

# NICHD Data Sharing Strategy: Update to NICHD Advisory Council

Office of Data Science & Sharing

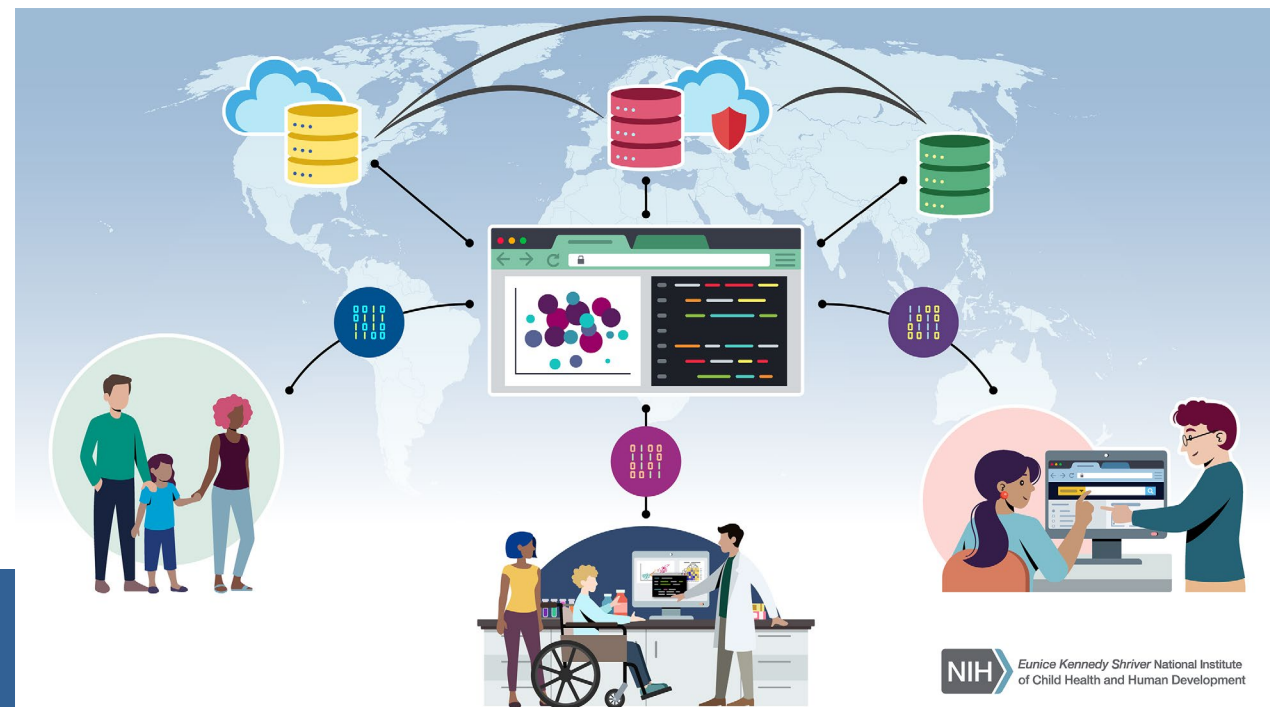
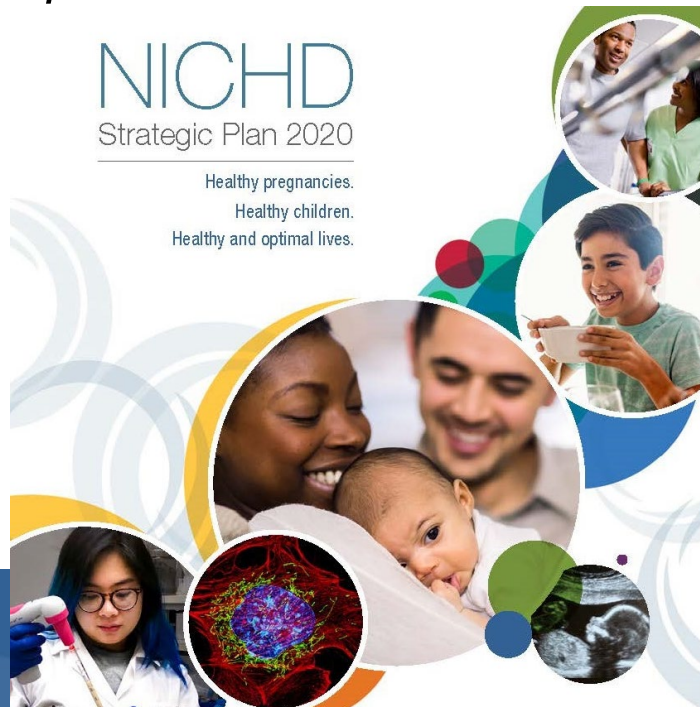
*June 3, 2024*



