NIH STRIDES Initiative
Federal Demonstration Partnership

Nick Weber

Program Manager, Cloud Services

Project Lead, NIH STRIDES Initiative

National Institutes of Health (NIH)

Center for Information Technology (CIT)

NIH National Institutes of Health
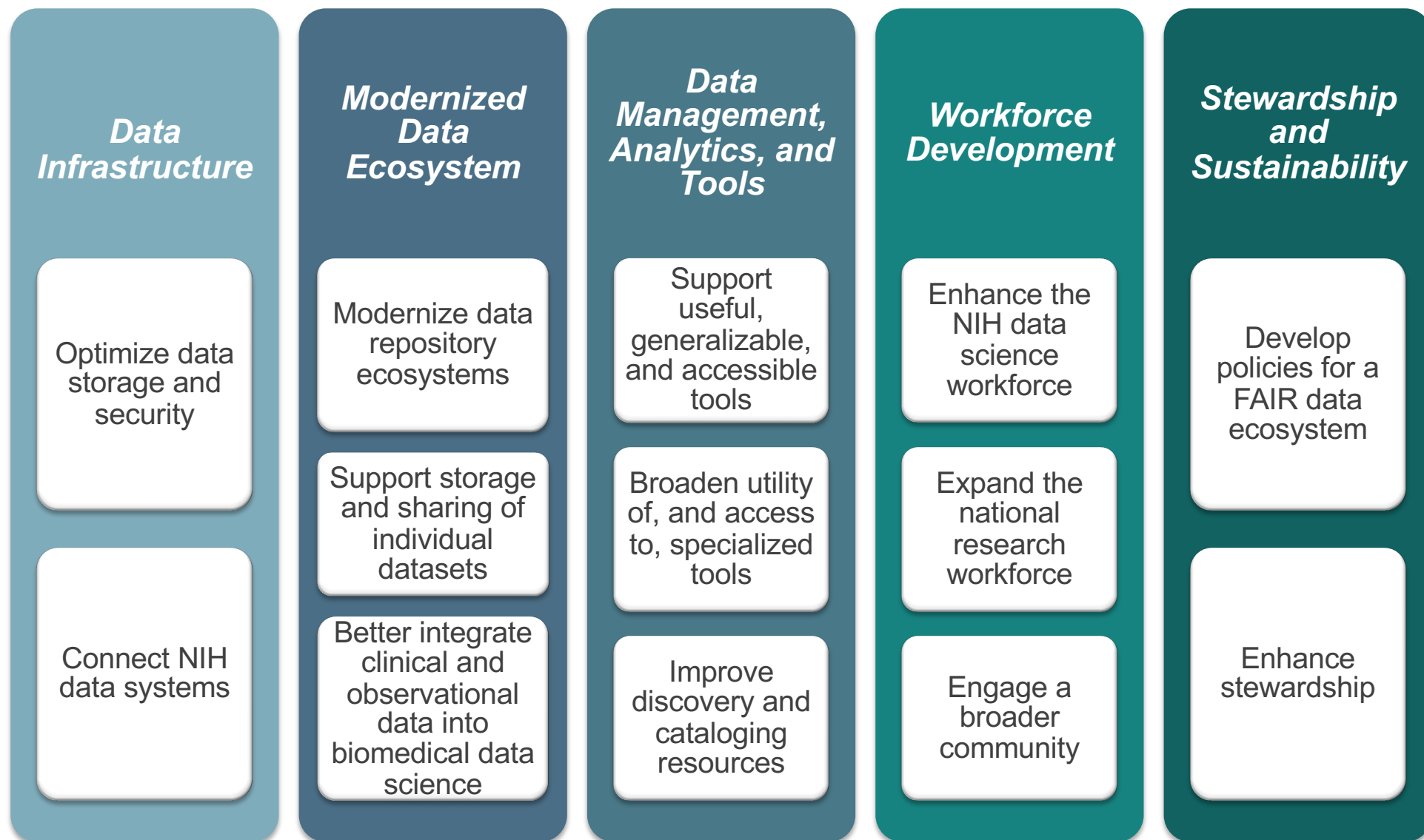
Agenda

- What is STRIDES?

- What have we done so far?

- Where are we headed?

