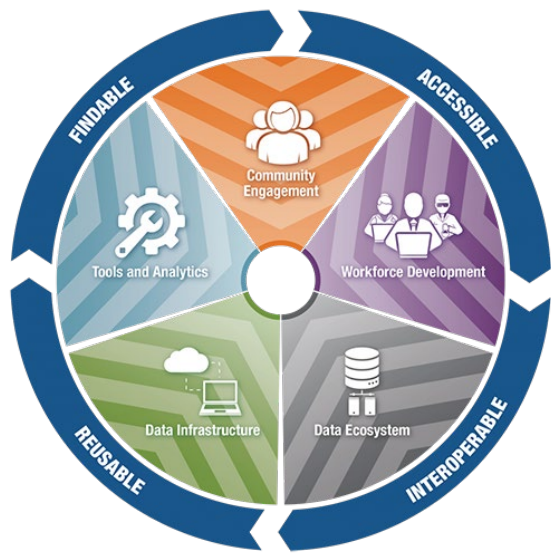


# 2023-2028 NIH-Wide Strategic Plan for Data

---

SUSAN GREGURICK, PHD, DIRECTOR ODSS/DPCPSI

COUNCIL OF COUNCILS: SEPTEMBER 7, 2023



*Progress towards the promise of data driven discovery requires a unified effort across the NIH Institutes, Centers, and Offices (ICOs) that is coordinated by and stimulated with the resources of the Office of Data Science Strategy (ODSS)*

## NIH Strategic Plan for Data Science

# ODSS Mission

Catalyze new capabilities in biomedical data science and related areas by providing NIH wide leadership and coordination to modernize the NIH data resource ecosystem, support a diverse and talented data science workforce, and build strategic partnerships to develop and disseminate advanced technologies and methods

# Some Examples of Impact since 2018

STRIDES Partnership with AWS, Google, Azure results in 248 petabytes of data available for research



New Cloud Lab training environment (since 6/23/23), over 200 trainees

Single-sign on and efficient data access across 30 NIH Data Platforms



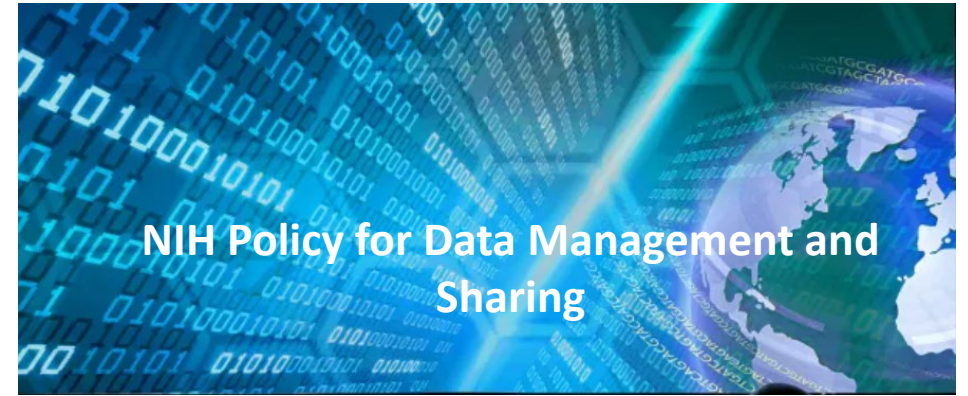
Over 100 software supplemental awards totaling over \$18M



Partnership with NSF, 17 awards across 11 NIH ICOs



21 scholars have joined NIH across 15 ICOs



**MITRE** Data Management Center of Excellence


Over 100 applicant teams and 11 awardees

DATAWorks!<sup>FASEB</sup>

Prize



# Updating the NIH Strategic Plan for Data Science



All-Hands Data Science 2-Day workshop early 2022  
NIH writing team draft summer 2022

Scientific Data Council input winter 2023  
NIH Institutes, Centers and Offices input spring 2023

NIH Leadership review summer 2023  
Community input fall 2023

# Overarching Principles of the Plan



Communicates achievable goals data science



Buildings on successes of current NIH Strategic Plan for Data Science



Includes goals and priorities of NIH Institutes, Centers, and Offices



Includes broad goals and specific priority activities



Provides accountability (e.g., measures of progress)



Input from internal and external stakeholders

# Feedback that helped shape the Plan

---

- ✓ Emphasize the NIH Policy for Data Management and Sharing as a key component of the NIH Data Science Strategy
- ✓ Identify the unique challenges and opportunities of human derived data for research, including SDoH/EDoH, standards, and Common Data Elements
- ✓ Greater emphasis in AI and ML methodologies that can enhance and accelerate basic science research
- ✓ Identify policy and governance blockers in important data science activities

