

'Good Enough' Practices to Build Better Research Data Managements Habits

Alaina Pearce

Provost-Endorsed Program in Research Data Stewardship

Build practical research data stewardship skills you can apply directly to your research.



What You'll Learn



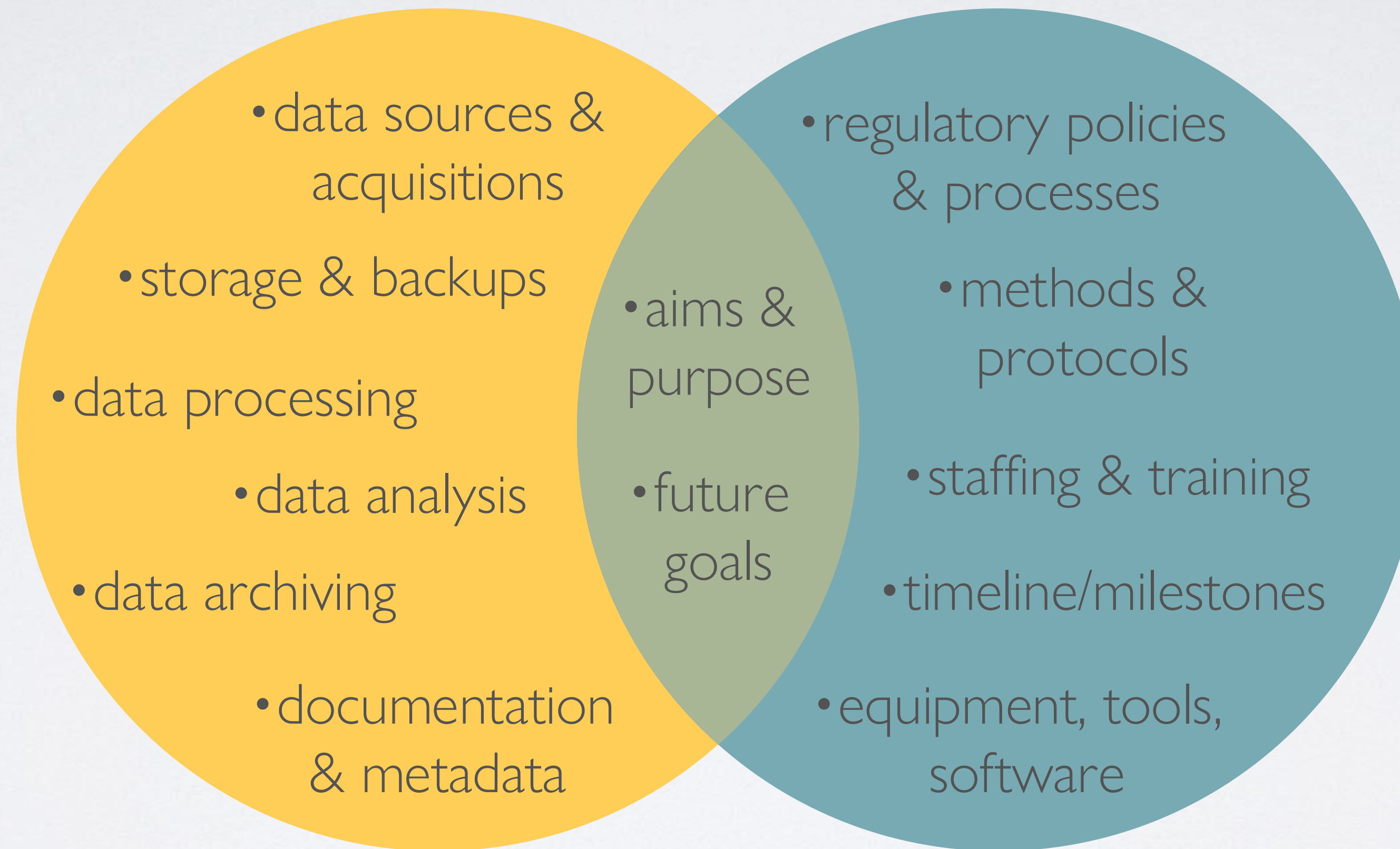
Program Components



Project vs Data Management

Data Management

Project Management



Goal: extract meaningful insight and information

Goal: meet project goals within set timelines

With what
time???



'Good Enough'

- (relatively) low effort
- shallow learning curve
- beneficial to current and future you
- increases 'openness' of research



Project Lifecycle

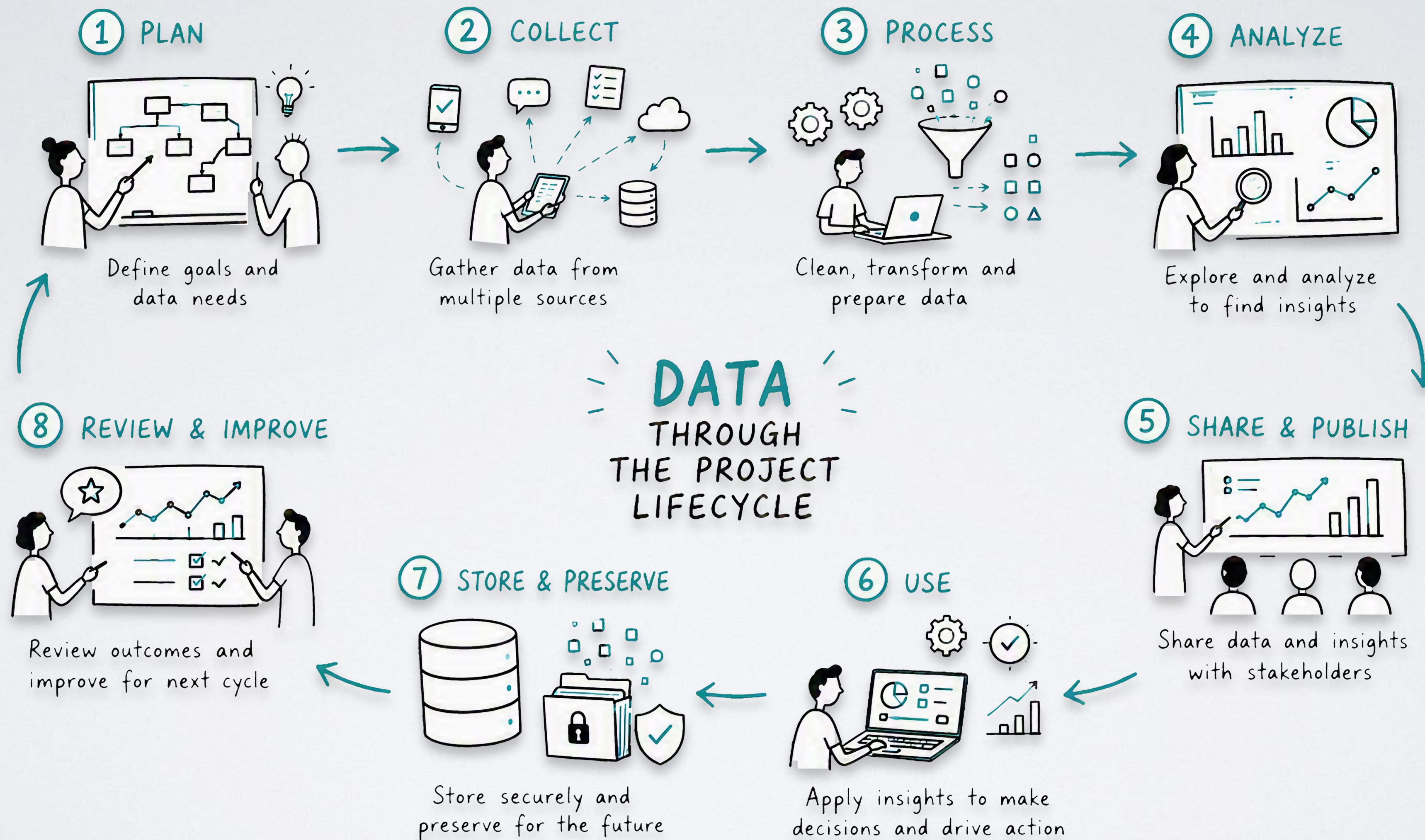


Image is AI generated and inspired by *The Turing Way* project illustration by Scriberia. Used under a CC-BY 4.0 license. DOI:

[10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807).

