



National Institutes of Health
Office of Data Science Strategy

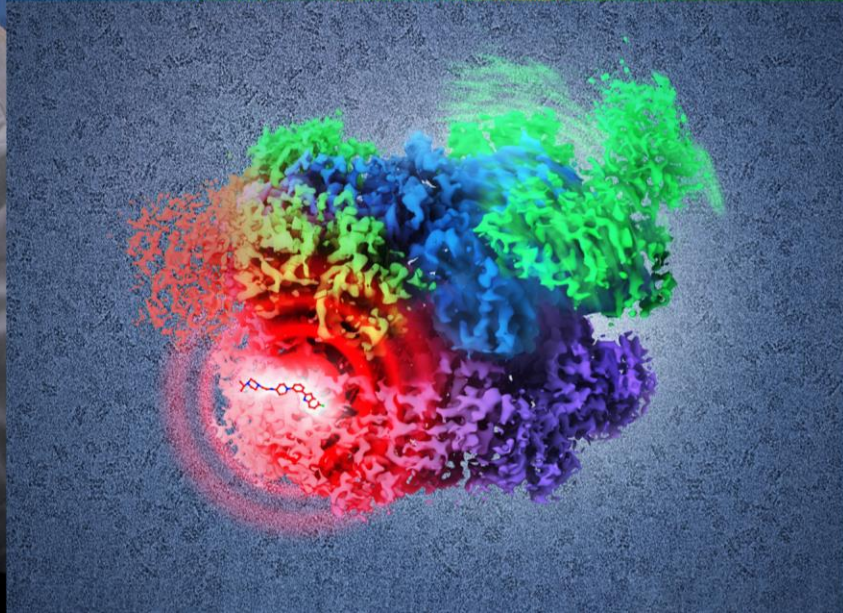
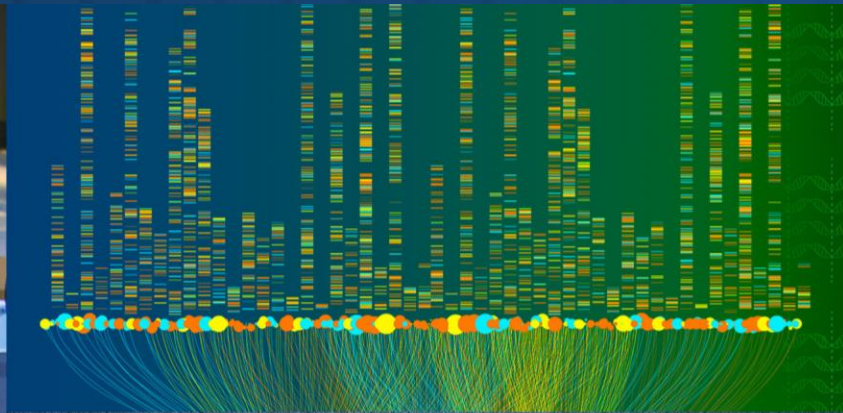
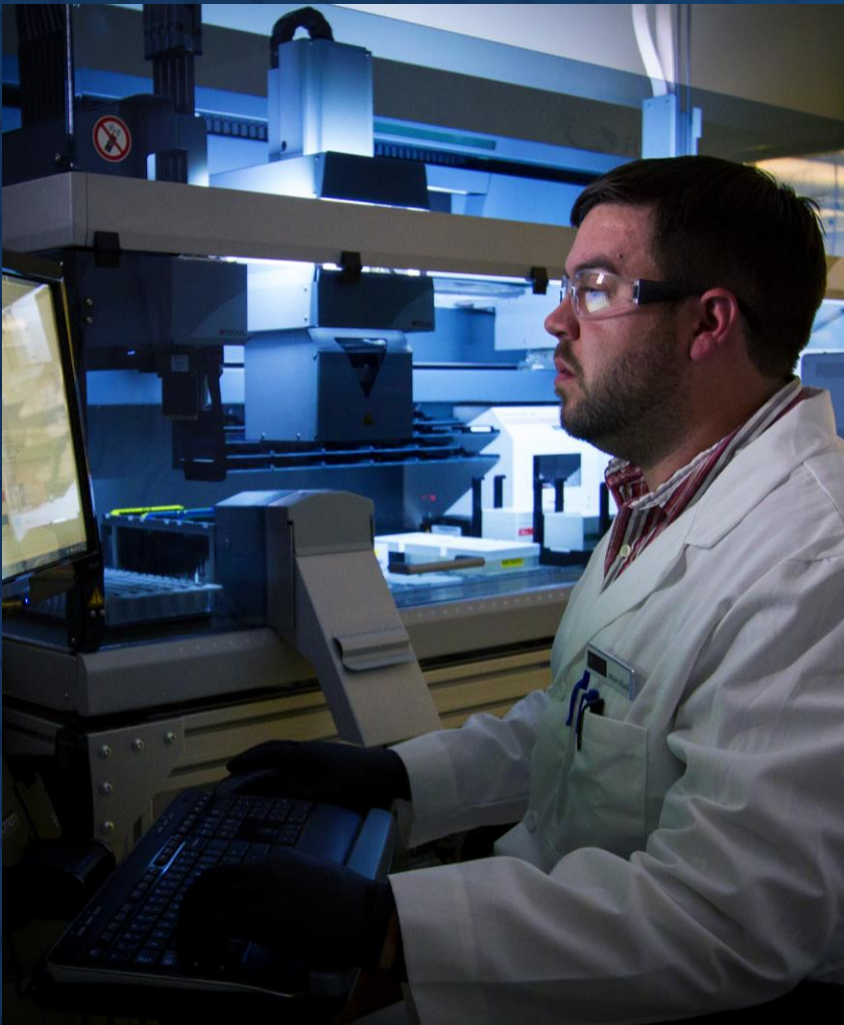
NOSI AI Supplement PI Meeting

