NIH Scientific Data Sharing

What you need to know and resources to help with implementation

Clista Clanton, MSLS
and
Jana Herrmann

December 1, 2022

NIH Data Management & Sharing Plan (DMSP) Policies

- What are they?
- Why is this important?
- When do they go into effect?
- Tools and resources to help

[NIH Data Management & Sharing Policy Overview]

Planning & Budgeting for Data Management & Sharing

Prospectively planning for how scientific data will be managed and ultimately shared is a crucial first step in optimizing the reach of data generated from NIH-funded research. Investigators and institutions are encouraged to consider these crucial elements early in research planning.

Tip: Consider consulting institutional resources such as librarians and data managers to help plan effectively!

Determine if your proposed research is subject to the DMS policy.

Identify appropriate methods/approaches and repositories for managing and sharing scientific data.

Develop a Plan for managing and sharing scientific data and submit this Plan within the funding application or proposal.

** Note that applications subject to both the DMS Policy and the GDS Policy will submit a single Plan.*

Estimate and request funds for data management and sharing activities if not already covered by institution or other sources.

What are the new NIH DMSP policies? Why are they important?

Final NIH Policy for Data Management and Sharing

Purpose: to promote the management and sharing of scientific data generated from NIH-funded or conducted research.

Clarifying Expectations for Sharing Scientific Data

The final DMS Policy **does not create a uniform requirement to share all scientific data**. Unlike a **requirement for submission of Plans**, which can be implemented across various funding mechanisms and types of research with little variation, appropriate data sharing is likely to be varied and contextual. **Through the requirement to submit a Plan**, **researchers are prospectively planning for data sharing**, which we anticipate will **increasingly lead researchers to integrate data sharing into the routine conduct of research**. Accordingly, we have included in the final DMS Policy an **expectation that researchers will maximize appropriate data sharing when developing Plans**.

Definition of "Scientific Data"

The final DMS Policy defines Scientific Data as: **"The recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications. Scientific data do not include laboratory notebooks, preliminary analyses, completed case report forms, drafts of scientific papers, plans for future research, peer reviews, communications with colleagues, or physical objects, such as laboratory specimens."** The data should be of sufficient quality to validate and replicate research findings. We encourage reasonable efforts to digitize data, recognizing that digitizing data may be a technical factor that may limit the sharing of data.

Timing of Submission of Data Management and Sharing Plans

The final DMS Policy **requires submission of a Plan for extramural grants at application**. This approach is more conducive **to achieving NIH's goal of promoting a culture in which data management and sharing are recognized to be an integral component of a biomedical research project, rather than an administrative or additive one.**

What are the new NIH DSMP policies? Why are they important?

Final NIH Policy for Data Management and Sharing

Assessment of Plans

The final DMS Policy maintains NIH Program Staff assessments of Plans' merits. However, **peer reviewers may comment on the proposed budget for data management and sharing, although these comments will not impact the overall score.** Over time, and through these reviews, we hope to learn more about what constitutes reasonable costs for various data management and sharing activities across the NIH portfolio of research.

NIH Institutes, Centers and Offices (ICO) Consistency of Data Sharing Expectations

This Policy affords NIH ICOs the opportunity to meet the goals of this Policy in ways that enhance their respective science. However, **we intend to promote consistency on some key tenets of the final DMS Policy, such as the requirement for submission of Plans and the timing of their submission. The DMS Policy represents the minimum requirements for the NIH, but NIH ICOs may expect more specificity in Plans.**

Data Derived from Human Participants

The final DMS Policy **does not introduce new requirements for protections for research with human participants.** Existing laws (e.g., Certificates of Confidentiality), regulations (e.g., the Common Rule), and policies (e.g., the NIH Genomic Data Sharing Policy) continue to apply. **However, through this Policy and associated supplemental information and other activities, NIH promotes thoughtful practices regarding the treatment of data derived from human participants**. **First, we encourage investigators to consider, while developing their Plans, how to address data management and sharing in the informed consent process, such that prospective participants will understand what is expected to happen with their data. Second, we note that any limitations on subsequent use of data (which may apply to non-human data as well) should be communicated to those individuals or entities preserving and sharing the scientific data. Finally, we highlight the importance of researchers considering whether, in choosing where and how to make their data available (if not already specified by an FOA or funding NIH ICO expectation), access to scientific data derived from humans should be controlled, even if de-identified and lacking explicit limitations on subsequent use.**

What are the new NIH DSMP policies? Why are they important?

Final NIH Policy for Data Management and Sharing

When Data Are Expected To Be Shared

"[s]hared scientific data should be made accessible as soon as possible, and no later than the time of an associated publication, or the end of the award/support period, whichever comes first." Data that do not form the basis of a publication produced during the award period should be shared by the end of the award period. A single research project may take advantage of both approaches. Namely, researchers may share data underlying publication during the period of award but may share other data that have not yet led to a publication by the end of the award period.