Project Lifecycle

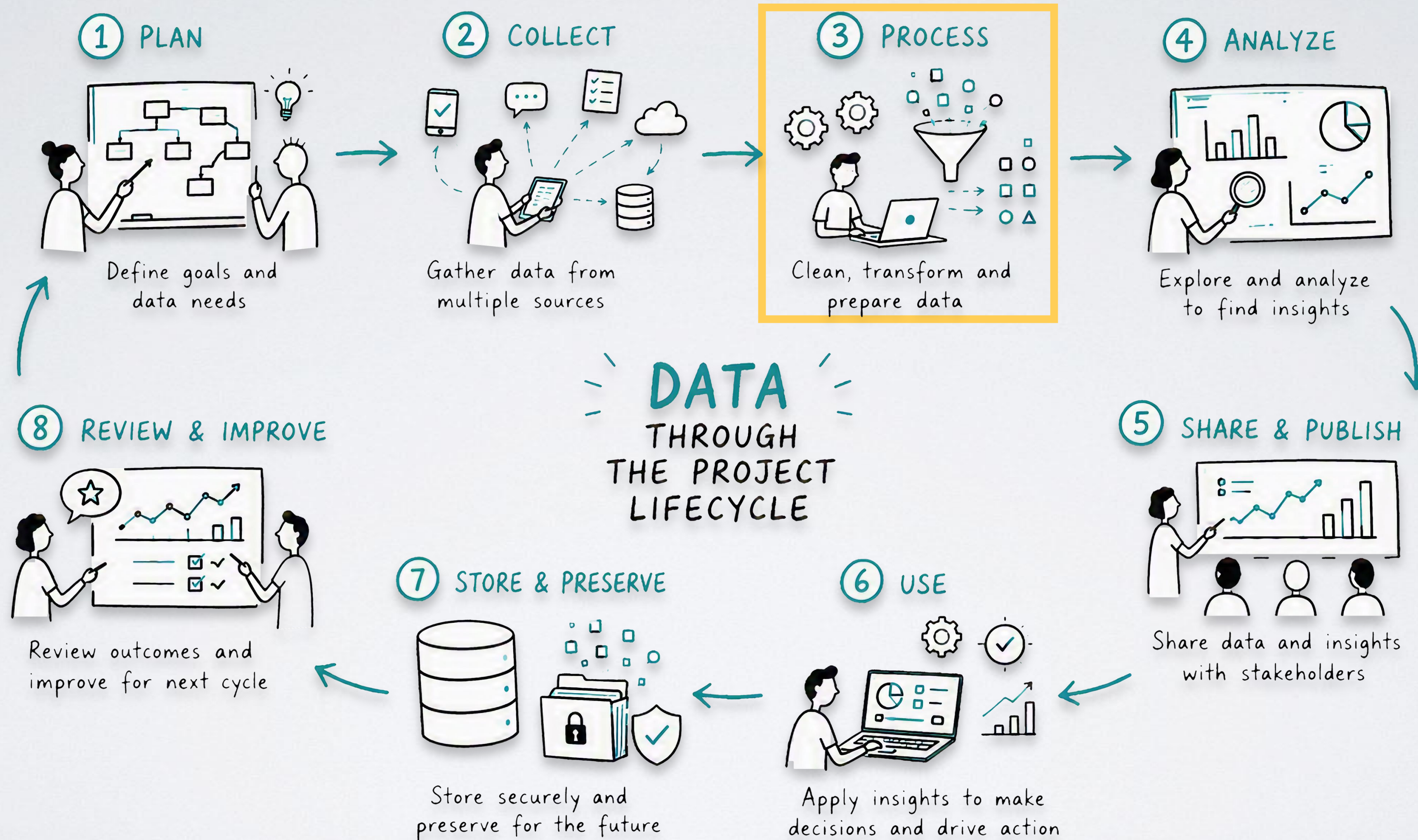
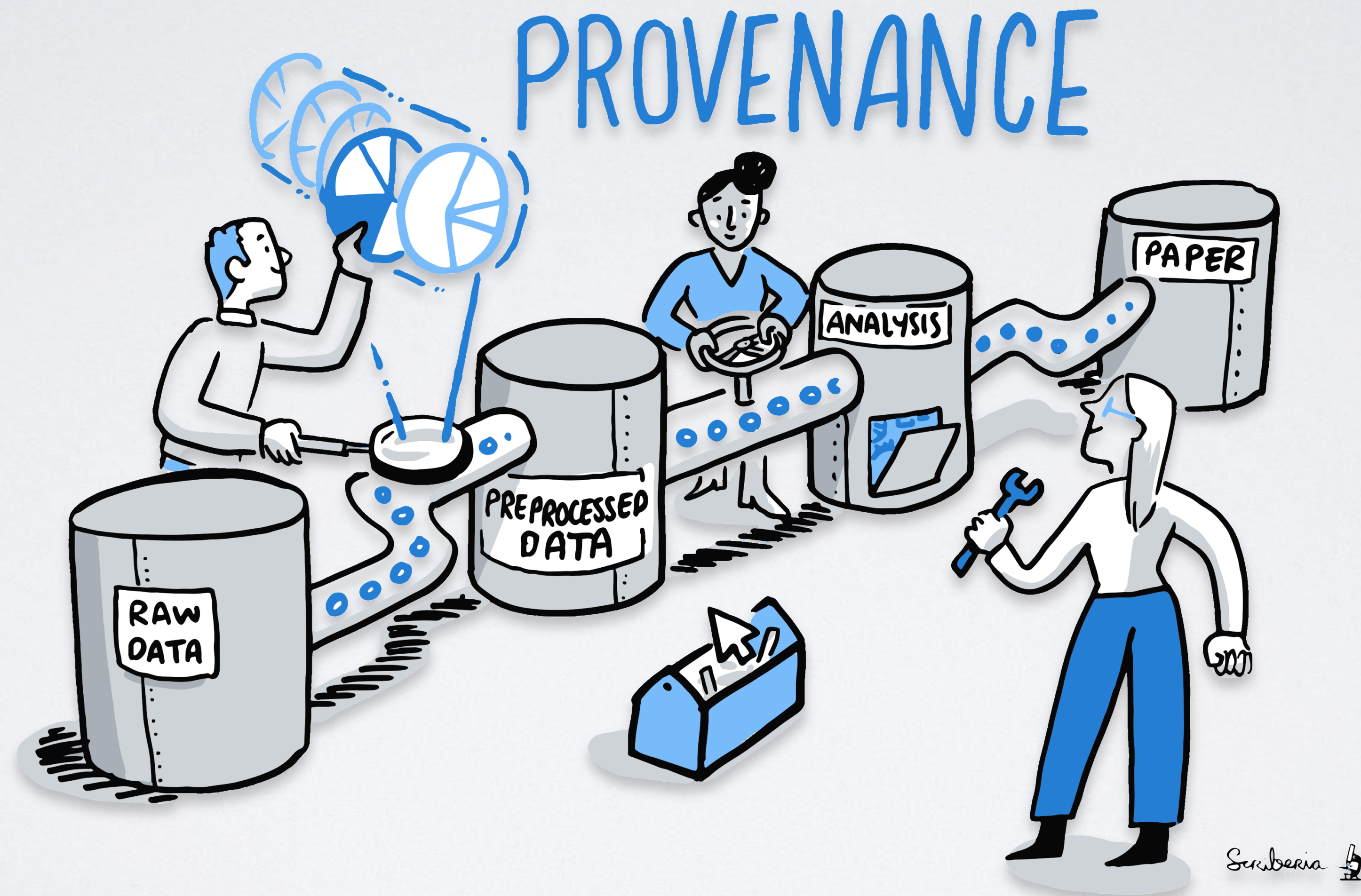


Image is AI generated and inspired by *The Turing Way* project illustration by Scriberia. Used under a CC-BY 4.0 license. DOI:

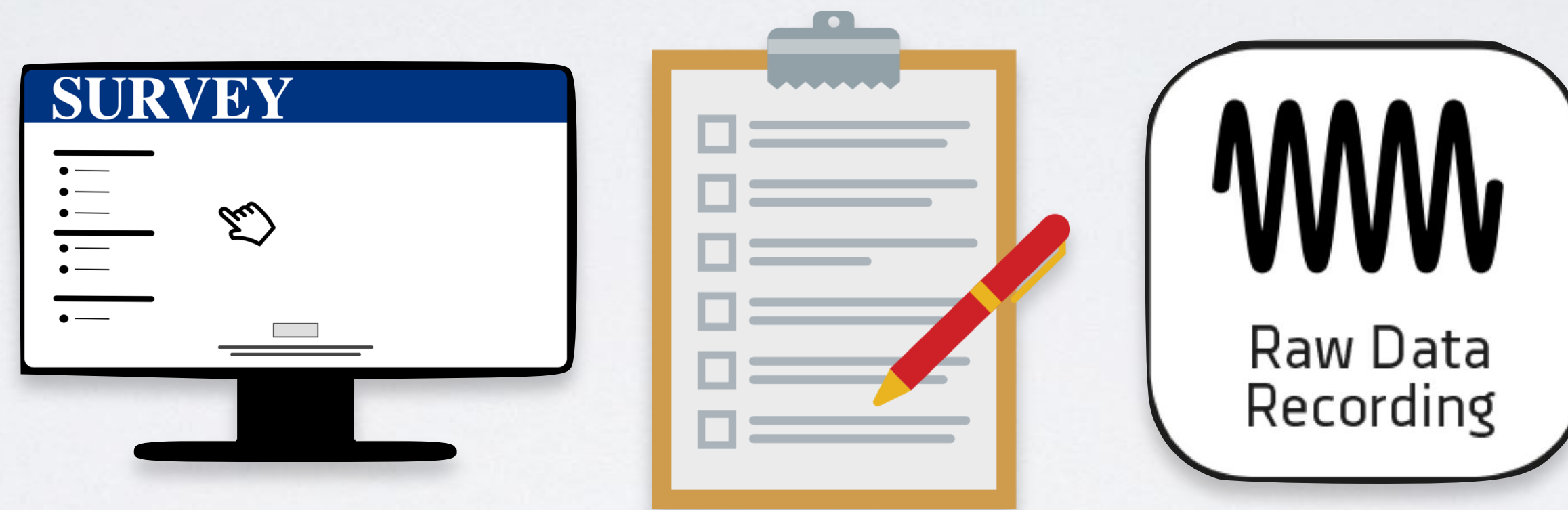
[10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807).

