



# NIH DATA SHARING & SUPPORTIVE INFRASTRUCTURE EFFORTS

**Lytic Jorgenson, PhD**

Acting Associate Director for Science Policy

**Michael Lauer, MD**

Deputy Director for Extramural Research

**Susan Gregurick, PhD**

Associate Director for Data Science

**NIH Advisory Committee to the Director**

June 10<sup>th</sup>, 2022



# **AN INTEGRATED VISION: NIH DATA MANAGEMENT & SHARING POLICY**

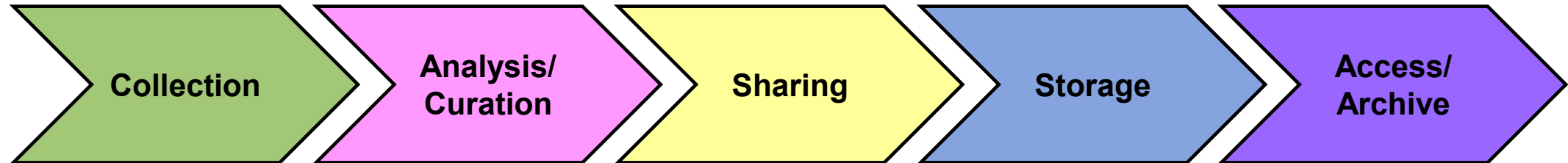
Lytic Jorgenson, PhD

Acting Associate Director for Science Policy



NIH VISION

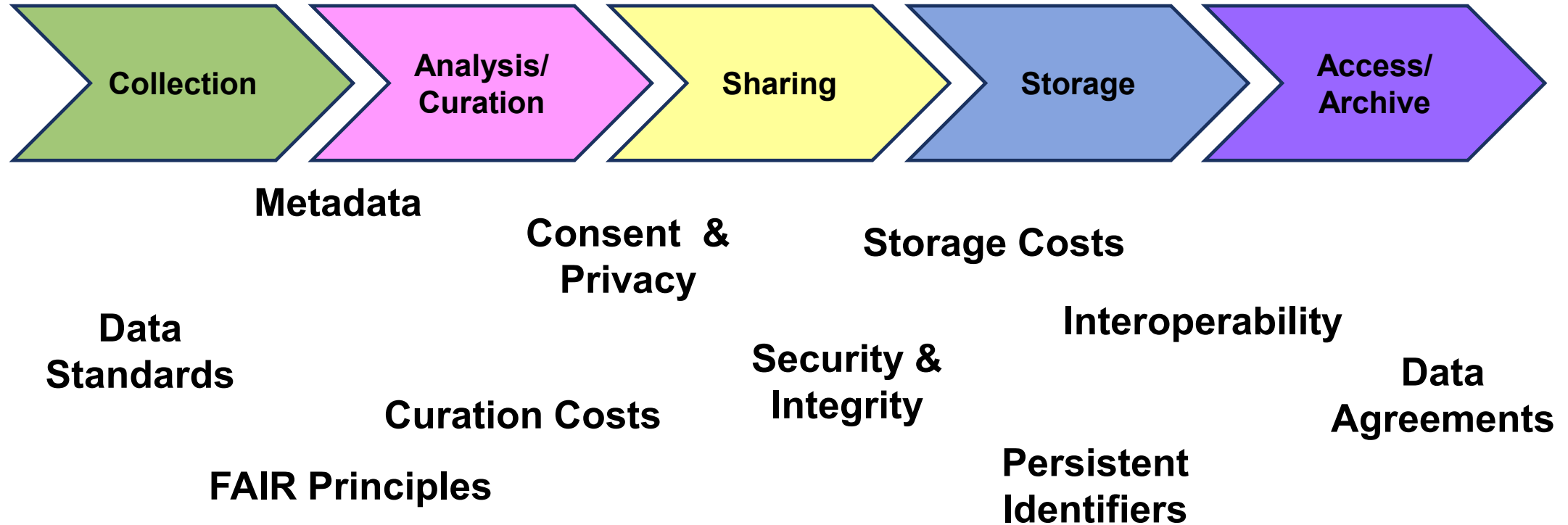
# INTEGRATED DATA FLOW



**Many subcomponents within each of these stages ... for example**

NIH VISION

# INTEGRATED DATA FLOW



NIH VISION

# INTEGRATED DATA FLOW

**Requires coordination and integration across  
policy, implementation, and infrastructure!**

Metadata  
Consent & Privacy  
Storage Costs  
Data Standards  
Curation Costs  
Security & Integrity  
Interoperability  
FAIR Principles  
Persistent Identifiers  
Data Agreements