NIH Data Science Strategic Plan Overview

Dr. Susan Gregurick, Associate Director For Data Science, NIH

27 March 2024

Progress is accelerated when advanced scientific methods, such as new data analytics, are applied to data that includes everyone, and when new discoveries are rapidly and equitably adopted in clinical care.



Current Challenges in Data Science



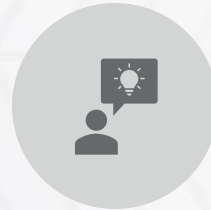
Need for the generation of FAIR data in a manner that fosters greater sharing and integration of scientific results



Need for cost effective strategies for sustainable, secure, and accessible biomedical data repositories and knowledgebases



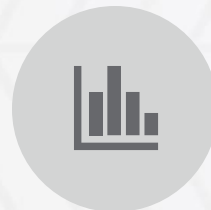
Need for acquisition and protection of data obtained from electronic health records and other real-world data that preserves privacy and enhances participant consent



Emergence of innovations in trustable artificial intelligence (AI) approaches that reduces bias and risks



Creating opportunities to explore new technologies and computing paradigms for biomedical research



Addressing disparities across institutions and regions in data science

Data Science in the next 5 years



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



Develop Programs to Enhance Human Derived Data for Research



Provide New Opportunities in Software, Computational Methods, and Artificial Intelligence



Support for a Federated Biomedical Research Data Infrastructure



Strengthen a Broad Community in Data Science

Goal 1

Capabilities to Sustain the NIH Data Management and Sharing Policy

Challenges

- Need for the generation of FAIR Data in a manner that will foster greater sharing and the integration of scientific results
- Need for cost effective strategies for sustainable, secure, and accessible biomedical data repositories and knowledgebases

Objectives to Address Challenges

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

Goal 2

Enhance Human Derived Data for Research

Challenges

- Need for acquisition and protection of data obtained from electronic health records, and other real-world data, that preserves privacy and enhances participant consent
- challenges in data quality, privacy and confidentiality, policy, regulatory, and ethical issues associated with healthcare and administrative data
- need to better understand the ethical, legal, and social implications of data linkage

Objectives to Address Challenges

- 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 for health equity

Interoperability with Human-derived Data

- **Data exchange standards:**

Fast Healthcare Interoperability Resources (FHIR®): a standards-based application programming interfaces (APIs) to facilitate health data collection and exchange across different health systems

Promotes the development of **patient-centric mobile health applications** or devices in the context of clinical research, as well as **clinical decision support**

Common data elements (CDE): standardized, precisely defined questions paired with a set of specific allowable responses, used systematically across different sites, studies, or clinical trials to ensure consistent data collection

U.S. Core Data for Interoperability (USCDI): ONC requirements for healthcare data

Request for Information: Common Data Elements (CDEs)

NIH is soliciting public input on

- A set of minimum core CDEs;
- Additional CDEs for social determinants of health (SDoH) and clinical domains including autoimmune diseases and immune-mediated diseases;
- Technologies, tools and policies that could facilitate the use of NIH CDEs.

