Personal Data Management during your PhD

Elgin Akin

Methods Workshop

Pre-Class Survey Results

Show Plots for:

- Currently used Cloud-based data backup
- Preferred Lab Notebook Style
- Current Personal Data Management Plan
- Preferred Citation/Reference Manager

The hubris of a PhD: **I'll remember everything...**

Data you will generate can be *Diverse* and *Specialized*

National Institutes of Health
Office of Intramural Research

2023 NIH Data Management and Sharing Policy

The NIH has issued a Data Management and Sharing (DMS) policy, effective January 25, 2023, to promote the sharing of scientific data. There are multiple benefits to sharing scientific data, and ultimately this will facilitate the development of treatments and products that improve human health.

Omics Data

Imaging Data

Additional Data

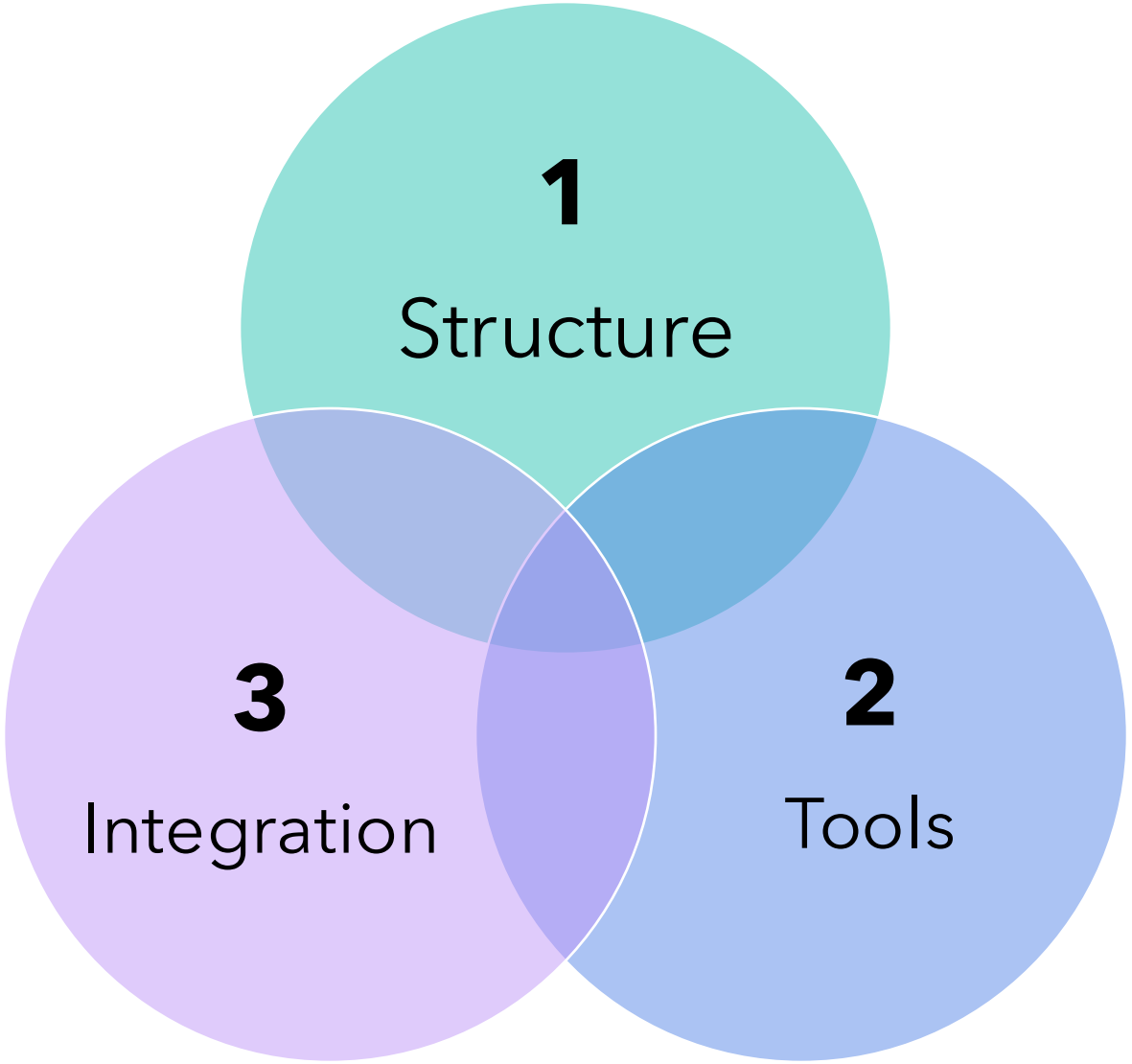
Biological Data

Phenotype Data

(your PI will take care of this…hopefully)

