

# Core Concepts of Research Data Management

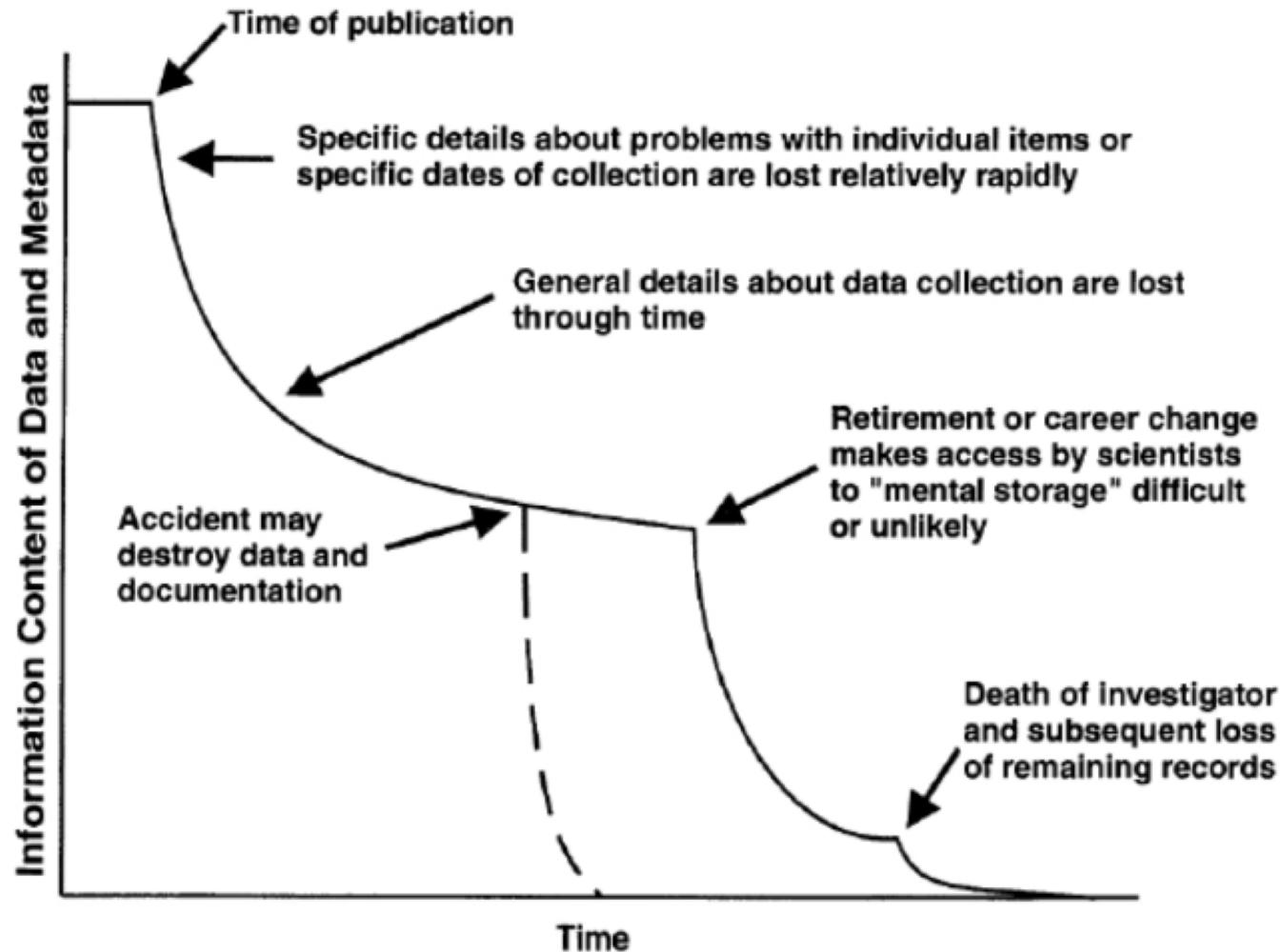
Yale UNIVERSITY LIBRARY

*Harvey Cushing/John Hay Whitney Medical Library*

# Why should you manage your data?

- Optimize space
- Save time
- Avoid mistakes
- Improve collaboration
- Verify the integrity of your data
- NIH and NSF requirements

