

NICHD Data Sharing Presentation to NICHD Advisory Council

Rebecca F. Rosen, PhD

Director, NICHD Office of Data Science & Sharing (ODSS)

January 25, 2023



*Eunice Kennedy Shriver National Institute
of Child Health and Human Development*

Discussion Topics

- NIH Office of Data Management and Sharing
- NICHD Policy for Data Science and Sharing
- Data Sharing Resources for NICHD Staff and Investigators



NICHD Office of Data Science and Sharing

New office in Office of the Director, established under the **NICHD Strategic Plan 2020**,
Scientific Stewardship Goal 2: Facilitating Data Sharing and Access to Biospecimens



**Valerie
Cotton**

Deputy Director



**Jianji
Yang**

Informatics
Program Manager



**Elizabeth
Clerkin**

Clinical Informatics
Project Manager



**Anne
Deslattes Mays**

NIH DATA Scholar



**Stephanie
Tison**

Presidential
Management Fellow



**Tarsheka
Thompson**

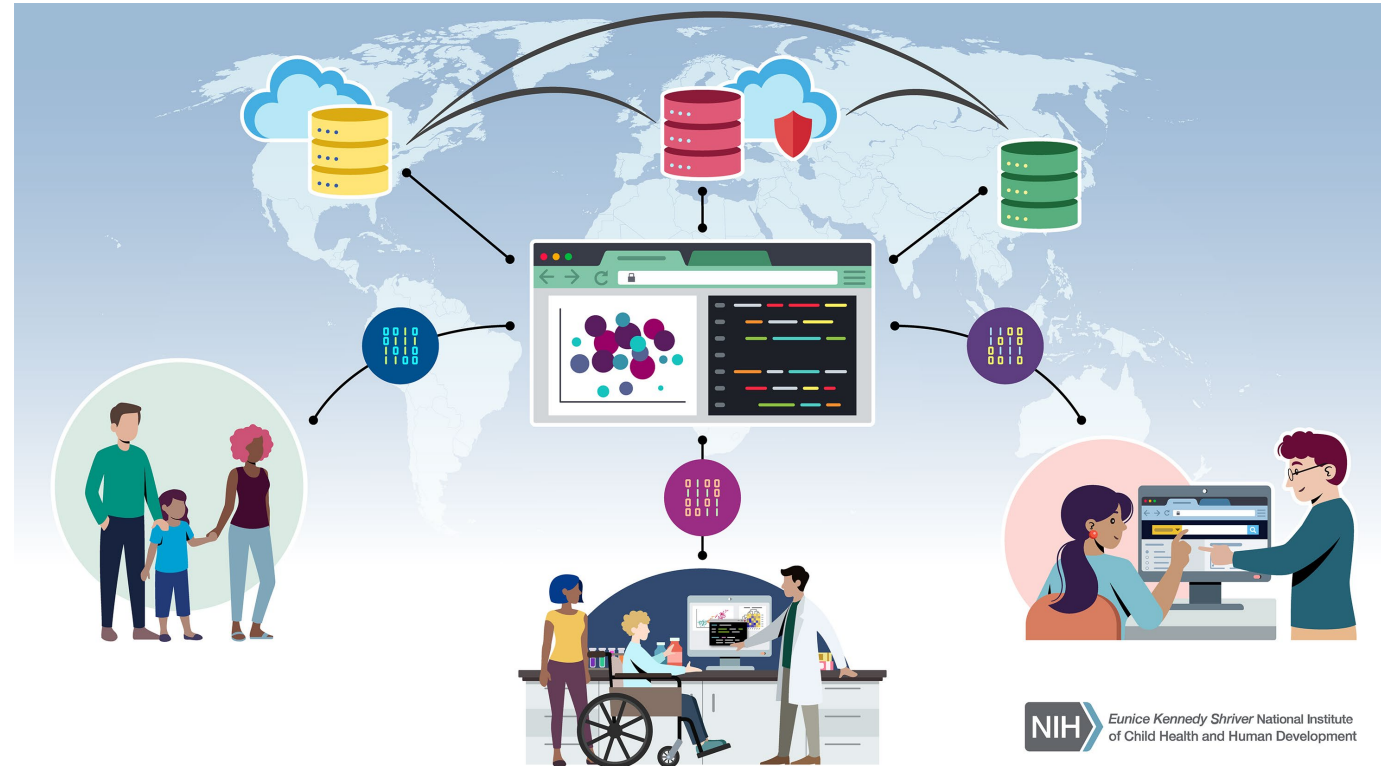
Support Integration
Specialist



NICHD ODSS Vision and Mission

Vision: A culture of responsible and innovative use of data and biospecimens that accelerates research and improves health for NICHD populations

