Research Data Services
UNIVERSITY OF WISCONSIN–MADISON

We **love** your data too.

data management plans | **personal consultations** | **education & training**

researchdata.wisc.edu & @UWMadRschSvcs

defining terms.

Photo by @kellysikkema on Unsplash

Data management → a set of practices to handle information collected and created during research; the compilation of many small practices that make data easier to find, easier to understand, less likely to be lost, and more likely to be usable in the future

Data sharing → broad umbrella term for the scholarly sharing of data between researchers, <u>both:</u>

- collaborative sharing between researchers within a research project
- broadly refers to sharing data through a repository (to make open or public!)

Open data → research data that is available for download via the internet without restriction or cost and licensed in such a way that allows for re-use and re-distribution by anyone for any purpose.

- Openness has both technical and legal requirements to it

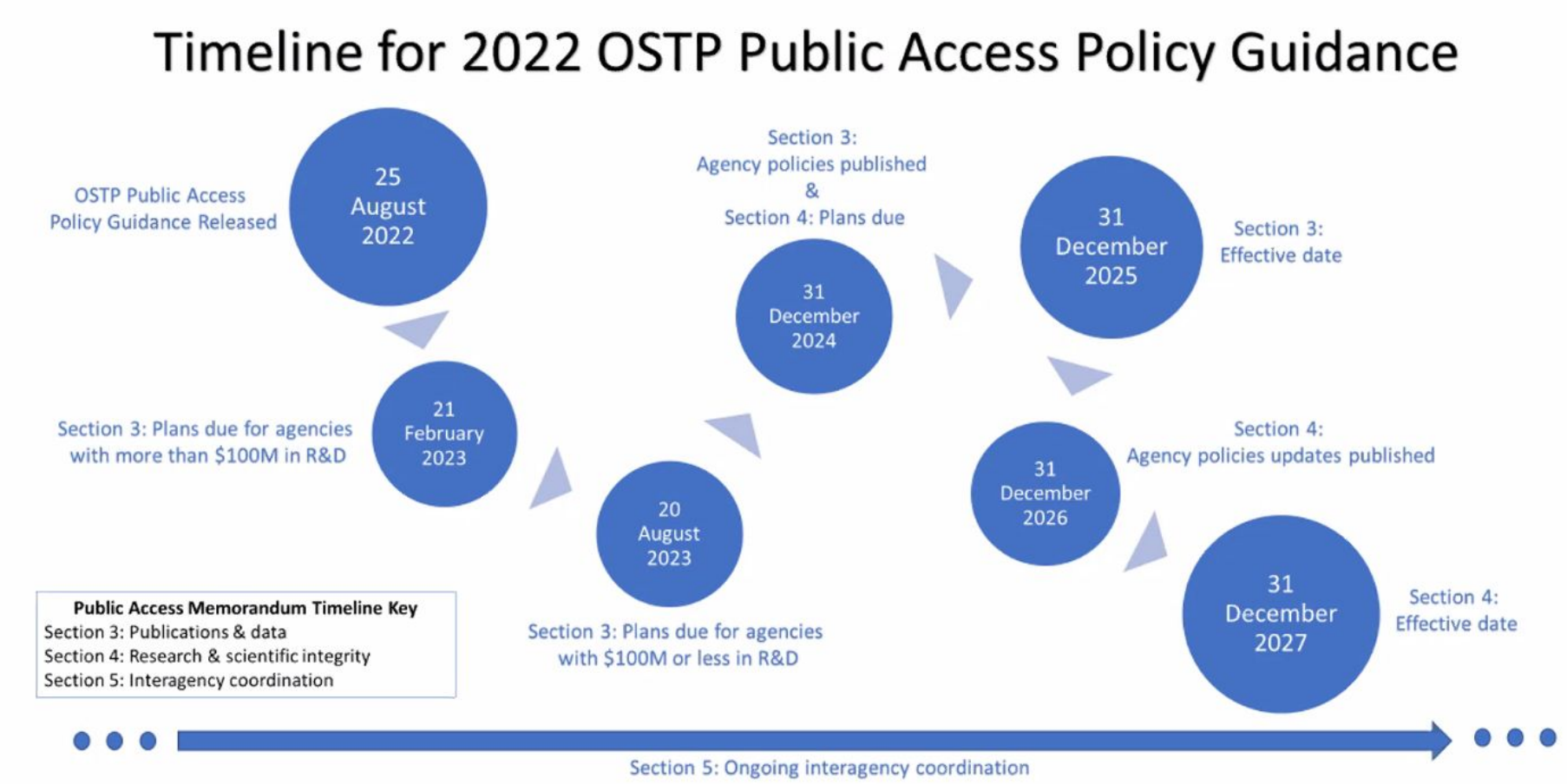
Public access to data → federal funding agency mandates that require making research data underlying publications from funded projects available to anyone for free ~~or at nominal cost~~ through a repository

- **Public access =/= open data**
- Open or moderated, restricted access data can be compliant
- Data should be shared as ethically appropriate for the dataset.
    - ‘as open as possible, as closed as necessary’.

Data repository → Managed third party platform/service with policies for collection, retention, etc. for research data that assumes stewardship of your deposited data, makes sure it is available, and conforms to best practices

- Frequently called a data archive, but is not archival storage solutions in a common IT sense
  - Think repository = making your data available to others
  - Archiving = storing it away long-term, little intention to access it
- Agencies are signaling that chosen repositories should be existing repositories that comply with the Desired Characteristics of Data Repositories for Federally Funded Research

new federal policy + past policy



Slide from “Community Forum on the 2022 OSTP Public Access Policy Guidance”, November 7, 2022

NIH rationale?

