EPA United States Environmental Protection Agency

ExpoCast
exposure forecasting

Integration of Exposure NAMs to identify contaminants of emerging concern: A case study with the Minnesota DOH

Kristin Isaacs
Center for Computational Toxicology and Exposure
Office of Research and Development, US EPA

Accelerating the Pace of Chemical Risk Assessment 6
October 15, 2021

https://orcid.org/0000-0001-9547-1654

Background

- The US Environmental Protection Agency's Center for Computational Toxicology and Exposure (CCTE) and the Minnesota Department of Health (MDH) are collaborating to use new chemical data generated from scientific approaches such as read-across, QSAR, high-throughput toxicology screening, and computational modeling of exposure and toxicokinetics to prioritize chemicals for further evaluation and inform risk assessment
- CCTE and MDH finalized a formal Cooperative Research and Development Agreement (CRADA) in 2019
  - CRADA has a goal of addressing up to five MDH chemical evaluation activities

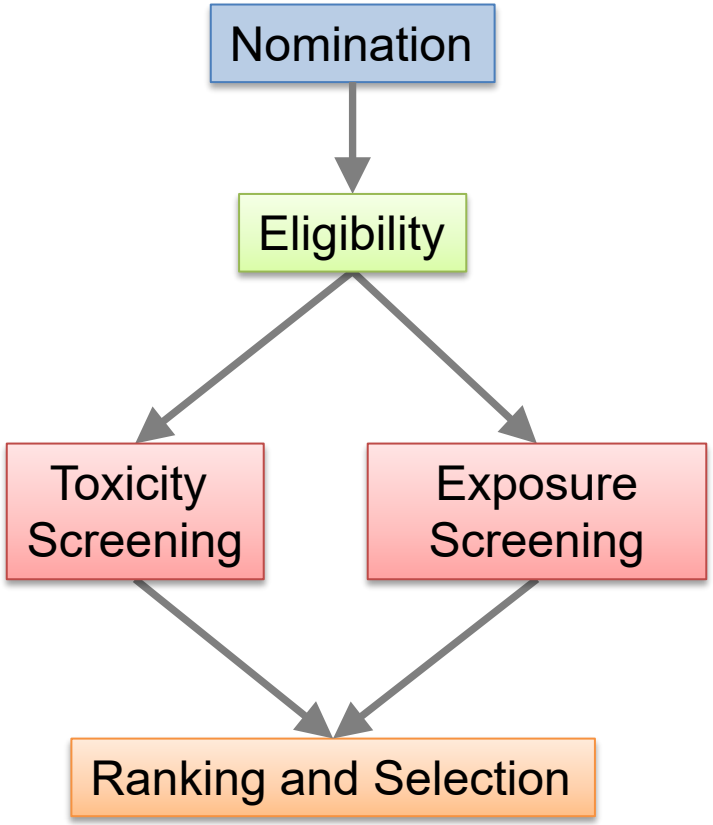
Problem: MDH CEC Initiative

- Through its **Contaminants of Emerging Concern** (CEC) initiative, the Minnesota Department of Health (MDH) collaborates with partners and the public to identify contaminants of interest in drinking water

- Substances that **have been released to, found in, or have the potential to enter Minnesota waters**, and:
  - Real or perceived health threat,
  - No current Minnesota human health-based guidance
  - New information that increases the level of concern

