
DATA INTEGRITY: ACQUISITION, MANAGEMENT, SHARING, AND OWNERSHIP

RESPONSIBLE CONDUCT OF RESEARCH

LAURA WAUGH, MIS
TEXAS STATE UNIVERSITY LIBRARIES
LWAUGH@TXSTATE.EDU

DEFINING RESEARCH DATA

Recorded factual material commonly accepted in the scientific or scholarly community
as necessary to validate research findings.



Kate Laskowski

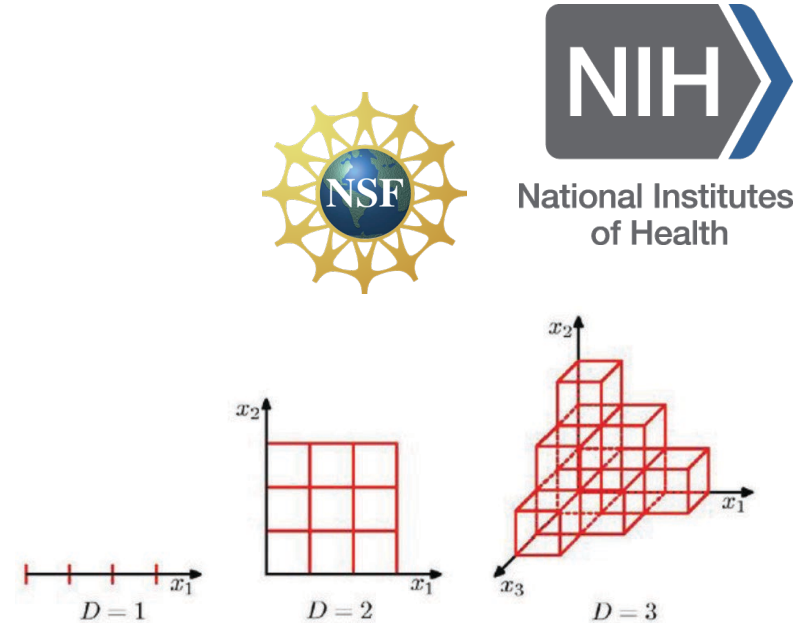
@KateLaskowski

I'm starting the year off with something I didn't expect to ever do: I'm retracting a paper. I recently discovered major problems in the raw data associated with it and so the results shouldn't be trusted.

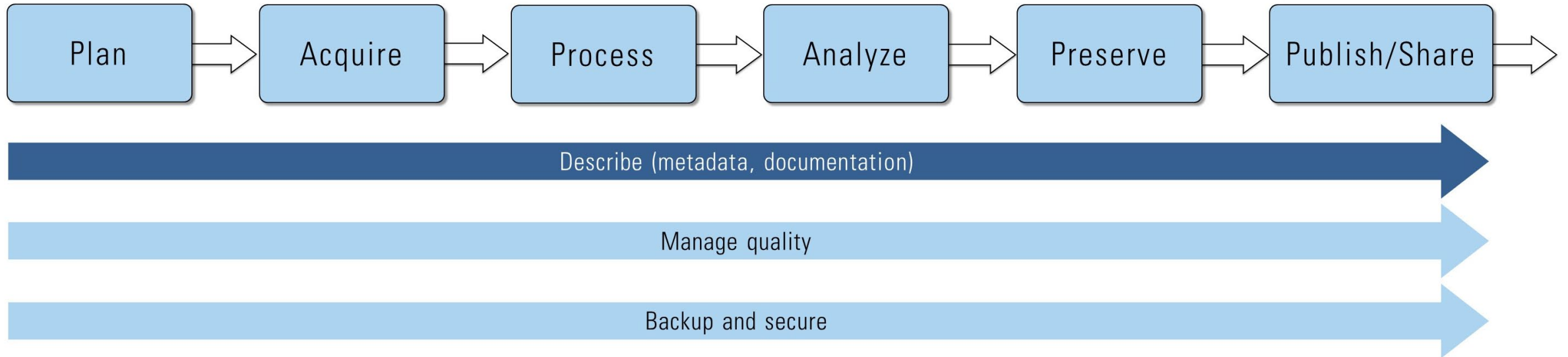
journals.uchicago.edu/toc/an/current...

