Interagency Coordinating Committee on the Validation of Alternative Methods

Data Sharing

Nicole C. Kleinstreuer, PhD
NICEATM Deputy Director

SACATM Meeting

September 18-19, 2017

Agency for Toxic Substances and Disease Registry • Consumer Product Safety Commission • Department of Agriculture
Department of Defense • Department of Energy • Department of the Interior • Department of Transportation
Environmental Protection Agency • Food and Drug Administration • National Institute for Occupational Safety and Health
National Institutes of Health • National Cancer Institute • National Institute of Environmental Health Sciences
National Library of Medicine • Occupational Safety and Health Administration

Strategic Goal: Foster the use of efficient, flexible, and robust practices to establish confidence in new methods

- Identify and collate sources of high-quality human toxicological data.
- Create centralized data access points that are publicly available and easily accessible.
- Actively solicit the submission and collation of parallel data from animal studies and alternative methods.
- Leverage partnerships and complimentary initiatives
  - FAIR Principles, Biomed21 recommendations, SEND reporting, NIH Data Commons, NCATS Data Translator, EPA Chemistry Dashboard, NLM Databases, etc.

FAIR PRINCIPLES	
Findable	A data object should be uniquely and persistently identifiable.
Accessible	Data is accessible by authorized users (human and machine) through a well-defined protocol.
Interoperable	(Meta) data assigned to the data object is syntactically parse-able and semantically machine accessible.
Reusable	Data objects must comply with the above three principles and sufficiently documented to allow integration/linkage with other data sources.

https://www.force11.org/fairprinciples

Wilkinson et al. 2016

State of FAIRness at NIEHS

- Microcosm of larger research enterprise
- Heterogenous mix of:
  - Data systems and technologies
  - Data management practices
  - Metadata capture and standards
  - Funding mechanisms for building & sustaining systems
  - Teams and talent around building & maintaining systems
  - Policies around usage of data
- Common data publication practices in place:
  - submission to *dbGAP, GEO, clinicaltrails.gov, …*

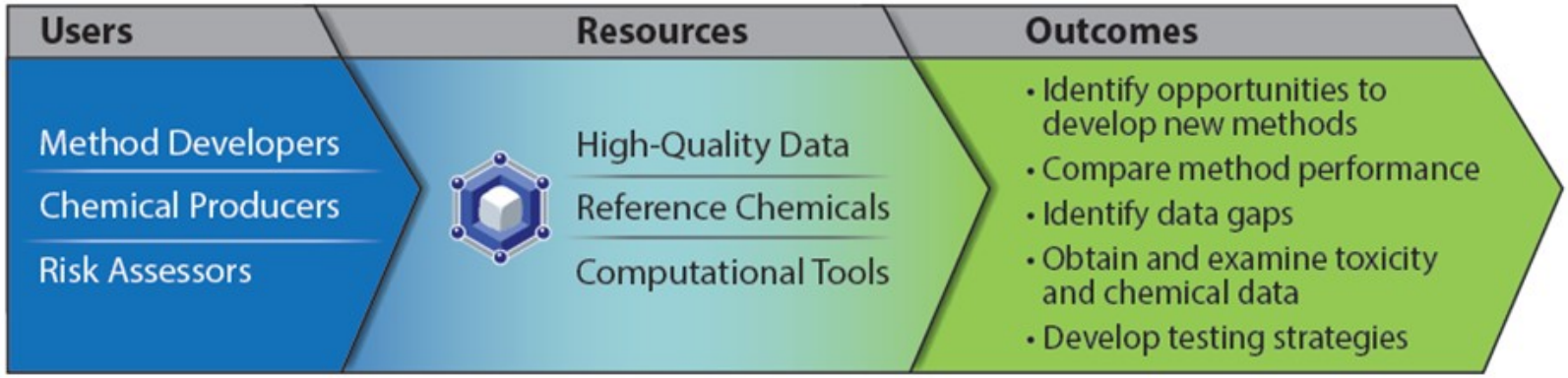
State of FAIRness at NICEATM

- **Integrated Chemical Environment (ICE)**
- https://ice.ntp.niehs.nih.gov/

  - Provides FAIR for ICCVAM data
  - Feeder into other applications and data systems
    - Chemical Effects in Biological Systems (CEBS)
    - CTD, EPA databases, etc.