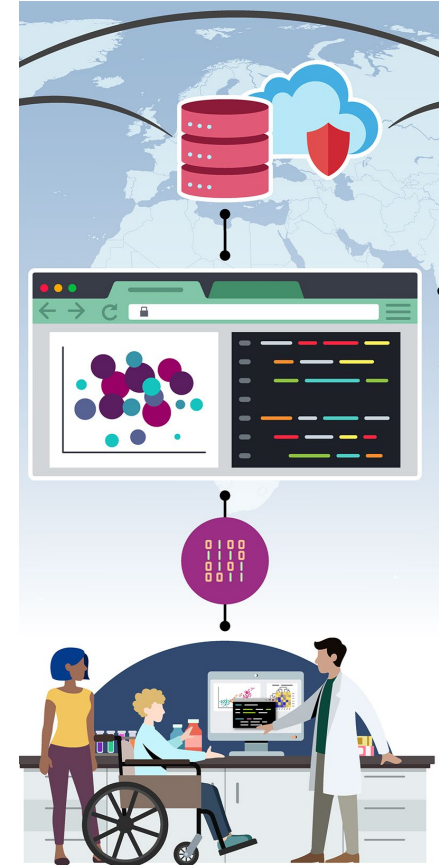
Mission: NICHD ODSS will develop a diverse, secure, and interoperable research data ecosystem and will advise on best practices for data collection, standards, management, sharing, and use across the research and funding lifecycles, in order to advance scientific discovery in support of NICHD's mission to understand human development, improve reproductive health, enhance the lives of children and adolescents, and optimize abilities for all.



NICHD Data and Specimen Ecosystem: Building out to Support Priority Use Cases

The Ecosystem is comprised of people, data, processes, and technologies to align with the NICHD Strategic Plan and support NICHD communities' data science and sharing needs:

- **Intramural and extramural investigators** share their data and biospecimens to encourage reproducibility and broad reuse
- The **researcher community** finds and uses NICHD data and specimens to enable new and innovative research
- The **researcher community** securely analyzes NICHD and other data using new tools, training resources, and powerful compute
- NICHD **participant communities** visualize outcomes of participation in research and benefit faster from new discoveries
- **NICHD staff** track researcher compliance with NIH policies and good stewardship of government funds (DASH)
- **NICHD leadership** assesses and visualizes the scientific return on investment in data and biospecimen sharing



“One NICHD” Approach to Facilitating Data Sharing and Access to Biospecimens

ODSS leads **NICHD-wide activities** across three strategic objectives:

Technical Infrastructure

Objective 2.1: Develop a secure, federated, and sustainable NICHD research data and specimen ecosystem

Data Ecosystem Working Group

Policy Implementation & Compliance

Objective 2.2: Create NICHD-wide guidance and implementation resources for NIH data and specimen sharing policies

Intramural Data Sharing Committee

Extramural Data Sharing Committee

Standards, Training, & Partnerships

Objective 2.3: Build informational resources and tools for NICHD data science and sharing activities



NICHD Data and Specimen Ecosystem Working Group – Established January 2022

Charge:

- Define the scope of the NICHD research data and specimen ecosystem
- Define the expectations and timeline for the build-out of the system, through membership subject matter expertise and engagement with NICHD staff and the research community

Division of Intramural Research

- Min Lee (DiPHR)
- Jennifer Weck (DiPHR)
- Ryan Dale (DIR)
- Asma Idriss (DIR)