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

# ***NICHD is committed to promoting data sharing to accelerate scientific progress***

- 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.
- [The 2020 NICHD Strategic Plan](#) established the [NICHD Office of Data Science & Sharing](#)
  - Vision: A culture of responsible and innovative use of data and biospecimens that accelerates research and improves health for NICHD populations: *children, pregnant and lactating people, people with disabilities*



# ***NICHD is committed to promoting data sharing to accelerate scientific progress...***

**...following a community-informed approach to enhance NICHD data sharing across the research and data lifecycles**

- ODSS regularly engages with Offices of Communications (OC), Health Equity (OHE), Global Health (OGH), Legislative, Public Policy, and Ethics (OLPPE), Science Policy, Reporting, and Analysis (OSPRA) on data sharing strategy
- ODSS data sharing strategy work is informed by three governing bodies comprised of Extramural and Intramural colleagues:

## **Extramural Data Sharing Committee**

Define and support the implementation of NIH data sharing policies for extramural staff and researchers

*Membership: Representatives from all extramural scientific branches and centers, grants management branch, and intramural data sharing experts*

## **Intramural Data Sharing Committee**

Define and support the implementation of NIH data sharing policies for intramural staff and researchers

*Membership: Principal Investigators from intramural clinical, population health, and basic research labs and offices*

## **Ecosystem Working Group**

Define (1) the scope of the NICHD research data and specimen ecosystem and (2) the expectations and timeline for the build-out of the ecosystem

*Membership: Representatives from all extramural scientific branches and centers, and intramural informatics experts*



# ***Data sharing is foundational to responsible and innovative development and use of Artificial Intelligence for NICHD research***



White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (October 2023)

All of government strategy aiming to harness AI for good and minimize its risks, through the engagement of the private sector, academia, and civil society.



OMB Memo on Advancing Governance, Innovation, and Risk Management for Agency use of Artificial Intelligence (March 2024)

Establishes new agency requirements and guidance for AI governance, innovation, and risk management



HHS AI Task Force developing an HHS AI Strategic Plan (due January 2025)

Led by the HHS Office of the Chief AI Officer, which released in 2021 the HHS Trustworthy AI Playbook, agency guidance for safe and confident use of AI solutions



NIH Strategic Plan for Data Science: Request for Comments on Draft Plan (closed March 2024)

Led by NIH Office of Data Science Strategy, will articulate an NIH AI Strategy that aligns with HHS AI Strategic Plan



# NICHD Data Management and Sharing Tips for Researchers & Staff

- Remember good data management & sharing is critical to **good research**
- Support the ultimate **goal** to accelerate scientific progress
- Use the DMS plan to clearly describe:
  - ***which data*** (*generate & share*)
  - ***when*** (*by publication or end of award*)
  - ***where*** (***established repositories***)
  - ***and how*** (*open/controlled access*)
- Refer to DMS Policy materials for privacy considerations, definitions etc.
- **Maximize the appropriate sharing of scientific data**
  - If sharing is limited or delayed, **explain why** (“justifiable limitations” – laws, IRB requirements etc.)