Problem: MDH CEC Initiative

MDH CEC Process



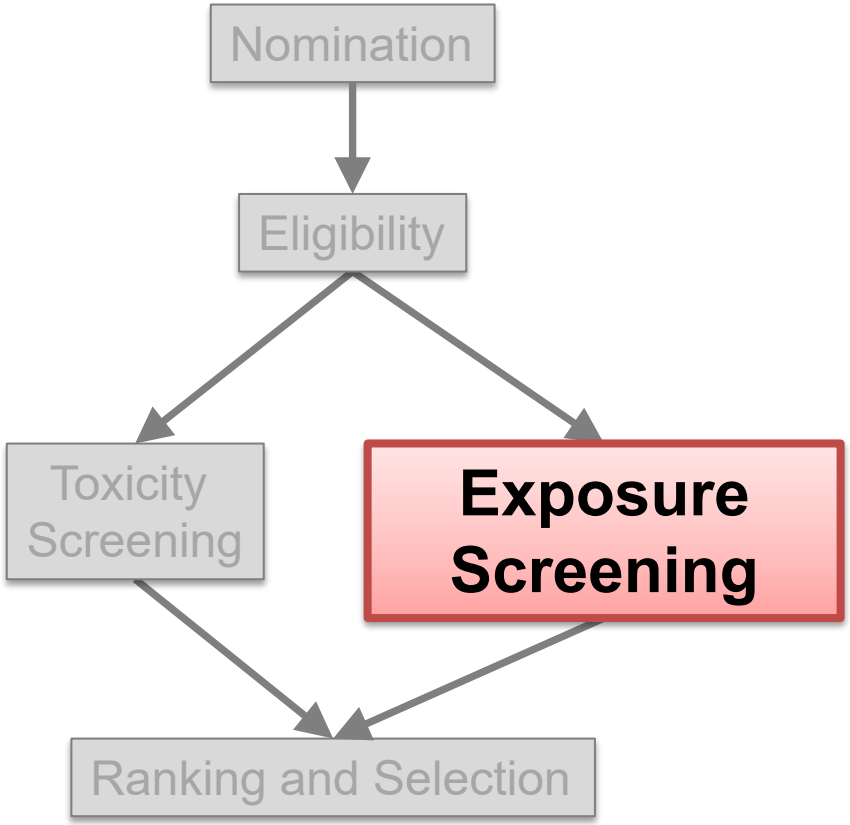
- Through its **Contaminants of Emerging Concern** (CEC) initiative, the Minnesota Department of Health (MDH) collaborates with partners and the public to identify contaminants of interest in drinking water

- Substances that **have been released to, found in, or have the potential to enter Minnesota waters**, and:
  - Real or perceived health threat,
  - No current Minnesota human health-based guidance
  - New information that increases the level of concern

- Substances selected via a nomination process, followed by:
  - **Screening-level evaluation and ranking of nominated chemicals based on exposure and toxicity potential**
  - Screening informs selection of contaminants for an in-depth toxicological review and guidance development

