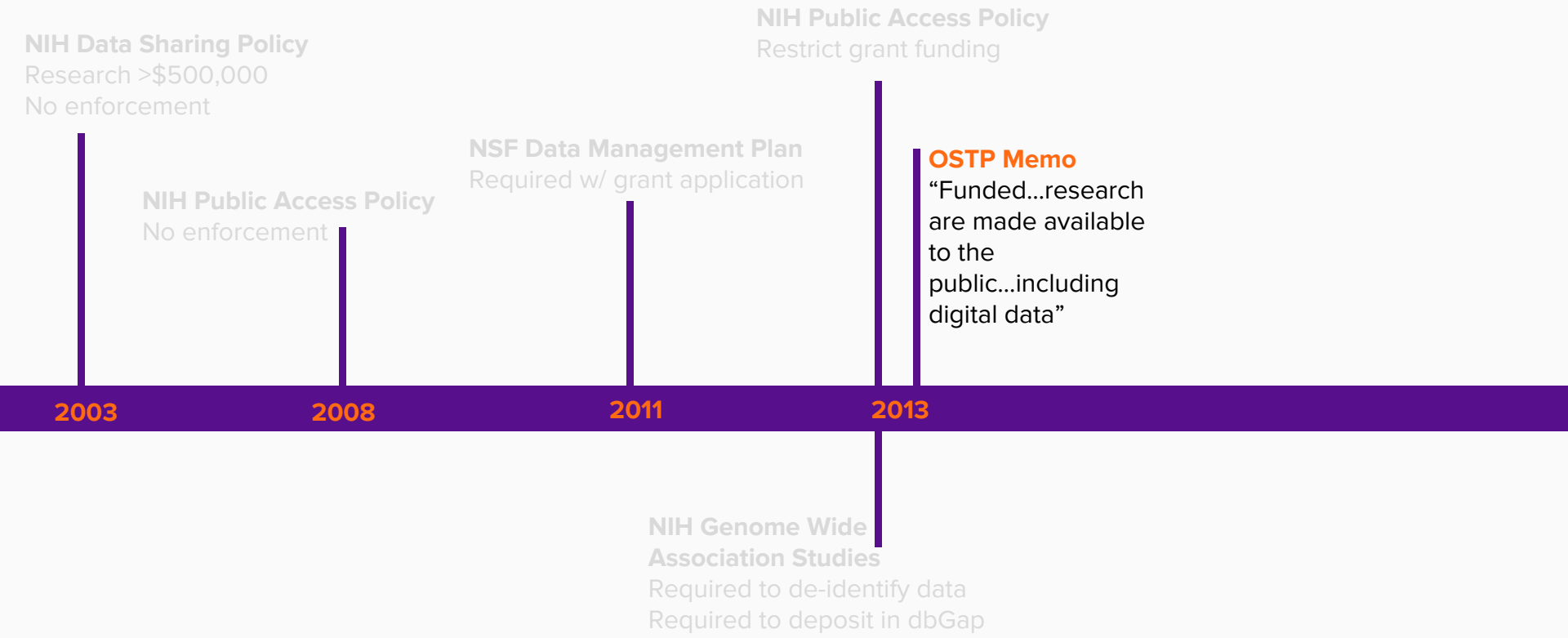
Timeline of funder data regulations



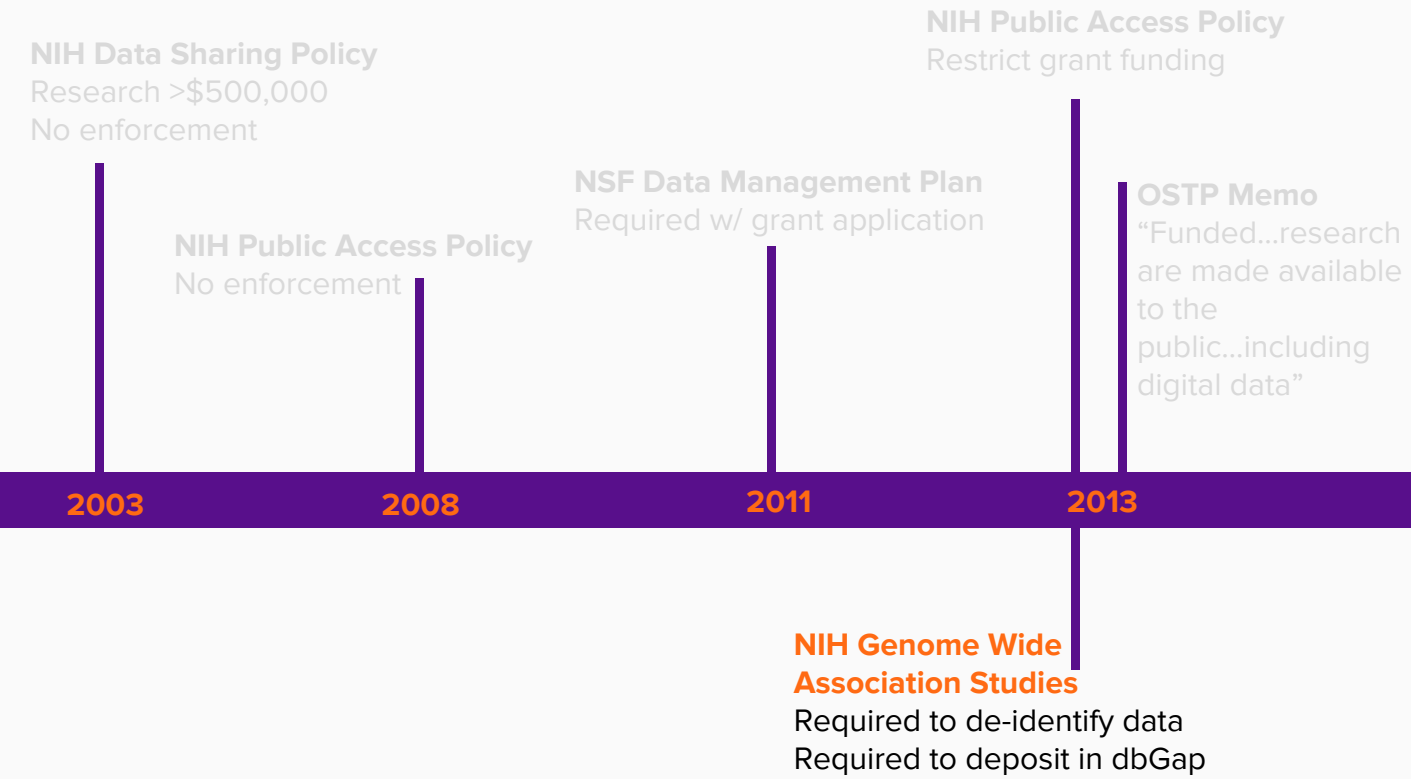
Timeline of funder data regulations



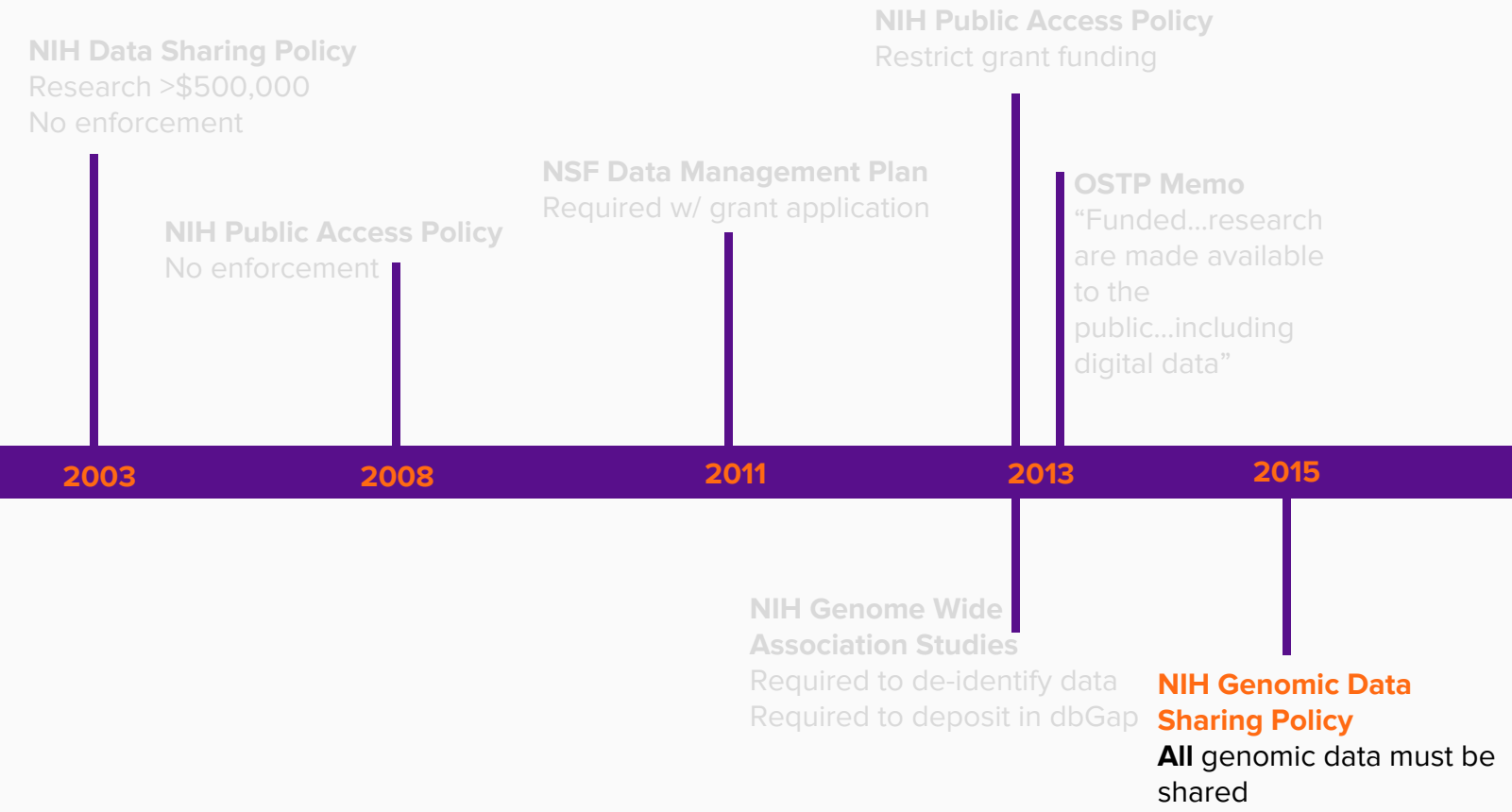
Timeline of funder data regulations



Timeline of funder data regulations



Timeline of funder data regulations



Timeline of funder data regulations

NIH Data Sharing Policy
Research >$500,000
No enforcement

NIH Public Access Policy
No enforcement

NSF Data Management Plan
Required w/ grant application

NIH Public Access Policy
Restrict grant funding

OSTP Memo
"Funded...research are made available to the public...including digital data"

***Coming soon* NIH Data Management Requirements**
What created the data?
What protocols were used?
Preservation?
Access to data?

2003 **2008** **2011** **2013** **2015** **2016**

NIH Genome Wide Association Studies
Required to de-identify data
Required to deposit in dbGap

NIH Genomic Data Sharing Policy
All genomic data must be shared

Timeline of funder data regulations

NIH Data Management Requirements

Full descriptions of the data and how it was collected

What **software/tools** were used to create the data?

What **protocols/steps** were used to create the data?

How will **long term preservation** of data be ensured?

How will **access** to data be provided?

NIH

National Institutes of Health

ICPSR DMP Elements

Data description	A description of the information to be gathered; the nature and scale of the data that will be generated or collected.	Yes	Expected Data
Existing data	A survey of existing data relevant to the project and a discussion of whether and how these data will be integrated.	Yes	Expected Data
Format	Formats in which the data will be generated, maintained, and made available, including a justification for the procedural and archival appropriateness of those formats.	Yes	Data Format and Dissemination
Metadata	A description of the metadata to be provided along with the generated data, and a discussion of the metadata standards used.	Yes	Data Format and Dissemination
Storage and backup	Storage methods and backup procedures for the data, including the physical and cyber resources and facilities that will be used for the effective preservation and storage of the research data.	Yes	Data Storage and Preservation of Access
Security	A description of technical and procedural protections for information, including confidential information, and how permissions, restrictions, and embargoes will be enforced.	Yes	Data Format and Dissemination
Responsibility	Names of the individuals responsible for data management in the research project.	Yes	Roles and Responsibility

NSF Requirements

1. the **types of data**, samples, physical collections, software, curriculum materials, and other materials to be produced in the course of the project;

2. the **standards to be used** for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);

3. **policies for access and sharing** including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;

4. **policies and provisions for re-use**, re-distribution, and the production of derivatives; and

5. **plans for archiving** data, samples, and other research products, and for preservation of access to them.

DMP Tool

NYU HEALTH SCIENCES LIBRARY

UNIVERSITY OF CALIFORNIA

DMPTool

Build Your Data Management Plan

CONTACT US

Think ahead. Plan ahead.

Discover the new, easy-to-use tool designed to build quality data management plans — for free

Go to the DMPTool

Learn about the DMPTool »

DISCOVER UC3

Explore our full suite of data services and tools

Learn More

RESOURCES & TRAINING

- Data management guides
- Webinars
- Slides and Marketing materials

View All

UC BY THE NUMBERS

- 862 DMPs created by UC researchers
- 980 users across UC

UC CONTACTS

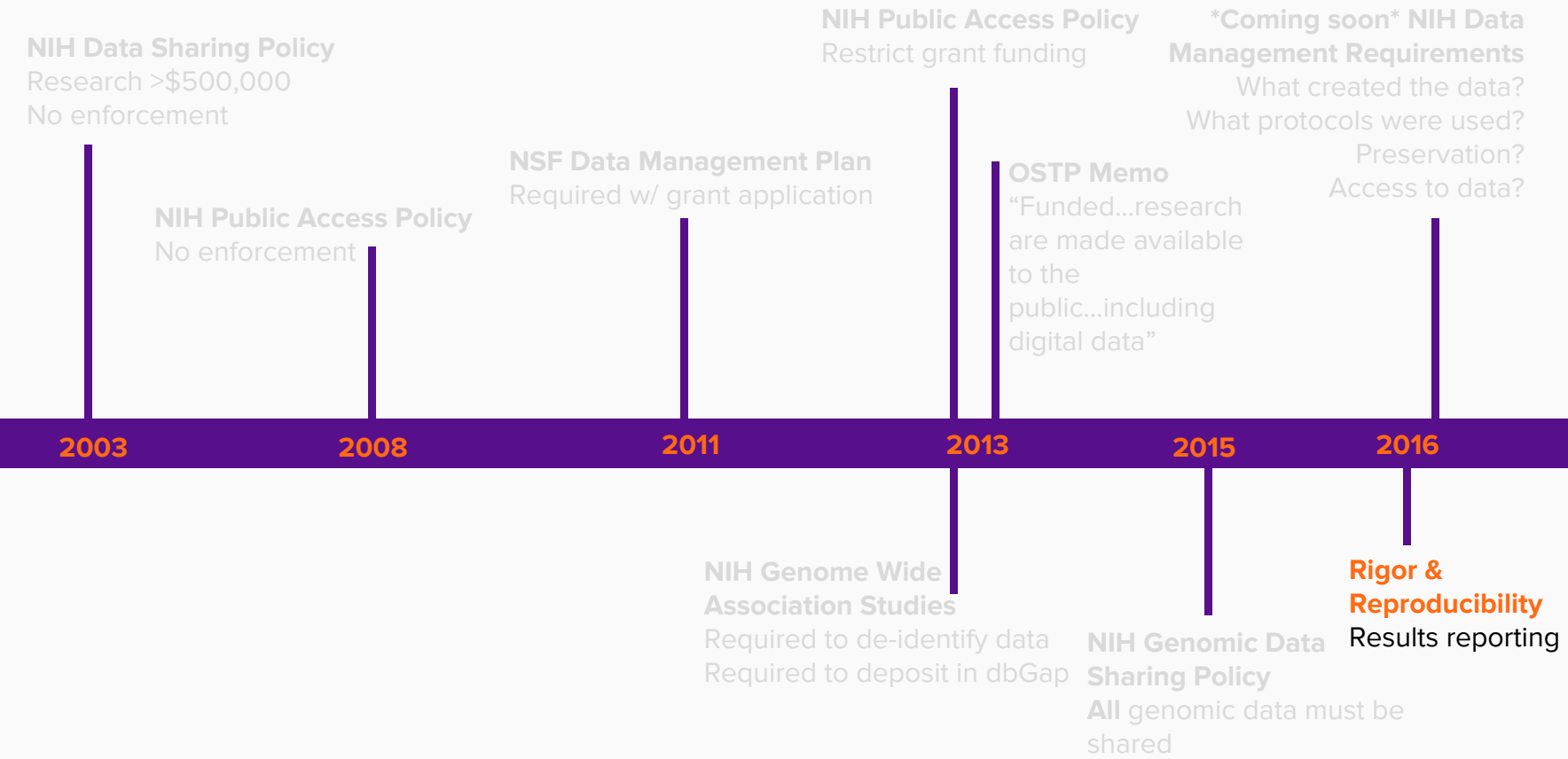
- Have a question?
- Get local campus help

RECENT DMP NEWS

- Get the latest information about data management and the DMPTool

View All

Privacy Policy | Accessibility Policy | Terms of Use | Contact Us | About



Timeline of funder data regulations

RDM regulations: Rigor & Reproducibility

New guidelines: January 25, 2016

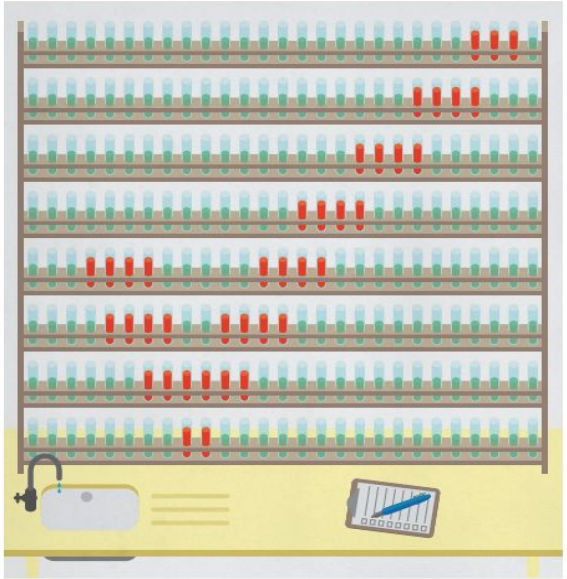
- Scientific premise must describe strengths/weaknesses of prior research
- Scientific rigor to ensure robust/unbiased experimental design, methodology, analysis, interpretation, reporting of results
- Consideration of relevant biological variables
- Authentication of key biological/chemical resources

RDM regulations: Rigor & Reproducibility

New guidelines: January 25, 2016

- Scientific premise must describe strengths/weaknesses of prior research
- Scientific rigor to ensure robust/unbiased experimental design, methodology, analysis, interpretation, reporting of results
  - **FULL TRANSPARENCY IN REPORTING EXPERIMENTAL DETAILS**
- Consideration of relevant biological variables
- Authentication of key biological/chemical resources

Rigor & Reproducibility



NIH plans to enhance reproducibility

Francis S. Collins and **Lawrence A. Tabak** discuss initiatives that the US National Institutes of Health is exploring to restore the self-correcting nature of preclinical research.

A growing chorus of concern, from scientists and laypeople, contends that the complex system for ensuring

shorter term, however, the checks and balances that once ensured scientific fidelity have been hobbled. This has compromised

outnumbered by the hundreds of thousands published each year in good faith.

Instead, a complex array of other factors seems to have contributed to the lack of reproducibi
ing of rese
increased
statements
details; an
basic elen
Crucial ex
are all too
ing, rando
calculatior
And some
sauce' to n
and withh
describe th
petitive ed
scientists
to further
Exacerb
and attitud
centres ar
ing agenci
the overva
high-profi
tres also p
in such jou
tenure, an
rewards[6].
Then th
not publis
researche
papers that point out scientific flaws in previously published work. Further compounding the problem is the difficulty of accessing unpublished data — and the failure of funding agencies to establish or enforce policies that insist on data access.

PRECLINICAL PROBLEMS

Reproducibility is potentially a problem in all scientific disciplines. However, human clinical trials seem to be less at risk because they are already governed by various regulations that stipulate rigorous design and independent oversight — including randomization, blinding, power estimates, pre-registration of outcome measures in standardized, public databases such as ClinicalTrials.gov and oversight by institutional review boards and data safety monitoring boards. Furthermore, the clinical trials community has taken important steps towards adopting standard reporting elements[7].