SOME CONTEXT

## DATA SHARING POLICIES AREN'T NEW

- 2003 NIH Data Sharing Policy (awards >500K)
- 2014 NIH Genomic Data Sharing Policy (human & non-human genomic data)
- 2015 NIH Intramural Human Data Sharing Policy
- 2016 NIH Policy on Dissemination of NIH-Funded Clinical Trial Information
- IC-specific policies and guidelines
- Program-specific policies and guidelines

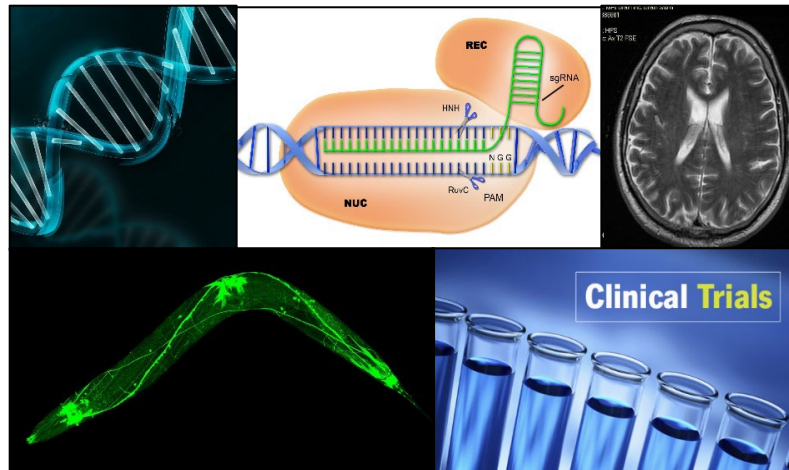
SOME CONTEXT

# DATA SHARING ADVANCES THE NIH MISSION

Sparks New Research Collaborations

Enhances Rigorous Study Design

Makes High-Value Datasets Available



Enables Unique Data Combinations

Facilitates Study Validation

Stimulates New Research Inquiries

**Maximizes Data Collection**

(reduces redundancy/maximizes participant contributions)

**Fosters Stewardship**

(provides transparency/accountability for taxpayer funds)

**Accelerates the Research Enterprise**

(for all the reasons stated above!)

# SOME CONTEXT DATA SHARING IS EXPECTED



*"Increasing the pool of researchers who can access data and decreasing the time it takes for them to review and find new patterns in that data is critical to speeding up development of lifesaving treatments for patients."*

- Joe Biden

## 21<sup>ST</sup> CENTURY CURES ACT

### GOALS OF THE LEGISLATION

#### RESEARCH



Remove barriers to research collaboration



Invest in STEM education



Provide new incentives for the development of rare disease drugs

#### GETTING TREATMENTS TO PATIENTS MORE QUICKLY



Foster coordination to find cures more quickly



Modernize clinical trials to increase access to drugs and treatments



Incorporate patient feedback in drug development and review process

#### KEEPING JOBS HERE AT HOME



Ensure U.S. remains a global leader in medical innovation, protecting and creating jobs at home



Encourage development of new medical apps to save lives and create jobs

#CURESatOne

E&C

*"[The NIH Director] may require recipients of NIH awards to share scientific data, to the extent feasible, generated from such NIH awards ..."*

- 21<sup>st</sup> Century Cures Act

*"Only now that the new Cures Act privacy protections are in place, are we moving forward on the exciting new authority to require data sharing."*

