



# EMERGING SCIENCE FOR ENVIRONMENTAL HEALTH DECISIONS

## AGENDA

### Integrating Environmental Health Data to Advance Discovery

JANUARY 10–11, 2013\* ■ THURSDAY 8:30–5:00, FRIDAY 8:30–12:45  
KECK CENTER, ROOM 100, 500 FIFTH STREET NW, WASHINGTON, DC

**THIS EVENT WILL BE WEBCAST**

**RESEARCH IN BIOMEDICAL SCIENCES** has undergone a dramatic transformation in the past two decades. Science is increasingly data-intensive, computational, interdisciplinary, and collaborative. This trend toward “Big Data” is pervasive throughout science and imposing new challenges for biomedical research along three dimensions, sometimes referred to as the three V’s: volume, velocity and variety. Only through the coordination of all three dimensions will the full potential of Big Data be realized. Whereas significant progress has been made in the development of digital technologies, community-wide principles and resource management for some large and rapidly expanding data

types such as genomic sequences, integration of existing heterogeneous data sets (the variety component of the three V’s) has lagged. This lag presents particular challenges for environmental health sciences, which is uniquely and inherently cross-disciplinary. This meeting aims to foster discussion about the need for enhanced data integration in environmental health sciences, evaluate the lessons that can be learned from integrative initiatives in other scientific domains, and strategize about how the community can take major steps toward improving data coordination and access to advance understanding about environmental effects on human health.

#### THURSDAY, JANUARY 10 (8:30AM-5PM)

- 8:30 **Welcome**—Carolyn Mattingly<sup>†</sup>, *North Carolina State University (NCSU)*
- 8:45 **Opening Remarks**—Allen Dearry, *National Institute of Environmental Health Sciences (NIEHS)*

#### SESSION 1 USING HETEROGENEOUS DATA TO ADVANCE DISCOVERY

Session 1 will provide a broad overview of the scientific opportunities and insights that arise from analysis of heterogeneous data sets. Speakers will draw examples from ongoing efforts, such as Sagebase and the Geisinger Health System, to integrate biological, epidemiological, medical, and other data to improve understanding of human health and disease.

- 9:10 **Scientific Opportunities from Heterogeneous Biological Data Analysis: Overcoming Complexity**—Stephen Friend, *Sage Bionetworks*

\* The committee and liaisons will meet the afternoon of January 11 until 3pm.

<sup>†</sup> indicates a member of the Standing Committee on Use of Emerging Science for Environmental Health Decisions.

- 9:55 **Doing Environmental Epidemiologic Research with Electronic Health Records**—Brian Schwartz, *Johns Hopkins Bloomberg School of Public Health (JHSPH)*
- 10:30 *Break*

#### SESSION 2 DATA INTEGRATION—STRATEGIES AND LESSONS LEARNED

This session will take a deeper look at ongoing data integration efforts or programs within a variety of health-related research fields. Speakers will discuss lessons-learned from these data integration efforts and suggest ways in which these lessons can inform the environmental health sciences community as the scope and diversity of environmental health data continues to expand.

- Moderator: Carolyn Mattingly<sup>†</sup>, *NCSU*
- 10:50 **Building Platforms for Data Integration: Diverse Perspectives**—C. Titus Brown, *Michigan State University (MSU)*
- 11:20 **Web-based Data Integration: Chemical Safety for Sustainability Dashboard**—Matthew Martin, *U.S. Environmental Protection Agency (EPA)*
- 11:50 *Lunch (provided for speakers and committee in Room 101; cafeteria is on 3rd floor)*

(continued)

