

# Supplemental Training in Making Data FAIR and AI/ML Ready

**Columbia University T32**

ADVANCED TRAINING IN ENVIRONMENTAL HEALTH AND DATA SCIENCE: MOLECULES TO POPULATIONS

**PIs:** Miller, Factor Litvak, Shaman

**Training Faculty:** Jeanette A Stingone & Marianthi-Anna Kioumourtzoglou

**Add'l Curriculum Committee:** Lena Mamykina, Kenneth Miller, Suzanne Bakken, Wes Greuber

# Parent NIEHS Training Grant

- Single, unified training program
- 18 pre-doctoral students & 8 post-doctoral fellows
- Intersection of Data Science & Environmental Health
- Training in advanced data analytics, environmental epidemiology, climate science, molecular mechanisms of disease, and the exposome

The integration of additional training in making diverse epidemiologic, toxicological, and clinical data FAIR and ready for use with AI/ML is a natural progression for our multi-disciplinary training program

# Framework for Training

Integrated into existing training seminar (E2C2)

Novel Curriculum that integrates

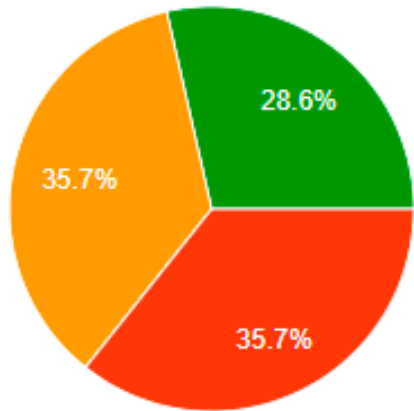
- Recorded didactic seminars (live or pre-recorded)
- Faculty-facilitated discussion to clarify concepts & discuss contextually-relevant examples
- Hands-on activities to promote competency & practical skills in data science use

## RESOURCES AVAILABLE AT COLUMBIA

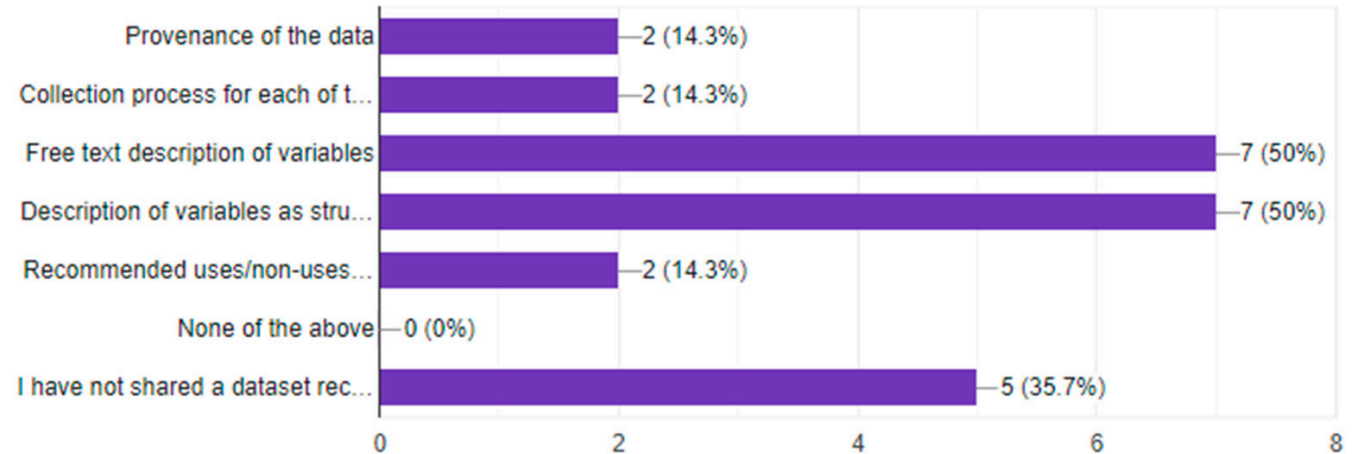
- Columbia University Data Science Institute
- NIH-Supported Administrative Core Centers
- OHDSI
- Neurobiology & Behavior PhD Program, Center for Theoretical Neuroscience
- Center for Teaching and Learning

# Needs Assessment helped shape training

Online Survey Distributed to training seminar attendees including trainees, faculty, staff, etc.



- Not at all relevant, I do not anticipate using these skills in my research and practice
- Somewhat relevant, I anticipate collaborating with individuals who need/have these skills as part of my research..
- Very relevant, I anticipate using these skills in my research and practice
- I'm not sure what AI/ML Readiness includes so I can't judge its relevance...



# Four Training Sessions with Different Modalities

- ❑ Overview of the FAIR Data Principles
  - ❑ Reading, Didactic Lecture and Discussion led by Columbia Faculty
- ❑ Ethics and Algorithmic Fairness
  - ❑ Postdoc-led Journal Club
- ❑ Using Git/GitHub for Reproducible Research in EHS
  - ❑ Guest Faculty Webinar (Recorded for Future Use)
- ❑ Using Ontologies to Facilitate Data Sharing
  - ❑ Flipped Training: Watch Intro Video and then Discussion and Hands-on Training Activity
  - ❑ Used GitHub to Disseminate Materials

# Challenges Faced

- Lack of in-person meetings due to continuation of COVID-19 pandemic
  - Zoom burnout
- Limited existing AI/ML resources directed at an EHS audience

# Plans for Dissemination

Needs Assessment already shared with other grantees via Slack

Creation of GitHub Repo that collates all developed materials (end of 2022)

Reading list, sample data, recorded lectures, assessments, etc.

**Idea for Group:** Joint symposium submission(s) to a national/international EHS meeting