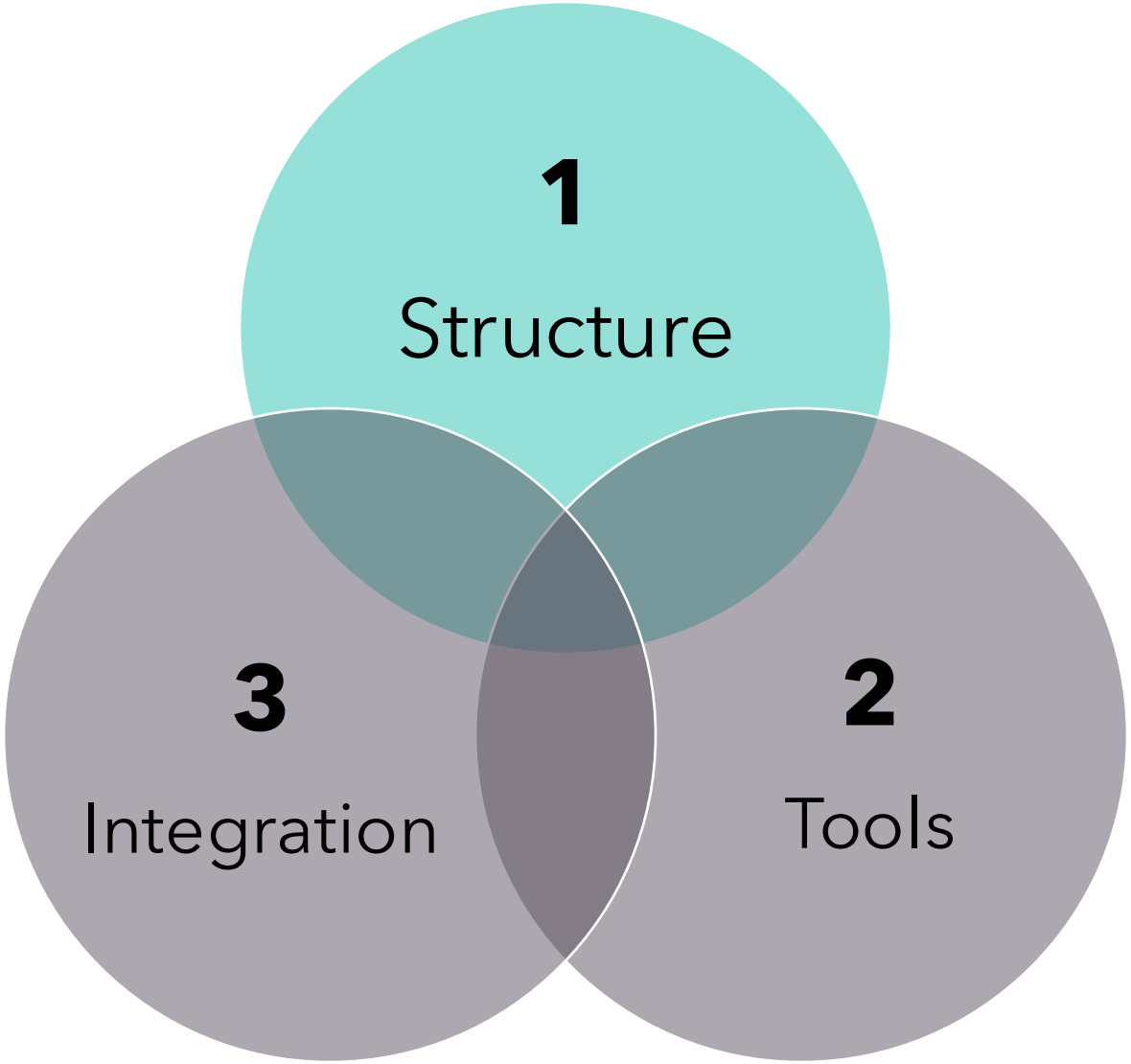
Reference: https://oir.nih.gov/sourcebook/intramural-program-oversight/intramural-data-sharing/2023-nih-data-management-sharing-policy#:~:text=The%20NIH%20has%20issued%20a,products%20that%20improve%20human%20health (accessed 18 September 2023)