Division of Extramural Research/NCMRR

- James Coulombe (DBSVB)
- Donna Mazloomdoost (GHDB)
- Zhaoxia Ren (OPTBB)
- Marion Koso-Thomas (PPB)
- Ashley Vargas (PGNB)
- Tammy Jenkins (Alt: ZSuzsanna Kocsis) (PTCIB)
- Mollie Minear (IDDB)
- Brett Miller (Alt: Virigina Salo) (CDBB)
- Toyin Ajisafe (NCMRR)
- Rebecca Clark (PDB)
- Jack Moye (MPIDB)
- Leigh Allen (CRB, FIB)



NICHD Extramural Data Sharing Committee

Established 2018

Led by ODSS/DEA, the EDSC collectively defines policy implementation strategies and informs the development of resources for NIH data sharing policies

Membership from 12 extramural scientific branches, NCMRR, DEA branches, OSPRA, DIR/DiPHR*

- Brett Miller (CDBB)
- Leigh Allen (CRB)
- Mahua Mukhopadhyay (DBSVB)
- Susan Taymans (FIB)
- Candace Tingen (GHDB)
- Jack Moyer (MPIDB)
- Katie Vance (OPPTB)
- Drew Bremer (PGNB)
- Zsuzsanna Kocsis (PTCIB)
- Rebecca Clark (PDB)
- Sujata Bardhan (IDDB)
- Guillermina Girardi (PPB)
- Maria Nurminskaya (NCMRR)
- Jennifer Weck (DiPHR)
- Ryan Dale (DIR)
- Rebekah Rasooly (DEA, co-Chair)
- Caroline Signore (DER)
- Maggie Young (GMB)
- Cathy Wedeen (SRB)
- Jennifer Guimond (OSPRA)



NICHD Intramural Data Sharing Committee

Established Jan 2022

Charge:

- Define the NICHD intramural implementation of NIH data sharing policies, and
- Identify and prioritize resources that will facilitate intramural researchers' ability to comply with NIH data sharing policies.

DiPHR

- **Diane Putnick**, *Epidemiology Branch*
- **Leah Lipsky**, *Social and Behavioral Sciences Branch*
- **Diana Blithe**, *Contraceptive Development Program*
- **Jennifer Weck**,* *Office of the DiPHR Director*

DIR

- **Dax Hoffman**, *Section on Neuroendocrinology*
- **Gigi Storz**, *Section on Environmental Gene Regulation*
- **Mary Dasso**, *Section on Cell Cycle Regulation*
- **Henry Levin**, *Section on Eukaryotic Transposable Elements*
- **Doreen Matthies**, *Unit on Structural Biology*
- **Vincent Schram**, *Microscopy and Imaging Core*
- **Ryan Dale**,* *Bioinformatics and Scientific Programming Core*
- **Alexander Sodt**, *Membrane Chemical Physics*
- **Karim Calis**, *Director of Clinical Research and Compliance*
- **Meg Keil**, *Associate Director of Nursing and Protocol Navigation*

*members of the NICHD Extramural Data Sharing Committee



Discussion Topics

- NICHD Office of Data Science and Sharing
- **NIH Policy for Data Management and Sharing**
- Data Sharing Resources for NICHD Staff and Investigators



NIH Policy for Data Management and Sharing

- Goes into effect on **January 25, 2023** for all NIH-supported research that generates scientific data
- **Requires** researchers to prospectively plan for how scientific data and metadata will be managed and shared through submission of a Data Management and Sharing Plan that considers any potential restrictions or limitations.
 - In the Plan, expects researcher to maximize data sharing, share data in established repositories at the time of publication or end of performance period, whichever comes first
- **Requires** compliance with approved Plans as a term and condition of award

Scientific Data: 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

Data Sharing: The act of making scientific data available for use by others (e.g., the larger research community, institutions, the broader public), for example, via an established repository.