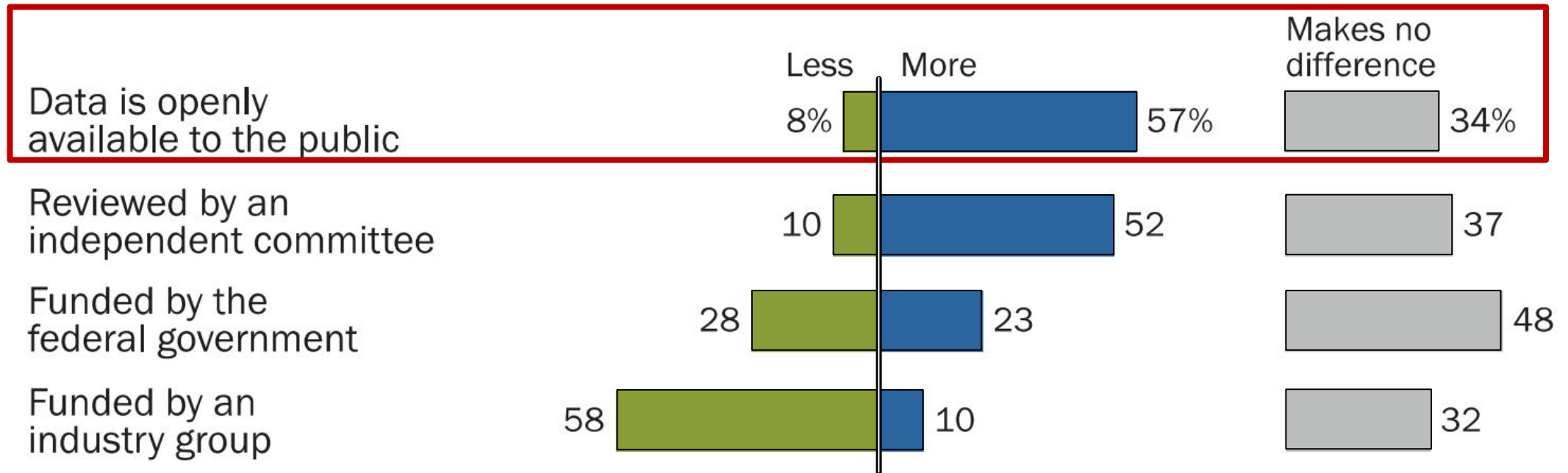
- Francis Collins



SOME CONTEXT

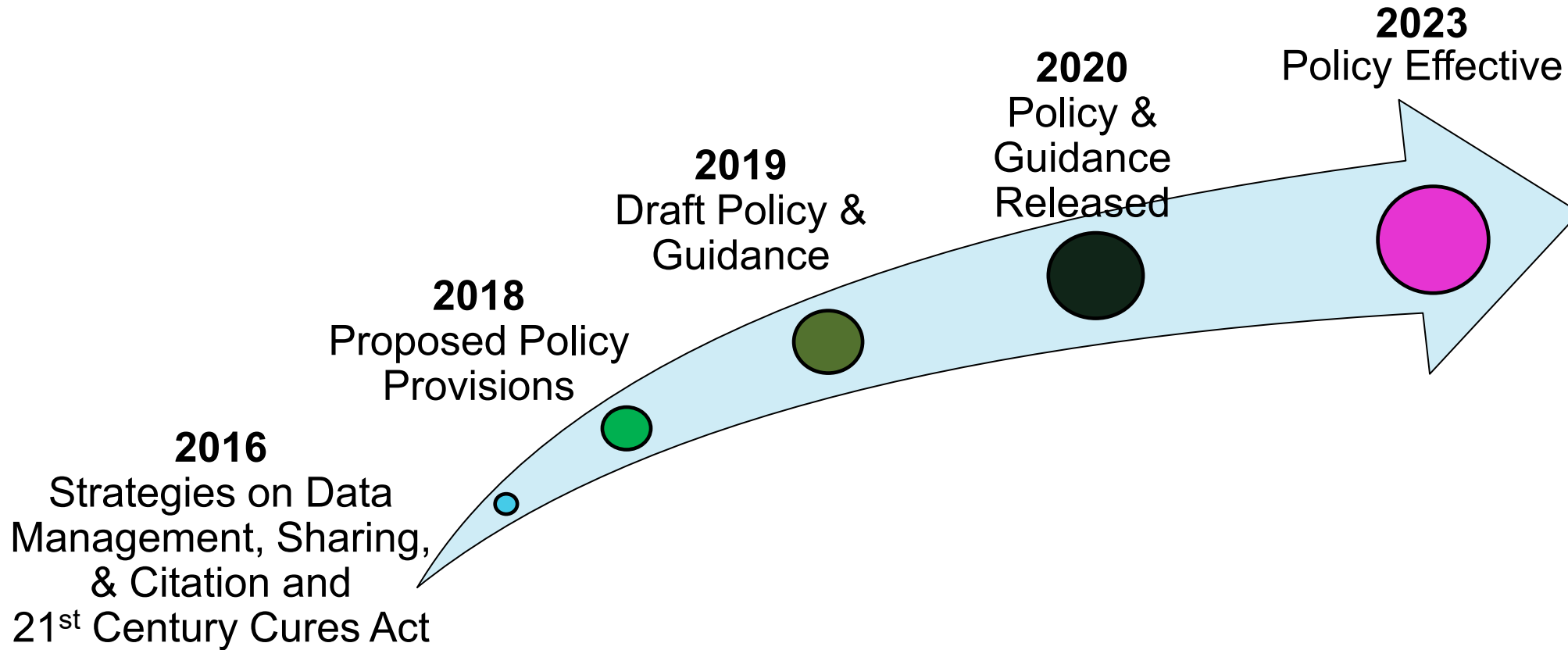
# DATA SHARING PROVIDES ENHANCED TRUST

*% of U.S. adults who say when they hear each of the following, they trust scientific research findings...*