WHY PLAN?

- Funder requirements
- Institutional policy
- Mitigate error and loss
- Avoiding unforeseen costs
- Be able to return the data
- Getting a handle on the complexity of data



MANAGEMENT OF RESEARCH OUTPUTS LONG TERM



BASIC ELEMENTS OF A PLAN

1. What are you creating / generating?
2. How is it securely handled during the project?
3. How is access and data integrity maintained long term?

Additional Details: Roles and responsibilities, systems used, documentation, security

GENERATING A FULL PICTURE OF THE DATA CHARACTERISTICS

#	Materials/Data Generated/Created	Format	Approx. Size	Shared	Privacy/IP needs

PLANNING AND WRITING RESOURCES

- Data Management Plan
- DMPTool
- Sample text & templates
- Login with TXST Email
- <https://dmptool.org/>



- 🌐 TXST Dataverse Repository
- 🌐 TXST Research Data Management Services
- 🌐 TXST Data Management Planning
- ✉ TXST Data Contact

PLANNING AND WRITING RESOURCES

Templates for TXST Researchers

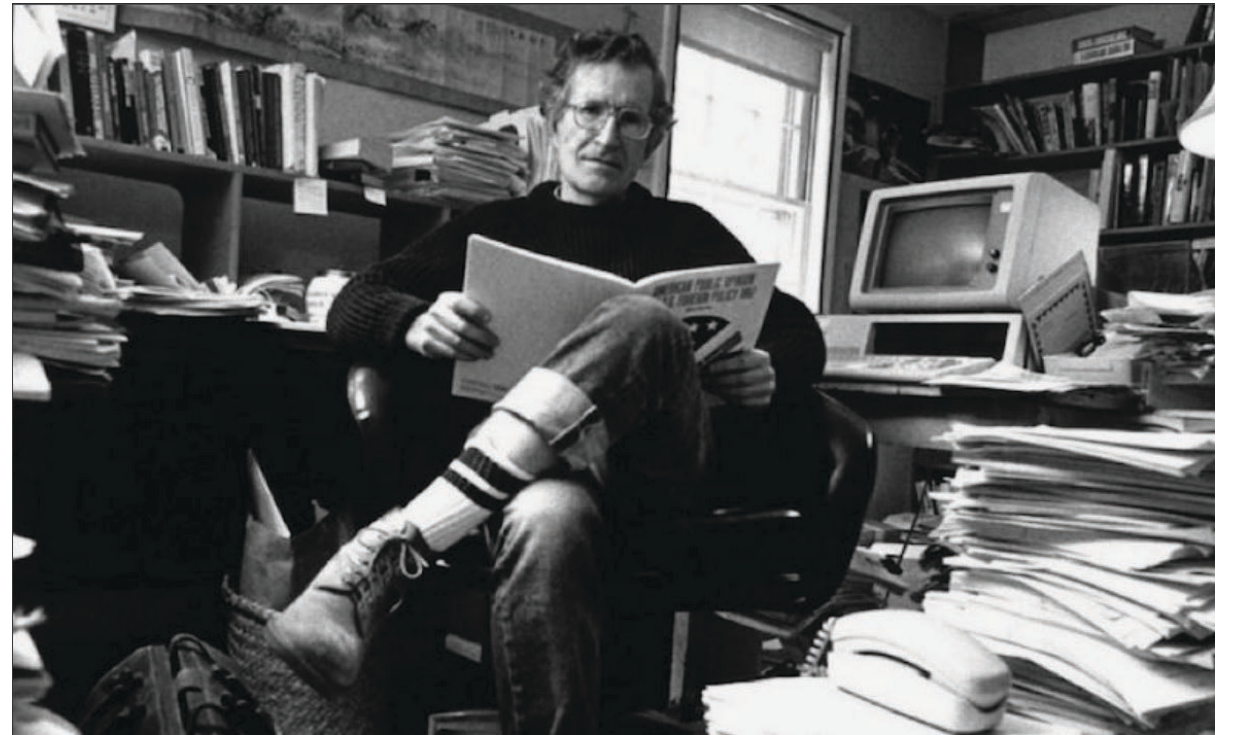
- Template Text: Data Sharing and Access
- Template Text: Metadata for Data Management
- <https://guides.library.txstate.edu/research-data/DMP>



FILE MANAGEMENT

- File naming conventions
- Version control
- Folder structure
- Workflows

(Document them!)



