



Current Topics and Best Practices in Research Ethics UCSF Winter/Spring 2024 RCR for Basic Sciences Postdocs

Tori Sharma and Ariel Deardorff
Session 3

New Guidance on Research Security

Word Cloud: One word about your knowledge on research security

Nobody has responded yet.

Hang tight! Responses are coming in.



Recent Developments in Research Security

- JASON Report (2019)
 - Full disclosure of all potential conflicts of interest
 - Recommends violations be investigated and adjudicated as research misconduct
- NSPM-33 (2021)
 - Requires digital persistent identifiers
 - Mandates Research Security Programs
 - Guidance on consequences for breaches
- CHIPS and Science Act (2022)
 - Establishes a clearing house for research security and integrity
 - Mandates research security training as part of RCR
 - Certification of grant applicants that they are not participating in malign foreign recruitment programs
- Research Security Programs Standard Requirements (2023)
 - Requires an institutional research security program (registered in SAM.gov)
 - Research security, cyber security, foreign travel, and export control



Encourages Appropriate Use of Funds



Prevents Deceptive Practices by Bad Actors



Protects IP & Innovations from Theft or Misuse



Protects Sensitive Information



Minimizes Foreign Government Interference

Collaboration and Knowledge

Integrity: Relies on trust to ensure attribution, acknowledgement, and dissemination of findings.

Security: Facilitates collaboration and protects IP and sensitive data.

Both: Promote collaboration and responsible knowledge sharing

Trust and Credibility

Integrity: Ensures processes and results are consistent with core values and ethics

Security: Ensures results are protected from unauthorized access or manipulation by external parties

Both: Help establish and maintain trust

Compliance

Integrity: Complies with guidelines and regulations governing ethics, human subjects, and animal welfare.

Security: Complies with regulations and best practices related to data, export controls, and IP protection.

Both: Require adherence to regulations, policies, and guidelines.

Protection of Research Assets

Integrity: Gives high confidence in accuracy and reliability of findings.

Security: Focuses on safeguarding data, IP, and sensitive information from unauthorized access, tampering, and theft.

Both: Aim to protect valuable research assets.

Core Values of Academic Research

Openness and transparency

Accountability and honesty

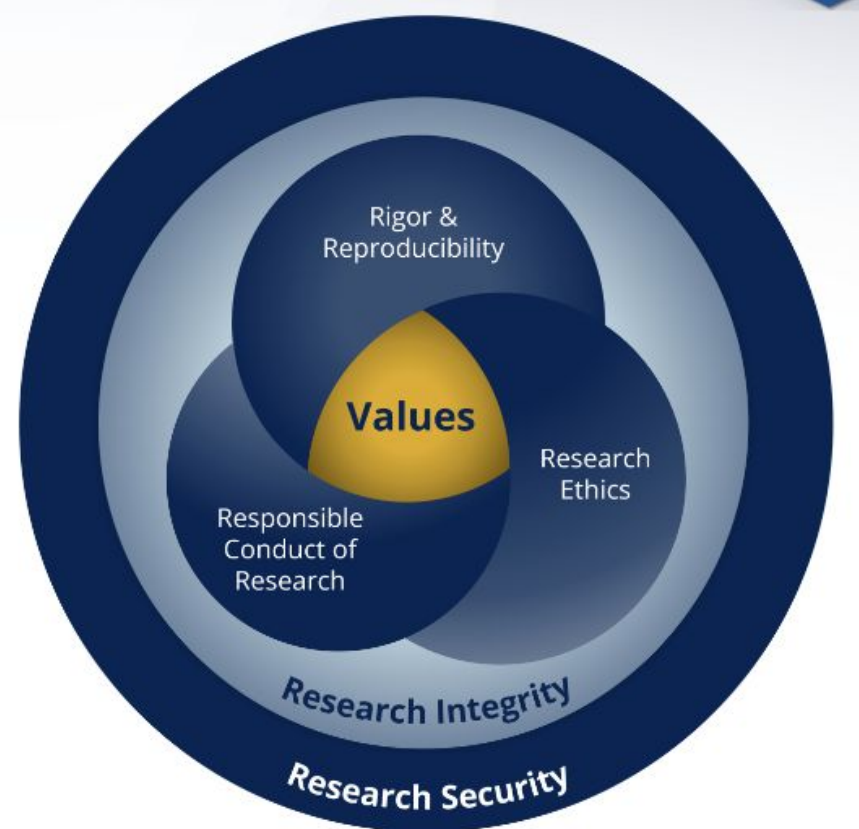
Impartiality and objectivity

Respect

Freedom of inquiry

Reciprocity

Merit-based competition



Countries of Particular Concern

The most recent Countries of Particular Concern designations were made by the Secretary of State on December 29, 2023:

Burma, People's Republic of China, Cuba, Eritrea, Iran, the Democratic People's Republic of Korea, Nicaragua, Pakistan, Russia, Saudi Arabia, Tajikistan, and Turkmenistan.