Problem: CEC Exposure Screening

MDH CEC Process

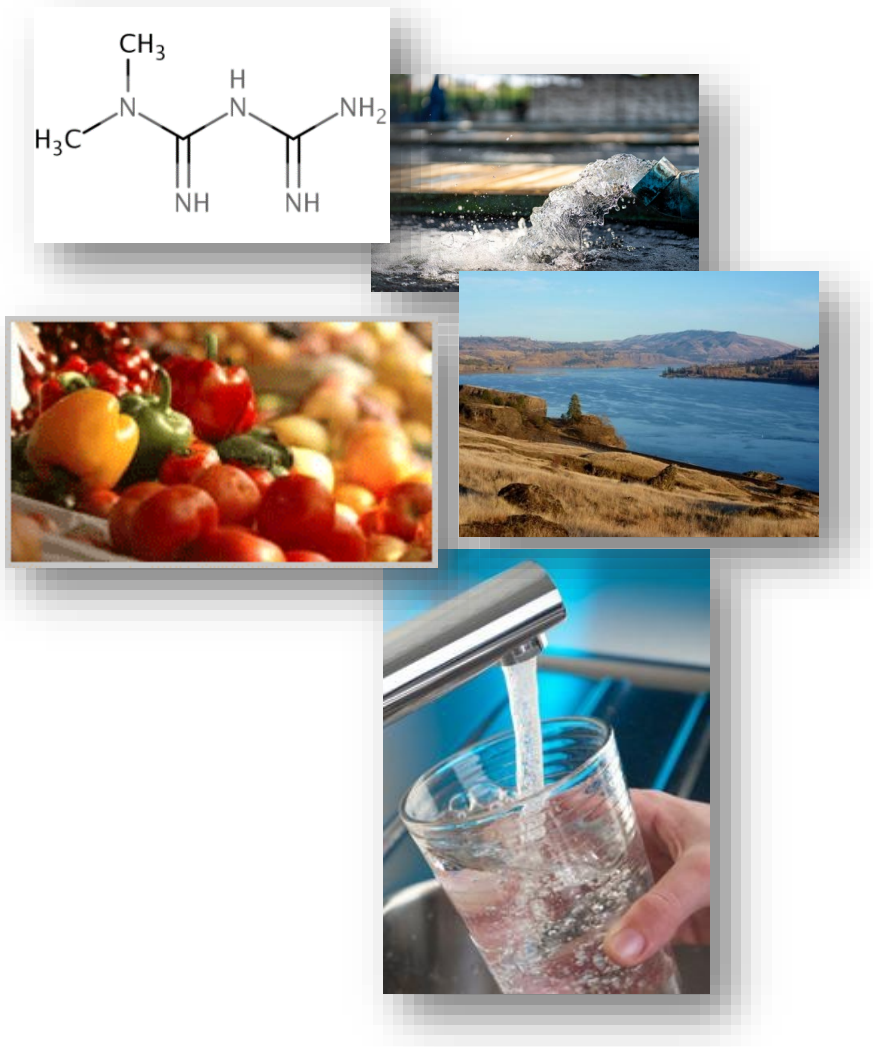


- **Exposure screening** was identified by MDH as a high-priority workflow for implementation under the CRADA
- Past approach: **manual** exposure screening by MDH staff
  - Data identification is time-consuming process (multiple days to a week for 1 chemical)
    - Disparate data sources
    - Synthesis can be challenging
  - Scoring is also manual: tedious/unreproducible
  - Many chemicals are **data-poor** based on traditional approaches (for example, existing regulatory exposure assessments, traditional monitoring data)

Approach

- Establish collaboration between MDH and CCTE accelerate the **exposure screening** process
- Develop a **proof-of-concept** automated workflow for scoring chemicals and reporting results according to MDH screening criteria
- Incorporate New Approach Methodologies (NAMs) for exposure from ORD's Exposure Forecasting (ExpoCast) project
- Apply workflow to two chemical lists
  - 87 chemicals previously manually evaluated by MDH (for assessment of workflow performance)
  - 171 proof-of-concept chemicals of interest to MDH and EPA

CEC Exposure Screening Criteria

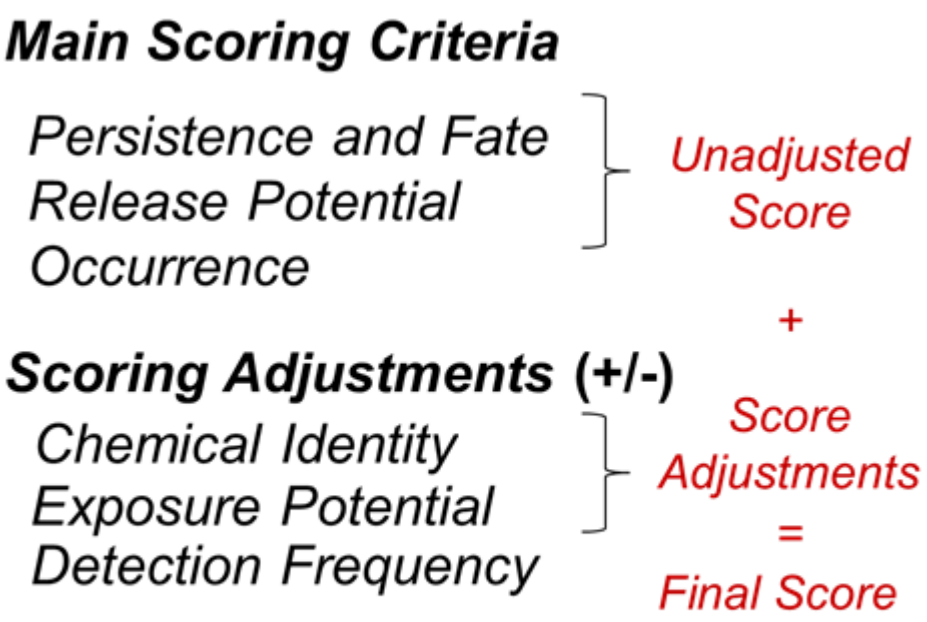


- Uses components of the US EPA's Office Water Candidate Contaminant List (CCL) methodology and incorporates the recommendations from MDH Stakeholder Task Group

- Considers data and criteria associated with multiple domains, including
  - Chemical identity and use
  - Chemical properties
  - Chemical emissions and disposal
  - Chemical occurrence in environment, drinking water, and food
  - Human exposure potential

- Incorporates MN information where possible

CEC Exposure Screening Criteria



- Uses components of the US EPA's Office Water Candidate Contaminant List (CCL) methodology and incorporates the recommendations from MDH Stakeholder Task Group

