NIEHS and Data Integration
Where We’ve Been and Where We’re Going

Linda Birnbaum, Ph.D., D.A.B.T., A.T.S.
Director
National Institute of Environmental Health Sciences
National Toxicology Program
Informing Environmental Health Decisions through Data Integration

Washington, D.C.
February 20-21, 2018
Data Science Landscape

High-Throughput Technologies

Computational Technologies & Infrastructure

Data- and Knowledge-Driven Discovery

FAIR Data
Findable, Accessible, Interoperable, Reusable
EHS and Emerging Data Science

- **2012-2017** – NIEHS Strategic Plan, Goal 7: Crosscutting Theme on Knowledge Management
- **2012** – NIEHS Data Sharing Strategies for Environmental Health Science Research Workshop
- **2013** – Identifying Opportunities for Global Integration of Toxicogenomics Databases Workshop
- **2013** – NIEHS-NCATS-UNC DREAM Toxicogenetics Challenge
- **2013** – NASEM ESEHD Integrating Environmental Health Data to Advance Discovery
- **2013** – Big Data to Knowledge (BD2K) Community-Based Data and Metadata Standards Efforts (R24):
  - **2014** – Development of a Framework for Environmental Health Science Language Workshop
  - **2014** – NTP Systematic Review and Evidence Integration for Literature-Based Environmental Health Science Assessments
- **2016** – Principles and Best Practices for Sharing Data from Environmental Health Research NASEM IOM Roundtable Workshop
Recent Investments

- **Office of Data Science (ODS)**
  - Leadership and support to enable discovery, access, and use of data needed to advance research, policy, and decision-making. Supports making NIEHS data FAIR.
  - 2018 focus on tools for data integration and metadata management

- **Office of Scientific Computing (OSC)**
  - Expertise, staff, and resources in scientific IT and applications

- **Director of Environmental Health Sciences Cyberinfrastructure**
  - Ensure coordination of OSC and ODS with CIO office and other IT programs

- **Other Staff Hires**
  - DNTP - Data Scientist and Developer
  - DERT - Data Scientist

- **Training**
  - E.g., Courses in machine learning, database structure, programming in “R”
Children’s Health Exposure Analysis Resource
Data Repository, Analysis, and Science Center

- Chemical, biological and epidemiologic measurements
- Familial relationships: mother-child; sibling linkages
- Multiple time-periods across fetal and child development
Comparative Toxicogenomics Database

- 30.5 million toxicogenomic connections:
  - Chemicals/drugs
  - Genes/proteins
  - Diseases
  - Taxa
  - Gene Ontology (GO) annotations
  - Pathways
  - Gene interaction modules

http://ctdbase.org/

NTP Toxico-Informatics

Suite of tools for management, analysis and visualization of chemical effects data

- BMD Express 2
- Tox21 Enricher
- Tox 21 Activity Profiler (beta)
- Tox 21 Curve Browser (beta)
- Drug Matrix/ToxFX
- Correlation Browser
- Integrated Chemical Environment (ICE)
- Chemical Effects in Biological Systems (CEBS)

Integration of chemical, gene expression, pathology, molecular, pharmacology, literature, and pathway data
NIH Data Commons

Co-location of large and/or highly utilized NIH funded data with storage and computing infrastructure and commonly used tools for analyzing and sharing digital objects to create an interoperable resource to enable investigators to collaborate and share.

- NIEHS Coordinating with Commons
  - Serve on Institutional Partners program
  - Advise on Helium Team (UNC) to ensure ability to move data & metadata between the NIEHS and NIH commons
Increasing Interoperability of NIEHS Data Systems

Application Programming Interfaces (API)

Deploying APIs on NIEHS Data Assets using common terminologies, e.g., for chemicals, chemical findings, and data set descriptors
Road Ahead: Towards the Information Commons and Knowledge Network

- Represents a Grand Challenge for Computer & Information Science
- Key are finding use cases that:
  - Enable science
  - Challenge the infrastructure
- NCATS Data Translator Program
  - Competency Questions
  - Identify Needs & Gaps
  - Across several use cases
    - Asthma
    - Fanconi
Can we identify sub-types of Asthma that inform treatment options and provide study populations of scientific interest?

- Integration of clinical (from NIEHS & UNC), air quality, socio-economic, patient surveys, genetic, and biomarker data.

Given a novel chemical, can we determine estimated properties that guide subsequent testing?

- Integration of public knowledgebases, NTP data sets, and chemical similarity models.
Workshop Charge

• …explore the promise and potential pitfalls of environmental health data integration

• …participants will work together to conceptualize future research directions

• Read the draft NIEHS Strategic Plan and Comment at: https://www.niehs.nih.gov/strategicplan
Thank you!

Provide your feedback on our Strategic Plan: www.niehs.nih.gov/strategicplan

http://www.niehs.nih.gov

National Institutes of Health • U.S. Department of Health and Human Services