Developing a personal data management Plan for your PhD



Developing a personal data management Plan for your PhD

Structure:
Projects
Experiments
Data

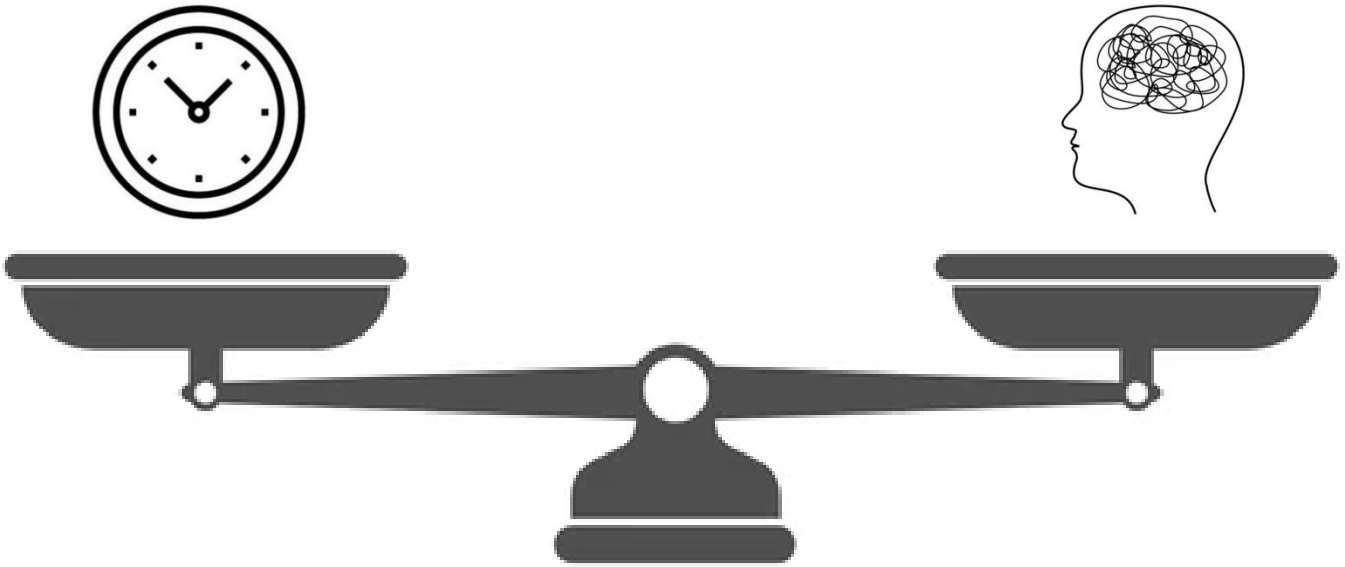


1 Developing A Structure

The "cost per use" heuristic

Time + Effort

'Usability'
(different for everyone)



1 Developing A Structure

Accessibility

Your Data

Future you
(and others)

1 Developing A (Personal) Structure

*each **project** and **experiment** should have an identifier*

P1

Project ID

E01

Experiment ID

.x

Optional:
experiment sub-ID
(Helpful when troubleshooting and optimizing experiments)

