

National Institutes of Health





Patient Partnerships



EHRs



Technologies



Genomics



Data Science

Where Biomedical Data is Today



Developing an NIH Strategic Plan for Data Science

Requested by Congress

Data resource ecosystem
and infrastructure
modernization

Data sharing, access,
and interoperability

EHR, clinical, and
observational data
availability enhancements



All while ensuring data confidentiality

Slide 4

TF12

Image Source: <https://insider.cit.nih.gov/about/who-we-are>

Todd Fernley, 13/12/2018

Strategic Plan for Data Science: Goals and Objectives

Data Infrastructure	Modernized Data Ecosystem	Data Management, Analytics, and Tools	Workforce Development	Stewardship and Sustainability
<ul style="list-style-type: none">•Optimize data storage and security•Connect NIH data systems	<ul style="list-style-type: none">•Modernize data repository ecosystem•Support storage and sharing of individual datasets•Better integrate clinical and observational data into biomedical data science	<ul style="list-style-type: none">•Support useful, generalizable, and accessible tools and workflows•Broaden utility of and access to specialized tools•Improve discovery and cataloging resources	<ul style="list-style-type: none">•Enhance the NIH data-science workforce•Expand the national research workforce•Engage a broader community	<ul style="list-style-type: none">•Develop policies for a FAIR data ecosystem•Enhance stewardship

Making Data FAIR

Findable

Accessible

Interoperable

Reusable

- To be **Findable**, data must have unique identifiers, effectively labeling it within searchable resources.
- To be **Accessible**, data must be easily retrievable via open systems and effective and secure authentication and authorization procedures.
- To be **Interoperable**, data should “use and speak the same language” via use of standardized vocabularies.
- To be **Reusable**, data must be adequately described to a new user, have clear information about data-usage licenses, and have a traceable “owner’s manual,” or provenance.

New: Office of Data Science Strategy

The NIH **Office of Data Science Strategy** in the Office of the Director:

- Provides leadership and coordination on the strategic plan for data science.
- Helps develop and implement NIH's vision for a **modernized** and **integrated** biomedical data ecosystem.
- Develops of a diverse and talented data science workforce.
- In coordination with the CIO, builds strategic partnerships to develop and disseminate advanced technologies and methods.

