National Institutes of Health
Office of Data Science Strategy

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

Susan K. Gregurick, Ph.D.
Associate Director for Data Science
National Institutes of Health

Council of Councils, January 2023

Agenda for Session

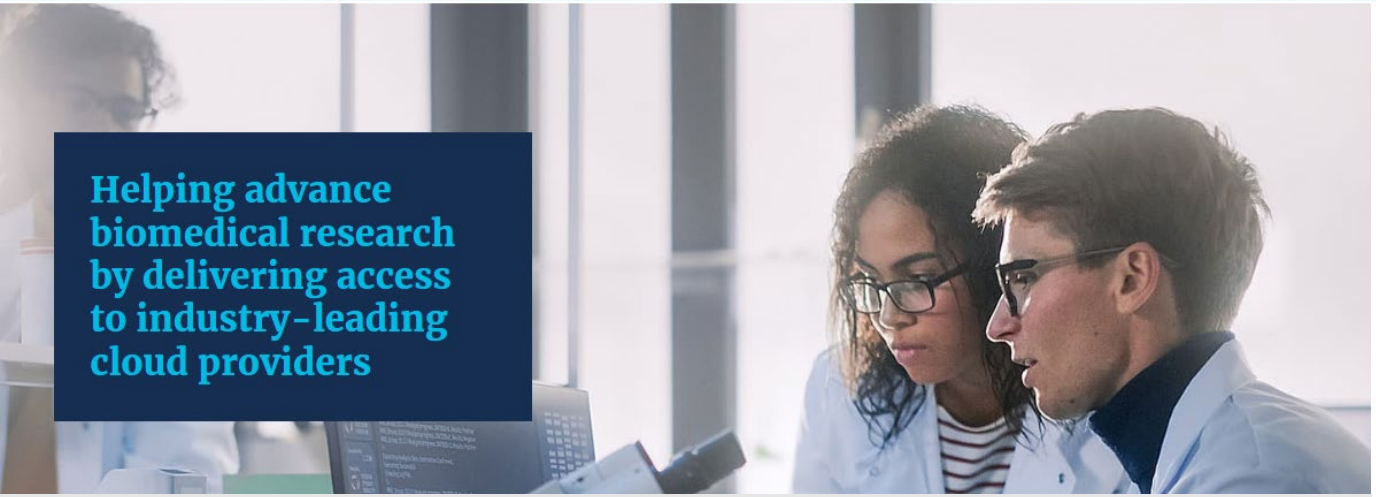
Introduction: Dr. Susan Gregurick, Associate Director for Data Science, NIH

Speaker: Dr. Ashok Krishnamurthy, Deputy Director of the Renaissance Computing Institute (RENCI), and Research Professor of Computer Science at University of North Carolina, Chapel Hill

Council member users:

- Dr. Kristin Ardlie, Director of the GTEx Laboratory Data Analysis and Coordination Center at the Broad Institute of MIT and Harvard
- Dr. Sachin Kheterpal, Associate Professor of Anesthesiology Sachin Kheterpal, M.D., MBA, is the Medical School's associate dean for research information technology, University of Michigan

- **Discussion**: Dr. Kevin Johnson (moderator), David L. Cohen University Professor, University of Pennsylvania



The STRIDES Initiative aims to help NIH and its institutions accelerate biomedical research by reducing barriers in utilizing commercial cloud services. This initiative aims to harness the power of the cloud to accelerate biomedical discovery. NIH and NIH-funded researchers can take advantage of STRIDES benefits.

Benefits:

- Discounts on partner services
- Professional services consultations
- Access to training
- Potential collaborative engagements

>200
Petabytes of Data

>995
NIH & NIH-funded Research Programs/ Projects

>4700
People Trained

274M
Compute Hours

$41M
Cost Savings

https://datascience.nih.gov/strides

Bringing NIH Programs to Cloud

ODSS has funded a three-year pilot STRIDES program aimed to partner NIH-supported extramural and intramural researchers with STRIDES effort

Pilot program provided STRIDES cloud credits, and this year the program included staffing support for intramural researchers

Program has supported a total of 39 investigators, 23 intramural investigators and 16 extramural investigators. Total support was $3.8M (FY20 - FY22)

Bringing NIH Programs to Cloud (continued)

Projects included:

- Telomere-to-Telomere Consortium Analyses on the NHGRI AnVIL/AnVIL
- Cloud migration of data and data analysis platform of The Environmental Determinants of Diabetes in The Young Study (TEDDY)
- NINDS Medical Image Analysis Program
- NCATS Rare Disease Alert System
- Clinical Center Small Bowel Segmentation Program
- NEI & NIBIB Cloud Computing for Optical Image Restoration and Intramural Training

Challenges investigators encountered:

- Lack of experience in cloud computing
- Challenges in setting up cloud environments
- Uncertainty of cloud costs, either overestimations or underestimations

Addressing Barriers to Cloud

A cloud testbed allowing researchers to experiment

Primary Cloud Lab Use Cases

Exploring the Cloud Consoles
Researchers can gain an understanding of the look and feel of cloud environments before they jump into a full STRIDES account for research

Supplementing Cloud Training
Researchers can use the sandbox to strengthen their understanding of cloud training or follow along with training content in a separate environment.

Experimenting with Simple Cloud Solutions
Researchers interested in solutions for specific scientific tasks can use the sandbox to build proof of concept or other simple solutions to understand LOE and other details for production.

Benchmarking Costs
Testing out different tools and configurations (instance types, sizes, etc.) to optimize research analyses