example:
***P2E21**.2*

Tubes and Reagents

Project + Experiment Directories

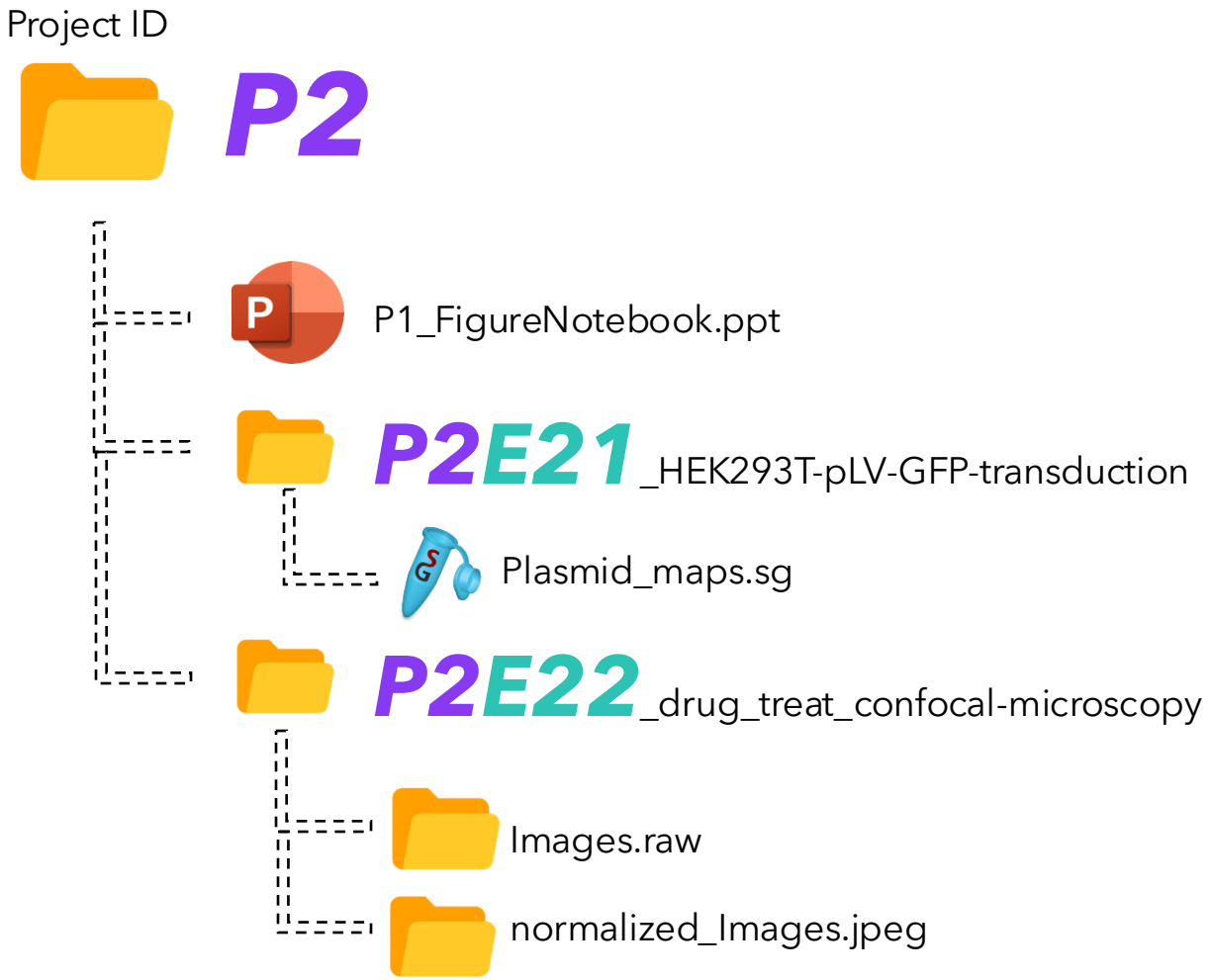
Notebooks + file projects

1 Developing A (Personal) Structure

*each **project** and **experiment** should have an <u>identifier</u>*

How do I **structure** projects? experiments?

Nested Directories w/ descriptive language



1 Developing A (Personal) Structure

PXXEXX.x

Notebook
- Literature?
- Hypothesis
- methods
- results
- detailed notes

Raw Data Directories
'straight from the instrument'
- structured by Project + experiment