Due Date: April 20, 2024.



Read the RFI here:
<https://bit.ly/3T3Q3U3>

Goal 3

Support for a Federated Biomedical Research Data Infrastructure

Challenges

- Creation of opportunities for exploration of new technologies and computing paradigms for biomedical research

Objectives to Address Challenges

- 1) Develop, test, validate, and implement ways to integrate NIH data and infrastructure
- 2) Ensure a robust and connected data resource ecosystem that includes collaborative data management platforms, curation, analysis, and sharing of data and metadata
- 3) Develop new capabilities for data search and discovery

Goal 4

Strengthen a Broader Community in Data Science

Challenges

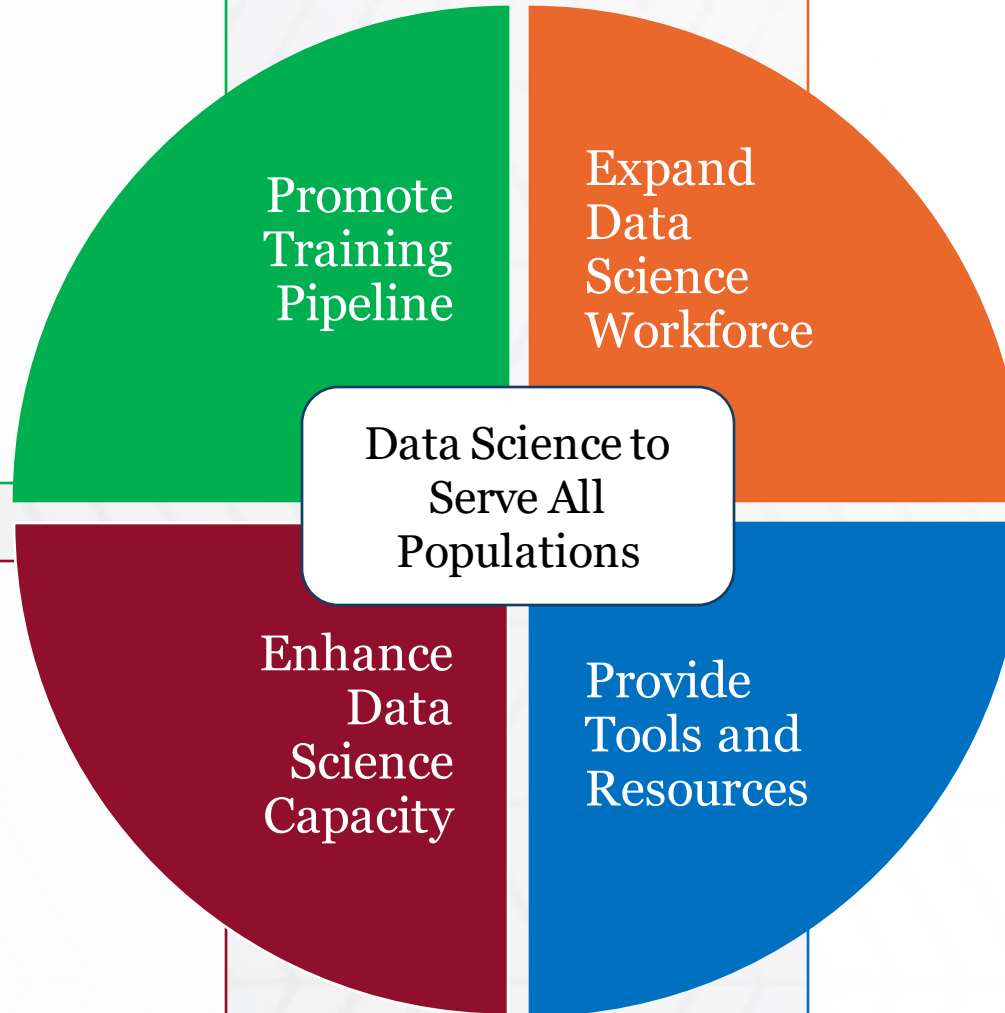
- Develop and nurture data science talent from a diverse array of scientific interests and institutions

Objectives to Address Challenges

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

Training, Workforce, and Community Engagement in FY23

- Institutional Training Awards:
 - 14 R25
 - 7 T32
- Individual Training Awards:
 - 9 Diversity K
 - 1 Diversity F
 - 7 Diversity Supplements
- Intramural Training:
 - GDSSP