FILE MANAGEMENT

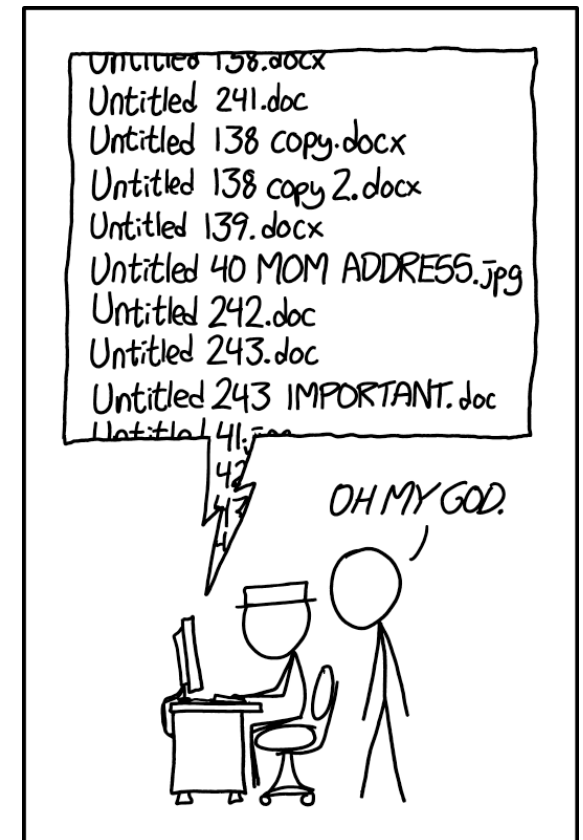
Suggestions

- Establish a file naming convention ***before*** you begin collecting data
- Ensure **everyone** on the team adheres to agreed file naming
- **Maintain** the organization across all folders, versions, devices, etc.
- *Document it!*

FILE MANAGEMENT

File Naming Convention Best Practices

- Project lead's last name or initials
- File creator's last name or initials
- Project name/acronym
- Date file created/generated (in YYYY-MM-DD format)
- Version number (with leading zeroes)



PRO TIP: NEVER LOOK IN SOMEONE ELSE'S DOCUMENTS FOLDER.

FILE NAMING CONVENTIONS EXAMPLES

- 2022-USDA-G0625-water-samples
 - 2022-06-USDA-G0625-MasterData-v0001
 - 2022-07-USDA-G0625-sampleTest-v0001
 - 2022-07-USDA-G0625-sampleTest-v0002
- 2023-07-DOE-Gonzalez_DesignData
 - 2023-07-08-DOQ-LG-water_sample-smr
 - 2023-07-07-DOA-LG-water_sample-smr
 - 2023-07-06-DOA-LG-water_sample-smr

DOCUMENTATION & STANDARDS

Descriptive Metadata

- Description of content
- Relationships between files
- Protocols, etc.

A close-up photograph of a nutrition label, likely for a food product like peanut butter. The label is white with black text and is slightly tilted. It shows the following information:

Amount Per Serving		Calories from Fat
		% Daily Value
Calories	25	
Total Fat	0g	0%
Saturated Fat	0g	
Trans Fat	0g	
Cholesterol	0mg	0%
Sodium	340mg	
Total Carbohydrate		
Dietary Fiber		
Sugars		

DOCUMENTATION & STANDARDS

How?

- Codebooks
- Readme files
- Tables of contents
- Computer - generated

This readme.txt file was generated on <YYYYMMDD> by <Name>

GENERAL INFORMATION

1. Title of Dataset

2. Author Information

Principal Investigator Contact Information

Name:
Institution:
Address:
Email:

Associate or Co-investigator Contact Information

Name:
Institution:
Address:
Email:

Associate or Co-investigator Contact Information

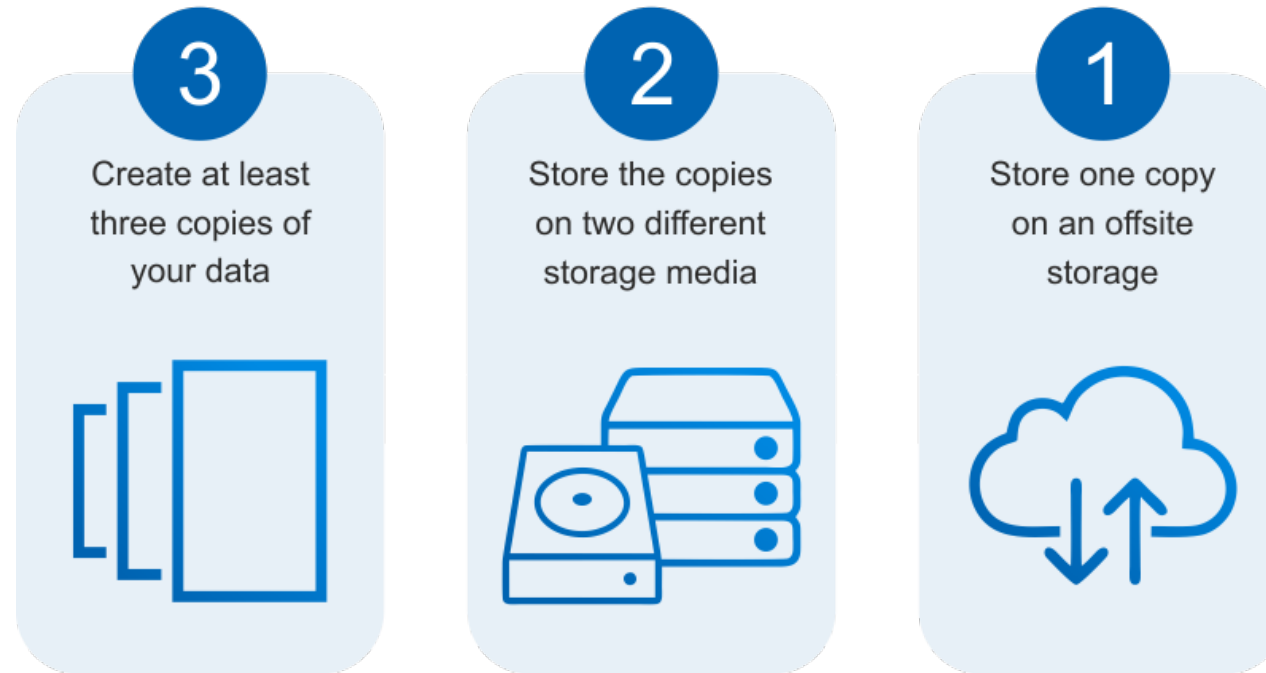
Name:
Institution:
Address:
Email:

3. Date of data collection (single date, range, approximate date) <suggested format YYYYMMDD>

4. Geographic location of data collection (where was data collected?):

5. Information about funding sources that supported the collection of the data:

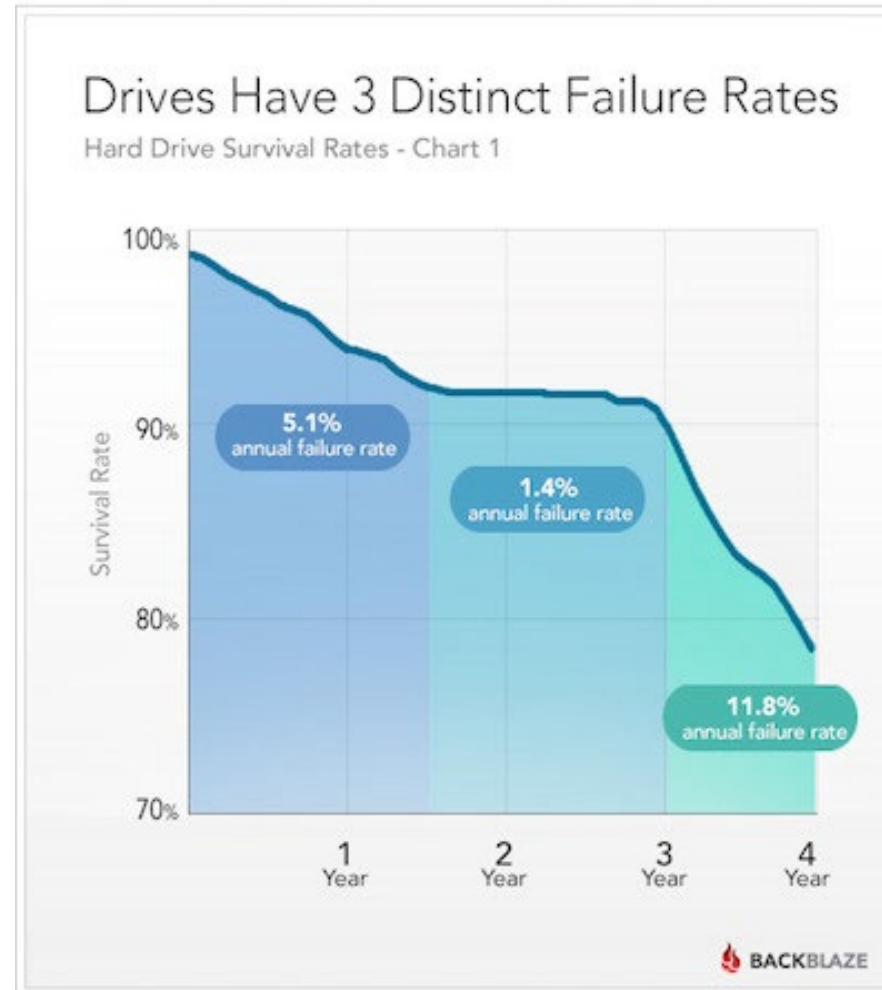
STORAGE AND BACKUP 3-2-1 RULE



 **Handy Backup**

scale, sensitivity, sustainability

WHAT COULD GO WRONG?



CONSIDERATIONS FOR BACK-UP

Who? How?
How often?

```
graph TD; A["Who? How? How often?"] --- B["Automate where possible."]; A --- C["Test your backups."]; A --- D["Pay close attention to raw data."]; A --- E["Keep an eye on the external."];
```

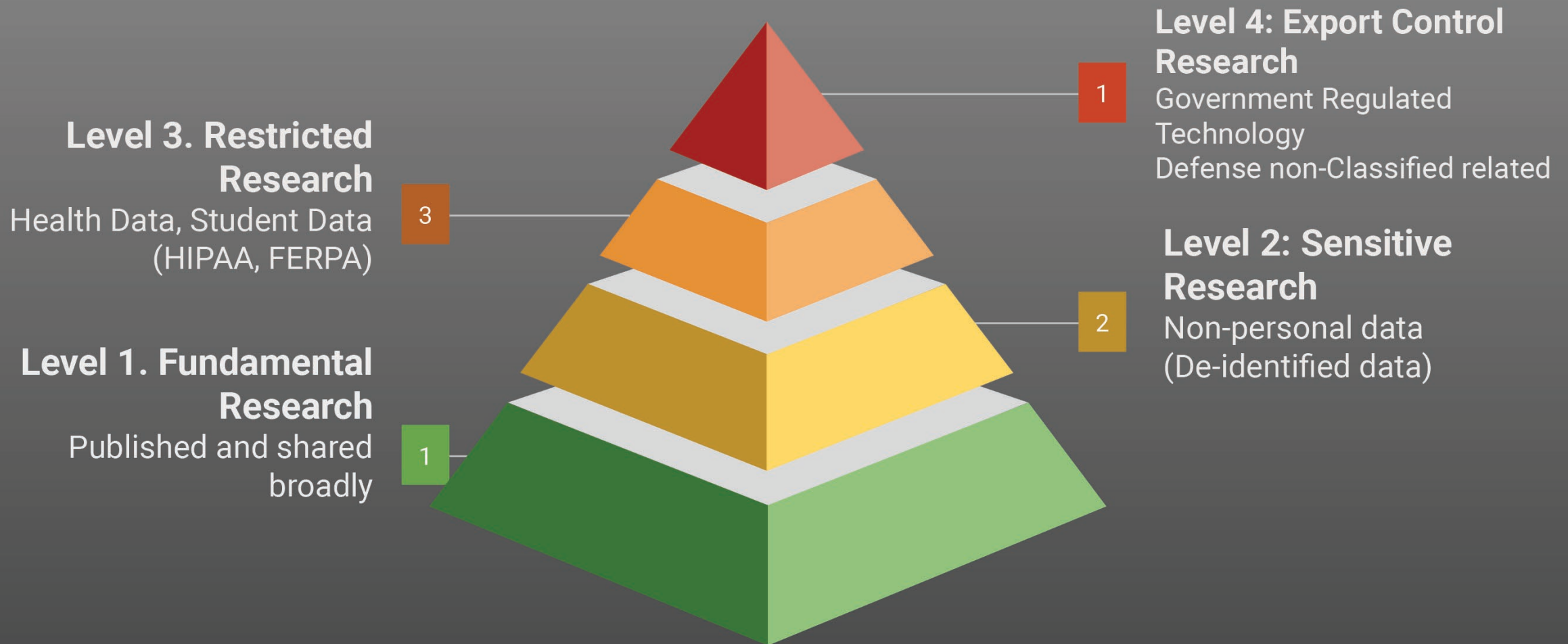
Automate where possible.

