

Building a Strong and Diverse Data Science Community

Alison Lin, PhD

*Lead, Training, Workforce Initiatives and Community
Engagement (TWICE)*

NIH Office of Data Science Strategy

February 28, 2024

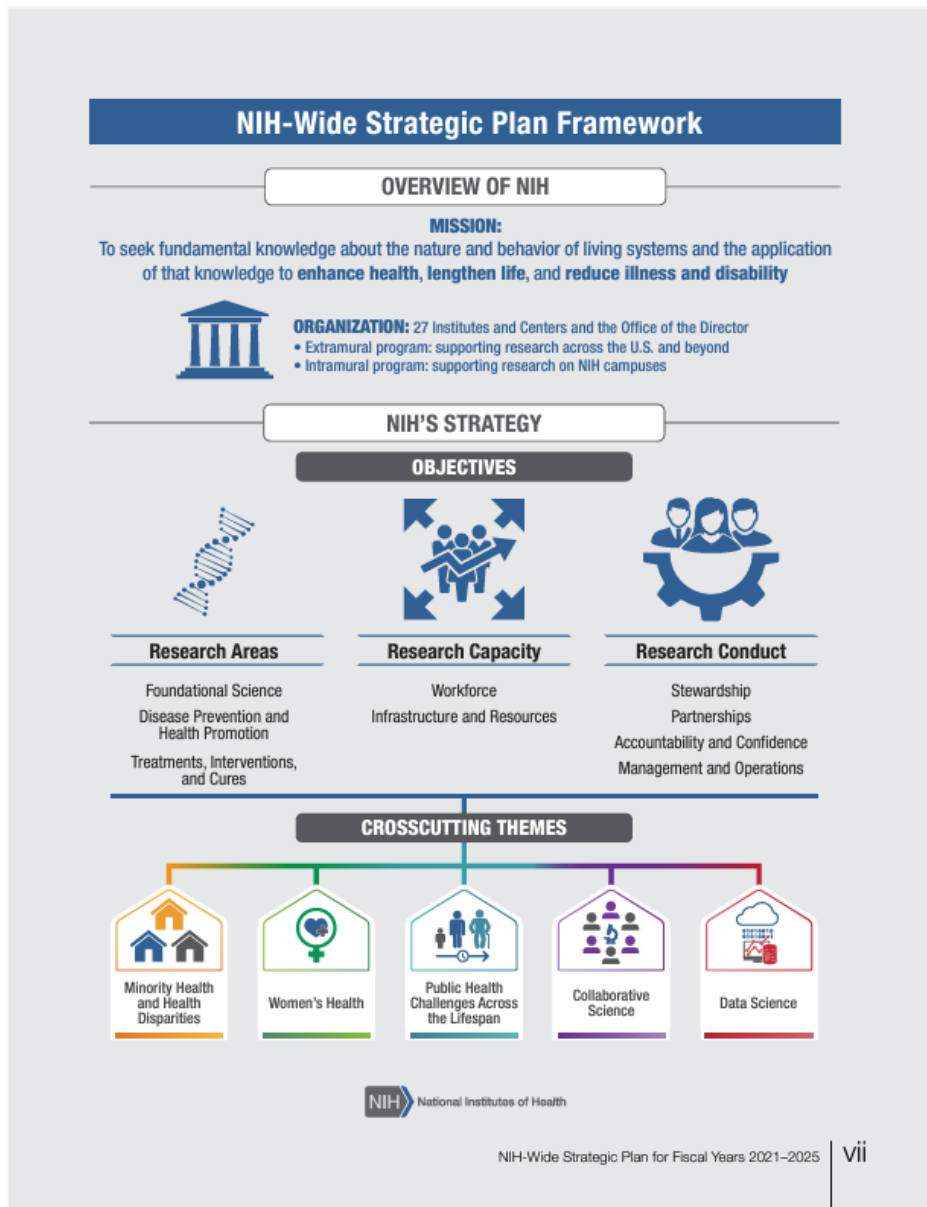
NIH Mission



To seek fundamental knowledge about the nature and behavior of living systems and the application of that knowledge to enhance health, lengthen life, and reduce illness and disability.