Structure of a Data Management & Sharing Plan

OMB No. 0925-0001 and 0925-0002 (Rev. 07/2022 Approved Through TBD)

- In general, 2 pages in length, but no hard limit
- Includes DMS Policy and Genomic Data Sharing
- May use optional [format page](#) or other templates
- Should address [6 Elements](#):
 1. Data Type
 2. Related Tools, Software and/or Code
 3. Standards
 4. Data Preservation, Access, and Associated Timelines
 - **Data Repositories** for sharing
 - **Timelines** for sharing
 5. Access, Distribution, or Reuse Considerations
 - Human data: **privacy protections**
 6. Oversight of Data Management and Sharing

DATA MANAGEMENT AND SHARING PLAN

If any of the proposed research in the application involves the generation of scientific data, this application is subject to the NIH Policy for Data Management and Sharing and requires submission of a Data Management and Sharing Plan. If the proposed research in the application will generate large-scale genomic data, the Genomic Data Sharing Policy also applies and should be addressed in this Plan. Refer to the detailed instructions in the application guide for developing this plan as well as to additional guidance on [sharing.nih.gov](#). The Plan is recommended not to exceed two pages. Text in italics should be deleted. There is no "form page" for the Data Management and Sharing Plan. The DMS Plan may be provided in the *format* shown below.

Public reporting burden for this collection of information is estimated to average 2 hours per response, including the time for reviewing instructions, searching existing data sources, gathering, and maintaining the data needed, and completing and reviewing the collection of information. An agency may not conduct or sponsor, and a person is not required to respond to, a collection of information unless it displays a currently valid OMB control number. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to: NIH, Project Clearance Branch, 6705 Rockledge Drive, MSC 7974, Bethesda, MD 20892-7974, ATTN: PRA (0925-0001 and 0925-0002). Do not return the completed form to this address.

Element 1: Data Type

A. Types and amount of scientific data expected to be generated in the project:

Summarize the types and estimated amount of scientific data expected to be generated in the project.

B. Scientific data that will be preserved and shared, and the rationale for doing so:

Describe which scientific data from the project will be preserved and shared and provide the rationale for this decision.

C. Metadata, other relevant data, and associated documentation:

Briefly list the metadata, other relevant data, and any associated documentation (e.g., study protocols and data collection instruments) that will be made accessible to facilitate interpretation of the scientific data.

Element 2: Related Tools, Software and/or Code:

State whether specialized tools, software, and/or code are needed to access or manipulate shared scientific data, and if so, provide the name(s) of the needed tool(s) and software and specify how they can be accessed.

Element 3: Standards:

State what common data standards will be applied to the scientific data and associated metadata to enable interoperability of datasets and [resources](#), and provide the name(s) of the data standards that will be applied and describe how these data standards will be applied to the scientific data generated by the research proposed in this project. If applicable, indicate that no consensus standards exist.

Element 4: Data Preservation, Access, and Associated Timelines



Budgeting for Data Management and Sharing

- Budget Form: The requested direct costs to support the activities proposed in the DMS Plan must be indicated as “Data Management and Sharing Costs”
- Budget Justification: A **brief summary of the DMS Plan** (type and amount of shared scientific data and name of repository) and a description of the requested DMS Costs for each category (curating/preparing, repositories, local management).

ALLOWABLE COSTS

- Curating data/ developing supporting documentation
- Preserving/ sharing data through repositories
- Unique/special local data management
- **IMPORTANT**: Must be incurred during the performance period

UNALLOWABLE COSTS

- Infrastructure costs typically included in indirect/F&A costs
- Costs associated with the routine conduct of research (e.g., costs of gaining access to research data)



Lifecycle of a Data Management & Sharing Plan for Extramural Grants

Plan Submission

- With application
- Full Plan as separate attachment submitted in “Other Plan(s)”
- Brief Plan description in Budget Justification

Plan Assessment

- Peer reviewers comment on (*not score*) budget (will not see attached DMS Plan)
- NIH Program Officers assess Plans
- Plans can be revised at JIT
- Plans must be approved prior to award

Plan Compliance

- Approved Plan becomes a Term and Condition of Award
- Grantee reports progress of approved Plan at annual RPPR
- Failure to comply may result in an enforcement action and affect future funding decisions

Peer reviewers only review/score DMS Plans, when data sharing is integral to the FOA