Implementing the Strategic Plan

I. Overarching Goals

i. Strategic Objectives

1. Implementation Tactics

a. Milestones and Metrics

Trans NIH Scientific Data Council
Chief Data Strategist

Office of Data Science Strategy
Implementation Teams



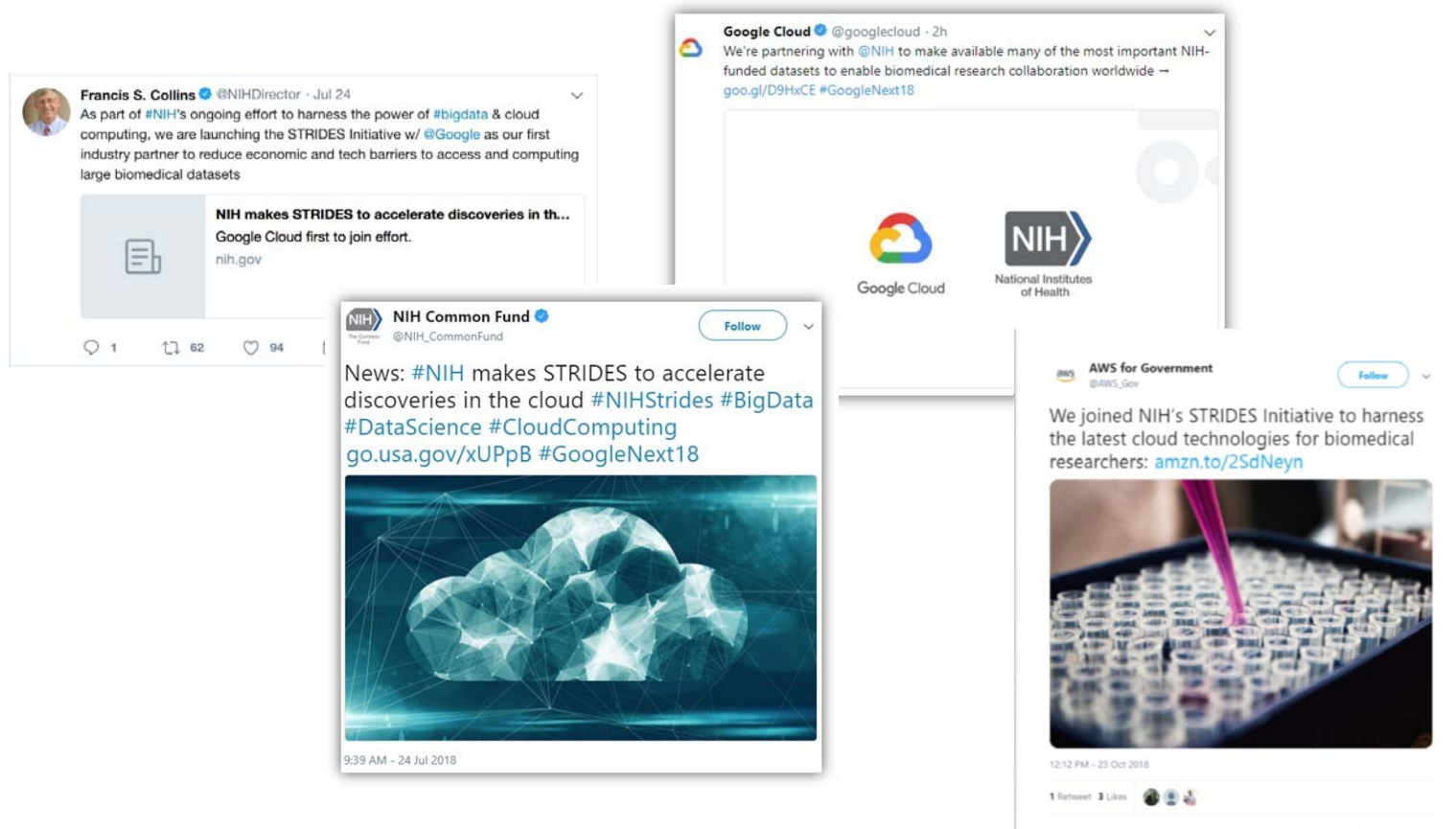
The STRIDES Initiative: Maximizing Research Data



The STRIDES Initiative allows NIH and NIH-funded researchers to take advantage of state-of-the-art data storage and computational capabilities, tools, and expertise.

The STRIDES Initiative: A Series of Partnerships

Launched in 2018 with Google Cloud and Amazon Web Services



The STRIDES Initiative: Objectives for Year One

- Make an initial collection of high-value datasets accessible on cloud
- Establish processes to operationalize STRIDES for use by NIH Institutes and Centers, as well as by NIH-supported researchers (informed by pilot activities)
- Develop communications/onboarding plans for broader community roll-out
- Plan for and begin providing cloud training to NIH researchers
- Identify and test possible areas of strategic collaboration with partners
- Continue to establish additional partnerships to enrich the ecosystem

STRIDES Representative Milestones:

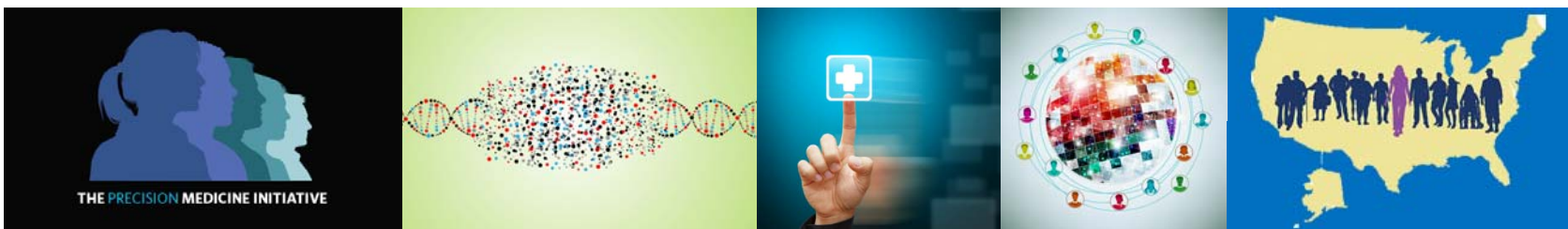
- ✓ 1. Establish partnerships with commercial cloud providers
- ⌄ 2. Enable NIH ICs to leverage the STRIDES vehicle using OT Authority
- ⌄ 3. Move and/or apply discounts for high-priority data sets
- ⌄ 4. Establish communication, outreach, and processes for ICs and universities to engage
- ⌄ 5. Develop technical and administrative frameworks/services to support NIH-managed data assets
- ⦿ 6. Coordinate broad-based training for the research community
- ⦿ 7. Adopt processes and approaches for general researcher-centric use of STRIDES