SOME CONTEXT

# LONG ARC OF ITERATION



SOME CONTEXT

# LONG ARC OF ITERATION

**NIH Data Science  
Policy Council WGs**

**2020  
Policy &  
Guidance  
Released**

**2023  
Policy Effective**

**Tribal Consultation**

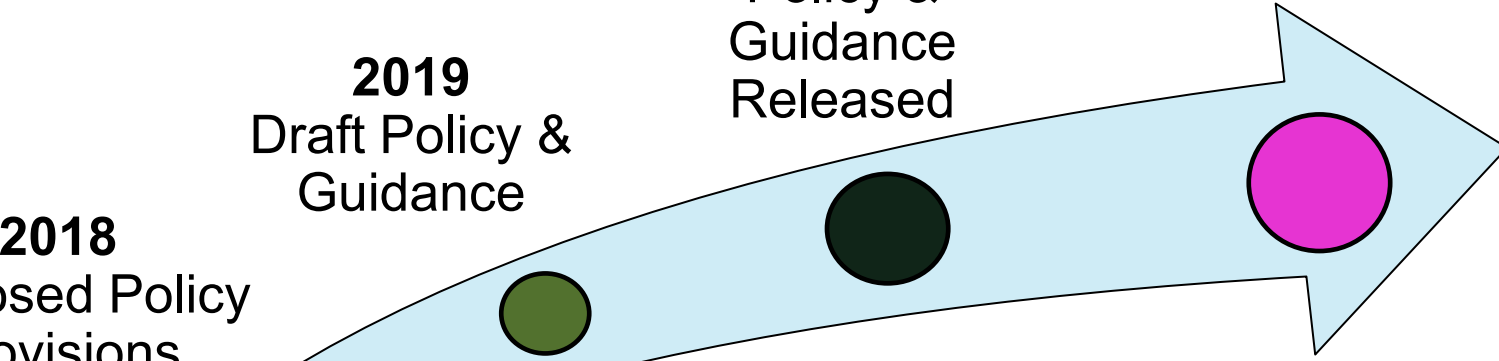
**2019  
Draft Policy &  
Guidance**

**2018  
Proposed Policy  
Provisions**

**Secretary's Advisory Committee for  
Human Research Protections**

**2016  
Strategies on Data  
Management, Sharing,  
& Citation and  
21<sup>st</sup> Century Cures Act**

**White House Subcommittee  
on Open Science**





## NIH POLICY FOR DATA MANAGEMENT & SHARING

- TWO BASIC REQUIREMENTS
  - Submission of a Data Management & Sharing “Plan” for all NIH-funded research
  - Compliance with the ICO-approved Plan
- Effective January 25, 2023 (*replaces 2003 Data Sharing Policy*)

# POLICY EXPECTATIONS

## SHARING SHOULD BE ...

- **The default practice**
  - Data sharing should be maximized
  - Justifiable limits for technical/ethical/legal factors
- **Responsibly implemented**
  - Outline protection of privacy, rights, and confidentiality
  - Abide by existing laws, regulations, and policies
- **Prospectively planned for at all stages of the research process**



# FLEXIBLE POLICY – KEY PARAMETERS

- **All data should be managed but not all data needs to be shared**
  - **What's in:** All NIH-supported research generating *scientific data* “Recorded factual material... of sufficient quality to validate and replicate research findings” – published or unpublished
  - **What's out:** lab notebooks, preliminary analyses, case report forms, physical objects
- **Data should be accessible as soon as possible**
  - No later than publication or end of award
  - Considerations regarding how long data should be shared (e.g., journal policies, repository policies)

