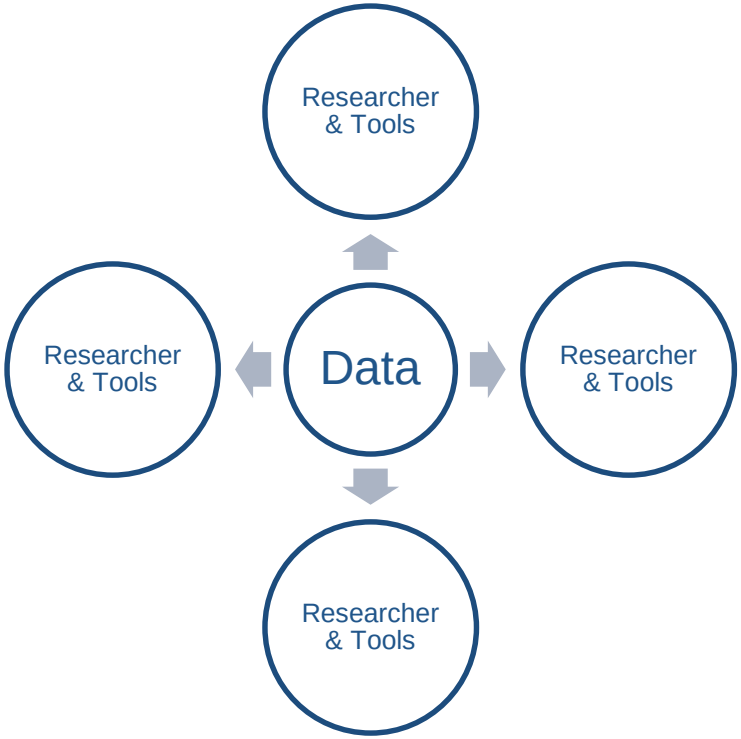
NIH STRIDES Initiative
NASEM – Neuroscience in the Cloud

Nick Weber
Program Manager, Cloud Services Office
Center for Information Technology, National Institutes of Health

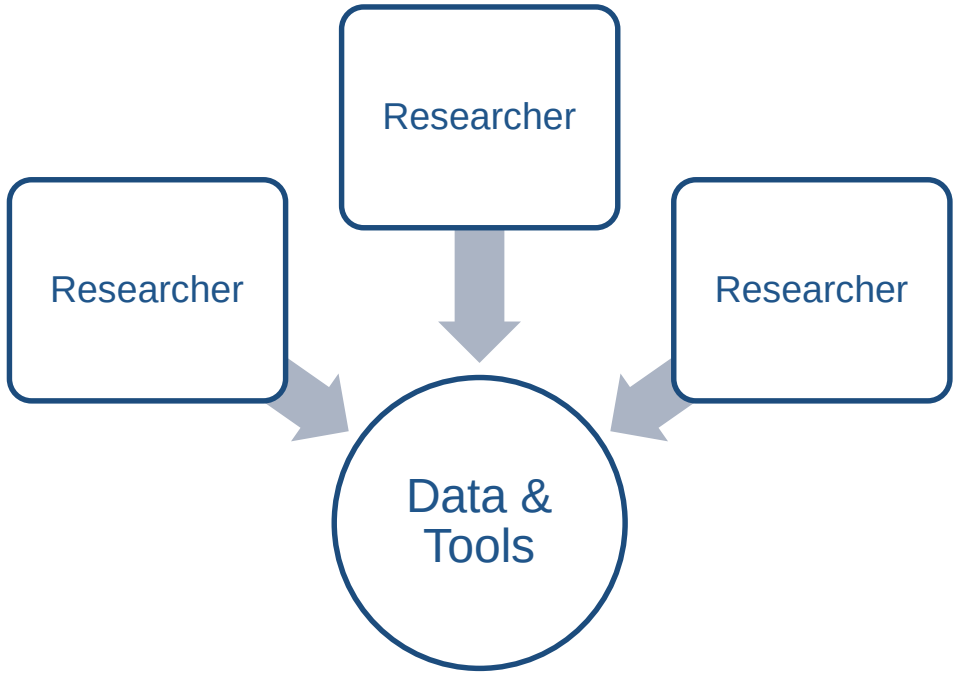
National Institutes of Health

September 24, 2019

A Shift in the Research Model



Traditional Model



Cloud Model

Use of Cloud for Research is Greatly Increasing

- 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/managed is often idiosyncratic to program
- Not enough attention is paid to data organization, structure, access, utility, findability, and overall provenance
- Data is often a byproduct, of research whereas end goal is scientific findings, journal articles
- *Result is a reduced ability to use/reuse the data, both within and across research programs*