***Helps us*** (*NIH, repository, researchers etc.*) ***work together to navigate perceived and/or real barriers***

**Data Sharing (NIH definition):** 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.



# NICHD Data Management & Sharing Plan Review

## *Common Issues*

### **Overall:**

- Elements (of the NIH-provided format page) are not used the same way
- Conflicting information (contradictions across elements)

---

### **Which data:**

- Not clear which data will be **generated** versus **shared**
- Lacking important details: **species/source, formats shared, amount, metadata**

---

### **When:**

- Missing adherence to [GDS Policy timelines](#) for human genomic data
- Duration follows **local** retention cycles (but should follow **repository** retention timelines)

---

### **Where:**

- Not using an established **repository** or not committing to one
- Planning to share only through “publication” and “conferences”
- Naming a data repository that is not appropriate (not broadly accessible, data type)
- When listing multiple repositories, not indicating which data goes to which
- Over-reliance on “generalist repositories” (should prioritize discipline-specific repositories)

---

### **How:**

- Plan for sharing is too restrictive, does not align to NIH's definition of "data sharing"
- Sharing “**by request**” or with **PI-control** (even if using a repository)

---

### **Justifiable**

### **Limitations:**

- Limitations are not adequately justified
- Vague reasons for not sharing (“ethical issues”, “privacy”, “sufficient quality,” “law”)
- No justification for not using a data repository, delaying sharing timeline

*NICHD presented these findings to the  
Federal Demonstration Partnership in May 2024*





# NICHD Data Management and Sharing *Online Resources*

## NICHD adheres to the NIH DMS Policy

NICHD has created several public resources **specific to the needs** of the NICHD research community

## [DMS Policy Resources on NICHD ODSS website](#)

- [Tips for Writing a DMS Plan](#) - describes what is expected for each DMS Plan Element & DMS Budget
  - *Updated with [Tips for Secondary Analysis Projects](#)*
- [Data Repository Finder](#) - includes data repositories that *in general* will accept data from NICHD researchers and shares information relevant to DMS Plan Elements
- [Example Plans](#) - based on four different NICHD project types
- [Data Standards](#) - new page that explains data standards and examples that may be relevant to NICHD research
  - *Metadata, vocabularies/terminologies/ontologies, common data elements, common data models, and more*

## 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 and standards for data collection, 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.

[Director Rebecca Rosen's presentation](#) (PDF 2.2 MB) from the January 2023 Council provides an overview of ODSS activities.

ODSS Staff

[Data Management and Sharing \(DMS\) Policy Resources](#)

[NICHD Data and Specimen Hub \(DASH\)](#)

[NICHD Data Ecosystem](#)

# Data Standards and NICHD DASH

## New Data Standards Webpage

### What data standards are relevant to NICHD?

The following existing data standards may be relevant to researchers who receive or are applying for NICHD support. Several categories of data standards exist, and sometimes a given standard falls into several categories. Although this information will be updated regularly, it is not intended to be comprehensive.

Expand All Collapse All

- Metadata Standards 
- Controlled Vocabularies, Terminologies, & Ontologies 
- Common Data Elements (CDEs) 
- Common Data Models (CDMs) 
- Other Standards and Resources 

## Updated NICHD DASH Codebook

- Researchers submitting data annotate variables with names and persistent identifiers for data standards
- Variable annotation improves the quality of submitted data and meaningful use of shared data for secondary analysis
- Mapping to published standards facilitates interoperability amongst shared datasets

N	O	P
Terminology Standard	"Other" Standard	Additional Standard Information
<i>Provide the name of the standard/code list that the variable was collected from (e.g. ICD, MedDRA, WHODrug, etc.); see Recommended_Standards tab for more information.</i>	<i>If selected "other" in column N, please provide the standard/code list used</i>	<i>Use to specify additional information about the standard; e.g. the standard code, the standard version, or specific CDEs, coding system URL, etc...</i>
LOINC		Antibody Screen RBC, PEG, 75265-9
RxNorm		aspirin[RxCUI 1191]
NIH Common Data Elements		Formula fed to infant name, NLM, id PQq_xX5JNo0, version 3.1
Other	ATC/DDD Index 2023	M01 ANTIINFLAMMATORY AND ANTIRHEUMATIC PRODUCTS