How Long Data Should Be Available

We have indicated a framework for helping researchers think through a minimum time period for data availability. **Existing requirements and expectations set forth through, for example, applicable record retention requirements, repository policies, and journal policies may guide researchers as they seek to define minimal periods for data availability. However, we encourage researchers to propose longer time periods that may be informed by other factors, such as anticipated value of the dataset for the scientific community and the public.**

Where to Share Scientific Data

The final DMS Policy **strongly encourages the use of established repositories to the extent possible. This reflects NIH's preference that scientific data be shared and preserved through repositories, rather than kept only by the researcher or institution and provided on request,** with the recognition that this is not always a practical or even a preferred approach. In addition, we have released the Supplemental Information to the NIH Policy for Data Management and Sharing: Selecting a Repository for Data Resulting from NIH-Supported Research, which will aid researchers as they choose suitable repositories for the preservation and sharing of data.

What are the new NIH DMSP policies? Why are they important?

Final NIH Policy for Data Management and Sharing

Page Limit and Template for Plans

In the final supplemental information, **we have noted the elements to be addressed in two pages or less, indicating that these descriptions need not be long narratives.** In addition, short Plans are anticipated to limit researcher burden.

The Acceptability of "To Be Determined" as a Response to Plan Elements

The final Supplemental **Information eliminates the language that a response of "to be determined" is acceptable**. We do not expect researchers to necessarily have all details at the application stage, but we encourage researchers to fill out Plans to the best of their knowledge and ability, so the Plans may be appropriately assessed. **We also note that adherence with NIH ICO-approved Plans is a requirement of the final DMS Policy.** As indicated in the final DMS Policy, **researchers will have opportunities to update their Plans throughout the course of their awards, subject to NIH ICO approval.**

The Use of Persistent Unique Identifiers (PIDs)

The final Supplemental Information asks researchers to describe how the scientific data will be findable and identifiable, i.e., via a persistent unique identifier or other standard indexing tools. This wording change is meant to highlight the importance of using a PID or other standard indexing tool so the data are findable, which is a key component of the FAIR (Findable, Accessible, Interoperable, and Reusable) Principles.

Data Security

We have removed the prompt for researchers to address provisions related to the security of scientific data. While we agree with the importance of appropriate data security measures, **we believe that technical provisions regarding data security are more appropriately addressed by the institutions and repositories preserving and sharing the scientific data.** While data may remain with an institution prior to submission to a data repository, the DMS Policy is not designed to set any new standards for institutional data security practices.

What are the new NIH DMSP policies? Why are they important?

Final NIH Policy for Data Management and Sharing

Timelines for Using Funds for Data Management and Sharing Activities

Personnel costs required to perform the types of data management and sharing activities described in the final Supplemental Information are allowable. Regarding the availability of data beyond the end of the project, which is crucial to achieving the goals of the DMS Policy, the final Supplemental Information clarifies that **fees for long-term data preservation and sharing are allowable, but funds for these activities must be spent during the performance period**, even for scientific data and metadata preserved and shared beyond the award period. **NIH funds cannot legally be spent after the award period.**

The recognition that more open sharing can lead to faster advances and treatments has led to an unprecedented worldwide effort to openly share publications and data related to both SARS-CoV-2 (the novel coronavirus that causes COVID-19) and coronaviruses more generally. While this is a specific example of an urgent public health need, patients, families, and patient advocacy groups consider the diseases and conditions that affect them to be of equal urgency, as do those who research these diseases and conditions and treat affected patients. With public input, NIH has worked to develop and refine this DMS Policy, the goal of which is to increase the sharing of scientific data generated from NIH-funded research to ultimately enhance health, lengthen life, and reduce illness and disability.

When do the new NIH DMSP polices go live?

On/After January 25, 2023

Applications for Receipt Dates BEFORE Jan 25 2023

Applications for Receipt Dates ON/AFTER Jan 25 2023

NIH has issued the Data Management and Sharing (DMS) policy (effective January 25, 2023) to promote the sharing of scientific data. Sharing scientific data accelerates biomedical research discovery, in part, by enabling validation of research results, providing accessibility to high-value datasets, and promoting data reuse for future research studies.

Under the DMS policy, NIH expects that investigators and institutions:

- Plan and budget for the managing and sharing of data
- Submit a DMS plan for review when applying for funding
- Comply with the approved DMS plan

Individual NIH Institutes, Centers, or Offices may have additional policies and expectations (see NIH Institute and Center Data Sharing Policies).

Select each step below to learn more.

Scope of DMSP

https://neuroscienceit.medium.com/how-to-determine-if-a-research-lab-is-right-for-you-2d18ccbed44a

- Applies to all research, funded in whole or in part by NIH, that results in the generation of "**scientific data**“

- "**Scientific data**" is defined as: "*the recorded factual material commonly accepted in the scientific community as of sufficient quality to validate and replicate research findings, regardless of whether the data are used to support scholarly publications*.“