Figure Notebook
- Clean
- markups
- presentable

Redundant Backups
Keep everything on the cloud

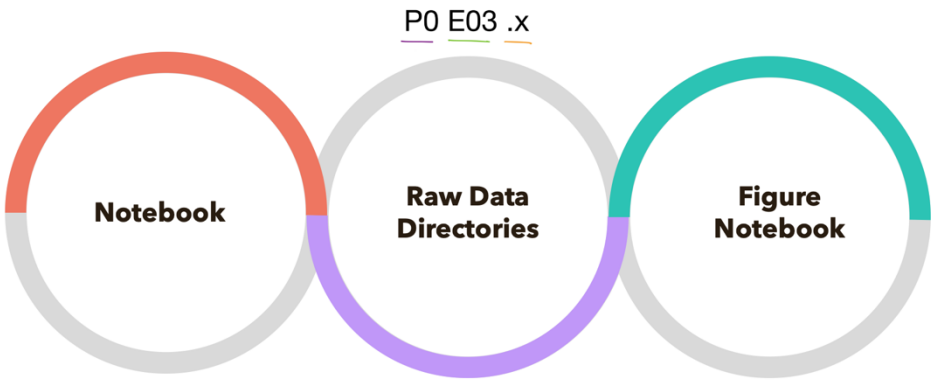
Summary: Developing A Structure

Project + Experiment Tracking

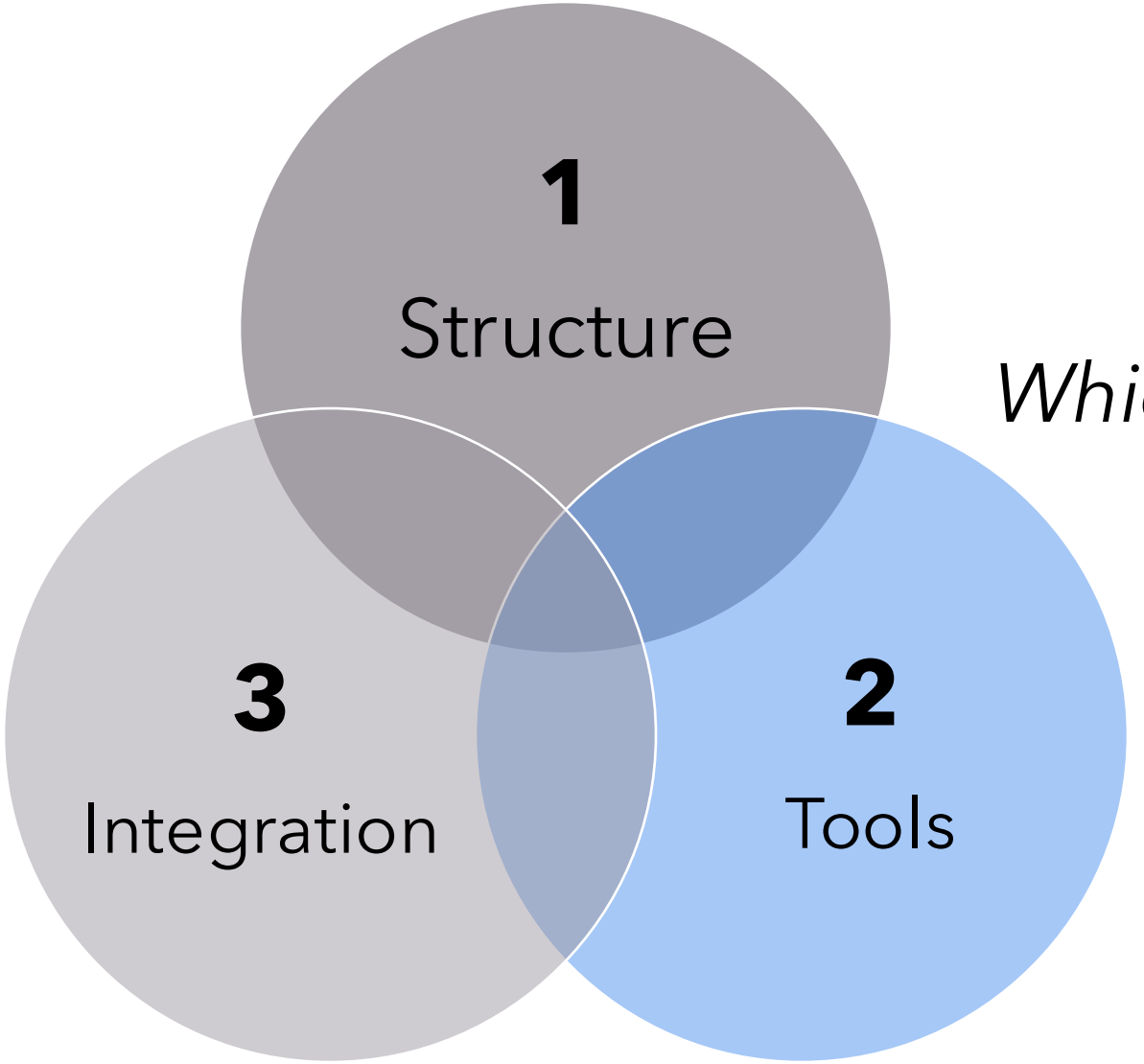
🤝 Compatibility + Accessibility

☁️ **Redundant Backups**

"cost per use



Developing a personal data management Plan for your PhD



Which tools do I use?

Lab notebooks

Electronic Laboratory Notebooks *All Cloud-backup Compatible*

OneNote

Benchling

Basic Customization

Obsidian

Any type

Medium to Advanced Customization

Electronic Laboratory Notebooks

OneNote

Pros

- Simple UI
- Multi Level organization
- Simple Backlinking Seamless Office
- Product Integration
- Cloud Backup Accessible
- Minutes to hours setup

Cons

- Table manipulation is difficult

-Limited to 1 template per ‘section’

- MacOS version of OneNote is limited in features

Benchling

Obsidian

Anytype

Electronic Laboratory Notebooks

Benchling

Pros

- Built with lab scientists in mind 💭 - Cloud Based
– Free integrated molecular biology tools
- Best table manipulation
- Database Ready (onedrive integration)

Cons

- Data ownership(?)

- Entirely Cloud Based

OneNote

Obsidian

Anytype

Electronic Laboratory Notebooks

Obsidian

Pros

- Complete Data Ownership
- Complex Back-linking
- Complex Template Usage
- Robust data type (.md)

Cons

- Steep learning curve
– Simple UI
- Tables are cumbersome (.md)

OneNote

Benchling

Anytype

Electronic Laboratory Notebooks

Anytype

Local Data (Cloud Backup)

Pros

- A blend of a smooth UX and powerful note-linking backend
- Completely customizable
- semi-complete data ownership

Cons

- Locked into AnyType Cloud backup service (free)
- Mild Learning Curve
- Hours to day – front end organization