NIH-Wide Strategic Plan for Fiscal Years 2021-2025

Data Science is a crosscutting theme at NIH



The 27 NIH Institutes and Centers

Data Science Topic Areas

Examples of data science areas include, but are not limited to:

- artificial intelligence;
- predictive analytics;
- machine learning;
- bioinformatics;
- cloud computing;
- computational science;
- software design and programming;
- supercomputing;
- statistics;
- clinical informatics;
- data visualization;
- modeling and simulation;
- data sharing and access;
- data management;
- data compression and standards;
- other data science topics

All data science training efforts are strongly encouraged to include activities that enhance awareness, knowledge and communication of:

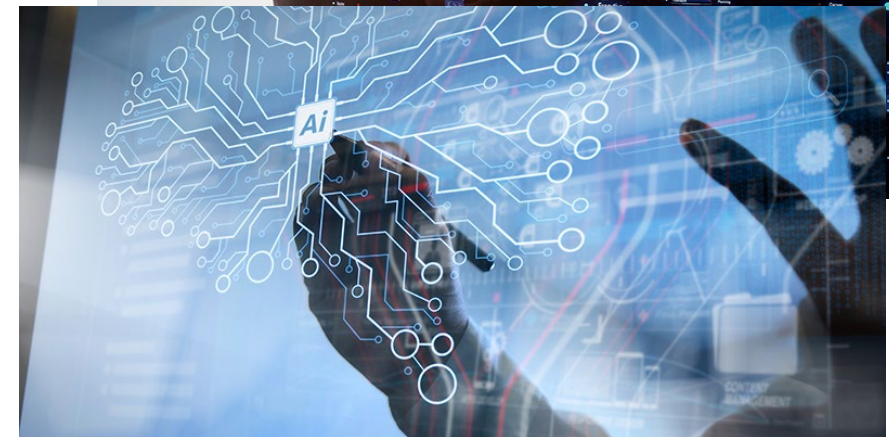
- data ethics;
- risk management of cybersecurity.



Mission and Goals of ODSS

The NIH Office of Data Science Strategy:

- Provides leadership and coordination on the strategic plan for data science
- Develops and implement NIH's vision for a modernized and integrated biomedical data ecosystem
- **Builds a strong and diverse data science community**
- Builds strategic partnerships to develop and disseminate advanced technologies and methods



Health Research Needs a Strong Data Science Community

In health research:

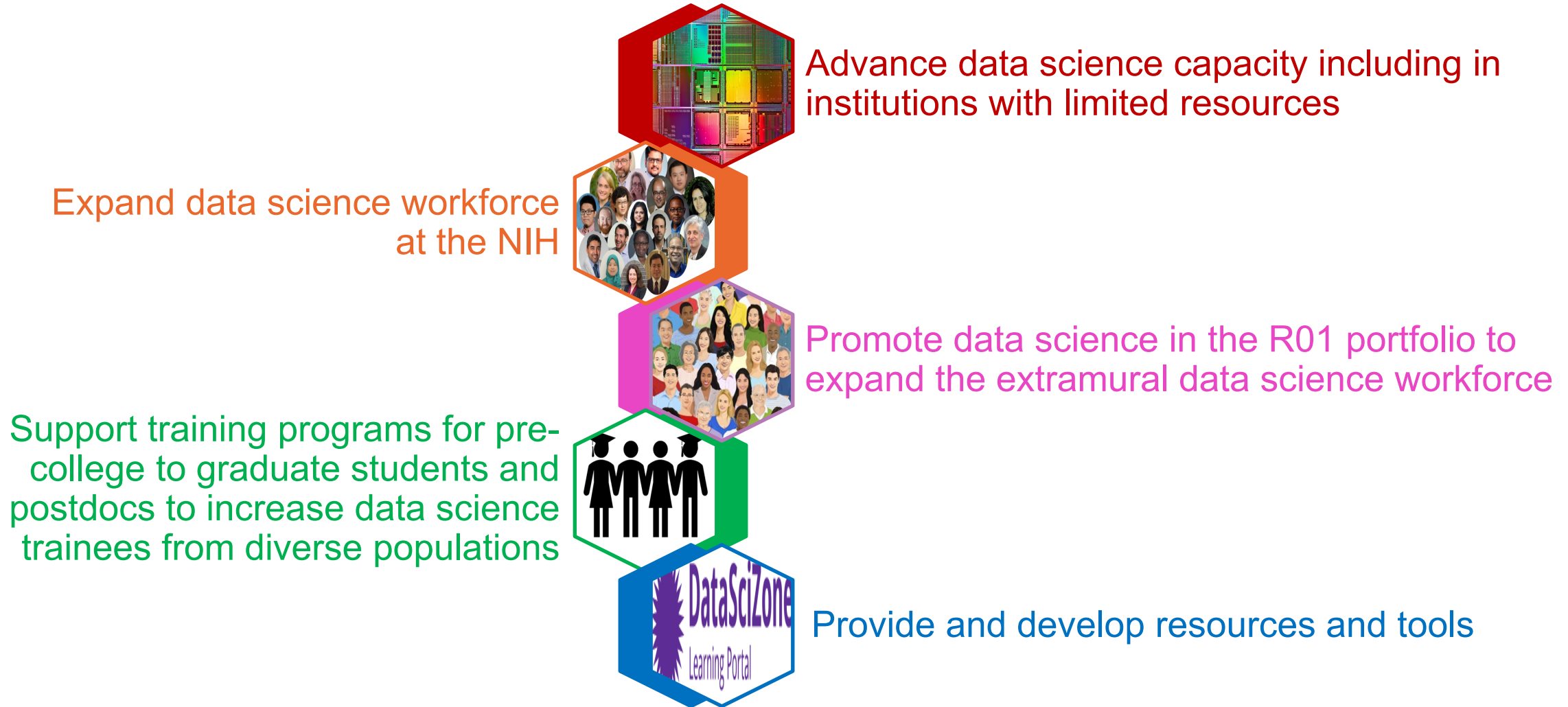
- Data science literate
 - Not intimidated by data science
 - Can read and understand reported outcomes resulting from data science approaches
 - Know where to find relevant resources
- Data science savvy – data science literate and
 - Will actively use data sciences approaches in research projects
 - Can initiate and/or participate in collaborations with data scientists
- Data scientist
 - Has skills and expertise in bioinformatics, artificial intelligence, clinical informatics, cloud computing, statistics, computational science, software design and programming, bioinformatics, visualization, machine learning, predictive analytics, supercomputing, modeling and simulation, digital health, data sharing and access, data management, and/or other data science areas
 - Can communicate what they learn and creatively display the information
 - Can formulate implications and implement follow up studies