Turning Research Data Into Knowledge and Discovery

The Science and Technology Research Infrastructure for Discovery, Experimentation, and Sustainability (STRIDES) Initiative

- State-of-the-art data storage and computational capabilities
- Training and education for researchers
- Innovative technologies such as artificial intelligence and machine learning

Partnerships with aws, Google Cloud, and other commercial providers

Shared Goals Among NIH, Research Institutions, Cloud Providers

Access

Large, rapidly expanding biomedical research datasets

Collaboration

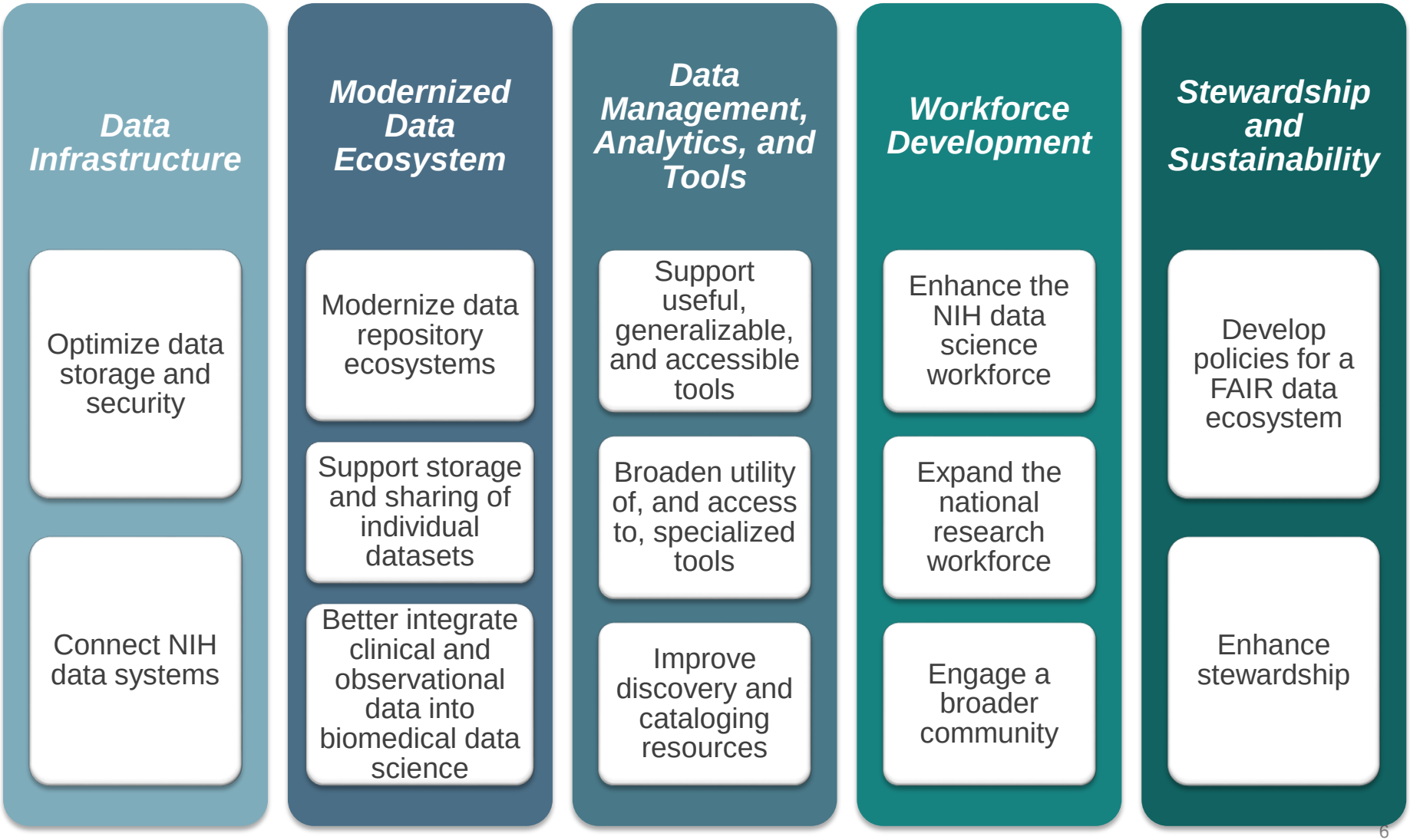
Need for scalable collaboration among researchers

