



# Research Data Curation: Services for UC San Diego

David Minor, Reid Otsuji

Research Data Curation Program

UC San Diego Library

March 2024

# What *is* data curation?

# Research process:

- Write a grant
- Create/collect data
- Analyze data
- Publish results
- Rinse and repeat

- Write a grant
- Create/collect data
- Analyze data
- Publish results
- Rinse and repeat



**Rinse and repeat**

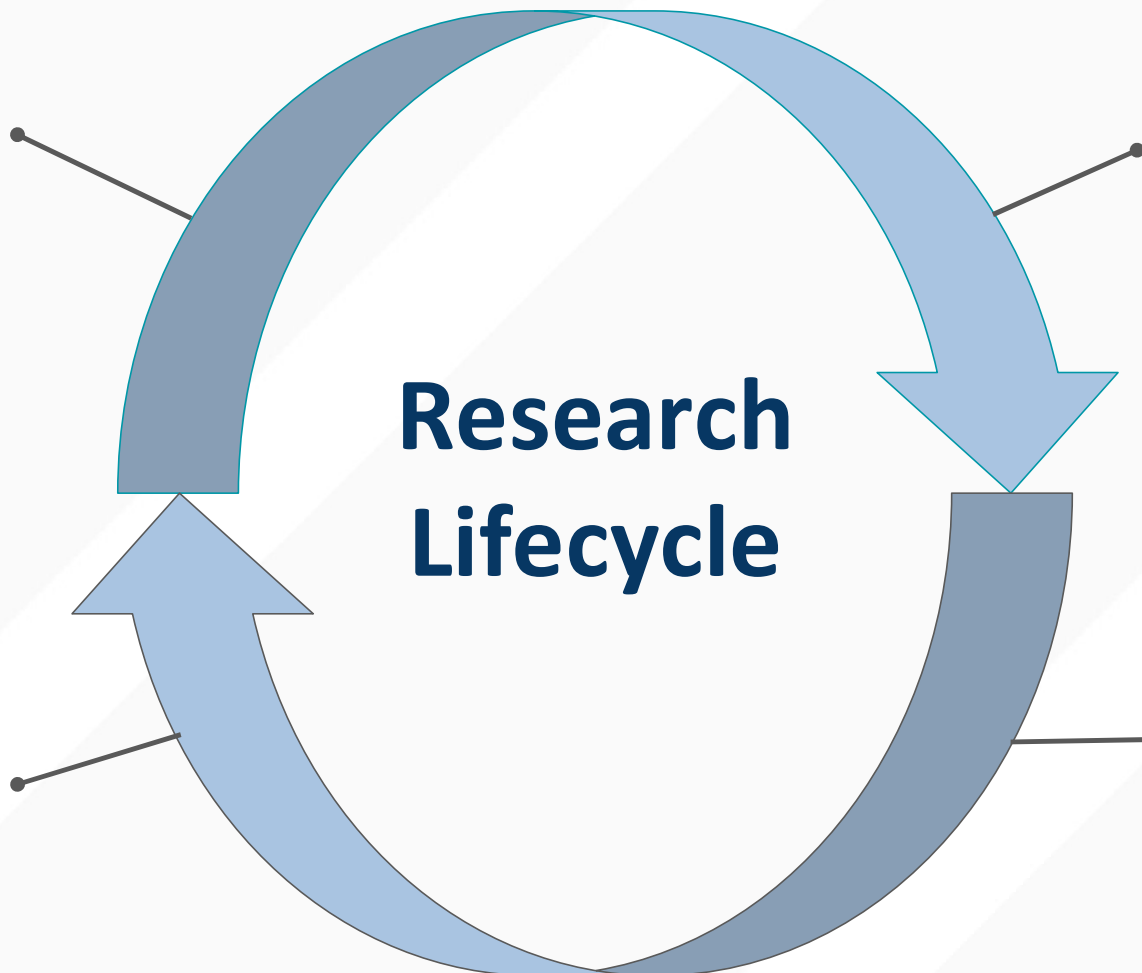
**Write a grant**

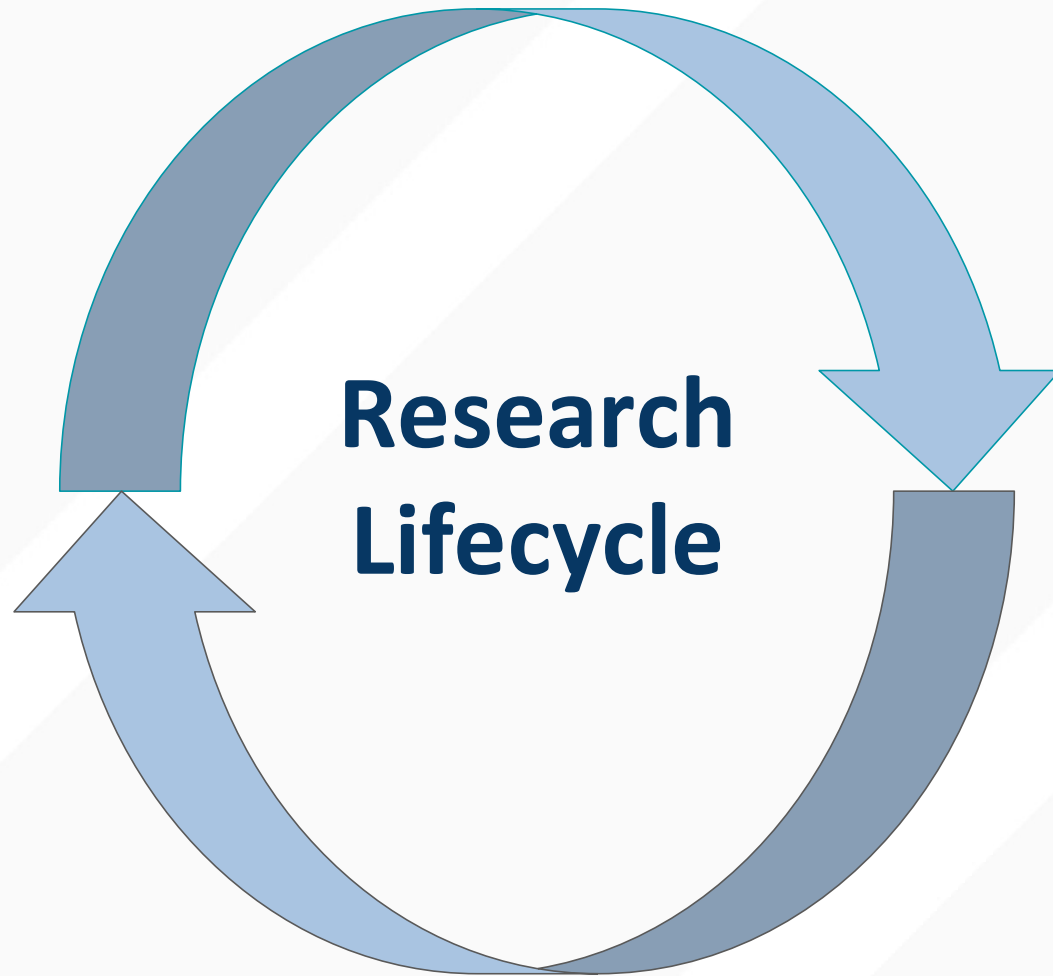
# Research Lifecycle

**Create/collect data**

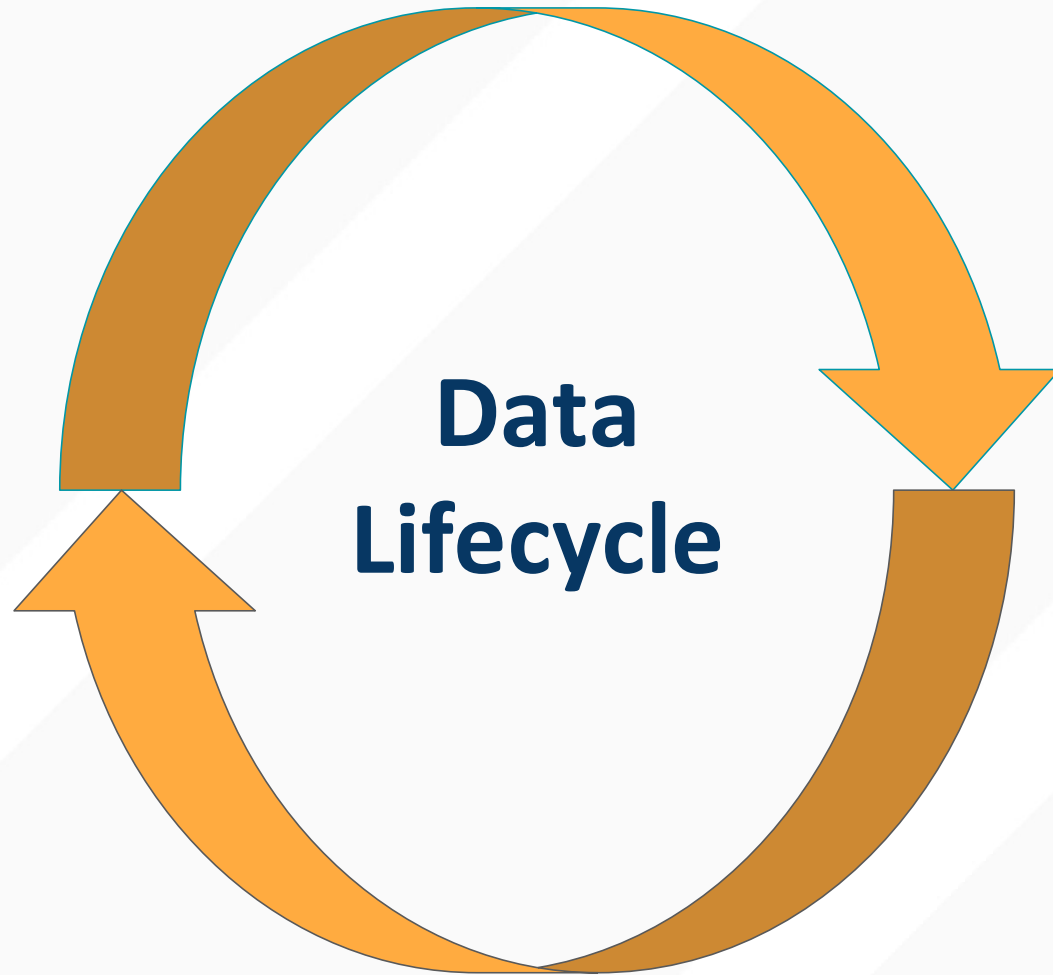
**Publish results**

**Analyze data**





**Research  
Lifecycle**



**Data  
Lifecycle**

**Data  
Preservation**

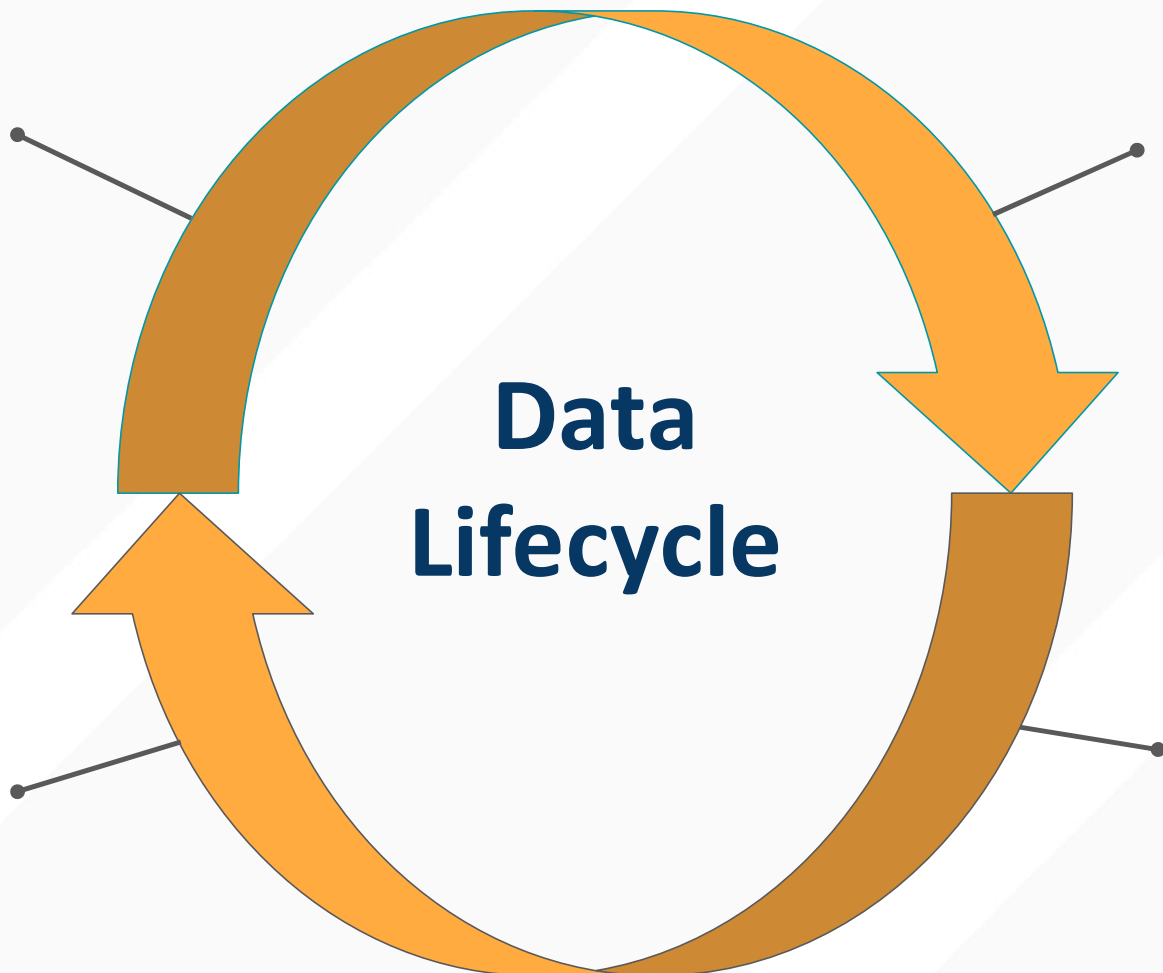
**Data Collection /  
Creation**

# **Data Lifecycle**

**Data Sharing**

**Data  
Processing**

**Data Analysis**





... creating, organizing and maintaining data so they can be discovered, accessed and used now and in the future. It involves collecting, structuring and indexing data as well as creating metadata.

**Why should anyone *care*  
about data curation?**

**It's a Good Thing.™**

To help make data **FAIR** –  
**F**indable, **A**ccessible,  
**I**nteroperable and **R**eusable

**Because of funder requirements**

# August 2022: OSTP issues “The Nelson Memo”



EXECUTIVE OFFICE OF THE PRESIDENT  
OFFICE OF SCIENCE AND TECHNOLOGY POLICY  
WASHINGTON, D.C. 20502

August 25, 2022

MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

FROM:

Dr. Alondra Nelson

A handwritten signature in blue ink that reads "Alondra Nelson".