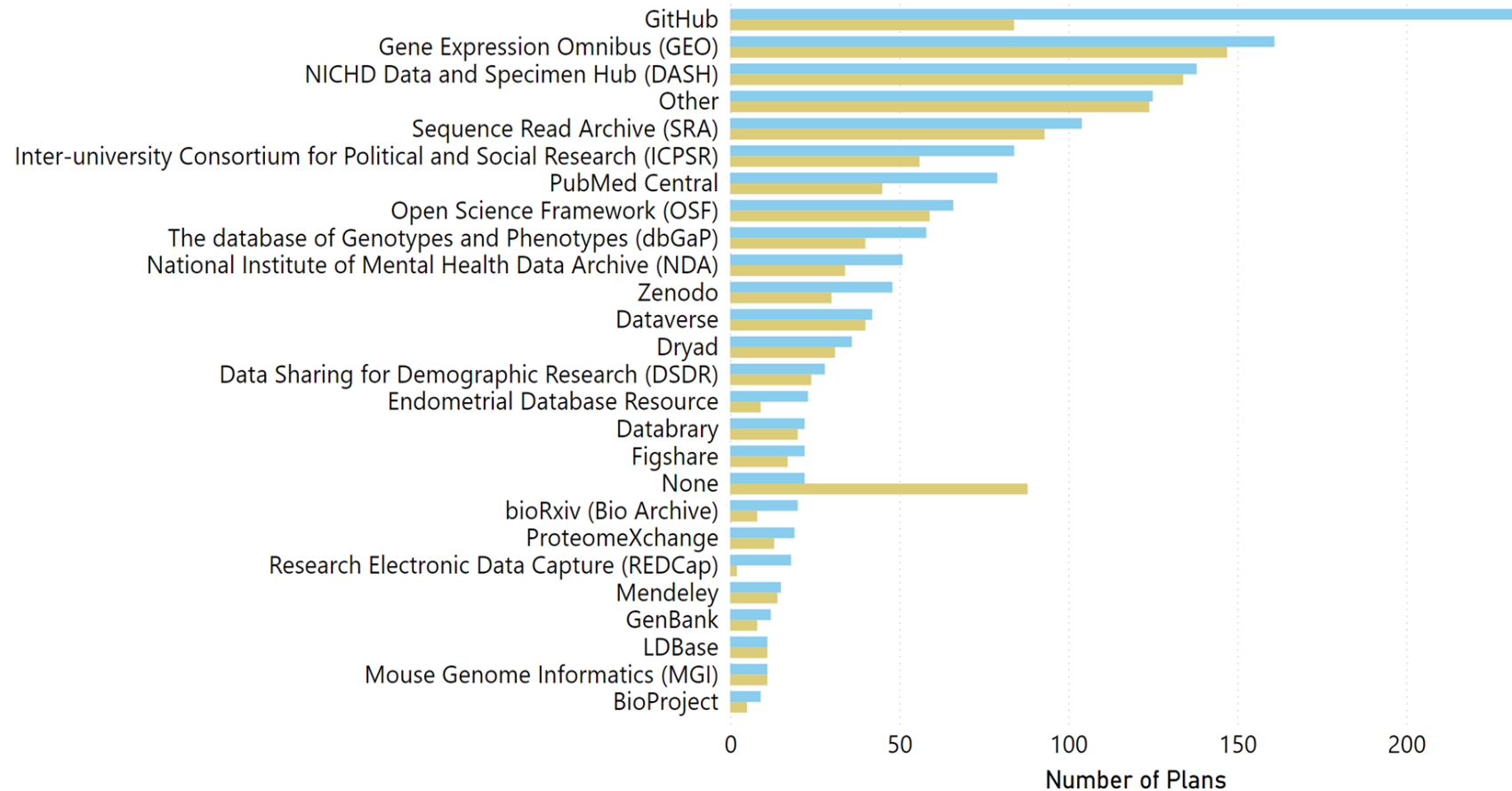


# NICHD Data Management & Sharing Plan Review: *Data Repositories*

- The NICHD-wide analysis detected patterns in repository use to provide better guidance to staff and researchers
- Several named repositories do not align with the “*Desirable Characteristics of Data Repositories for Federally Funded Research*” (described in 2022 White House OSTP memo), including:
  - **GitHub** – good for code/documentation, not good for data sharing (no curation, no central interface for findability, costs for ongoing storage)
  - **PubMed (or Journals “with publication”)** – not a data repository, need a plan for data not associated with publications
- Many researchers did not name any repository for sharing data
  - Not acceptable

Repository Frequency in DMS Plans

● Entire Plan ● Question 4



# Building out Data Repository Capacity: *The NICHD Data Ecosystem*

ODSS implemented a **human-centered design** approach to designing and developing the ecosystem to support NICHD communities' data science and sharing needs:

## Synergistic Activities:

### 1. Assessing all NICHD-relevant data repositories to inform:

- **Sustainability:** Data repositories that do or can meet NICHD needs for long-term data sharing and data security
- **Interoperability:** Approach to building connections between NICHD, NIH, and external data repositories and tools

### 2. Collecting and prioritizing **user stories**, and then implement work to address researcher, staff, participant, and community needs

## Repository Assessment variables address four domains:

- **Supporting DMS Policy** (submitting/sharing data)
- **Using Data** (e.g, metrics of data FAIRness, system interoperability, analysis tools/environment)
- **Sustainability** (capacity to meet long-term needs)
- **Data Governance and Security**



**Many others....**

*NICHD coordinates with the NIH Office of Data Science Strategy to leverage trans-NIH resources, including repositories, tools, standards, & trainings*



# NICHD Data Ecosystem: *Focusing on three types of interoperability*

## Governance

Rules and controls that define and enforce appropriate data collection, access, sharing, linking, and use.  