“Alone we can do so little; together we can do so much.”

- Helen Keller

Training, Workforce Initiatives and Community Engagement (TWICE)

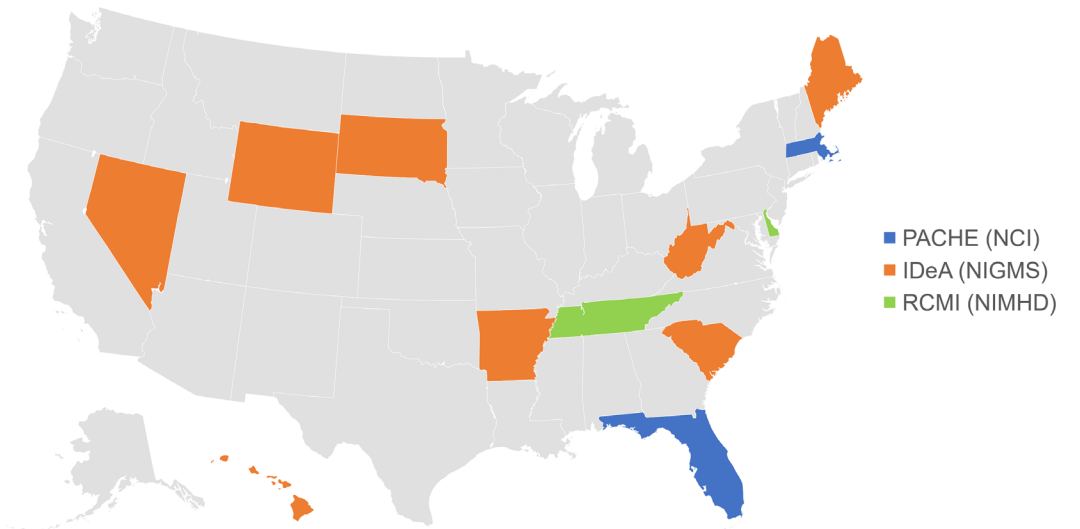
Our priority is to serve you:



NOT-OD-23-123: Supplement to Promote Capacity Building

- Enhance institutional data science capacity through:
 - Growing human capital with data science competencies.
 - Developing or expanding institutional infrastructure for conducting data science-relevant research.
 - Building data science partnerships.
- Eligible parent awards:
 - NIMHD: Research Centers in Minority Institutions (RCMI).
 - NIGMS: Institutional Development Award (IDeA).
 - NCI: Partnerships to Advance Cancer Health Equity (PACHE).
- <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-23-123.html>

Locations of Awards for Administrative Supplements to Enhance Institutional Data Science Capacity (NOT-OD-23-123)



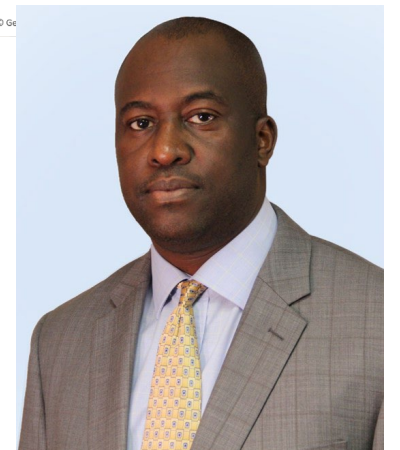
Next Submission Date:

April 1, 2024

ODSS Contact:

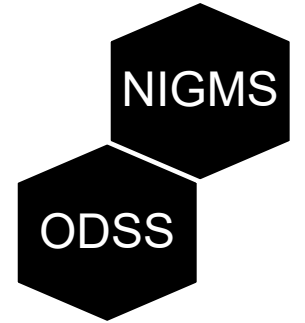
Dr. Raphael Isokpehi

ds-workforce@nih.gov



Co-funding Support for Science Education Partnership Awards (SEPA)

The NIH SEPA program supports STEM and Informal Science Education activities for pre-kindergarten to grade 12 (P–12) students from diverse backgrounds. ODSS supports co-funding of SEPA programs to help grow a pool of well-prepared young students in data science.

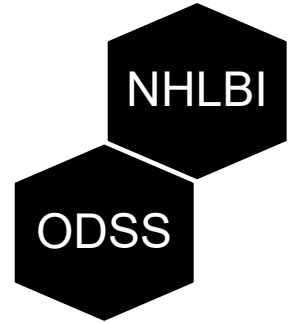


- The Knox Scholars Data Science Research Program from ***Health Resources in Action, Inc.*** supports high school students from the Boston area who are underrepresented in STEM fields (Black and Latinx youth, first-generation college students, low-income students). Students are provided internship opportunities in the local industry, college access information and professional mentorship
- The Data Detectives: Using Real Data to Solve Real Community Health Problems from ***Emory University*** provides underrepresented middle school students with a curriculum focused on using population-level Big Data for community health needs assessment, planning, analysis, evaluation and application.



Co-funding Support for Training and Career Development Awards

ODSS supports co-funding of K awardee Dr. Gabriel Tajeu, Ph.D.



- K01 recipient from Temple University
- Associate Professor in Public Health
- Research topic: Machine Learning to Reduce Hypertension Treatment Clinical Inertia
- Developed and validated ML algorithms to predict uncontrolled blood pressure among adults with hypertension
- Identified barriers and facilitators to implementing machine learning-based clinical decision support tools in healthcare settings
- Contributed valuable insights into the relationship between alcohol use and blood pressure, the development of interventions to decrease racial bias among healthcare staff, and the cost-effectiveness of hypertension treatment strategies



Graduate Data Science Summer Program (GDSSP)



National Institutes of Health



- In collaboration with the NIH Office of Intramural Training and Education (OITE).
- Summer internship in the Intramural Research Program (IRP) at NIH.
- Must be current Master's student at the time of start of program.
- Must be able to commit 10 weeks in summer full-time.
- More information: <https://www.training.nih.gov/research-training/grads/summer-internship-program-sip/gdssp/>

Work at NIH: DATA Scholar Program

Data and Technology Advancement National Service Scholar Program



- Recruit talents with **advanced data science expertise** to the NIH to use transformative approaches that lead to increased efficiency, innovative research, tool development and analytics.
- One to two years commitment.
- In addition to their own project, scholars participate in workgroups and collaborations, and contribute in many ways to the NIH
- <https://datascience.nih.gov/data-scholars-2023>

Open for Applications: March 2024

Contact:

Dr. Bryan Kim

ds-workforce@nih.gov



DATA Scholar Highlight

Vision for
Improving
Interoperability
for Eye Health
Data (NEI)