- Equitable delivery of federally funded research results and data
- Foster public trust in research
  - reproducibility & transparency in research
- Maximize research participants' contributions
- Speed research/discovery, especially during crises like COVID

NIH Data Management & Sharing Policy Overview

Required for competing applications submitted after January 25, 2023

- Submit a 2-page plan outlining how scientific data from their research will be managed and shared
  - Scientific data: data of sufficient quality to validate and replicate research findings
- **Data should be shared as soon as possible, and no later than the time of an associated publication or end of performance period (whichever comes first)**
  - Data must be shared regardless of whether the data are used to support scholarly publications
- This plan represents the minimum requirements
  - NIH ICOs many expect more specificity in their plans
- Complete list of NIH activity codes subject to the DMS Policy
  - Review your FOA to determine if the DMS Policy applies to your application

Elements to include in a DMP/DMSP

1. **Data Types:** Data to be preserved and shared (file types and sizes)
2. **Related Tools, Software, Code:** Tools and software needed to access & manipulate data, (open and/or proprietary)
3. **Common Data Standards:** Standards applied to data/metadata (protocols, data dictionaries, common data elements)
4. **Data Preservation, Access, Timelines:** Repository to be used, persistent unique identifiers used by repo, and when the data will be made available (when the work is published or the award/support period ends)
5. **Access, Distribution, Reuse Conditions:** Legal, technical or ethical considerations that may affect the extent of data sharing
6. **Oversight of Data Management:** Name and ORCID of Lead PI, a statement on broad compliance oversight from UW-Madison

Writing a Data Management & Sharing Plan

Should my data be shared?

Under element 5, a compelling rationale for limited sharing of some or all of the data produced can be explained. NIH examples include:

- informed consent will not permit or will limit the scope or extent of sharing and future research use
- existing consent prohibits sharing or limits the scope or extent of sharing and future research use
- privacy or safety of research participants would be compromised or place them at greater risk of re-identification or suffering harm, and protective measures such as de-identification and Certificates of Confidentiality would be insufficient
- explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure
- restrictions imposed by existing or anticipated agreements

FAQ - Managing and Sharing Scientific Data

Submission & Review of DMS Plans

- Submit DMS Plans and budget requests as part of the funding application or proposal.
- *Peer Review will not see or review DMS Plans, but will consider any related budget items.*
- NIH program staff will review the DMS Plan for acceptability and may request modifications prior to award as appropriate.
- Provide updates on data management and sharing activities in annual progress reports.
- If plans change over the course of the project, work proactively with NIH Program Officer to obtain review and approval of modifications.

Submission & Review of Data Management and Sharing Plans

Budgeting for Data Management & Sharing

What to include in the budget justification summary regarding DMS costs

- include the requested dollar amount and a brief justification of the proposed activities that will incur costs (no more than half a page)
  - curating data and developing supporting documentation (*data activities)
  - local data management considerations
  - preserving and sharing data through established repositories

Personnel effort included and budgeted as part of the DMSP

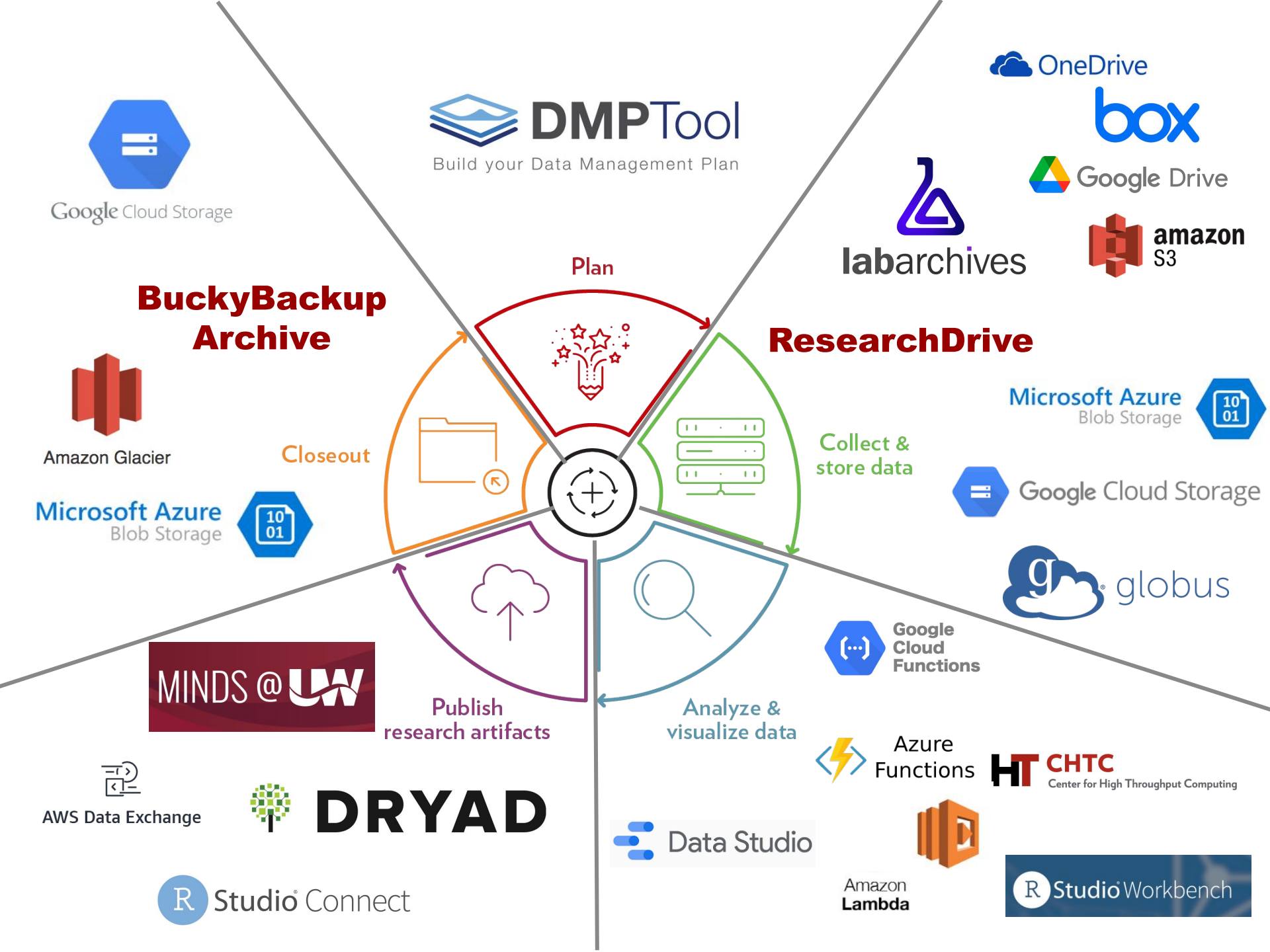
- salary and fringe benefits corresponding to the time it takes personnel to undertake data activities *
- must be included in the single line item on the R&R Budget Form in section F. Other Direct Costs