<https://www.state.gov/countries-of-particular-concern-special-watch-list-countries-entities-of-particular-concern/> accessed February 15, 2024



UCSF Research Security Program

Data Management and Sharing at UCSF

Ariel Deardorff, UCSF

- Data Services Librarian
- UCSF Library - [Data Science Team](#)
- Data Management + Sharing Educator
- [Open Science Researcher](#)

Want to meet with me?

- Send me an email at ariel.deardorff@ucsf.edu
- [Book a free consultation](#)



Additional Contributors

Erin Foster

RDM Program Service Lead
Library/ResearchIT

Ann Glusker

Sociology, Demography &
Quantitative Research
Librarian

Anna Sackmann

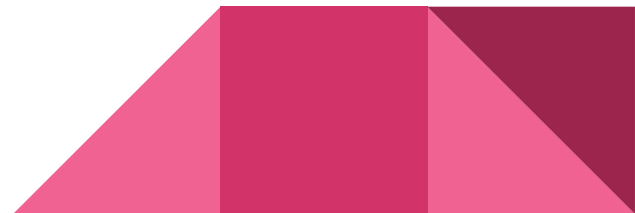
Data Services Librarian

Elliott Smith

Emerging Technologies &
Bioinformatics Librarian

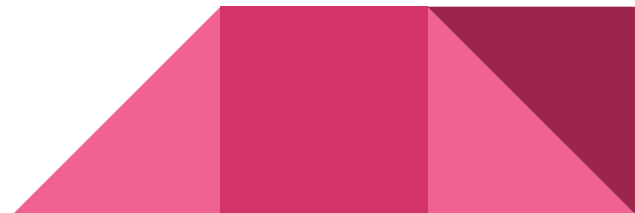
Samantha Teplitzky

Open Science Librarian



Topics on Data Management for Today:

1. Data Management and Responsible Conduct of Research
2. Data Management Planning
3. Documentation and Description
4. Storage and Backup (and Transfer)
5. Sharing and Publishing





Support Your Data

Credit: <https://pitt.libguides.com/managedata/understanding>

- Data is a scholarly product
- Data are easily lost
- Data requirements may be externally imposed (DMP requirements, etc.)
- Good management prevents errors and increases the quality of analysis
- Good management allows others to validate and replicate findings, enhances collaboration
- When shared, data can lead to unintended valuable discoveries
- RDM SAVES TIME



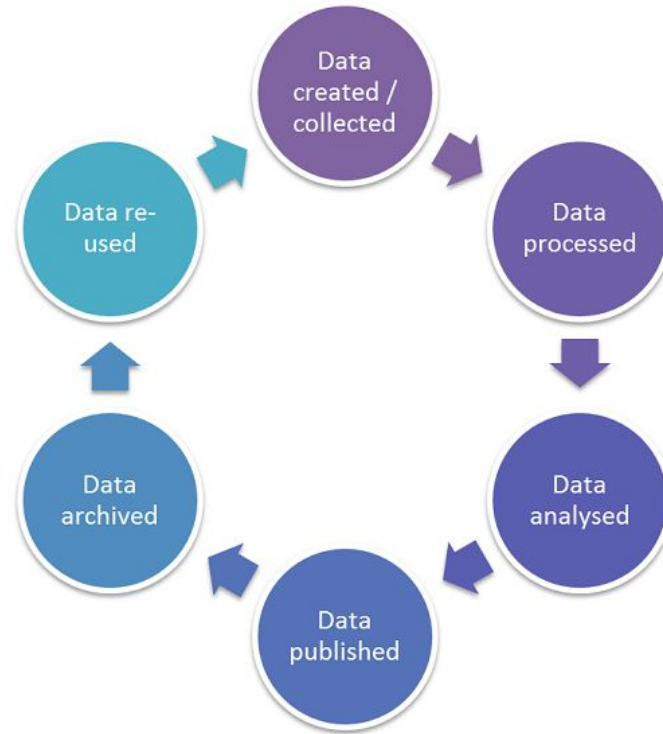
Avoid becoming a
data horror story!

What is a data management plan?

A **living** document that addresses how you will manage and secure your data throughout the lifecycle of a research project

Also, a required document for

- NIH grants
- NSF grants
- Eventually, all federal grants



Why make a DMP?

Circumvent future problems by anticipating issues—a plan can help ensure that:

- you have **adequate technological resources** (e.g. storage space, support staff time)
- your data will be **robust and free from versioning errors** and gaps in documentation
- your data is **backed up** and **safe** from sudden loss or corruption
- you can meet **legal and ethical** requirements
- you are able to **share your finalized data publicly**, if you and/or your funder desires
- your data will **remain accessible and comprehensible** in the near, middle, and distant future.

Remember you can review and revise your plan— living document, iterative

What about this new NIH DMP requirement?