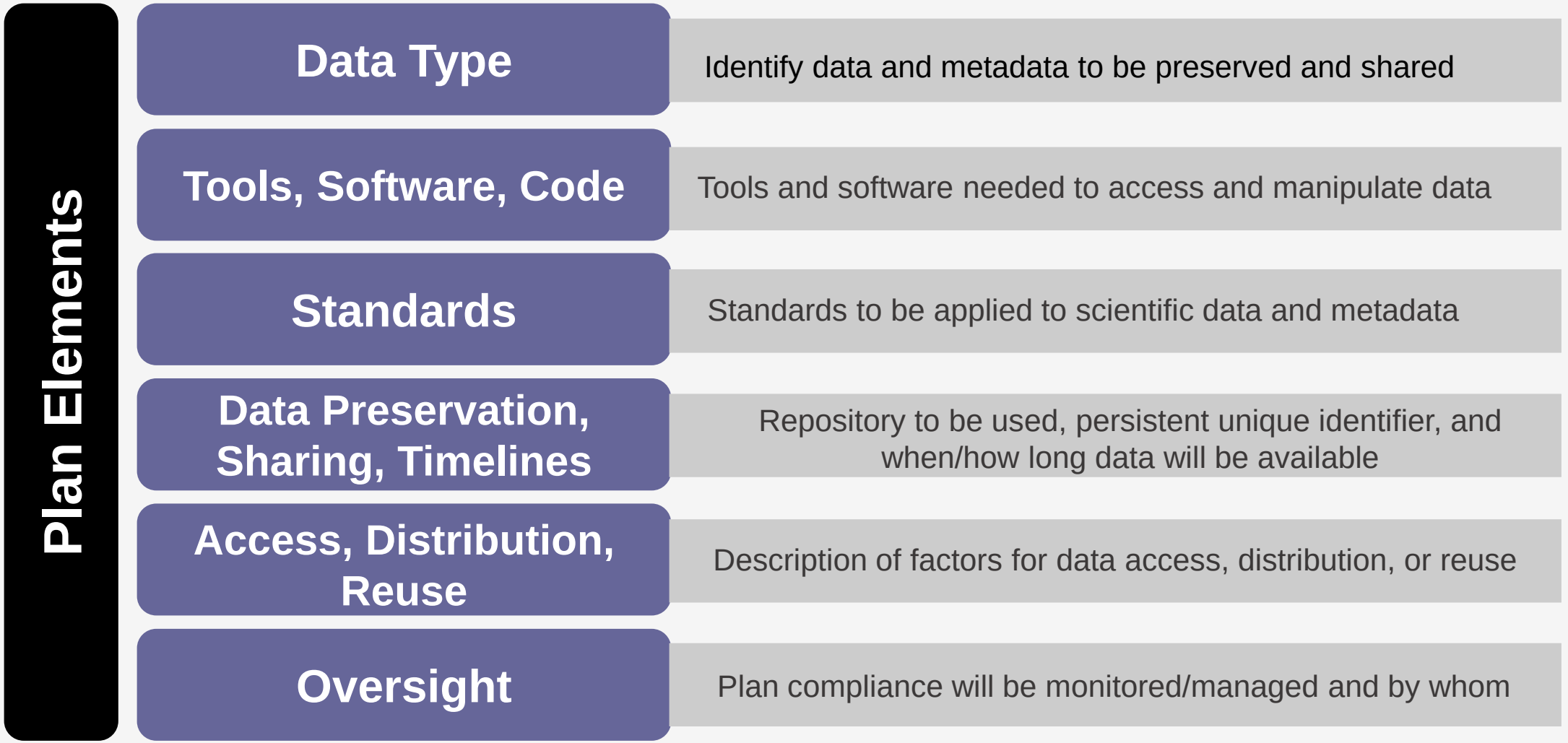
- **Does not apply to funding that does not generate data**

Exclusions from the DMSP

Scientific data not included:

- Data **not** necessary for or of sufficient quality to validate and replicate research findings,
- Laboratory notebooks
- Preliminary analyses
- Completed case report forms
- Drafts of scientific papers
- Plans for future research
- Peer reviews
- Communications with colleagues
- Physical objects (e.g., laboratory specimens)

Elements of a DMS Plan



Tools and resources to help

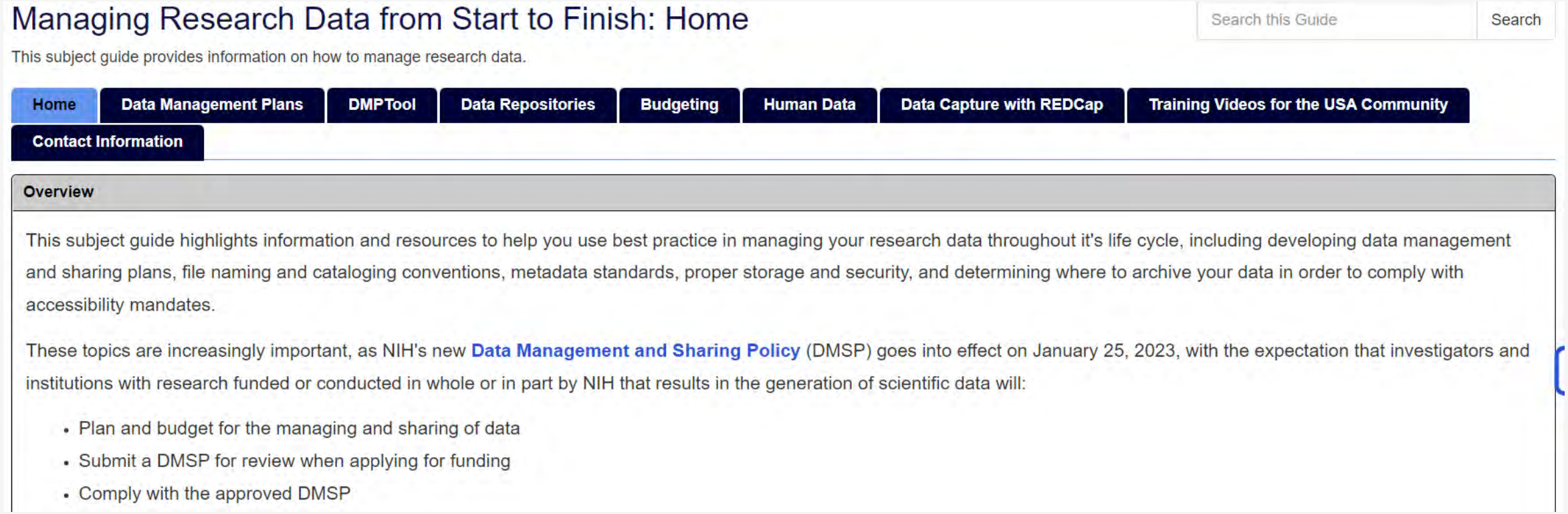
DMPTool (https://dmptool.org/)

