

# Finding a Repository for NIDDK Study Datasets

Jeran Stratford

31 May 2023



# Planning for Data Management and Sharing



- **Who**
  - Investigators applying for NIH Funding
- **What**
  - Scientific Data
- **Where**
  - Quality data repository(ies)
- **When**
  - As soon as possible, or no later than the time of a publication or at the end of the award, whichever comes first

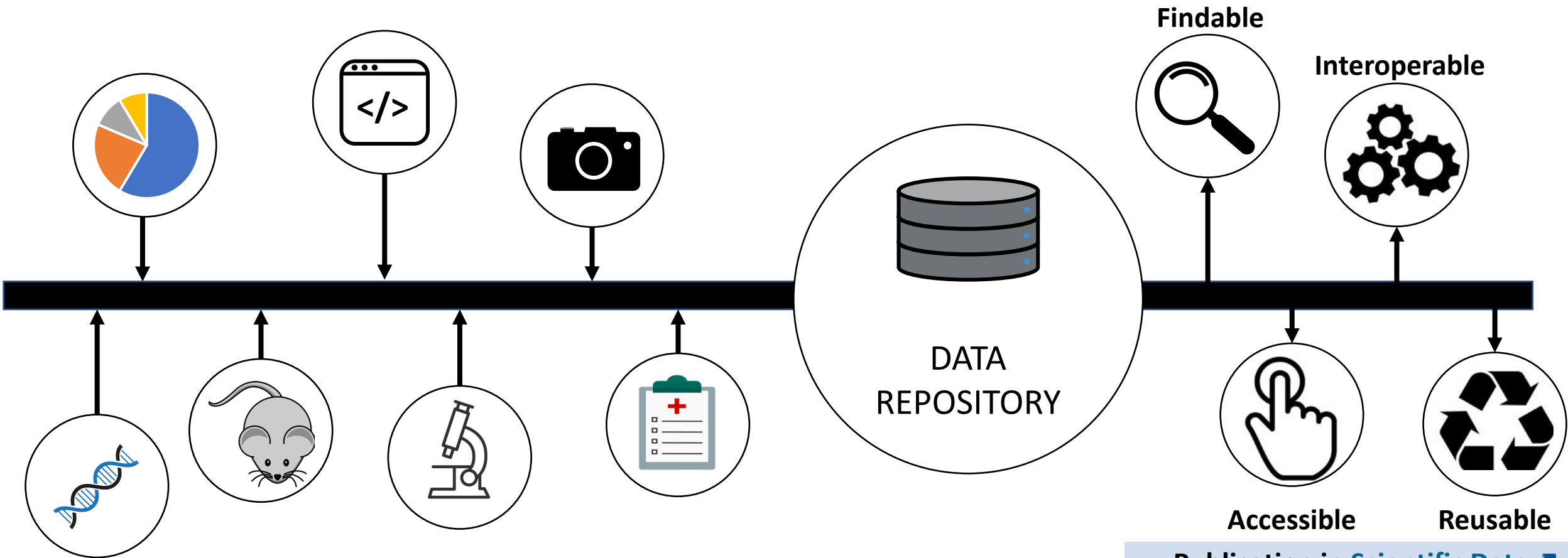
## Data preservation, Access, And Associated Timelines

- Repository to be used, persistent unique identifier, and when/ how long data will be available

See [NIDDK DMS Webinar 1: Writing a DMS Plan](#) to review

# The Data Repository

A data repository is a database infrastructure that collects, manages, and stores data sets for data analysis, sharing, and reporting.