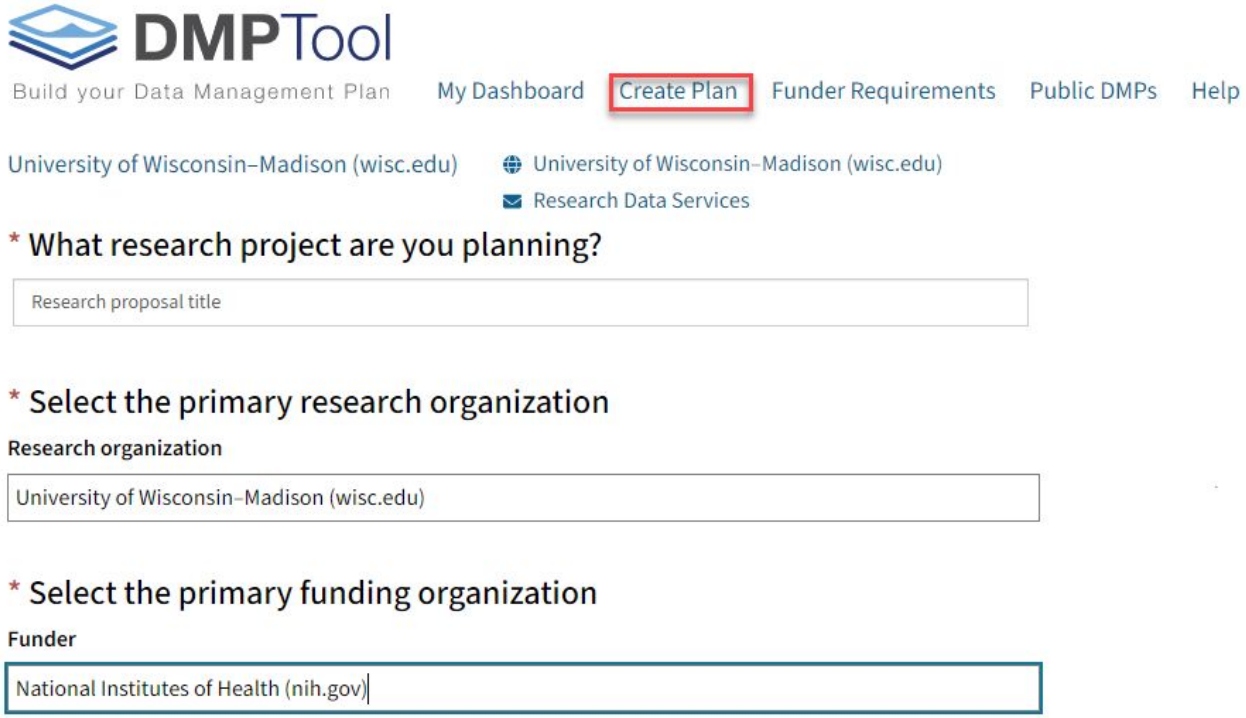
All allowable costs must be incurred during the performance period

generalist repository ecosystem initiative (GREI)

- Office of Data Science Strategy aims for GREI
  - "...further enhance the biomedical data ecosystem and help researchers find and share data from NIH-funded studies in generalist repositories."
  - The 6 repositories will work together to "establish consistent metadata, develop use cases for data sharing, train and educate researchers on FAIR data and the importance of data sharing,"
  - Increase discoverability of data, improve reproducibility and reuse of data
- NIH is working with 6 data repositories
  - Dryad (UW-Madison is a member!)
  - Dataverse
  - Figshare
  - Mendeley Data
  - Open Science Framework (OSF)
  - Vivli

https://datascience.nih.gov/news/nih-office-of-data-science-strategy-announces-new-initiative-to-improve-data-access