# Research Data Entropy



# Research Data

Raw data, summary data, & computational effort

Yale UNIVERSITY LIBRARY

*Harvey Cushing/John Hay Whitney Medical Library*

# Data Management

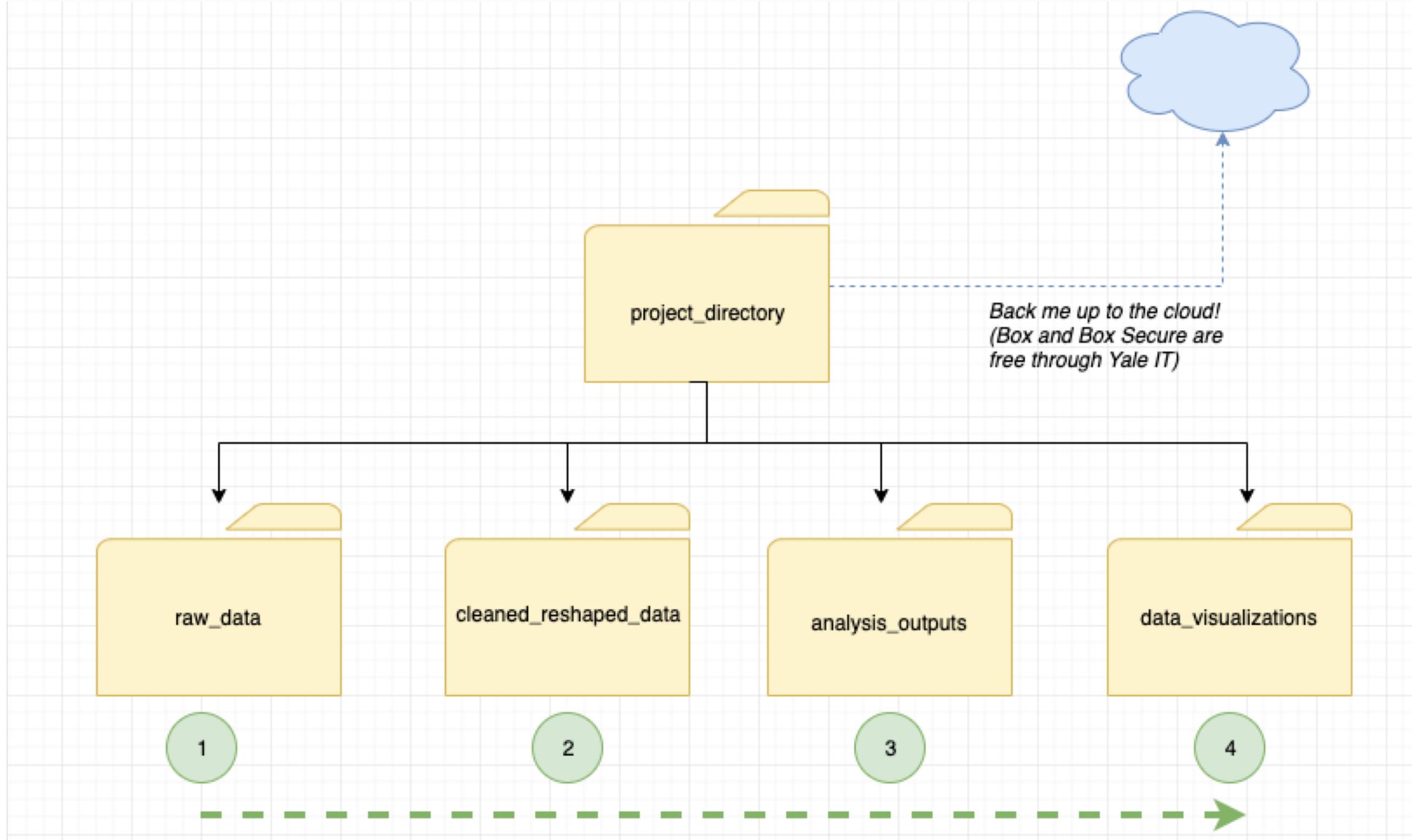
Data organization, backup, storage & collaboration