# PLAN SUBMISSION & REVIEW (\*EXTRAMURAL)

## Plan Submission

With application

Brief Plan description in Budget Justification

Full Plan as a separate attachment

## Plan Assessment

Peer review comment on (not score) budget

NIH program staff assess Plans

Plans can be revised

## Plan Compliance

Incorporated into Terms and Conditions

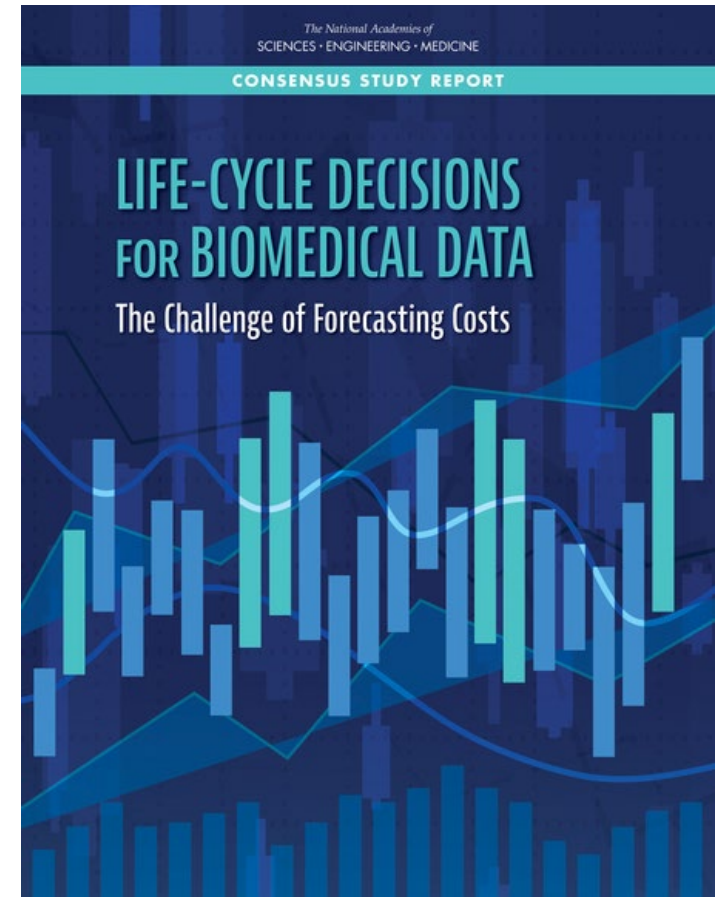
Monitored at regular reporting intervals – mechanisms and tools to support oversight under development

Compliance may factor into future funding decisions

*\*Analogous requirements for contracts, OTAs, IRP*

# ASSESSING DATA SHARING COSTS

- **Framework for cost-effective decision-making regarding preserving, archiving, and accessing biomedical data**
  - Identifies steps and decision points to assist in planning for costs
  - Outlines cost components associated with biomedical information resources
  - Provides interactive framework analyzing cost drivers
- **Relevant life-cycle costs delineated, as well as assumptions underlying the models**





NASEM REPORT

# INCENTIVIZING GOOD SHARING

## Changing the Culture of Data Management and Sharing

A WORKSHOP

Apr 28, 2021 11AM – 4PM ET

Apr 29, 2021 11AM – 3PM ET

The National  
Academies of  
SCIENCES  
ENGINEERING  
MEDICINE

- **Implementation requires a system-wide culture shift**
  - Need for aligned incentives and resources
- **Impactful data sharing is key to successful policy implementation**
  - Practices mindful of secondary data users are necessary for useful data sharing
  - Metrics are needed to assess value
  - Data citation key for adoption and monitoring
  - Trust needs to be earned for success

# FROM POLICY TO IMPLEMENTATION

- Some implementation considerations built into the Policy provisions (i.e., Plan review, compliance infrastructure, etc.)
- **Trans-NIH engagement** in implementation approach, including:





# SUSTAINABLE AND EFFICIENT IMPLEMENTATION STRATEGIES

Michael Lauer, MD

Deputy Director for Extramural Research



# WHAT WE'RE HEARING

- **One person's flexibility is another's ambiguity**
  - *How to interpret scientific data definition?*
  - *Where should data be shared?*
  - *How does this relate to other ICO data sharing policies?*
- **Where is the money coming from?**
  - *Is NIH going to raise budget caps?*
  - *What happens after the award ends?*
  - *How will program officers assess plans?*
- **Bottom line: who is paying attention to this?**
  - *Will Plans be public?*
  - *How will NIH enforce compliance?*



## WHAT'S A GOOD PLAN?

- **Recommended elements of a Plan:**
  - **Data type** - Data to be preserved and shared
  - **Related tools, software, code** - Tools and software needed to access/manipulate data
  - **Standards** - Standards to be applied to scientific data/metadata
  - **Data preservation, access, timelines** - Repository to be used, persistent unique identifier, and when/how long data will be available
  - **Access, distribution, reuse considerations** - Factors for data access, distribution, or reuse
  - **Oversight of data management** - How Plan compliance will be monitored/ managed and by whom

## COMMUNITY RESOURCES

# WHO'S GOING TO PAY?



- **Reasonable costs allowed in budget requests**
  - Curating data/developing supporting documentation
  - Preserving/sharing data through repositories
  - Local data management considerations
- **NOT considered data sharing costs**
  - Infrastructure costs typically included in indirect costs
  - Costs associated with the routine conduct of research (e.g., costs of gaining access to research data)