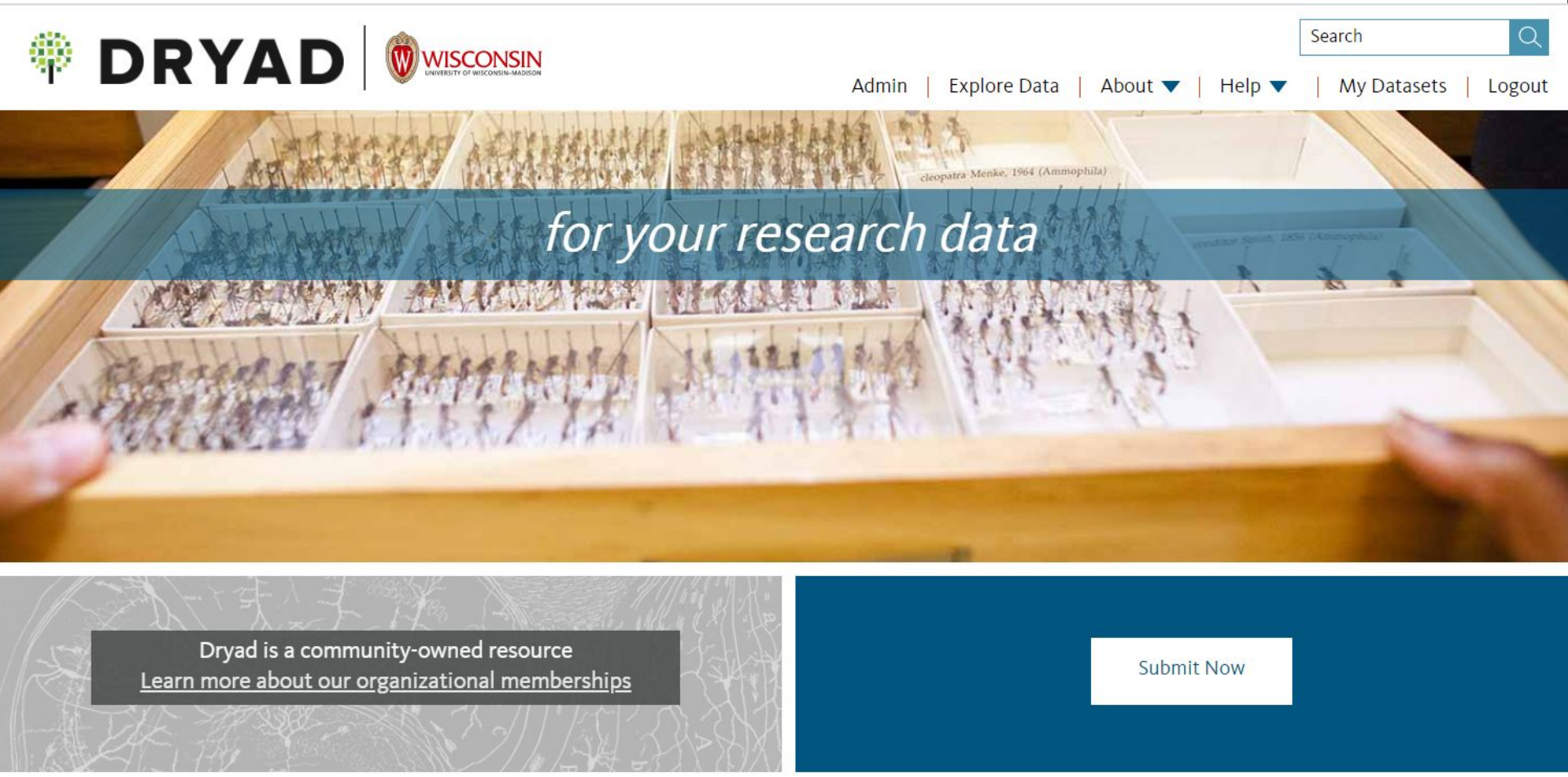

DMPTool
Build your Data Management Plan
Google Cloud Storage
OneDrive
box
Google Drive
labarchives
amazon S3
BuckyBackup Archive
ResearchDrive
Plan
Amazon Glacier
Closeout
Microsoft Azure Blob Storage
Collect & store data
Microsoft Azure Blob Storage
Google Cloud Storage
globus
MINDS @ UW
Publish research artifacts
Analyze & visualize data
Google Cloud Functions
Azure Functions
CHTC
Center for High Throughput Computing
AWS Data Exchange
DRYAD
Data Studio
Amazon Lambda
RStudio Connect
RStudio Workbench