Yale UNIVERSITY LIBRARY

*Harvey Cushing/John Hay Whitney Medical Library*

# Folder organization

- Organize your folders in a way that makes sense to you
- Folder structures should be wider than they are deep
- Consider using folder organization template
- Documentation can capture organization rules



# Naming files

- File names should be consistent
- Short but descriptive (<25 characters)
- Use date format ISO 8601: YYMMDD
- Include a version number
- Document your file naming rules

# Spreadsheet Organization

- Columns contain variables (the measured elements)
- Rows contain observations
- Each cell should contain only one element
- Keep “master copies” of your raw data (create new Excel tabs or files as necessary)

# Data Backups

- Data should be stored in multiple locations to help ensure its preservation
- Code42 (formerly called CrashPlan) is a Yale ITS backup service that allows data to be backed up automatically and continuously with no user intervention required

# Version control

Version control allows you to see the change history of a file, and to restore a file to an previous iteration.

1. Manual version control
  - Add dates and/or v1, v2, final, draft to file name
  - Change log
2. Software with version control capabilities
  - Box and Google Drive
3. Version control systems (VCS)
  - Git

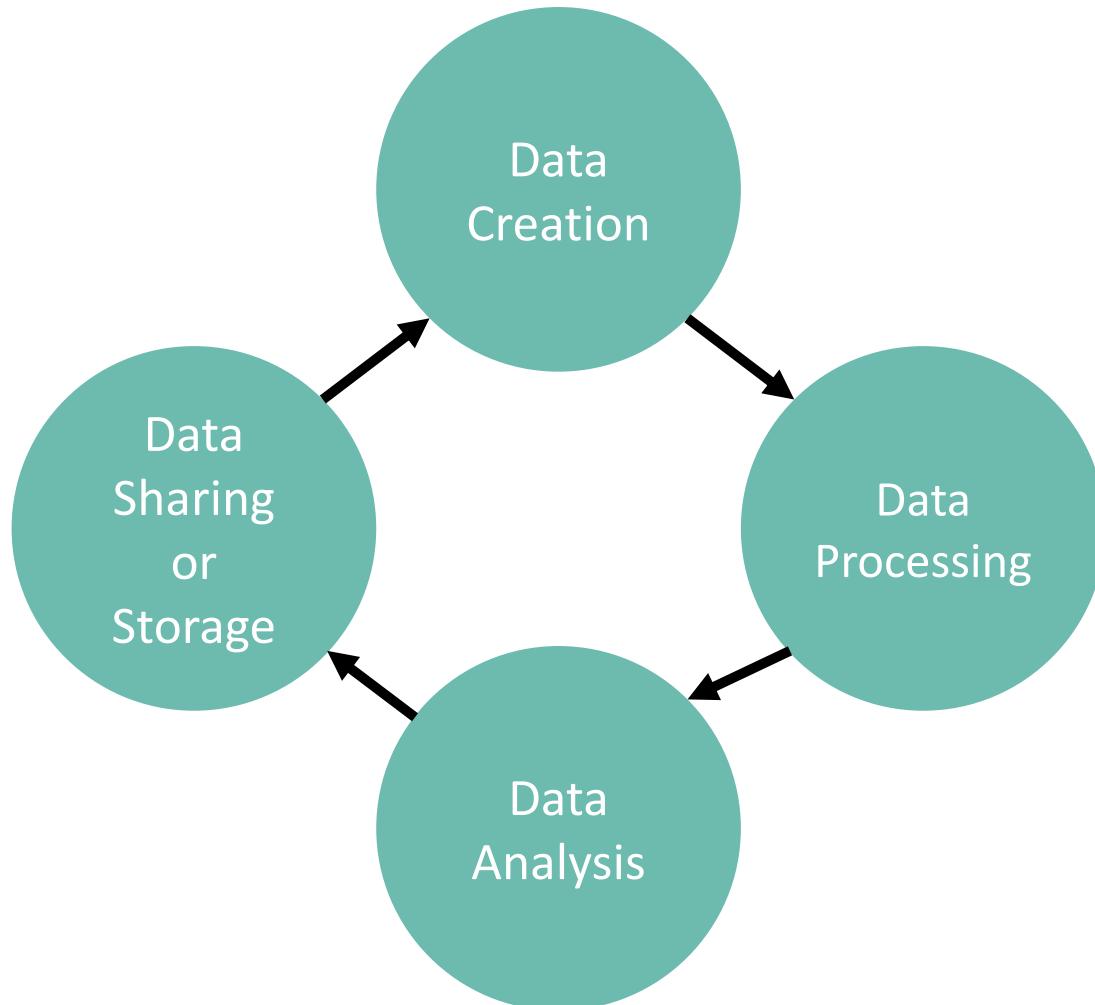
# Research Data Management

Leveraging data management for better research

Yale UNIVERSITY LIBRARY

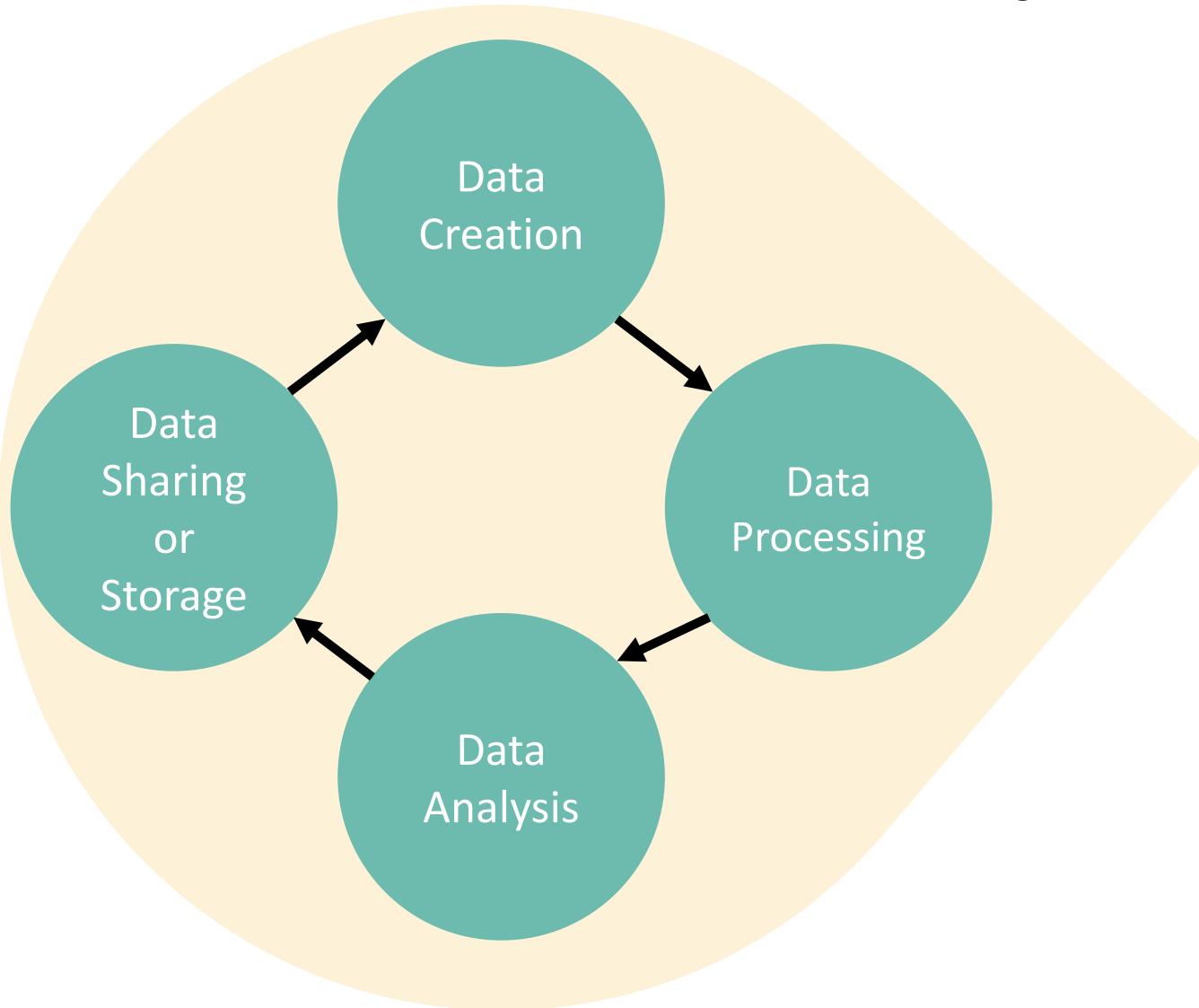
*Harvey Cushing/John Hay Whitney Medical Library*

