

Research Data Management Support at UW-Madison
March 16, 2022 via Microsoft Teams

Tobin Magle

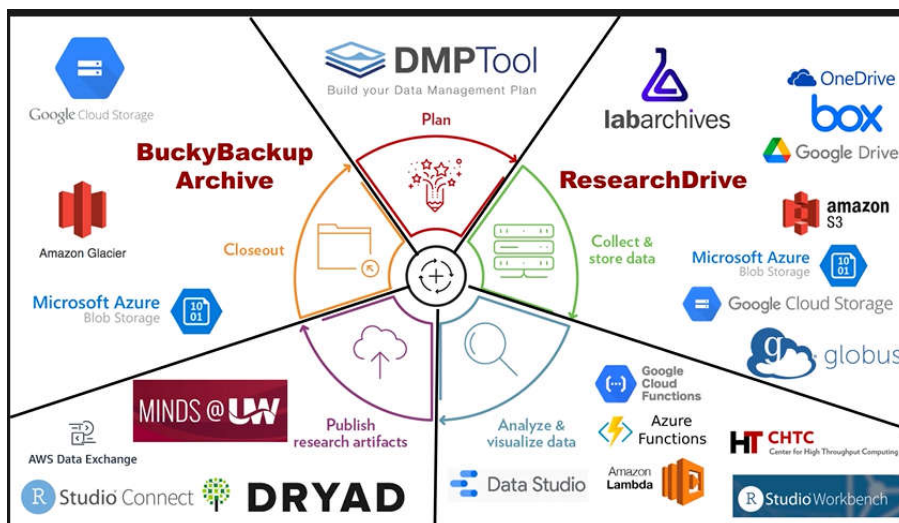
Manages Lab Archives, Electronic Lab Notebook, etc.

Think about your research workflow:

- What is going well
- What is the hardest part
- What tools/technologies are missing
- Where are you spending time where technology could help lighten your load

Research lifecycle:

- Plan
- Collect & store data
- Analyze & visualize data
- Publish research artifacts
- Closeout



PLAN

- Make decisions BEFORE you start collecting data
- A number of funding agencies require Data Management Plans
 - NSF (National Science Foundation)
 - NIH (National Institutes of Health) (starting January 25, 2023)
 - Must be submitted with grant applications

Data Management Plan:

- Data inventory
- Describing data
- Plan for preserving
- See DMP Tool (DMPtool.org)

Collect and Store:

- Do you have enough storage space
- Is your data backed up
- Access
- Compliance considerations
- See electronic lab notebook service: <https://eln.wisc.edu/request-eln-access/> (Lab Archives)
 - Supports restricted data
 - Now has a tool for inventory management (separate product, UW does not have a license for this “extra”; \$99/year/user?)
- See ResearchDrive service (5TB free; hosted in University data centers)
 - <https://it.wisc.edu/services/researchdrive/>
 - Can include links in ELN to files stored here
- Other Cloud Storage (\$) <https://it.wisc.edu/services/>
 - Amazon S3 (AWS) <https://it.wisc.edu/services/amazon-web-services/>
 - Microsoft Azure Blob Storage <https://it.wisc.edu/services/microsoft-azure/>
 - Google Cloud Storage <https://it.wisc.edu/services/google-cloud-platform/>

QUESTION: how can lab managers use this from a quality control standpoint to monitor instrument performance over time?

- Cloud compute
 - Amazon Lambda
 - Google Cloud Functions
 - Azure Functions
- Can use Python to create the code to have this done automatically
 - Implementation through Tobin’s group, but Data Science Hub would be needed to help write the actual code <https://datascience.wisc.edu/hub/>

Data Storage Finder <https://storage.researchdata.wisc.edu>

Describe your use case

UW-Madison data storage services

Select All Clear Selections

Clear Answers

1. What is your University affiliation?

Faculty or research principal investigator (PI)
 University staff
 Student
 Member of the UW-Madison Health Care Component (HCC)

2. What is the classification of your data?

Public
 Internal
 Sensitive
 Restricted

3. Does your data need to meet any

Amazon EBS Cloud-based block storage from Amazon Web Services	Amazon EFS Cloud-based NFS file storage from Amazon Web Services	Amazon S3 Cloud-based S3 object storage from Amazon Web Services	Azure Blob Storage Cloud-based object storage from Microsoft Azure	Azure Disk Storage Cloud-based block storage from Microsoft Azure
Azure Files Cloud-based file storage from Microsoft Azure	Block Storage Campus hosted high performance SAN storage for applications and databases	Box Cloud-based storage and collaboration service.	Bucky Backup Campus hosted data backup and archive service using the IBM SpectrumProtect platform	CCI Virtualization Campus hosted private cloud for virtual servers and appliances using the VMware platform
Electronic Lab Notebooks Cloud-based digital research notebook services from LabArchives	File Storage (Shared Drive) Campus hosted secure, network attached file storage	G Suite (Google Drive) Cloud-based collaboration services from Google.	Google Cloud Storage Cloud-based object storage from Google Cloud Platform.	MINDS@UW Campus hosted digital archive for the UW System libraries

Globus (glow-bus) – data transfer tool

- Easy way to transfer large files (incl. restricted data)
- Biotech Center is now using this tool

Analyze and Visualize

- How do you analyze your data

Center for High Throughput Computing

- <https://chtc.cs.wisc.edu/>
- Works with ResearchDrive and Globus

Google Data Studio – quick visualizations

Publishing Data

- Dryad via the UW <https://datadryad.org/>
- Contact libraries for assistance

Minds@UW

- Institutional repository
- <https://www.library.wisc.edu/digital-library-services/minds/>
- Can handle larger data sets
- <https://minds.wisconsin.edu/>

Closeout

- What do you do with your research when the project is done
- Julia, ELN, Jupiter Notebooks, R Studio Connect (contact Tobin for assistance)

QUESTION: code and data licenses? CCO is the way to go for data (full public access)

Cloud archival services

- Also for data you need to keep, but don't access frequently
- Useful if you already use cloud resources
- Same storage, different tiers
- Trade offs: price for access speed



Amazon Glacier



Google Cloud Storage



QUESTION: how to overcome barriers for getting all interested parties on board (many want to stick with what they know, ex. Excel)? Appeal to the person and how it will help THEM vs. benefits to the lab. PI really needs to be on board and if they will say “you must do this”, obviously easier. Talk about publication requirements – this will make that process significantly easier in the long run.

- QUESTION: who should we talk to about transitioning to ResearchDrive? Mike Laid
<https://it.wisc.edu/services/researchdrive/>

Researcher Toolkit

- <https://researchertoolkit.wisc.edu>