Innovation

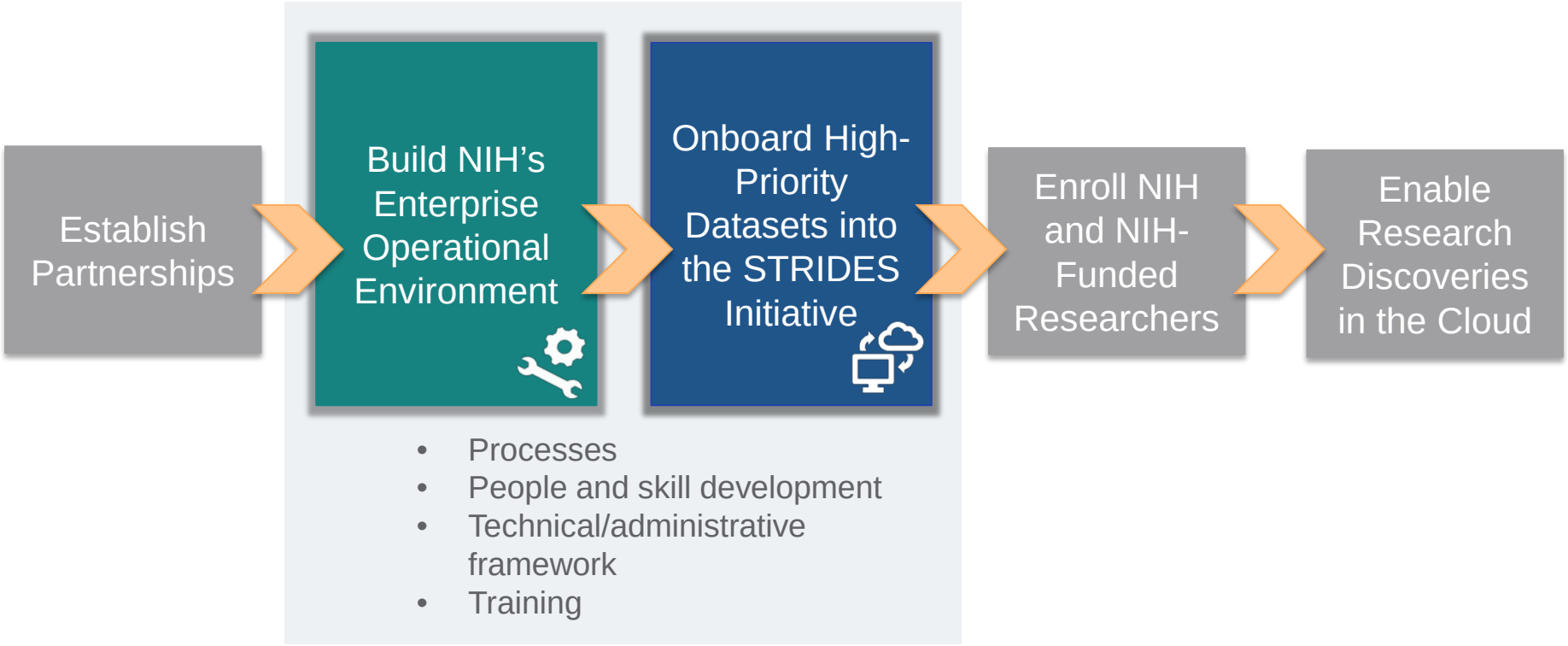
Data localization; new capabilities constantly emerging

