



Workshop on Developing a Data Science Competent EHS Workforce

August 14-15, 2018

NIEHS Building 101, Rodbell Auditorium, Research Triangle Park, N.C.

AGENDA

Tuesday, August 14

8:30 – 8:45 a.m. **Opening Remarks and Purpose of the Workshop**
Carol Shreffler, Ph.D., NIEHS
Linda Birnbaum, Ph.D., Director, NIEHS and NTP

SESSION ONE: Understanding the Current State of Data Science in the EHS Domain As It Relates to Training

Goal: Through the evaluation of representative use cases, identify current limitations and rate-limiting steps for data science training in EHS. Presentations should include a high-level description of the persons or things involved, a sequence of actions and process flow, key competencies, and key obstacles. Discussion will entail a deep dive into use cases to identify relevant issues and challenges involved in solving data-intensive EHS research questions.

Moderator: *Charles Schmitt, Ph.D., NIEHS Office of Data Science*

8:45 – 8:50 a.m. **Introduction to Session One**
Charles Schmitt, Ph.D., NIEHS Office of Data Science

8:50 – 9:20 a.m. **Use Case 1: Tools, Skills, and Training for Quantitative Trait Locus Mapping**
Daniel Gatti, Ph.D., The Jackson Laboratory

9:20 – 9:50 a.m. **Use Case 2: Targeted Learning for Exposome Science**
Alan Hubbard, Ph.D., University of California, Berkeley

9:50 – 10:05 a.m. **Break**

10:05 – 10:35 a.m. **Use Case 3: Mapping What Matters: The Promise of Geospatial Health Informatics**
Marie Lynn Miranda, Ph.D., Rice University

10:35 – 11:05 a.m. **Use Case 4: Maximizing the Impact of Exposure Data in Children's Environmental Health Research**
Susan Teitelbaum, Ph.D., Icahn School of Medicine at Mount Sinai

11:05 a.m. – noon **Panel Discussion**

Noon – 1:00 p.m. **Lunch**

1:00 – 1:15 p.m. **Overview of Big Data Science Training**
Carol Shreffler, Ph.D., NIEHS

SESSION TWO: Relating BD2K Accomplishments to EHS Training Goals

Goal: Examine existing training resources relevant to the intersection of EHS and data science, including existing data science training resources (e.g., BD2K).

Moderator: *Amy Herring, Sc.D., Duke University*

- 1:15 – 1:20 p.m.** **Introduction to Session Two**
Amy Herring, Sc.D., Duke University
- 1:20 – 1:50 p.m.** **Applying the FAIR Principles to Online Data Science Training Resources**
John Van Horn, Ph.D., University of Southern California
- 1:50 – 2:20 p.m.** **BD2K Training at the Harvard T.H. Chan School of Public Health**
John Quackenbush, Ph.D., Harvard T.H. Chan School of Public Health and Dana-Farber Cancer Institute
- 2:20 – 2:35 p.m.** **Break**
- 2:35 – 3:05 p.m.** **The Physician Data Scientist: An Unexpected Journey**
Jonathan Chen, M.D., Ph.D., Stanford University
- 3:05 – 3:35 p.m.** **The Training of Next Generation Data Scientists**
Lana Garmire, Ph.D., University of Hawaii
- 3:35 – 4:05 p.m.** **NLM Training in Bioinformatics – Biomedical Data Science Training for Diverse Backgrounds and Career Paths**
Mark Craven, Ph.D., and Chris Bradfield, Ph.D., University of Wisconsin-Madison
- 4:05 – 5:00 p.m.** **Panel Discussion**
- 5:00 p.m.** **Adjourn Day One**

Wednesday, August 15

- 8:30 – 8:35 a.m.** **Day Two Opening Remarks: Chris Duncan, Ph.D., NIEHS**

SESSION THREE: Building EHS Training in Big Data Science

Goal: Examine existing NIEHS training resources (e.g., trainee pipelines) relevant to data science and discuss the next steps to increase training emphasis in big data. This may include both training for quantitative big data scientist career paths and needed big data skills for biomedical scientists in other environmental health programs. Identify the limitations and rate-limiting steps in data science training in EHS and make recommendations for priority areas.

Moderator: *Cheryl Walker, Ph.D., Baylor College of Medicine*

- 8:35 – 8:40 a.m.** **Introduction to Session Three**
Cheryl Walker, Ph.D., Baylor College of Medicine
- 8:40 – 9:10 a.m.** **Experiences With Environmental Bioinformatics Training**
Fred Wright, Ph.D., North Carolina State University
- 9:10 – 9:40 a.m.** **Data Science Training in Epidemiology**
Jim Gauderman, Ph.D., University of Southern California
- 9:40 – 9:55 a.m.** **Break**
- 9:55 – 10:25 a.m.** **Building Multidisciplinary Environmental Health Data Science Training Resources**
Chirag Patel, Ph.D., Harvard University
- 10:25 a.m. – 12:20 p.m.** **Panel Discussion on the Training Intersection of Data Science and Environmental Health Sciences**
Additional Panelists:
Wesley Gray, Ph.D., Southern University and A&M College
Ronald Hines, Ph.D., U.S. Environmental Protection Agency
Jeanette Stingone, Ph.D., Icahn School of Medicine at Mount Sinai
- 12:20 – 12:30 p.m.** **Meeting Wrap-Up and Closing Remarks**
Gwen Collman, Ph.D., Director, NIEHS/DERT
- 12:30 p.m.** **Adjourn Meeting**