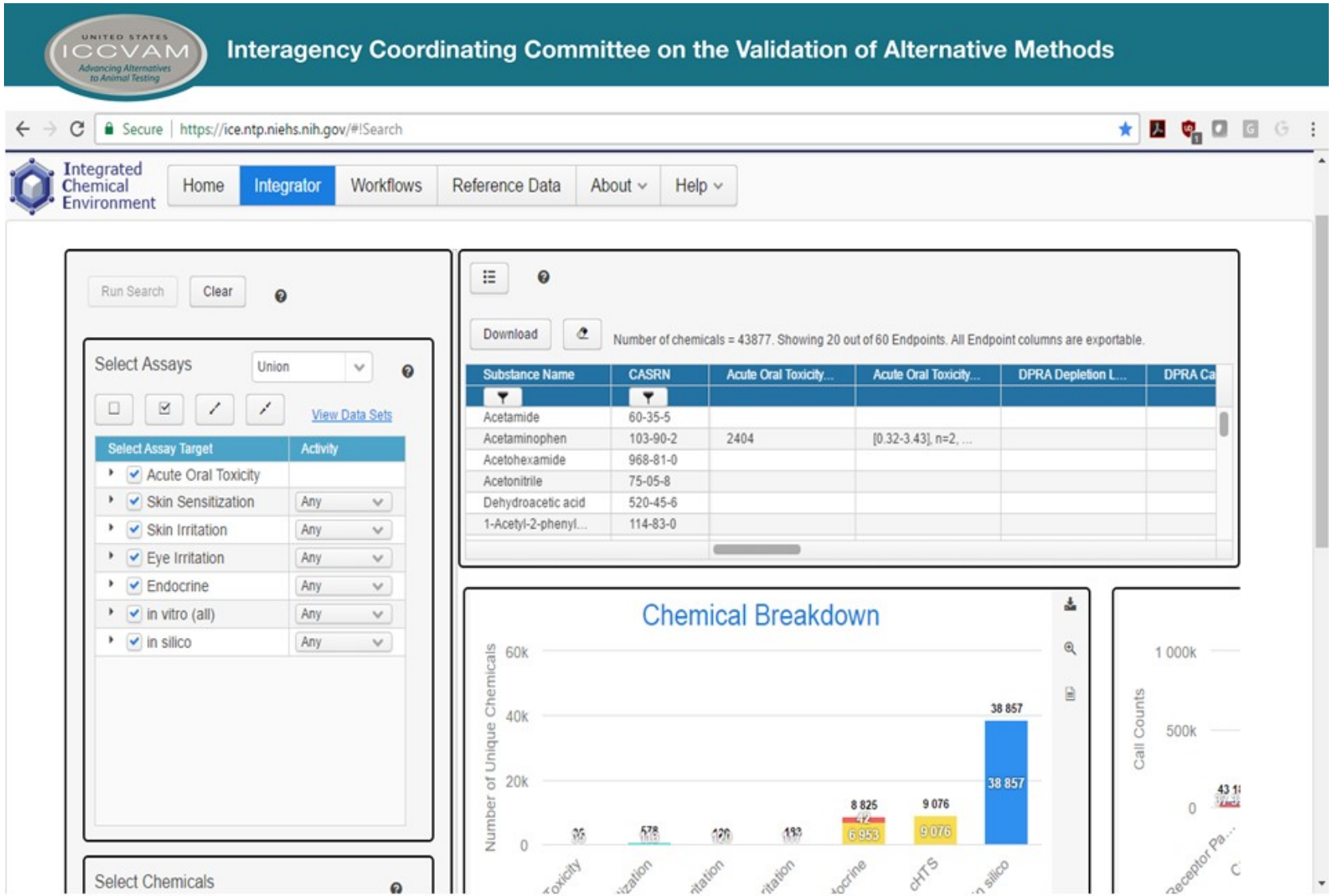
  - Significant ongoing efforts to improve data utility
    - Intuitive user interface
    - Web-based visual analytics
    - Improved search functionality
    - Improving capture of data provenance
    - Web application programming interfaces (APIs)
    - Many more…