- Questions/Discussion
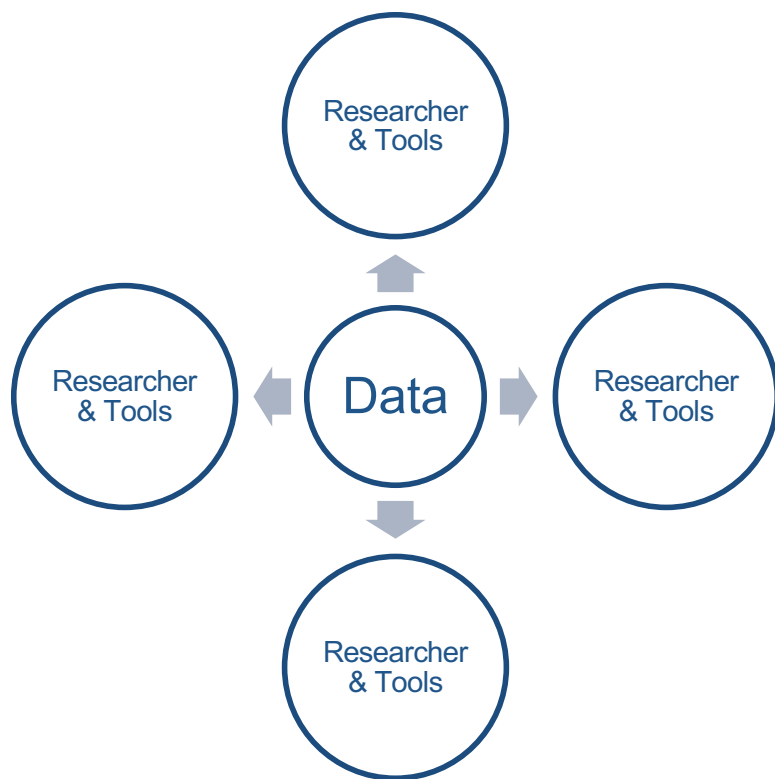
What is **STRIDES?**

Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability

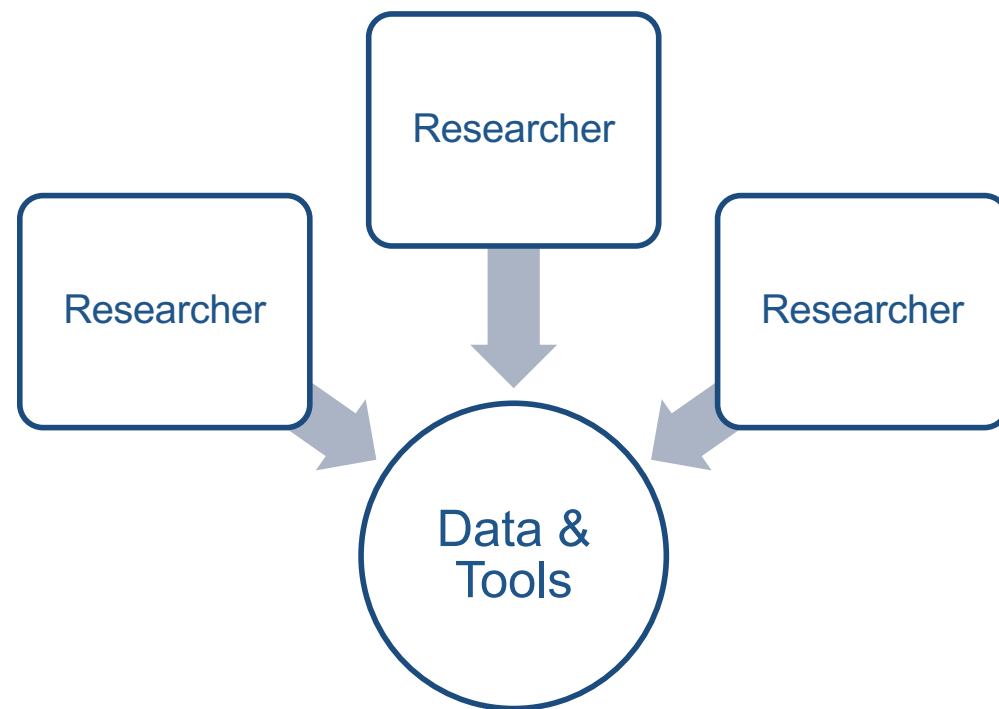# Strategic Plan for Data Science: Goals and Objectives