Data Processing Pipelines



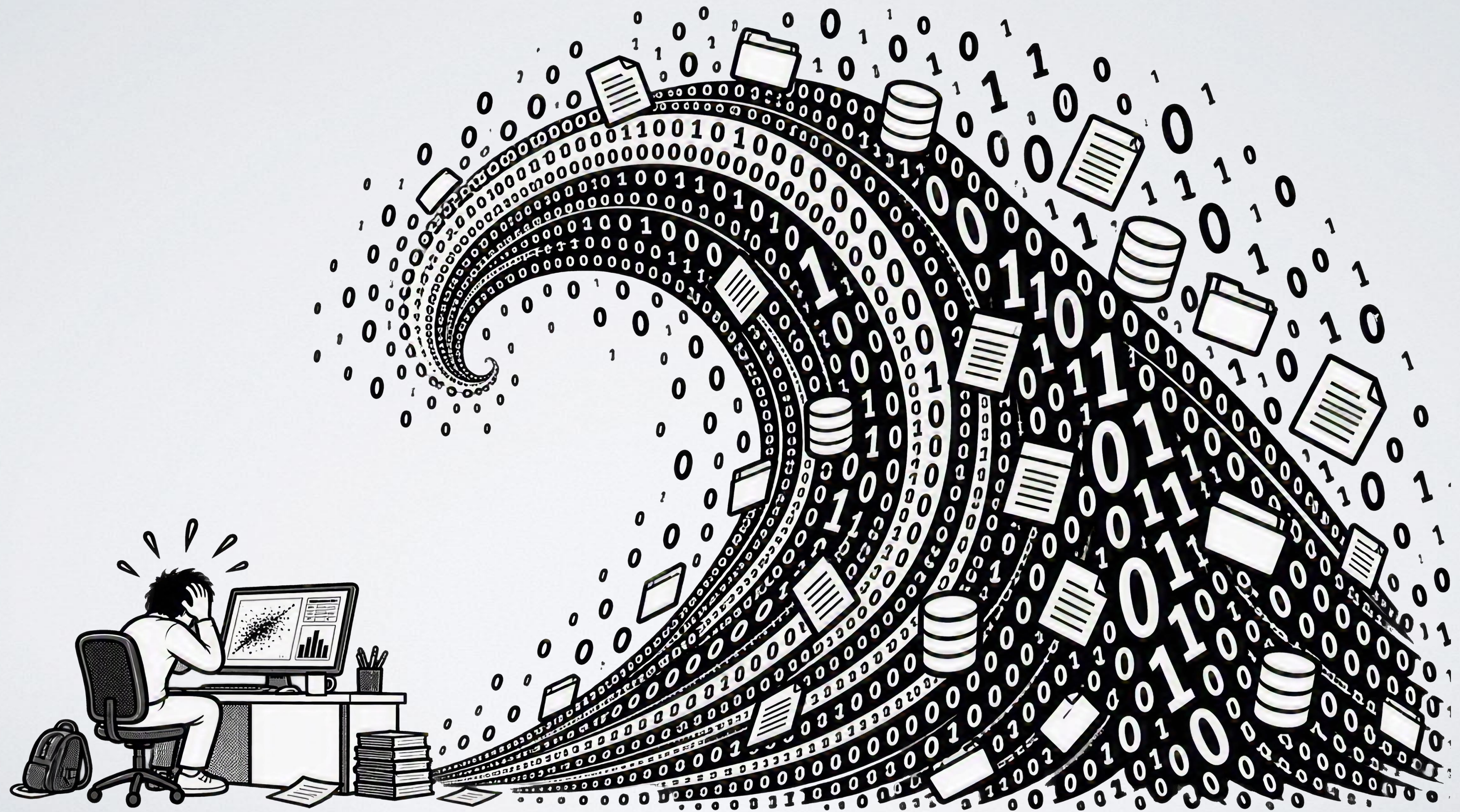
1. Preserve Raw Data

Raw Data: data as it was originally collected



Save in data in its original form and DO NOT alter or 'improve' it

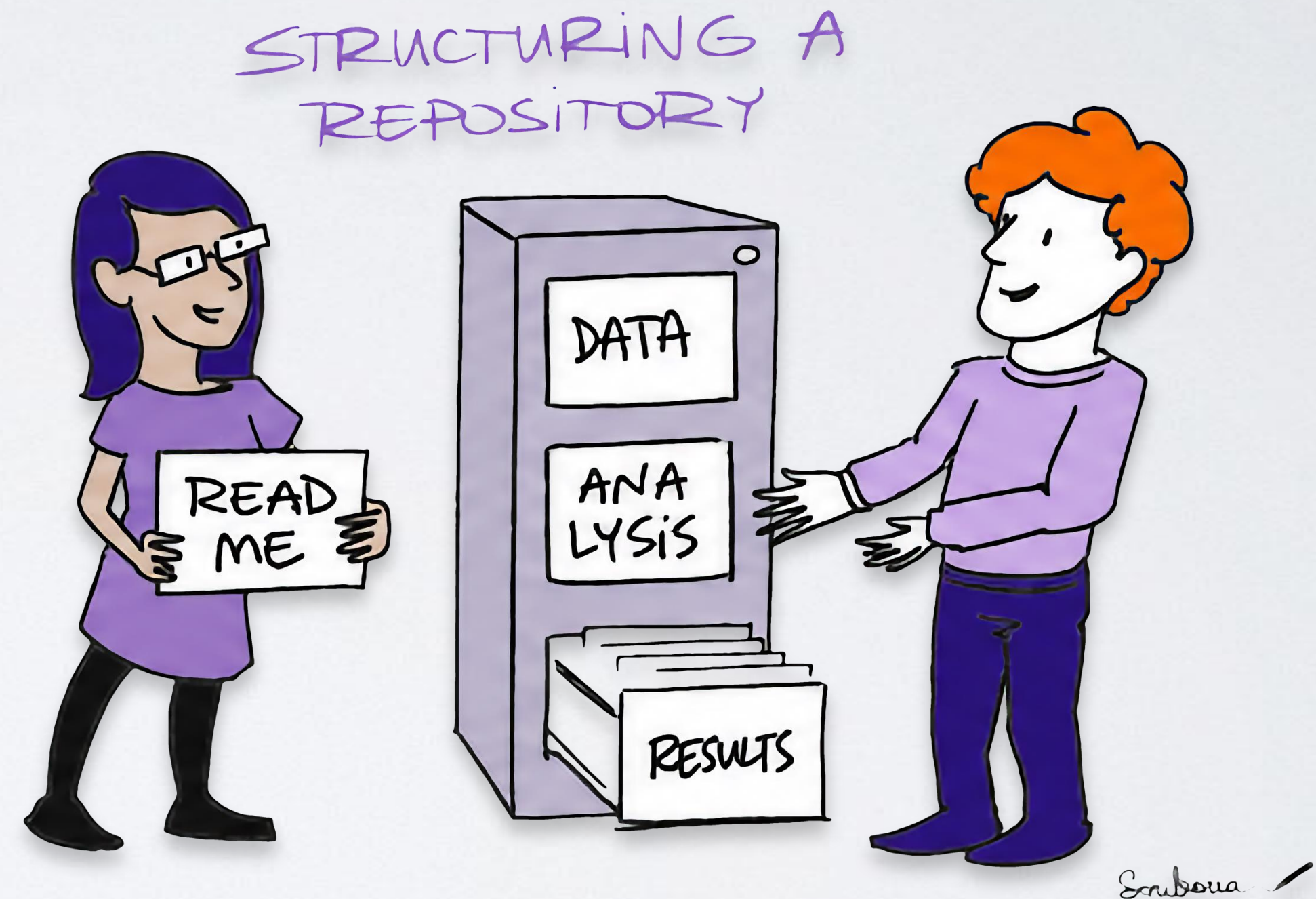
Data Tsunami



2. Create a Central Hub

Directory Structures: organization of files into a hierarchical structure

- Create a directory for each project
- Use a consistent structure
- Separate data management from project management

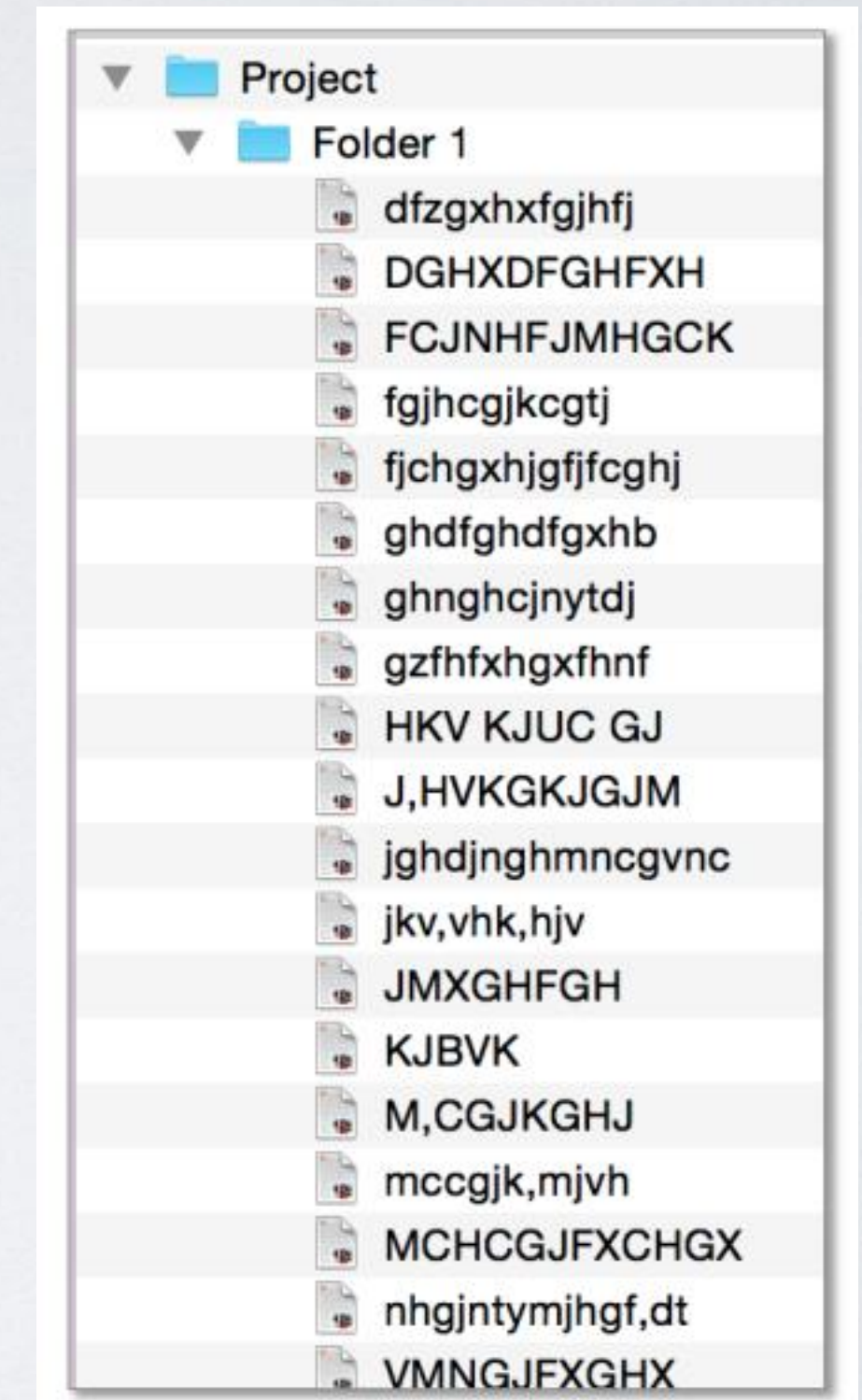
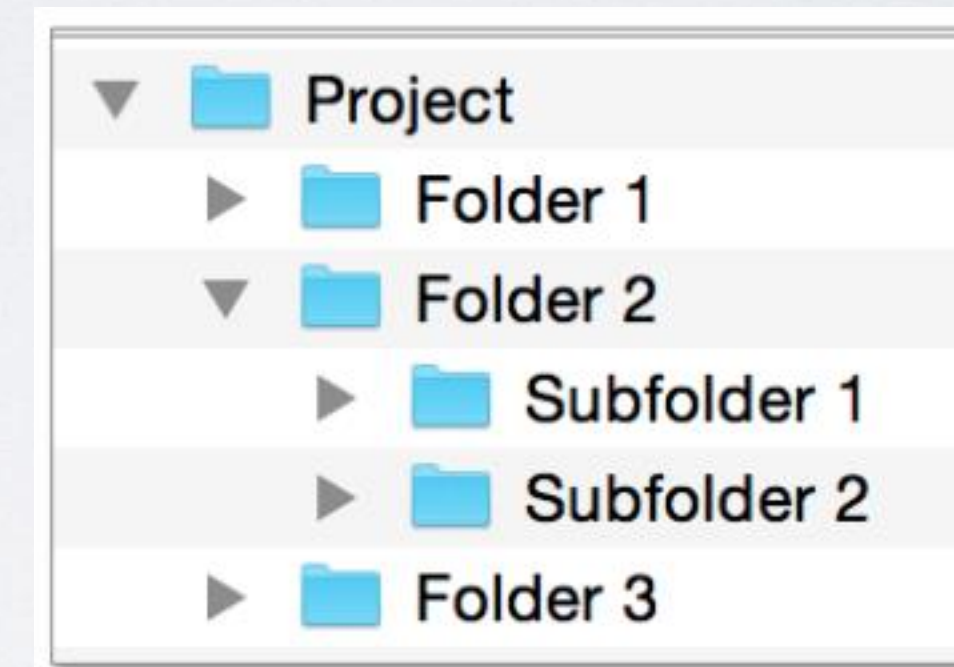


The Turing Way project illustration by Scriberia. Used under a CC-BY 4.0 license. DOI: [10.5281/zenodo.3332807](https://doi.org/10.5281/zenodo.3332807).