The DMPTool is a free, open-source, online application that helps researchers create data management plans. The tool provides a click-through wizard for creating a DMP that complies with funder requirements. It also has direct links to funder websites, help text for answering questions, and data management best practices resources.

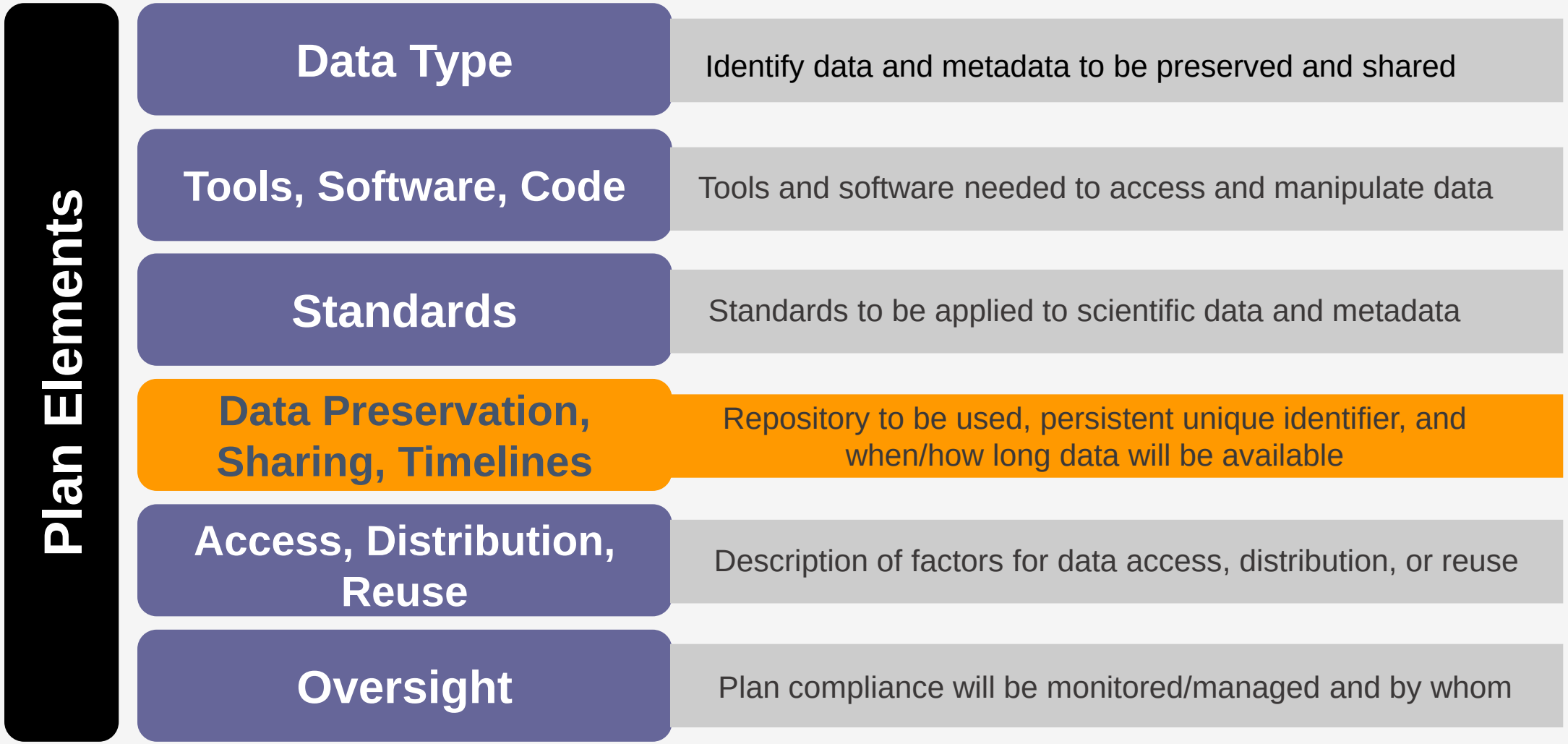
- Free
- When you log in, you will be directed to “My dashboard.” From here, you can create, edit, share, download, copy, or remove any of your plans. You will also see plans that have been shared with you by others.
- If others at your institution/organization have chosen to share their plans internally, you will see a second table of organizational plans. This allows you to download a PDF and view their plans as samples or discover new research data. Additional samples are available in the list of public plans.
- Dusty Layton serving as the interim DMPTool administrator until the Biomedical Library hires the Research Data & Scholarly Communications Librarian, a new position.