DMPTool

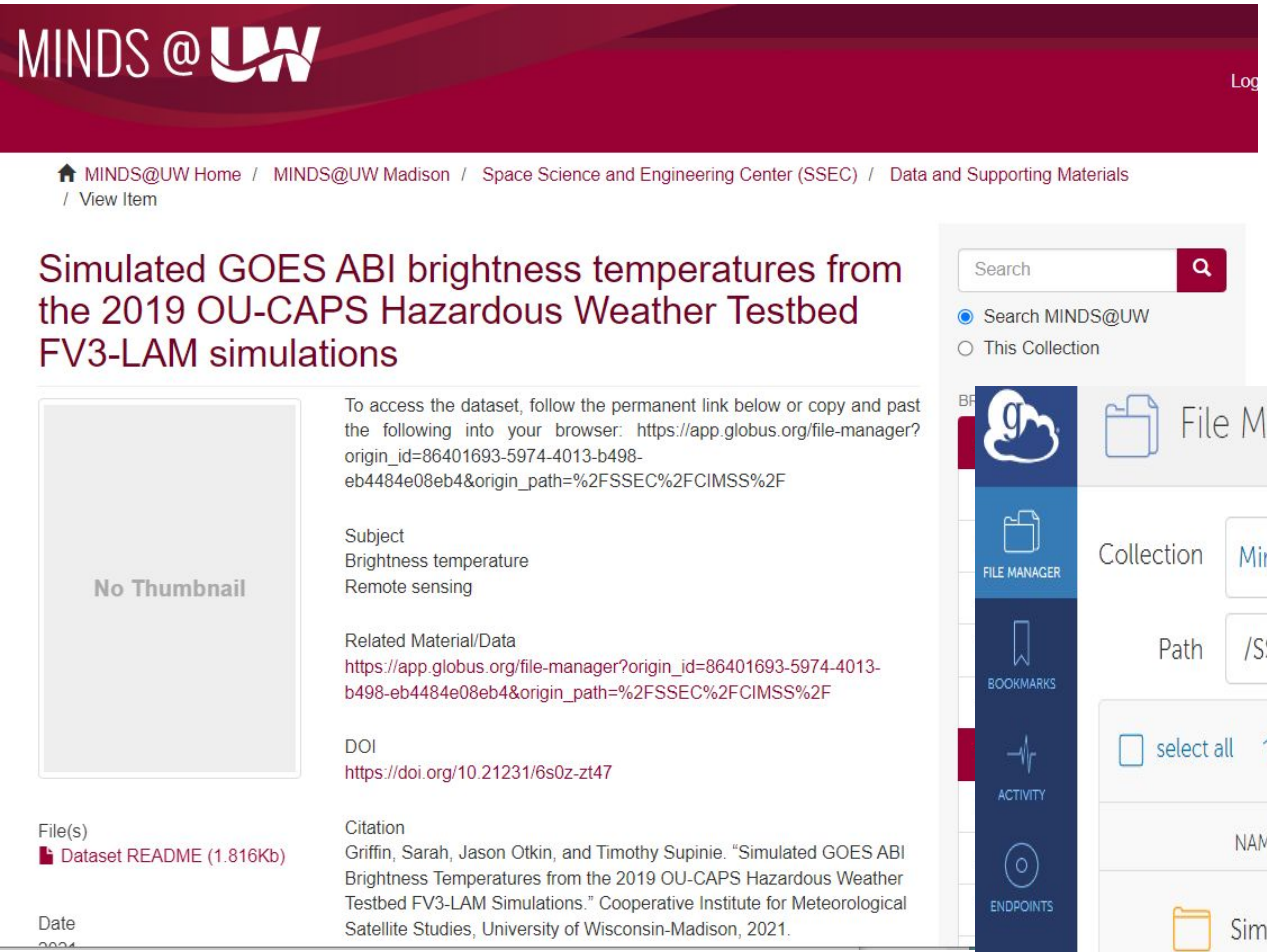
https://dmptool.org/

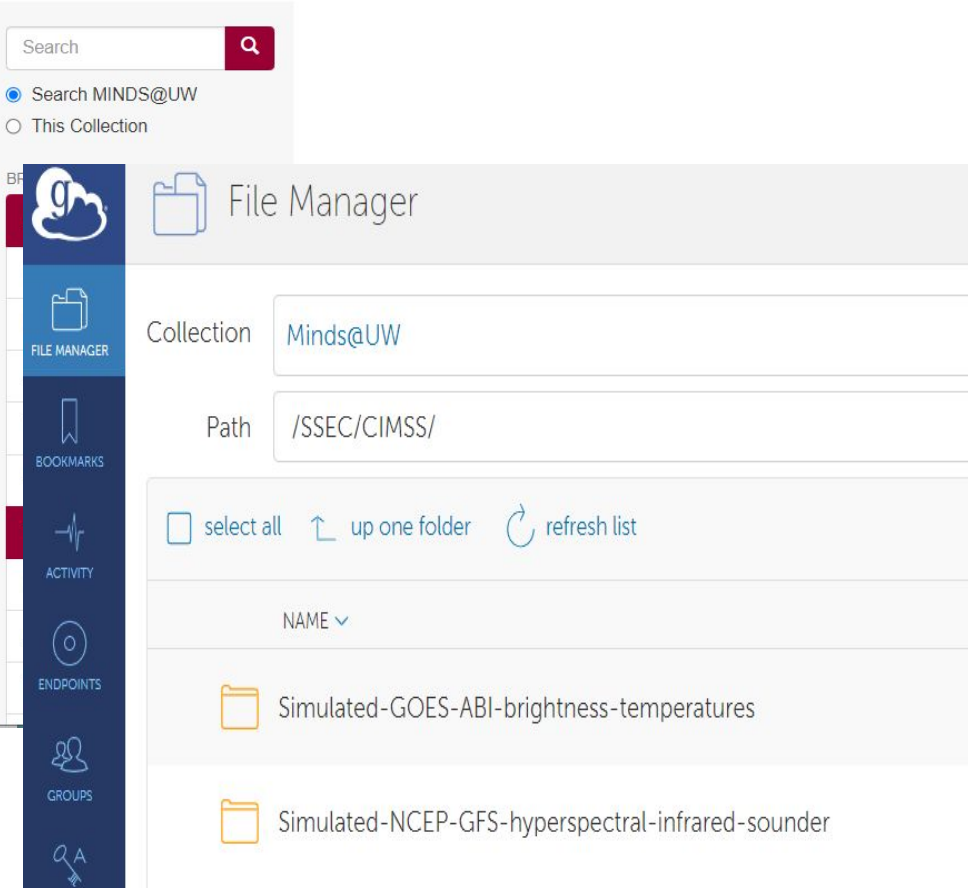
Dryad



https://datadryad.org/

MINDS@UW





minds.wisconsin.edu

ICPSR

ICPSR
Sharing data to advance science

Log In | Giving

Studies, publications, variables, webpages

Home | Find Data | Share Data | Membership | Summer Program | Teaching & Learning | Data Management | About

Find Data

Search | view all

search tips

Browse

Topics / Series / Thematic data collections

- List studies for which online analysis is available
- List self-published data, including replication datasets
- List studies with learning guides

New/Updated Data Releases:

- In the last week

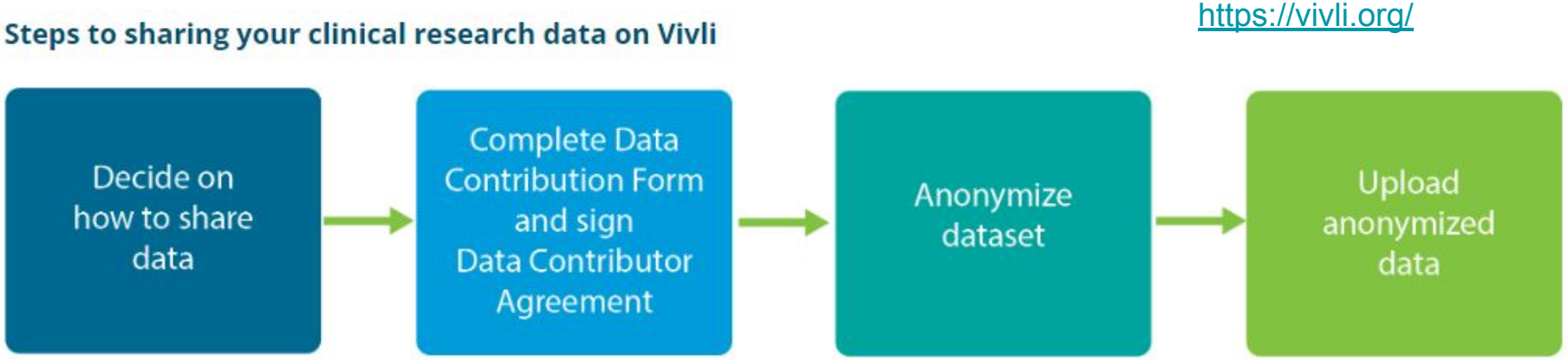
Statistics

16,575 studies

5,888,749 variables

Most Popular Search Terms

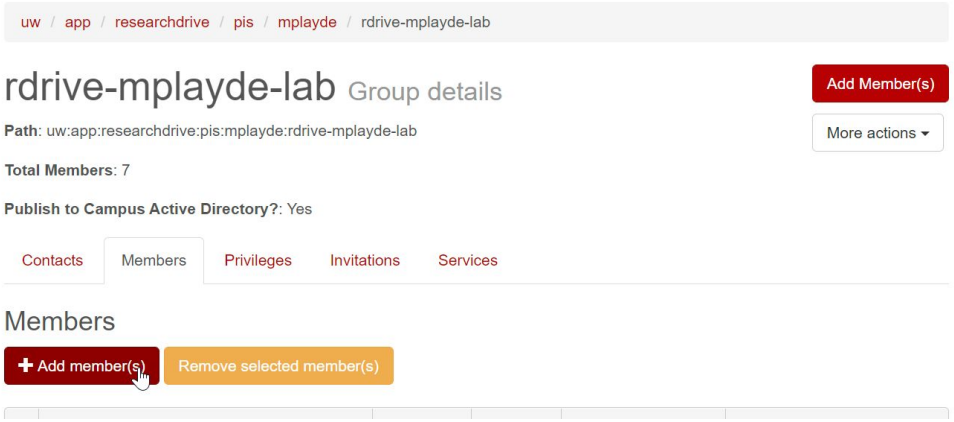
https://vivli.org/

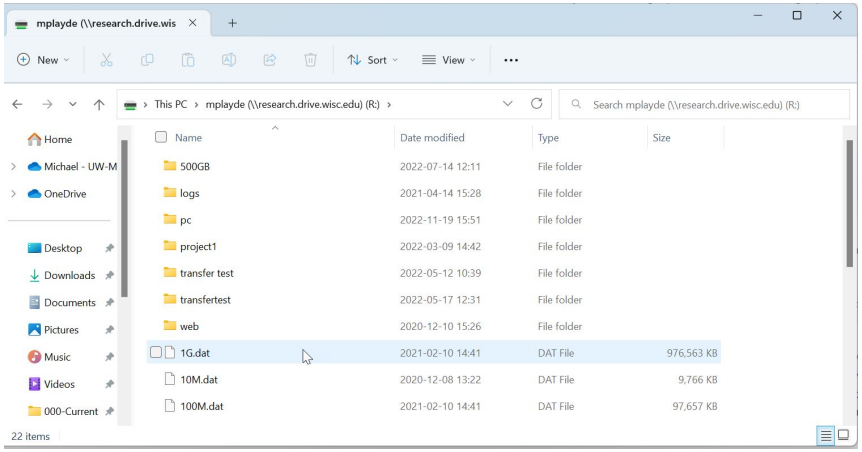


- Managed access repository
  - Users request access to the data for a specific research proposal
- Data collected
  - anonymized individual participant-level data (IPD)
  - raw data collected during a clinical trial
- Vivli Data Use Agreement, user agrees to:
  - maintain data security
  - refrain from any attempts to reidentify research participants
  - submit a valid scientific question, including a statistical analysis plan
- Vivli agrees to store data for 10 years

ResearchDrive?

- File storage service for research data - 5 TB per PI
- Share data with UW colleagues or external collaborators
- Approved tool for exchanging HIPAA data
- Cross platform - Linux, Mac OS, Windows
- High speed connectivity on campus or remotely accessible via VPN and the web.



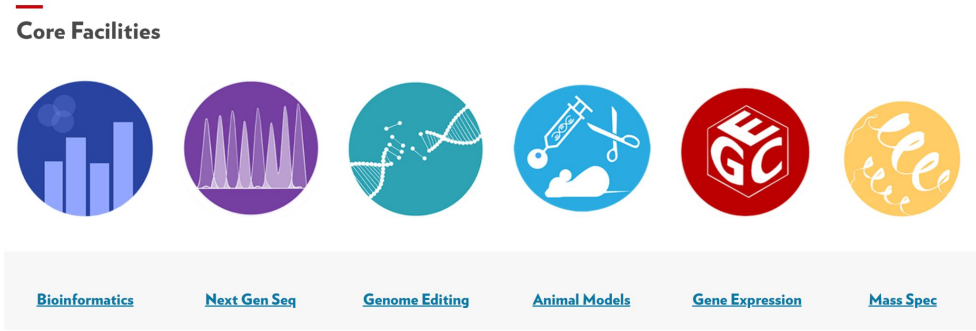


it.wisc.edu/services/researchdrive

Gather data

Research Cores

- Data from instruments
- Biotech Center and UW-CCC



Data Protection

- Automated data protection
- Snapshots and offsite mirroring of data
- Ransomware protection including offline backups

Analyze & visualize data

Endpoint Computers

- UW Endpoint Management and Security standards
- Can easily share data with other departments

