The EPA CompTox Chemicals Dashboard: An Integration Hub for Data Supporting Computational Toxicology

Antony Williams
National Center for Computational Toxicology
U.S. Environmental Protection Agency, RTP, NC

This work was reviewed by the U.S. EPA and approved for presentation but does not necessarily reflect official Agency policy.
National Center for Computational Toxicology established in 2005 to integrate:
- High-throughput and high-content technologies
- Modern molecular biology
- Data mining and statistical modeling
- Computational biology and chemistry

Currently staffed by ~60 employees as part of EPA’s Office of Research and Development
Home of ToxCast & ExpoCast research efforts
Key partner in U.S. Tox21 federal consortium
• Tens of thousands of chemicals in commerce and 100s more introduced every year

• Testing is expensive and slow with only a small fraction of chemicals fully evaluated for potential human health effects

• NCCT researchers integrate advances in biology, chemistry, and computer science to prioritize chemicals based on risk

• Underpinnings of our computational toxicology approaches
  – Data – high quality, curated data sourced from public resources and literature
  – Transparency – FAIR data available for download, reuse and repurposing
  – Prediction models – transparent, openly available (Github)
Publicly accessible website delivering access to:

- >875,000 chemicals with >25 million property data points
- >1 million toxicity data points from 30 public resources and >65,000 literature articles
- Millions of “Biological assay” data points for 1000s of chemicals
- Information about chemicals in consumer products
- Links to other agency websites and public data resources
- Integrated “literature” searches for ~30 million abstracts
CompTox Chemicals Dashboard
Rich data content, Powerful Tools
Substance Identifiers for Semantic Mapping
Take Home Messages

• FAIR and Open Data is critical to building scientific data hubs for the community

• Transparency – in data and predictive models is the new approach to science and should be embraced

• Data QUALITY is key and community collaboration and crowdsourcing is critical to success

• Interoperability is enabled by the adoption of open standards – especially ontologies and taxonomies