Test your backups.

Pay close attention to raw data.

Keep an eye on the external.

SECURITY MEASURES: KNOW YOUR DATA...



MEASURES TO PROTECT PRIVACY

- De-identification
- Data Use Agreements
- Controlled Access
- Encryption

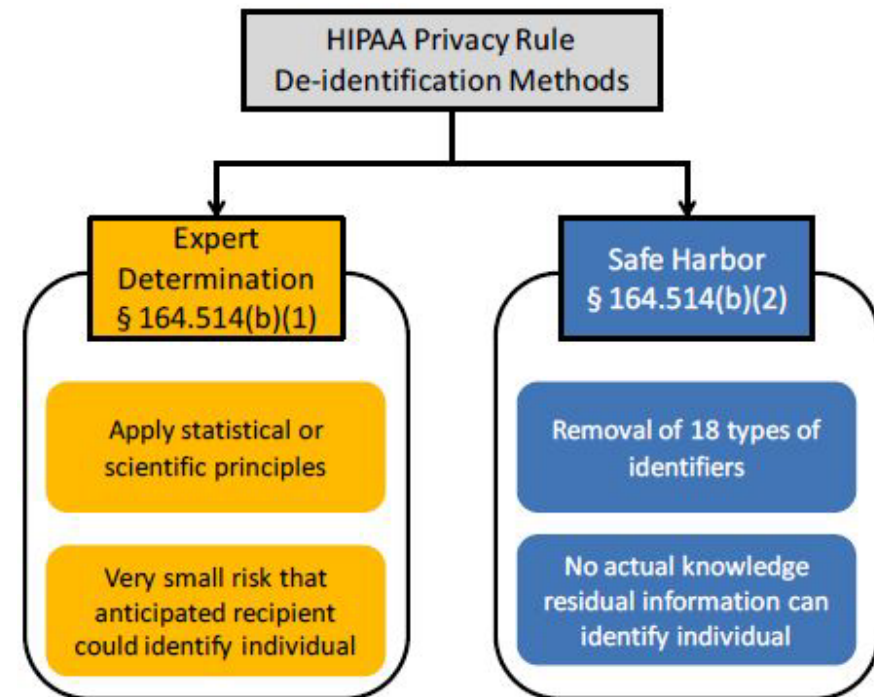


Figure 1. Two methods to achieve de-identification in accordance with the HIPAA Privacy Rule.

POST PROJECT ELEMENTS

- Curation: selecting and documenting
- Archiving: securing in a space long term
- Repository: a secure space to archive
- Preservation: actions for sustaining integrity