“All of Us Program” Overview

- A study launched by NIH in support of the Precision Medicine Initiative
- Will engage at least one million volunteers living in the U.S. to provide genetic data, biological samples, and other health information over many years.
- Will serve as a national research resource to inform thousands of studies, covering a wide variety of health conditions that researchers will use to learn more about how individual differences in lifestyle, environment, and biological makeup can influence health and disease.

Goals:

- Better predict disease risk
- Better understand how disease occurs
- Find improved diagnosis and treatment strategies



All of Us
RESEARCH PROGRAM

**The future
of health
begins
with you**

JoinAllofUs.org

NIH
National Institutes of Health

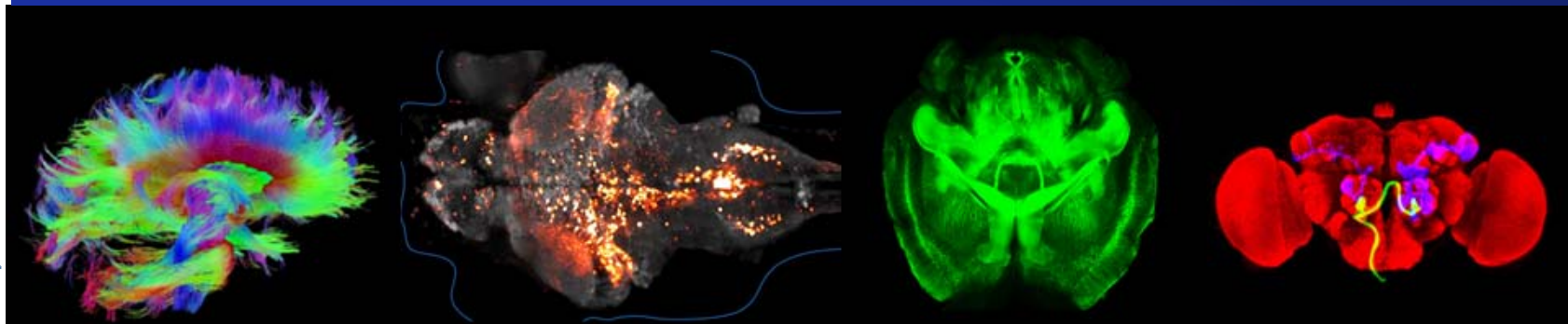
The advertisement features a central collage of diverse individuals. At the top left is a large portrait of a smiling Black woman with curly hair and a nose ring, wearing a red and black plaid shirt. To her right are two smaller portraits: a woman with short dark hair and a man with short dark hair. Below the large portrait are three more portraits: a woman with glasses and a red top, an elderly man with white hair in a green suit, and a woman with curly blonde hair in a white lab coat. The bottom right corner is a dark blue rectangle containing the website URL and the NIH logo.

NIH and the U.S. BRAIN Initiative



- A focus on circuits and networks
- Measure the fluctuating electrical and chemical patterns within circuits
- Understand how all of this helps generate our unique thoughts and actions