Data Infrastructure	*Modernized Data Ecosystem*	*Data Management, Analytics, and Tools*	*Workforce Development*	*Stewardship and Sustainability*
Optimize data storage and security	Modernize data repository ecosystems	Support useful, generalizable, and accessible tools	Enhance the NIH data science workforce	Develop policies for a FAIR data ecosystem
Connect NIH data systems	Support storage and sharing of individual datasets	Broaden utility of, and access to, specialized tools	Expand the national research workforce	Enhance stewardship
	Better integrate clinical and observational data into biomedical data science	Improve discovery and cataloging resources	Engage a broader community	

A Shift in the Research Model

Traditional Model

Cloud Model

National Institutes of Health

Making Data FAIR

Findable
Data must have unique identifiers, effectively labeling it within searchable resources

Accessible
Data must be easily retrievable through open systems, and require effective and secure authentication and authorization procedures

Interoperable
Data should "use and speak the same language" through the use of standardized vocabularies

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

National Institutes of Health

Use of Cloud Across NIH is Greatly Increasing

NIH programs are using / planning to use the cloud to store and compute on data

- Supports increasing size and complexity of data
- Has robust compute and analytical tools that are constantly evolving
- Provides the ability to share information among geographically distributed groups
- Allows researchers to focus on what they do best!

BUT… using the cloud doesn't address all research challenges

- The way the data are stored and managed is often unique to each NIH program
- Not enough attention is paid to data organization, structure, access, utility, findability, and reusability
- Data is often a byproduct, whereas end goal is scientific findings, journal articles
- *Result is a reduced ability to use/reuse the data, both within and across programs*

NIH National Institutes of Health

STRIDES Overview

○ A series of public-private sector relationships enabling both NIH and NIH-funded researchers to obtain cloud-based data storage and computational capabilities at cost-effective rates

○ Aimed at significantly lowering the barriers to entry for accessing and computing against biomedical research data

– Administrative, technical, financial obstacles

○ Includes mechanisms to explore unique opportunities to collaborate with commercial partners on the development of new ways to access and use high-value research data

– Training, consultation, and collaboration components

– Access to innovative technologies such as artificial intelligence and machine learning

NIH National Institutes of Health

STRIDES Partnerships

- Agreements with Google Cloud (July 2018) and Amazon Web Services (Oct. 2018)

- Other Transaction mechanism used

- Additional partnerships anticipated

> "NIH must weave its existing data-science efforts into the larger data ecosystem and fully intends to take advantage of … the commercial sector through a variety of innovative public-private partnerships."
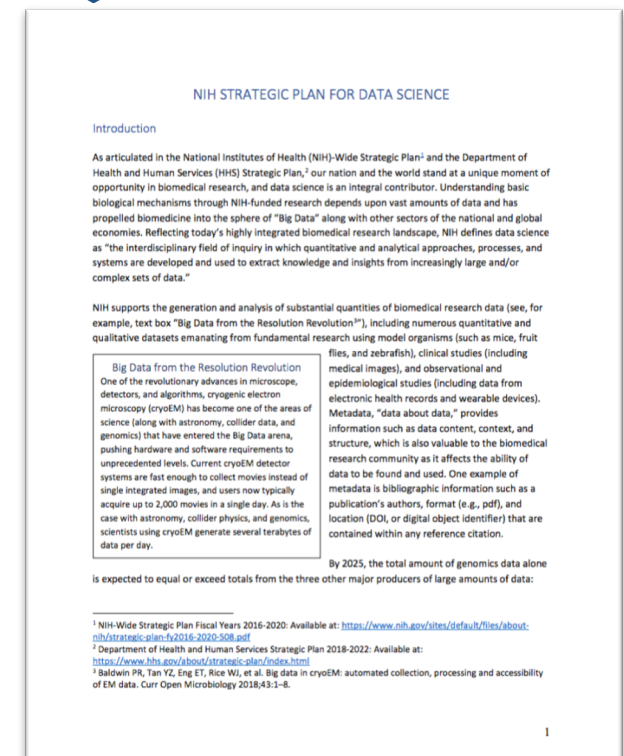
Shared Goals Among NIH, Research Institutions, Cloud Providers