> "a complex array of other factors seems to have contributed to the **lack of reproducibility**. Factors include poor training of researchers in experimental design; increased emphasis on making provocative statements rather than presenting technical details; and **publications that do not report basic elements of experimental design**"

http://www.nature.com/news/policy-nih-plans-to-enhance-reproducibility-1.14586

ClinicalTrials.gov

First Received Date ICMJE	June 20, 2013
Last Updated Date	September 21, 2015
Start Date ICMJE	October 2013
Primary Completion Date	July 2015 (final data collection date for primary outcome measure)
Current Primary Outcome Measures ICMJE **(submitted: July 1, 2013)**	• For Phase I of the Study: Metrics Used to Understand Diabetes Control [Time Frame: 4 months] [Designated as safety issue: No] Identification of common factors patients use to understand their diabetes and diabetes control via a qualitative analysis of the patient interview resp • For Phase II of the Study: Change in Hemoglobin A1C [Time Frame: 6 months following enrollment] [Designated as safety issue: No] Change in A1C between enrollment and 6-months compared between study arms.
Original Primary Outcome Measures ICMJE **(submitted: June 24, 2013)**	Metrics Used to Understand Diabetes Control [Time Frame: 4 months] [Designated as safety issue: No] Identification of common factors patients use to understand their diabetes and diabetes control via a qualitative analysis of the patient interview respor
Change History	Complete list of historical versions of study NCT01886170 on ClinicalTrials.gov Archive Site
Current Secondary Outcome Measures ICMJE **(submitted: July 1, 2013)**	• For Phase I of the study: Feedback on alternative formats [Time Frame: 4 months] [Designated as safety issue: No] qualitative and quantitative analysis of the feedback received on the alternative communication formats reviewed with participants during the intervi • For Phase II of the Study: Understanding of diabetes control [Time Frame: At the time of enrollment] [Designated as safety issue: No] Accuracy of participant knowledge of level of current diabetes control
Original Secondary Outcome Measures ICMJE **(submitted: June 24, 2013)**	Feedback on alternative formats [Time Frame: 4 months] [Designated as safety issue: No] qualitative and quantitative analysis of the feedback received on the alternative communication formats reviewed with participants during the interview

The Final Rule

Data Elements Required in Final Rule	Provision No. in 42 CFR 11.48(a)	ClinicalTrials.gov PRS Pre-Final Rule Status		Comments
		Required	Optional	
Other measure(s)			X	Sub-element of Baseline Measure Information, (2)(iii). Any other measure(s) that were assessed at baseline and are used in the analysis of the primary outcome measure(s).
Name and Description of the Measure, including any categories that are used to submit Baseline Measure Data	(2)(iii)(A)	X		
Measure Type and Measure of Dispersion	(2)(iii)(B)	X		
Unit of Measure	(2)(iii)(C)	X		
Baseline Measure Data	(2)(iv)	X		
Number of Baseline Participants (and Units)	(2)(v)			If different from Overall Number of Baseline Participants or Overall Number of Units Analyzed
Outcomes and Statistical Analyses				
Outcome Measure Arm/Group Information	(3)(i)	X		
Analysis Population Information	(3)(ii)	X		
Number of Participants Analyzed	(3)(ii)(A)	X		
Number of Units Analyzed	(3)(ii)(B)	X		If the analysis is based on a unit other than participants, a description of the unit of analysis (e.g., eyes, lesions)
Analysis Population Description	(3)(ii)(C)		X	If Number of Participants Analyzed or Number of Units Analyzed differs from the number of human subjects or units assigned to the arm
Outcome Measure Information	(3)(iii)	X		
Name of the Specific Outcome Measure	(3)(iii)(A)	X		
Description of the Metric Used	(3)(iii)(B)		X	
Time Point(s) at which the Measurement was Assessed	(3)(iii)(C)	X		
Outcome Measure Type	(3)(iii)(D)	X		
Measure Type and Measure of Dispersion or Precision	(3)(iii)(E)	X		

https://prsinfo.clinicaltrials.gov/FinalRuleChanges-16Sept2016.pdf

Animal Research

OPEN ACCESS Freely available online PLoS one

Survey of the Quality of Experimental Design, Statistical Analysis and Reporting of Research Using Animals

Carol Kilkenny[1]*, Nick Parsons[2], Ed Kadyszewski[3], Michael F. W. Festing[4], Innes C. Cuthill[5], Derek Fry[6], Jane Hutton[7], Douglas G. Altman[8]

1 The National Centre for the Replacement, Refinement and Reduction of Animals in Research, London, United Kingdom, 2 Warwick Medical School, University of Warwick, Coventry, United Kingdom, 3 Pfizer Global Research and Development, Groton, Connecticut, United States of America, 4 Animal Procedures Committee, London, United Kingdom, 5 School of Biological Sciences, University of Bristol, Bristol, United Kingdom, 6 Animals Scientific Procedures Inspectorate, Home Office, Shrewsbury, United Kingdom, 7 Department of Statistics, University of Warwick, Coventry, United Kingdom, 8 Centre for Statistics in Medicine, University of Oxford, Oxford, United Kingdom

Detailed information collected from 271 publications:

59% stated hypothesis and number/characteristics of animals

13% used randomization

14% used blinding

30% of publications that used statistical methods did not describe methods

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0007824

Reproducibility of Preclinical Research

nature International weekly journal of science

Drug development: Raise standards for preclinical cancer research

C. Glenn Begley & Lee M. Ellis

Scientists in haematology and oncology departments at Amgen tried to confirm findings from 53 “landmark” studies

Findings confirmed in only 6 (11%) cases.

http://www.nature.com/nature/journal/v483/n7391/full/483531a.html

Principles and Guidelines for Reporting Preclinical Research

Joint workshop June 2014: NIH, NPG, Science
Consensus from journal editors:

Rigorous statistical analysis

Transparency in reporting

Data and material sharing

Consideration of refutations

Consider establishing best practice guidelines for:

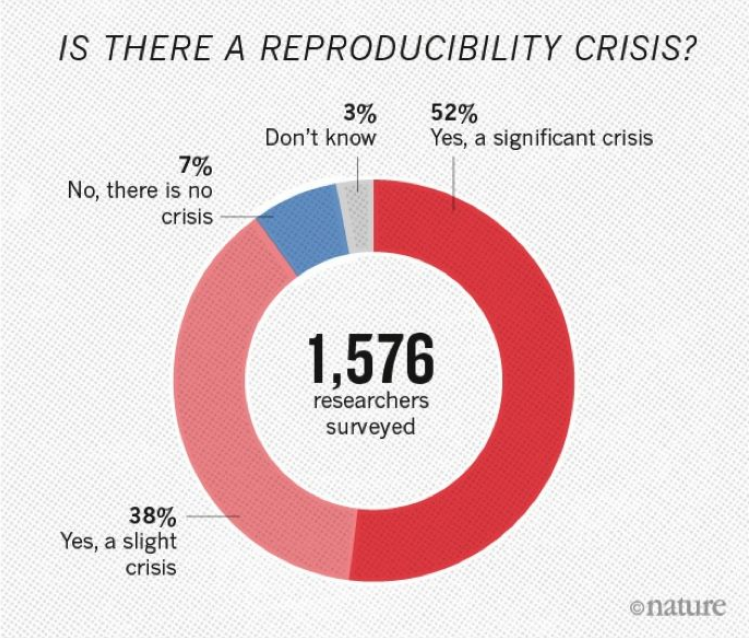
- Image based data
- Antibodies
- Cell lines
- Animals

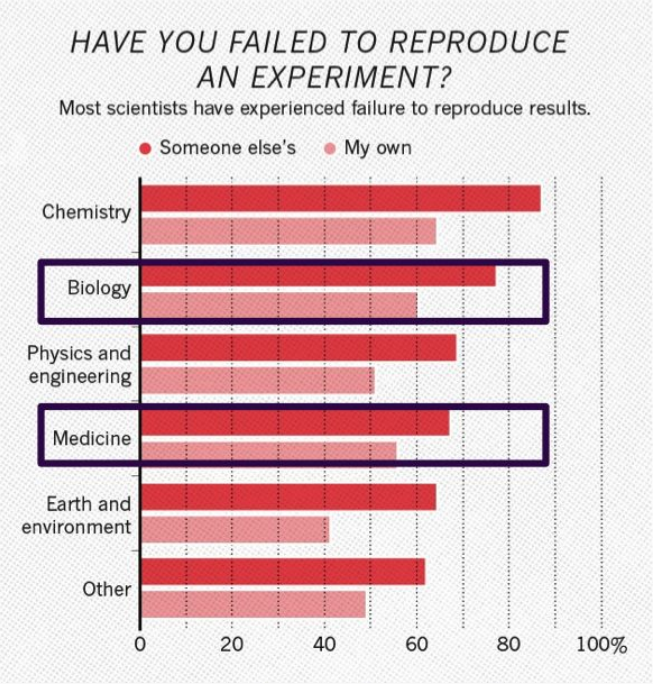
http://www.nih.gov/about/reporting-preclinical-research.htm

What do scientists think?

NATURE | NEWS FEATURE

1,500 scientists lift the lid on reproducibility





http://www.nature.com/news/1-500-scientists-lift-the-lid-on-reproducibility-1.1997017

Resource for Comparing Federal Policies

NYU HEALTH SCIENCES LIBRARY

http://datasharing.sparcopen.org/

SPARC

Who We Are | What We Do | Why It Matters | Become a Member

Browse Data Sharing Requirements by Federal Agency

This community resource for tracking, comparing, and understanding both current and future U.S. federal funder research data sharing policies is a joint project of SPARC & Johns Hopkins University Libraries. Click the icons below to select up to three agencies to view or compare. Click here to download the full data set.

Search for an agency...

AHRQ

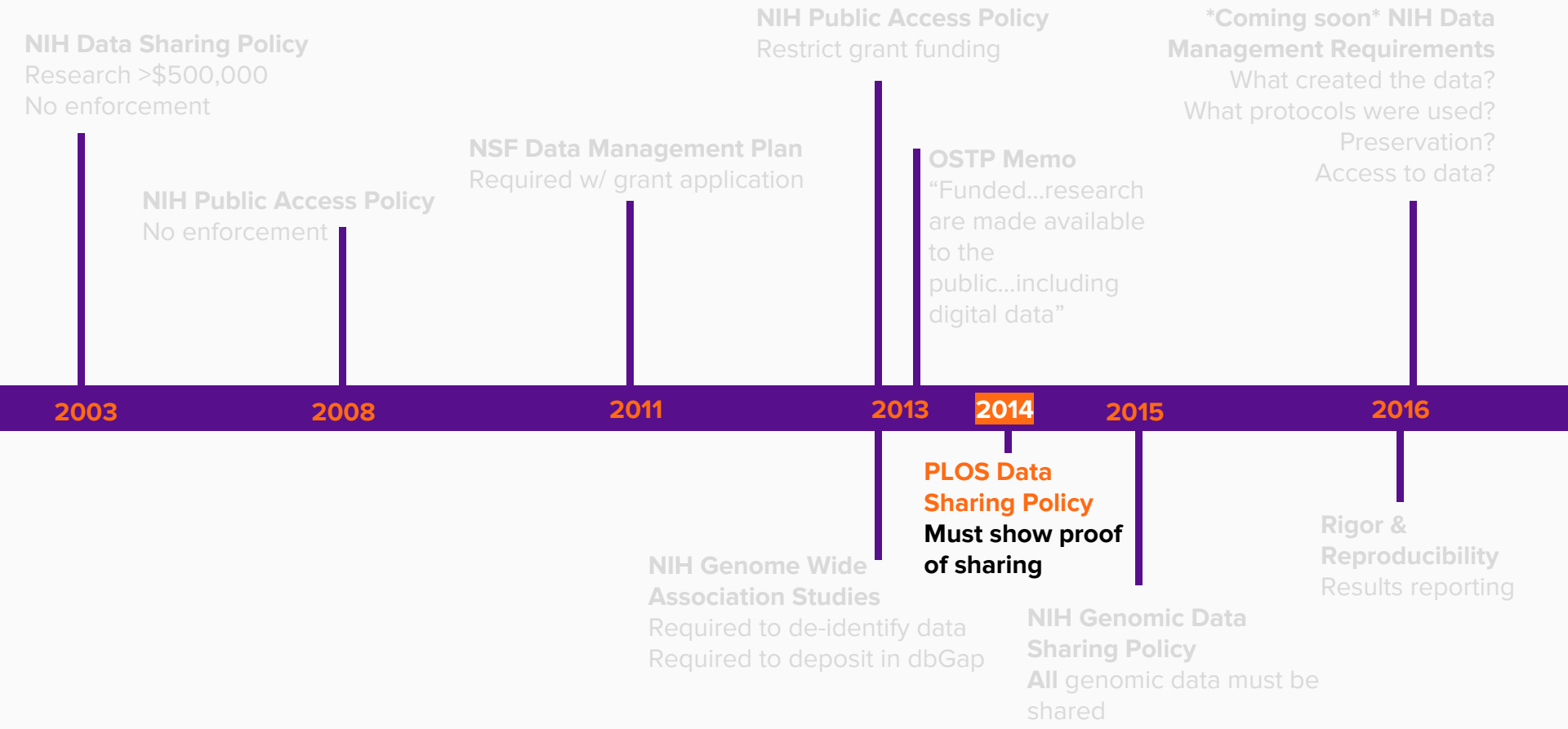
ASPR

CDC

Dept. of Defense

Dept. of Education

RDM publisher regulations



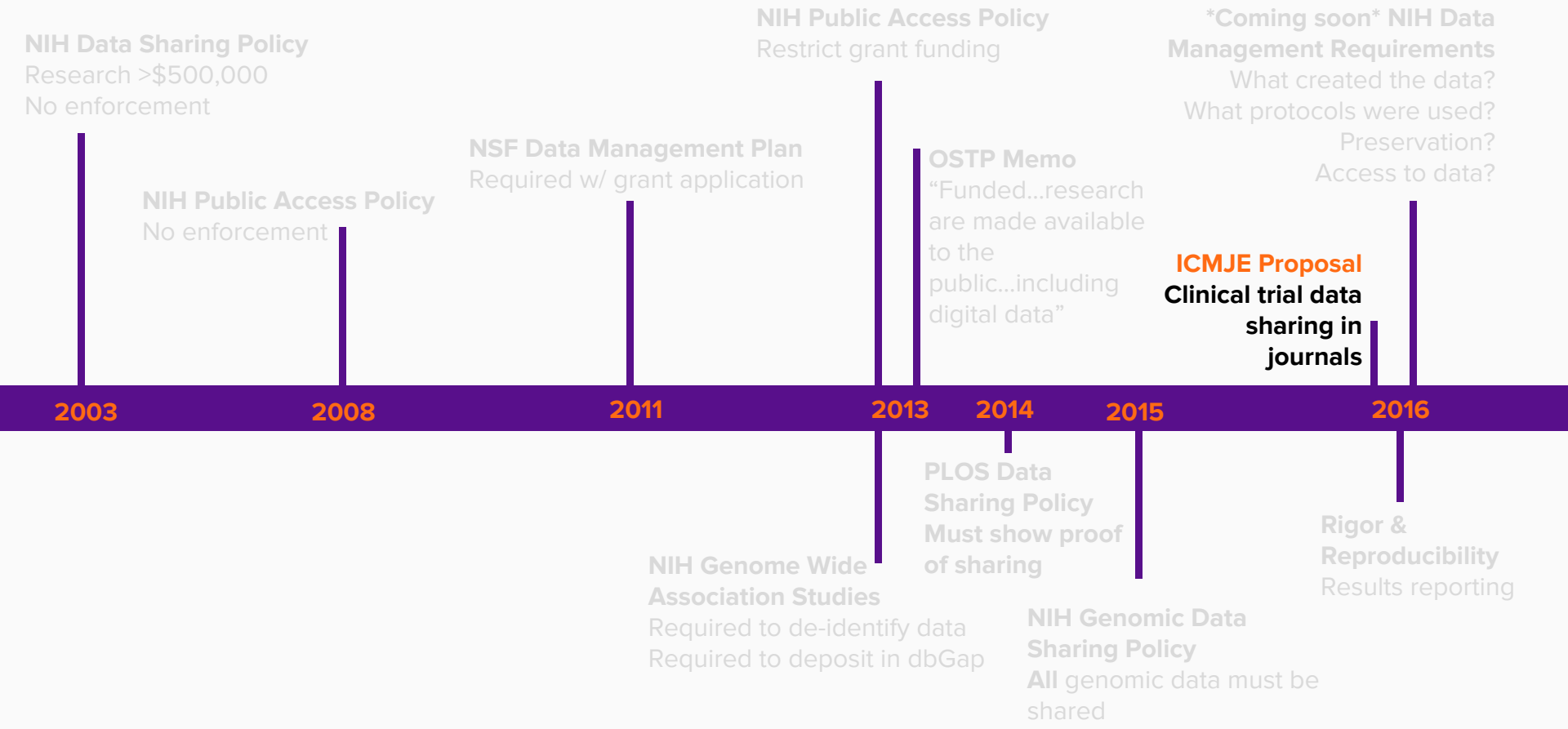
Timeline of publisher data regulations

PLOS Data Sharing Policy

PLOS

"Refusal to share data…in accordance with this policy will be ***grounds for rejection…***

…must specify that data are deposited publicly and list the name(s) of repositories along with digital object identifiers or accession numbers"



Timeline of publisher data regulations

ICMJE Proposal

Clinical trial data sharing

De-identified data underlying the results

No later than 6-months after publication

Open for comments (ended April 30)

The NEW ENGLAND JOURNAL of MEDICINE

EDITORIAL

Sharing Clinical Trial Data — A Proposal from the International Committee of Medical Journal Editors

Darren B. Taichman, M.D., Ph.D.Joyce Backus, M.S.L.S.Christopher Baethge, M.D.Howard Bauchner, M.D.Peter W. de Leeuw, M.D.Jeffrey M. Drazen, M.D.John Fletcher, M.B., B.Chir., M.P.H.Frank A. Frizelle, M.B., Ch.B., F.R.A.C.S.Trish Groves, M.B., B.S., M.R.C.Psych.Abraham Haileamlak, M.D.Astrid James, M.B., B.S.Christine Laine, M.D., M.P.H.Larry Peiperl, M.D.Anja Pinborg, M.D.Peush Sahni, M.B., B.S., M.S., Ph.D.Sinan Wu, M.D.

N Engl J Med 2016; 374:384-386 | January 28, 2016 | DOI: 10.1056/NEJMe1515172

The International Committee of Medical Journal Editors (ICMJE) believes that there is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk. In a growing consensus, many funders around the world — foundations, government agencies, and industry — now mandate data sharing. Here we outline the ICMJE's proposed requirements to help meet this obligation. We encourage feedback on the proposed requirements. Anyone can provide feedback at www.icmje.org by 18 April 2016.

The ICMJE defines a clinical trial as any research project that prospectively assigns people or a group of people to an intervention, with or without concurrent comparison or control groups, to study the cause-and-effect relationship between a health-related intervention and a health outcome. Further details may be found in the *Recommendations for the Conduct, Reporting, Editing and Publication of Scholarly Work in Medical Journals* at www.icmje.org.

As a condition of consideration for publication of a clinical trial report in our member journals, the ICMJE proposes to require authors to share with others the deidentified individual-patient data (IPD) underlying the results presented in the article (including tables, figures, and appendices or supplementary material) no later than 6 months after publication. The data underlying the results are defined as the IPD required to reproduce the article's findings, including necessary metadata. This requirement will go into effect for clinical trials that begin to enroll participants beginning 1 year after the ICMJE adopts its data-sharing requirements. (The ICMJE plans to adopt data-sharing requirements after considering feedback received to the proposals made here.)

ICMJE Proposal: Result

THE WATCHDOGS

New science data-sharing rules are two scoops of disappointment

By ADAM MARCUS @armarcus and IVAN ORANSKY @ivanoransky / JUNE 6, 2017

https://www.statnews.com/2017/06/06/data-sharing-rules-disappoint/

RDM regulations: Fear of retraction

Use specific cases of retraction to resonate with researchers

http://www.retractionwatch.com

Use examples that identify prominent researchers or publications

Isolate examples where data management is the trigger for lost credibility

Fear of retraction: Lost credibility

NYU HEALTH SCIENCES LIBRARY

Retraction Watch

NEJM paper on sleep apnea retracted when original data can't be found

with 4 comments

The authors of a paper in the *New England Journal of Medicine* are retracting it, after being unable to find data supporting a table that required corrections.

The NEW ENGLAND JOURNAL of MEDICINE

Discovered multiple errors in table

Fear of retraction: Lost credibility

NYU HEALTH SCIENCES LIBRARY

Retraction Watch

NEJM paper on sleep apnea retracted when original data can't be found

with 4 comments

The authors of a paper in the *New England Journal of Medicine* are retracting it, after being unable to find data supporting a table that required corrections.

The NEW ENGLAND JOURNAL of MEDICINE

Discovered multiple errors in table

Did not alter conclusions of article

Fear of retraction: Lost credibility

NYU HEALTH SCIENCES LIBRARY

Retraction Watch

NEJM paper on sleep apnea retracted when original data can't be found

with 4 comments

The authors of a paper in the *New England Journal of Medicine* are retracting it, after being unable to find data supporting a table that required corrections.

The NEW ENGLAND JOURNAL of MEDICINE

Discovered multiple errors in table

↓

Did not alter conclusions of article

↓

BUT could not locate raw data

Fear of retraction: Lost credibility

NYU HEALTH SCIENCES LIBRARY

Retraction Watch

NEJM paper on sleep apnea retra found

with 4 comments

The authors of a paper i *Medicine* are retra supporting a f

NGLAND NAL of MEDICINE

RETRACTION

ons of article

could not locate raw data

NYU HEALTH SCIENCES LIBRARY

RDM Climate

In Perspective

- It's an uphill battle
- Researchers say they will comply...but don't
- PLOS only game in town
- Even genomic researchers don't always share

?

Questions

EXERCISE 2

The interview

Researcher: Study issues

Errors entering data from forms into spreadsheet

Researcher: Study issues

Errors entering data from forms into spreadsheet

Despite verbal instruction from research coordinator, research team is still collecting data inconsistently

Researcher: Study issues

Errors entering data from forms into spreadsheet

Despite verbal instruction from research coordinator, research team is still collecting data inconsistently

Difficulty locating specific participants and variables within files

Researcher: Study issues

Errors entering data from forms into spreadsheet

Despite verbal instruction from research coordinator, research team is still collecting data inconsistently

Difficulty locating specific participants and variables within files

Push back from researchers who have asked for their data -- currently unusable

The interview: Tasks

Based on the article in Exercise 1, **what questions could you ask** the researcher about their RDM?

What is important to know about their data and research practices?

Can you think of **others at your institution** that could help support the researcher's needs?

The interview: Possible questions

NYU HEALTH SCIENCES LIBRARY

- What is the size of the data?
- What formats?
- Where is your data stored?
- What is the workflow?
- How do you collect the data?
- Do you reuse data?
- Who collects the data?
- Who has access rights to the data?
- What do you share with collaborators?
- How long do you want to save your data?

Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. **Data documentation best practices**
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

Data Management SNAFU

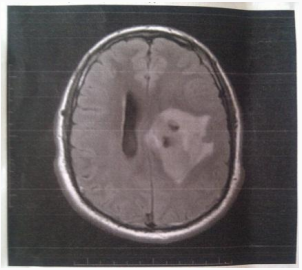
https://www.youtube.com/watch?v=66oNv_DJuPc

Data to do list

➔ Determine how, what, where, who will work with data (the workflow)	**CREATE DOCUMENTATION**
➔ Develop a system for naming files and folders	**CREATE DOCUMENTATION**
➔ Select and name variables to be collected	**CREATE DOCUMENTATION**

A simple workflow

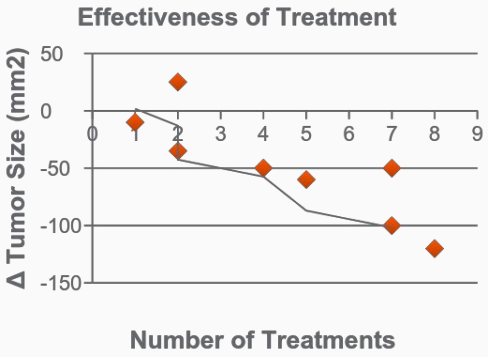
Create



Process

Patient ID	TumorArea PreTreat (mm^2)	TumorArea PostTreat (mm^2)	Number Therapy Sessions
1001	454	317	4
1002	234	82	7

Analyze



Workflows: Standard Operating Procedures

NYU HEALTH SCIENCES LIBRARY

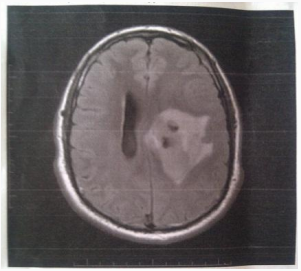
Purpose	What is being documented?
Responsibility	Who will carry out the procedure
Procedure	Step-by-step instructions of what is being done (e.g. experiment, treatment, interview)
Version number	Tracking documentation as it develops
Data updated	Dating documentation and data

How will data be created?