All While Ensuring Data Integrity and Confidentiality

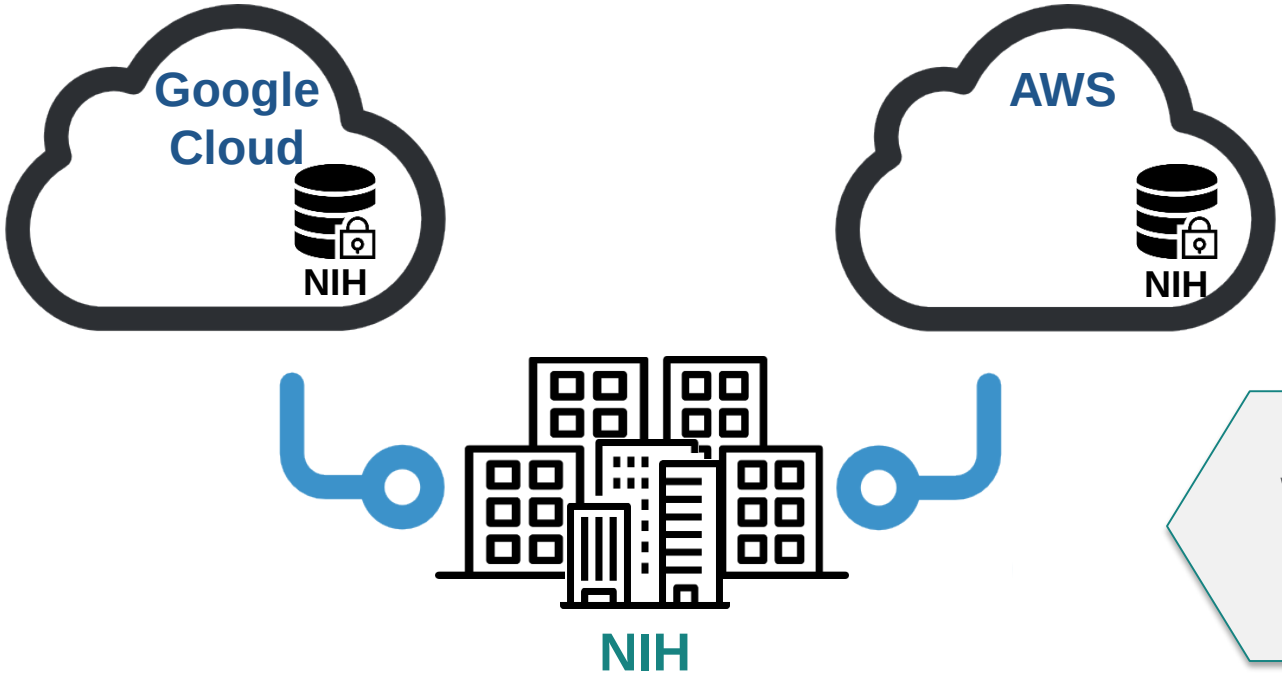
Strategic Plan for Data Science: Goals and Objectives



The STRIDES Initiative Today



Operationalizing Cloud for NIH-Managed Data



WORK IN PROGRESS

- Enhancing network connectivity from NIH to Cloud Service Providers (CSPs)
- Appling cybersecurity controls (Authority To Operate)
- Optimizing cloud environments with assistance from the CSPs
- Leveraged federated login and Identity and Access Management services for a more consistent and maintainable authentication and authorization

Training is Critical

Training Logistics
- Available both in-person and online
- Courses generally full- or multi-day
- Specialized “tracks” and certification paths available

Training Courses
- Architecting with Google Cloud platform
- Design and Process
- Google Cloud Platform Bootcamp for Big Data Summer Fellows
- AWS Bootcamp for Big Data Summer Fellows
- AWS Cybersecurity Overview

Training Audiences
- Research staff
- Data scientists/bioinformaticians
- Technical staff

First set of NIH trainings sold out in 1 hour!

16 spots available; 125 people interested

Professional Services Can Jumpstart Projects

- Professional services include enterprise-level customer support, dedicated account managers, and on-site engagements
- Benefits include greater insight into cost, usage, and consumption patterns, with suggestions for cost, performance, and cybersecurity optimization
- TOPMed, University of Washington results from professional services—**10x speedup in their data processing pipeline** based on engineering support

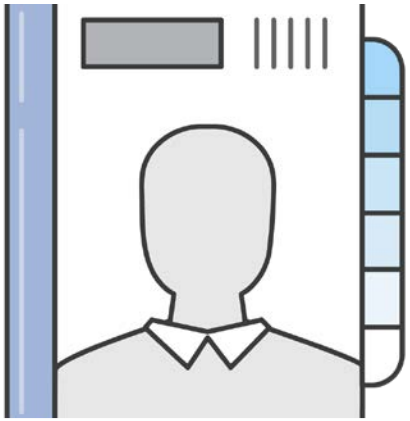
I am **very grateful** to have **this level of access to talented engineers** from Google to help the team, and **I'm excited to see what other possibilities come** from this model when deployed with other groups!