Computing Centers

- Easy to move data in/out of computing centers
- Platform X and SSCC

Share data with collaborators

UW-Madison NetID

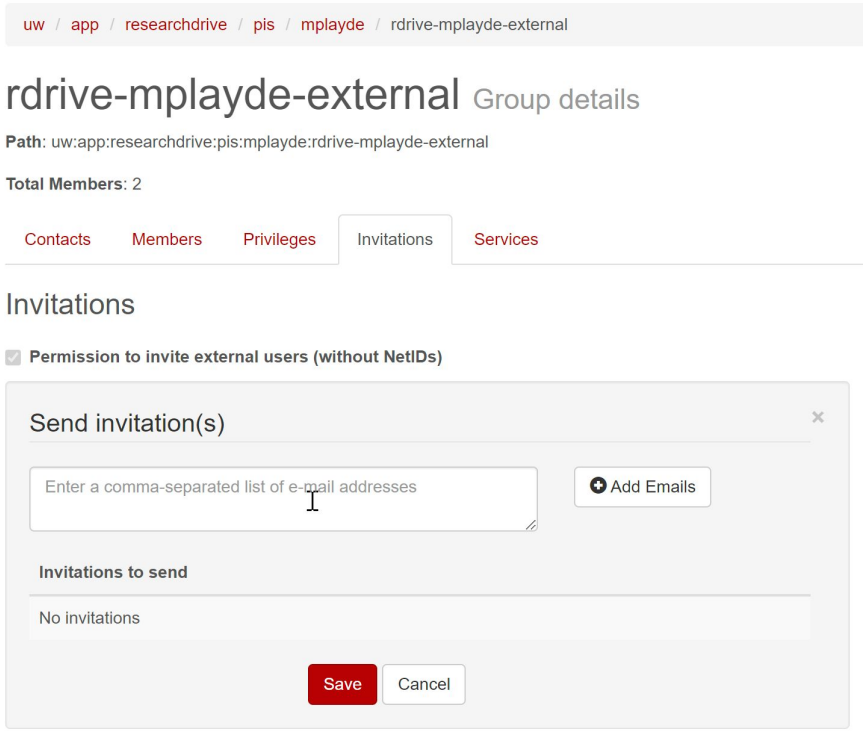
- Share data with anyone with a NetID

External Collaborators

- Send invites to create a NetID with VPN and Duo-MFA

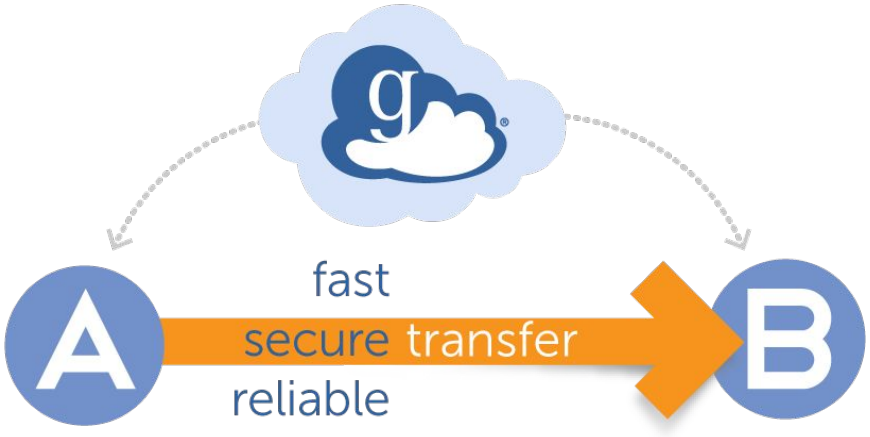
Remote Access

- Collaborators use UW VPN and Duo-MFA



Globus?

- Web-based research collaboration platform that supports secure data sharing and transfers
- Connects disparate storage systems together through Globus endpoints