2. Create a Central Hub

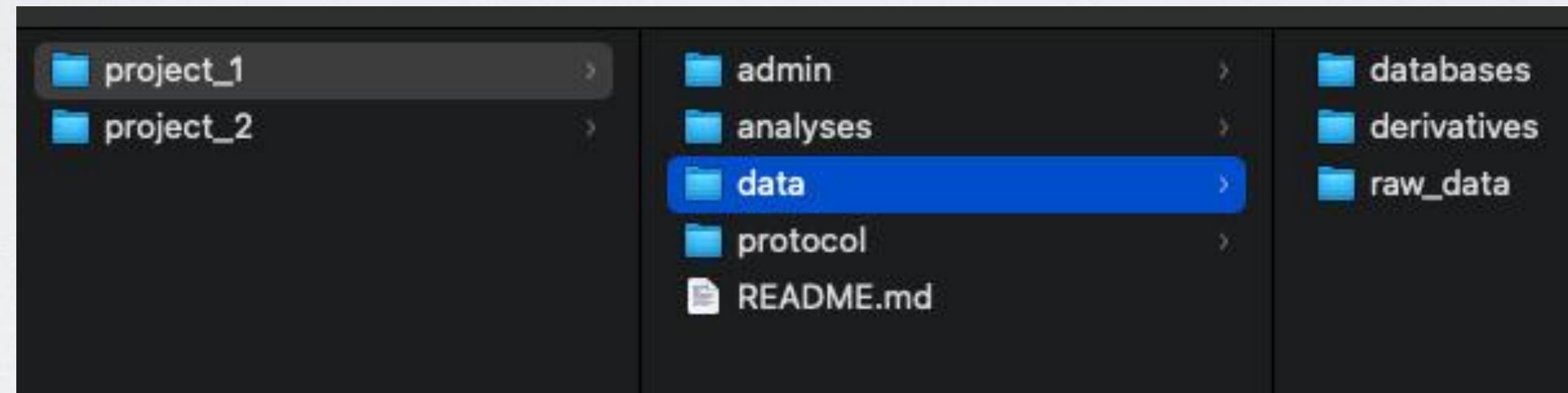
Directory Structures: organization of files into a hierarchical structure

- Create a directory for each project
- Use a consistent structure
- Separate data management from project management
- Keep subfolder categories narrow to limit number of files in each one



2. Create a Central Hub

Directory Structures: organization of files into a hierarchical structure

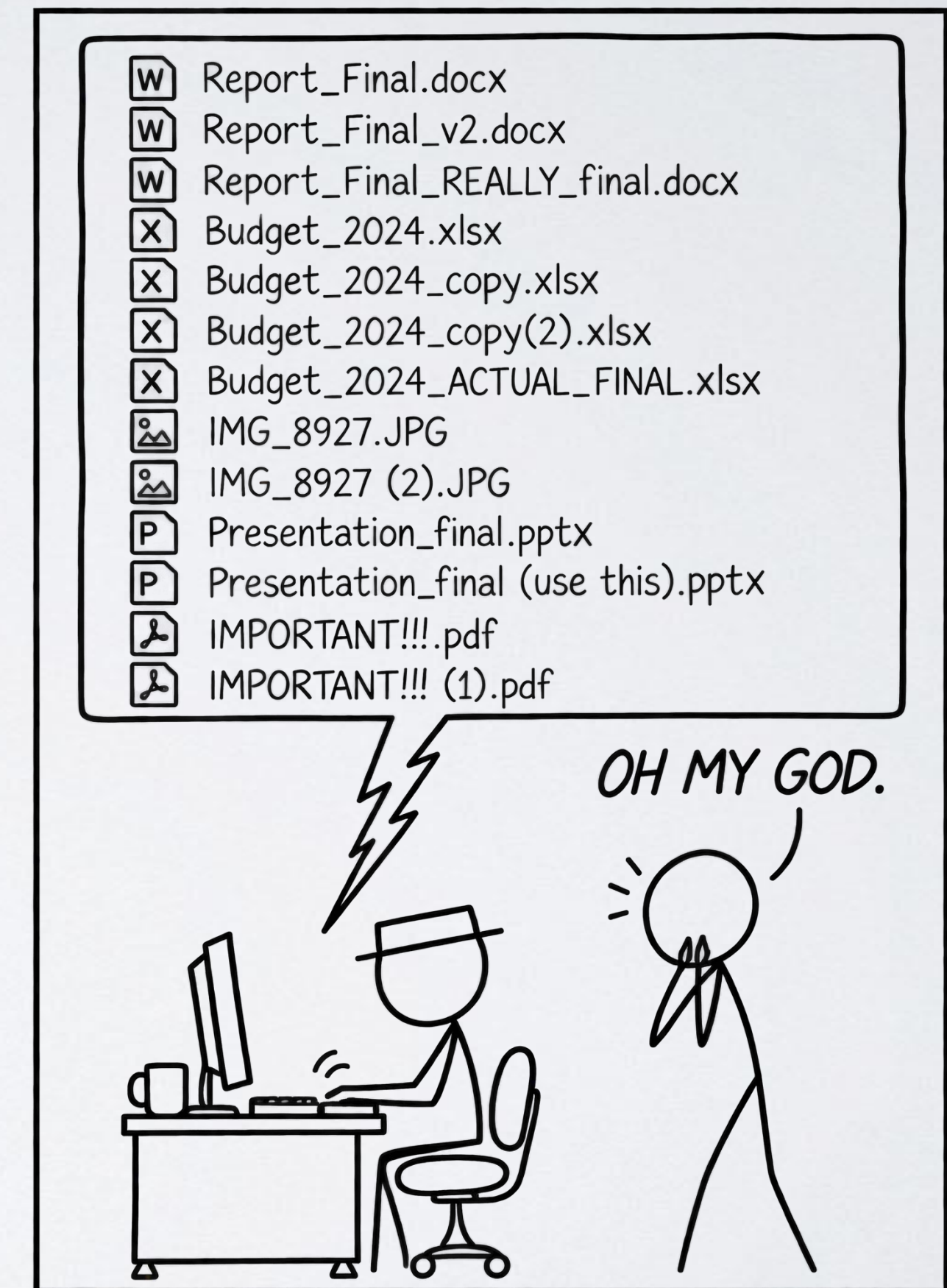


3. Use Meaningful Names

Leverage filenames to help you manage complex projects

Goals:

- Identify file/contents in a clear way
- Have a consistent approach across projects and collaborators
- Should be meaningful but brief

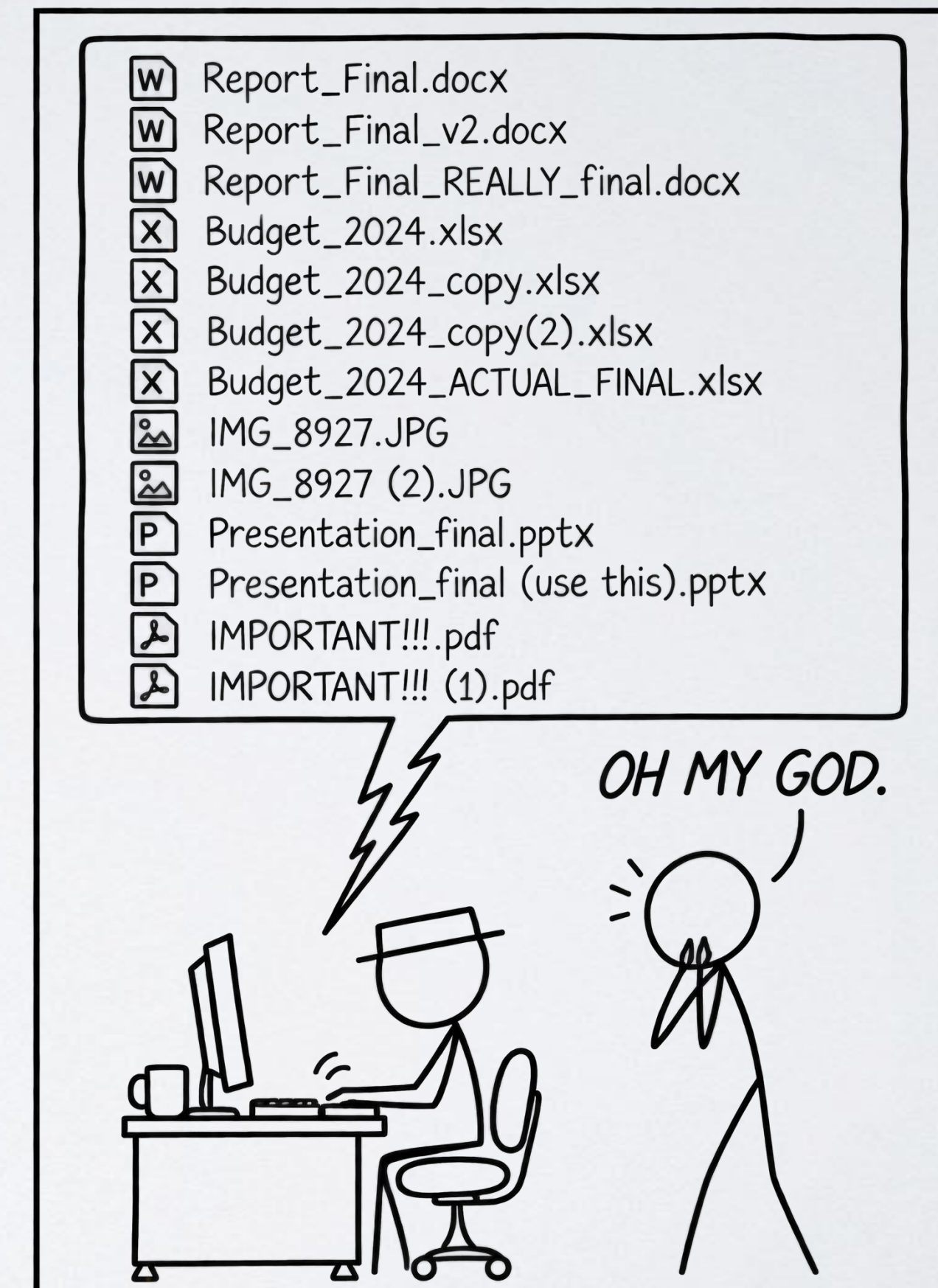


DIGITAL ARCHAEOLOGY IS NOT FOR
THE FAINT OF HEART.

3. Use Meaningful Names

Leverage filenames to help you manage complex projects

- Human Readable: names should clearly describe content in the simplest way possible (e.g., 'code', 'data')
- Computer Readable: ability of a computer to parse a name
 - Use '-' or '_' in place of spaces
 - No special characters (e.g., '&', '#', '^', etc)



DIGITAL ARCHAEOLOGY IS NOT FOR
THE FAINT OF HEART.