OneNote

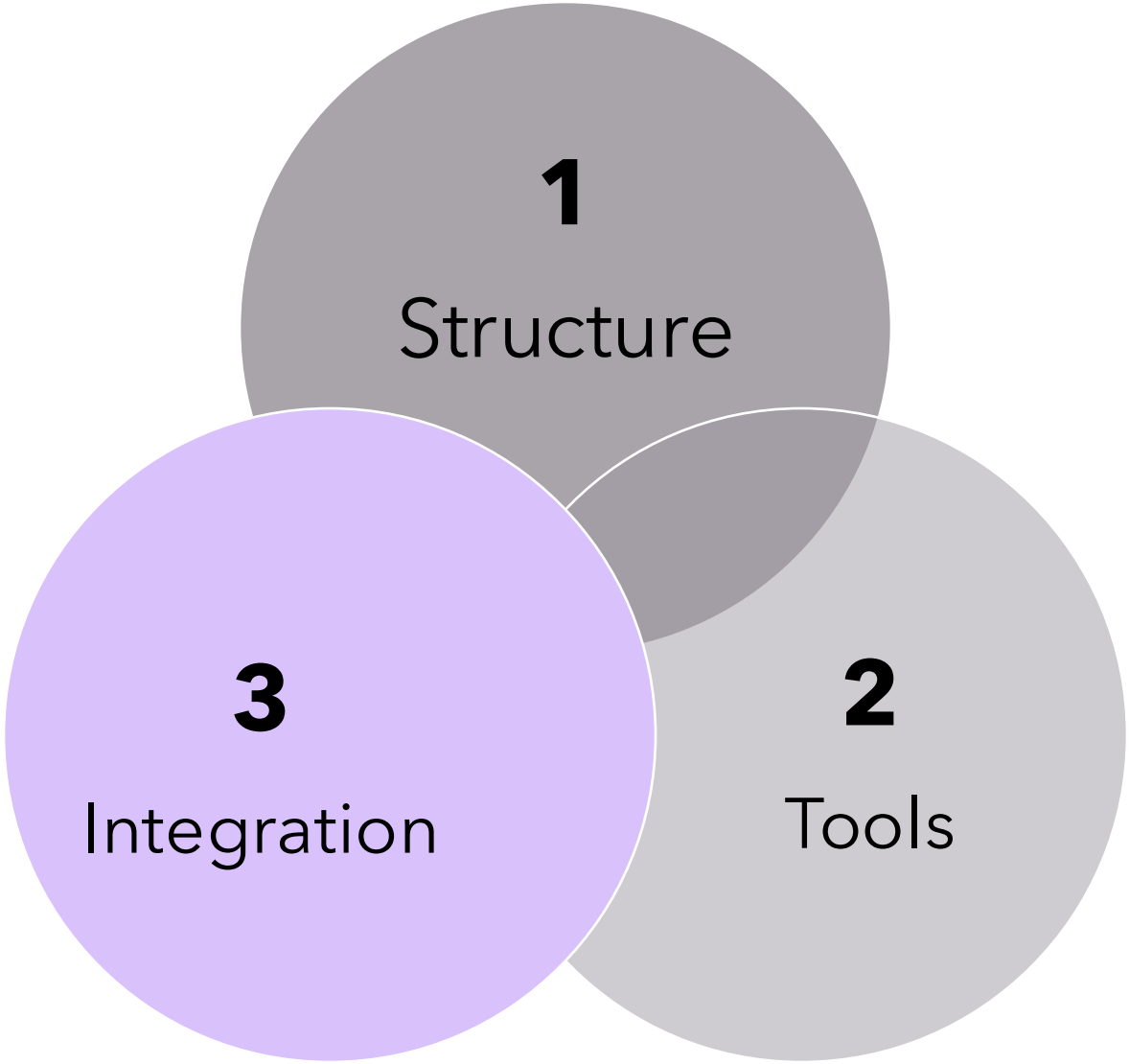
Benchling

Obsidian

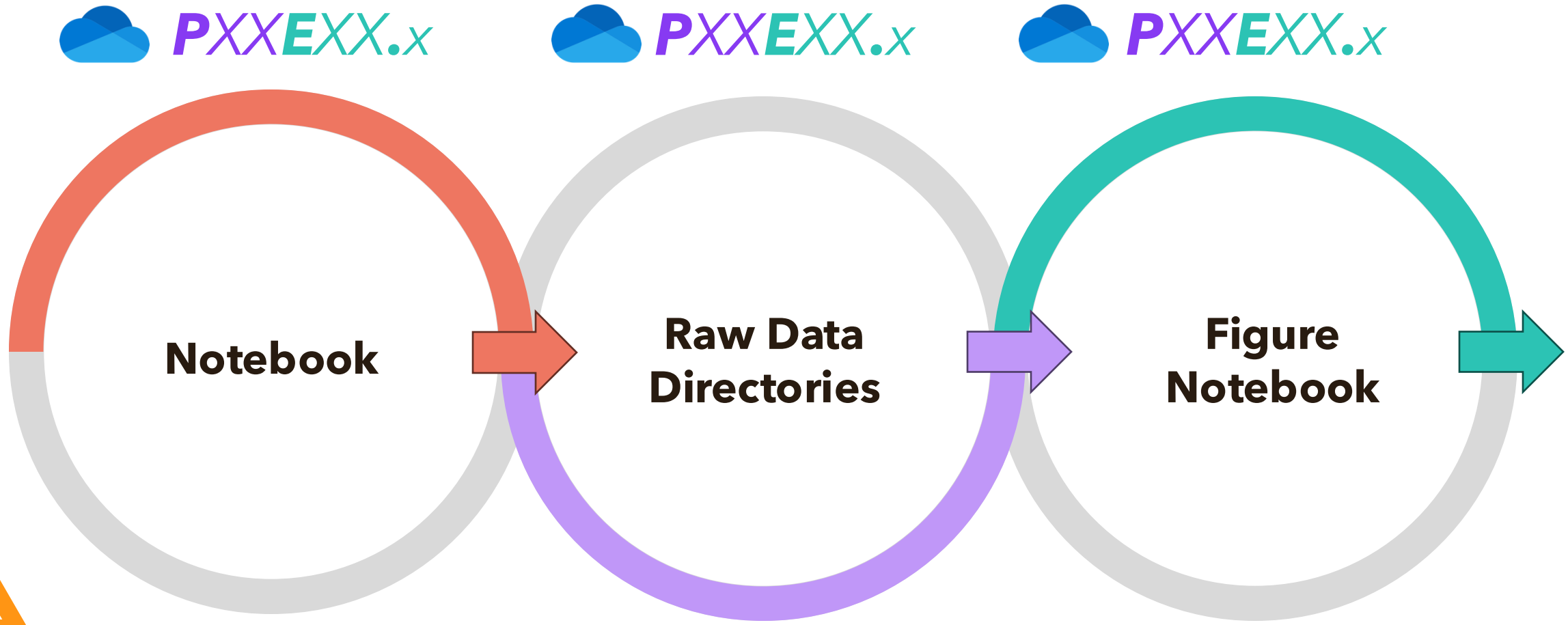
Summary: Notebooks

	Pros	Cons
 **OneNote**	• Easy to use and structure • Office 365 Synced + suite compatible Page Links • Easy Collaboration/Sharing	• Limited customization • Sub-Optimal Searching (requires some digging)
 **Benchling**	• Cloud Only • Excellent Molecular