- Considers data and criteria associated with multiple domains, including
  - Chemical identity and use
  - Chemical properties
  - Chemical emissions and disposal
  - Chemical occurrence in environment, drinking water, and food
  - Human exposure potential

- Incorporates MN information where possible

- Evaluates and scores chemicals using algorithm developed by MDH (primary unadjusted score + score adjustments= final score)

Eight Classes of NAMs for Exposure from the ExpoCast Project

- ***Chemical descriptors*** that provide information on chemicals in an exposure context (e.g., how chemicals are used)
- ***Machine-learning approaches*** that use these descriptors to fill gaps in existing data
- ***High-throughput exposure models*** that address various pathways
- ***High-throughput measurements*** that fill gaps in monitoring data
- High-throughput approaches that measure or predict chemical ***toxicokinetics***
- New ***evaluation frameworks*** that integrate models and monitoring to provide consensus exposure predictions

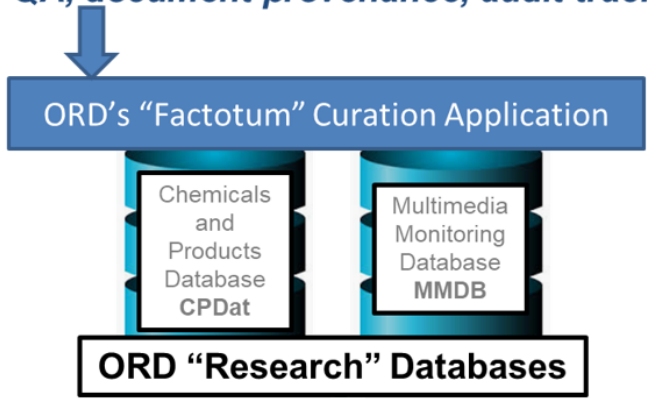
All these pieces together provide can accelerate high-throughput risk-based chemical prioritization

Workflow Design and Implementation

Data Curation

MN-specific documents and other source documents extracted and curated into ORD's research databases via the Factotum curation application.

QA, document provenance, audit tracking

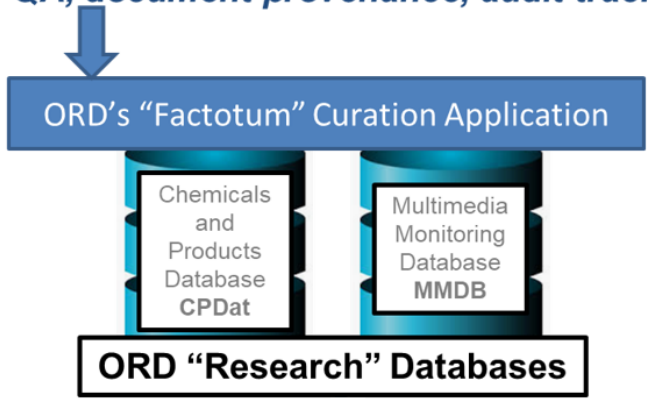


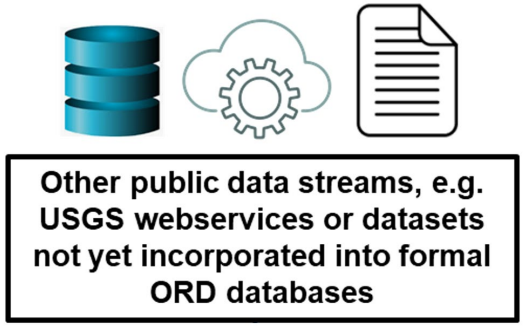
Workflow Design and Implementation

Data Curation

MN-specific documents and other source documents extracted and curated into ORD's research databases via the Factotum curation application.