3. Use Meaningful Names

Leverage filenames to help you manage complex projects

- Human Readable: names should clearly describe content in the simplest way possible (e.g., 'code', 'data')
- Computer Readable: ability of a computer to parse a name
 - Use '-' or '_' in place of spaces
 - No special characters (e.g, '&', '#', '^', etc)
- Sortable: help you find what you need in the future
 - Dates: YYYY-MM-DD
 - Study IDs: Pad with zeros



fig_1.pdf
fig_10.pdf
fig_11.pdf
fig_12.pdf
fig_2.pdf
fig_3.pdf
fig_4.pdf
fig_5.pdf
fig_6.pdf
fig_7.pdf
fig_8.pdf
fig_9.pdf

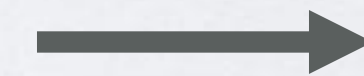
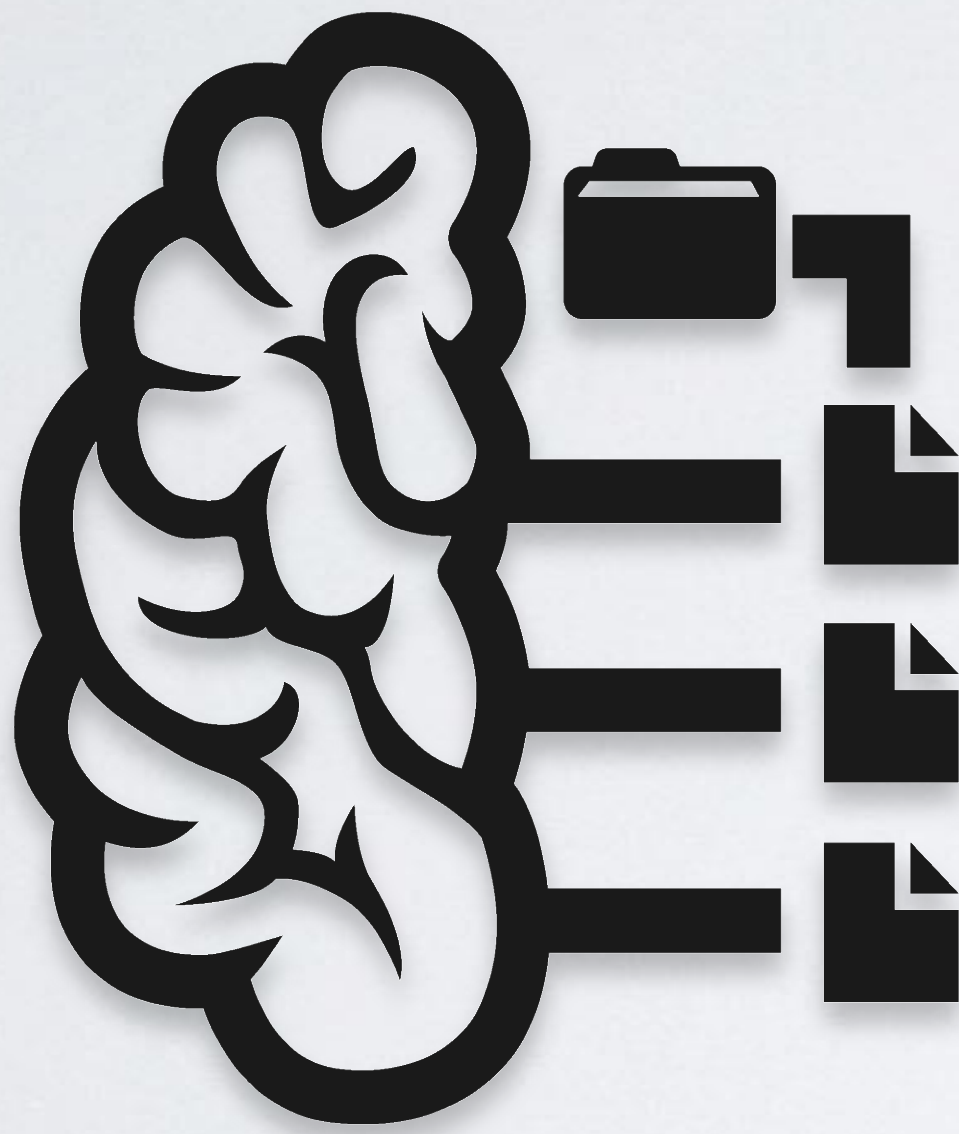


fig_01.pdf
fig_02.pdf
fig_03.pdf
fig_04.pdf
fig_05.pdf
fig_06.pdf
fig_07.pdf
fig_08.pdf
fig_09.pdf
fig_10.pdf
fig_11.pdf
fig_12.pdf

3. Use Meaningful Names

Leverage filenames to help you manage complex projects



Key-Value Pairs in the Brain Imaging Data Structure (BIDS):

key1-value_key2-value_suffix.extension

key-value pair naming:

keys: label or identifier that describes the value that follows

value: specific value for that file

Labels: value that is a string (e.g., 'followup', 'baseline')

Indices: value that is a zero-padded indexes (001, 002)

suffix: follows all key-value pairs and often describes the type of data

3. Use Meaningful Names

Leverage filenames to help you manage complex projects

Do **NOT** Use

- Spaces
- Periods (except for file extensions)
- Other special characters (&, *, ^, etc)

DO Use

- CamelCase
- snake_case
- kabob-case
- YYYYMMDD date format
- Pad numbers with zeros (e.g., 001)