- Oregon Health and Science University
- Uses her knowledge of standard health care terminologies and common data models
- Enhance data standardization and interoperability in ocular health
- Drives community consensus building and develops strategies to advance ocular health care through data standardization



Dr. Michelle Hribar



New Pilot: NIH DataPath Program

*Goal: To recruit and develop **early career** data science talents from diverse backgrounds to contribute their expertise to NIH operations while gaining valuable experience within the NIH environment.*

- Collaborate with the **U.S. Digital Corps Program.**
- Recruit **postbacs** and **post Masters** data talents to work at the NIH for 2 years.
- The program will:
 - Provide immediate data science capabilities for the NIH.
 - Foster a pipeline of skilled professionals who can contribute to the future of data-driven research and operations at NIH.



UNITED STATES
**DIGITAL
CORPS**

Five Fellow Tracks:

- Cybersecurity.
- Data Science and Analytics.
- Design.
- Product Management.
- Software Engineering.

Open for Applications:
Fall 2024

Contact:
Evelyn Botchway
ds-workforce@nih.gov



Other NIH Opportunities: Pathways



- Current Students: internship opportunities.
- Recent Graduates: within two years of degree or certificate completion.
- Presidential Management Fellow Program: administered by the U.S. Office of Personnel Management (OPM); NIH has participated since 1985; for candidates with advanced degree.
- More information: <https://hr.nih.gov/jobs/student/pathways-students-and-recent-graduates>

Helping advance biomedical research by delivering access to industry-leading cloud providers.



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.

Enroll Now

Gain access to

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

>200
Petabytes of
Data

274M
Compute
Hours

>995
NIH & NIH-funded
Research
Programs/
Projects

\$41M
Cost Savings

>4700
People Trained

<https://datascience.nih.gov/strides>

ODSS Common Data Elements (CDE) Workshop

Advancing the Use and Development of Common Data Elements in Research

<https://www.scgcorp.com/AdvancingUseofCDEs24/>

March 6th – 7th, 2024

Hybrid (Virtual and In-person)

Workshop Sessions

I: The Value of Common Data Elements (CDEs)

II: Current NIH Resources for CDEs

III: Overcoming Barriers in CDE Adoption, Mapping, and Use in Community Research

IV: Technical Implementation Aspects of Mapping, Transformation, and Harmonization

V: Approaches to Improve Interoperability

VI: Use Cases for Preparing and Applying CDEs for Intelligent Technologies (AI/ML)

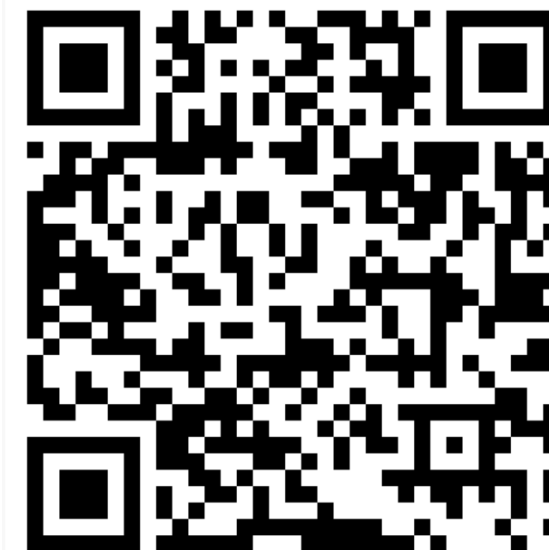
Venue (In-Person Attendees)

Ruth Kirschstein Auditorium, Natcher Conference Center (Building 45)

National Institutes of Health (Main Campus)

9000 Rockville Pike Bethesda, MD 20892

Register Here



Request for Information (RFI)

Proposed Use of Common Data Elements (CDEs) for NIH-Funded Clinical Research and Trials

- 1. Recommended CDEs for NIH-funded clinical research/trials, including a set of minimal core CDEs**
 - Minimal core CDEs required for all NIH-funded clinical research/trials
 - CDEs for autoimmune Diseases, immune-mediated diseases, clinical domains in categories
 - High level CDEs for SEDH domain
- 2. Technology standards for using NIH CDEs**
 - Strategies for ensuring CDEs are machine-readable and/or machine computable
 - Tools and processes to assist in data transformation and mapping
 - Resources to streamline and design upstream data collection that support downstream data harmonization
 - Resources needed to ensure that CDEs are tagged with appropriate terminology codes
- 3. NIH policies and governance on CDEs**
 - Enabling the adoption of the proposed requirement to use minimum core CDEs
 - Difficulties or obstacles foreseen in the adoption of the proposed requirement to use minimum core CDEs
 - Useful policies and governance on establishing new CDEs and data sharing and management