QA, document provenance, audit tracking



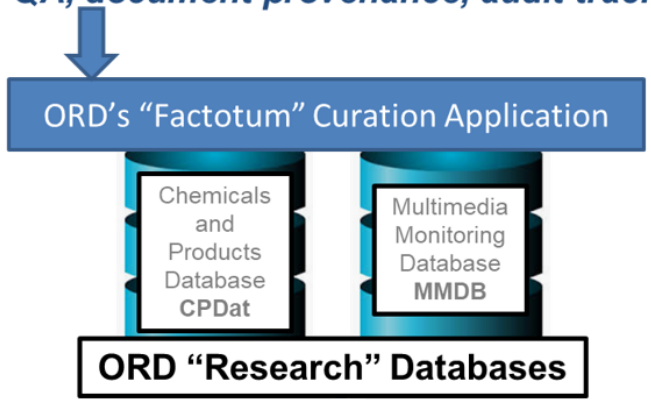


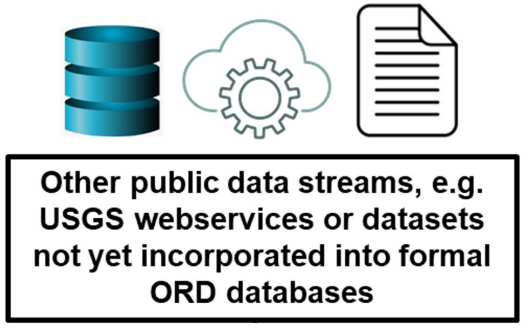
Workflow Design and Implementation

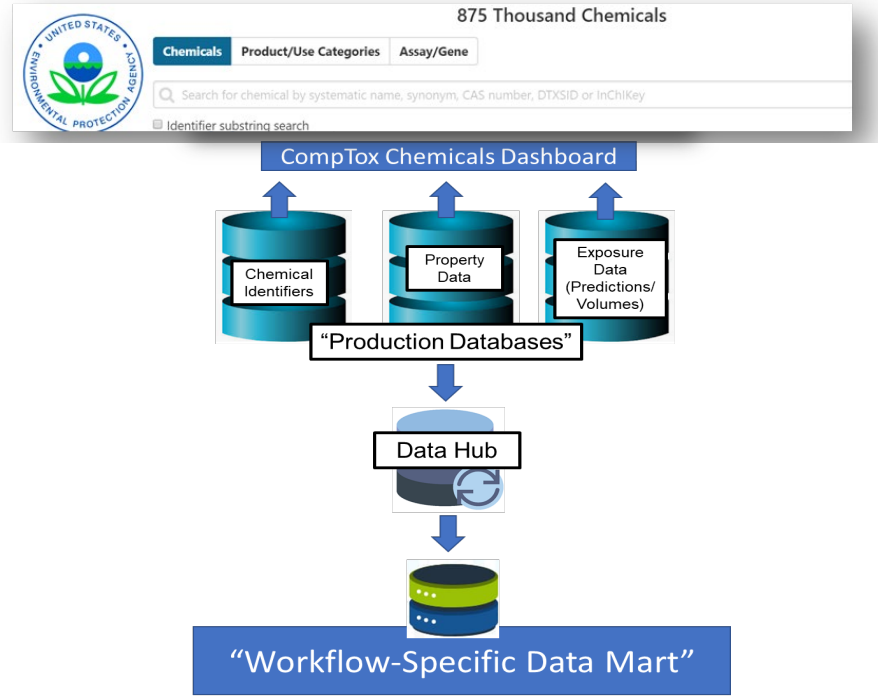
Data Curation

MN-specific documents and other source documents extracted and curated into ORD's research databases via the Factotum curation application.

QA, document provenance, audit tracking







Workflow Design and Implementation

Data Curation

MN-specific documents and other source documents extracted and curated into ORD's research databases via the Factotum curation application.

QA, document provenance, audit tracking

ORD's "Factotum" Curation Application

Chemicals and Products Database **CPDat**

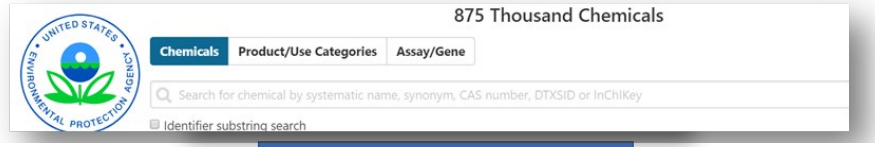
Multimedia Monitoring Database **MMDB**

ORD "Research" Databases

Other public data streams, e.g. USGS webservices or datasets not yet incorporated into formal ORD databases

R

- Data retrieval and caching



CompTox Chemicals Dashboard

Chemical Identifiers

Property Data

Exposure Data (Predictions/ Volumes)

"Production Databases"

Data Hub

"Workflow-Specific Data Mart"

Workflow Design and Implementation

United States Environmental Protection Agency

Data Curation

MN-specific documents and other source documents extracted and curated into ORD's research databases via the Factotum curation application.