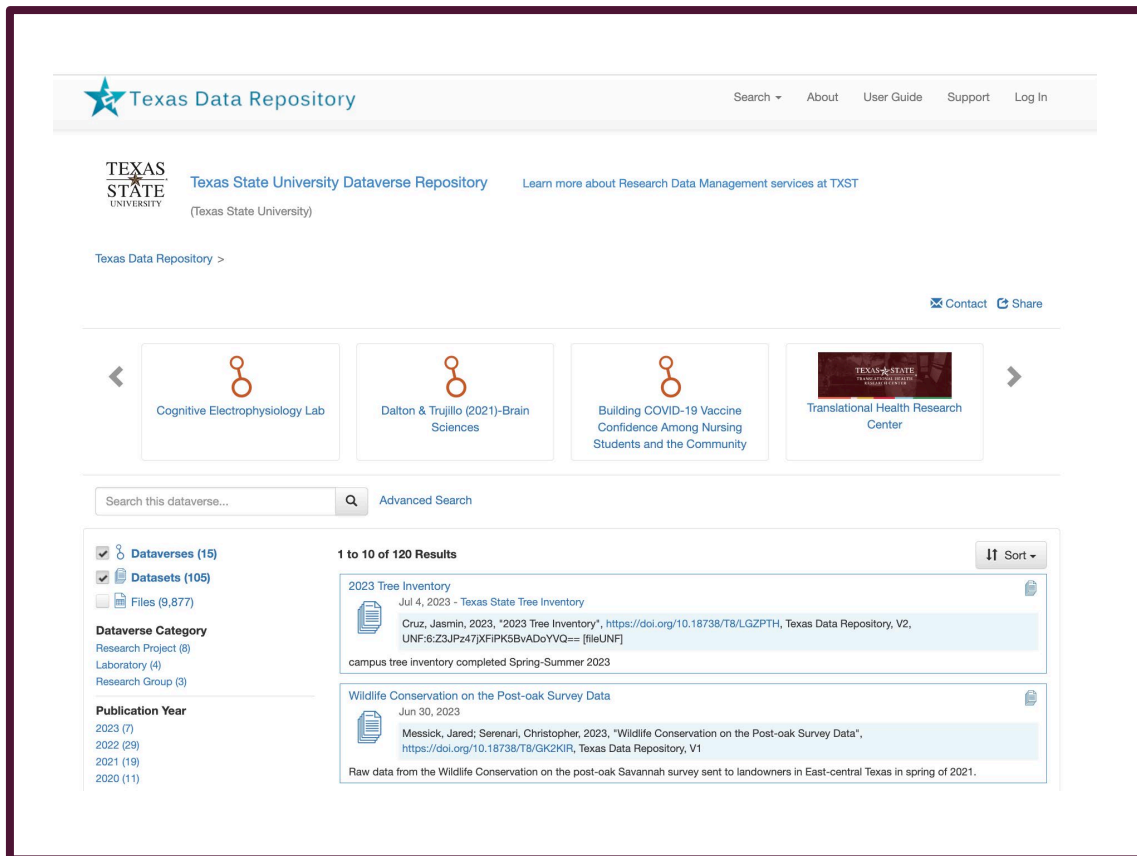
TEXAS STATE DATAVERSE REPOSITORY

TXST Dataverse

- <https://dataverse.tdl.org/dataverse/txst>

General Info

- Login with TXST NetID and password
- Published datasets shared in compliance with grant funders
- Version control, metadata, DOIs
- Share during and after data collection



QUICK RECAP

- Know your data
- Document *before, during, and, after*
- Organize and document consistently
- Ensure you can describe, replicate, and share

THANK YOU!

Get in touch any time

- Laura Waugh, lwaugh@txstate.edu
- Xuan Zhou, x_zhou@txstate.edu

SUPPLEMENTAL INFO & ACTIVITIES

Who owns the data?

It may depend on who sponsors the research.

As employees of the university, you are working for hire for the university, which, in most cases, owns the rights to the data. In federally sponsored research, the university owns the data but allows the principal investigator on the grant to be the steward of the data. The PI takes responsibility for the collection, recording, storage, retention, and disposal of data.

SUPPLEMENTAL INFO & ACTIVITIES

How does a researcher ensure reliable methodology?

Before a study is undertaken, researchers should:

- Have a testable hypothesis.
- Design a study to eliminate or remove potential bias.
- Choose methods of analysis before the study is undertaken.
- Enlist statistical expertise.
- Establish a plan for researchers' involvement.
- Follow appropriate legal guidelines for working with animals and humans.

SUPPLEMENTAL INFO & ACTIVITIES

Who should be listed and included in the grant application?

Everyone involved!

- Consultants in the research gathering, data analysis, methodology, and other steps.
- Post-doctoral, graduate, undergraduate students, and university staff that will be assisting in the grant process.