Hands On - Directory Structures and File Naming

4. Preserve the Journey

Version control: tracking and managing changes to documents or code

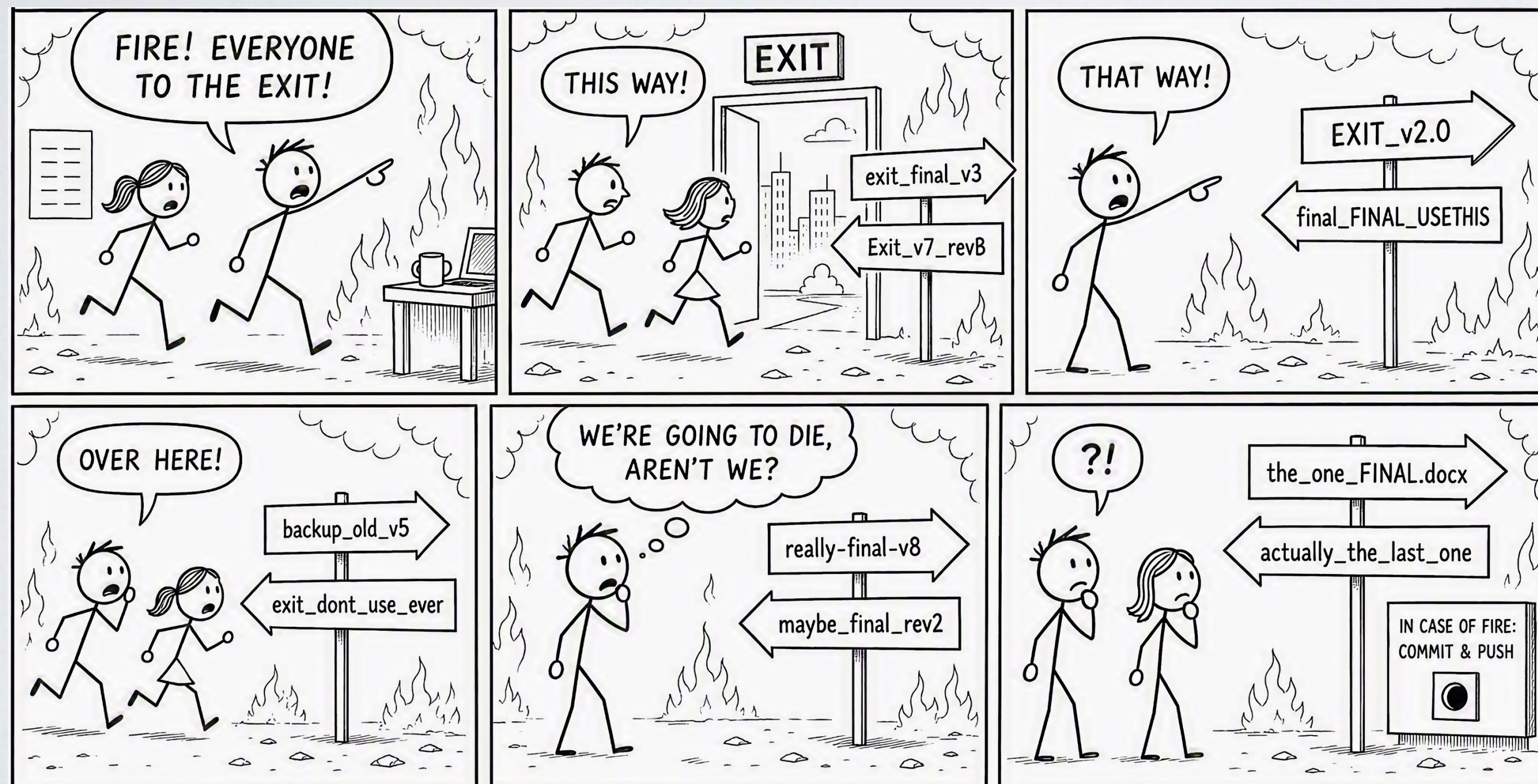


Image is AI generated and inspired by @redpenblackpen

4. Preserve the Journey

Version control: tracking and managing changes to documents or code

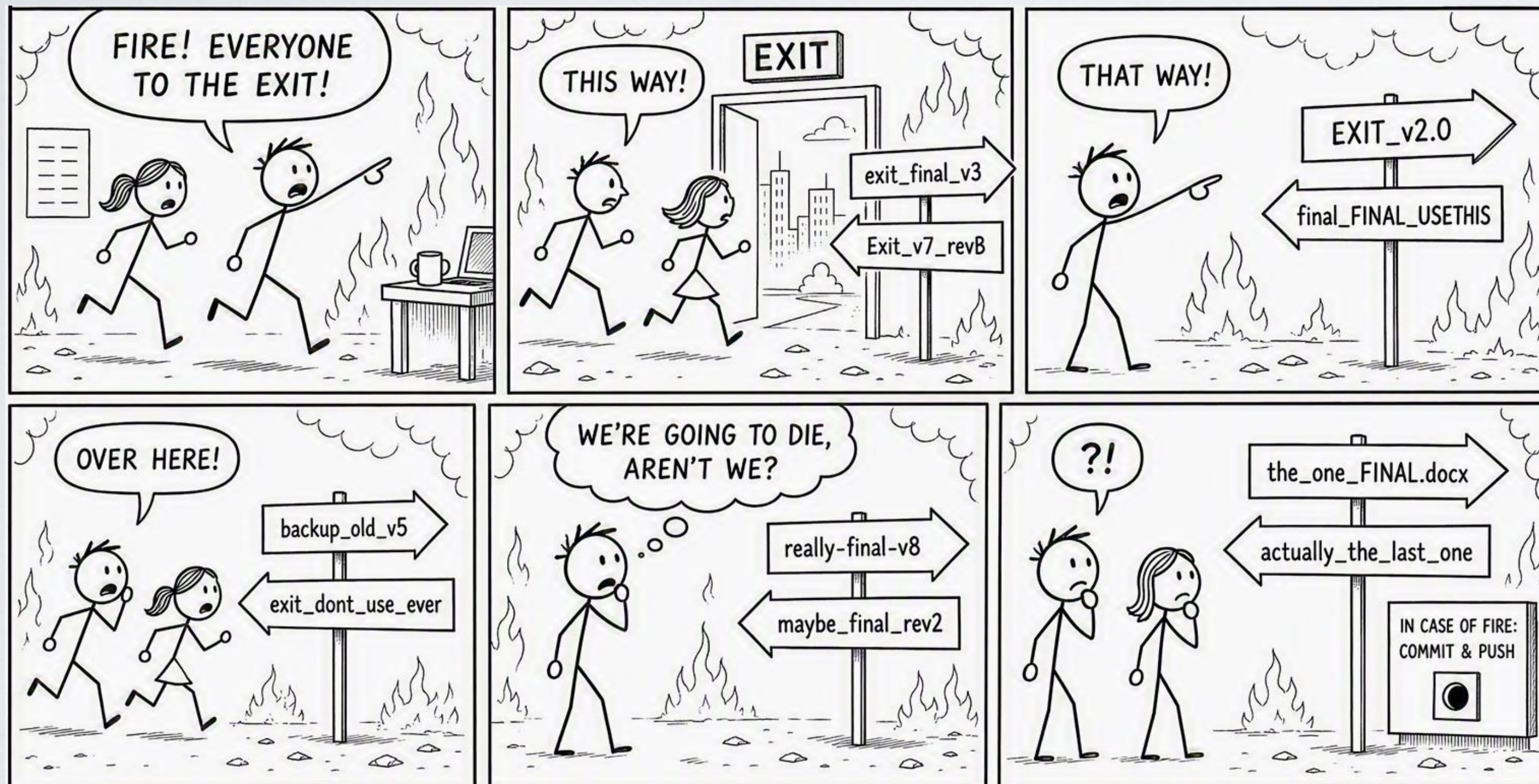


Image is AI generated and inspired by @redpenblackpen

- Manual: use file naming to document drafts (e.g., dates, version numbers)
- Software: git, GitHub, subversion
- Allows you to trace your steps

4. Preserve the Journey

Version control: tracking and managing changes to documents or code

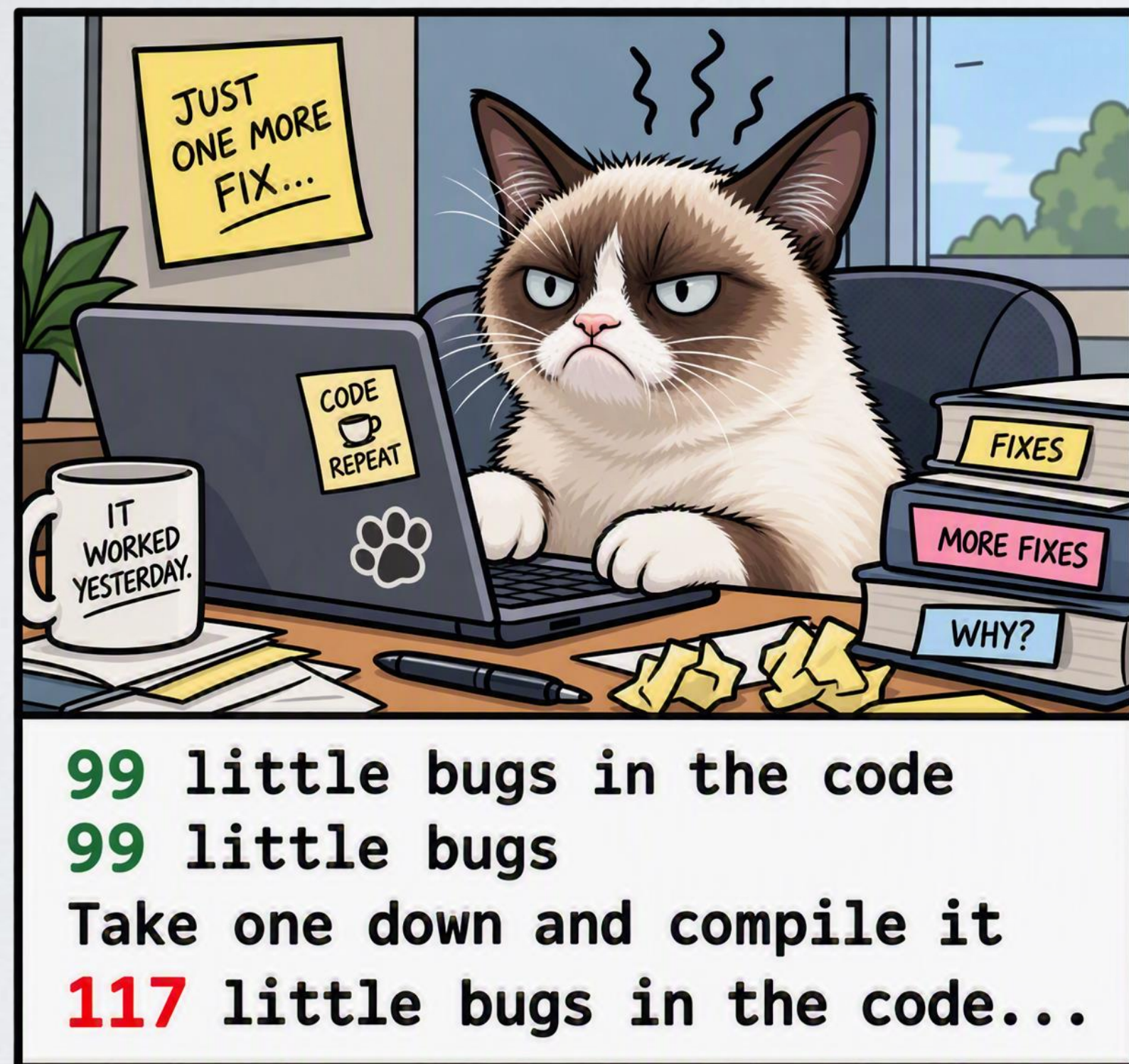
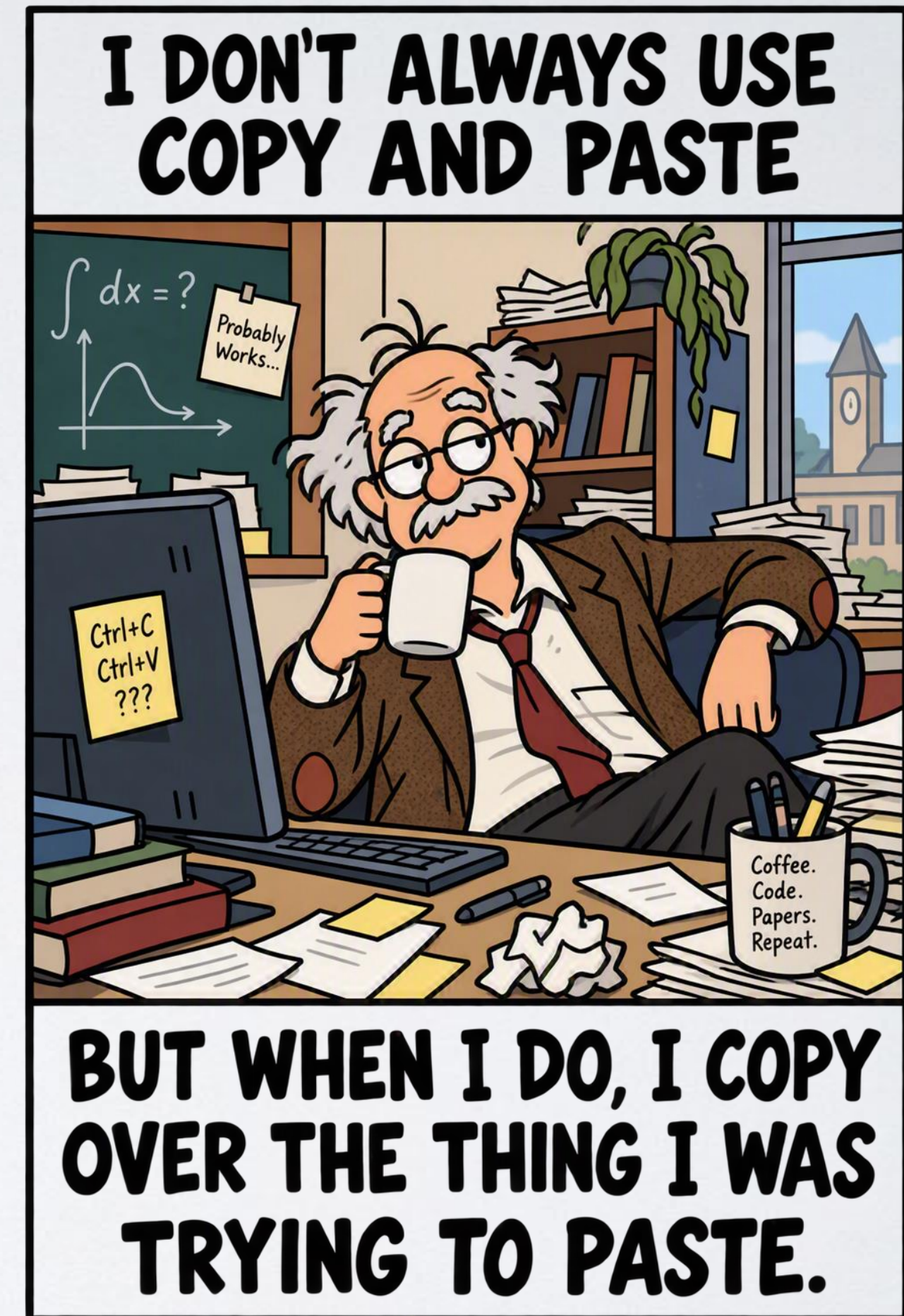


Image is AI generated

- Manual: use file naming to document drafts (e.g., dates, version numbers)
- Software: git, GitHub, subversion
- Allows you to trace your steps