# Research Data Lifecycle



The research data lifecycle model maps out a research project with a data-focused perspective.

# Research Data Lifecycle



- Data Security
- Documentation
- Operational Data Storage

# Data Storage

- Is your data low, moderate or high risk?
- How actively are you using your data?
- Is this your data or your lab's data?
- Is there a data usage agreement that states how data should be handled?

Data Storage Options @ Yale: [docs.ycrc.yale.edu/data](https://docs.ycrc.yale.edu/data)

# Data Security

- Protect data against loss due to accidental or technical problems
- Yale Data Classification Policy: low, moderate, and high risk data

[cybersecurity.yale.edu/protectyourdata](https://cybersecurity.yale.edu/protectyourdata)

# Data Sharing

## Transferring working data

- Yale Secure File Transfer  
(a.k.a. Filelocker)
- Yale Box and Secure Box



## Sharing data with a wider audience

- NIH data repositories
- Disciplinary data repositories
- Dryad @ Yale



# Research Data Services

@ the Cushing/Whitney Medical Library

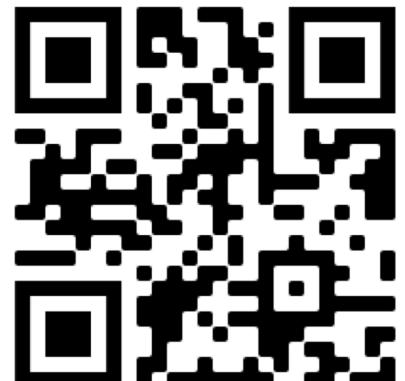
Yale UNIVERSITY LIBRARY

*Harvey Cushing/John Hay Whitney Medical Library*

# Classes @ the Cushing/Whitney Medical Library

- Research Data Documentation
- Introduction to R
- Working with Data
- Unix Shell: Working with Data

<https://library.medicine.yale.edu/research-data/classes-materials>



# LinkedIn Learning @ Yale

<https://your.yale.edu/yale-link/linkedin-learning>

The screenshot shows the LinkedIn Learning platform interface. At the top, there's a navigation bar with the LinkedIn Learning logo, a search bar containing "research data", and buttons for Home, In Progress, Saved, and Me. Below the search bar, a sidebar on the left encourages users to "Set a weekly goal" with a trophy icon and a "Set a goal" button. The main content area displays a list of search results under the heading "research data", which includes "qualitative research data analysis", "qualitative research data analysis coding", "research data management", "data research analyst", "research data science", "data research", "data analysis research", "secondary data research", and "quantitative data analysis academic research spss". At the bottom of this list is a "See all results for 'research data'" link. To the right, a sidebar titled "Saved" and "From your org" lists two courses: "R Statistics Essential Training" (5h 41m left) and "C Essential Training: 1 The Basics" (2h 16m left). The bottom right corner shows course thumbnails for "FEATURED COURSES".