Examples: policies, regulations, consent, data use agreements

## Data

Semantics, formats, structure, and mappings.  
Examples: data and metadata standards, common data elements, record linkage

## System

System to system connectivity.  
Examples: web services/APIs, single sign-on, cloud-based analysis platforms

## Example Interoperability-Enabled Workflows

**Researchers gain authorized access to sensitive datasets by following data and repository-specific governance requirements**

**Researchers integrate disparate datasets by mapping common data elements, standards, geographic areas, or individual participant identities**


**Authorized researchers co-analyze large datasets by calling repository APIs from cloud analysis tools, using centrally issued identity and entitlement tokens**



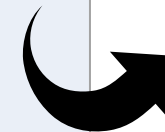
# Our Human Centered Design Approach

Understand users' needs and experiences to create effective solutions:

***All work is driven by user stories***

User Story	Current Problem	User Goal
<b><i>What does the user want to be able to do?</i></b>	<b><i>Why can't the user do this today?</i></b>	<b><i>What is the user's ultimate goal?</i></b>
As a researcher/clinician, I want to combine participant-level data collected from multiple studies to merge multiple data types for each participant and avoid working with inflated sample sizes to effectively study COVID in children 	We believe the same children were recruited for multiple studies with different data collection protocols, but we don't have a way to identify which children are the same without sharing personally identifiable information (PII)	My goal is to link data for the same children across multiple studies and data repositories without sharing PII

***Can record linkage solve this?***



NICHD User Stories:

- The foundation for formal use cases describing desired functionality for technical development work
- Foster collaborative solutioning and adoption of new development work and associated standards
- Shared broadly on NICHD's public GitHub repository (<https://github.com/NIH-NICHD-Ecosystem>)