~Alastair Thomson, NHLBI CIO

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

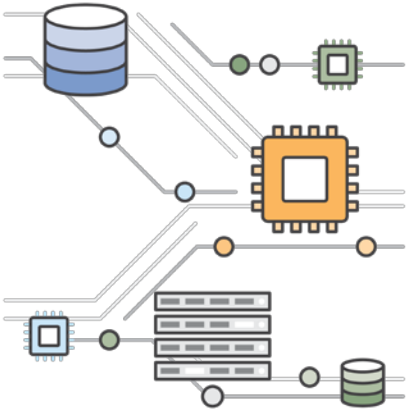
Inclusive of standard introductory content as well as customized training for biomedicine (in-person and online)

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

STRIDES Benefits Already Realized

Cost savings from programs are projected to be multiple millions of dollars per year

15 programs onboarded with more than 30 petabytes of data

8 training sessions offered for 129 NIH attendees; plans for hundreds more!

NIH National Institutes of Health

Onboarding/Enrollment Options

NIH IC or Program

Examples: NHLBI, Common Fund

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

NIH IC enrolls researchers and establishes accounts

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

Research Institutions

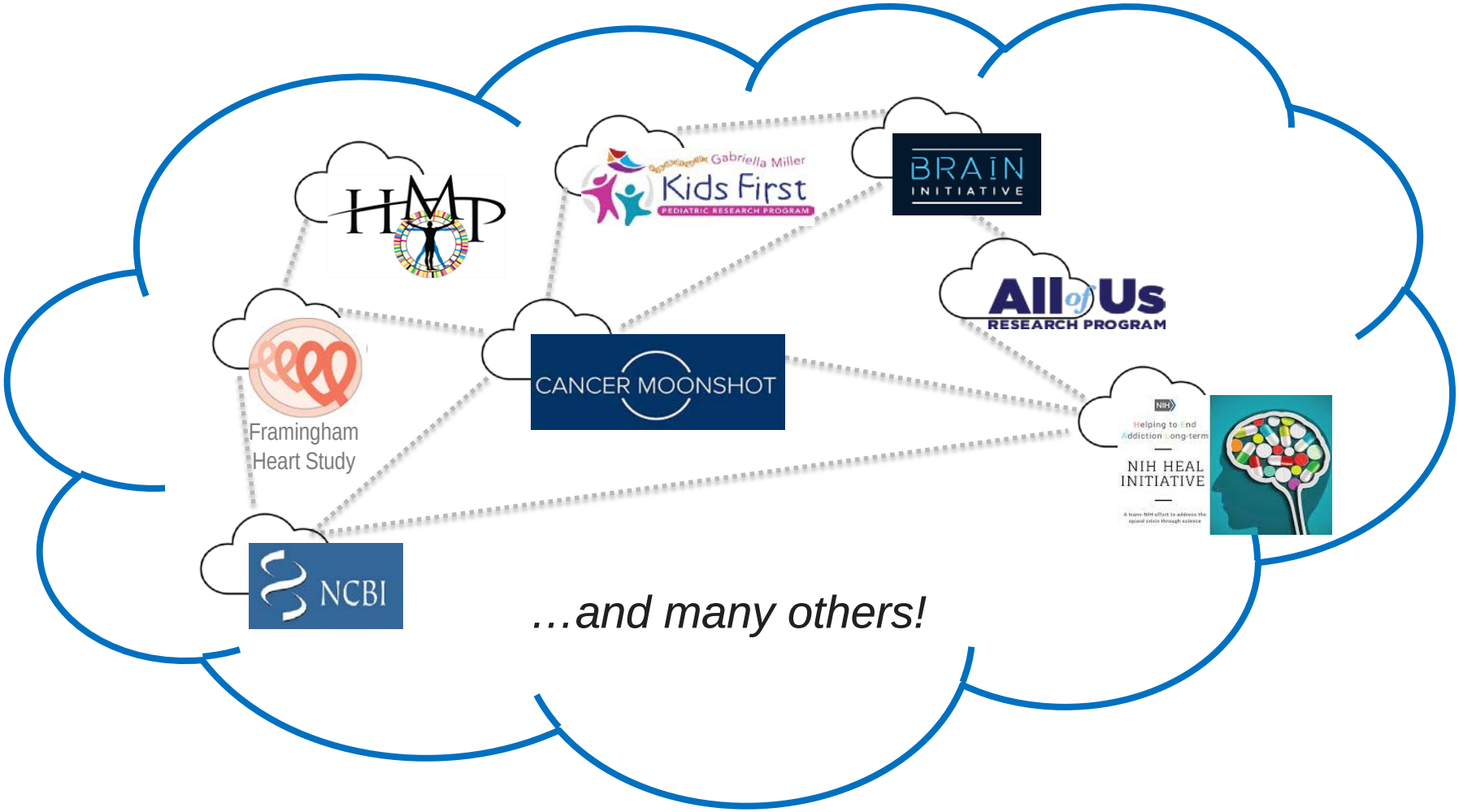
Example: Johns Hopkins University

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

Institution enrolls NIH-funded researchers and establishes accounts

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

Envisioning a Future of Interconnected Data Sets



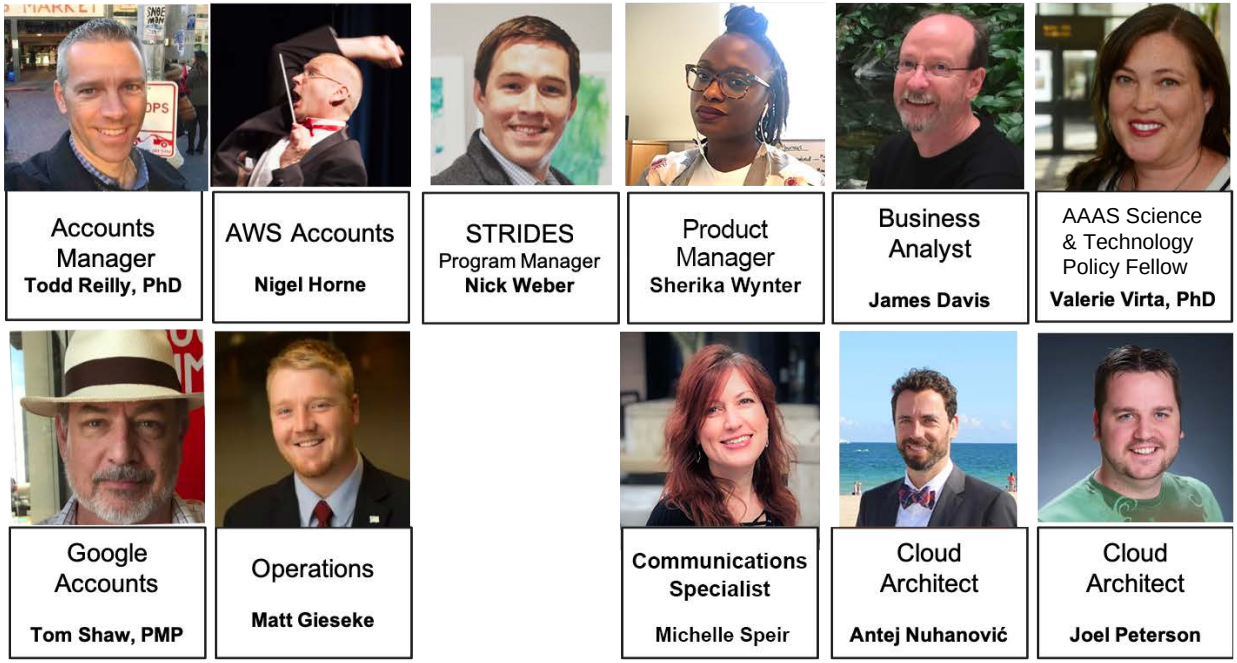
Further Questions?

Contact:
Nick Weber, Program Manager | nick.weber@nih.gov

Website:
https://datascience.nih.gov/strides

NIH National Institutes of Health

Acknowledgements and Group Contact Information



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 Cloud, AWS, Carahsoft, Four Points Technology
- *..and many others!*