- Effective for all NIH grants starting **January 25, 2023**
- Requires researchers seeking NIH funding to prospectively **submit a 2-page plan outlining how scientific data from their research will be managed and shared**
- Researchers should “**maximize the appropriate sharing of scientific data**”
- Data should be shared **as soon as possible**, and no later than the time of an associated publication or end of performance period (whichever comes first)
- This plan represents the **minimum requirements**. NIH ICOs may expect more specificity in their plans - check funding announcements for info
- Read the [full policy](#)

Elements of the new NIH Plan

In the two-page document researchers will be asked to describe their:

1. Data type
2. Related tools, software, and/or code
3. Standards (file formats, data dictionaries, etc)
4. Data preservation, access, and associated timelines
5. Access, distribution, or reuse considerations
6. Oversight of data management and sharing

What is the overall goal here?



Increase public access to
results of federally funded
research



Increase transparency and
reproducibility of research
results



Make data available for
discovery and reuse

Get started, using the DMP Tool

dmptool.org/ is:

- Free
- Open source
- Has templates
- User friendly
- Has examples



Build your Data Management Plan

[Funder Requirements](#)

[Public DMPs](#)

[Help](#)



103,197 Users



386 Participating Institutions



99,505 Plans

Data Management at UCSF

Some resources and examples

- [UK Data Service Data Management Checklist](#)
- Open Science Framework [DMP overview](#)
- [ICPSR Data Management page](#)
- Generic [NIH DMP example](#)
- Example from the [University of Minnesota](#)



Are currently following a data management plan (DMP)?

Yes, a formal DMP



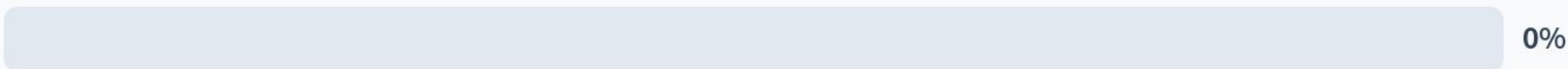
Yes, an informal DMP



No



Not sure



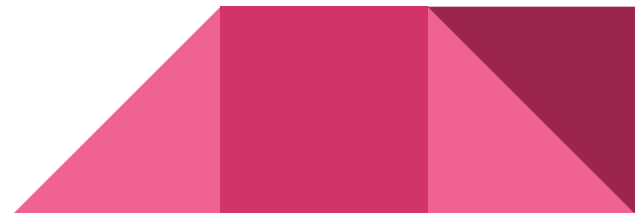
New Sharing Mandate

OSTP Nelson Memorandum:

Ensuring Free, Immediate, and Equitable Access to Federally Funded Research

(effective 2026):

“Scientific data underlying peer-reviewed scholarly publications resulting from federally funded research should be made freely available and publicly accessible by default at the time of publication...”

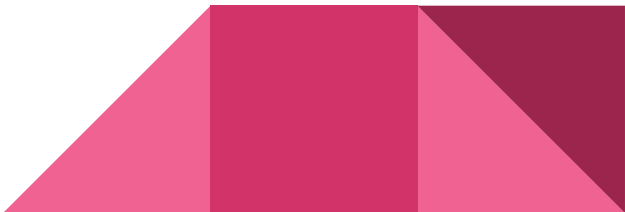


Documentation + Description

Persistent identifiers (PIDs): enable data discovery, interoperability, reuse

- [ORCID](#) (Open Researcher and Contributor ID)
- [RRID](#) (Research Resource ID)
- [DOI](#) (Digital Object Identifier)
- [ARK](#) (Archival Resource Key)
- [ROR ID](#) (Research Organization Registry ID)

OSTP memo: include PIDs for authors, affiliations, funders, & research outputs



Documentation + Description

File formats: Best practices

- Non-proprietary
- Unencrypted
- Uncompressed
- In common use in the discipline
- Open, documented, supported, interoperable standards



Portrait of scribe Jean Miélot by Jean Le Tavernier. Bibliothèque nationale de France

Documentation + Description

Excel: Cautionary cases

- **Hidden formula errors:** [The Reinhardt-Rogoff error – or how not to Excel at economics](#)
- **Template limitations:** [Why you should never use Microsoft Excel to count coronavirus cases](#)
- **Autoformatting:** [Excel autocorrect errors still plague genetic research, raising concerns over scientific rigour](#)



Documentation + Description

Some preferred file formats:

- Tables, databases or other structured information: .csv, XML, .json
- Textual information: .txt, HTML, LaTeX, XML, PDF/A
- Software and code: open source/open standard
- Still images: .tiff or .png
- Moving images: MPEG-4 (.mp4)
- Audio: .WAV or .AIFF



Documentation + Description

File names should be sortable and searchable:

- Project or experiment name or acronym
- Researcher name/initials
- Date or date range of experiment
- Type of data
- Sample identifiers, strains, etc.
- Version or run number