# NICHD Record Linkage Implementation Checklist



Developed in response to the pediatric COVID data linkage need, the NICHD checklist was utilized to address a user story to link data across NIH data repositories from participants with Down syndrome



## Implementation Checklist

### Governance Considerations :

- Determine the scope of linkage (which datasets to link) → Final Research data shared through multiple NIH data repositories (INCLUDE Data Hub, NIMH Data Archive)
- Obtain approval to link → Data contributors determine if linkage is appropriate (obtain additional approvals, as needed), NIH program approval
- Identify policies and rules that apply to each dataset, specific data type(s), or participant population(s) → NIH Genomic Data Sharing Policy, consent-based data use limitations, other policies
- Establish which party should link the data → Approved researchers will link/merge datasets
- Use a variety of controls to mitigate re-identifiability risk → Share linkages after data access committee approval, sign Data Use Agreement (prohibits re-identification), standard de-identification of all datasets

### Technical Considerations:

- Collect & standardize PII elements for high linkage quality
- Select a technology that meets basic requirements
- Consider PPRL Tool Sustainability for Long-term Implementations



**NDA:** The National Institutes of Mental Health (NIMH) Data Archive  
**GUIDs:** Global Unique Identifiers





# NICHD Data Repository Finder

<https://data-repository-finder.ll.mit.edu/>



**Data Repository Finder** helps researchers find data repositories where they can share data as they develop and implement their NIH Data Management and Sharing Plans. Repositories are added on an ongoing basis.

## Online tool developed to address multiple user stories:

- As an NICHD researcher, I want to find the best data repositories for the sharing scientific data that I am generating in my project and understand the repository's submission and sharing practices so I can complete my Data Management and Sharing Plan
- As an NICHD program officer, I want to verify that researchers are using data repositories appropriate for the scientific data they plan to share and that their DMS Plan is consistent with the repository's expectations and requirements.



- Includes data repositories that generally accept data from NICHD researchers (without special approval)
- **Shares information related to writing and implementing a Data Management & Sharing Plan**

## Filters

Start typing in a box to view related terms. Terms are an exact match to repository tags; select all related terms within a filter box for a broad search?

Learn More ▼

Filter by **Organism**. Select all that apply.

Organisms

Zebrafish ✕ Mouse ✕

Filter by **Research Area**. Select all that apply.

Research Areas

Filter by **Data Type**. Select all that apply.

Data Types

## Search Results (8)

1

2

3

COMPARE



FaceBase

**Affiliation:** University of Southern California

Compare

FaceBase is a collaborative NIDCR-funded project that houses comprehensive data in support of advancing research into craniofacial development and malformation. It serves as a community resource by curating large datasets of a variety of types from the craniofacial research community and sharing them via this website.