Publication in [Scientific Data](#)

# Selecting a Data Repository

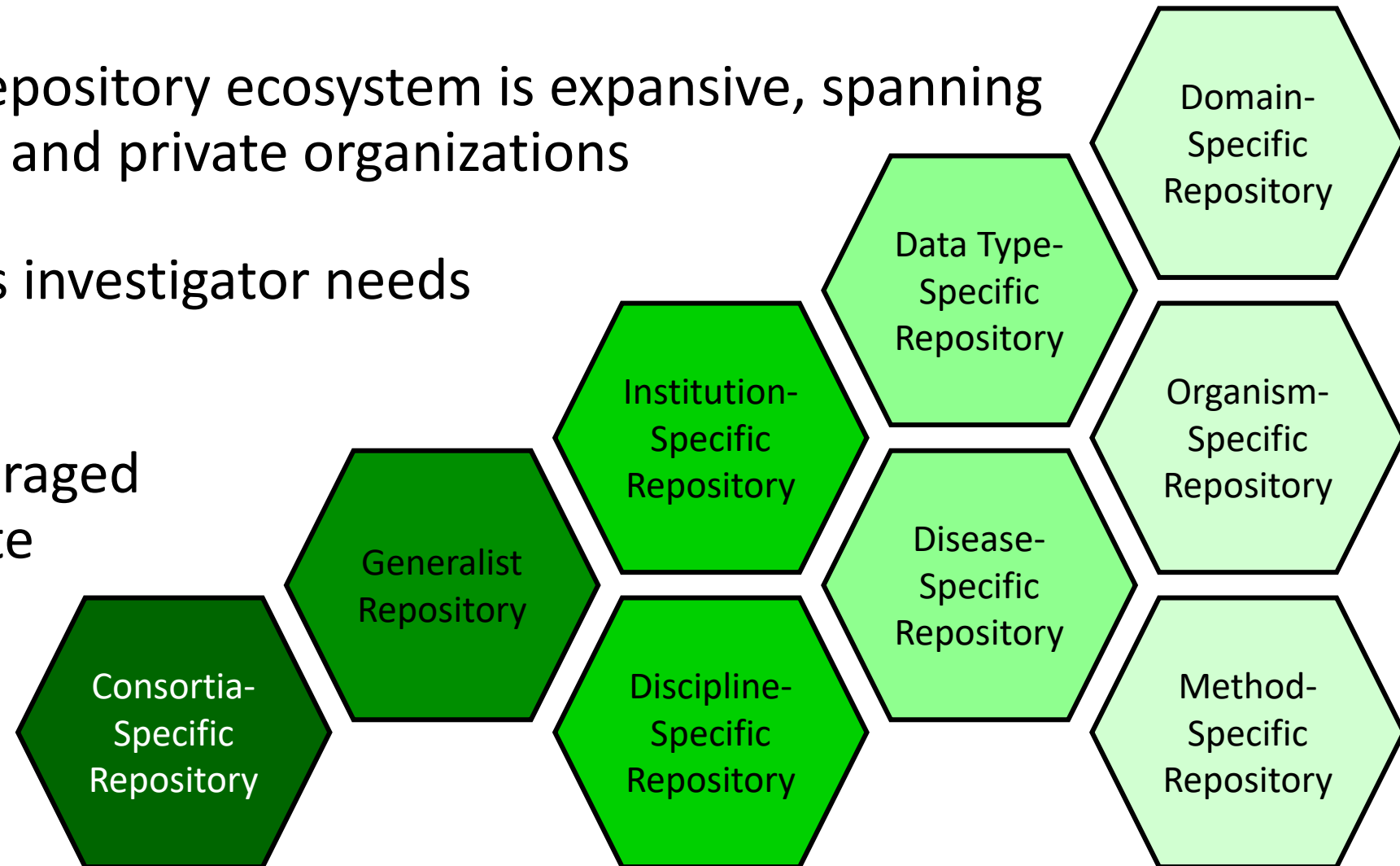
## Desirable Characteristics for All Data Repositories

- ✓ 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

See [Supplementary Guide Notice \(NOT-OD-21-016\)](#) for details

# Data Repository Ecosystem

- The biomedical data repository ecosystem is expansive, spanning NIH-hosted and public and private organizations
- Continually evolving as investigator needs evolve
- Researchers are encouraged to select an appropriate data repository