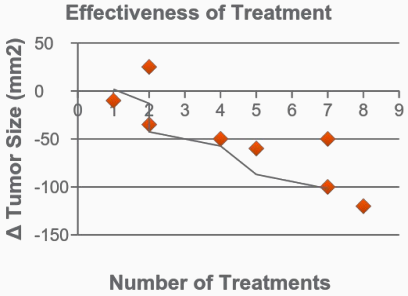
NYU HEALTH SCIENCES LIBRARY

What processes produce the data?

What transformations does the data go through?



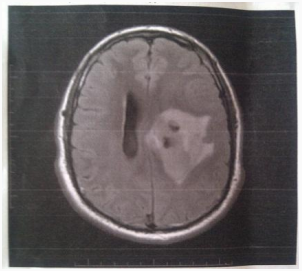
ID	TumorArea PreTreat (mm^2)	TumorArea PostTreat (mm^2)	Number Therapy Sessions
1001	454	317	4
1002	234	82	7



What data will be created?

What are the products of each step of the study?

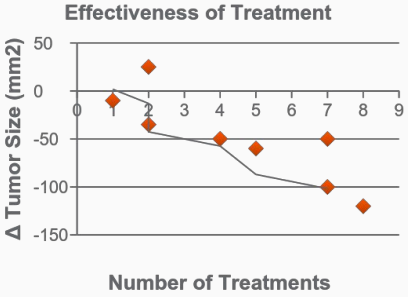
Image Files



Spreadsheets

ID	TumorArea PreTreat (mm^2)	TumorArea PostTreat (mm^2)	Number Therapy Sessions
1001	454	317	4
1002	234	82	7

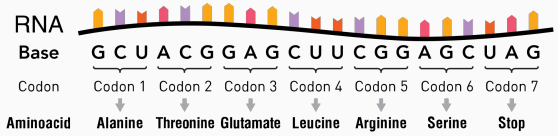
Graphs



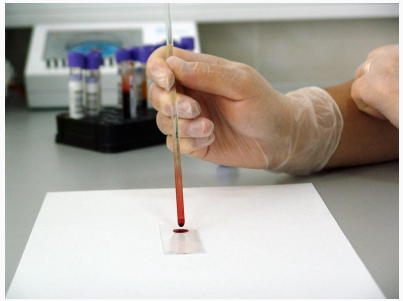
Data

SubjectID	Age	SBP	DBP
001	30	130	70
002	24	145	80
003	28	120	180

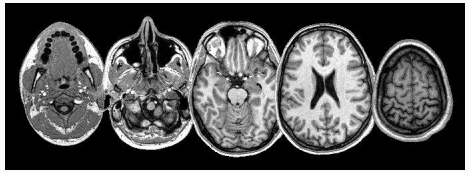
Tables of numbers



Sequences, base pairs

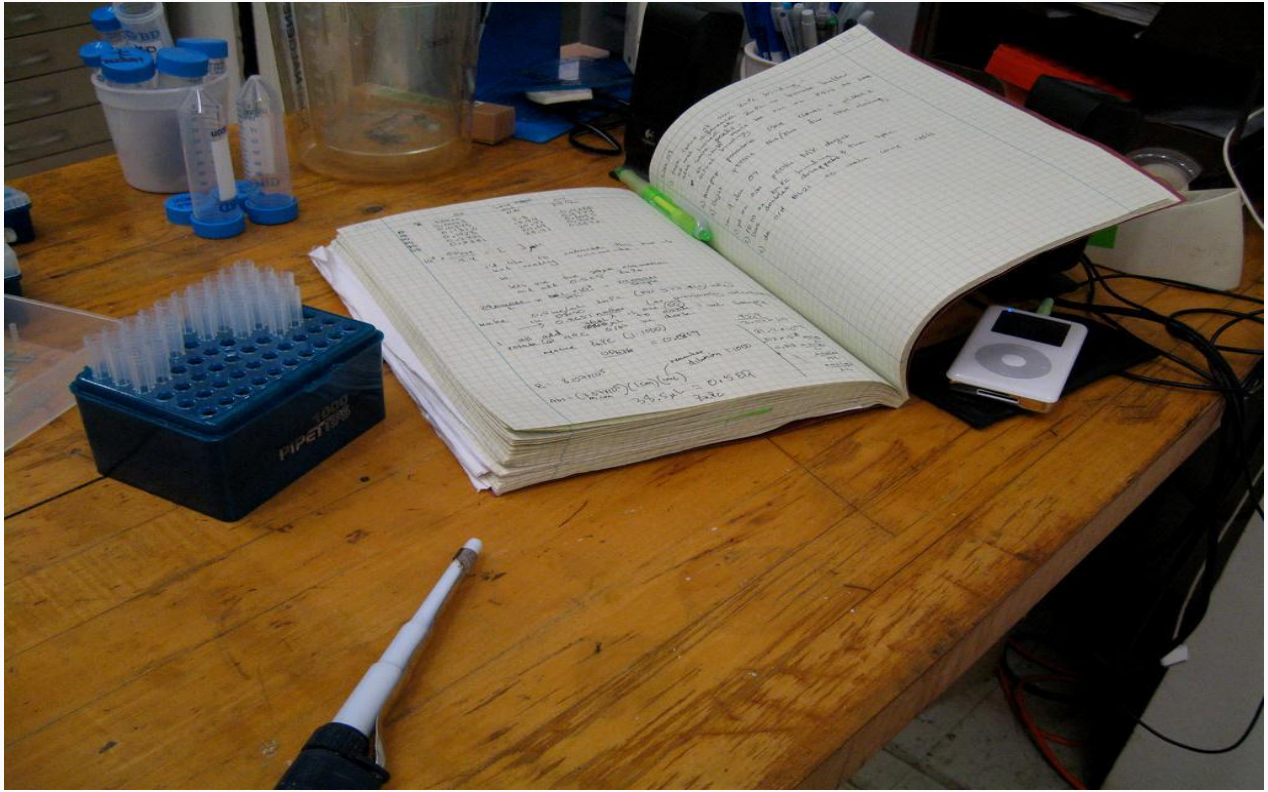


Samples, specimens, slides



Audio, video, imaging

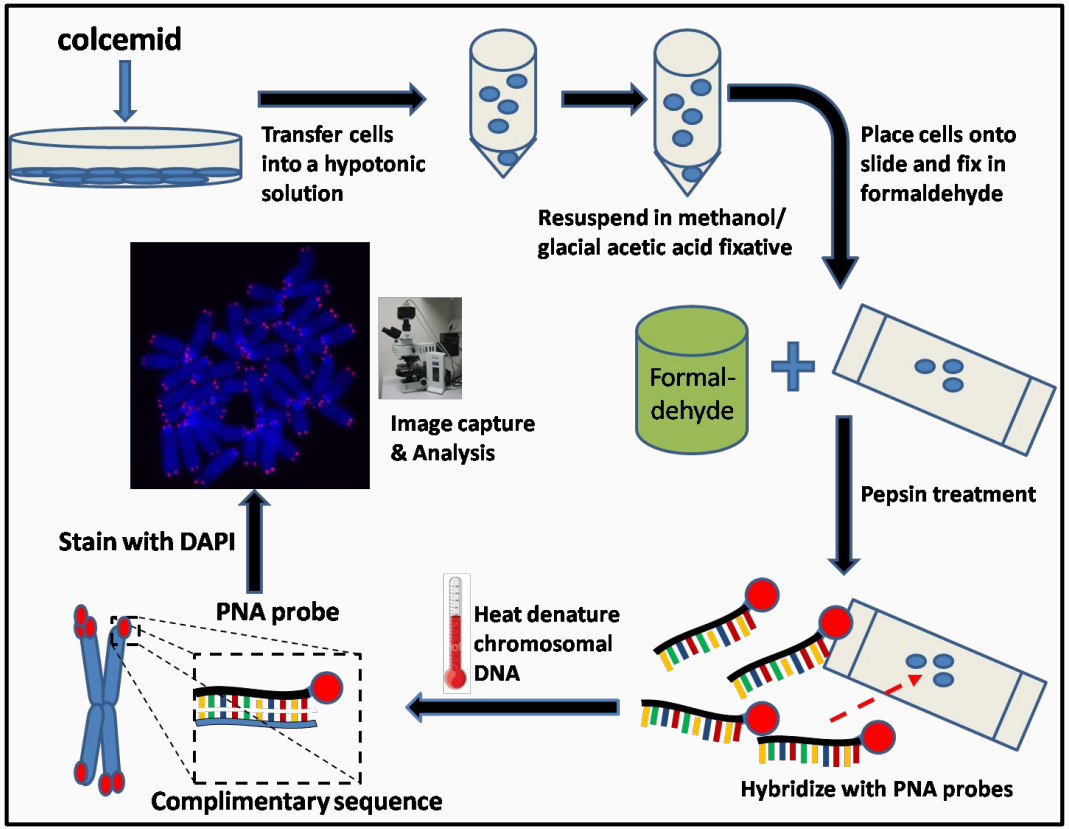
But what about?



and...

```
 *
 * However it's more likely that you'll just use
 * {@link ng.directive:ngApp ngApp} or
 * {@link angular.bootstrap} to simplify this process for you.
 *
 * @param {!string} name The name of the module to create or retrieve.
 * @param {!Array.<string>=} requires If specified then new module is being created. If
 *        unspecified then the module is being retrieved for further configuration.
 * @param {Function=} configFn Optional configuration function for the module. Same as
 *        {@link angular.Module#config Module#config()}.
 * @returns {module} new module with the {@link angular.Module} api.
 */
return function module(name, requires, configFn) {
  var assertNotHasOwnProperty = function(name, context) {
    if (name === 'hasOwnProperty') {
      throw ngMinErr('badname', 'hasOwnProperty is not a valid {0} name', context);
    }
  };

  assertNotHasOwnProperty(name, 'module');
  if (requires && modules.hasOwnProperty(name)) {
    modules[name] = null;
  }
  return ensure(modules, name, function() {
    if (!requires) {
      throw $injectorMinErr('nomod', "Module '{0}' is not available! You either misspelled " +
         "the module name or forgot to load it. If registering a module ensure that you " +
         "specify the dependencies as the second argument.", name);
    }

    /** @type {!Array.<Array.<*>>} */
    var invokeQueue = [];

    /** @type {!Array.<Function>} */
    var configBlocks = [];

    /** @type {!Array.<Function>} */
    var runBlocks = [];

    var config = invokeLater('$injector', 'invoke', 'push', configBlocks);
```

and…


colcemid
Transfer cells into a hypotonic solution
Resuspend in methanol/ glacial acetic acid fixative
Place cells onto slide and fix in formaldehyde
Formal-dehyde
Pepsin treatment
Hybridize with PNA probes
Heat denature chromosomal DNA
Complimentary sequence
PNA probe
Stain with DAPI
Image capture & Analysis

and…

NYU HEALTH SCIENCES LIBRARY

Study Type ICMJE	Interventional
Study Phase	*Not Provided*
Study Design ICMJE	Allocation: Randomized Endpoint Classification: Efficacy Study Intervention Model: Parallel Assignment Masking: Open Label Primary Purpose: Supportive Care
Condition ICMJE	Type 2 **Diabetes**
Intervention ICMJE	Behavioral: telephonic Between 4-8 phone calls each year for health behavior counseling to improve HbA1c
Study Arm (s)	• Experimental: Telephonic Tailored telephonic intervention to improve HbA1c for participants in the **diabetes** registry Intervention: Behavioral: telephonic • Active Comparator: Standard registry People with **diabetes** who are in the A1c registry may receive letters from the DOHMH to promote improved A1c and also give lists of bronx resources for healther foof and activites Intervention: Behavioral: telephonic
Publications *	• Chamany S, Walker EA, Schechter CB, Gonzalez JS, Davis NJ, Ortega FM, Carrasco J, Basch CE, Silver LD. Telephone Intervention to Improve Diabetes Control: A Randomized Trial in the N Am J Prev Med. 2015 Dec;49(6):832-41. doi: 10.1016/j.amepre.2015.04.016. Epub 2015 Jul 29. • Davis NJ, Schechter CB, Ortega F, Rosen R, Wylie-Rosett J, Walker EA. Dietary patterns in Blacks and Hispanics with diagnosed diabetes in New York City's South Bronx. Am J Clin Nutr. 201 10.3945/ajcn.112.051185. Epub 2013 Feb 27.

* **Includes publications given by the data provider as well as publications identified by ClinicalTrials.gov Identifier (NCT Number) in Medline.**

Recruitment Information

Recruitment Status ICMJE	Completed
Enrollment ICMJE	941
Completion Date	June 2012
Primary Completion Date	November 2011 (final data collection date for primary outcome measure)
Eligibility Criteria ICMJE	Inclusion Criteria: • Subjects will be those patients with diabetes who speak English and/or Spanish and reside in the South Bronx. • Subjects will be adults, > 18 years, with diabetes, who become part of the NYC registry by virtue of having a reported HbA1c >7% to the DOHMH. • The sampling frame for this study comprises virtually all adults with diabetes in the South Bronx. Exclusion Criteria: • Age < 18 years • A1c < = 7 % • Refuses informed consent and HIPAA consent • Cognitive dysfunction as assessed by telephone

Where should the data live during the experiment?

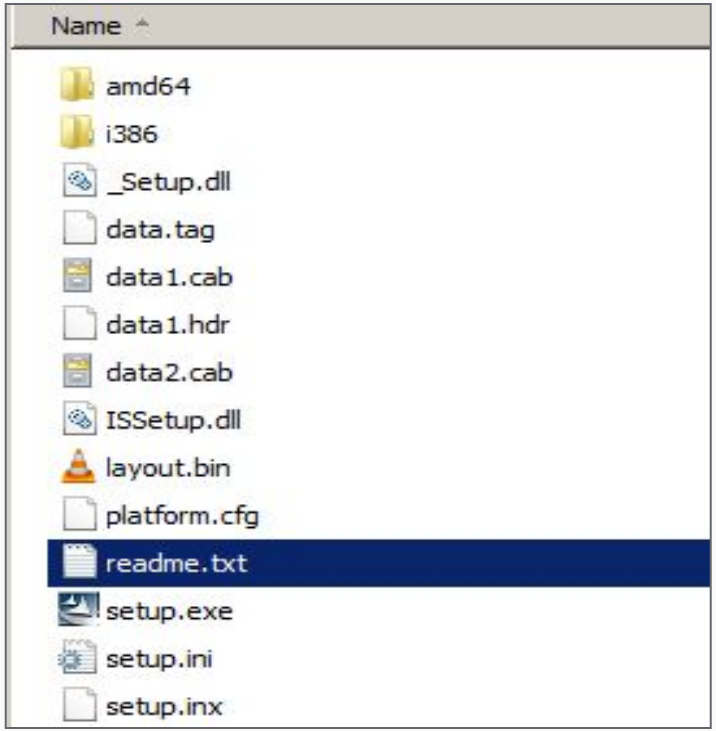
Where is it stored in different parts of the workflow?

Where is it backed up?

Where should different versions be stored? How should they be managed?

And document it!

- All processes should be documented
- Few researchers will do this
- Starting point: readme.txt



Anything is better than nothing!

Who needs access to the data?

Should data be restricted to a limited set of people?

Should files or folders be password protected?

Should simultaneous access be restricted?

Who is responsible for the quality of the data?

EVERYONE

Who is responsible for the quality of the data?

~~EVERYONE~~

(but everyone means no one)

Who is responsible for quality?

NYU HEALTH SCIENCES LIBRARY

Useful for one person to take ownership of:

- Adhering to naming conventions
- Minimum documentation
- Access controls
- Versioning
- Data validation
- Backing up data

?

Questions

Data to do list

➔ ~~Determine how, what, who, where will work with data (the worklow)~~	~~CREATE DOCUMENTATION~~
➔ Develop a system for naming files and folders	CREATE DOCUMENTATION
➔ Select and name variables to be collected	CREATE DOCUMENTATION

File names

sam_1262011.tif

File names

sam_1262011.tif

12 June, 2011?
December 6, 2011?
January 26, 2011?

File names

sam_1262011.tif

Scanning acoustic microscope?
S-Adenosyl methionine?
Sam Lee?

12 June, 2011?
December 6, 2011?
January 26, 2011?

File names

NYU HEALTH SCIENCES LIBRARY

sam_1262011.tif

Scanning acoustic microscope?	12 June, 2011?
S-Adenosyl methionine?	December 6, 2011?
Sam Lee?	January 26, 2011?

Unambiguous dates, the **ISO standard**:

YYYYMMDD *or* YYYY-MM-DD

e.g. 20120612 = June 6, 2012

YYYYMMDDTHH:MM:SS

e.g. 20120612T14:03:12 = June 6, 2012 2:03:12 pm

File names: Example

1 rat heart

File names: Example

1 rat heart

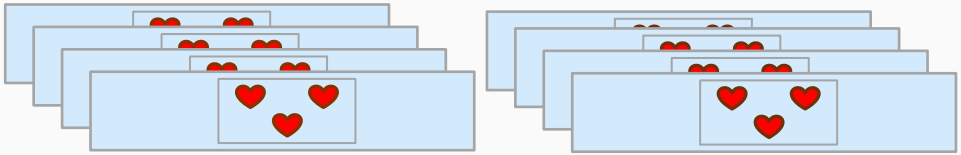
100s
of slices

File names: Example

1 rat heart

100s of slices

100s of slides

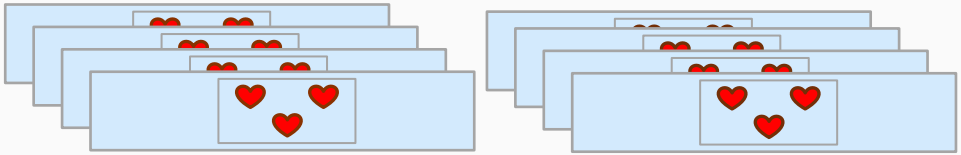


File names: Example

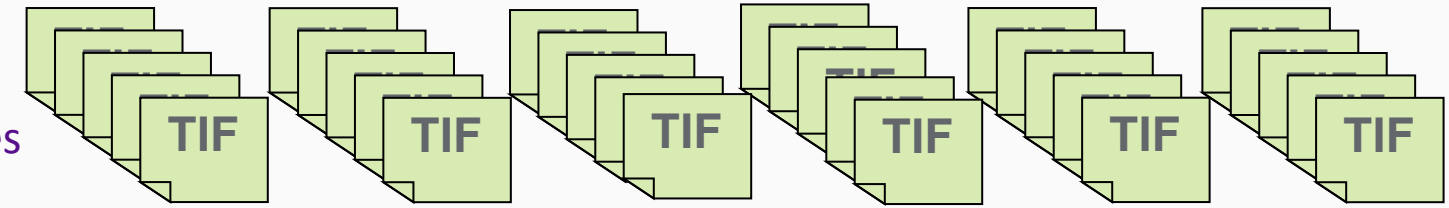
1 rat heart

100s of slices

100s of slides



1000s of image files

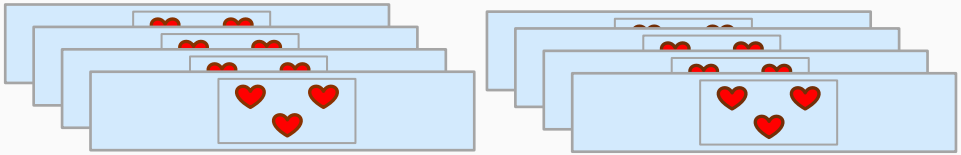


File names: Example

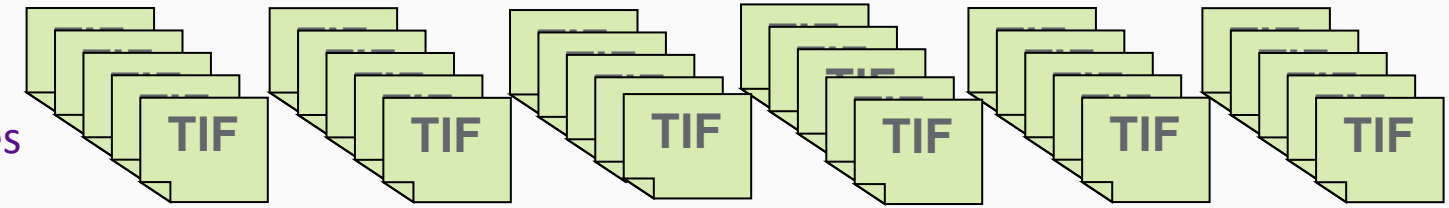
1 rat heart

100s of slices

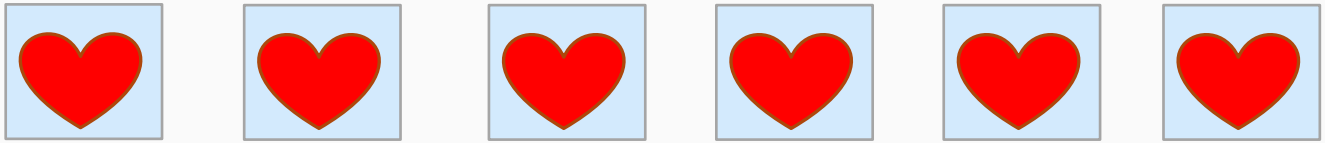
100s of slides



1000s of image files



100s of *huge* images



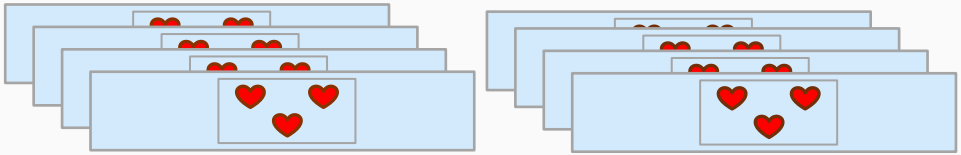
File names: Example

1 rat heart

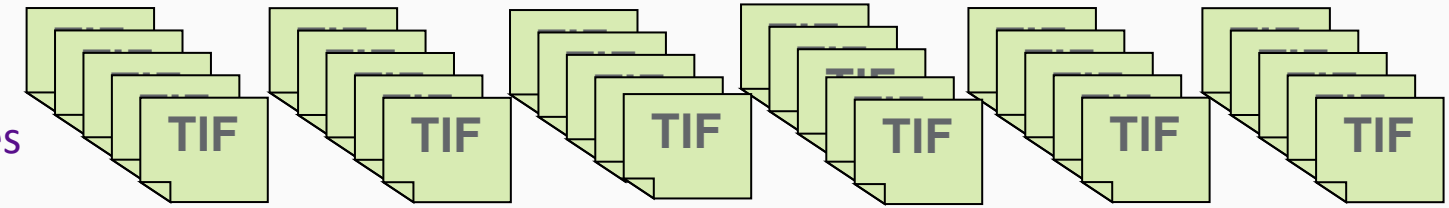
100s of slices

3 postdocs

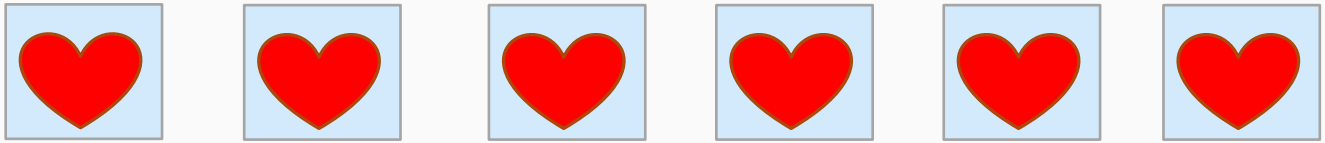
100s of slides



1000s of image files



100s of *huge* images

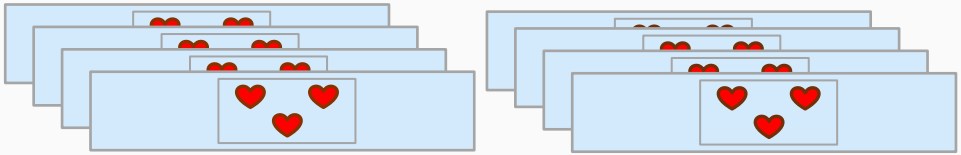


File names: Example

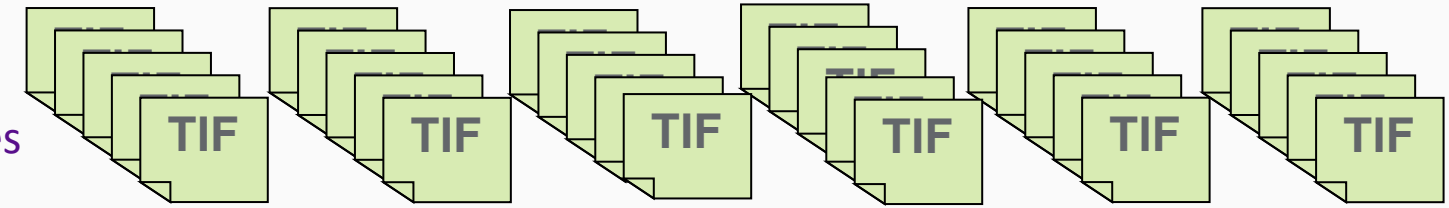
1 rat heart

100s of slices

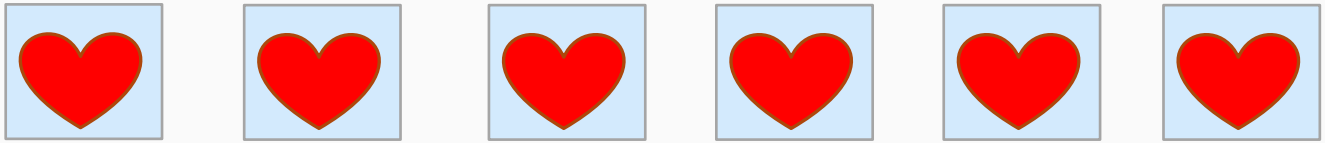
100s of slides



1000s of image files



100s of *huge* images



3 postdocs

5-7 experiments/week

When file names go wrong...

NYU HEALTH SCIENCES LIBRARY

ura Gel 5 - 3.10.11	1sc
ura Gel 6 - 3.10.11	1sc
ura Mon R	1sc
Restriction Digest 5.26.11 DNA 1 & 2	1sc
awesome 2010	1sc
awesomer 2010	1sc
Deb 2010-03-09 yeast gel	1sc
7.22.10 Gel 1=60bp wt706 Tmp Tm & cycle	1sc
7.22.10 Gel 2=60bp Temp-wt001& Neal	1sc
7.22.10 Gel 2=60bp wt001 & Neal Temp	1sc
7.22.10 Gel 3=100bp wt706 Tmp, Tm &cycle	1sc
7.22.10 Gel 4=100bp wt001 & Neal Temp	1sc
ura gel 1 Mon R starred	1sc
ura gel 2 Monday Ravenclaw un-starred	1sc
Dpn Gel 5 10 WT	1sc
a	1sc
attractive	1sc
group	1sc
joe	1sc
lab 2010-10-04 StatiionC	1sc

Good file naming

AtherRat_ex012_norm_lipitor_056_raw.tif

AtherRat = experiment series name
ex012 = experiment number 12
norm = normal heart, no atherosclerosis
lipitor = treatment given
056 = slice number
raw = data stage

In the folder:

NYU HEALTH SCIENCES LIBRARY

Name	Date modified	Type
AtherRat_ex012_ather_lipitor_126.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_lipitor_127.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_lipitor_128.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_lipitor_129.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_001.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_002.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_003.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_004.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_005.tif	5/9/2014 7:55 PM	TIFF imag
AtherRat_ex012_ather_notreat_006.tif	5/9/2014 7:55 PM	TIFF imag

File names should:

Embody their content, including major parameters

AtherRat_ex012_ather_lipitor_128.tif

DataDictionary_SmokingCessation.csv

DataCollection_Subject001_DepressionScale_20160102.csv

File names should:

Have non-cryptic/intuitive names *where possible:*

AtherRat_SOP_DataValidation_v01.docx

MLACE_RDMClass_20160304.pptx

DataCollection_ReadMe.txt

File names should:

NYU HEALTH SCIENCES LIBRARY

Be extensible. “`ex001`” **not** “`ex1`”

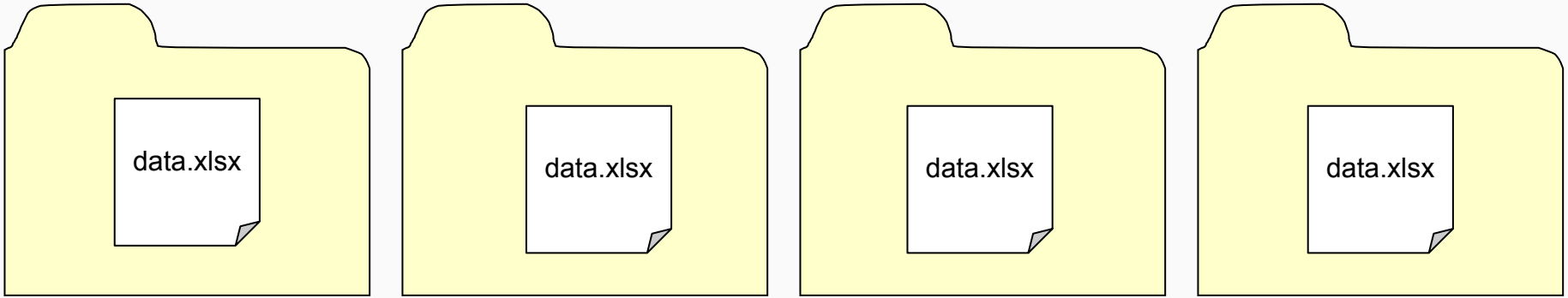
AtherRat_ex001_lipitor.tif
AtherRat_ex002_lipitor.tif
AtherRat_ex003_lipitor.tif
AtherRat_ex004_lipitor.tif
AtherRat_ex005_lipitor.tif
AtherRat_ex006_lipitor.tif
AtherRat_ex007_lipitor.tif
AtherRat_ex008_lipitor.tif
AtherRat_ex009_lipitor.tif
AtherRat_ex010_lipitor.tif

VS.

AtherRat_ex1_lipitor.tif
AtherRat_ex10_lipitor.tif
AtherRat_ex2_lipitor.tif
AtherRat_ex3_lipitor.tif
AtherRat_ex4_lipitor.tif
AtherRat_ex5_lipitor.tif
AtherRat_ex6_lipitor.tif
AtherRat_ex7_lipitor.tif
AtherRat_ex8_lipito.tif
AtherRat_ex9_lipitor.tif

File names should:

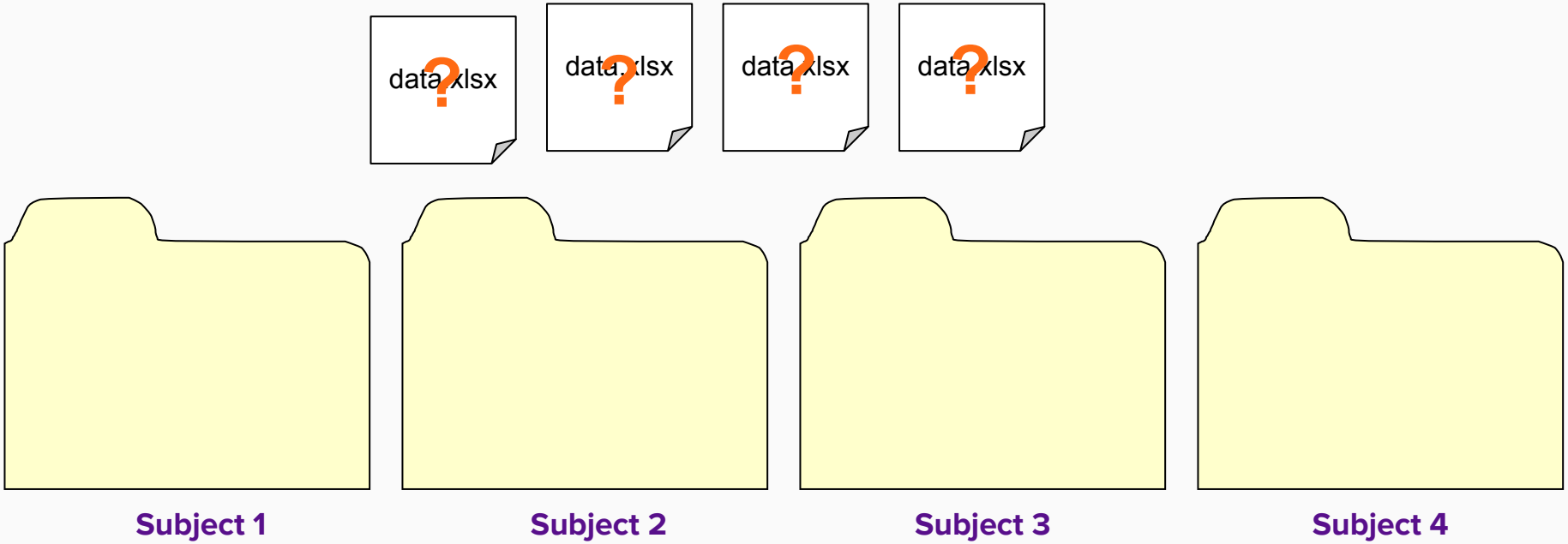
Be unique, where possible and practical. Avoid 20 files called "**data.xlsx**" in different folders



Subject 1 **Subject 2** **Subject 3** **Subject 4**

File names should:

Be unique, where possible and practical. Avoid 20 files called "**data.xlsx**" in different folders



File names should:

Be wary of using special characters – restrict file names to **numbers**, **letters**, and **underscores**

File names should:

NYU HEALTH SCIENCES LIBRARY

Use **underscore ("_")** instead of space to separate words in file names

DataAnalysis_AlcoholConsumption_2016.spss

DataDictionary_SmokingCessationForm.csv

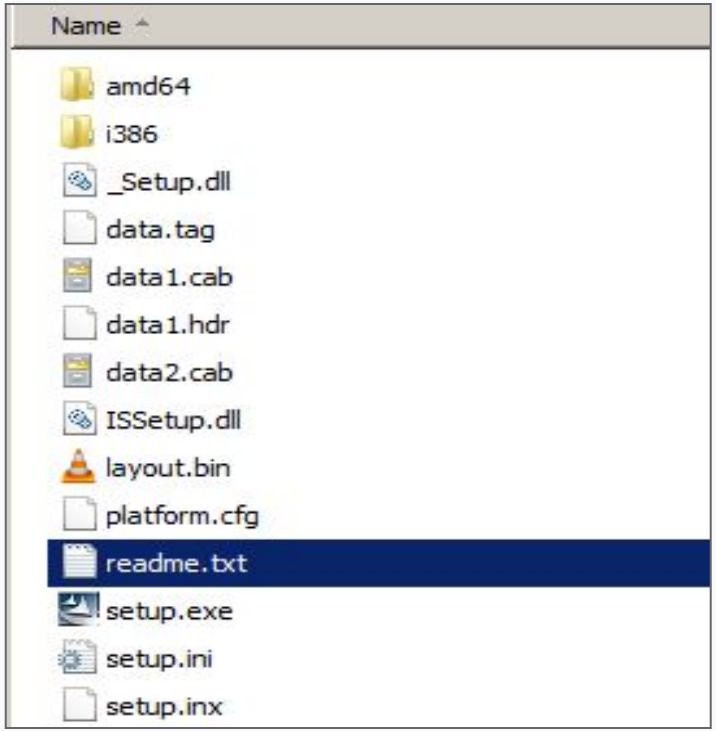
INSTEAD OF

~~**Data dictionary for smoking cessation collection form.csv**~~

~~**Analyzed alcohol consumption data.spss**~~

And document it!

- Selecting meaningful names is part of documentation
- Document the naming convention and file structure
- Starting point: readme.txt



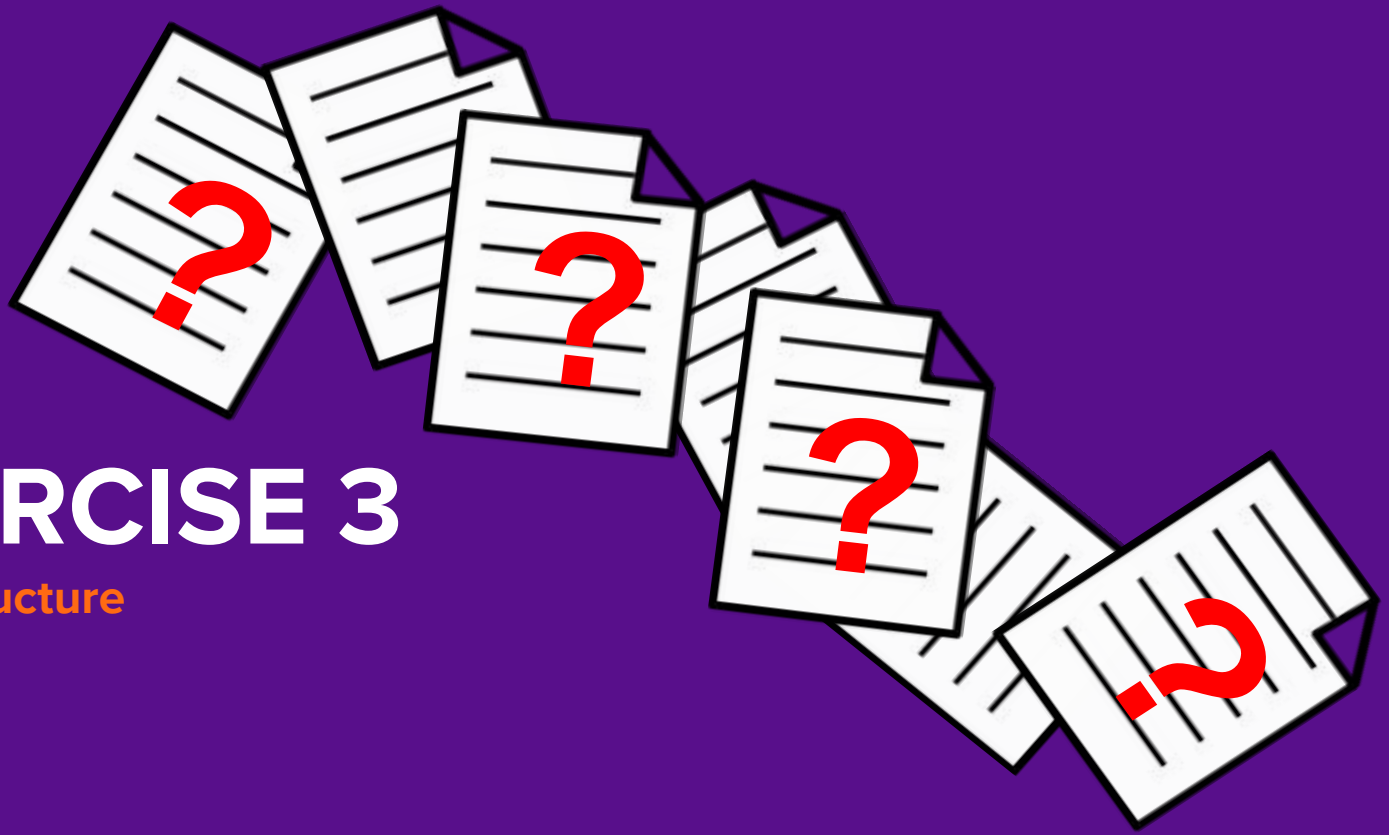
Anything is better than nothing!

?

Questions

Break

NYU HEALTH SCIENCES LIBRARY



EXERCISE 3

File structure

File structure: Tasks

How can you **improve** the file names?

How would you **organize** these files?

File structure: Revised file names

NYU HEALTH SCIENCES LIBRARY

File Name	Contents	Revised file name
BASS M.xsl	Mother bioassay of saliva samples	BioassayMotherSaliva.csv
BASS I.xsl	Infant bioassay of saliva samples	BioassayInfantSaliva.csv
BASS M: Scan10001.jpeg	Scan of saliva sample for participant 10001	BioassayMotherSaliva_Scan_10001.tiff
BASS M: Scan10002.jpeg	Scan of saliva sample for participant 10002	BioassayMotherSaliva_Scan_10002.tiff
BASS M: Scan10003.jpeg	Scan of saliva sample for participant 10003	BioassayMotherSaliva_Scan_10003.tiff
BASS I: Scan Sally Smith.jpeg	Scan of saliva sample for infant of Sally Smith	BioassayInfantSaliva_Scan_10001.tiff
BASS I: Scan Tina Tartare.jpeg	Scan of saliva sample for infant of Tina Tartare	BioassayInfantSaliva_Scan_10002.tiff
BASS I: Scan J Fine.jpeg	Scan of saliva sample for infant of J Fine	BioassayInfantSaliva_Scan_10003.tiff
TL Data.xsl	All Timeline Interview Data	TimelineInterview_Data_Final.csv
TL Jim1M.wpd	Jim's timeline interview data file (first round)	TimelineInterviewData_Jim_v1.csv
TL Jim1+2M.xsl	Jim's timeline interview data file (second round merged with first round)	TimelineInterviewData_Jim_v2.csv
Analysis New.sas	Depression Test data analyzed using ANOVA	DepressionTest_Analysis_ANOVA.sas
Analysis Old.sas	Older version of analyzed Depression Test using T-Test	DepressionTest_Analysis_TTest.sas
Analysis.sas	Analyzed Depression test data using logistic regression	DepressionTest_Analysis_LogisticRegression.sas

Data to do list

➔ ~~Determine how, what, who, where will work with data (the worklow)~~	~~CREATE DOCUMENTATION~~
➔ ~~Develop a system for naming files and folders~~	~~CREATE DOCUMENTATION~~
➔ Select and name variables to be collected	CREATE DOCUMENTATION

Identify appropriate variables

Question:

"I want to analyze patient lab tests with type II diabetes"

Before starting the study:

Walk through the process in detail. What will be collected?

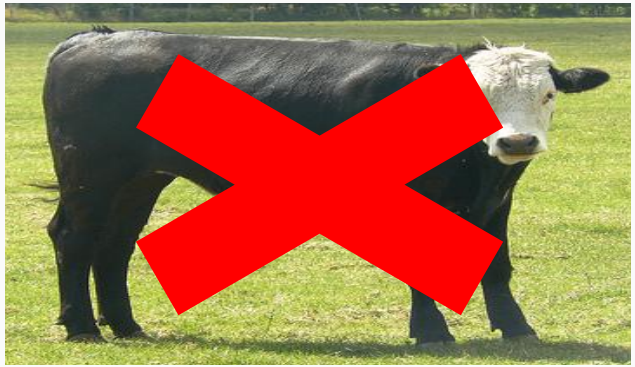
How will it be analyzed?

What are the exact variables you will need to perform the analysis?

Storing data in its rawest form

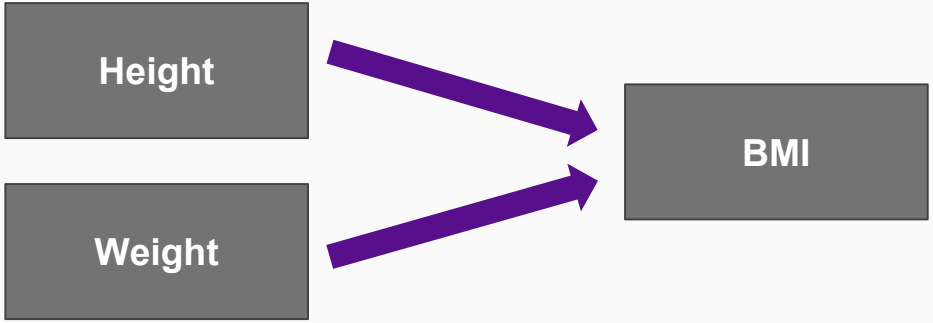
BMI

Storing data in its rawest form



BMI

Storing data in its rawest form

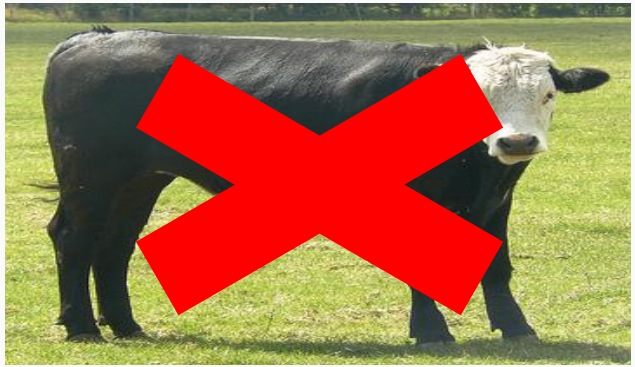


Storing data in its rawest form

Spike time

Spike peak voltage

Storing data in its rawest form



- Spike time
- Spike peak voltage
- ✗ Spike duration
- ✗ Spike threshold

Capture the most precise value

Avoid early categorization where **precise measurements** are possible

Ideal	**Acceptable**	**Bad**
Date of Birth Date of Visit	Age at time of visit	Age group (adult, child etc)

Use unambiguous variable names

Follow similar rules to file names:

- Use underscores
- Make as intuitive as possible
- Avoid similar names
- Create documentable rules for variables

Unambiguous example

Variable description:

"Pulse Oximetry Percentage"

Possible variable names:

pulse_ox_pct

pulse_oximetry

?