**Organisms:** Human Mouse Zebrafish Chimpanzee Chick Xenopus

**Research Areas:** Facial Development and Research Craniofacial Research Craniofacial Development and Malformation

**Data Types:** Genotypes Phenotypes Genetics Sequence Biology RNA Sequencing Expression Data Genomics

Craniofacial Phenotype scRNA-seq Image



Controlled access option



Unclear data volume limitations



Persistent ID used for datasets



Metadata required



Accepts individual-level data



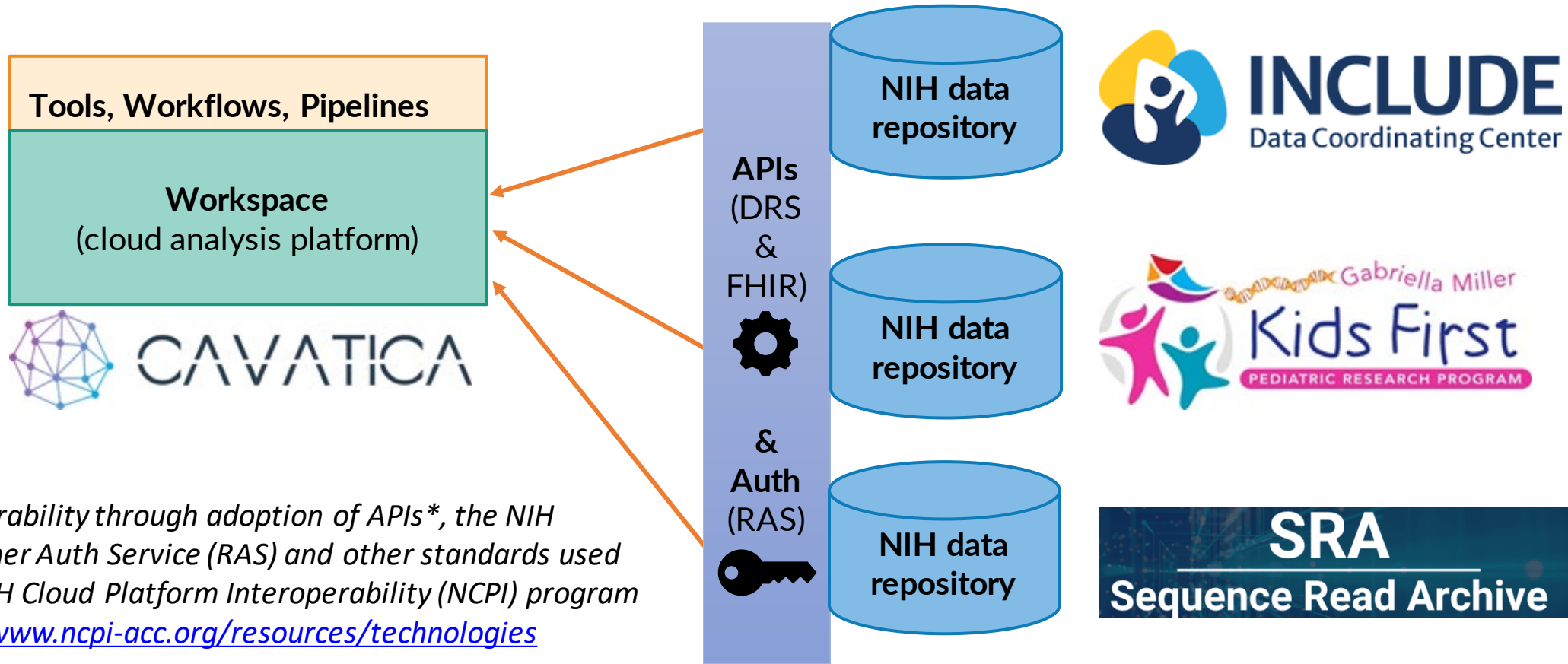
No fee for submission



# Standards-Based Interoperability of Data Repositories and Workspaces



To address a user story to co-analyze large human-derived datasets from three NIH controlled-access data repositories, NICHD worked with NIH and extramural colleagues to enable **collaborative co-analysis of clinical and genomics data in a secure cloud workspace**