**THURSDAY, JANUARY 10 (8:30AM-5PM)**  
(CONTINUED)

**SESSION 2 DATA INTEGRATION—STRATEGIES AND LESSONS LEARNED (CONTINUED)**

- 1:00 **Cancer Genomics Big Data: From The Cancer Genome Atlas to the Cancer Information Donor and Cancer Commons**—Barbara Wold, *California Institute of Technology*
- 1:30 **Data Interconnectivity and Federation: National Database of Autism Research**—Gregory K. Farber, *Office of Technology Development and Coordination, NIMH*
- 2:00 **The Encyclopedia of DNA Elements (ENCODE): An Experience of Data Integration from Genomics**—Ewan Birney, *European Bioinformatics Institute (by videoconference)*
- 2:30 *Break*

**SESSION 3 ENVIRONMENTAL HEALTH SCIENCE DATA STREAMS**

Session 3 will provide snapshots of areas of science and their associated data sets that can help advance environmental health research. Speakers will discuss the unique challenges—from availability to accessibility to complexity—that could impact the ability to integrate environmental health data.

- 2:45 **Brief Overview of Environmental Health Science Data Streams**—George Daston<sup>†</sup>, *Proctor & Gamble (P&G)*
- 3:00–5:15 **Key Environmental Health Science Data Streams: Challenges and Opportunities**
- **Air Pollution Data**—Francesca Dominici, *Harvard School of Public Health (3:00)*
  - **Health Data**—Brian Schwartz, *JHSPH (3:25)*
  - **Genomics Data and Pathway Data**—Carolyn Mattingly, *NCSU (3:50)*
  - **Exposure Data**—Elaine Cohen Hubal, *EPA (4:15)*
  - **Cheminformatics for Toxicology**—Ann Richard, *EPA (4:40)*

For more information and to subscribe for updates, please visit <http://nas-sites.org/emergingscience/>  
Emerging Science meetings are free and open to the public.

**About the Committee**

At the request of the National Institute of Environmental Health Sciences (NIEHS), the National Research Council formed the Standing Committee on Use of Emerging Science for Environmental Health Decisions to facilitate communication among government, industry, environmental groups, and the academic community about scientific advances that may be used in the identification, quantification, and control of environmental impacts on human health.

**FRIDAY, JANUARY 11 (8:30AM-12:45PM)**

**SESSION 3 ENVIRONMENTAL HEALTH SCIENCE DATA STREAMS (CONTINUED)**

- 8:30 **Environmental Health Science Data Stream Panel Discussion**  
**Moderator:** George Daston<sup>†</sup>, *P&G*
- Panelists from Thursday afternoon's presentations will reflect upon the infrastructure and technical needs for data integration that could be informed by other integration efforts and programs as well as the needs specific to the environmental health that may require unique, innovative solutions. Some key questions included for discussion are:
- Should different data streams be prioritized for integration?
  - Are there common themes among diverse data streams that challenge their integration?

9:40 *Break*

**SESSION 4 DEFINING A ROLE FOR THE FEDERAL GOVERNMENT**

The final session will address the role of the federal government in developing strategies and coordinating and managing efforts to integrate heterogeneous environmental health science data sets. Participants will discuss steps that may require centralized or top-down management as well as what may be best accomplished from the bottom-up.

**Moderator:** Ivan Rusyn<sup>†</sup>, *University of North Carolina at Chapel Hill (UNC-CH)*

- 10:00 **Community of Science: Strategies for Coordinating Integration of Big and Heterogeneous Data**—Kevin T. Gallagher, *U.S. Geological Survey*
- 10:30 **Federally Supported Data Integration Efforts: National Science Foundation and Cross-Agency Initiatives**—Suzanne Iacono, *National Science Foundation (NSF)*
- 11:00 **Panel Discussion**  
**Panelists:** C. Titus Brown, *MSU*; Stephen Friend, *Sage Bionetworks*; Suzanne Iacono, *NSF*; Carolyn Mattingly, *NCSU*; Kevin T. Gallagher, *USGS*; Rick Woychik, *NIEHS*
- 12:10 **Closing Remarks**—Stan Ahalt, *UNC-CH*
- 12:45 **Adjourn**—Committee and Liaisons meet until 3pm