Questions

EXERCISE 4

Identifying data management errors

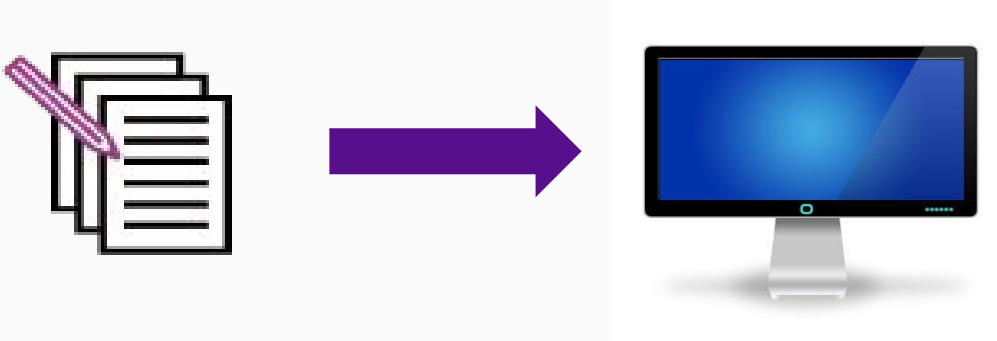
Data management errors: Paper form

How can the questions in the **paper form** be improved to support better data management?

Data management errors: Spreadsheet

What errors can you identify in the spreadsheet?

How can these errors be fixed?



Paper form:

Identifying errors

Exercise 3: Maternal Smoking During Pregnancy & Newborn Neurobehavior: Data Collection

Clinical Evaluation:

Subject ID:_____

Participant:_____

Child:_____

Date of Interview:_____

Age:_____

Gender:

- Male
- Female

BMI:_____

Vital Signs (enter vital signs, and others as appropriate to the case):

Date	Time	Pulse	Blood Pressure	Respiratory Rate	Temperature	Pulse Oximetry

Timeline Interview

Were there any times when you had nothing at all to drink, not even a drop of alcohol?

What was the longest period of total abstinence?

What was the longest period you were drinking?

Depression Test

	Rarely	Sometimes	Occasionally	Most of the time
I was bothered by things that usually don't bother me				
I did not feel like eating				

Physical data: Identifying errors

NYU HEALTH SCIENCES LIBRARY

	A	B	C	D	E	F	G	H	I	J	K	L
1		Maternal Smoking During Pregnancy and Newborn Neurobehavior										
2	Subject ID	Participant	Child	DOI	Age	BMI	Blood Pressure	Pulse1	Pulse2	LPAbstin	LP Drink	Bothered by things that usually don't bother me
3	10001	Smith Sally	38 weeks	2011	41	22	120/60	40	9	1	0	Rarely
4	10002	Tina Tartare	B Tartare	06-Jur	3 weeks	33	130 over 64		9	4 years	2 years	Mostly
5	10003	J Fine	P Fine	11-05-04	29	28	120-65	60	9	48	12	Sometimes
6	10001	Lucy Bordeaux	40 weeks	2001 - 3 - 4	37	25	160/70	95	10	8 days	17 years	"
7	10005	Rebecca Alice	"	05-11-11	25	20	125/60	49	9	20 months	2 days	N/A
8	10002	Polyanne Trudy	"	September 25, 1999	2 months	22	130/40	"	9	Missing	"	Mostly
9	10006	Breer, Diane	Breer, Judy	01-Apr-00	34	24	121/62	53	9	89 days	6 weeks	Occasionally
10	10007	Participant 7	Y	Jan-00	44	N/A	130/60	46	9	12	22	Occasionally

Document variable names

- What do the variable names mean?
- What does each variable contain?
- How do variables relate to each other?
- Are there a limited set of possible values?

Name	Type	Description	Possible values
Stain	Text	Stain used on cell sample	IO = Iodine; EY = Eosin Y; MB = Methylene blue;

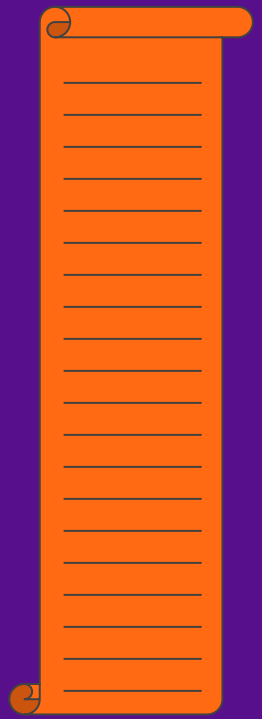
Data dictionary

Data Dictionary for COMIRB protocol #_____.
NB: Missing numeric data are coded -999
Values below detection limits are coded -666
Missing categorical data are coded 6=UNK 7=Refused

Variable name	units	format	Description & Additional info (assay/machine/algorithm)
ID			Unique patient identifier
age	years		Age at date of consent
sex		1=male 2=female	
race		1=White 2=Black 3=Asian 4=Pacific Islander 5=Mixed Race 6=UNK 7=Refused	
ethnicity		1=Hispanic 0=Non-Hispanic 6=UNK 7=Refused	
HTX		0=no 1=yes 6 =UNK	Hypertension indicator
BMD_hip	g/cm^2		Hologic, total hip
BMD_troch	g/cm^2		Hologic, trochanter subregion
BMD_LS	g/cm^2		Hologic, L2-L4
E2	pmol/L		E_2 (estradiol), Diagnostic Systems Laboratories (DSL)

EXERCISE 5

Building a data dictionary



Data dictionary: Tasks

Create a **data dictionary** to help clarify the variables collected in the **paper form**

Include corrections you made

Variable Name	Units	Format	Description
ID			Unique Patient Identifier
Age	years		Age at date of consent
Sex		1 = male 2 = female 3 = other	
Race		1 = White 2 = Black 3 = Asian 4 = Pacific Islander 5 = Mixed race 6 = UNK 7 = Refused	

Data dictionary

NYU HEALTH SCIENCES LIBRARY

Variable	Type	Definition	Values
DOB	Date	The date of birth of the participant (Mother)	YYYY-MM-DD
ParticipantFirstName	Text	The first name of the participant	N/A
ParticipantLastName	Text	The last name of the participant	N/A
Height	Numeric	Height of child measured in feet	N/A
Weight	Numeric	Weight of child measured in lbs	N/A
SysBloodPressure	Numeric	Systolic blood pressure of child	N/A
DiastBloodPressure	Numeric	Diastolic blood pressure of child	N/A
MotherAge	Numeric	Age of mother in years	N/A
LongestPeriodAbstinence	Numeric	The longest period of abstinence of the participant measured in days (up to 7), weeks (up to 4), months (up to 12), and years (no limit)	Days (max 7) Weeks (max 4) Months (max 12) Years (no limit)
LongestPeriodDrinking	Numeric	The longest period of uninterrupted drinking of the mother measured in days (up to 7), weeks (up to 4), months (up to 12), and years (no limit)	Days (max 7) Weeks (max 4) Months (max 12) Years (no limit)
DepressBotheredByThings	Numeric	Center for Epidemiologic Studies: Depression: The mother was bothered by things that don't usually bother her	1. Rarely 2. Sometimes 3. Occasionally 4. Most of the time

Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. **Standards**
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

What are data standards?

Standards provide guidance to research communities on:

What to collect (**guideline**)

How to represent what is collected (**terminology**)

How to encode data for transmission (**data model**)

Why use standards?

Guidance on recording/collecting data or metadata

Provide a common language for researchers

Facilitates data interoperability

Why use standards?

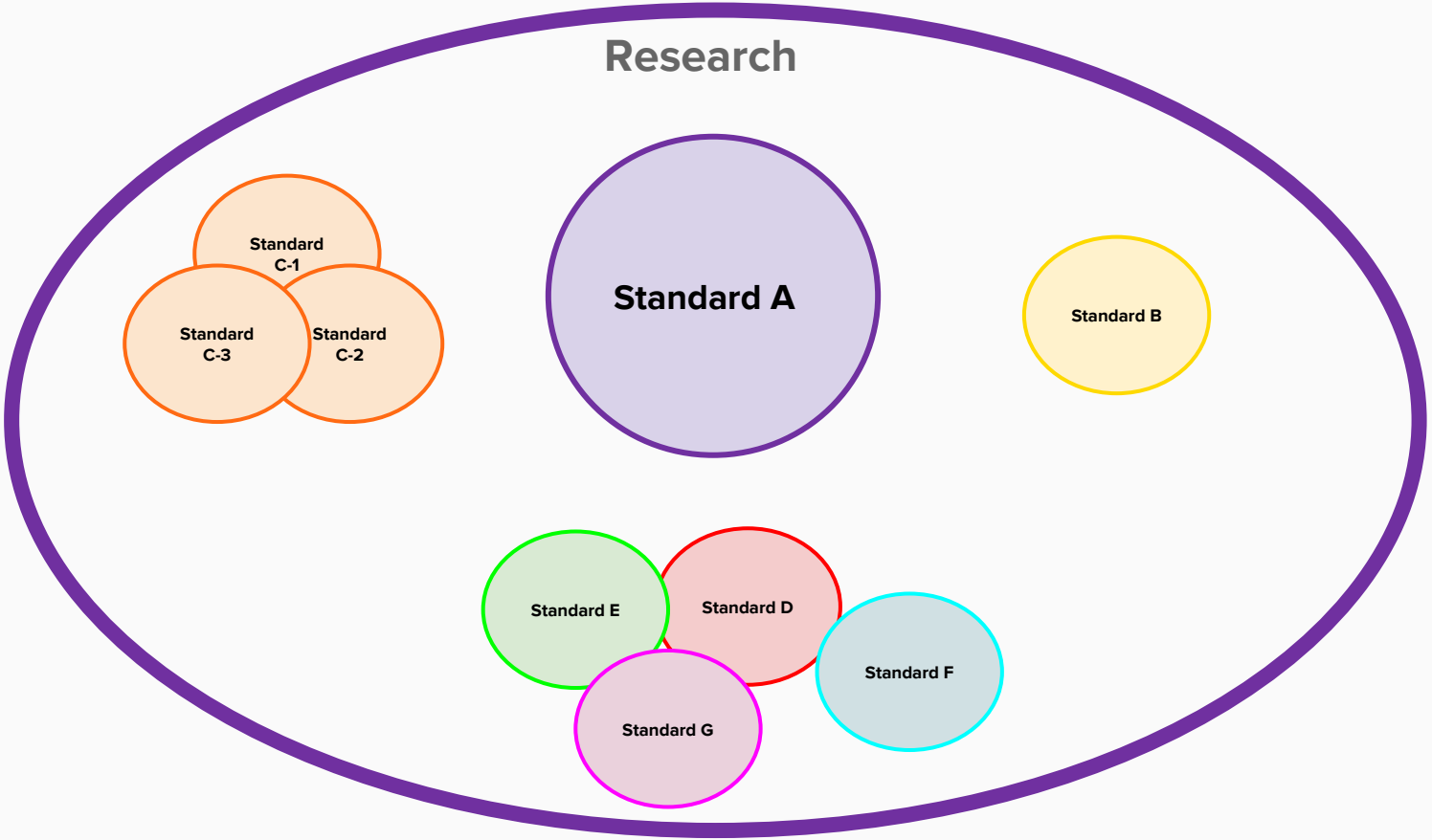
Guidance on recording/collecting data or metadata

Provide a common language for researchers

Facilitates data interoperability

CAUTION: Benefits depend on standard's level of adoption

What do standards cover?



Standards can be Broad or Narrow

NYU HEALTH SCIENCES LIBRARY

In Vivo Experiments: ARRIVE Standard

Example:

Experimental Outcomes

Spinal Cord Injury Experiments MIASE Standard

Example:

Electroporation Device Name

Standards can be Broad or Narrow

NIH: Common Data Elements

NYU HEALTH SCIENCES LIBRARY

NIH Common Data Element (CDE) Resource Portal

Home | Resource Summaries | Glossary

Home

Summary Table for NIH CDE Initiatives

This table lists summary information for NIH CDE Initiatives. More information on NIH CDE Initiatives: Subject Areas, Detailed Summaries.

Show 50 entries

Search:

Link to Homepage	Link to CDEs	Brief Summary	Subject Area	Number of Elements	CDE Resource Contact
Standardized Asthma Outcomes for Clinical Research	Asthma CDEs	The standardized asthma outcomes for clinical research represent recommendations for core (required in future studies), supplemental (to be used according to study aims), and emerging (requiring validation and standardization) outcomes for 7 domains of asthma clinical research outcome measures. More...	Asthma. More...	10 (adults), 25 (children)	NHLBI, NIAID
Chronic Low Back Pain CDEs	cLBP	Recommended minimum dataset for research on chronic low back pain. More...	Chronic low back pain. More...	40	NCCAM
Early Detection Research Program	EDRN	CDEs for use in describing samples and data collected as part of cancer biomarker research. More...	Cancer. More...	1,600	NCI
National Ophthalmic Disease Genotyping Network	eyeGENE	As part of eyeGENE, common data elements have been developed for collecting phenotypic data associated with more than 30 inherited ophthalmic diseases. More...	Ophthalmology. More...	300+	NEI
Global Rare Diseases Patient Registry and Data Repository	GRDR	CDEs to facilitate standardized data collection into the GRDR and to assist organizations in establishing rare disease registries that contribute information to GRDR. More...	Rare diseases. More...	70	ORDR

https://www.nlm.nih.gov/cde/summary_table_1.html

NIH: Common Data Elements

NYU HEALTH SCIENCES LIBRARY

CDEs | Forms | Boards | Quick Board (0) | Help | Log In

Search | Pin All | Hide Filters | Table View | Export

120 results for **All Terms** | NIDA | All Topics | All Statuses (0.03 secs)

Filter by:

Classification
- NIDA (120)
  - Clinical Research (120)
  - Electronic Health Records (31)

Registration Status
- Standard (4)
- Qualified (116)

Person Weight Value Qualified

Matched by: **Classification**

The number that describes the vertical force exerted by the mass of an individual as a result of gravity.

Used By: NIDA NCI
Steward: NCI
Source: caDSR

Gender Code Qualified

Matched by: **Classification**

The code representing the gender of a person.

Used By: NIDA NINR
Steward: NCI
Source: caDSR

Value	Code Name	Code
0	Unknown	C17998
1	Male	C20197
2	Female	C16576
9	Not specified	C38046

Birth Date Qualified

Matched by: **Classification**

Person's Birthdate

Used By: NIDA
Steward: NCI
Source: caDSR

Patient Age Year Count Qualified

Matched by: **Classification**

the patient's age in number of years.

Used By: NIDA
Steward: NCI

https://cde.nlm.nih.gov/cde/search?selectedOrg=NIDA

FAIRsharing.org

NYU HEALTH SCIENCES LIBRARY

FAIRsharing.org
standards, databases, policies

Standards | Databases | Policies | Collections | Add/Claim Content | Stats | Log in or Register

FAIRsharing is here! From our first incarnation, BioSharing.org, which focussed on the life sciences, we are growing into FAIRsharing.org, to serve users across all disciplines and support Findable, Accessible, Interoperable and Reusable (FAIR) data.

A curated, informative and educational resource on data and metadata *standards*, inter-related to *databases* and data *policies*.

Find

Recommendations

Standards and/or databases recommended by journal or funder data policies.

Discover

Collections

Standards and/or databases grouped by domain, species or organization.

Learn

Educational

About standards, their use in databases and policies, and how we can help you.

Search FAIRsharing | Search

☑ Standards ☑ Databases ☑ Policies ☑ Collections/Recommendations

Advanced Search

Fine grained control over your search.

Search Wizard

ask FAIRsharing

Let us guide you to your results.

Terminology artifacts

NYU HEALTH SCIENCES LIBRARY

MeSH:

Investigative Techniques

- Accelerometry
  - Actigraphy
- Airway Extubation
- Animal Experimentation
  - Animal Use Alternatives +
  - Rotarod Performance Test
  - Vivisection
- Animal Identification Systems
- Anthropometry
  - Body Weights and Measures +
  - Cephalometry
  - Odontometry +
  - Pelvimetry
- Artifacts
- Autoanalysis
- Automation, Laboratory
- Autopsy
- Biological Assay
  - Limulus Test
- Biomedical Enhancement
  - Genetic Enhancement
- Bioprinting
- Bioprospecting
- Bone Demineralization Technique
- Catheterization
  - Angioplasty +
  - Balloon Embolectomy
  - Balloon Occlusion +
  - Balloon Valvuloplasty
  - Cardiac Catheterization +
  - Catheterization, Central Venous
  - Catheterization, Peripheral +
  - Urinary Catheterization +

Terminology artifacts

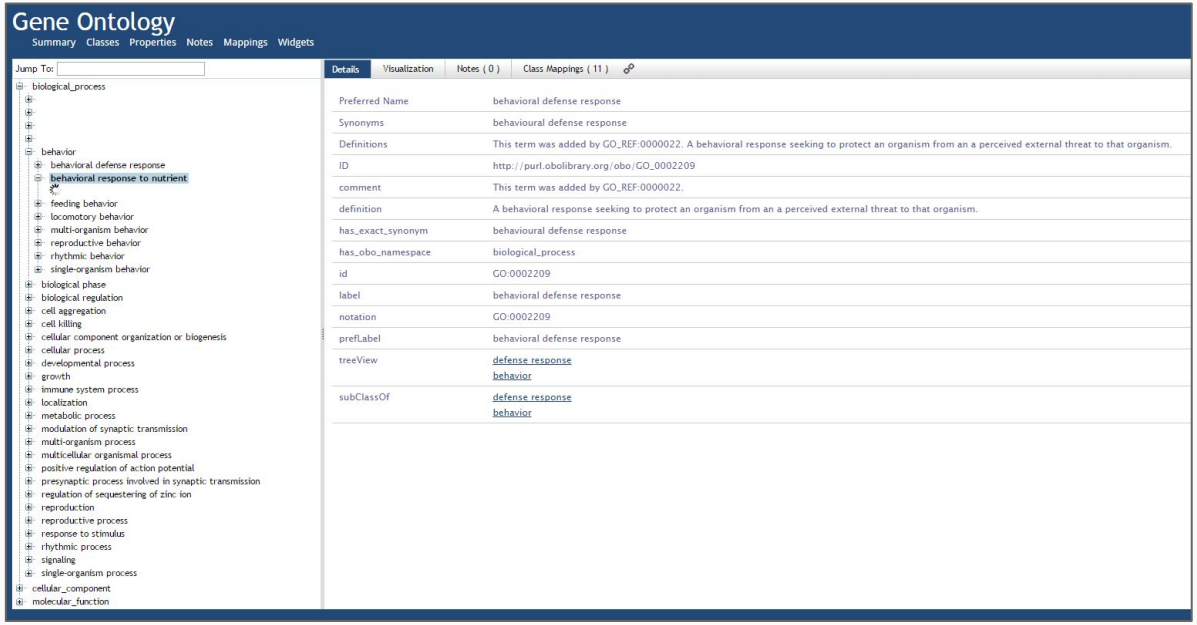
NYU HEALTH SCIENCES LIBRARY

MeSH:

Investigative Techniques

- Accelerometry
  - Actigraphy
- Airway Extubation
- Animal Experimentation
  - Animal Use Alternatives +
  - Rotarod Performance Test
  - Vivisection
- Animal Identification Systems
- Anthropometry
  - Body Weights and Measures +
  - Cephalometry
  - Odontometry +
  - Pelvimetry
- Artifacts
- Autoanalysis
- Automation, Laboratory
- Autopsy
- Biological Assay
  - Limulus Test
- Biomedical Enhancement
  - Genetic Enhancement
- Bioprinting
- Bioprospecting
- Bone Demineralization Technique
- Catheterization
  - Angioplasty +
  - Balloon Embolectomy
  - Balloon Occlusion +
  - Balloon Valvuloplasty
  - Cardiac Catheterization +
  - Catheterization, Central Venous
  - Catheterization, Peripheral +
  - Urinary Catheterization +

Gene Ontology:



Reporting guidelines

NYU HEALTH SCIENCES LIBRARY

Clinical Data Acquisition Harmonization (CDASH):

5.0 CDASH Domain Tables

5.1 Common Identifier Variables

The following apply across all of the data collection domains.