Tools and resources to help

[Managing Research Data from Start to Finish](https://libguides.southalabama.edu/research_data_mgt) subject guide (https://libguides.southalabama.edu/research_data_mgt)



Elements of a DMS plan: Data preservation, sharing, timelines



NIH Data Sharing Landscape

NIH encourages the use of established data repositories

Improves the **FAIRness** of Data **(F**indable, **A**ccessible, **I**nteroperable, **R**eusable)

Factors for choosing a repository: **sensitivity of data**, **size of dataset**, and **complexity of data**

Desirable Characteristics for All Data Repositories

- Unique Persistent Identifiers
- Long-Term Sustainability
- Metadata
- Curation and Quality Assurance
- Free and Easy Access
- Broad and Measured Reuse
- Clear Use Guidance
- Security and Integrity
- Confidentiality
- Common Format
- Provenance
- Retention Policy

Additional considerations for human data

- **Clear Use Guidance**
- **Retention Guidelines**
- **Fidelity to Consent**
- **Restricted Use Compliant**
- **Privacy**
- **Plan for Breach**
- **Download Control**
- **Violations**
- **Request Review**

https://denovaresearch.com/what-are-clinical-trial-facilities/

Finding and selecting a repository

For some programs and types of data – NIH Institutes, Centers, or Offices (ICO) and Funding Opportunity Announcements (FOAs) identify or designate particular NIH-supported data repositories (or sets of repositories) to be used for sharing data.

NIH ICO/ FOAs

Designated repository

Institute or Center	Data Sharing Policy Name	Description of Data Sharing Policy	Repositories
HEAL	HEAL Public Access and Data Sharing	Through the NIH HEAL Initiative Public Access and Data Sharing Policy (the Policy), NIH seeks to create an infrastructure that addresses the need for researchers, clinicians, and patients to collaborate on sharing their collective data and knowledge about opioid misuse and pain to provide scientific solutions to the opioid crisis. Under the Policy, applicants for extramural research funding (grants, cooperative agreements, contracts, and other transactions; "Applicants") for NIH HEAL Initiative Research Projects are required to submit a Public Access and Data Sharing Plan that (1) describes their proposed process for making resulting Publications and, to the extent possible, the Underlying Primary Data immediately and broadly available to the public or (2), if applicable, provides a justification to NIH if such sharing is not possible. Underlying Primary Data should be made as widely and freely available as possible while safeguarding the privacy of participants and protecting confidential and proprietary data.	Various HEAL-Compliant repositories
NCI	Cancer Moonshot℠ Public Access and Data Sharing Policy	The primary goal of NCI's Cancer Moonshot℠ is to significantly accelerate cancer research discovery and meaningful implementation. The Cancer Moonshot Public Access and Data Sharing Policy addresses the recommendation of the Blue Ribbon Panel's Enhanced Data Sharing working group to the National Cancer Advisory Board that researchers, clinicians, and patients should collaborate in sharing their collective data and knowledge about cancer to accelerate progress towards improving cancer outcomes. Under this policy, applicants for Cancer Moonshot Research Projects are required to submit a "Public Access and Data Sharing Plan" that describes their proposed process for making, to the extent possible, resulting Publications and the Underlying Primary Data immediately and broadly available to the public. Investigators applying for Cancer Moonshot funds must provide a justification to NCI if such sharing is not possible. Through the Cancer Moonshot Public Access and Data Sharing Policy (the "Policy"), NCI seeks to create an infrastructure that addresses the recommendation of the Blue Ribbon Panel's Enhanced Data Sharing working group to the NCAB that researchers, clinicians, and patients collaborate in sharing their collective data and knowledge about cancer to accelerate progress towards improving cancer outcomes. The NCAB subsequently accepted this recommendation and recommended it to NCI.	Genomic Data Commons, dbGaP, TCIA
NCI	NCI Clinical Trial Access Policy	NCI believes that the full value of NCI-supported Interventional Clinical Trials can be realized only if the results of clinical trials are published as rapidly as possible. The Clinical Trial Access Policy aims at ensuring public availability of results from NCI-supported clinical trials from all NCI-funded research grants, cooperative agreements, and/or contracts that support covered interventional clinical trials. Review the NCI Clinical Trial Access Policy for expectations of the policy.	Clinical Trial Databases
	ENCODE Consortia	Requires resource producers to release primary data along with an an initial interpretation, in the form of genome features, to the appropriate public	

Finding and selecting a repository

- NIH Supported subject-specific, open-access repositories
- Primary consideration should be given to data repositories that are discipline or data-type-specific

**** First Choice Whenever Possible ****

70+ NIH Subject Repositories →

Examples

- Metabolomics Workbench (MetWB)
- Stimulating Peripheral Activity to Relieve Conditions Portal (SPARC)
- BioSystics Analytics Platform (BioSystics-AP)
- National COVID Cohort Collaborative (N3C)
- Natural Products Magnetic Resonance Database (NP-MRD)
- ETC…..