Image by Tony Webster

<https://www.flickr.com/photos/87296837@N00/23665323559>

Documentation + Description

Naming conventions:

- Limit to under 50 characters if possible (some sources say 32)
- Use only letters, numbers, underscores, hyphens, or camelCase
- Avoid spaces or special characters: ~ ! @ # \$ % ^ & * () ; < > ? , [] { } ' "
- Date format (ISO 8601 Standard): YYYYMMDDThhmmss
- For sequential numbers use leading zeros: "001, 002" instead of "1, 2"
- Use versioning (v1, etc.) instead of "final," "finalfinal,"
"finalfinal_usethisone"



Documentation + Description

File name example

MGE_NkA1_JDS_104C1_20181107T141524_003_v2.tif

ExperimentName_InstrumentID_ResearcherInitials_CellLine_CaptureDateTime_ImageID_version

Make sure your chosen template orders your data files in the desired way.



Do you use a file naming convention

YES!

0%

Sort of

0%

No

0%

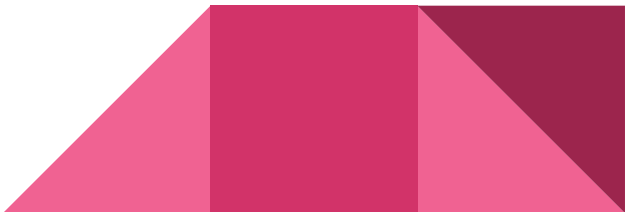
This conversation scares me

0%



Documentation + Description

Basic metadata (“Information about the data”):

- **Who** created the data
 - **What** the data file contains
 - **When** the data were generated
 - **Where** the data were generated
 - **Why** the data were generated
 - **How** the data were generated
- 

Documentation + Description

Metadata standards: specific standards differ by discipline and repository

- FAIR Principles: <https://www.go-fair.org/fair-principles/>
- Standards by discipline:
 - DCC: <https://www.dcc.ac.uk/guidance/standards/metadata>
 - FAIRsharing: <https://fairsharing.org/standards/>
 - EQUATOR: <https://www.equator-network.org/reporting-guidelines/>

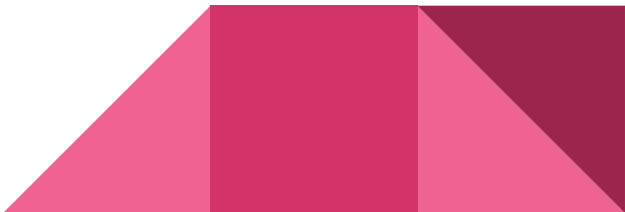
What not to do:

Biased **O**utcomes reporting **G**uidelines for **U**nderwhelming **S**tudies (**BOGUS**)



Documentation + Description

README.txt elements

- Title of dataset
 - Name and contact information for PI and responsible researcher
 - File name template, elements and codes
 - File formats present
 - Column headings for tables: variable names, units, etc.
 - Data processing: How final data were derived from raw data
 - Versioning: Change log for documenting file versions
- 

Documentation + Description

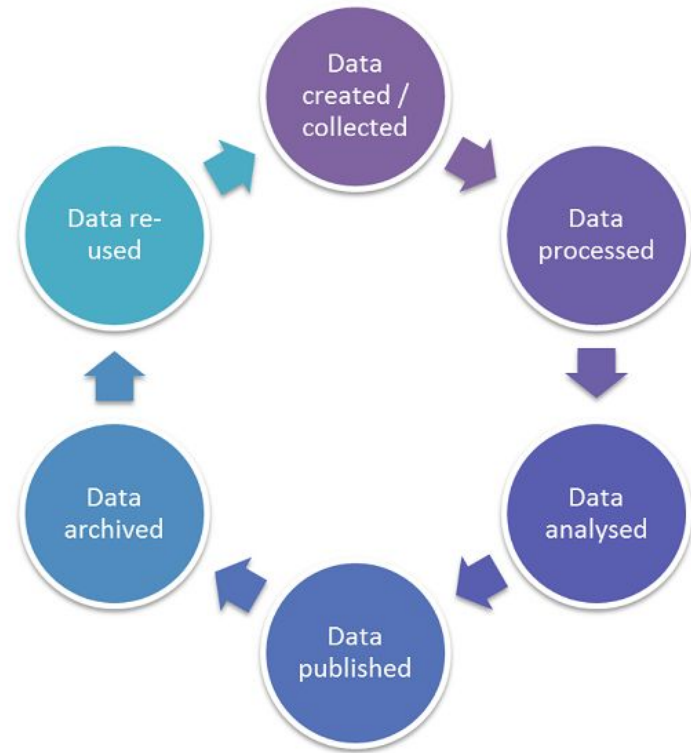
README.txt cont.