Discussion Topics

- NICHD Office of Data Science and Sharing
- NIH Policy for Data Management and Sharing
- **Data Sharing Resources for NICHD Staff and Investigators**
 - **Staff and Investigator resources for implementing the DMS Policy**
 - **Data ecosystem development for sharing data and tools**
 - **Training resources for equitable access to data and tools**



NICHD-Developed Staff and Researcher Resources for DMS Policy Implementation

Primary Goals: Prepare NICHD extramural staff to field questions from researchers and review DMS Plans, and support extramural and intramural DMS Plan development

- DMS Policy Resources on NICHD ODSS website
- Tips for Writing a DMS Plan (public resources)
- Example DMS Plans (public documents)
- NICHD Data Repository Finder (public web tool)
- NICHD Extramural DMS Policy FAQs (internal)
- NICHD Intramural DMS Policy FAQs (internal)
- NICHD DMS Plan Assessment Guidance (internal)

Resources are available on an internal Teams and/or Public website

Office of Data Science and Sharing (ODSS)



ODSS was established in 2021 to lead and coordinate NICHD's activities within data science, bioinformatics, data sharing policy and compliance, and emerging technologies.

ODSS's vision is to enable a culture of responsible and innovative use of data and biospecimens that accelerates research and improves health for NICHD populations. The office's mission is to:

- Develop a diverse, secure, and interoperable research data ecosystem
- Advise on best practices for data collection, standards, management, sharing, and use across the research and funding lifecycles
- Advance scientific discovery in support of NICHD's mission to understand human development, improve reproductive health, enhance the lives of children and adolescents, and optimize abilities for all

ODSS is a trusted informational resource for NICHD staff and researchers on all NIH data and specimen sharing policies.

ODSS serves as NICHD's primary liaison with the NIH Office of the Director's Office of Data Science and Strategy, to ensure engagement in large NIH data-science and emerging technology programs and ensure alignment with NIH, HHS, and federal programs and policies.

ODSS Staff



NICHD Data and Specimen Hub (DASH)



Data Management and Sharing Policy Resources



NICHD Data Ecosystem



NICHD Data Ecosystem Strategy

The NICHD Data Ecosystem is comprised of **people, data, processes, and technologies** to align with the NICHD Strategic Plan and support NICHD communities' data science and sharing needs

OPPORTUNITIES

- The DMS Policy will increase the scale and scope of data available for responsible and innovative use
- NIH Office of Data Science Strategy coordinates the development of trans-NIH ecosystem resources, including repositories, tools, standards, and trainings
- NICHD has **existing investments in modern data sharing infrastructure** and serves in leadership roles on trans-NIH ecosystem interoperability activities

CHALLENGES & NEEDS

- NICHD-relevant data are and will be shared in many diverse NIH data systems
- NICHD needs increased **visibility into the current state of these data systems** and how diverse **NICHD communities use or aspire to use** the systems and the data
- Democratizing access to the NICHD Data Ecosystem requires active **engagement with and training** of diverse researchers, developers, and other community members



DASH Data and Specimen Hub

- Centralized resource for researchers to share de-identified data from studies funded by NICHD
- Serves as a portal for requesting biospecimens from selected studies in DASH
- Aims to accelerate scientific findings to ultimately improve human health



<https://dash.nichd.nih.gov>

DASH Content

Top Study Topics

Breastfeeding & Breast Milk*	Labor & Delivery
Cerebral Palsy	Newborn Screening
Child Health*	Obesity & Overweight
Early Learning	Pelvic Floor Disorders
Early Labor and Birth	Pharmacology
Fertility Problems	Pregnancy*
High-Risk Pregnancy	Preterm Labor & Birth*
HIV/AIDS*	Stillbirth
Infant Care & Health*	Women's Health*
Infant Mortality	
Infertility & Fertility	

*biospecimens available

Currently Available Biospecimens

Amniotic fluid	Hair
Blood	Lymphocytes
Breast Milk	Meconium
Buffy Coat	Nail
Cord Blood (Buffy Coat, RBC, Plasma, Serum)	Saliva
DNA/RNA/Proteins	Serum/Plasma
Environmental Samples	Tissue samples
Erythrocytes (RBC)	Urine
	Vaginal Fluid