Access

Large, rapidly expanding biomedical research datasets

Collaboration

Need for scalable collaboration among researchers

Innovation

Data localization; difficulty sharing datasets with others

All While Ensuring Data Integrity and Confidentiality

National Institutes of Health

STRIDES Benefits to Research Programs

Cost Discounts

Significant savings on full catalog of services, including compute, storage, and analytics
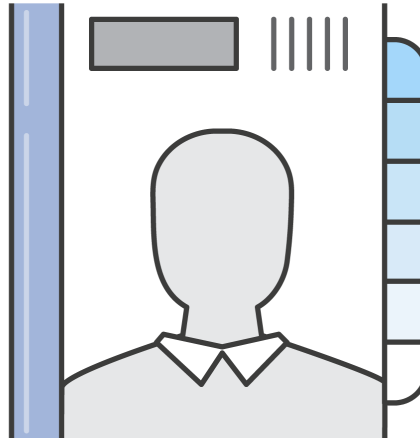
Professional Services

Range of engagements, from consultations to custom-scoped collaborative development efforts
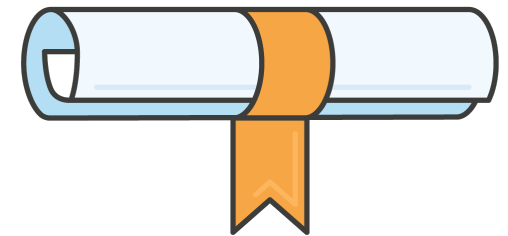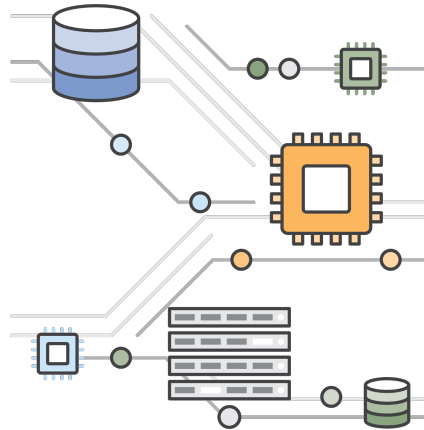
Training

Inclusive of standard introductory content as well as customized training for biomedicine (in-person and online)

NIH National Institutes of Health

STRIDES Benefits to Government
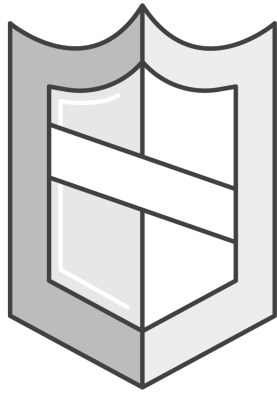
Additional Data Protections

More options to secure data and systems, using modern cyber-security capabilities

Ecosystem Development

Connecting data sets, tools, resources, and researchers in new ways

Reporting Insight

New insights into cost and usage of data sets and resources, to inform sustainability efforts

NIH National Institutes of Health

What have we done so far?

Accomplishments to-date

National Institutes of Health

Established STRIDES

Goal

Establish STRIDES, execute partner agreements, and provide NIH Institutes/Centers/Offices (ICOs) with a mechanism to access cloud services

Results

- Partnerships established with Google Cloud and Amazon Web Services (AWS)
- Other Transaction Authority (OTA) extended to all 27 ICOs

Examples:
– Common Fund
– NHLBI
– NLM/NCBI
– NCI

National Institutes of Health

Established Initial Guidelines and Processes

Goal

Provide guidelines and processes to efficiently support the adoption of STRIDES for NIH and NIH-funded researchers

Results

- Established processes for onboarding, setting up accounts, and billing
- Developed and published general STRIDES content to aid the decision-making process for active and potential participants

National Institutes of Health

Provided Training

Goal

Provide training to researchers that use STRIDES to access cloud resources

Results

- Held 8 training sessions with a total of 129 NIH attendees
  - 3-5 day, multi-modal (in-person and online trainings)
- Training course examples:
  - Architecting with Google Cloud platform
  - GCP Bootcamp for Big Data Summer Fellows
  - AWS Bootcamp for Big Data Summer Fellows
  - AWS Cybersecurity Overview