- DATA Scholar Program:
 - 10 new scholars
 - 7 scholars completed program
 - ODSS gained FTE slot
- DataPath Program:
 - Concept approved and program launched
 - 3 new fellows
- Leadership Scholars Program

- 12 Enhancing Capacity Supplements
- 4 NARCH awards
- 1 Tribal Epidemiology Center
- 4 Cloud-Based Learning Modules supplements (IDeA)
- 1 Education Hub (NHGRI)
- 2 DS-I Africa awards
- 2 FIC LAUNCH program

- Completed Phase 1 of DataSciZone
- Assumed management of the NIH Coursera Program
- Conducted codethon in collaboration with ASBCB

Goal 5

New Opportunities in Software, Computational Methods, and AI

Challenges

- Emergence of innovations in trustable artificial intelligence (AI) approaches that reduces bias and risks
- Multi-dimensional data integration remains a significant challenge for biomedical and behavioral research

Objectives to Address Challenges

- 1) New opportunities to improve biomedical AI analysis
- 2) Develop cutting edge software technologies
- 3) Support FAIR software sustainability

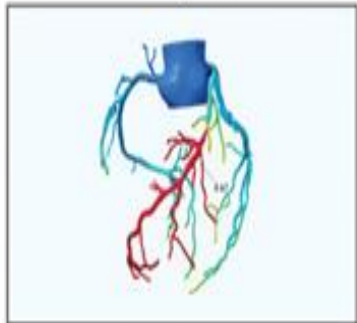
Top Medical AI Procedures and Adoption



Coronary Artery Disease
0501T-0504T



Example product:
HeartFlow Analysis



Diabetic Retinopathy
92229



Example product:
LumineticsCore



Coronary Atherosclerosis
0623T-0626T



Example product:
Cleerly



Liver MR
0648T, 0649T



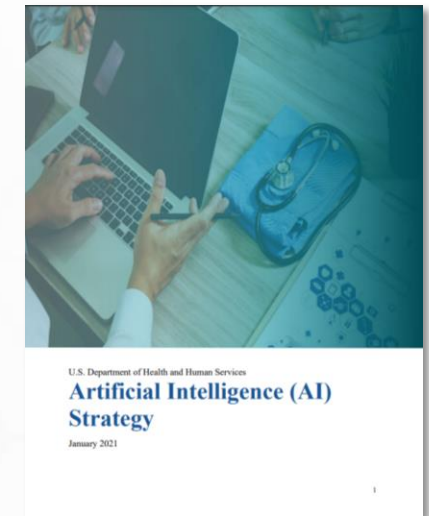
Example product:
Perspectum LiverMultiScan



- Human–AI collaboration enables empathic conversations on mental health
- Use of GPT4 to Diagnose Complex Clinical Cases
- Characterizing the Clinical Adoption of Medical AI Devices through U.S. Insurance Claims

NEJM AI 2023; 1 (1) DOI: 10.1056/Aloa230003

Accelerating Trustworthy AI



OCTOBER 30, 2023

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

<https://www.ai.gov/strategic-pillars/advancing-trustworthy-ai/>

<https://www.hhs.gov/sites/default/files/hhs-ai-strategy.pdf>

<https://www.hhs.gov/sites/default/files/hhs-trustworthy-ai-playbook.pdf>



<https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government>

NIH Office of Data Science Strategy

Goals of the Executive Order on Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence

Objectives

Establish Policies on AI in Health and Human Services

Advance Quality and Safety of AI in Health

Leverage Grantmaking to advance AI use

Promote compliance with non-discrimination and privacy laws

EO Implementation timeline



- HHS EO obligations will be met throughout process
- HHS AI Task Force will brief senior leadership every three months
- **Note:** Resource needs to execute plan not yet met; ability to meet ambitions will be directly impacted by investment levels

Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence



Promoting responsible innovation, competition, and collaboration, includes **investments in AI-related education, training, development, research, and capacity**, while simultaneously tackling novel intellectual property (IP) questions and other problems to protect inventors and creators.



Establish a program to identify and **attract top talent in AI**, and other critical and emerging technologies, at universities, research institutions, and the private sector overseas and expedited adjudication of visa petitions and applications for overseas talent.



Launch a pilot program **implementing the National AI Research Resource (NAIRR)**.



Support **2024 Leading Edge Acceleration Project** cooperative agreement awards to improve healthcare-data quality, support the responsible development of AI tools for clinical care, real-world-evidence programs, population health, public health, and related research.