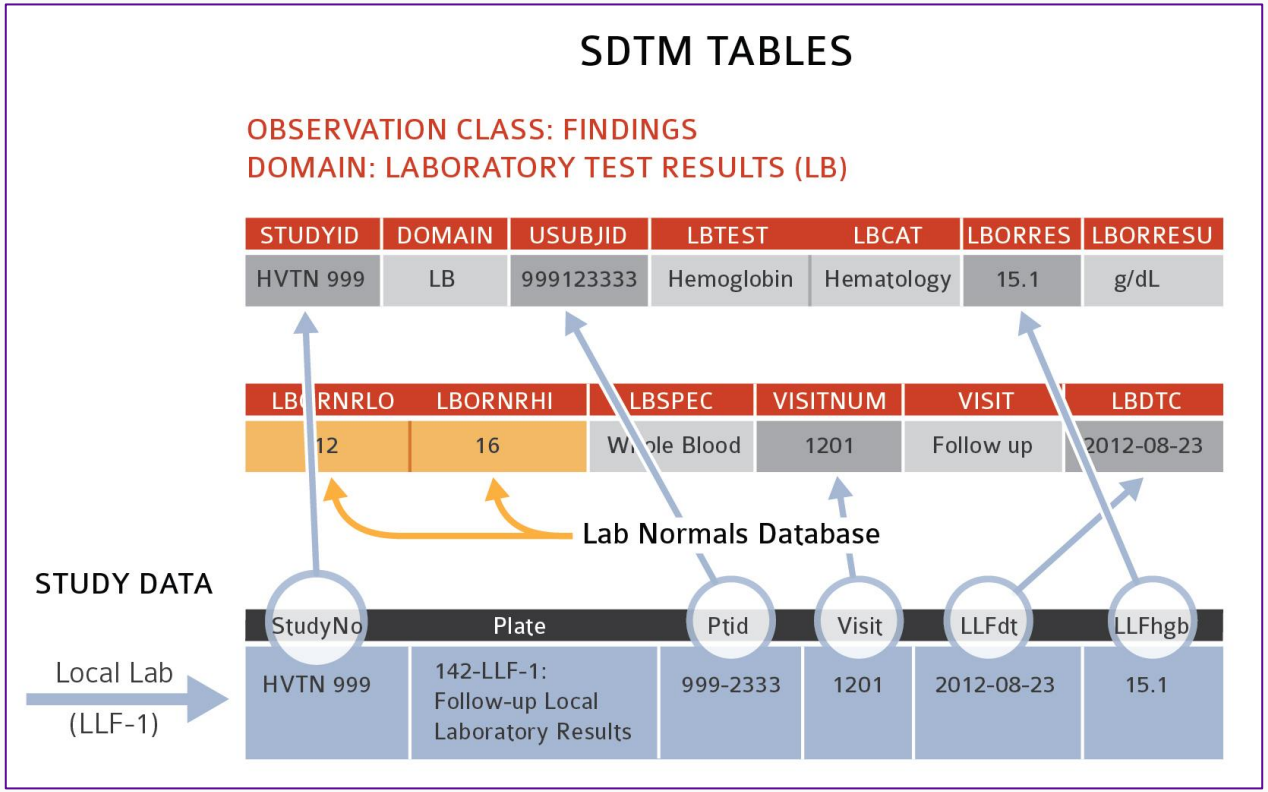
	Question Text	Prompt	SDTM or CDASH Variable Name	BRIDG	Definition	CRF Completion Instructions	Information for Sponsors	Core
1	What is the sponsor identifier?	Sponsor	SPONSOR	Organization.identifier*	A unique identifier for a study sponsor. An individual, company, institution, or organization that takes responsibility for the initiation and management of a clinical trial, although may or may not be the main funding organization. If there is also a secondary sponsor, this entity would be considered the primary sponsor. A corporation or agency whose employees conduct the investigation is considered a sponsor and the employees are considered investigators.	Not applicable	This is typically pre-printed. It may be used as an identifier in external data warehouses (e.g. Janus) and in electronic medical records or other partnerships for sharing data. This field does not map directly to an SDTM variable. *See the BRIDG model for complete path.	O
2	What is the study identifier?	Protocol *or* Study	STUDYID	DocumentIdentifier. identifier*	Unique identifier for a study.	Not applicable.	This is typically pre-printed/pre-populated. *See the BRIDG model for complete path.	HR

CDISC: Commonly Used Controlled Terminology

NYU HEALTH SCIENCES LIBRARY

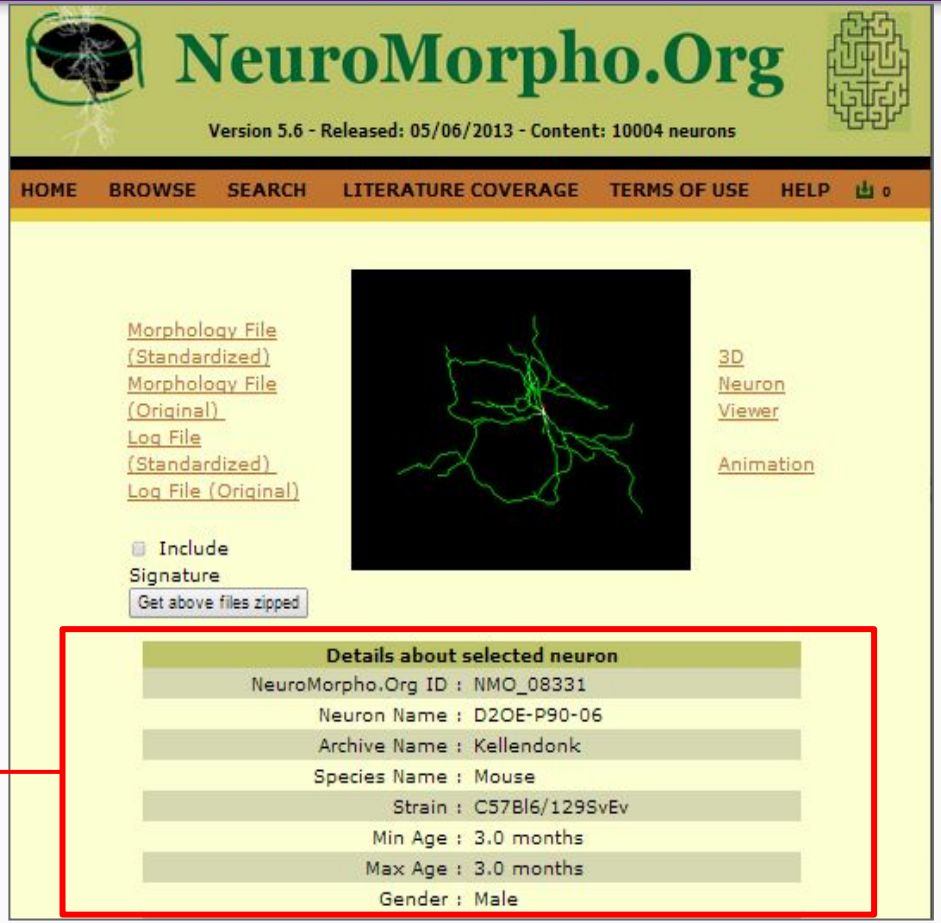
CDASH Data Collection Field	CDASH Definition	CDISC Approved Terminology Code list	Commonly Used Terms from the CDISC Terminology Code lists See code list for full list of values		
			Description	**CDASH Abbreviation**	**CDISC Submission Value**
			rectal	- - -	Rectal
DAORRESU	Unit of Drug Dispensed or Returned	Unit Code list C71620 Extensible	bag	- - -	BAG
			bottle	- - -	Bottle
			box	- - -	BOX
			capsule	cap	Capsule
			container	- - -	Container
			disk	- - -	Disk
			package	- - -	Package
			packet	- - -	Packet
			patch	- - -	Patch
			tablet	tab	Tablet
			tube	- - -	Tube
			vial	- - -	Vial
EGORRESU	ECG Original Units	Units Code list C71620 Extensible	millisecond	msec	msec
			second	sec	sec
			beats per minute	- - -	BEATS/MIN
EXDOSU	Units for Exposure	Unit Code list C71620 Extensible	tablet	tab	TABLET
			capsule	cap	CAPSULE
			puff	- - -	PUFF
			milliliter	mL	mL
			microgram	ug	ug
			milligram	mg	mg

CDISC: Standard Data Tabulation Model (SDTM)



https://hvtnews.files.wordpress.com/2013/10/5-7-originalsize.jpg

Standard?



Structured metadata

Structured metadata -- Neuromorpho

NYU HEALTH SCIENCES LIBRARY

- Neuromorpho ID (UID)
- Neuron Name
- Archive (researcher) name
- Species
- Strain of species
- Age range
- Gender
- Weight range
- Developmental stage
- Primary/Secondary/Tertiary brain regions
- Primary/Secondary/Tertiary Cell classes
- Original data format
- Experiment condition
- Experiment protocol
- Staining method
- Slicing direction / thickness
- Tissue shrinkage
- Objective type
- Magnification
- Reconstruction method
- Dates of deposition / upload
- Associated publications
- Web URL of archives (if available)
- Any additional information about the reconstruction

Structured metadata -- Neuromorpho

NYU HEALTH SCIENCES LIBRARY

- Neuromorpho ID (UID)
- Neuron Name
- Archive (researcher) name
- Species
- Strain of species
- Age range
- Gender
- Weight range
- Developmental stage
- Primary/Secondary/Tertiary brain regions
- Primary/Secondary/Tertiary Cell classes
- Original data format
- Experiment condition
- Experiment protocol
- Staining method
- Slicing direction / thickness
- Tissue shrinkage
- Objective type
- Magnification
- Reconstruction method
- Dates of deposition / upload
- Associated publications
- Web URL of archives (if available)
- Any additional information about the reconstruction

?

Questions

Course Schedule

NYU HEALTH SCIENCES LIBRARY

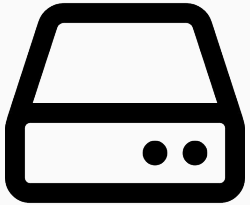
1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. **Storage and preservation**
9. Providing access to data
10. Strategies for implementing RDM
11. Wrap up

Storage Solutions

What is available at your institution?

What will provide researchers with **secure storage**?

What are your institution's **policies** around storage?

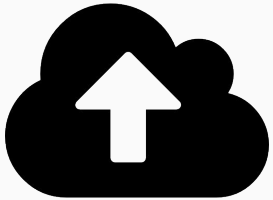


Storage Options

Proprietary cloud options -- what to look for:

Data ownership policies

Picking **>1 cloud option**



Where is data stored during different stages of the workflow?

Where is data stored during different stages of the workflow?

Backup Plan

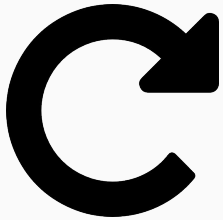
How often will data be backed up?

Who is responsible for the data?

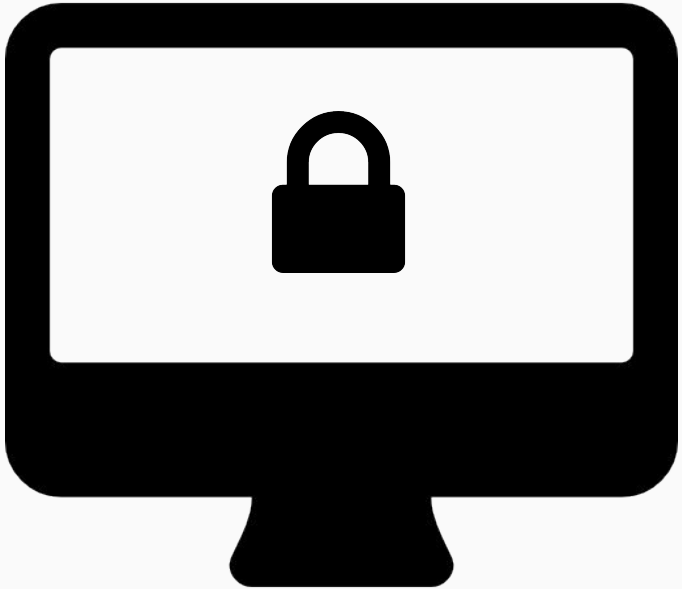
How many copies will be made?

Where are those copies stored?

How will data be dispersed geographically?

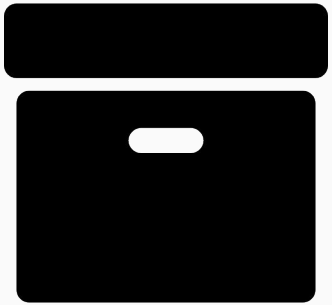


Data Security

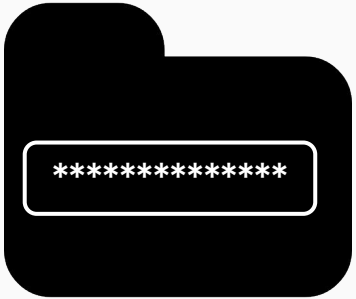


Data Security: Extra steps

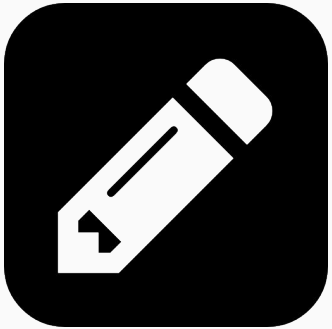
Lock machines



Password protect files



Use agreements



storage ≠ preservation

Preservation: Hardware obsolescence

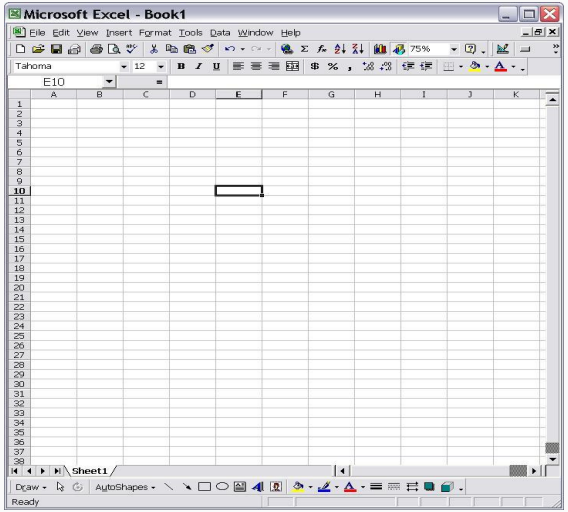
Preservation: Software obsolescence

Preservation: Data formats

Collection

VS

Dissemination

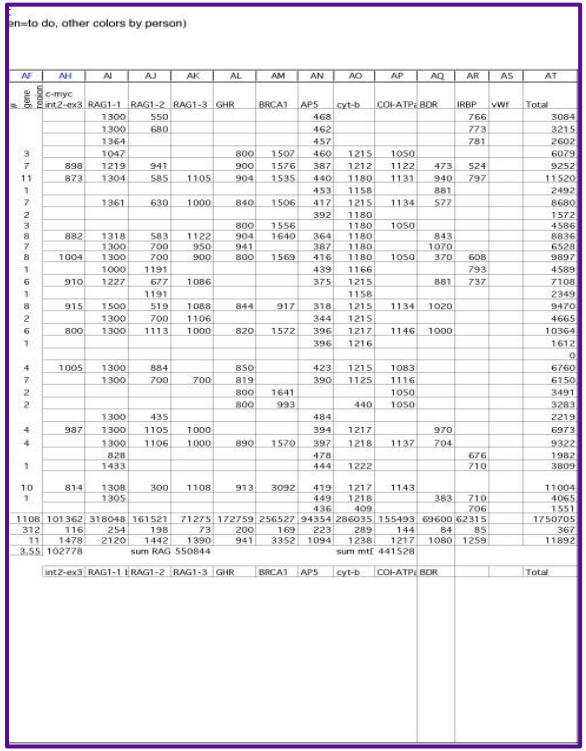


Preservation: Open data formats

Preservation: Open data formats aren't always realistic…

NYU HEALTH SCIENCES LIBRARY

Microsoft Excel

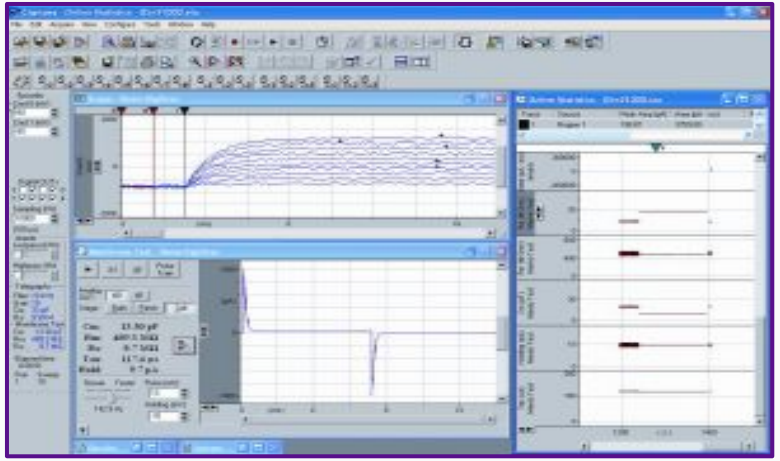


on=to do, other colors by person)

AF	AH	AI	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT
gene	c-myc int2-ex3	RAG1-1	RAG1-2	RAG1-3	GHR	BRCA1	AP5	cyt-b	COI-ATP	BDR	IRBP	vWF	Total
		1300	550				468				766		3084
		1300	680				462				773		3215
		1364					457				781		2602
3		1047			800	1507	460	1215	1050				6079
7	898	1219	941		900	1576	387	1212	1122	473	524		9252
11	873	1304	585	1105	904	1535	440	1180	1131	940	797		11520
1							453	1158		881			2492
7		1361	630	1000	840	1506	417	1215	1134	577			8680
2							392	1180					1572
3					800	1556		1180	1050				4586
8	882	1318	583	1122	904	1640	364	1180		843			8836
7		1300	700	950	941		387	1180		1070			6528
8	1004	1300	700	900	800	1569	416	1180	1050	370	608		9897
1		1000	1191				439	1166			793		4589
6	910	1227	677	1086			375	1215		881	737		7108
1			1191					1158					2349
8	915	1500	519	1088	844	917	318	1215	1134	1020			9470
2		1300	700	1106			344	1215					4665
6	800	1300	1113	1000	820	1572	396	1217	1146	1000			10364
1							396	1216					1612
													0
4	1005	1300	884		850		423	1215	1083				6760
7		1300	700	700	819		390	1125	1116				6150
2					800	1641			1050				3491
2					800	993		440	1050				3283
		1300	435				484						2219
4	987	1300	1105	1000			394	1217		970			6973
4		1300	1106	1000	890	1570	397	1218	1137	704			9322
		828					478				676		1982
1		1433					444	1222			710		3809
10	814	1308	300	1108	913	3092	419	1217	1143				11004
1		1305					449	1218		383	710		4065
							436	409			706		1551
1108	101362	318048	161521	71275	172759	256527	94354	286035	155493	69600	62315		1750705
312	116	254	198	73	200	169	223	289	144	84	85		367
11	1478	2120	1442	1390	941	3352	1094	1238	1217	1080	1259		11892
3.55	102778		sum RAG	550844				sum mtE	441528				
	int2-ex3	RAG1-1	RAG1-2	RAG1-3	GHR	BRCA1	AP5	cyt-b	COI-ATP	BDR			Total

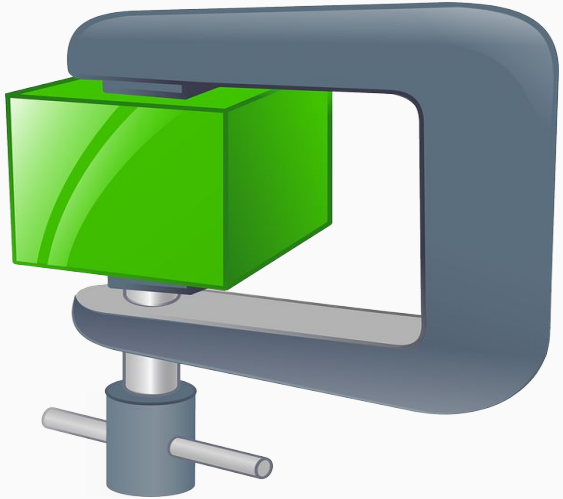
≠

Molecular Devices pClamp Software



If data is irreplaceable...

Preservation: Encryption & Compression



Preservation: Data ownership

Researcher's can't assume they own their data

All should review:

- **Funder** policies on data ownership
- **Institution** policies on data ownership

Review with office of research or intellectual property

Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. **Providing access to data**
10. Strategies for implementing RDM
11. Wrap up

NIH Data Sharing Repositories

NYU HEALTH SCIENCES LIBRARY

NIH Data Sharing Repositories

This table lists NIH-supported data repositories that accept submissions of appropriate data from NIH-funded investigators (and others). Also included are resources that aggregate information about biomedical data and information sharing systems. The table can be sorted according by name and by NIH Institute or Center and may be searched using keywords so that you can find repositories more relevant to your data. Links are provided to information about submitting data to and accessing data from the listed repositories. Additional information about the repositories and points-of-contact for further information or inquiries can be found on the websites of the individual repositories.

Show 50 entries Search:

IC	Repository Name	Repository Description	Data Submission Policy	Access to Data
NCI	The Cancer Imaging Archive (TCIA)	The Cancer Imaging Archive (TCIA) is a large archive of medical images of cancer accessible for public download. All images are stored in DICOM file format. The images are organized as "Collections", typically patients related by a common disease (e.g. lung cancer), image modality (MRI, CT, etc) or research focus.	How to Submit Data to TCIA	How to Access TCIA Data
NCI (NHGRI, NIGMS)	PeptideAtlas	PeptideAtlas is a multi-organism, publicly accessible compendium of peptides identified in a large set of tandem mass spectrometry proteomics experiments. Mass spectrometer output files are collected for human, mouse, yeast, and several other organisms, and searched using the latest search engines and protein sequences.	How to Submit Data to PeptideAtlas	How to Access PeptideAtlas Data
NHGRI	FlyBase: A Drosophila Genomic and Genetic Database	Drosophila Genomic and Genetic database that includes proteomics data, microarrays and Tiling BAC's.	How to Submit Data to FlyBase	How to Access FlyBase Data
NHGRI	The Zebrafish Model Organism Database (ZFIN)	ZFIN serves as the zebrafish model organism database. It aims to: a) be the community database resource for the laboratory use of zebrafish, b) develop and support integrated zebrafish genetic, genomic and developmental information, c) maintain the definitive reference data sets of zebrafish research information, d) to link this information extensively to corresponding data in other model organism and human databases, e) facilitate the use of zebrafish as a model for human biology, and f) serve the needs of the research community.	How to Submit Data to ZFIN	How to Access ZFIN Data
NHGRI	WormBase	WormBase is an international consortium of biologists and computer scientists dedicated to providing the research community with accurate, current, accessible information concerning the genetics, genomics and biology of C. elegans and related nematodes.	How to Submit Data to WormBase	How to Access WormBase Data

Research Data Repositories

NYU HEALTH SCIENCES LIBRARY

re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

Home | Search | Browse | Suggest | FAQ | About | Schema | Contact | Imprint

Search for Repositories (1132 Reviewed Repositories)

Search

Subject: Add subjects — Basic Biological and Medical Rese...
Content Type: Add content types
Country (of the responsible institutions): Add countries

Certificates | Open Access | Persistent Identifier

Clear

ICPSR
Inter-university Consortium for Political and Social Research
A partner in social science research

Find & Analyze Data — quick search for data — GO
Membership in ICPSR
Deposit Data
ICPSR Summer Program
Teaching & Learning with ICPSR
Data Management & Curation

Mission
ICPSR advances and expands social and behavioral research, acting as a global leader in data stewardship and providing rich data resources and responsive educational opportunities.