research data

qualitative research data analysis

qualitative research data analysis coding

research data management

data research analyst

research data science

data research

data analysis research

secondary data research

quantitative data analysis academic research spss

See all results for "research data"

Home In Progress Saved Me

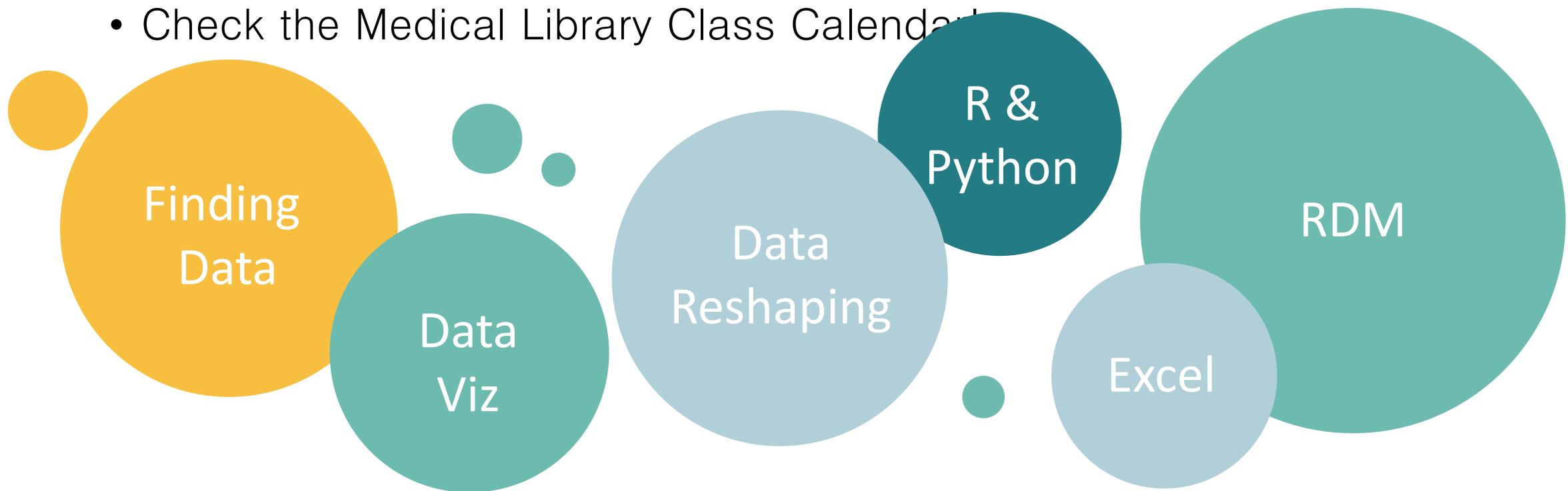
R Statistics Essential Training COURSE + 5h 41m left

C Essential Training: 1 The Basics COURSE + 2h 16m left

FEATURED COURSES

# Office Hours

- Drop-in and ask questions about research data, research data management and data visualization
- Schedule changes on a monthly basis
  - Check the Medical Library Class Calendar



# Find More Research Help

- Bioinformatics Services (Medical Library)
- Statistical Support Services (CSSSI)
- Yale Center for Research Computing (YCRC)
- Research at Yale ([research.yale.edu](http://research.yale.edu))
- Yale Center for Analytical Sciences (YCAS)
- Yale Center for Clinical Investigation (YCCI)



Contact  
[medicaldata@yale.edu](mailto:medicaldata@yale.edu)

Yale UNIVERSITY LIBRARY

*Harvey Cushing/John Hay Whitney Medical Library*