www.braininitiative.nih.gov



Cancer Moonshot Initiative: Goals

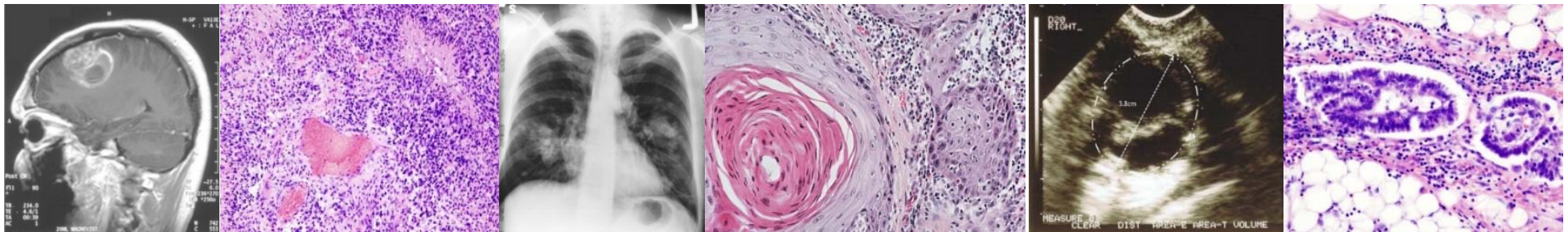
Cooperative endeavor launched in 2016 to:

- Accelerate progress in cancer, including prevention and screening
 - From cutting edge basic research to wider uptake of standard of care
- Encourage greater cooperation and collaboration
 - Within and between academia, government, private sector
- Enhance data sharing