- Include for each dataset or group of similar datasets
- Helps you and other researchers understand the data
- Guides:
 - <https://datamanagement.hms.harvard.edu/readme-files>
 - <https://data.research.cornell.edu/content/readme>
- Example README files



Your **data** and **documentation** likely live in several different places depending on where you are at in the research process

Let's talk UCSF resources...



Documentation in your lab notebook:

- Your lab notebook is a complete record of procedures, reagents, data, and calculations to pass on to other researchers.
- It is a source of knowledge for your lab, and a legal document for patents.
- Notebooks should be retained 6 years beyond the close of a study
 - Unless the funding mandates a longer retention period, the data concerns clinical research involving minors or pregnant participants, or there is a patent application.
- Want to use an Electronic Lab Notebook? Work with IT Customer Solutions Group for an [IT Consultation Request](#)
 - help.ucsf.edu -> Consulting & Development -> IT Consultation Request
- UCSF has [guidance for maintaining lab notebooks](#)

Ex: Protocol hosted on protocols.io

OPEN ACCESS

✓ protocols.io



mcSCRB-seq protocol [↗](#)

Version 2

Nature Communications

Johannes Bagnoli¹, Christoph Ziegenhain¹, Aleksandar Janjic¹, Lucas Esteban Wange¹, Beate Vieth¹, Swati Parekh¹, Johanna Geuder¹, Ines Hellmann¹, Wolfgang Enard¹

¹Ludwig-Maximilians-Universität München

[dx.doi.org/10.17504/protocols.io.p9kdr4w](https://doi.org/10.17504/protocols.io.p9kdr4w)

[Human Cell Atlas Method Development Community](#)



Aleksandar Janjic

Ludwig-Maximilians-Universität...



ABSTRACT

Single-cell RNA sequencing (scRNA-seq) has emerged as a central genome-wide method to characterize cellular identities and processes. Consequently, improving its sensitivity, flexibility and cost-efficiency can advance many research questions. Among the flexible plate-based methods, "Single-Cell RNA-Barcoding and Sequencing" (SCRB-seq) is one of the most sensitive and efficient ones. Here, we systematically evaluated experimental conditions of this protocol and find that adding polyethylene glycol considerably increases sensitivity by enhancing cDNA synthesis. Furthermore, using Terra polymerase increases efficiency due to a more even cDNA amplification that requires less sequencing of libraries. We combined these and other improvements to a new scRNA-seq library protocol we call "molecular crowding SCRB-seq" (mcSCRB-seq), which we show to be the most sensitive and one of the most efficient and flexible scRNA-seq methods to date.

EXTERNAL LINK

<https://www.nature.com/articles/s41467-018-05347-6>

THIS PROTOCOL ACCOMPANIES THE FOLLOWING PUBLICATION

Johannes W. Bagnoli, et al (2018) Sensitive and powerful single-cell RNA sequencing using mcSCRB-seq. *Nature Communications* 9:2937. doi: [10.1038/s41467-018-05347-6](https://doi.org/10.1038/s41467-018-05347-6)

UCSF subscribes to Protocols.io meaning that our researchers can store protocols privately, and use their protocol service to upload and transform your protocols.

Source: [dx.doi.org/10.17504/protocols.io.p9kdr4w](https://doi.org/10.17504/protocols.io.p9kdr4w)