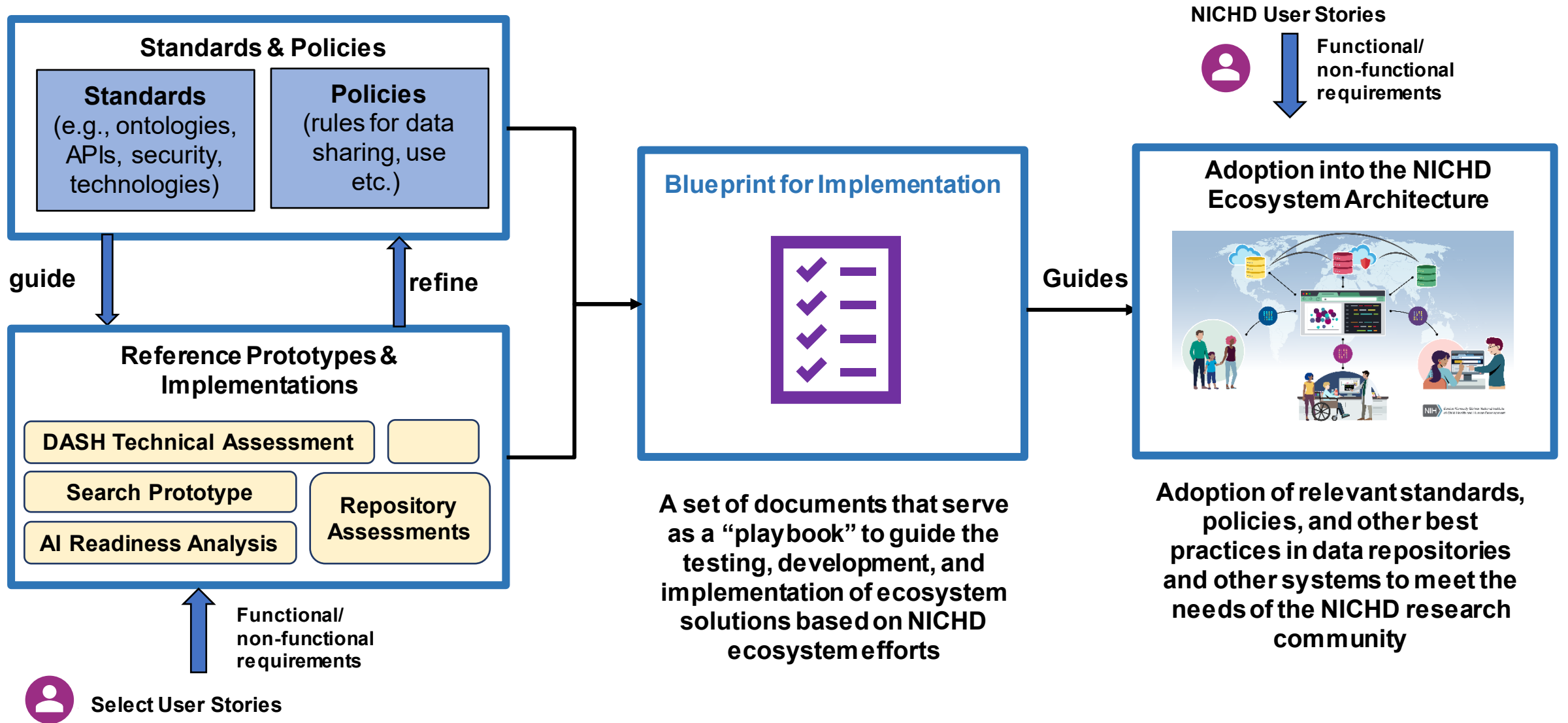


*Interoperability through adoption of APIs\*, the NIH Researcher Auth Service (RAS) and other standards used in the NIH Cloud Platform Interoperability (NCPI) program*  
<https://www.ncpi-acc.org/resources/technologies>

\*API: Application Programming Interface, a way for two or more computer programs to communicate with each other



# NICHD Data Ecosystem Strategy & Blueprint



# Looking Forward

As we continue to build out NICHD data sharing capacity as a fundamental strategy for **responsible and innovative data science**, ODSS will:

- Coordinate development of community-driven Data Management and Sharing Policy guidance, data repository enhancements, and ecosystem interoperability development and standards with the NIH Office of the Director (Offices of Data Science Strategy, Extramural Research, and Science Policy), and other NIH Institutes, Centers, and Offices
- Continue to document diverse NICHD community data sharing user stories; reference user stories to communicate NICHD research community needs to NIH and HHS colleagues (including but not limited to development and implementation of AI strategies)
- Continue to broadly share on GitHub (1) NICHD user stories for data sharing and data science and (2) NICHD and community efforts to address user stories
- Learn how NICHD efforts are responding to our communities' needs





**THANK YOU!**