NICHD Data and Specimen Hub (DASH) An NICHD-Wide Clinical Data Repository



- DASH recently released several new DMS Policy-specific features:
 - DASH is integrated with the NIH Researcher Auth Service
 - DASH is assigning digital object identifiers to DASH studies
 - DASH updated the “External Resources” webpage as a prototype for the new DASH Catalog to support findability of NICHD-funded data assets
- All NICHD intramural clinical researchers will share data in DASH
- In 2023-2024, DASH will develop additional DMS Policy-specific features, such as:
 - Faster data submission, curation, and QA processes to support 4-month submission to release
 - Dataset versioning and multiple publication-related data released for a clinical study
 - Streamlined dataset submission from NICHD’s intramural clinical data management system



NICHD Data Ecosystem Analysis

Led by ODSS with guidance from NICHD Ecosystem Working Group, started August 2022

1. NICHD DASH will be a central component within the Data Ecosystem
2. Assessment of NICHD-relevant data systems
3. Development of researcher, staff, participant, and community use cases
4. Assessment and uses cases will inform the identification of:
 - Data repositories that do or can meet an NICHD threshold for long-term data sharing [**Sustainability**]
 - NICHD approach to **building connections** between DASH and other NICHD, NIH, and external systems [**Interoperability**]



Many others....

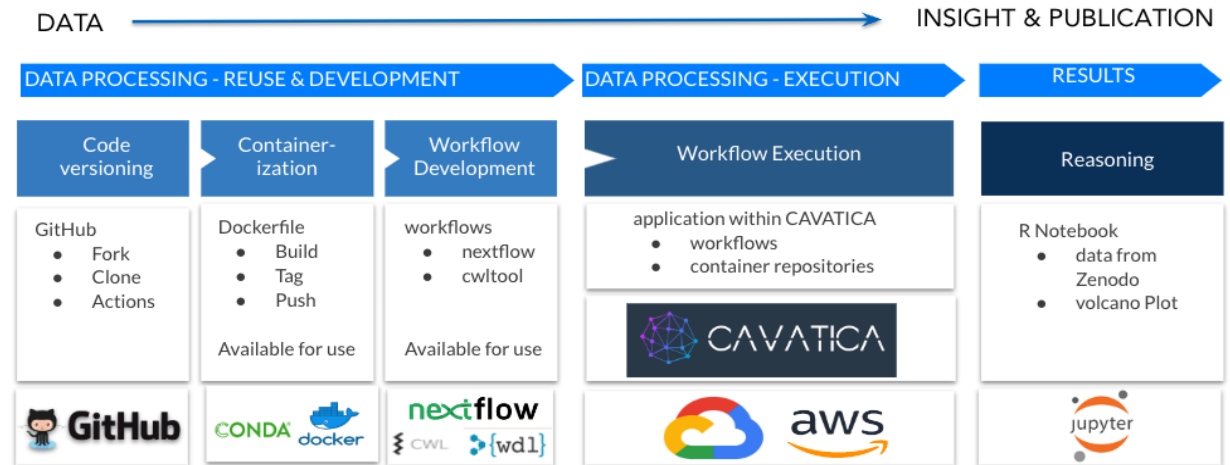


Researcher Training to Facilitate Use of the Ecosystem

In-Person/Online Course: Elements of Style in Workflow Creation and Maintenance

by NICHD DATA Scholar, Anne Deslattes Mays

- Goal: Building solutions from data to insight & publication by containerizing at the process level, stitching together with workflow languages, and analyzing with JupyterLab Notebooks
- Audience: Broad (no previous command line skills) to build literacy, capacity, skill, and understanding to perform analysis on any platform.
- Taught 4 times in 2022 to INCLUDE & Kids First researchers
- 2023 Plans: Continued engagements at AACR, BioIT World, and other scientific meetings



5 Day course available for self-study in an NICHD Online Repository: <https://github.com/NIH-NICHD>

- Day 1: Reasoning
- Day 2: Code Versioning
- Day 3: Containerization with Environment Control
- Day 4: Workflow Development
- Day 5: Workflow Execution



THANK YOU!