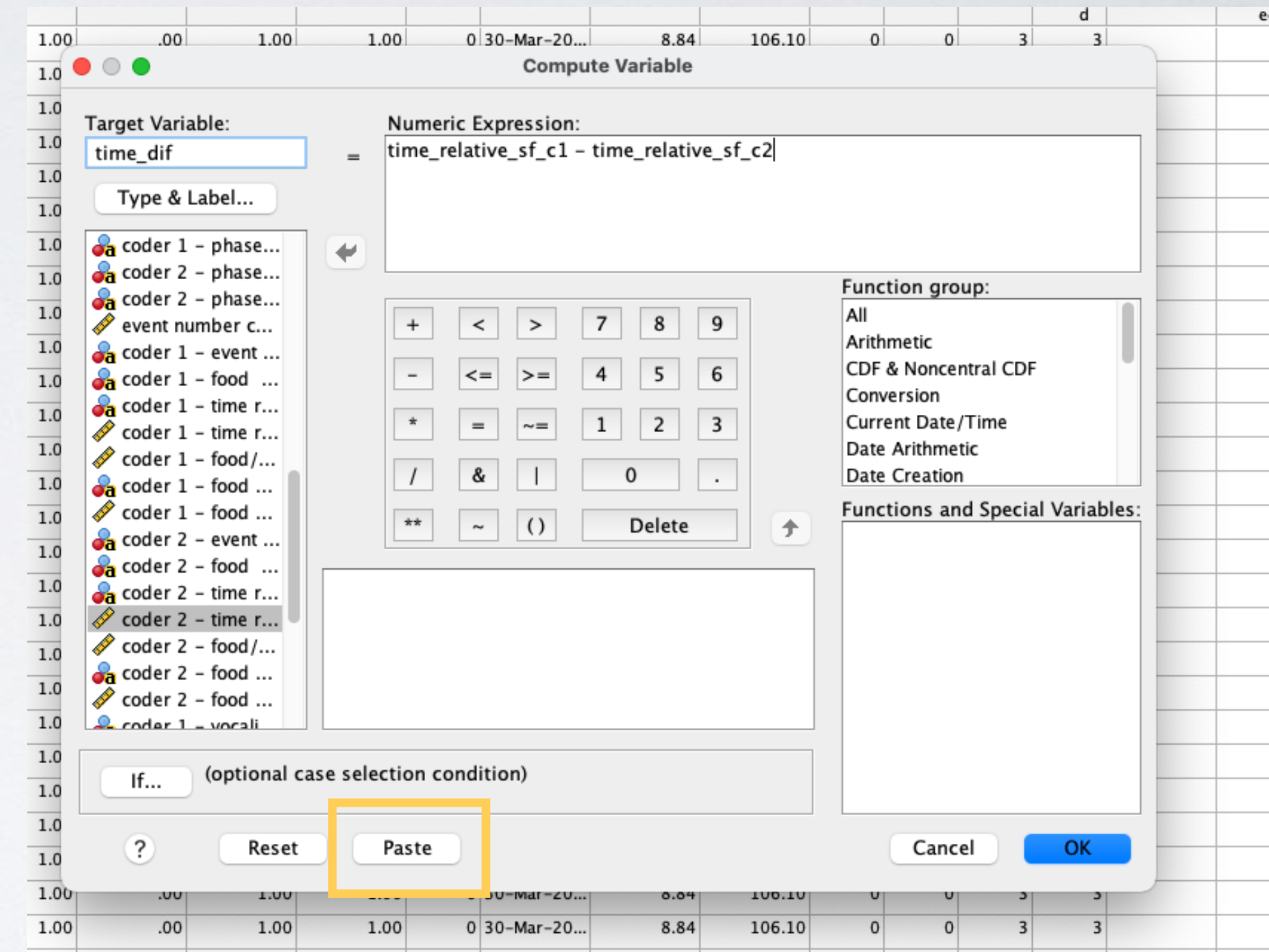
5. Avoid manual manipulations

- Manual data manipulations leave no trace
 - Hard to reproduce
 - Error prone
- Alternatives:
 - Save Syntax in SPSS
 - Include calculations in variable descriptions
 - Script data cleaning

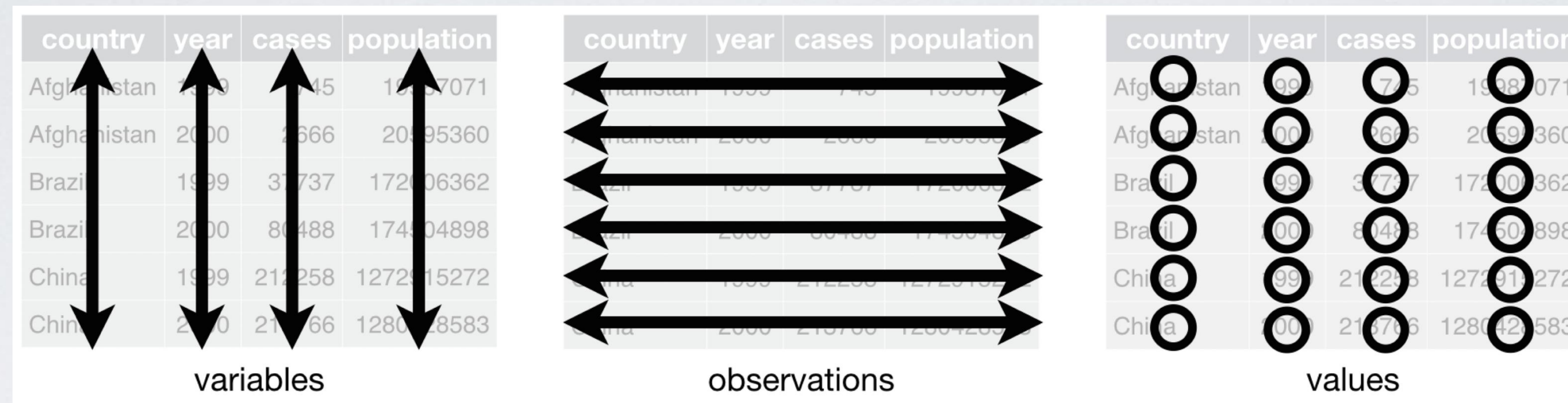


5. Avoid manual manipulations

- Manual data manipulations leave no trace
 - Hard to reproduce
 - Error prone
- Alternatives:
 - Save Syntax in SPSS
 - Include calculations in variable descriptions
 - Script data cleaning



6. 'Tidy' your data



- Every variable is in its own column
- Each participant/sample is in its own row
- Each value is in its own cell

6. 'Tidy' your data

- Use open file formats — csv, html, txt, jpeg

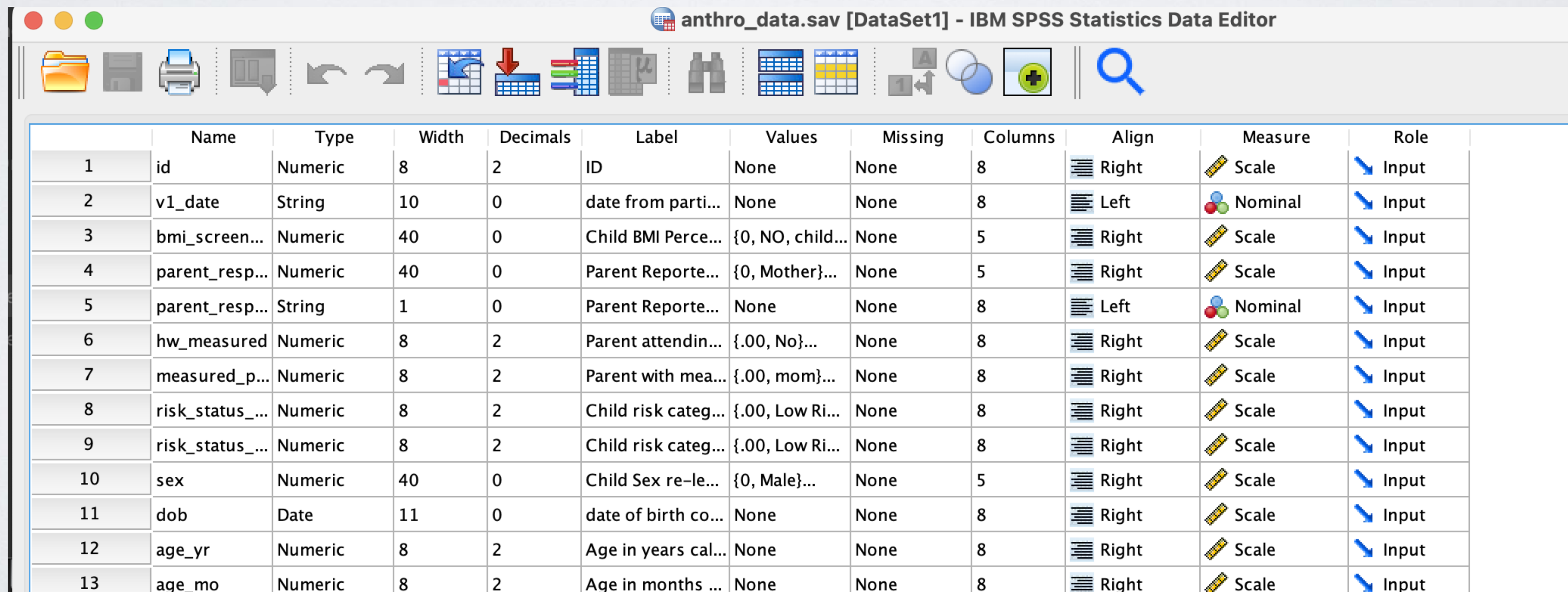
6. 'Tidy' your data

- Use open file formats — csv, html, txt, jpeg
- Create a data dictionary

column	variable	label	value_labels	type	n_na	range
1	id	ID	NULL	double	0	c(1, 133)
2	v1_date	date from participant contacts databases ('verified_visit_da	NULL	character	0	c("2018-01-31", "2022-05-07")
3	bmi_screenout	Child BMI Percentile Screen Out	c(`YES, child is overweight, so	double	0	c(0, 1)
4	parent_respondent	Parent Reported: Parent relationship to child re-leveled in R	c(Mother = 0, Father = 1, Oth	double	0	c(0, 1)
5	parent_respondent_o	Parent Reported: Parent specify relationship to child if other	NULL	character	0	c("", "")
6	hw_measured	Parent attending Visit 1 had measured height and weight	c(No = 0, Yes = 1)	double	0	c(1, 1)
7	measured_parent	Parent with measured BMI at Visit 1	c(mom = 0, dad = 1)	double	0	c(0, 1)
8	risk_status_mom	Child risk categor: Low risk: Mom BMI < 26, High Risk: Mom	c(`Low Risk` = 0, `High Risk` =	double	0	c(0, 1)
9	risk_status_both	Child risk category: Low Risk: Mom and Dad BMI < 25, High	c(`Low Risk` = 0, `High Risk` =	double	0	c(0, 2)
10	sex	Child Sex re-leveled in R to start with 0	c(Male = 0, Female = 1)	double	0	c(0, 1)
11	dob	date of birth converted to format yyyy-mm-dd in R	NULL	double	0	c(14333, 16391)
12	age_yr	Age in years calculated from dob and start_date	NULL	double	0	c(7, 8.99)
13	age_mo	Age in months calculated from dob and start_date	NULL	double	0	c(84, 107.9)
14	ethnicity	Parent Reported: Child ethnicity	c(`NOT Hispanic or Latino` = 0	double	0	c(0, 0)
15	race	Parent Reported: Child race -- Note: prefer not to answer (p	c(`White/Caucasian` = 0, `Am	double	0	c(0, 2)
16	income	Parent Reported: Yearly household income -- Note: prefer n	c(`Less than \$20,000` = 0, `\$20	double	3	c(0, 5)
17	parent_ed	Parent Reported: Parent education re-leveled in R to start w	c(`High School or GED (12 yea	double	0	c(0, 5)