Interagency Coordinating Committee on the Validation of Alternative Methods



https://ice.ntp.niehs.nih.gov/

- Integrated Chemical Environment: ICE
  - Structured format designed for ease of use
  - Allows access to data for multiple regulatory endpoints
  - Query by CASRN or established reference chemical lists
  - Flexible, exportable results
    - Both in computer-friendly format and a human-friendly summary format for quickly comparing data between chemicals



https://ice.ntp.niehs.nih.gov/

Challenges facing FAIR

- IR (interoperable, reusable) are especially challenging
  - Depends on the questions being asked
  - Requires agreement and coordination across parties
- Not all data can be moved
  - Issues based on policy, size
  - Computation must be moved to the data
- Finite resources
  - Community must guide what data and type of access is prioritized

BioMed21 (26-27June 2017)

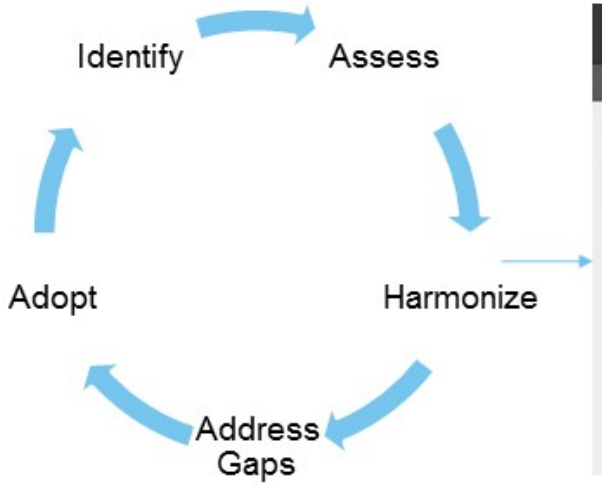
http://www.biomed21.org

- Standardization of data reporting (e.g. SEND)
- Data availability and accessibility
- Data quality and relevance
- Ontological classification and data integration
- Multi-layer communication protocols
- Systems biology approaches
- Establishing partnerships

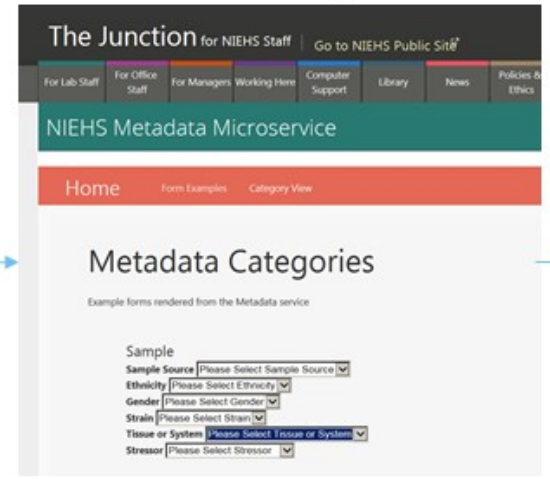
BioMed21: Major Recommendations

- International and inter-agency collaboration is critical: formal collaboration between major organizational and funding bodies should be established
- Funding should be prioritized for researching human-based biology and promoting open access data
- Human data should be collected in collaborative, open-access high-quality databases
- Common reporting formats and common ontologies should be established for collecting and collating human biology information, from different 'omics technologies to human clinical data
- There is a need to establish formal processes for cross-sector communication
- There is an immediate need for the creation of case studies to demonstrate applications and benefits of predictive, mechanism-based approaches

Towards Common Terminology



Live catalog



- NIEHS Data Commons
- Integrate with future systems
- Harmonized use by ICE, CEBS

Effort being led by Charles Schmitt, NIEHS/ODS

Evaluating Data Quality

- Understanding reproducibility, relevance, and readability
- Establishing minimum criteria for animal study data based on guideline protocols
  - Ex: Uterotrophic database (*Kleinstreuer et al. 2016*)
- Semi-automated retrieval and evaluation of published literature
  - Project with Oak Ridge National Labs (ORNL) to apply text-mining approaches to identify high-quality data

Interoperability Across Systems

Consistent & compatible web-APIs

CEBS

Integrated Chemical Environment

ICE

Others…

Data Commons

NIEHS Data Systems

Consistent data set access & retrieval

CSS Dashboard

PubChem

Many others…

TOXNET Toxicology Data Network

External Data Systems

Acknowledgments

- NICEATM staff
- ICCVAM members
- BioMed21 participants
- NIEHS DNTP
  - Mary Wolfe, John Bucher, Andy Shapiro
- NIEHS Office of Data Science
  - **Charles Schmitt,** Stephanie Holmgren