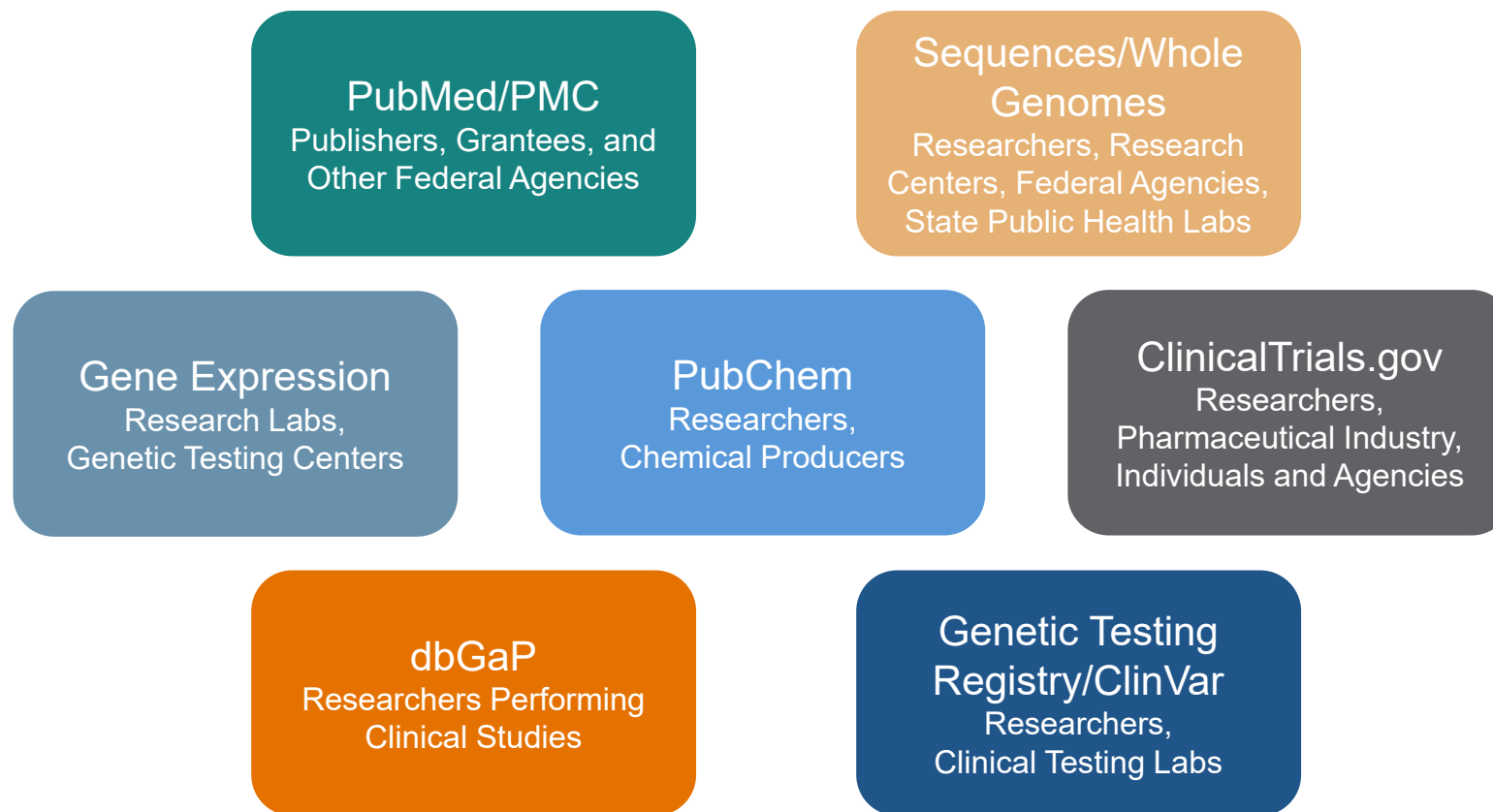


The Cancer Genome Atlas

- Coordinated effort to accelerate understanding of cancer through genome analysis to improve diagnosis, treatment, and prevention
- Provides analysis of > 20 types of cancer, including
 - Leukemia
 - Brain
 - Breast
 - Lung
 - Colon
 - Ovary
 - Bladder
 - Thyroid



More Than 3,000 Different Groups & Individuals Provide Submissions to the National Library of Medicine Daily



The HEAL Initiative



Helping to End
Addiction Long-term

NIH HEAL INITIATIVE

A trans-NIH effort to address the
opioid crisis through science



PAIN
MANAGEMENT

OVERDOSE
REVERSAL

OPIOID
ADDICTION
TREATMENT



National Institutes of Health

Trans-Omics for Precision Medicine (TOPMed)

Very Large, Diverse Data
(4 PB – 20 PB)

Whole Genomic Sequencing of
up to 150,000 individuals
and other –Omics, Molecular,
Behavioral, Imaging,
Environmental
and Clinical Data

Studies Focused on
Heart , Lung, Blood and Sleep
Disorders



Accelerate systems
medicine and emerging
precision medicine to
predict, prevent, diagnose,
and treat HLBS disorders
based on a patient's unique
genes, environment, and
molecular signatures

Alzheimer's Disease



Overview of Sharing Publication Data

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

https://www.nlm.nih.gov/NIHbmic/nih_data_sharing_repositories.html

Options for scaled implementation for orphan datasets

Datasets up to **2 gigabytes**

PubMed Central

- PMC stores publication-related supplemental materials and datasets directly associated publications. Up to 2 GB.
- Generate Unique Identifiers for the stored supplementary materials and datasets.

Datasets up to **20*gigabytes**

Use of commercial and non-profit repositories

- Assign Unique Identifiers to datasets associated with publications and link to PubMed
- Store and manage datasets associated with publication, up to 20* GB.

High Priority Datasets **petabytes**

STRIDES

- Store and manage large scale, high priority NIH datasets (Partnership with STRIDES)
- Assign Unique Identifiers, implement authentication, authorization & access control

NIH Data Science Fellowships – Happening NOW

Coding it Forward

- Civic Digital Fellowship; ~10 **undergraduate** fellows for 2019
- Student-led non-profit to place tech-savvy students in federal agencies; pipeline to public service for technology students
- 10 week summer program; placement in administrative or funding offices
- ODSS will coordinate central NIH activities so fellows connect with each other on campus

<https://www.codingitforward.com/>

Graduate Data Science Summer Program

- 13-15 **Masters** level interns for summer 2019
- Pilot driven by discussion with local universities consortium
 - UVA, George Mason, George Washington, UMD, University of Delaware/Georgetown, Johns Hopkins
- Open for application for students from any university

https://www.training.nih.gov/data_science_summer

Planning for a Trans-NIH Data Management and Sharing Policy

- NIH solicited community input for policy foundation
 - **October 2018:** Input requested on key provisions for potential data management and sharing policy (NOT-OD-19-014)
 - Two public webinars with ~800 participants (combined)
 - 189 submissions from national and international stakeholders
- Identified need for appropriate infrastructure; policy and implementation to go ‘hand-in-hand’
- **Next steps**
 - Develop draft policy for data management and sharing and related guidance
 - Release draft for community input (target date summer 2019)
 - ***Release final policy by end of 2019 calendar year***

National Institutes of Health