6. 'Tidy' your data

- Use open file formats — csv, html, txt, jpeg
- Create a data dictionary



The screenshot displays the IBM SPSS Statistics Data Editor window for a file named 'anthro_data.sav'. The window title bar includes the file name and the application name. Below the title bar is a toolbar with various icons for file operations, editing, and analysis. The main area of the window shows a data dictionary table with 13 columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The table lists 13 variables, each with its corresponding data type, width, decimal places, label, values, missing values, column count, alignment, measure, and role.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	id	Numeric	8	2	ID	None	None	8	Right	Scale	Input
2	v1_date	String	10	0	date from parti...	None	None	8	Left	Nominal	Input
3	bmi_screen...	Numeric	40	0	Child BMI Perce...	{0, NO, child...	None	5	Right	Scale	Input
4	parent_resp...	Numeric	40	0	Parent Reporte...	{0, Mother}...	None	5	Right	Scale	Input
5	parent_resp...	String	1	0	Parent Reporte...	None	None	8	Left	Nominal	Input
6	hw_measured	Numeric	8	2	Parent attendin...	{.00, No}...	None	8	Right	Scale	Input
7	measured_p...	Numeric	8	2	Parent with mea...	{.00, mom}...	None	8	Right	Scale	Input
8	risk_status_...	Numeric	8	2	Child risk categ...	{.00, Low Ri...	None	8	Right	Scale	Input
9	risk_status_...	Numeric	8	2	Child risk categ...	{.00, Low Ri...	None	8	Right	Scale	Input
10	sex	Numeric	40	0	Child Sex re-le...	{0, Male}...	None	5	Right	Scale	Input
11	dob	Date	11	0	date of birth co...	None	None	8	Right	Scale	Input
12	age_yr	Numeric	8	2	Age in years cal...	None	None	8	Right	Scale	Input
13	age_mo	Numeric	8	2	Age in months ...	None	None	8	Right	Scale	Input

6. 'Tidy' your data

- Use open file formats — csv, html, txt, jpeg
- Create a data dictionary
- One piece of information per cell

height		height_ft	height_in
5 ft 6 in		5	6
5 ft 2 in		5	2
7 ft		7	0
5 ft 11 in		5	11

6. 'Tidy' your data

- Use open file formats — csv, html, txt, jpeg
- Create a data dictionary
- One piece of information per cell
- Do not use highlighting/font color as data

height			
5 ft 6 in	height_ft	height_in	check_height
5 ft 2 in	5	6	0
7 ft	5	2	0
5 ft 11 in	7	0	1
	5	11	0

6. 'Tidy' your data

- Use open file formats — csv, html, txt, jpeg
- Create a data dictionary
- One piece of information per cell
- Do not use highlighting/font color as data

7. Metadata Magic

Metadata: the who, what, when, where, and why of your data

Easiest: when in doubt, document

- Data dictionaries
- Standard operating procedures manuals
- Lab notebooks
- changelog file (document versions)
- README
 - Description of folders/files
 - Can provide instructions on use of code/data
 - License information

**METADATA IS A LOVE
NOTE TO THE FUTURE!**



7. Metadata Magic

Metadata: the who, what, when, where, and why of your data

Medium Effort: Data Manual

- Larger
- More verbose and detailed
- Can include science/rational/citations
- Like a user manual for data

METADATA IS A LOVE
NOTE TO THE FUTURE!



7. Metadata Magic

Metadata: the who, what, when, where, and why of your data

Bigger Lift: Structured Metadata

- Often laid out in fields
- Can require use of shared vocabularies
- Often field/data type specific

**METADATA IS A LOVE
NOTE TO THE FUTURE!**



'Good Enough' Practices

1. Preserve Raw Data
2. Create a Central Hub
3. Use Meaningful Names
4. Preserve the Journey
5. Avoid Manual Manipulations
6. 'Tidy' Your Data
7. Metadata Magic



Provost Endorsement: Research Data Stewardship



1. Register



2. Workshop Completion

Tools to Put Research Data Stewardship into Practice

Resources can help you navigate requirements, plan for your data, choose appropriate infrastructure, and share your work.

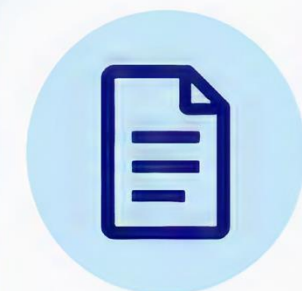


What does my funder require?

Stay informed about federal and institutional requirements for data management and sharing.



Federal policy tracking resources
e.g., NIH, NSF, OSTP



Funder guidance and templates
Find requirements and sample language.



How should I plan for my data?

Create a data management plan (DMP) to organize your data and meet funder and institutional expectations.



DMPTool
Plan smarter.
Research further.

Get step-by-step guidance, templates, and examples to help you build a strong DMP.



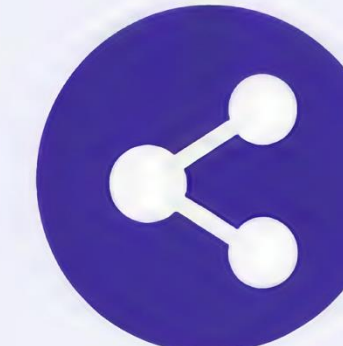
Where should my data live?

Find the right storage solution for your data based on size, sensitivity, and project needs.



Penn State Data Storage Finder
Identify the best storage option for your research data.

Explore options for short-term and long-term storage, including services for sensitive data.



How can I share and preserve my data?

Increase the visibility and impact of your research by sharing your data in trusted repositories.



Repository-finding resources
Find a trusted repository for your discipline or funder.



Penn State repositories
Share your data in ScholarSphere and Data Commons, our institutional repositories.



Not sure which resource applies?

Research Data Stewardship can help you navigate the options.

Get in touch to discuss your needs or book a consultation.

Connect with Research Data Stewardship

Questions about research data? We're here to help.



BOOK A CONSULTATION

Get individualized help with data management, requirements, storage, documentation, sharing, repositories, or other research data questions.



Scan the QR code
to request a consultation.



EXPLORE TRAINING & RESOURCES

Find workshops, guidance, tools,
and other research data
stewardship resources.



Workshops
& programs



Guidance
& templates



Tools
& services

Visit our website

libraries.psu.edu/research-data-stewardship



STAY CONNECTED

Join the Research Data Stewardship
mailing list for training, resources,
and program updates.



Scan the QR code
to join our mailing list.

OR



Email us at
ul-rds@lists.psu.edu

Impactful research is built on strong research data stewardship.

Research Data Stewardship Workshops

Practical training to strengthen research data practices.



FALL 2026

20 workshops this fall
All online | Free

Build your skills. Advance your research.



Provost-Endorsed Program Workshops

Core workshops for the Provost-Endorsed Program in Research Data Stewardship. Open to all.

SEPT 16

What Researchers Need to Know About Data Management and Sharing Policies
1:00–2:00 p.m.

SEPT 17

“Good Enough” Practices that Build Better Research Data Management Habits
1:00–2:30 p.m.

OCT 8

Creating Data Others Can Understand: Metadata and Documentation
1:00–2:30 p.m.



Repositories & Open Access

Learn about sharing, preserving, and increasing the visibility of your research.

OCT 20

Open Access from Every Angle: Navigating Funder Requirements and Penn State’s OA Policy with the Researcher Metadata Database and ScholarSphere
1:00–2:00 p.m.

OCT 29

Sharing Data in ScholarSphere
1:00–1:45 p.m.

NOV 5

Where Should Your Data Live? Selecting the Right Repository
1:00–1:45 p.m.



RECR Credit Workshops

Eligible for credit through the Responsible and Ethical Conduct of Research (RECR) program.

OCT 1

“Good Enough” Research Data Management
11:00 a.m.–noon

OCT 8

Layers of Reproducibility: Practical Skills for More Transparent Research
11:00 a.m.–noon



Statistical & Analytical Methods

Build your data analysis skills with hands-on, practical training.

SEPT 22

Data Analysis Planning: Why is it Essential?
2:00–4:00 p.m.

OCT 7

Quantitative Methods for Undergraduate Students
1:00–2:30 p.m.

OCT 15

Introduction to SPSS
2:00–4:00 p.m.

OCT 22

Introduction to Minitab
2:00–4:00 p.m.



Research Reproducibility and R Workshop Series

A five-part series on R and RStudio for research reproducibility.

OCT 14

Introduction to R and RStudio
1:00–3:00 p.m.

OCT 21

Data Wrangling in R
1:00–3:00 p.m.

OCT 28

Data Management and Reproducibility in R and RStudio
1:00–3:00 p.m.

NOV 4

Data Visualizations in R
1:00–3:00 p.m.

NOV 11

Functions and Reproducible Workflows in R
1:00–3:00 p.m.



Explore the full schedule & register

Scan the QR code or visit the Penn State News article for the complete list of Fall 2026 workshops, descriptions, and registration information.

psu.edu/news/university-libraries/story/university-libraries-announces-research-data-stewardship-fall-2026



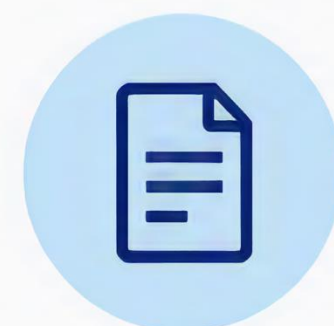
Open to Penn State undergraduate and graduate students, postdoctoral scholars, staff, and faculty.

Provost-Endorsed Program in Research Data Stewardship

Build practical research data stewardship skills you can apply directly to your research.



What You'll Learn



UNDERSTAND REQUIREMENTS

Interpret and apply Penn State and federal research data policies to research data stewardship decisions.



BUILD BETTER PRACTICES

Implement research data stewardship practices that support rigorous and reproducible research workflows.



DEVELOP A STRATEGY

Create a practical research data stewardship approach for your project or research group.

Program Components



Workshops

Three interactive workshops:

- 1 Penn State & Federal Research Data Policies
- 2 Research Data Management
- 3 Data Documentation



Real-World Applications

Two applied activities:

- 1 Create a reusable dataset
- 2 Prepare a Research Data Strategy for yourself or your research group

Learn more & register



Scan the QR code
or visit our website

Provost Endorsement: Research Data Stewardship



1. Register



2. Workshop Completion