Deputy Assistant to the President and Deputy Director for Science and Society  
Performing the Duties of Director  
Office of Science and Technology Policy (OSTP)

SUBJECT:

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

# OSTP 2022 Nelson Memo Directives

## **Publications** must be made available:

- freely and publicly
- without embargo or delay
- in machine-readable formats

## **Data** must be made available:

- without embargo or delay if supporting a publication
- per agency guidelines if not associated with a publication
- in a way that *maximizes appropriate sharing*
- in a repository that meets [specific criteria](#)

## **Metadata** must be made available:

- for both articles and data
- at time of deposit in a public access repository.

## **Identifiers** must be used:

- unique persistent identifiers must be used for individuals, awards, organizations and outputs.

# Timeline for new funder requirements

<b>Publication &amp; Data Access Policies</b>	<ul style="list-style-type: none"><li>• Draft plans by: February* <b>2023</b></li><li>• Published by: December <b>2024</b></li><li>• Effective by: December <b>2025</b></li></ul>
<b>Metadata &amp; Persistent Identifier Policies</b>	<ul style="list-style-type: none"><li>• Draft plans by: December <b>2024</b></li><li>• Published by: December <b>2026</b></li><li>• Effective by: December <b>2027</b></li></ul>



# New NIH Data Management & Sharing Policy

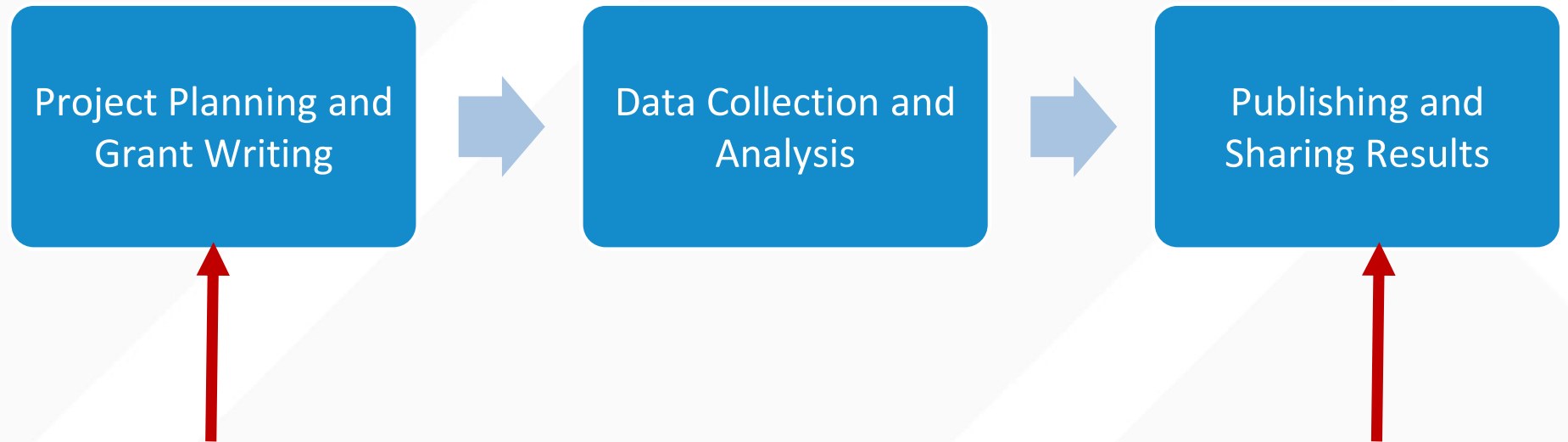
- Effective **January 25, 2023**
- Requires researchers seeking NIH funding for research to prospectively **submit a 2-page plan outlining how scientific data from their research will be managed and shared**
- Researchers should “**maximize the appropriate sharing of scientific data**”
- Data should 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 opportunity announcements