# NIH & Other Resources for Finding a Repository

- NIH and/or Institutes, Centers, and Offices and Funding Opportunity Announcements may identify a particular repository
- Primary consideration should be given to repositories that are discipline or data-type specific to support effective data discovery and reuse
- Secondary consideration to generalist or institutional repositories

See [Selecting a Data Repository](#) for details

## NIH-Supported Repositories

- Easy-to-filter list of 70+ [NIH Repositories](#)

Institute or Center	Repository Name	Repository Description
All		Diabetes
NIDDK	<a href="#">Accelerating Medicines Partnership in Common Metabolic Diseases (AMP-CMD) Knowledge Portal</a>	The Common Metabolic Disease analyzes genetic association computational prediction met open-access portal. The aim molecular basis of complex d cardiovascular and cerebroma

## Other Repository Resources

- [Generalist repositories](#)
- [Springer Nature's Data Repository Guidance](#)
- [Registry of Research Data Repositories](#)

See [Repositories for Sharing Scientific Data](#)

# NIDDK-Specific Guidance on Repository Selection

Home \ Research & Funding \ Research Resources \ NIDDK Data Management & Sharing \ NIDDK Guidance for Writing a DMS Plan

NIDDK Data  
Management &  
Sharing

NIDDK Guidance for Writing a  
DMS Plan

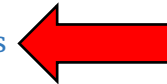
NIDDK DMS Tools &  
Examples

## NIDDK Guidance for Writing a DMS Plan

The National Institutes of Health (NIH) Data Management and Sharing (DMS) policy expects that researchers maximize the appropriate sharing of scientific data. The DMS Plan should describe how the scientific data generated will be managed and shared. NIH has outlined six elements that should be included in each submitted DMS Plan. The [Writing a Data Management & Sharing Plan](#) NIH section of [sharing.nih.gov](https://www.sharing.nih.gov) describes each of the recommended elements.

The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) provides Institute-specific DMS Guidance, which builds upon the NIH guidance, for the following DMS Plan Elements:

1. Data Type
2. Related Tools, Software, and/or Code
3. Standards
4. Data Preservation, Access, and Associated Timelines
5. Access, Distribution, or Reuse Considerations
6. Oversight of Data Management and Sharing



## Repositories and tools to find repositories resource

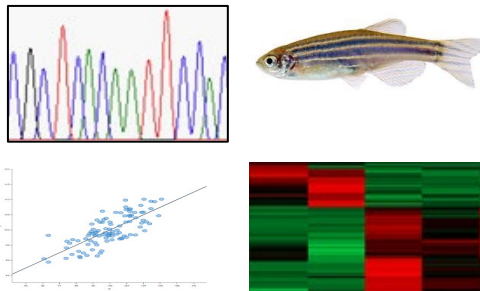
- [NIDDK Central Repository](#)
- [dkNET](#)  (lists repositories used by NIDDK-supported researchers)
- [NIH-supported repositories](#)
- [NIH GREI Repositories](#) 

See [NIDDK Data Preservation Guidance](#) for details



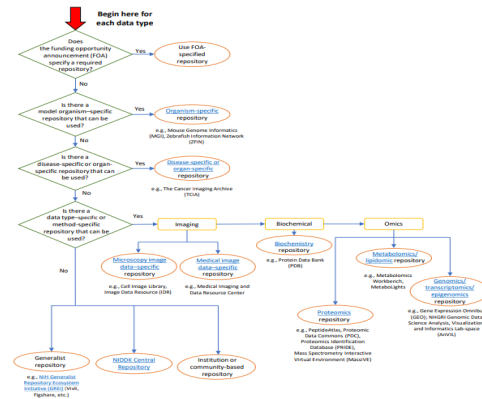
# NIDDK Repository Selection Considerations Tool