## COMMUNITY RESOURCES

# WHERE SHOULD THE DATA GO?

- **Encourages use of established repositories**
- **Helps investigators identify appropriate data repositories**
  - E.g., use of persistent unique identifiers, attached metadata, facilitates quality assurance
  - Refers to list of [NIH-supported Data Repositories](#)
- **NIH ICs may designate specific data repository(ies)**



# CONSENT LANGUAGE FOR SHARING

- **Sharing data and biospecimens requires good consent practices**
  - Uphold individual autonomy, strengthen trust in research
  - Communicate clearly the potential risks, benefits
- **NIH heard the community and developed, with your help:**
  - “Points to consider” for investigators, IRBs when modifying consent language
  - Sample consent language for data, biospecimen storage, sharing





# COMMUNITY RESOURCES COORDINATION ACROSS NIH



Search



NIH Staff  | [FAQ](#) | [Contacts & Help](#)

[DATA MANAGEMENT AND SHARING POLICY](#)

[GENOMIC DATA SHARING POLICY](#)

[OTHER SHARING POLICIES](#)

[ACCESSING DATA](#)

[ABOUT](#)

## Expediting the Translation of Research Results to Improve Human Health.

### LATEST NEWS & EVENTS

Gearing Up for 2023: Implementing the  
NIH Data Management and Sharing  
Policy

[View More](#)

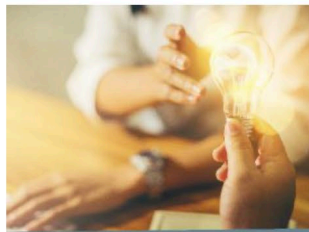
<https://sharing.nih.gov>

# COMMUNITY RESOURCES

# COORDINATION ACROSS NIH

## Sharing Scientific Data

Sharing scientific data accelerates biomedical research discovery, enhances research rigor and reproducibility, provides accessibility to high-value datasets, and promotes data reuse for future research studies. Under the NIH Data Management & Sharing Policy, investigators are empowered to choose the most appropriate methods for sharing scientific data. Learn more about methods for data sharing and selecting data repositories.



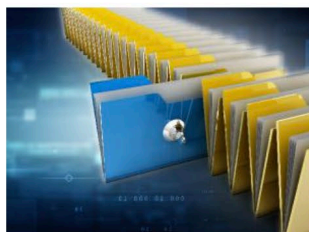
### Data Sharing Approaches

Get familiar with how and when NIH expects data to be shared and learn how to safeguard the privacy of human participants while sharing scientific data.



### Selecting a Data Repository

Learn how to evaluate and select appropriate data repositories.



### Repositories for Sharing Scientific Data

Browse a list of NIH-affiliated repositories for sharing scientific data.

<https://sharing.nih.gov>

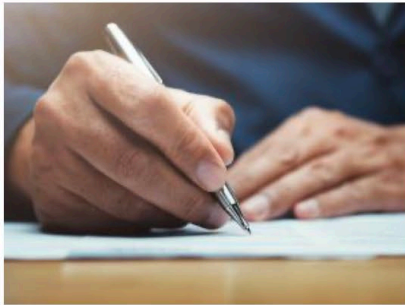
# COMMUNITY RESOURCES

## COORDINATION ACROSS NIH

[Home](#) > [Data Management and Sharing Policy](#) > [Planning and Budgeting for Data Management & Sharing](#)

### Planning and Budgeting for Data Management & Sharing

NIH expects applicants to submit a plan for how they will manage and share their data and allows applicants to include certain costs associated with data management and sharing in their budget.



#### Writing a Data Management & Sharing Plan

Learn what NIH expects Data Management & Sharing plans to address.

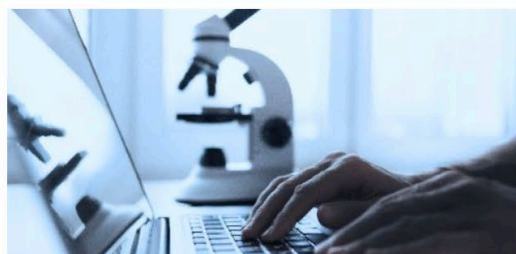


#### Budgeting for Data Management & Sharing

Find out what data sharing related costs may be requested in an application for funding.