News
NCAA's newest newest Division I Academic Progress Rate (APR) data will be released this spring
Synergies for Europe's Research Infrastructures in the Social Sciences (SERISS) announces course on "Designing questionnaires for cross-cultural surveys"
New Releases through 2016-04-17

Website Search — GO — search for data

DSA 2014–2015 | ICSU WORLD DATA SYSTEM

*Open*DOAR — *Directory of Open Access Repositories*
Home | Find | Suggest | Tools | FAQ | About | Contact Us

The Directory of Open Access Repositories - *Open*DOAR

Search for repositories | Search repository contents | List of repositories | Repository Statistics

*Open*DOAR is an authoritative directory of academic open access repositories. Each *Open*DOAR repository has been visited by project staff to check the information that is recorded here. This in-depth approach does not rely on automated analysis and gives a quality-controlled list of repositories.

OpenDOAR has over 2600 listings!

DRYAD — About | For researchers | For organizations | Contact us | Log in | Sign up

DRYAD LAB
DryadLab is a collection of free, openly-licensed, high-quality, hands-on, educational modules for students to engage in scientific inquiry using real data.
Learn More »

Submit data now
How and why?

Search for data
Enter keyword, author, title, DOI, etc — Go
Advanced search

Browse for data
Recently published | Popular | By author | By journal

Recently published data

Beijbom O, Treibitz T, Kline DI, Eyal G, Khen A, Neal B, Loya Y, Mitchell BG, Kriegman D (2016) Data from: Improving automated annotation of benthic survey images using wide-band fluorescence. *Scientific Reports* http://dx.doi.org/10.5061/dryad.t4362

May MR, Hoehna S, Moore BR (2016) Data from: A Bayesian approach for detecting the impact of mass-extinction events on molecular phylogenies when rates of lineage diversification may vary. *Methods in Ecology and Evolution* http://dx.doi.org/10.5061/dryad.v16ns

Muylaert RL, Stevens RD, Ribeiro MC (2016) Data from: Threshold effect of habitat loss on bat richness in cerrado-forest landscapes. *Ecological Applications*

Latest from @datadryad

Tweets by @datadryad
Dryad Retweeted — Nathan Whelan @pcdarfke: I also agree that large datasets should be in repositories like @figshare, @datadryad, or NCBI. Not in a journal's online supplement/website
Dryad @datadryad: Ecology and Evolution in an Open World (or: why supplementary data are evil) onlinelibrary.wiley.com/doi/10.1002/ec...

Research Data Repositories: Figshare

NYU HEALTH SCIENCES LIBRARY

figshare | My data | search figshare (titles, tags, authors, etc.) | Browse | Upload | K. Read

My data | Projects | Activity

0% of private storage used

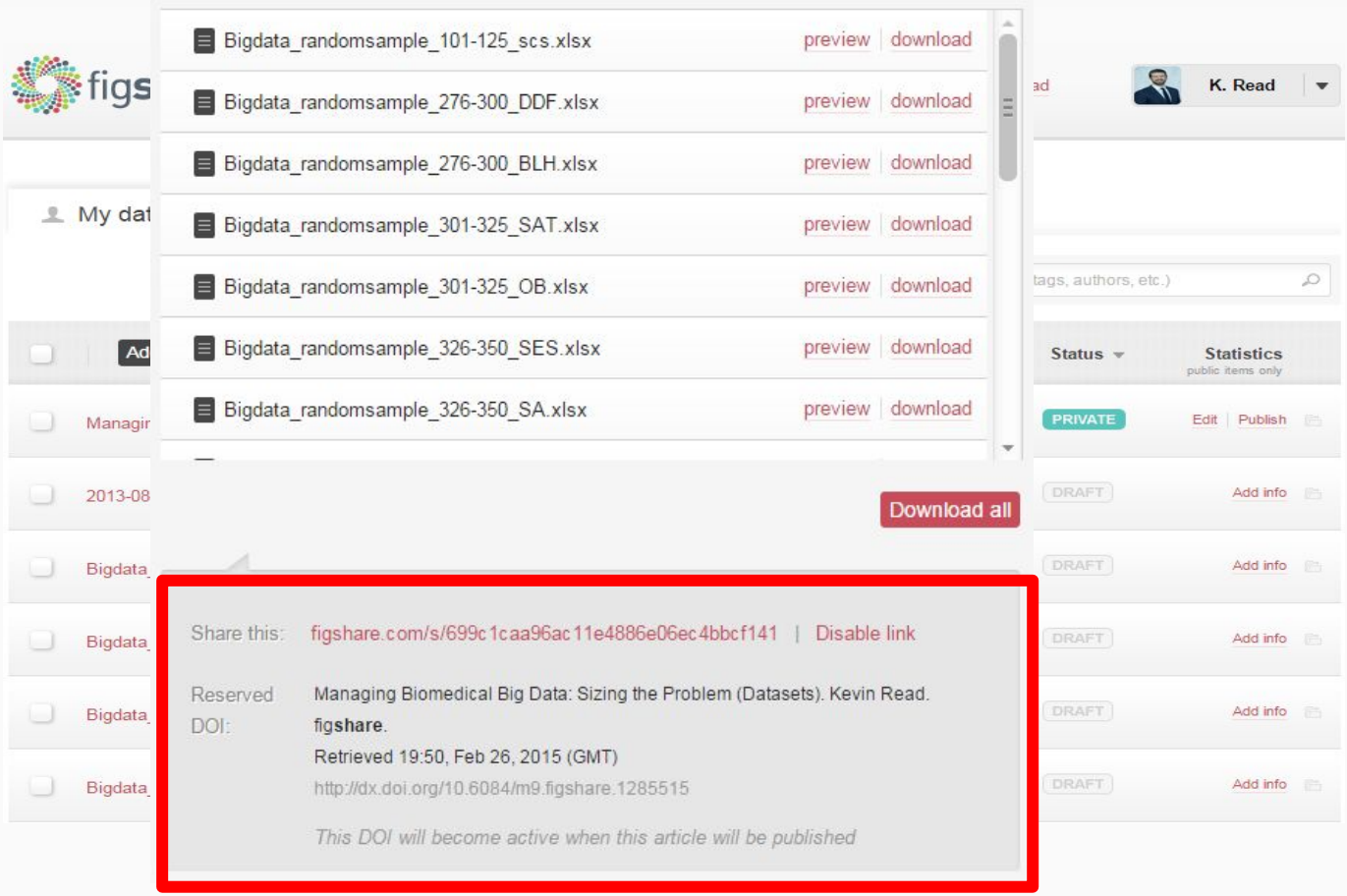
search my data (titles, tags, authors, etc.)

Add to Fileset | Batch edit

	Type (mouseover)	Date	Status	Statistics (public items only)
Managing Biomedical Big Data: Sizing the Problem (Datasets)	FILESET (20)	07.01.2015 20:27	PRIVATE	Edit | Publish
2013-08-07_Bigdatastudy_dataanalysis.xlsx	DATASET	07.01.2015 20:27	DRAFT	Add info
Bigdata_randomsample_351-375_TE.xlsx	DATASET	07.01.2015 20:27	DRAFT	Add info
Bigdata_randomsample_351-375_PML.xlsx	DATASET	07.01.2015 20:27	DRAFT	Add info
Bigdata_randomsample_326-350_SA.xlsx	DATASET	07.01.2015 20:26	DRAFT	Add info
Bigdata_randomsample_326-350_SES.xlsx	DATASET	07.01.2015 20:26	DRAFT	Add info

Research Data Repositories: Figshare

NYU HEALTH SCIENCES LIBRARY



Research Data Publications

NYU HEALTH SCIENCES LIBRARY

SCIENTIFIC DATA

Home | About | For Authors | For Referees | Advisory and Editorial Board | Open Access | FAQ

Sign up for Scientific Date e-alert Facebook Twitter

Submit to *Scientific Data* in three simple steps:

1. DESCRIBE
Write a detailed description of your dataset. We have templates to help you and a detailed guide to authors.

2. DEPOSIT
See our list of recommended repositories. We will help you find the right place for your data.

3. SUBMIT
Submit online and get the credit you deserve for your data!

Get credit where credit's due and share your data.

Sample Data Descriptors

Proteomic profiles of human embryonic stem cells, induced-pluripotent stem cells and

Sequencing of genomes, transcriptomes and methylomes of wild *Arabidopsis thaliana*

Log on BioMed Central

(GIGA)n SCIENCE

Search GigaScience

Home | Articles | Authors | Reviewers | About this journal | My GigaScience

Access vs *meaningful* access

Meaningful access means…

1. Depositing data in a location where other researchers will find it
2. Providing well-documented data
3. Using standards where possible to make data interoperable

Data Curation Repository Checklist

1. Is the repository reputable?
2. Will it accept the data a researcher wants to deposit?
3. Will the data be kept safe legally?
4. Will the repository sustain the value of the data?
5. Will the repository support analysis of data use, and track usage?

http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data#5

Discovery vs Preservation

?

Questions

EXERCISE 6

Sharing data

Sharing data: Tasks

Where could the researchers of the paper share their data?

What are some of the concerns with sharing data of this kind?

Why did you choose one option over another?

Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. **Strategies for implementing RDM**
11. Wrap up


OPEN
for business

Understand your environment

Identify institutional gaps

Identify institutional gaps

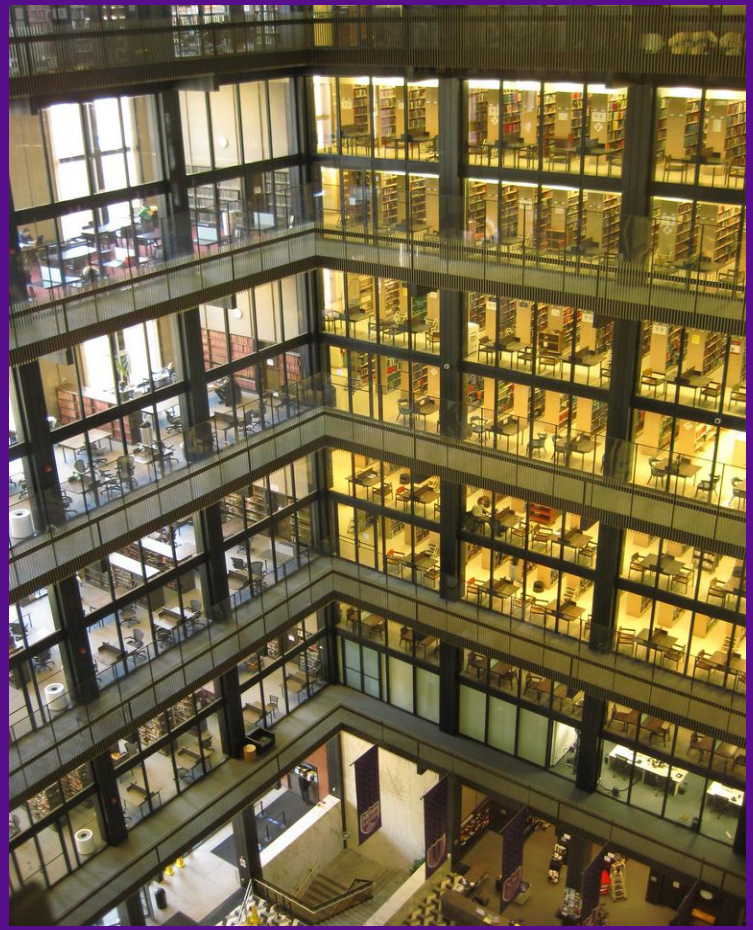
↓

Avoid turf wars

Seek out partnerships for complementary skills

Seek out partnerships for branding

Affiliated Libraries



Your IT Department

NYU HEALTH SCIENCES LIBRARY

NYU HEALTH SCIENCES LIBRARY

Office of Science & Research

Partnerships: Other Potential options

Clinical and Translational Science Awards

Office of Scholarly Communication

Institutional review board

Postdoctoral/Graduate Student Offices

Specific research departments

Partnerships: Other Potential options

Clinical and Translational Science Awards

Office of Scholarly Communication

Institutional review board

Postdoctoral/Graduate Student Offices

Specific research departments

Use data interviews to get started

Find opportunities to talk about data

Find bout

OSR Weekly Announcements - Week of 3/17/2014

OSR Broadcast Administrator [OSRBroadcastAdministrator@nyu...

To: Hanson, Karen

Tuesday, March 18, 2014 10:3

PLOS Announces New Data Sharing Mandate

On February 24, 2014, PLOS announced a more stringent data sharing policy, where researchers are now **required** to share data, related metadata and methods relating to the article: *"PLOS journals require authors to make all data underlying the findings described in their manuscript fully available without restriction, with rare exception."* To be accepted for publication, you must share your data in one of the following ways:

1. Provide a DOI or accession number for data from a publicly accessible data repository
2. Include smaller datasets within the supplementary files
3. Make sensitive data available on request (through a third party)

Note: In no case is it acceptable for the data to be available solely through the author(s).
If you are publishing in a PLOS journal, here are some options for data sharing that will meet the new requirements:

1. Faculty digital archive
   - a service provided to faculty by NYU Division of Libraries and IT Services
   - supports public sharing of data sets in many formats
   - supplies a unique digital identifier for all content

Find bout

OSR Weekly Announcements - Week of 3/17/2014

OSR Broadcast Administrator [OSRBroadcastAdministrator@nyu...

To: Hanson, Karen

Tuesday, March 18, 2014 10:3

PLOS Mandate

On February 24, 2014, P policy, where researchers are now **required** to share o the article: *"PLOS journals require authors to make heir manuscript fully available without restriction, with on, you must share your data* in one of the following

1. Provide a DOI o ccessible data repository
2. Include smaller
3. Make sensitive y)

Note: In no case is it acc ough the author(s).
If you are publishing in a sharing that will meet the new requirements:

1. Faculty digital archive
   - a service provided to faculty by NYU Division of Libraries and IT Services
   - supports public sharing of data sets in many formats
   - supplies a unique digital identifier for all content

National Institutes of Health

Opportunities: Electronic Lab notebooks

GitHub

eCAT

Scalability is crucial

Avoid library jargon

Avoid library jargon

↓

Metadata

Controlled Vocabulary

Avoid library jargon

~~Metadata~~

~~Controlled Vocabulary~~

Education:

Don’t reinvent the wheel

Education: Don’t reinvent the wheel

Preparing Librarians to Learn and Teach RDM

http://compass.iime.cloud/mix/G3X5E/

RDM Teaching Toolkit:

https://figshare.com/articles/Research_Data_Management_Teaching_Toolkit/5042998

Hands on RDM Course:

https://osf.io/fms4u/

Retraction Watch:

http://retractionwatch.com/

Retraction Watch

NEJM paper on sleep apnea retracted when original data can't be found

with 4 comments

The authors of a paper in the *New England Journal of Medicine* are retracting it, after being unable to find data supporting a table that required corrections.

The NEW ENGLAND JOURNAL of MEDICINE

Data Horror Stories:

https://pinboard.in/u:dsalo/t:horrorstories

dsalo + horrorstories 316

« earlier

Can I run old 16-bit programs like Superbase in Windows 10? | Technology | The Guardian
datacuration horrorstories software fileformats
3 days ago by dsalo copy to mine

Man accidentally 'deletes his entire company' with one line of bad code | News | Lifestyle | The Independent
our old friend rm -rf

WHERE WERE THE BACKUPS?!
horrorstories software datacuration
17 days ago by dsalo copy to mine

Outdated and Vulnerable WordPress and Drupal Versions May Have Contributed to the Panama Papers Breach – WordPress Tavern
security horrorstories recordsmgmt
25 days ago by dsalo copy to mine

Concerns about image manipulation? Sorry, the data were lost in a flood - Retraction Watch at Retraction Watch
datacuration horrorstories disasterplanning
4 weeks ago by dsalo copy to mine

Former VW employee says he was fired after questioning deletion of documents | Ars Technica
horrorstories recordsmgmt
6 weeks ago by dsalo copy to mine

Hospital pays $17k for ransomware crypto key | Ars Technica
351 644 ransomware horrorstories security
10 weeks ago by dsalo copy to mine

Out of a Rare Super Bowl I Recording, a Clash With the N.F.L. Unspools - The New York Times
video preservation horrorstories
12 weeks ago by dsalo copy to mine

404 Error: Why are Madison's open data and civic hacking communities almost dead? | Local News | host.madison.com
opendata horrorstories datacuration
january 2016 by dsalo copy to mine

Cartoons: PhD Comics

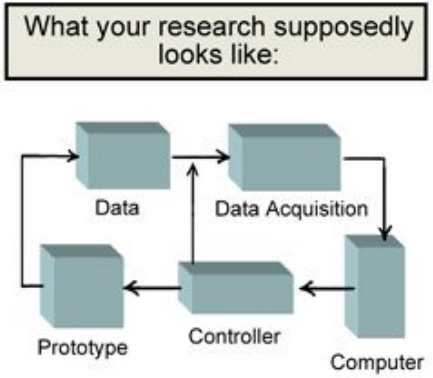


Figure 1. Experimental Diagram

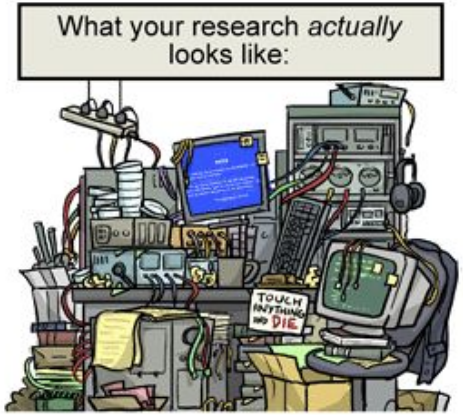


Figure 2. Experimental Mess

Videos:

Data Sharing and Management Snafu in 3 Short ... - YouTube
https://www.youtube.com/watch?v=N2zK3sAtr-4

?

Questions

EXERCISE 7

Planning your strategy

Your strategy: Tasks

Select one recorder from your table (use flip charts)

Think about service **you** can start when you get back from MLA

Consider:

1. One current or proposed service
2. Who can you partner with?
3. What actions do you need to take?
4. What challenges do you foresee?

NYU HEALTH SCIENCES LIBRARY

Who are the contacts at your institution related to managing research data? Who could partner with?

- Contacts for questions about retaining or destroying data
- Contacts for questions about sponsored research data
- Contacts for questions about de-identifying data
- Contacts for questions about storing, backing up, and securing data
- Contacts for questions about archiving and preserving data
- Contacts for questions about depositing data in a repository
- Contacts for questions about describing data
- Contacts for questions about sharing data
- Contacts for questions about licensing data
- Contacts for questions about data ownership
- Contacts for questions about analyzing data
- Contacts for questions about visualizing data

What local resources and tools related to data management at your institution?

- Data storage options
- Describing and annotating data tools
- e-Lab notebooks
- Data backup tools
- Resources and tools for sharing data
- Resources and tools for de-identifying data
- Tools for analyzing data
- Tools for visualizing data
- Resources and tools for publishing data
- Resources and tools for archiving and preserving data
- Resources and tools for citing and licensing data

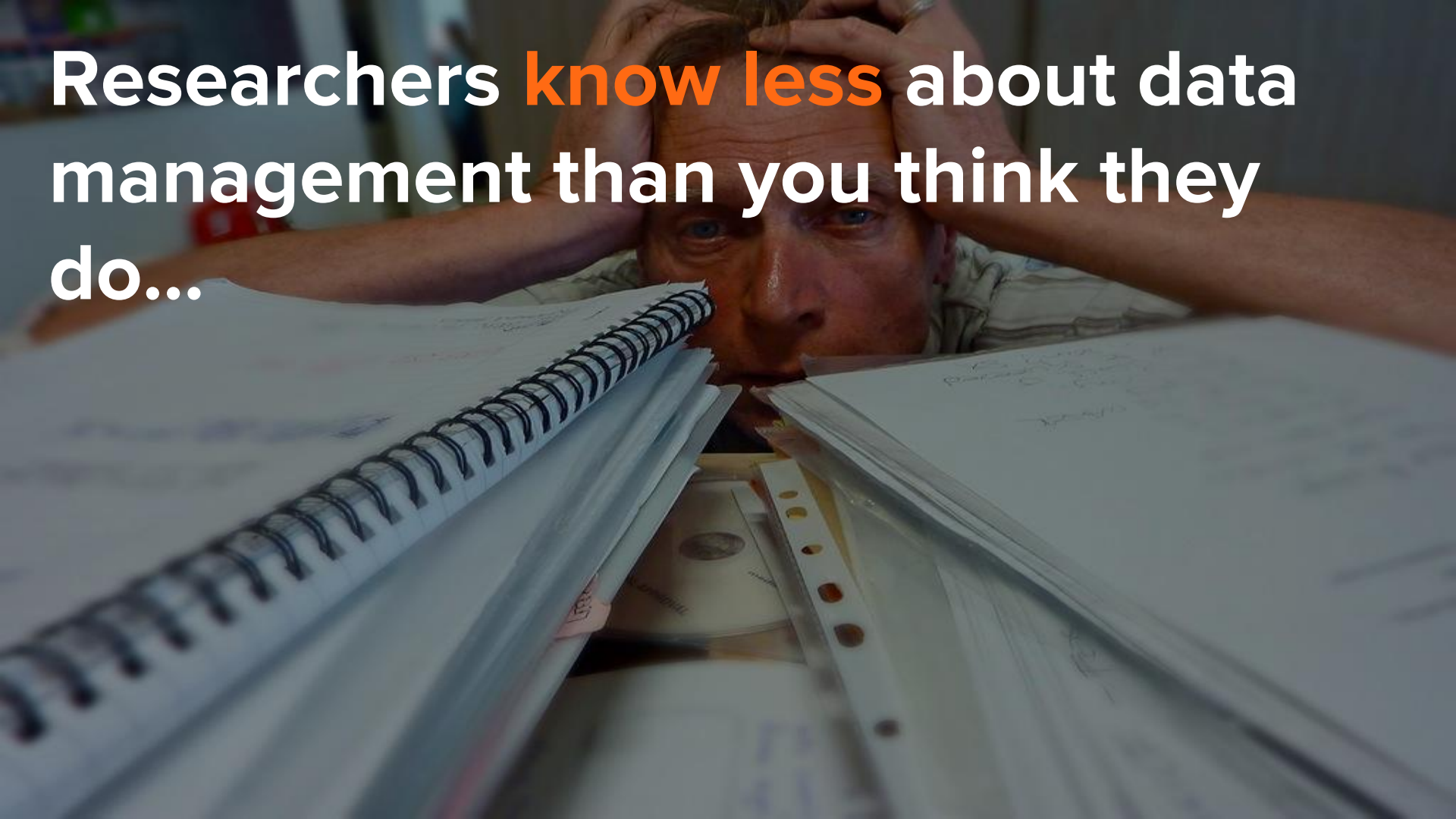
Creamer A. Customizing and Using the New England Collaborative Data Management Curriculum. 2014. E-Science Community Blog. Available from: http://esciencecommunity.umassmed.edu/2014/04/16/customizing-and-using-the-new-england-collaborative-data-management-curriculum/

Course Schedule

NYU HEALTH SCIENCES LIBRARY

1. Introduction
2. Current library roles in RDM
3. Story of data
4. Understanding your research community
5. RDM climate
6. Data documentation best practices
7. Standards
8. Storage and preservation
9. Providing access to data
10. Strategies for implementing RDM
11. **Wrap up**

Changing how people see the library isn’t easy!