## Identify Scientific Data to be Shared



Investigators will have previously identified the scientific data types that will be shared

## Identify Appropriate Repository



Identify repository for each data type guided by the provided decision tree

## Evaluate Additional Considerations



Repository Requirements

- Eligibility
- Deidentification
- File Formats / Size
- Data Linkage
- Metadata

## Data Management and Sharing Plan



Include the selected repository and justification in the “Data Preservation, Access and Associated Timelines” element of the DMS Plan

[NIDDK Repository Selection Considerations Tool](#)



# Factors to Consider: Type of Access



## Sharing through **controlled access** if...

- There are explicit limitations on sharing
- Data pose particular risks to participants or groups
- Data cannot be sufficiently de-identified
- Unanticipated risks are discovered after initial planning



## Sharing through **unrestricted access** if....

- Explicitly consented for unrestricted sharing
- Data are de-identified **and** institutional review finds unrestricted sharing poses very low risk

# Unique Persistent Identifier (PID)

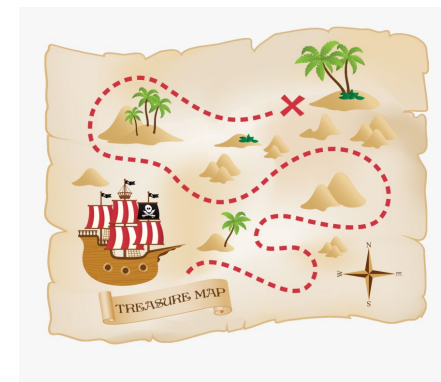
**Definition: A Unique Persistent Identifier** assigns datasets a citable, unique persistent identifier, such as a digital object identifier (DOI) or accession number, to support data discovery, reporting, and research assessment.



**Data**



**Metadata**



**PID**

**The use of PIDs is encouraged, and their use is a desirable characteristic of data repositories.**

# Summary

- Repositories are a key component to enable the FAIR (Findability, Accessibility, Interoperability, and Reusability) and maximal sharing of scientific data
- Choosing an appropriate repository is a multifaceted decision
- NIDDK has developed institute-specific guidance about finding and selecting a repository
- The NIDDK DMS webpage has tools, resources, and examples to aid investigators in choosing an appropriate repository(ies)
- If you have any questions, please reach out to your program officer

# Additional Resources

## NIDDK

- [Data Preservation, Access, and Associated Timelines | NIDDK Guidance for Writing a DMS Plan \(niddk.nih.gov\)](https://www.niddk.nih.gov/health-research/data-management/data-preservation/data-preservation-timelines)
- [Selecting a Data Repository | NIDDK DMS Tools & Examples \(niddk.nih.gov\)](https://www.niddk.nih.gov/health-research/data-management/data-repository)

## NIH

- [Selecting a Repository for Data Resulting from NIH-Supported Research | NOT-OD-21-016 \(grants.nih.gov\)](https://www.grants.nih.gov/grants-apply/data-repository)
- [Writing a Data Management & Sharing Plan | Data Sharing \(nih.gov\)](https://www.nih.gov/data-sharing)
- [Selecting a Data Repository | Data Sharing \(nih.gov\)](https://www.nih.gov/data-sharing/data-repository)
- [Repositories for Sharing Scientific Data | Data Sharing \(nih.gov\)](https://www.nih.gov/data-sharing/repositories)
- Sharing Data From Human Participants ([Data Sharing Approaches | Data Sharing \(nih.gov\)](https://www.nih.gov/data-sharing/data-sharing-approaches))
- [NIH Institute and Center Data Sharing Policies | Data Sharing](https://www.nih.gov/data-sharing/policies)
- [Home Page | Data Sharing \(nih.gov\)](https://www.nih.gov/data-sharing)

## OTHER

- [FAIR Principles | Scientific data](https://www.fair4science.org/en/scientific-data) 

# Questions