Biology Suite • Easy Collaboration/Sharing	• Cloud Only
 **Obsidian**	• Local (with flexible backup options) • *each note is a .md file*	• Notebook sharing is bulky • Version control method is user-dependent
 **AnyType**	• Local (with specific cloud sync)	• Notebook sharing is bulky

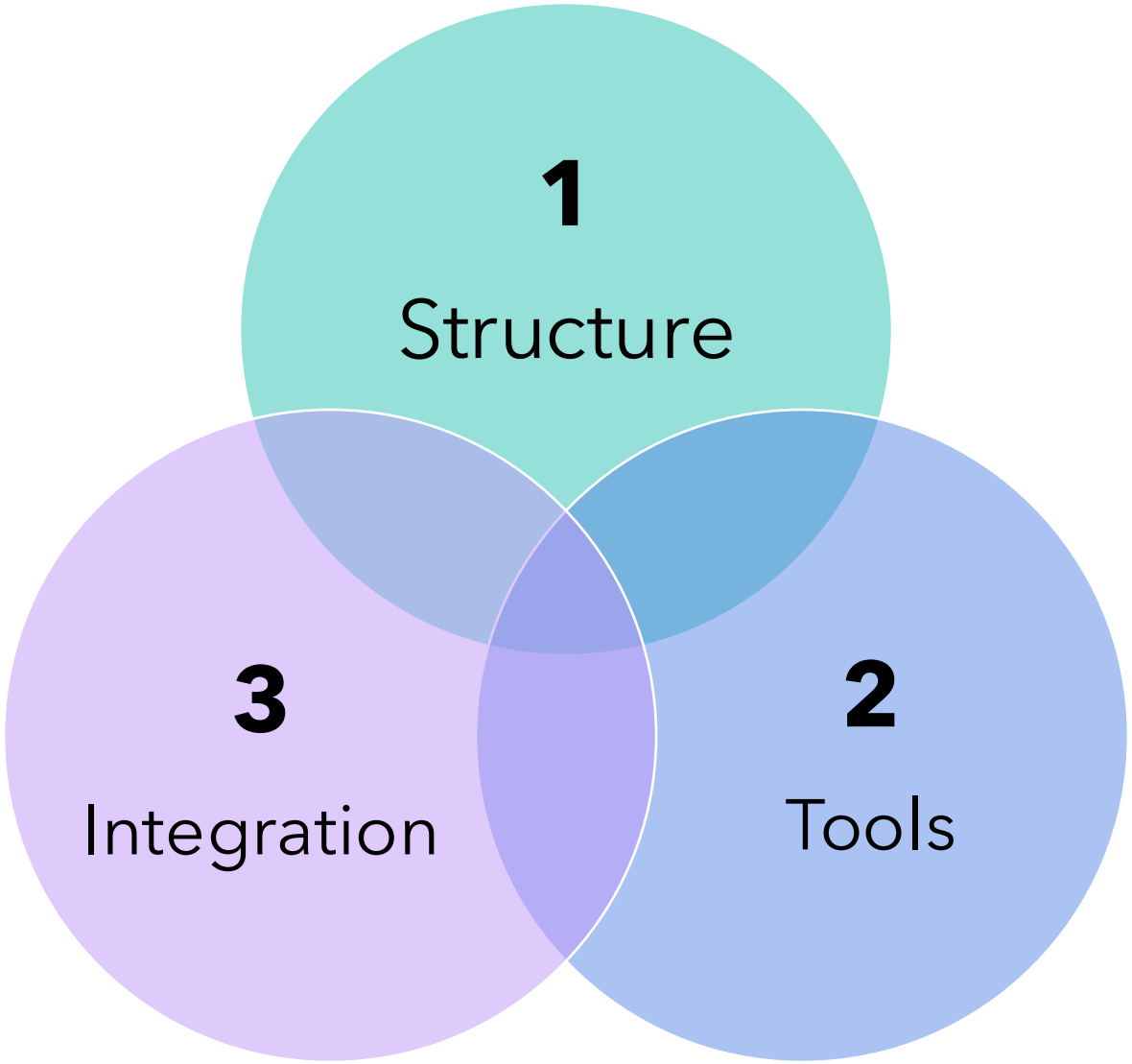
Developing a personal data management Plan for your PhD