First set of trainings sold out in 1 hour!
16 spots available; 125 people interested

National Institutes of Health

Facilitated Access to Professional Services

Goal

Provide professional engineering services to those using STRIDES

Results

- Professional services include enterprise-level customer support, dedicated account managers, and on-site engagements
- TOPMed, University of Washington results from professional services—**10x speedup in their data processing pipeline** based on engineering support

> I am **very grateful** to have **this level of access to talented engineers** from Google to help the team, and **I'm excited to see what other possibilities come** from this model when deployed with other groups!
>
> ~Alastair Thomson, NHLBI CIO

Where are we headed?

Ideas for FY2020 and beyond

NIH National Institutes of Health

Onboarding/Enrollment Options

NIH IC or Program
Examples: NHLBI, Common Fund

Research Institutions
Example: Johns Hopkins University

NIH IC funds STRIDES for specific program/data resource
- o Determines who can leverage discounted cloud services
- o Determines limits/parameters

Institution establishes own "STRIDES agreement" that leverages NIH pricing and discounts
- o Determines who can leverage discounted cloud services
- o Determines limits/parameters

NIH IC enrolls researchers and establishes accounts

Institution enrolls NIH-funded researchers and establishes accounts

STRIDES team handles invoicing and billing
- o NIH IC pays for cloud services using appropriated funds
- o STRIDES team provides overall usage reporting to funding NIH IC

Institution handles invoicing and billing
- o Institution pays for cloud services using NIH award funds
- o STRIDES team provides overall usage reporting to funding NIH IC

National Institutes of Health

Onboard Research and Academic Institutions

Objective

Onboard any NIH-funded institution that wants to use STRIDES

Ideas

- National campaign to enroll NIH-funded institutions using established NIH channels
  - Established Office of Extramural Research (OER) communication channels (e.g., "Open Mike" blog, NIH Extramural Nexus Newsletter, etc.)
  - Research administrative venues (e.g., American Association of Medical Colleges, Society of Research Administrators, etc.)
- Provide guidance to program and grants management officials on incorporating agreement terms into both new and existing awards
- Different approaches for different targeted institutions (e.g., HBCUs, Tribal Colleges, top computational institutions)

Continue to Onboard NIH Extramural Programs

Objective

Continue to onboard any NIH extramural programs and researchers that want to use STRIDES

Ideas

- Announce campaign to onboard NIH research programs using established NIH channels
    - Provide information in specific Funding Opportunity Announcements, on Office of Data Science Strategy website, and other communication channels
    - Share examples of success stories on funding high-value research in the cloud (intramural example)
    - Target cloud-savvy extramural programs to transition to STRIDES

Provide Training for Researchers

Objective

Jumpstart training for all researchers who engage with STRIDES

Ideas

- Potential ideas for offering training:
  - Conference/regional-based trainings
  - Training center grants
  - Piggyback on other NIH data science training programs
  - Encourage training as part of institutional onboarding with STRIDES
  - Focus on cloud-specific trainings and/or include data science best practices (e.g., FAIR)