Ex: Data Dictionary in REDCap

 Print page

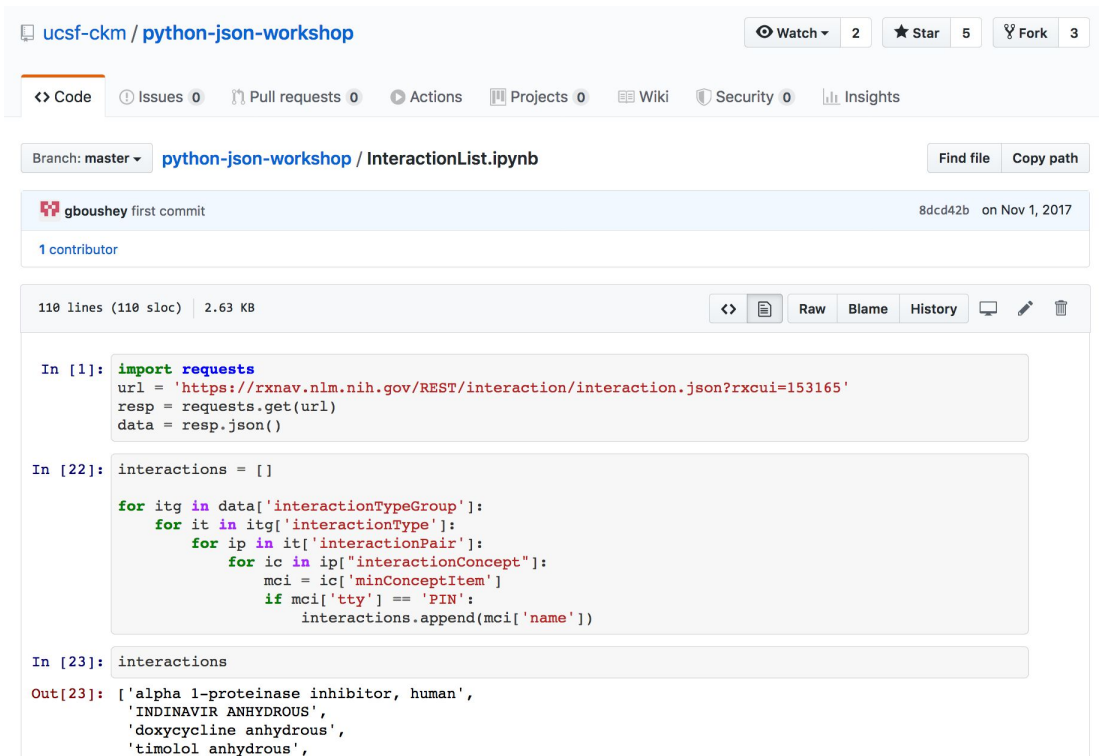
Data Dictionary Codebook

#	Variable / Field Name	Field Label <i>Field Note</i>	Field Attributes (Field Type, Validation, Choices, Calculations, etc.)														
Instrument: Demographics (demographics)																	
	1 study_id	Study ID	text														
 	2 date_enrolled	Section Header: <i>Consent Information</i> Date subject signed consent <i>YYYY-MM-DD</i>	text (date_ymd)														
 	3 first_name	Section Header: <i>Contact Information</i> First Name	text, Identifier														
 	4 last_name	Last Name	text, Identifier														
 	5 dob	Date of birth	text (date_ymd), Identifier														
 	6 ethnicity	Ethnicity	radio <table><tr><td>0</td><td>Hispanic or Latino</td></tr><tr><td>1</td><td>NOT Hispanic or Latino</td></tr><tr><td>2</td><td>Unknown / Not Reported</td></tr></table> Custom alignment: LH	0	Hispanic or Latino	1	NOT Hispanic or Latino	2	Unknown / Not Reported								
0	Hispanic or Latino																
1	NOT Hispanic or Latino																
2	Unknown / Not Reported																
 	7 race	Race	dropdown <table><tr><td>0</td><td>American Indian/Alaska Native</td></tr><tr><td>1</td><td>Asian</td></tr><tr><td>2</td><td>Native Hawaiian or Other Pacific Islander</td></tr><tr><td>3</td><td>Black or African American</td></tr><tr><td>4</td><td>White</td></tr><tr><td>5</td><td>More Than One Race</td></tr><tr><td>6</td><td>Unknown / Not Reported</td></tr></table>	0	American Indian/Alaska Native	1	Asian	2	Native Hawaiian or Other Pacific Islander	3	Black or African American	4	White	5	More Than One Race	6	Unknown / Not Reported
0	American Indian/Alaska Native																
1	Asian																
2	Native Hawaiian or Other Pacific Islander																
3	Black or African American																
4	White																
5	More Than One Race																
6	Unknown / Not Reported																

UCSF has an enterprise instance of RedCap

Ex: Code in a Jupyter Notebook on Github

UCSF now has a secure version of GitHub for researchers who want to work with sensitive data.



ucsf-ckm / python-json-workshop

Watch 2 Star 5 Fork 3

Code Issues 0 Pull requests 0 Actions Projects 0 Wiki Security 0 Insights

Branch: master python-json-workshop / InteractionList.ipynb Find file Copy path

gboushey first commit 8dcd42b on Nov 1, 2017

1 contributor

110 lines (110 sloc) 2.63 KB

<> Raw Blame History

```
In [1]: import requests
url = 'https://rxnav.nlm.nih.gov/REST/interaction/interaction.json?rxcur=153165'
resp = requests.get(url)
data = resp.json()

In [22]: interactions = []

for itg in data['interactionTypeGroup']:
    for it in itg['interactionType']:
        for ip in it['interactionPair']:
            for ic in ip["interactionConcept"]:
                mci = ic['minConceptItem']
                if mci['tty'] == 'PIN':
                    interactions.append(mci['name'])

In [23]: interactions

Out[23]: ['alpha 1-proteinase inhibitor, human',
          'INDINAVIR ANHYDROUS',
          'doxycycline anhydrous',
          'timolol anhydrous']
```

Source: Boushey, Geoff. Python-JSON-Workshop
<https://github.com/ucsf-ckm/python-json-workshop/blob/master/InteractionList.ipynb>

Want to make sure you are documenting the right stuff?

Reporting guidelines for publishing your research:

- [CONSORT](#) standard for clinical trials
- [PRISMA](#) standard for systematic reviews
- [ARRIVE](#) for animal pre-clinical research

Standards for different data types and experiments:

- [MIAME](#) for microarray experiments
- [NWB](#) for Neurophysiology data

Other reporting guidelines and metadata standards available at [EQUATOR network](#) or [Fairsharing.org](#)

What could go wrong?