Finding and selecting a repository

Non Subject-Specific Repositories

Generalist Data Repositories →

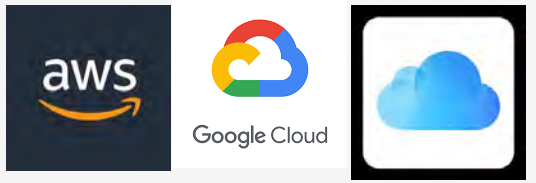
Generalist Repository Ecosystem Initiative - GREI

- **Dataverse**
- **Dryad**
- **Figshare**
- **Mendeley Data**
- **Vivli**
- **Open Science Framework**
- Synapse
- Zenodo
- IEEE Dataport

PubMed → Datasets (≤2 GB) supplementary material to accompany articles submitted to PubMed Central

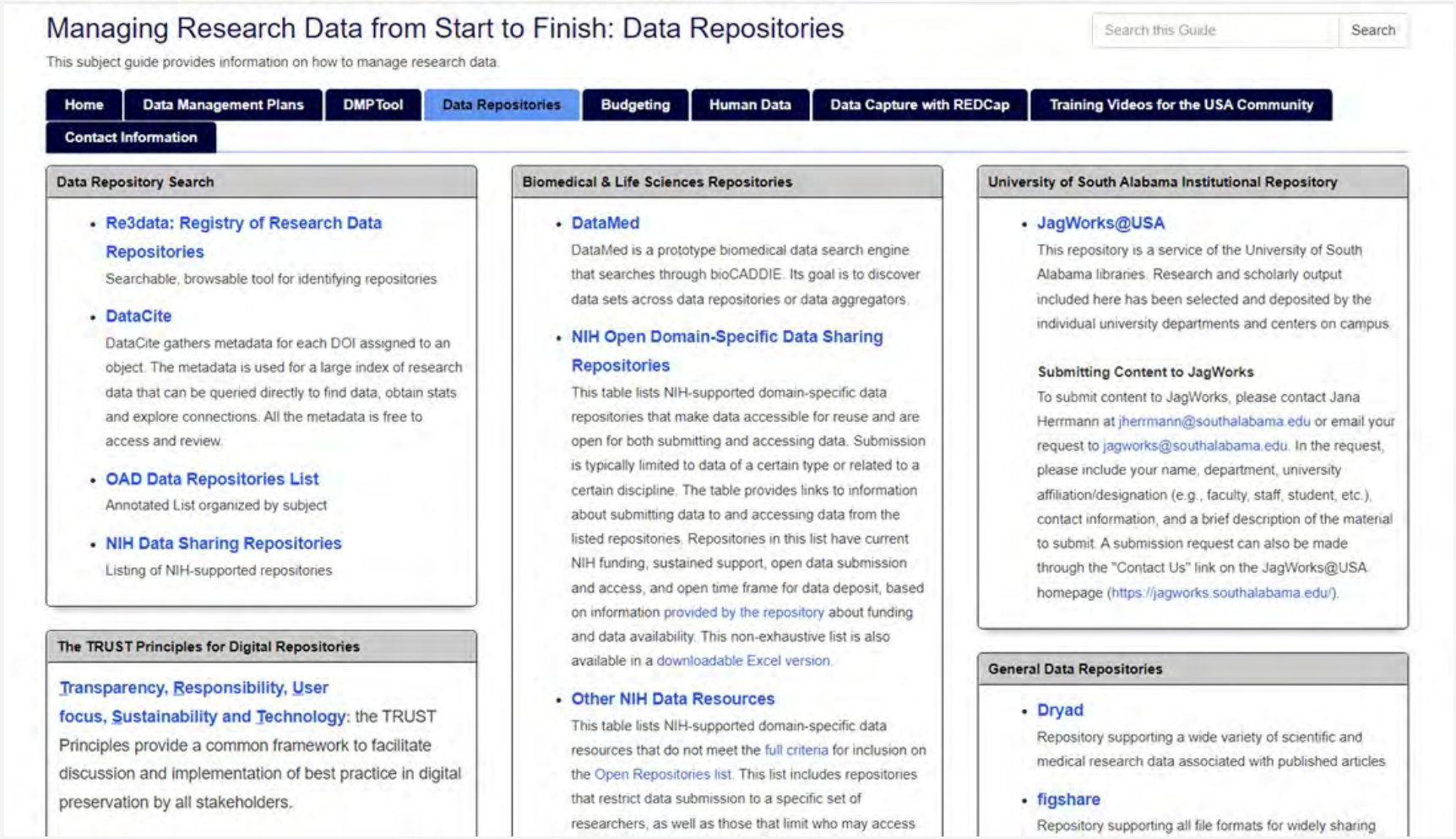
Institutional Repository → JagWorks@USA

Cloud-Based Servers → Large datasets (petabyte- scale)



Tools and resources to help

Research Data Management Subject Guide - Data Repositories



Managing Research Data from Start to Finish: Data Repositories

Search this Guide | Search

This subject guide provides information on how to manage research data.