**Challenge and Opportunity:** Generation and dissemination of FAIR data in a manner that will foster greater sharing and add value to the NIH research investments, including cost effective strategies for sustainable, secure, and accessible biomedical data repositories and knowledgebases

**Goals over next five years:**

1. Support the biomedical community to manage, share, and sustain data
2. Enhance FAIR data and greater data harmonization
3. Strengthen NIH's data repository and knowledgebase ecosystem

**Overall Impact:**

- ✓ *Established guidelines, processes, data sharing tools and training in data management*
- ✓ *Increased development and use of community agreed upon standard schemas and metadata, enhance automated ontologies and automated curation processes, includes agreed upon core common data elements*
- ✓ *Ensure the long-term sustainability of NIH data assets*

**Goal 1:**

*Capabilities to Sustain the NIH Policy for Data Management and Sharing*

## Goal 2:

### *Programs to Enhance Human Derived Data for Research*

**Challenge and Opportunity:** Acquisition and protection of data obtained from electronic health records, and other real-world data, including data captured outside of traditional healthcare settings, that preserves privacy and promotes participant consent

#### **Goals over next five years:**

1. Improve access to and use of clinical and Real-World Data
2. Adopt Health IT standards for research
3. Enhance the adoption of Social and Environmental Determinants of Health (SDoH) for health equity
4. Cross-disciplinary training to empower clinical data science

#### **Overall Impact:**

- ✓ *Increased support for research in clinical and healthcare data science, including new methods for privacy protection, participant informed consent, and data governance*
- ✓ *Increased support for developing tools to collect and analyze data from wearable devices and other new RWD technologies*
- ✓ *New programs to integrate social and environmental determinants with clinical data*



## Goal 3:

### *New Opportunities in Methods and Artificial Intelligence*

**Challenge and Opportunity:** Emergence of innovations in trustable AI approaches that reduces bias and risks and are fair, validated, and explainable. Creation of opportunities for exploration of new technologies and computing paradigms for biomedical research

#### **Goals over next five years:**

1. New Opportunities to enhance Artificial Intelligence (AI), including ethical AI for biomedicine
2. Develop and validate cutting edge software technologies
3. Supporting FAIR software sustainability

#### **Overall Impact:**

- ✓ *New opportunities to support collaborations to develop socio- technical- solutions, including guidelines and principles, for ethical AI, including new technologies and methods to take advantage of foundational models.*
- ✓ *Enhance support for FAIR software development, including supporting partnerships between biomedical and data science researchers*

Goal 4:  
*Support for a  
Federated  
Biomedical  
Research Data  
Infrastructure*

**Challenge and Opportunity:** Challenges in finding, accessing and comparing data from different NIH supported databases, including data on health outcomes and risk factors

**Goals over next five years:**

- Enhance utilization of cloud and hybrid computing architectures
- Support efficiencies and sharable technologies across in NIH data platforms
- Expand researchers' ability to access data and ensure accountabilities for privacy protection and cybersecurity of systems
- Ensure a robust and connected data resource ecosystem that includes interoperability across NIH supported cloud platforms
- Develop new capabilities for data search and discovery

**Overall Impact:**

*Enhanced and increased NIH supported data systems for researchers to search, discover, access, and analyze data across these resources, with enhance the accuracy, validity, transparency, and reproducibility of these capabilities.*

Goal 5:  
*Strengthen a  
Broader  
Community of  
Data Science*

**Challenge and Opportunity:** *Decrease disparities across institutions, regions, and global partners in data science*

**Goals over next five years:**

1. Increase training opportunities in data science
2. Develop and advance initiatives to expand the data science workforce
3. Enhancing data science collaboration within the NIH intramural Research Program
4. Broaden and champion capacity building and community engagement efforts

**Overall Impact:**

*Increased use of data science approaches in biomedical and behavioral research, enhanced diversity of data scientist workforce, and growth in data science skills among clinician scientists.*

# Five Year Outcomes

- Innovative approaches to data curation, harmonization, and validation
- Increase support for communities to develop and implement new Common Data Elements and standards in priority disease areas
- Increase support for research on clinical and healthcare data science, including methods for privacy protection, participant informed consent, and data governance
- Increase support for tools to collect and analyze data from wearable devices and other new RWD technologies
- Develop new research, training programs, and collaborations in AI and Bioethics
- New capabilities for researchers to search, discover, access, and analyze data
- Engage researchers and communities in data science training across biomedical, social, environmental, and behavioral disciplines