Harmonizing Tox Data to Enhance Confidence and Utility

Aswani Unnikrishnan[1*], Kim T. To[1], Alexandre Borrel[1], Victoria Hull[1], Amber Daniel[1], Emily Reinke[1], Nicole Kleinstreuer[2]

[1]*Inotiv, Research Triangle Park, NC, USA;* [2]*NIH/NIEHS/DTT/NICEATM, Research Triangle Park, NC, USA*

*Presenting author

Leveraging diverse, toxicologically relevant data from multiple sources is essential for chemical risk characterization and the development and validation of new approach methodologies. However, inconsistencies in formatting and terminology present significant challenges to data integration and analysis. Here we present a comprehensive approach to clean and aggregate toxicologically relevant data from disparate sources (e.g., in vivo and in vitro studies, physico-chemical properties, in silico models), for inclusion in NICEATM's Integrated Chemical Environment (ICE: https://ice.ntp.niehs.nih.gov/). The approach incorporates subject matter expertise with computational techniques like programmatic databases and processing scripts to streamline data harmonization. The processing pipeline is applied across millions of data points via a series of quality control steps to identify potential errors or inconsistencies (e.g., duplicate entries, missing values) and appropriately rectify them. Pipelined data are organized using standardized terminology to facilitate comparisons across datasets, derive meaningful insights, and uphold FAIR principles: findability, accessibility, interoperability, and reusability. This approach highlights data transparency and curation for ensuring reliability and integrity through metadata annotations of data provenance, assumptions, and quality assurance practices. The iterative nature of data cleanup and aggregation processes emphasizes the need for ongoing collaborations across stakeholder groups to continually refine and validate data. This presentation will feature examples of practical applications of this approach from ICE and demonstrate effective methods for making data high-confidence and robust. This project was funded with federal funds from the NIEHS, NIH under Contract No. HHSN273201500010C.