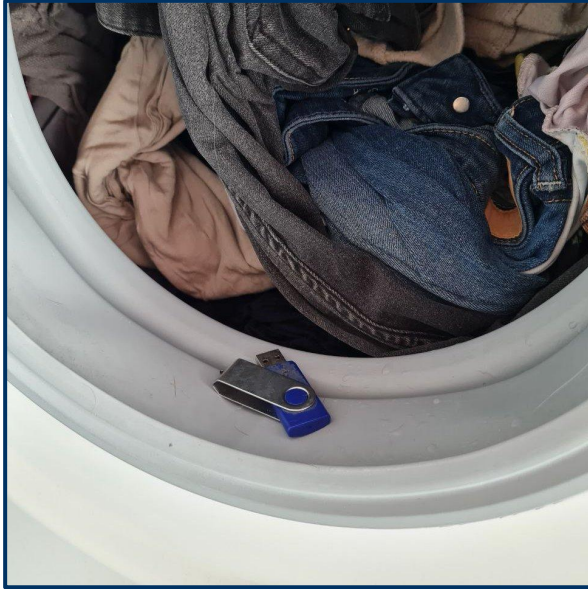


Photo credit: Juli Tkotz (@juli_tkotz)



Photo credit: Citizen Reporter



Photo credit: Amy Neeser

Here + Near + Far (3-2-1 rule)

Here = local, working copy (e.g., laptop, lab computer)

Near = other local, external copy or in a remote location (e.g., external hard drive, network attached storage)

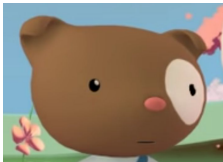
Far = remote copy (e.g., cloud storage, network attached storage)



Where does your data live?

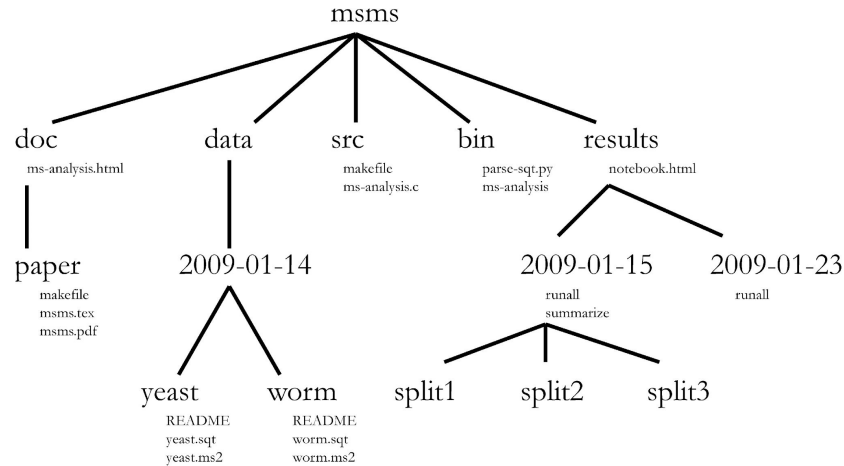
UCSF Secure Storage Options:

- [IT Data Center Services](#)
- [REDCap/Qualtrics](#)
- [RAE](#) (formerly MyResearch)
- (Secure) [Box](#)
- OneDrive in [Office 365](#)
- [Cloud computing options](#)



Maybe a flash drive isn't the best option. I think I will move my data to Box and store a copy on my lab server.

How are your files organized?



Micah Allen
@micahgallen

Follow

My universal directory structure, forged from years of bad organization:

```
/Projects
.../inProgress
...../ProjectName
...../docs
...../code
...../data
...../figures
.../published
.../submitted
```

You're welcome.

Sharing

Defined as “making data for scholarly research available (and accessible) for use”.

- Followed closely on the heels of public access to research articles
- Response to lack of availability of primary data
- Seen as a mechanism to ensure reproducibility and transparency in the research process



Sharing

Your funder may require it!

- National Science Foundation
- National Institutes of Health
- OSTP Memorandum

Your publisher may require it!

- PLoS
- Springer/Nature
- Wiley



Types of repositories



Domain specific
(e.g., ICPSR, QDR)



Institutional
(e.g., Dryad @ UC)



Generalist
(e.g., Figshare,
Zenodo)



How do you choose a data repository?



Is there a repository named in your funding announcement?



Is there a repository specific to your discipline or data type?



Is there a generalist repository available?

How to find a data repository for your discipline

- Places to Search:

- [NIH-Sponsored Repositories](#)
- [Scientific Data Guide to Data Repositories](#)
- [UCSF Library Guide to Repositories](#)

- To consider:

- Do you want to restrict access to your data or make it completely open?
- Does the repository charge any fees?