Researchers know less about data management than you think they do...

There is a community out there

Handout

https://docs.google.com/presentation/d/1XdwDoVyM2bpCQRXIP-PI0TNJlkL6vpyypS1EiwbILbI/edit?usp=sharing

Research Data Management Resources & Tools

Created by Kevin Read, Alisa Surkis

Background Information

Data Management for Librarians
https://www.mendeley.com/groups/2956801/data-management-for-librarians/
Group of librarians sharing data literature

Escience Portal
http://esciencelibrary.umassmed.edu/data-management
Website collaborative group providing RDM resources

Digital Curation Centre How-to Guides
http://www.dcc.ac.uk/resources/how-guides
How-to gudies on a variety of RDM issues

Education

New England Collaborative Research Data Management Curriculum Case Studies
http://library.umassmed.edu/necdmc/research_cases

Digital Curation Centre Training
http://www.dcc.ac.uk/training

Data Management for Clinical Research
https://www.coursera.org/learn/clinical-data-management/home/info

Research Data Management and Sharing
https://www.coursera.org/learn/data-management/

Standards & Repositories

Re3data
http://www.re3data.org/
Registry of data repositories

NIH Data Sharing Repositories
https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
NIH-supported data repositories

Biosharing.org
https://biosharing.org/
Registry of biomedical data standards

NIH Common Data Element Resource Portal
https://www.nlm.nih.gov/cde/summary_table_1.html
NIH-supported common data elements

Standards & Repositories

NIH Data Sharing Policies
https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_policies.html

SPARC* Data Sharing Policies Comparison Tool
http://datasharing.sparcopen.org/

PLOS Data Sharing Policy
http://journals.plos.org/plosone/s/materials-and-software-sharing

Data Management Resources

Conducting Data Interviews
http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4511052/

Data Curation Profiles Toolkit
http://datacurationprofiles.org/

PhD Comics
http://phdcomics.com/comics.php

Data Horror Stories
https://pinboard.in/u:dsalo/t:horrorstories

Retraction Watch
http://retractionwatch.com/

Data Asset Framework
http://www.data-audit.eu/

Stay informed...

JISC RDM Listserv
https://www.jiscmail.ac.uk/cgi-bin/webadmin?A0-RESEARCH-DATAMAN

RDAP Listserv
http://mail.asis.org/mailman/listinfo/rdap

#datalibs on Twitter

?

Questions

NYU HEALTH SCIENCES LIBRARY

kevin.read@med.nyu.edu

alisa.surkis@med.nyu.edu

References

- U.S. Joint Chiefs of Staff JP2-0. “Relationship of Data, Information, and Intelligence”. https://commons.wikimedia.org/wiki/File:Relationship_of_data,_information_and_intelligence.png
- mrwynd. “USB”. https://www.flickr.com/photos/mrwynd/4938515289
- Steve Baker. “The messy desk syndrome open to the public”. https://www.flickr.com/photos/littlebiglens/8240611316
- Bryn Nelson. “Empty Archives”, September 10, 2009. Nature, v. 461, p.160. http://www.nature.com/news/2009/090909/pdf/461160a.pdf
- National Science Foundation. “Dissemination and Sharing of Research Results”. http://www.nsf.gov/bfa/dias/policy/dmp.jsp
- Font Awesome Cheatsheet. http://fortawesome.github.io/Font-Awesome/3.2.1/cheatsheet/
- NYU Health Sciences Library Data Management Guide. http://hslguides.med.nyu.edu/data_management
- Habib M’henni. “Wikimania 2012, Education I Session bis”. https://commons.wikimedia.org/wiki/File:Wikimania_2012,_Education_I_Session_bis.JPG
- NYU Health Sciences Library. https://hsl.med.nyu.edu/services
- Gary Wong. “Baiyoke Bangkok View”. https://www.flickr.com/photos/koolgary/2460635163/
- Thomas Splettstoesser. “RNA-codons-aminoacids”. https://commons.wikimedia.org/wiki/File:RNA-codons-aminoacids.svg
- https://pixabay.com/static/uploads/photo/2012/02/27/15/32/blood-17305_960_720.jpg
- Genesis 12. “MRI Head 5 Slices”. https://en.wikipedia.org/wiki/File:MRI_Head_5_slices.jpg
- J Biochemist. “Lab bench”. https://www.flickr.com/photos/proteinbiochemist/3167660996
- Jclam. “General workflow for Q-FISH with cultured cells”. https://commons.wikimedia.org/wiki/File:Q-FISH_workflow.png
- Kovah. “Angular JS Code I”. https://www.flickr.com/photos/kovah/15332449387
- Francisco Arevalo. “Rat”. https://thenounproject.com/search/?q=rat&i=15130
- Anthony Bossard. “Oscilloscope”. https://thenounproject.com/search/?q=oscilliscope&i=375822
- Alex Auda Samora. “Test Tubes”. https://thenounproject.com/search/?q=test+tube&i=98063
- ESO/G. Bono & CTIO. “Carina Dwarf Galaxy”. https://en.wikipedia.org/wiki/File:Carina_Dwarf_Galaxy.jpg
- NASA / Paul Riedel. “Atomic Laboratory Experiment on Atomic Materials”. https://commons.wikimedia.org/wiki/File:Atomic_Laboratory_Experiment_on_Atomic_Materials_-_GPN-2000-000663.jpg
- Chrislb. “MultiLayerNeuralNetwork”. https://commons.wikimedia.org/wiki/File:MultiLayerNeuralNetwork_english.png
- Arbeck. “Data Mining”. https://commons.wikimedia.org/wiki/File:Data_Mining.svg
- Local Studies NSW “Storytelling, Concord Library” https://www.flickr.com/photos/local_studies_nsw/6518792855
- iz west. “MRI”. https://www.flickr.com/photos/53133240@N00/7694882446

References

- Nathanael Burton. "Brain tumor MRI scans". https://www.flickr.com/photos/mathrock/7374758900
- Synaptidude “LTP exemplar” http://commons.wikimedia.org/wiki/File:LTP_exemplar.jpg
- Mccapdevila. “Current clamp recording of neuron” 2012. http://en.wikipedia.org/wiki/File:Current_Clamp_recording_of_Neuron.GIF
- Ling C, Hendrickson ML, Kalil RE (2012) Morphology, Classification, and Distribution of the Projection Neurons in the Dorsal Lateral Geniculate Nucleus of the Rat. PLoS ONE 7(11): e49161. doi:10.1371/journal.pone.0049161
- Uwe Kils. “Iceberg” http://commons.wikimedia.org/wiki/File:Iceberg.jpg
- Cow concept: Dorothea Salo, “Save the Cows”, 2009. http://www.slideshare.net/cavlec/save-the-cows-data-curation-for-the-rest-of-us-1533252
- AJ Yakstrangler. “Tithby” 2011. www.flickr.com/photos/yakstrangler/6030261340.
- BobPetUK. “Raw Minced Beef” 2010. www.flickr.com/photos/22179048@N05/5195112462/
- Like_the_Grand_Canyon. “McDonalds Hamburger Royal Bacon” 2008. www.flickr.com/photos/like_the_grand_canyon/3022123379
- Royal Siam Beauty. “Asian Female Scientist with Laboratory Test Tube of Green Solution.” https://www.flickr.com/photos/86552932@N03/9599521036
- CGUTREC. “IMG_4736”. https://www.flickr.com/photos/66019482@N07/6035342419
- Seattle Municipal Archives. “City water testing laboratory, 1948”. https://www.flickr.com/photos/seattlemunicipalarchives/3739366791
- Umass CVIP. “PVLSI Scientists at Work”. https://www.flickr.com/photos/umasstechcast/2295248279/
- https://pixabay.com/en/flowchart-diagram-drawing-concept-311347/
- Berrout J, Mamenko M, Zaika OL, Chen L, Zang W, et al. (2014) Emerging Role of the Calcium-Activated, Small Conductance, SK3 K^+Channel in Distal Tubule Function: Regulation by TRPV4. PLoS ONE 9(4): e95149. doi:10.1371/journal.pone.0095149
- Tim Vickers. “Anti-lipoic acid immunoblot”. http://commons.wikimedia.org/wiki/File%3AAnti-lipoic_acid_immunoblot.png
- Hundahl CA, Fahrenkrug J, Hay-Schmidt A, Georg B, Faltoft B, et al. (2012) Circadian Behaviour in Neuroglobin Deficient Mice. PLoS ONE 7(4): e34462. doi:10.1371/journal.pone.0034462
- Akhmedov K, Rizzo V, Kadakkuzha BM, Carter CJ, Magoski NS, et al. (2013) Decreased Response to Acetylcholine during Aging of Aplysia Neuron R15 . PLoS ONE 8(12): e84793. doi:10.1371/journal.pone.0084793
- Korkotian E, Bombela T, Odegova T, Zubov P, Segal M (2013) Ethanol Affects Network Activity in Cultured Rat Hippocampus: Mediation by Potassium Channels. PLoS ONE 8(11): e75988. doi:10.1371/journal.pone.0075988
- Pilly PK, Grossberg S (2013) Spiking Neurons in a Hierarchical Self-Organizing Map Model Can Learn to Develop Spatial and Temporal Properties of Entorhinal Grid Cells and Hippocampal Place Cells. PLoS ONE 8(4): e60599. doi:10.1371/journal.pone.0060599
- Winden KD, Karsten SL, Bragin A, Kudo LC, Gehman L, et al. (2011) A Systems Level, Functional Genomics Analysis of Chronic Epilepsy. PLoS ONE 6(6): e20763. doi:10.1371/journal.pone.0020763

References

- Duncan Hall. “Postdoc Hell”. https://www.flickr.com/photos/dullhunk/3986237289
- Victorgrigas. “India Victor Grigas 2011-12”. http://commons.wikimedia.org/wiki/File:India_Victor_Grigas_2011-12.jpg
- Antgirl. “Notebook tower”. https://www.flickr.com/photos/kaienong/6384812301
- Victorgrigas. “India Victor Grigas 2011-13”. http://commons.wikimedia.org/wiki/File%3AIndia_Victor_Grigas_2011-13.jpg
- Sydney Uli. “Medical Students”. https://www.flickr.com/photos/sydneyuni/7001513971
- Bejjani A, O'Neill J, Kim JA, Frew AJ, Yee VW, et al. (2012) Elevated Glutamatergic Compounds in Pregenual Anterior Cingulate in Pediatric Autism Spectrum Disorder Demonstrated by 1H MRS and 1H MRSI. PLoS ONE 7(7): e38786. doi:10.1371/journal.pone.0038786
- Rubio-Moscardo F, Setó-Salvia N, Pera M, Bosch-Morató M, Plata C, et al. (2013) Rare Variants in Calcium Homeostasis Modulator 1 (CALHM1) Found in Early Onset Alzheimer’s Disease Patients Alter Calcium Homeostasis. PLoS ONE 8(9): e74203. doi:10.1371/journal.pone.0074203
- REDCap. http://www.project-redcap.org/
- Karen Hanson, Kevin Read, Alisa Surkis. How to Avoid a Data Management Nightmare. https://www.youtube.com/watch?v=nNBiCcBlwRA
- Woshi yufu. “795px-Brueghel-tower-of-babel. https://www.flickr.com/photos/yufujamar/3383400631/
- Stroud LR, Paster RL, Papandonatos GD, et al. Maternal smoking during pregnancy and newborn neurobehavior: A pilot study of effects at 10–27 days. *The Journal of pediatrics*. 2009;154(1):10-16. doi:10.1016/j.jpeds.2008.07.048.
- College Degrees360. “Confused”. https://www.flickr.com/photos/83633410@N07/7658298768/
- Smithsonian Institution. “Elizabeth Lee Hazen (1888-1975) and Rachel Brown (1898-1980)” http://www.flickr.com/photos/25053835@N03/5493818989
- HeroesWiki. http://heroeswiki.com/images/thumb/6/66/Hgp_logo.jpg/250px-Hgp_logo.jpg
- Spithoven A. "DNA". The Noun Project. https://thenounproject.com/search/?q=genomics&i=186662
- Tahil T. "Lock". The Noun Project. https://thenounproject.com/search/?q=protective&i=89842
- Krisada. "Collaboration". The Noun Project. https://thenounproject.com/search/?q=collaboration&i=28324
- Voet T, Kumar P, Van Loo P, et al. Single-cell paired-end genome sequencing reveals structural variation per cell cycle. Nucleic Acids Research. 2013;41(12):6119-6138. doi:10.1093/nar/gkt345
- Alzheimer’s Disease Neuroimaging Initiative. http://www.adni-info.org/
- Oliver Kittler. “Unlocked”. The Noun Project. https://thenounproject.com/kittler/collection/locked-unlocked-lock/?oq=lock&cidx=1&i=424410
- Oliver Kittler. “Locked”. The Noun Project. https://thenounproject.com/kittler/collection/locked-unlocked-lock/?oq=lock&cidx=1&i=424411
- FDA. http://www.fda.gov/ucm/groups/fdagov-public/documents/image/ucm218078.png
- David Richfield. “FlattenedRoundPills”. https://en.wikipedia.org/wiki/File:FlattenedRoundPills.jpg

References

- https://pixabay.com/static/uploads/photo/2014/04/02/17/04/museum-307833_960_720.png
- O'Bryant SE, Xiao G, Barber R, Huebinger R, Wilhelmsen K, et al. (2011) A Blood-Based Screening Tool for Alzheimer's Disease That Spans Serum and Plasma: Findings from TARC and ADNI. PLoS ONE 6(12): e28092. doi:10.1371/journal.pone.0028092
- Swaminathan S, Huentelman MJ, Corneveaux JJ, Myers AJ, Faber KM, et al. (2012) Analysis of Copy Number Variation in Alzheimer's Disease in a Cohort of Clinically Characterized and Neuropathologically Verified Individuals. PLoS ONE 7(12): e50640. doi:10.1371/journal.pone.0050640
- Cai S, Huang L, Zou J, Jing L, Zhai B, et al. (2015) Changes in Thalamic Connectivity in the Early and Late Stages of Amnestic Mild Cognitive Impairment: A Resting-State Functional Magnetic Resonance Study from ADNI. PLoS ONE 10(2): e0115573. doi:10.1371/journal.pone.0115573
- "NIH Data Sharing Repositories". https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html
- Scientific Data. http://www.nature.com/sdata/
- Kafkas Ş, Kim J-H, McEntyre JR (2013) Database Citation in Full Text Biomedical Articles. PLoS ONE 8(5): e63184. doi:10.1371/journal.pone.0063184
- ICPSR Data Management & Curation. http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/elements.html
- National Science Foundation. "Proposal Preparation Instructions". http://www.nsf.gov/pubs/policydocs/pappguide/nsf15001/gpg_2.jsp#dmp
- DMPTool. https://dmp.cdlib.org/
- Open Science Collaboration. PSYCHOLOGY. Estimating the reproducibility of psychological science. Science. 2015 Aug 28;349(6251):aac4716. doi:10.1126/science.aac4716. PubMed PMID: 26315443.
- Anderson CJ, Bahník Š, Barnett-Cowan M, Bosco FA, Chandler J, Chartier CR, Cheung F, Christopherson CD, Cordes A, Cremata EJ, Della Penna N, Estel V, Fedor A, Fitneva SA, Frank MC, Grange JA, Hartshorne JK, Hasselman F, Henninger F, van der Hulst M, Jonas KJ, Lai CK, Levitan CA, Miller JK, Moore KS, Meixner JM, Munafò MR, Neijenhuijs KI, Nilsonne G, Nosek BA, Plessow F, Prenoveau JM, Ricker AA, Schmidt K, Spies JR, Stieger S, Strohminger N, Sullivan GB, van Aert RC, van Assen MA, Vanpaemel W, Vianello M, Voracek M, Zuni K. Response to Comment on "Estimating the reproducibility of psychological science". Science. 2016 Mar 4;351(6277):1037. doi: 10.1126/science.aad9163. PubMed PMID: 26941312.
- Etz A, Vandekerckhove J. A Bayesian Perspective on the Reproducibility Project: Psychology. PLoS One. 2016 Feb 26;11(2):e0149794. doi: 10.1371/journal.pone.0149794. eCollection 2016. PubMed PMID: 26919473; PubMed Central PMCID: PMC4769355.
- Gilbert DT, King G, Pettigrew S, Wilson TD. Comment on "Estimating the reproducibility of psychological science". Science. 2016 Mar 4;351(6277):1037. doi: 10.1126/science.aad7243. PubMed PMID: 26941311.
- SPARC. "Browse Data Sharing Requirements by Federal Agency". http://datasharing.sparcopen.org/

References

- Retraction Watch. “NEJM paper on sleep apnea”. http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/
- Karen Hanson, Alisa Surkis, Karen Yacobucci. Data Sharing and Management Snafu in 3 Short Acts. https://www.youtube.com/watch?v=N2zK3sAtr-4
- ClinicalTrials.gov. “Bronx A1C: Bring it down for health”. https://clinicaltrials.gov/ct2/show/NCT00797888?term=diabetes+and+new+york&rank=6
- Glenn Gaudette. Presentation at UMass’ 2012 New England eScience Symposium. http://escholarship.umassmed.edu/escience_symposium/2012/program/9
- New England Collaborative Data Management Curriculum. http://library.umassmed.edu/necdmc/modules
- AACC website. https://www.aacc.org/publications/cln/articles/2012/april/quality-mistakes
- https://pixabay.com/en/monitor-flatscreen-widescreen-309523/
- biosharing.org. https://biosharing.org/
- Medical Subject Headings. “Investigative Techniques.” http://www.ncbi.nlm.nih.gov/mesh/?term=investigative+techniques
- Gene Ontology. http://bioportal.bioontology.org/ontologies/GO/?p=classes&conceptid=http%3A%2F%2Fpurl.obolibrary.org%2Fobo%2FGO_0051780
- Clinical Data Acquisition Standards Harmonization (CDASH) http://www.cdisc.org/system/files/all/generic/application/pdf/cdash_std_1_1_2010_04_19_review.pdf
- SDTM Tables. https://hvtnews.files.wordpress.com/2013/10/5-7-originalsize.jpg
- NIH. “Summary Table for NIH CDE Initiatives”. http://www.nlm.nih.gov/cde/summary_table_1.html
- Neuromorpho.org. http://neuromorpho.org/neuroMorpho/neuron_info.jsp?neuron_name=D2OE-P90-06
- Normann. “Computer Model 1981” 2014 http://www.flickr.com/photos/26009408@N00/8600423939
- Weiler, Aron.” Encrypted file” 2014. http://www.flickr.com/photos/57523780@N00/52632856
- RRZEicons. “Text-xml”. https://commons.wikimedia.org/wiki/File:Text-xml.svg
- RRZEicons. “Text-csv-text”. https://commons.wikimedia.org/wiki/File:Text-csv-text.svg
- PDF Tricks. http://jon.dehdari.org/tutorials/pdf_tricks.html
- Hopstarter. “Image TIFF Icon”. http://www.iconarchive.com/show/soft-scraps-icons-by-hopstarter/Image-TIFF-icon.html
- BIO RAD. “CFX Manager™ Software”. http://www.bio-rad.com/en-mx/product/cfx-manager-software
- Molecular Devices pClamp 10 Software. https://www.moleculardevices.com/systems/conventional-patch-clamp/pclamp-10-software
- George Hodan. “Stars in the Night Sky”. http://www.publicdomainpictures.net/view-image.php?image=67167

References

- Yuri Somoilov. System Lock. https://www.flickr.com/photos/yusamoilov/13334048894
- https://pixabay.com/p-149782/?no_redirect
- Re3data.org: Registry of Research Repositories. http://service.re3data.org/search
- ICPSR: Inter-University Consortium for Political and Social Research. https://www.icpsr.umich.edu/icpsrweb/landing.jsp
- OpenDOAR: Directory of Open Access Repositories. http://www.opendoar.org/
- DRYAD. http://datadryad.org/
- Government of Canada: Open Data Portal. http://open.canada.ca/data/en/dataset
- Statistics Canada/Statistique Canada: Data Sets and Research Tools. http://www.statcan.gc.ca/eng/rdc/data
- Canadian Research Data Centre Network. http://www.rdc-cdr.ca/datasets-and-surveys
- Read, Kevin (2015): Sizing the Problem of Improving Discovery and Access to NIH-funded Data: A preliminary study. Figshare. https://dx.doi.org/10.6084/m9.figshare.1285515.v1 Retrieved: 01 56, May 09, 2016 (GMT)
- Gigascience. http://gigascience.biomedcentral.com/
- Digital Curation Centre: Where to Keep Research Data. http://www.dcc.ac.uk/resources/how-guides-checklists/where-keep-research-data#5
- GotCredit. Share. https://www.flickr.com/photos/jakerust/16649920968
- opensource.com. “Building an open source business”. https://www.flickr.com/photos/opensourceway/4427310974
- Calamity_sal. “library life”. https://www.flickr.com/photos/calamity_sal/3375002696
- Torkild Retvedt. “Server room”. https://www.flickr.com/photos/torkildr/3462607995
- Lane F. Kinkade. The Noun Project. Microscope. https://thenounproject.com/search/?q=microscope&i=113889
- PLOS. “PLOS’ New Data Policy: Public Access to Data”. Feb 24, 2014. http://blogs.plos.org/everyone/2014/02/24/plos-new-data-policy-public-access-data-2/
- “NEJM paper on sleep apnea retracted when original data can’t be found”. Retraction Watch. http://retractionwatch.com/2013/10/30/nejm-paper-on-sleep-apnea-retracted-when-original-data-cant-be-found/
- Data Horror Stories. https://pinboard.in/u:dsalo/t:horrorstories
- PhD Comics. http://www.phdcomics.com/comics/archive.php?comicid=961
- Michael Dalby. “Librarians”. https://www.flickr.com/photos/medalby/274218052/
- Christian Gauthier. “Drowning under a mountain of paper”. https://www.flickr.com/photos/wheatfields/4774087006
- Portage. https://portagenetwork.ca/