<https://sharing.nih.gov>

## Explore the areas in which NIH has sharing policies.



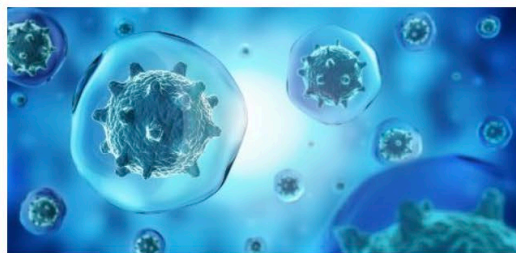
**Scientific Data**



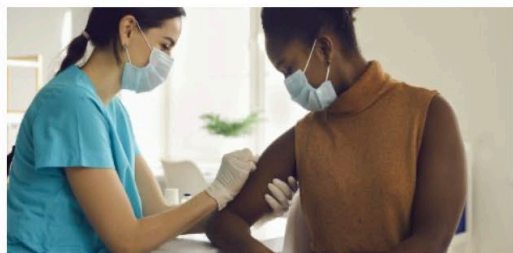
**Genomic Data**



**Research Tools**



**Model Organisms**



**Clinical Trials** ↗

### **Research Publications**

NIH expects that all peer-reviewed manuscripts be publicly available on PubMed Central.

Not sure where to start?

[Find which policies apply to you](#)



# ROADMAP TO 2023 & BEYOND

- **Out now!**

- Webinars & FAQs
- DRAFT Guidance for Researchers working with AI/AN Participants

- **Before 2023:**

- Additional supplemental information (e.g., cost considerations, protecting privacy)
- Sample plans
- Harmonization of GDS Policy

- **Beyond 2023:**

- Ongoing assessment of the Policy for short- and long-term goals
- Incentives for data sharing



OPEN MIKE

*Helping connect you with the NIH perspective, and helpi*





# MAXIMIZING VALUE TO THE RESEARCH COMMUNITY THROUGH INFRASTRUCTURE EFFORTS

Susan Gregurick, PhD

Associate Director for Data Science



# NIH DATA SHARING INFRASTRUCTURE

NIH strongly encourages  
**open access Data Sharing Repositories**  
as a first choice.

[https://www.nlm.nih.gov/NIHbmic/nih\\_data\\_sharing\\_repositories.html](https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html)

Datasets up to **2 gigabytes**

## PubMed Central

Stores publication-related supplemental materials and datasets directly associated publications.



Datasets up to **20 gigabytes**

## Generalist Repositories

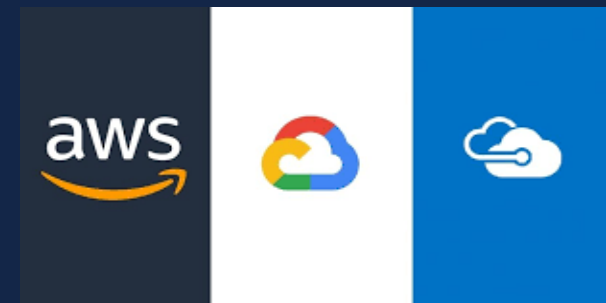
Datasets associated with publications or otherwise and links to PubMed.



High priority datasets **petabytes**

## Cloud Partners (STRIDES Program)

Store and manage large scale, high priority NIH datasets.



# SUPPORT FOR NIH DATA REPOSITORIES

NIH supports a variety of data repositories and knowledgebases of **differing sizes** and **complexities** and at **different levels of maturity**

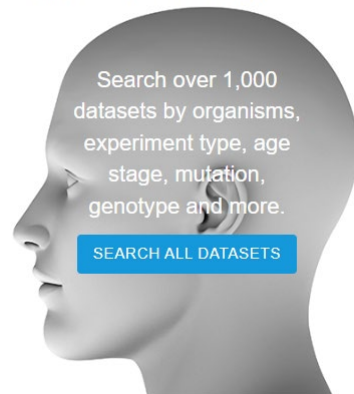
- Each has the **potential** to bring **value** to a given research area, but tend to be at **different stages** of maturity demonstrating that they have the appropriate practices in place to reliably manage the data they ingest and make available
- **Spectrum of ability** and **readiness** to adhere to the characteristics that are desirable for a data repository that are aligned with **FAIR** (**F**indable, **A**ccessible, **I**nteroperable, and **R**eusable) and **TRUST** (**T**ransparency, **R**esponsibility, **U**ser focus, **S**ustainability, and **T**echnology) principles
- **Developing metrics** for evaluating the **usage**, **utility**, and **impact** of a given repository is **evolving** and likely a function of several aspects



# POSITIONING REPOSITORIES FOR SHARING



Explore our repository:



DBAASP<sub>v3.0</sub>

Database of antimicrobial activity and structure of peptides



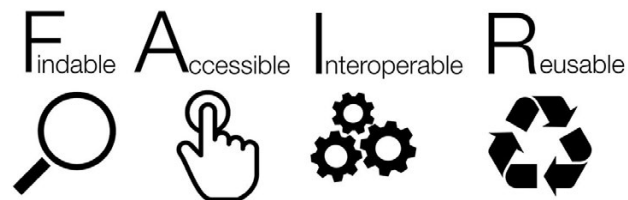
## 17 Awards in 2021:

2 IDeA States  
7 Female PI's  
5 intramural

8 addressing FAIR and TRUST  
6 addressing FAIR, TRUST, and Metrics  
2 addressing FAIR  
1 addressing TRUST

**Biomedical focus areas:** traumatic brain injuries, obesity nutrition, mental health, immune response

**Data types:** imaging, behavioral measures, clinical, EHRs, -omics, speech and language



# DATA REPOSITORY (DR) & KNOWLEDGEBASE (KB) PROGRAM

Fill a scientific need or gap

Encourage adoption of good data management practices

Engage the research community to contribute and use data

Govern data life-cycle and preservation

[PAR-20-089](#) and [PAR-20-097](#)

An NIH program to support investigator-initiated, sustainable data resource development driven by critical research needs



**Pan-Neurotrauma Data Commons**  
U24NS122732-01

**Principal Investigator(s):**  
ADAM R. FERGUSON (contact), PHD  
Karim Fouad, PHD  
Jeffrey S. Grethe, PHD  
Vance P Lemmon, PHD

**Co-Investigators**  
John Bixby, PHD  
Ubbo Visser, PHD  
Michael Beattie, PHD  
Jacqueline Bresnahan, PHD  
J Russell Hule, PHD  
Abel Torres-Espin PHD

**Consultants**  
Maryann Martone, PHD  
Alison Callahan, PHD

**Federal Agency Information**  
9. Awarding Agency Contact Information  
ERNA Petrich  
NATIONAL INSTITUTE OF NEUROLOGICAL DISORDERS AND STROKE  
erna.petrich@nih.gov  
301-496-9245  
10. Program Official Contact Information  
LINDA LOUISE Bambrick  
NATIONAL INSTITUTE OF NEUROLOGICAL DISORDERS AND STROKE



# GENERALIST REPOSITORY ECOSYSTEM INITIATIVE

*Applications from Repositories Working Together to:*



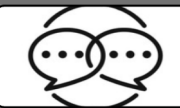
Implement consistent capabilities  
(NOT-OD-21-016)



Create better access to & discovery  
of NIH funded data



Conduct outreach & train on FAIR data  
practices



Engage the research community

*Expected Outcomes*



Make data sharing easier



Improve discoverability



Increase reproducibility of research



Encourage secondary use of data



# GENERALIST REPOSITORY ECOSYSTEM INITIATIVE *(CONT.)*

Align with  
Desirable  
Characteristics for  
Data Repositories

Implement Browse  
& Search for NIH  
Funded Data

Develop Consistent  
Metadata Models

Conduct Limited  
Q/AC of the NIH  
Funded Data

Enable  
Connectivity of  
Digital Objects

Use Case Support  
Including  
(X-Repository Use  
Cases)

Implement Open  
Metrics

Develop  
Educational  
Materials

Conduct Broad  
Outreach  
(Workshops)

Commit to  
“Co-opetition”

**Openly Share Software & Work Products Developed Under Award**

# DATAWorks!<sup>FASEB</sup> Prize

**\$500,000 Total Available**

*Up to 12 monetary prizes recognizing team achievement in data sharing or reuse practices*

**Entries Open:** May 11, 2022

**Entries Close:** July 19, 2022

**DATA SHARING AND  
REUSE**

**Learn More & Enter**

**[www.herox.com/dataworks](http://www.herox.com/dataworks)**

*DataWorks! Prize is a partnership between FASEB and NIH*



# ODSS DATA SHARING & REUSE SEMINAR SERIES

Highlighting exemplars of data sharing/reuse monthly on 2nd Friday

## Past Speakers:



**Karen E. Adolph, PhD**  
Databrary: Secure and Ethical  
Sharing of Research Video as Data  
and Documentation



**Purvesh Khatri, PhD**  
Adventures of a Data Parasite:  
Accelerating Clinical Translation  
Using Heterogeneity in Public  
Data



**Alexander Ropelewski**  
The Brain Image Library:  
A Resource for Sharing  
Microscopy Data

<https://datascience.nih.gov/nih-data-sharing-and-reuse-seminar-series>

NIH VISION

## DATA SHARING EFFORTS



**Scientific data are the catalyst for biomedical breakthroughs and treatments**



**Integrated policies, resources, and infrastructure are key to sharing and reuse**



**NIH aims to promote effective data sharing as the rule, not the exception**



**ACD DISCUSSION – 45 MINUTES**