https://it.wisc.edu/services/globus

Data transfer

Research Cores

- Cryo-EM and Biotech Center

Computing Centers

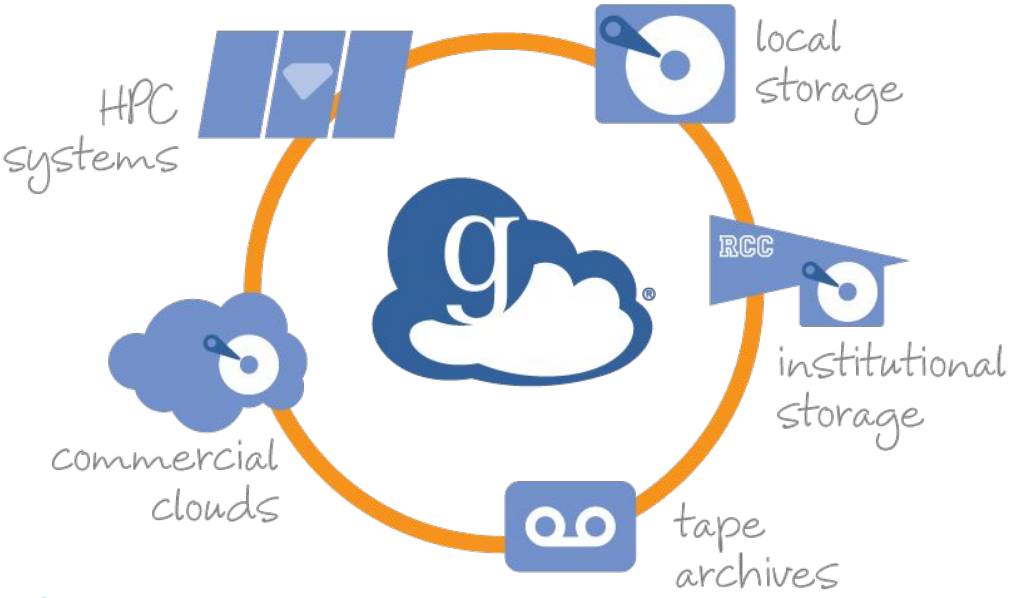
- SSCC and Platform X

Campus Storage

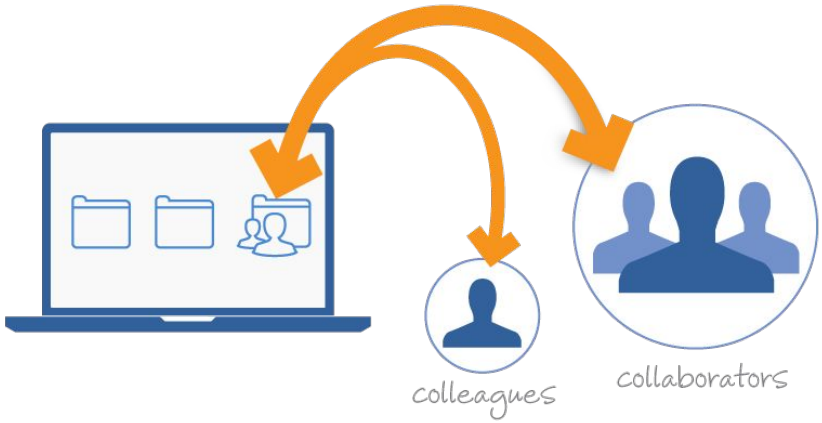
- ResearchDrive, Shared Drive,and S3

Public Clouds

- AWS, Google Cloud, and Microsoft Azure

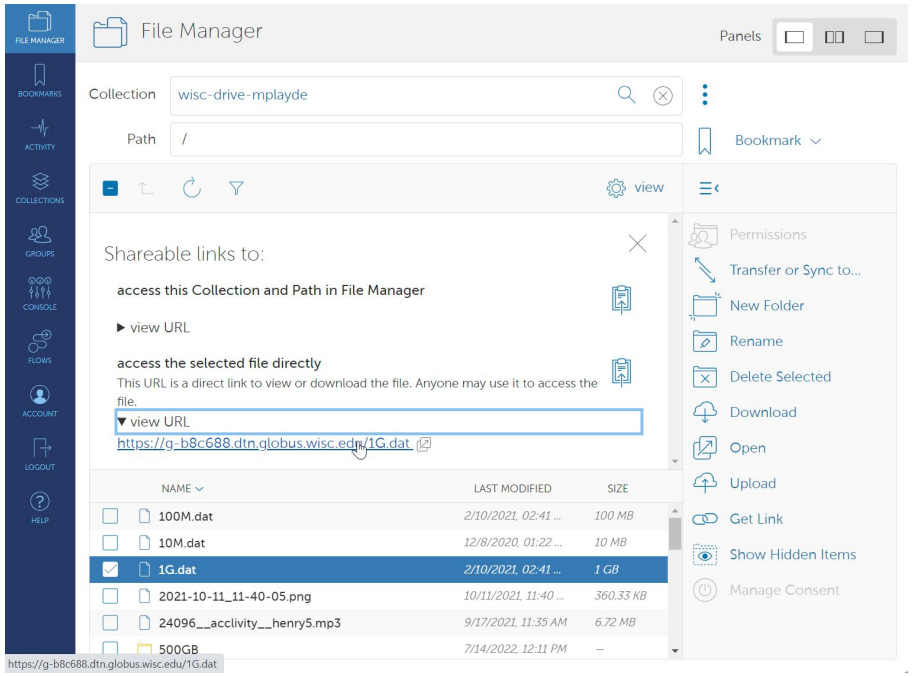


Data sharing



Securely exchange data with external collaborators

- Allows sharing between institutions using existing identities
- Approved tool for exchanging HIPAA data
- Supports sharing files with web links



Publish research artifacts

Publish open access or moderated data sets through the web

- Deposit large data sets in Minds@UW with Globus
- Globus platform integration for web applications and data portals

NIH DMSP Drop-ins & DMS Plan Office Hours

Drop-in Information Sessions

Health Sciences Learning Center 3330

- Every Wednesday, January 18 - March 15, 2023 (Noon-1PM)

DMS Plan Open Office Hours

Health Sciences Learning Center 2320

- Every Tuesday, January 10 - March 28 (11:30AM-1:30PM)
- Every 3rd Tuesday, January – March (4:30PM-6:30PM)

QUESTIONS?

TRISHA ADAMUS

adamus@wisc.edu

CAMERON COOK

cameron.cook@wisc.edu

MICHAEL LAYDE

michael.layde@wisc.edu

researchdata.wisc.edu

Photo by Wilhelm Gunkel on Unsplash

Researcher Toolkit

Your Guide to Researcher Resources at UW-Madison

Lifecycle | Resource List | Acronyms | Feedback | About

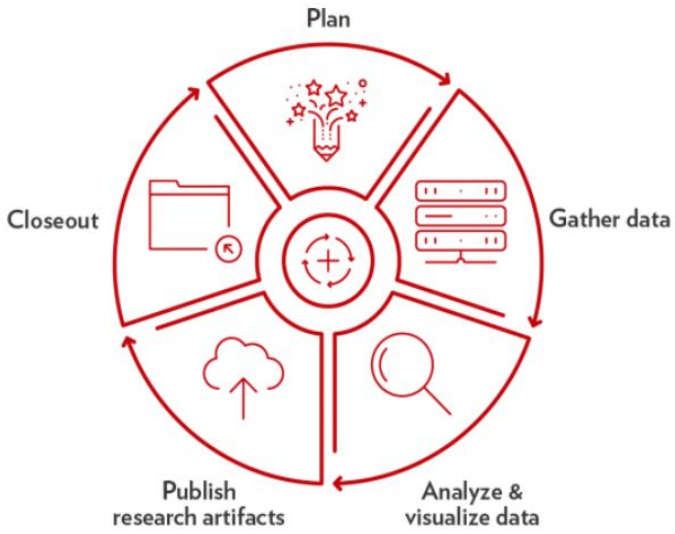
What is the research lifecycle?

The research lifecycle

The research lifecycle includes planning, collecting & storing data, analyzing & visualizing data, publishing research artifacts, and closing out the findings.

Plus outreach, compliance, reporting, and security happening throughout the process.

About Research Lifecycle Components



https://researchertoolkit.wisc.edu/

UW-Madison resources.

Research Data Services - data management help, writing DMPs, data sharing/archiving

Data Science Hub - Coding Meetup, training, office hours

- Data & Software Carpentry - sign up for the Data Science Hub newsletter

CALS Statistical Consulting - study design, data analyses, software use

Social Science Computing Cooperative - software training, SILO

UW-Madison Libraries - finding data, copyright, publishing questions, etc.

IRB, IRB Guidance

MINDS@UW - data sharing via UW institutional repository

Dryad - data repository available via UW-Madison membership

Researcher Toolkit

Data Storage Finder