Some ideas for disciplinary repositories

Research Type	Repository	Considerations
Clinical Research	Vivli	Restricted Access, \$
Neuroscience	OpenNeuro	Requires BIDS data format
Microscopy	Image Data Resource	
HIV	National Addiction & HIV Data Archive Program	Restricted Access
Qualitative data	Qualitative Data Repository	Restricted Access, \$
Alzheimer's	AD Knowledge Portal	Restricted Access

Spotlight on Dryad – generalist repository



DRYAD

UCSF Library

[Admin](#)[Explore Data](#)[About ▼](#)[Help ▼](#)[My Datasets](#)[Logout](#)

for your research data

Dryad is a community-owned resource
[Learn more about our organizational memberships](#)

[Submit Now](#)

Data Sharing Requires Advanced Planning

Project Planning Stage:

- Check sponsor agreements for guidance
- Make a Data Management Plan
- Incorporate data sharing language into your IRB & consent forms
- Research data repositories – pay attention to any required standards
- Include costs for data de-identification & curation into your grants

Publishing and Sharing Stage:

- Prepare a shareable version of your dataset
- Upload to your chosen repository
- Get your DOI or accession number

UCSF now has step-by-step guidance

UCSF Data Resources

Research ▼ EHR Reporting Team ▼ Education & Training Administration Self-service Analytics (SSA) ▼ Process & Policy ▼ Events & Community ▼

SSA Frequently Asked Questions

[Home](#) > [Process & Policy](#) > Sharing De-identified Data for Publication

Sharing De-identified Data for Publication

Overview

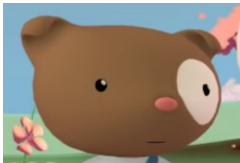
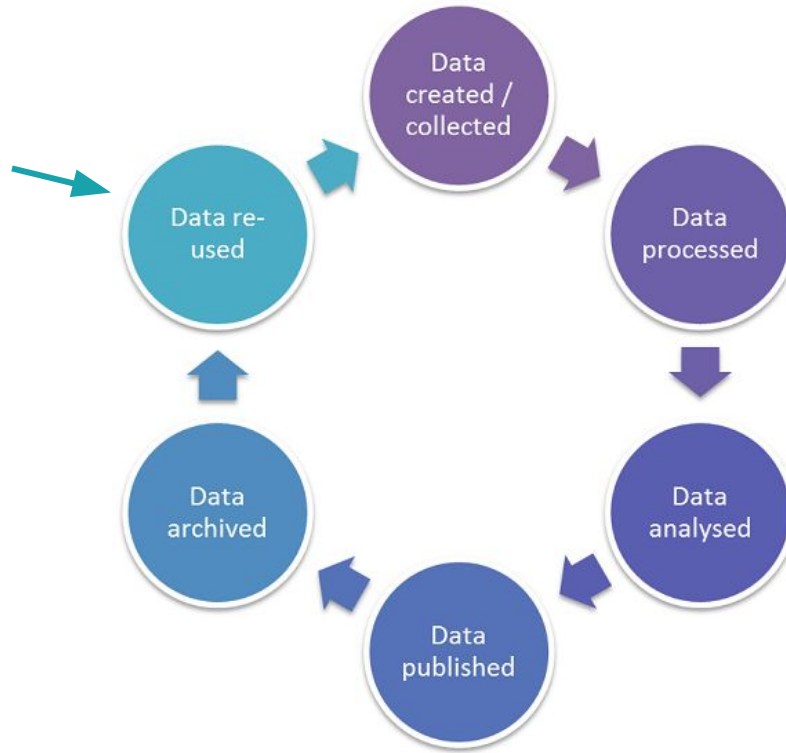
As part of a push to improve research reproducibility an increasing number of health science funders and publishers are asking researchers to share the de-identified data underlying their research. NIH already has a [genomic data sharing requirement](#), and recently announced new [requirement for data management and sharing plans](#) that will go into effect in 2023. This page explains how to plan for data sharing so that you can meet these requirements while also following UCSF guidance on privacy and data security.

Note that sharing [P4 data types](#) including identified PHI/PII/RHI data does not fall under this process, and researchers interested in sharing this kind of data will need to follow the [guidance on this page](#).

If you have questions about sharing models or algorithms please reach out to Industry Contracts at Industrycontracts@ucsf.edu

data.ucsf.edu/sharing-de-identified-data-publication

This is the goal



I think I will deposit a copy of my data in Dryad so researchers don't need to email me.

UCSF Resources

- [NIH DMS Policy Overview Page](#)
 - Includes FAQs, templates and sample plans
- [UCSF Guide to De-Identified Data Sharing](#)
- [Guide to Reproducible Data Management](#)
- [UCSF Secure Research Guide](#)
- [Book a 1:1 consultation with Ariel](#)
- [Bay Area Open Science Group](#)