Home | Data Management Plans | DMPTool | Data Repositories | Budgeting | Human Data | Data Capture with REDCap | Training Videos for the USA Community | Contact Information

Data Repository Search

- **Re3data: Registry of Research Data Repositories**
  Searchable, browsable tool for identifying repositories
- **DataCite**
  DataCite gathers metadata for each DOI assigned to an object. The metadata is used for a large index of research data that can be queried directly to find data, obtain stats and explore connections. All the metadata is free to access and review.
- **OAD Data Repositories List**
  Annotated List organized by subject
- **NIH Data Sharing Repositories**
  Listing of NIH-supported repositories

The TRUST Principles for Digital Repositories

Transparency, Responsibility, User focus, Sustainability and Technology: the TRUST Principles provide a common framework to facilitate discussion and implementation of best practice in digital preservation by all stakeholders.

Biomedical & Life Sciences Repositories

- **DataMed**
  DataMed is a prototype biomedical data search engine that searches through bioCADDIE. Its goal is to discover data sets across data repositories or data aggregators.
- **NIH Open Domain-Specific Data Sharing Repositories**
  This table lists NIH-supported domain-specific data repositories that make data accessible for reuse and are open for both submitting and accessing data. Submission is typically limited to data of a certain type or related to a certain discipline. The table provides links to information about submitting data to and accessing data from the listed repositories. Repositories in this list have current NIH funding, sustained support, open data submission and access, and open time frame for data deposit, based on information provided by the repository about funding and data availability. This non-exhaustive list is also available in a downloadable Excel version.
- **Other NIH Data Resources**
  This table lists NIH-supported domain-specific data resources that do not meet the full criteria for inclusion on the Open Repositories list. This list includes repositories that restrict data submission to a specific set of researchers, as well as those that limit who may access

University of South Alabama Institutional Repository

- **JagWorks@USA**
  This repository is a service of the University of South Alabama libraries. Research and scholarly output included here has been selected and deposited by the individual university departments and centers on campus.

  Submitting Content to JagWorks
  To submit content to JagWorks, please contact Jana Herrmann at jherrmann@southalabama.edu or email your request to jagworks@southalabama.edu. In the request, please include your name, department, university affiliation/designation (e.g., faculty, staff, student, etc.), contact information, and a brief description of the material to submit. A submission request can also be made through the "Contact Us" link on the JagWorks@USA homepage (https://jagworks.southalabama.edu/).

General Data Repositories

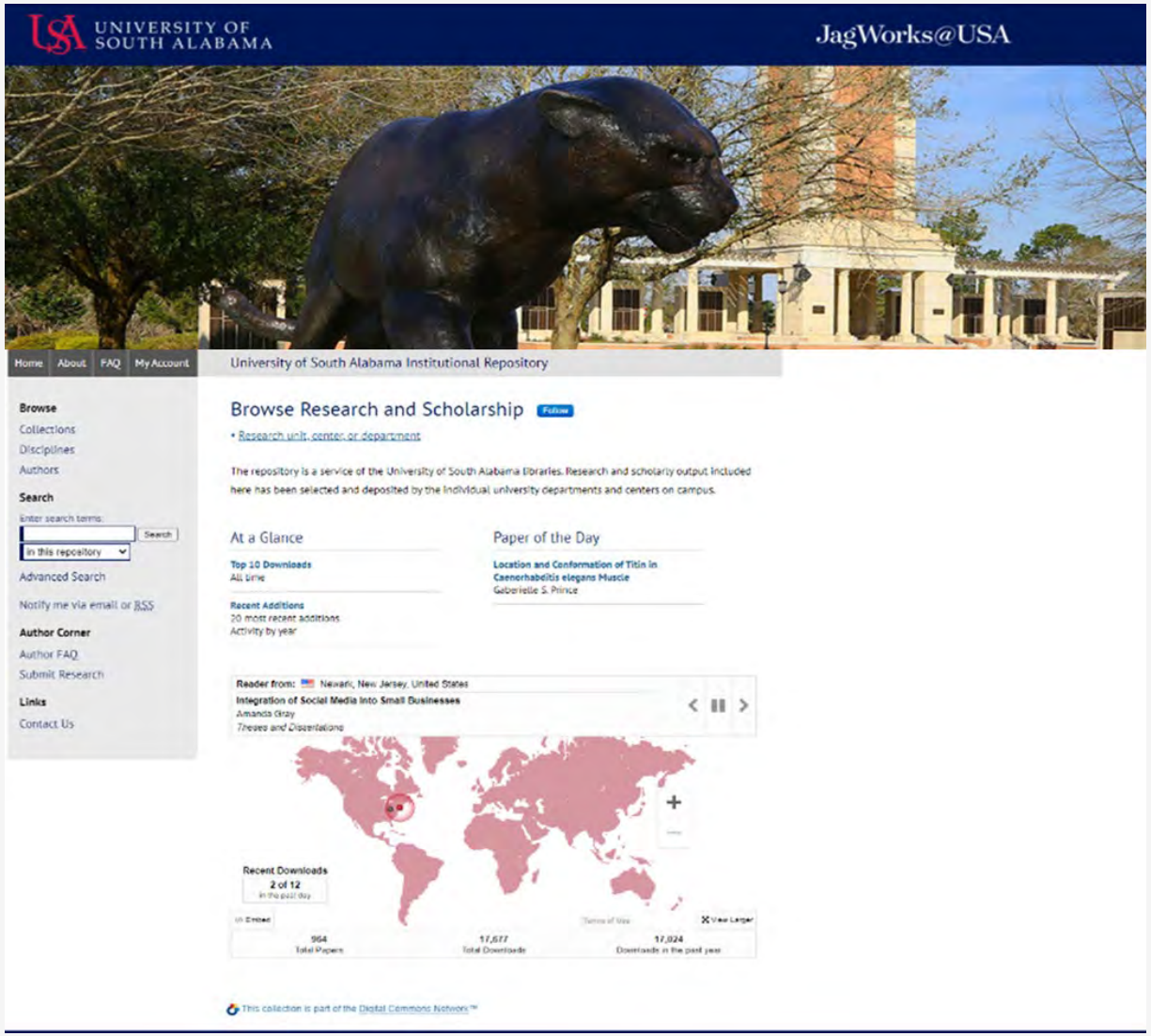
- **Dryad**
  Repository supporting a wide variety of scientific and medical research data associated with published articles
- **figshare**
  Repository supporting all file formats for widely sharing