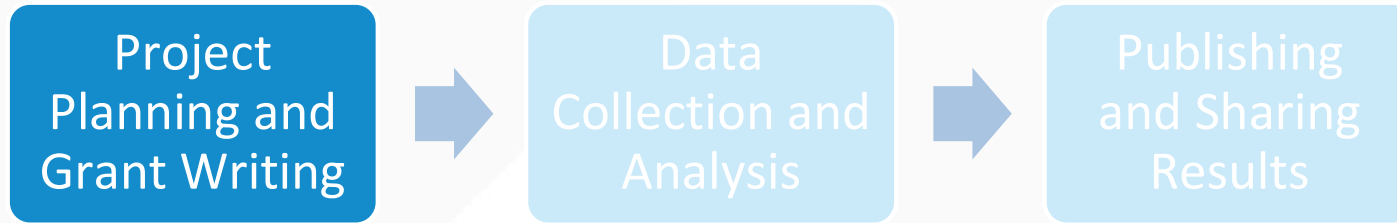
# Planning for Data Sharing

# Prepare in advance for data sharing



# Prepare in advance for data sharing





- 1. Check sponsor agreements**
- 2. Incorporate data sharing into your consent forms**
- 3. Decide on potential data repositories**
- 4. Incorporate costs for data sharing into grant**
- 5. Write that data management and sharing plan!**

# 1. Check sponsor agreements

- **Are you required to share your data in a specific repository?**
- **Is data sharing restricted?**
- **Is there a mandated timeline for sharing?**
- **Are there recommended file formats or standards?**

## 2. Incorporate data sharing into your consent forms

- **Get consent from research subjects (in advance!) to share their data**
- **Use the IRB consent form templates**

### 3. Decide on potential data repositories

**Some questions to consider:**

- **What files will you need to submit?**
- **Are there required standardized file formats/metadata schemas?**
- **What is the turnaround time?**
- **Is there a cost for data sharing?**



## 4. Incorporate costs for data sharing into your grants

**Some things to consider when writing your grants:**

- **Who will be preparing your datasets for sharing?**
- **Will you need to have them certified de-identified?**
- **Is there a cost associated with your chosen data repository?**

**Note that the new NIH requirements permit all of these costs in grants. Expected to become norm for funders.**

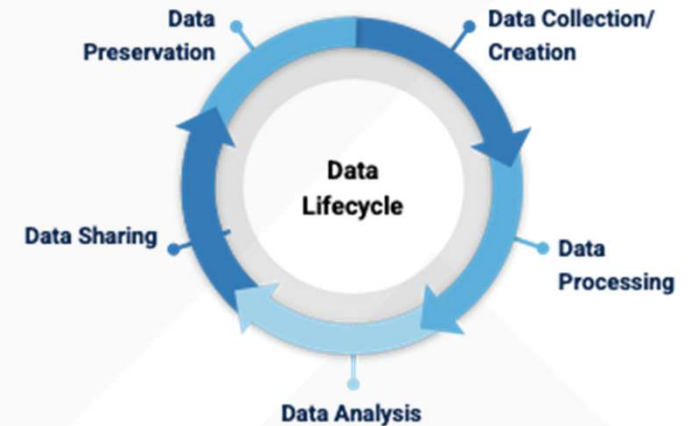
## 5. Write your data management and sharing plan

- **Summarize everything in your plan**
- **Follow the appropriate funder format**
- **Use the DMPTool - <https://dmptool.org>**
  - **Built-in templates to meet funder requirements**
  - **NIH policy specific templates**