Accelerate the National Institutes of Health Artificial Intelligence/Machine Learning Consortium to Advance Health Equity and Researcher Diversity (AIM-AHEAD) program and showcasing current AIM-AHEAD activities in underserved communities.

HHS AI Task Force will:



Establish policies to both harness the potential and manage the risks of AI through the safe, equitable, responsible, and ethical adoption and use of AI in health and human services



Advance Quality and Safety of AI in health through assurance standards and quality management processes



HHS will evaluate and deploy its policy and funding tools (e.g., grantmaking, contracting) as appropriate to advance and manage risk in the safe, equitable, responsible, and ethical development and use of AI across the health and human services delivery value chain



Provide public education across the healthcare ecosystem and constituents – from individuals to organizations and states – on AI development and use in health and human services delivery



Where AI offers the opportunity to accelerate the advancement of HHS' mission and the tools meet the Department's quality and equity standards, deploy and evaluate AI capabilities across HHS to drive process innovation and modernization

Core Team and Project Management Office

- Support Task Force (TF), including both Steering Committee and Work Groups
- Ensure internal accountability and drive progress
- Function as connective tissue, both across USG and with industry



Susan Gregurick



Syed Mohiuddin



Greg Singleton



Troy Tazbaz



Micky Tripathi



Jesse Isaacman-Beck



Sebastian Jayaraj



Jenna Osborn



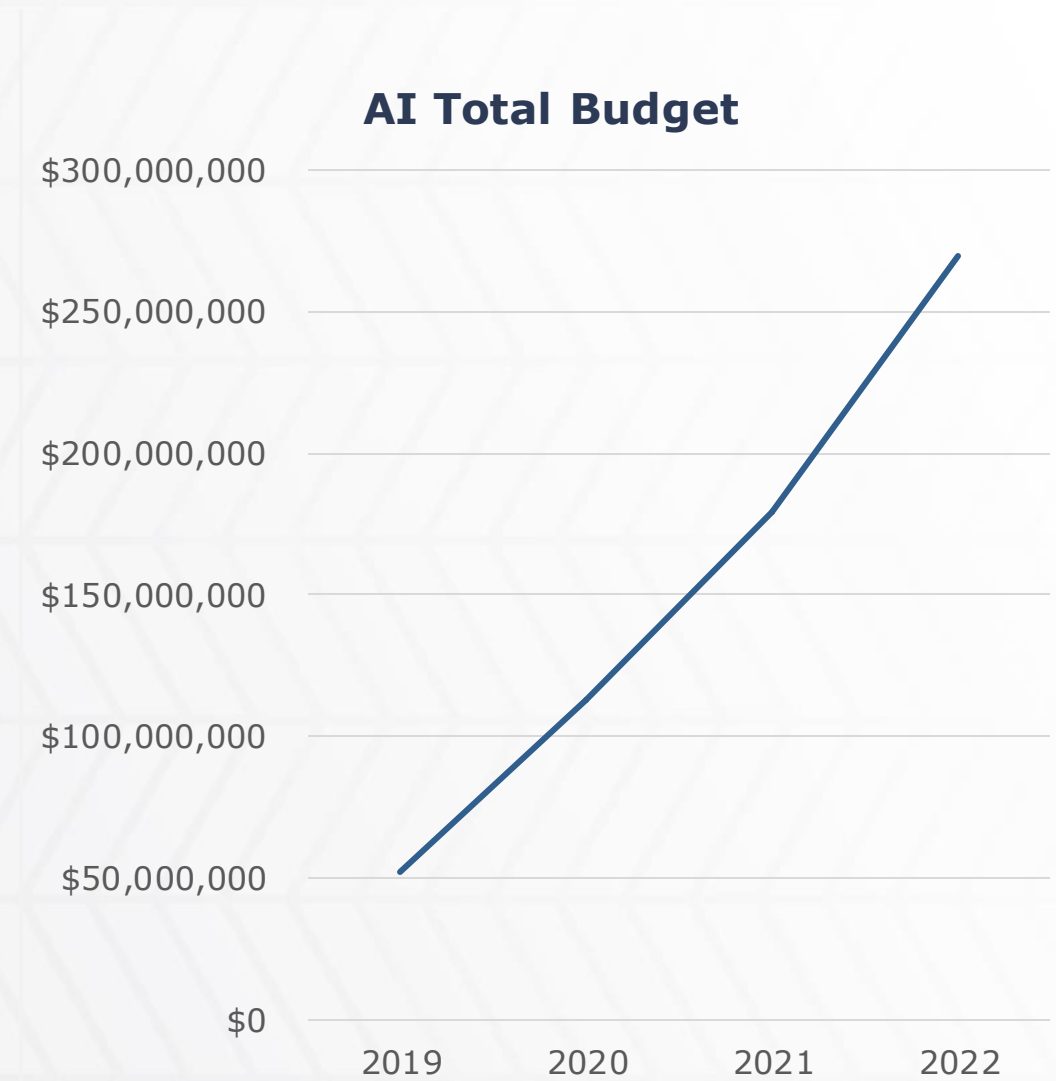
Tricia Lee Rolle



Akshay Venugopalan

Examples of Progress

- **Bridge2AI** to generate new “flagship” datasets and best practices for machine learning analysis.
- **AIM-AHEAD** to enhance participation of underrepresented communities in AI/ML research.
- **ScHARe** to test AI bias mitigation strategies and to advance health disparities research.
- **DEMONSTRATE** to guide healthcare providers and systems in safe opioid prescribing.
- **CARD** to extract insights on disease risk and protective factors from large networks of data
- **Improved operations in health** through developing AI computation tools to improve screening



Support for AI-Enabled research

- Validation of Digital Health and Artificial Intelligence Tools for Improved Assessment in Epidemiological, Clinical, and Intervention Research (NOT-CA-24-031)
- Explainable *Artificial Intelligence* for Decoding and Modulating Neural Circuit Activity Linked to Behavior (NOT-MH-23-110)