RFI Link



<https://datascience.nih.gov/cde-rfi>

Motivation and Inspiration

The thing that keeps us going – Why we do what we do

- Think about the “why” – it needs to come from within
- Remember the “why”
- Know that the “why” may evolve and it’s OK
- Think about the “why”

My first “why”: curiosity

My sustaining “why”: science enables me to serve the people

What is your “why”?



**"Sorry, pal, right metaphor,
wrong motivation."**

Persistence

It's OK to fail – as long as you don't stop there



- Keep trying
- Maintain your goals, change your approach
- Seek support

Know Your Support Network

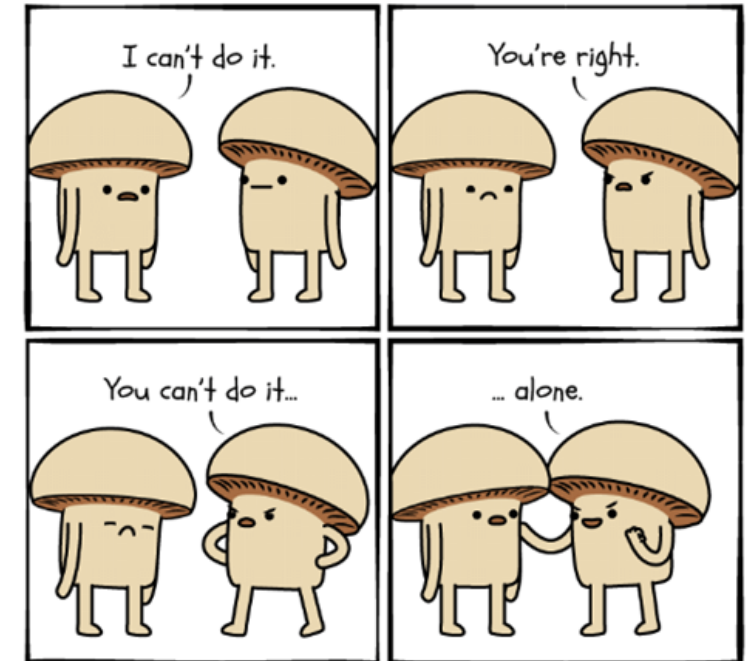
Know your support network

Build your support network

Maintain your support network



- You can't do it alone
- Family and friends
- Mentors and advocates
- Colleagues and collaborators



@mushroommovie

Upcoming Opportunities

NIH ODSS Webpage

<https://datascience.nih.gov>

Apply for Capacity Building and Training

- Administrative Supplements to Enhance Institutional Data Science Capacity ([NOT-OD-23-123](#)), Due Date **April 1, 2024**
- Broadening Opportunities for Computational Genomics and Data Science Education (UE5, [RFA-HG-23-002](#)), Due Date **June 11, 2024**



Come Work at NIH

- Data and Technology Advancement (DATA) National Service Scholar [Program](#): Advanced data experts, 1-2 years. Applications open in March.
- DataPath Fellows: early career technologists, 2 years. Applications open in the Fall.
- [Fellowship and Job Opportunities | Data Science at NIH](#)

Contact

ds-workforce@nih.gov

Share Your Thoughts



NIH Strategic Plan for Data Science 2023-2028:
Request for Information ([RFI](#))

Deadline **March 15, 2024**

<https://grants.nih.gov/grants/guide/notice-files/NOT-OD-24-037.html>



TWICE Team and Contact



Dr. Alison Lin
Lead



Dr. Raphael Isokpehi
Program Director



Dr. Bryan Kim
Program Director



Evelyn Botchway
Program Analyst



Nicholas Andrade
Training Specialist

- **Contact:** ds-workforce@nih.gov ←
- ODSS webpage: <https://datascience.nih.gov/>
- **Data science funding opportunities:** <https://datascience.nih.gov/nih-grants-and-funding-opportunities>
- Repositories for sharing scientific data: <https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data/repositories-for-sharing-scientific-data>
- **Data science job opportunities at NIH (federal):** <https://datascience.nih.gov/jobs>