# Wait, what's a **Data** **Management and Sharing Plan?**

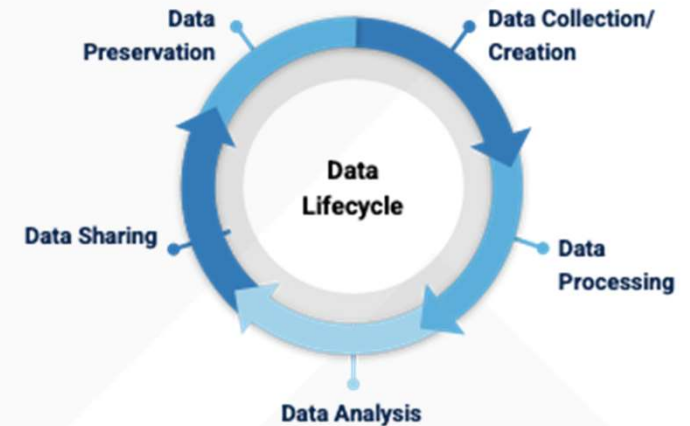
# Data Management and Sharing Plan (DMSP)

- A document that addresses how you will manage and secure your data throughout the lifecycle of a research project.
- Is a *required* document for grants and can be a *living* document for research planning purposes.



# OK, but what makes a *good* DMSP?

- A good DMSP has a clear, organized and effective system to manage data throughout the project.
- It includes plans for the data after the research is complete.
- The most important components of most required plans are:
  - Data sharing and data preservation.





**Follow your data management and sharing plan!**



1. **Prepare your data for sharing**
2. **Upload to your chosen repository and get your DOI or accession number**

# 1. Prepare your data for sharing

- **Make sure your data is properly de-identified to protect patient privacy**
  - **PHI needs to be removed for data to be de-identified**
- **Working with qualitative data? Don't forget about indirect identifiers. Learn more about qualitative data de-identification [with this guide.](#)**
- **Have questions about PHI and data privacy? You can always reach out to the [UC San Diego Health Sciences Office of Privacy and Compliance](#)**



## Other things to remember

- **Don't forget to package all the necessary documentation with your data!**
  - **What about those data dictionaries, analysis scripts, data guides, interview protocols, etc?**
- **Open/common data formats are best. Try saving files as CSVs or PDFs rather than proprietary software formats.**

## 2. Upload your data to your chosen repository

- **Once your data is submitted you will receive a DOI (digital object identifier)**
- **These can be used by other researchers to cite your data.**
- **Link them in your paper and grant progress reports!**

### **Example data citations:**

Krauth, David; Apollonio, Dorothy (2015), Data from: Accuracy of popular media reporting on tobacco cessation therapy in substance abuse and mental health populations, Dryad, Dataset, <https://doi.org/10.5061/dryad.j646b>

Wilson, C. (2012). Identify polymorphisms associated with risk for the development of myopericarditis following smallpox vaccine [Data set]. ImmPort. <http://doi.org/10.21430/M389L5FQCN>

*Whew!*

Who can help with all this?

# Library: Research Data Curation Services

- **The Library's Research Data Curation Program offers free services to help with:**
  - **Finding data standards**
  - **Selecting data repositories**
  - **Planning for data sharing**
  - **Minting DOIs**

<https://library.ucsd.edu/research-and-collections/research-data/index.html>

# Library: Scholarly communications services

- **Publishing opportunities & UCSD journal editor resources**
- **UC Transformative Publish & Read Agreements**
- **Copyright & author rights**
- **Open educational resources/open textbooks (OER)**
- **Research impact & scholarly identity**
- **Open access advocacy and guidance**

<https://ucsd.libguides.com/ScholComm>

# Campus Research IT Services

## Support for many research services, e.g.:

- **Assistance with compute resources**
  - **UC San Diego Research Cluster**
  - **Cloud Onboarding and Integration**
  - **National resources**
- **Data management and storage**
  - **Universal Scale Storage (USS) Brokering Program with SDSC**
  - **Globus Data transfer**
- **Staff to help support systems integration**

# Contact Us

## **UC San Diego Library Research Data Curation Program:**

- <https://library.ucsd.edu/research-and-collections/research-data>

## **Scholarly Communications team:**

- <https://ucsd.libguides.com/ScholComm>

## **Research IT team:**

- <https://research-it.ucsd.edu>