USA Institutional Repository

JagWorks@USA Repository – open access, digital archive provided by the USA Libraries (https://jagworks.southalabama.edu)

- Anyone affiliated with USA or USA Health Services can submit content.
  - Faculty
  - Researchers
  - Staff
  - Students (with approval)

- JagWorks meets NIH desirable characteristics of data repositories



What can be included at JagWorks@USA?

Wide range of content and materials

Examples: (not an exhaustive list)

Theses/Dissertations

Conference presentations/posters

Journal articles

Journals published at USA

Datasets

Images

Accreditation documentation

Open educational resources

Podcasts

Textbooks

Training materials/SOPs

Accommodates most common file types

Examples: (not an exhaustive list)

- .doc/.docx
- .mp3/.mp4
- .xls/.xlsx
- .jpg/.jp2/.jpx
- .pdf
- .tiff
- .gif
- .png
- .bmp
- .eps
- .rtf
- .zip

Limitations on data sharing

Justifiable ethical, legal, and technical factors:

- Informed consent will not permit or limits scope of sharing or use
- Privacy or safety of research participants would be compromised and available protections insufficient
- Explicit federal, state, local, or Tribal law, regulation, or policy prohibits disclosure
- Restrictions imposed by existing or anticipated agreements with other parties

Reasons NOT acceptable to limit sharing:

- Data are considered too small
- Data will not be widely used
- Data are not thought to have a suitable repository

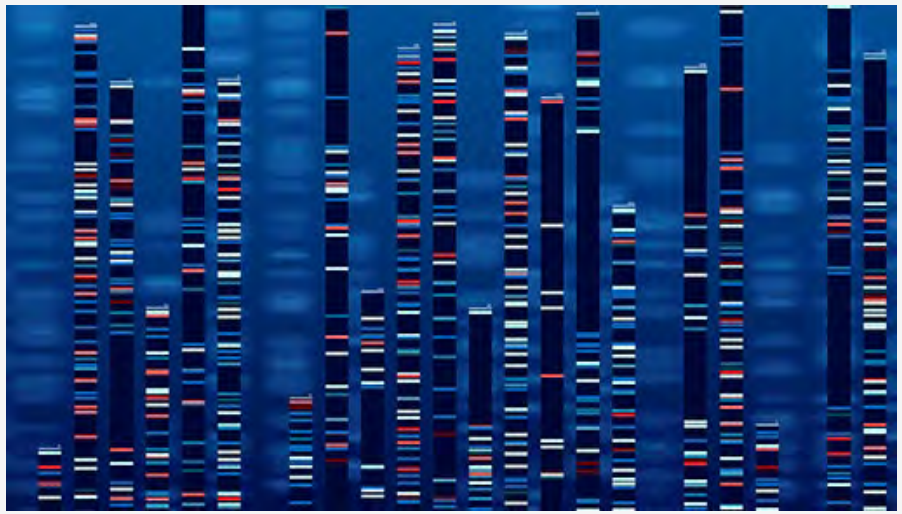
Others considerations for data sharing

❖ Proprietary Data

- **Small Business Innovation Research** (SBIR) and **Small Business Technology Transfer** (STTR) awardees may withhold applicable data for 20 years after the award date per SBIR/STTR funding agreement.
- Must submit DMS Plan

❖ Genomic Data

- Separate DMS Plan for Human and Non-human Genomic Data
- Specific repositories for genomic data



https://www.genome.gov/about-nhgri/Policies-Guidance/Data-Sharing-Policies-and-Expectations

For more information and questions?

NIH Data Management and Sharing Policy website:

https://sharing.nih.gov/data-management-and-sharing-policy

NIH DMS FAQs:

https://sharing.nih.gov/faqs#/data-management-and-sharing-policy.htm

NIH Genomic Data Sharing Policy website:

https://sharing.nih.gov/genomic-data-sharing-policy

Managing Research Data from Start to Finish subject guide:

https://libguides.southalabama.edu/research_data_mgt

If you have questions or would like a consult regarding your DSMP planning process, please contact the appropriate person:

Data Management & Sharing Plan Development:

Clista Clanton (cclanton@southalabama.edu)

DMPTool:

Dusty Layton (dlayton@southalabama.edu)

JagWorks Institutional Repository:

Jana Herrmann (jherrmann@southalabama.edu)

NIH Data Management & Sharing Policies:

Gina Hedberg (ghedberg@southalabama.edu)