3 Integration



Specific Use case



Specific Pipeline Use case:

PXXEXX.x

1.
Lab Notebook

organized
by project
and experiment

2.
Raw Data

Folders organized
by project
and experiment

3.
Figure Notebook

Single ppt per
project
*multiple if
dealing with images

iCloud

Project Scripts

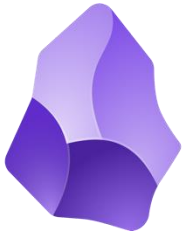
OneDrive

Office 365

SharePoint

OneDrive

Used for ~.5 years now,
Nearly perfect…



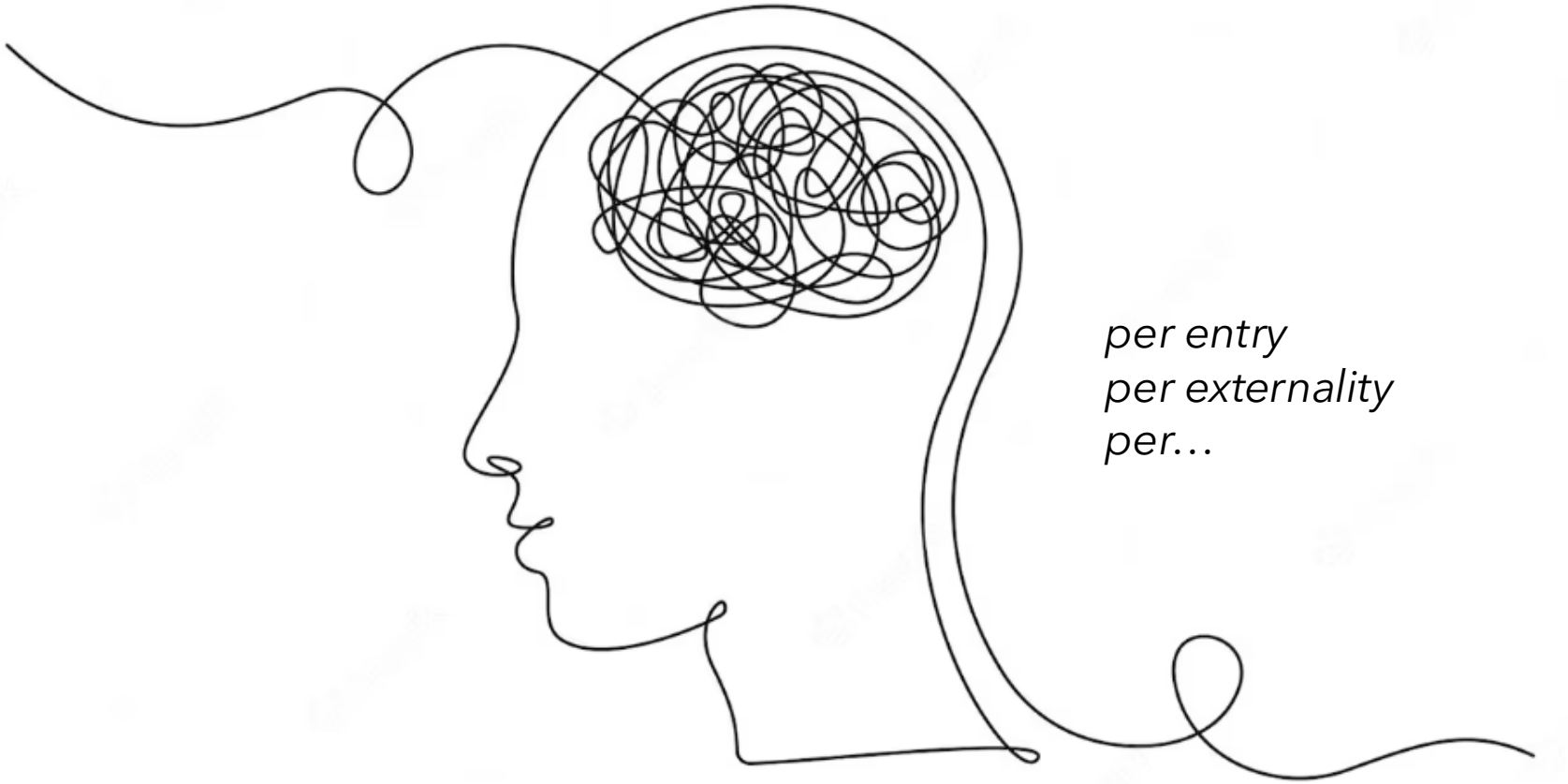
VS

Used for 1.5 years,
its (nearly) perfect for me

1. Laboratory Notebook

2. Knowledge Base
   Literature
   Experiment Results
   Writing

The "Cost per use" heuristic



per entry
per externality
per…

Acknowledgements

Soumia + Kristen

Jill – hubris

Parsa – a paper notebook success.

Jerome

- Additional Resources:

https://elginakin.github.io/posts/content/PhDTools/