Envisioning a Future of Interconnected Data Sets

...and many others!
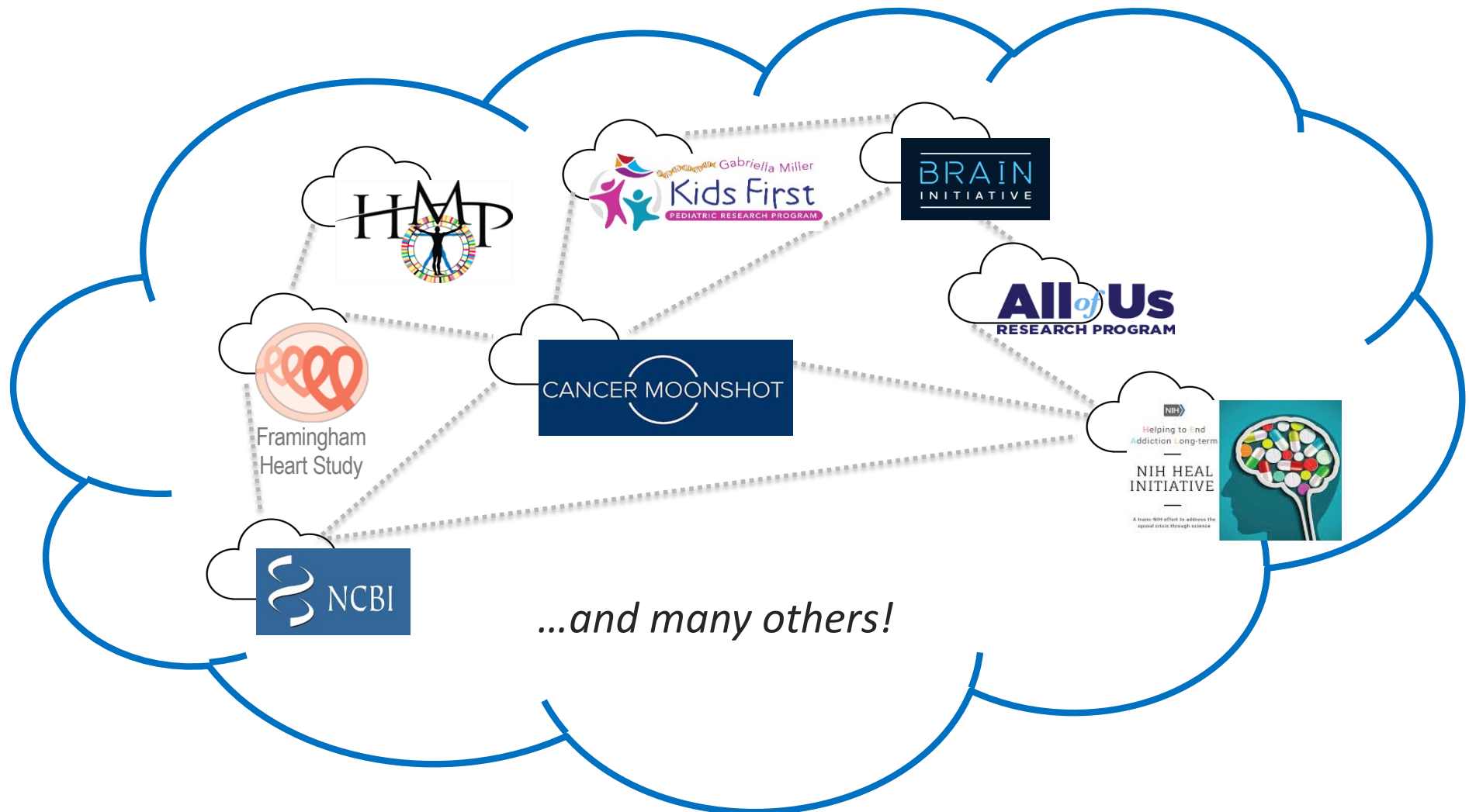
Acknowledgements & Contact Information

| Accounts Manager Todd Reilly, PhD | AWS Accounts Nigel Horne | STRIDES Program Manager Nick Weber | Product Manager Sherika Wynter | Business Analyst James Davis | Science Policy Fellow Valerie Virta, PhD |

| Google Accounts Tom Shaw, PMP | Operations Matt Gieseke | | Communications Specialist Michelle Speir | Cloud Architect Antej Nuhanović | Cloud Architect Joel Peterson |

For questions, please contact:
strides@nih.gov

- Oversight & Policy: Andrea Norris, Larry Tabak, Jim Anderson, Betsy Wilder, Susan Gregurick, Jess Mazerik, Taylor Gilliland, Belinda Seto, Michelle Bulls, Adam Graham

- Negotiations & Awards Management: Kate O'Sullivan, Jeff Snyder, Teresa Marquette, Ann Gawalt, Chris Hammond, Kristin Wegner

- Technology Advice: Vivien Bonazzi, Tony Kerlavage, Jim Ostell, Alastair Thomson

- Communications: Jennifer Morgan Gray, Rachell Britt, Alexis Williams, Sisley Chung

- STRIDES Partners: Google, AWS, Carahsoft, Four Points Technology

- *..and many others!*

Questions?

Contact:

Nick Weber, STRIDES Program Lead | nick.weber@nih.gov