- Advancing Data Science Research in HIV: Responding to a Dynamic, Complex, and Evolving HIV Epidemic with *Artificial Intelligence*/Machine Learning (NOT-MH-23-350)
- Advanced Computational Approaches to Elucidate Disease Pathology and Identify Novel Therapeutics for Addiction (NOT-DA-21-004)
- [Harnessing *Artificial Intelligence* and Polypharmacology to Discover Pharmacotherapeutics for Substance Use Disorders \(RFA-DA-25-054 R43/R44 and NOT-CA-24-031 and RFA-DA-25-053 R41/R42\)](#)



NIH Contributions to the NAIRR Pilot

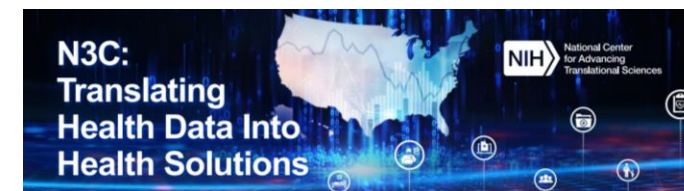
Governance	<ul style="list-style-type: none">• Experience developing and overseeing federated interoperability
NAIRR Open	<ul style="list-style-type: none">• Integration of ImmPort datasets into the NAIRR• Integration of Health Equity Action Network (HEAN) datasets and SchARE analysis tools into the NAIRR
NAIRR Secure	<ul style="list-style-type: none">• Integration of the Medical Imaging and Data Resource Center (MIDRC) and National COVID Cohort Collaborative (N3C) into NAIRR Secure.
Software Stack	<ul style="list-style-type: none">• Coordinate with NSF and DOE a NAIRR software stack community workshop.
Classroom	<ul style="list-style-type: none">• NIH Cloudlab and other platform tools leveraged in NAIRR
Outreach	<ul style="list-style-type: none">• Leverage NIH networks to attract diverse users and data.

NIH Data and Computational Infrastructure Ecosystem



STRIDES Initiative






STRIDES / CloudLab



New Opportunities in Artificial Intelligence

Next 5 Years

The activities below will introduce new opportunities to support collaborations in developing socio-technical solutions, including guidelines and principles, for ethical AI, including new technologies and methods for foundational models.

-  **Develop** social and technical solutions for ethical AI
-  **Create** and validate an approach for using synthetic clinical datasets for AI
-  **Leverage** new technologies and methods for AI and foundational models to accelerate biomedical and behavioral research
-  **Develop** new AI technologies that will enable the translation of data to knowledge
-  **Enhance** NIH capabilities in AI through partnerships across federal agencies and communities

NIH Office of Data Science Strategy

datascience.nih.gov

A modernized, integrated, FAIR biomedical data ecosystem



[@NIHDataScience](https://twitter.com/NIHDataScience)



linkedin.com/showcase/nih-office-of-data-science-strategy

datascience@nih.gov

**Thank you for your
time and attention**