QA, document provenance, audit tracking

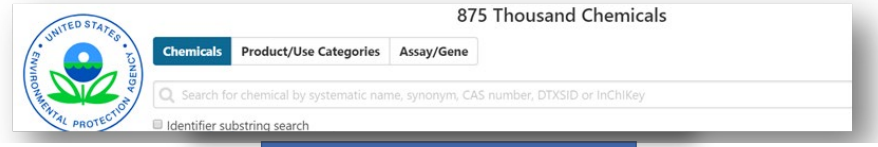
ORD's "Factotum" Curation Application

Chemicals and Products Database **CPDat**

Multimedia Monitoring Database **MMDB**

ORD "Research" Databases

Other public data streams, e.g. USGS webservices or datasets not yet incorporated into formal ORD databases



CompTox Chemicals Dashboard

Chemical Identifiers

Property Data

Exposure Data (Predictions/ Volumes)

"Production Databases"

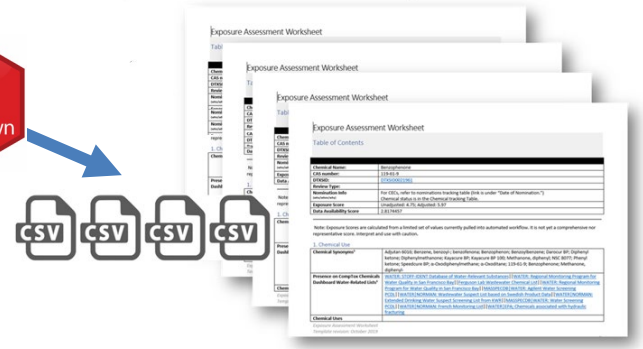
Data Hub

"Workflow-Specific Data Mart"

- Data retrieval and caching
- Chemical scoring
- Summary report and data table generation

rmarkdown

Automated Reporting and Data Generation for In-Depth Assessment



Main Scoring Criteria

Persistence and Fate
Release Potential
Occurrence
→ *Unadjusted Score*

\+

Scoring Adjustments **(+/-)**

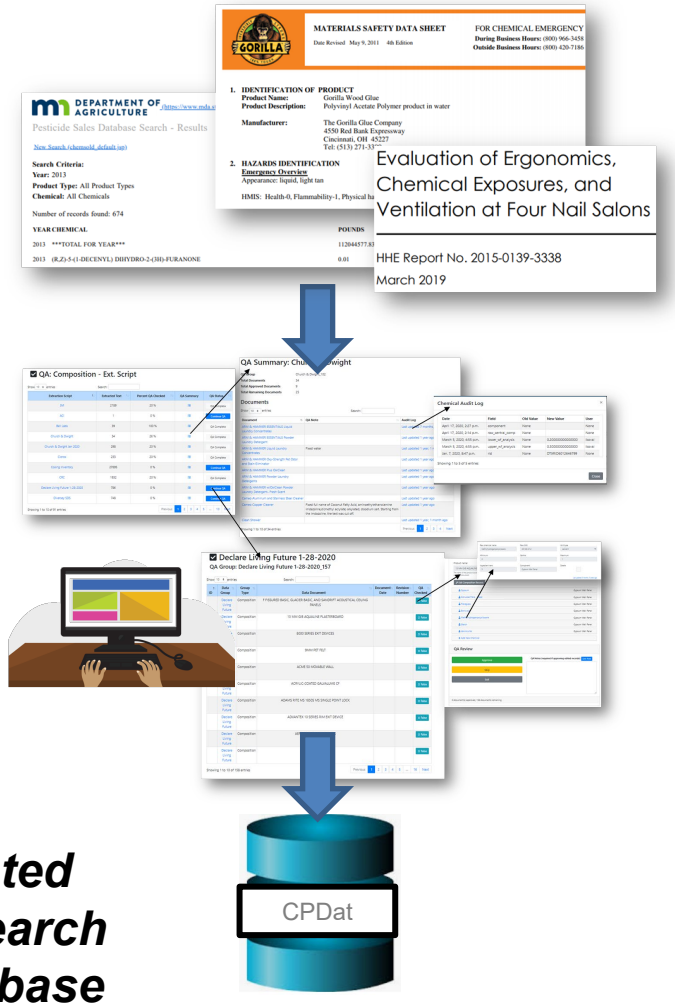
*Chemical Identity**
Exposure Potential
Detection Frequency
→ *Score Adjustments*

=

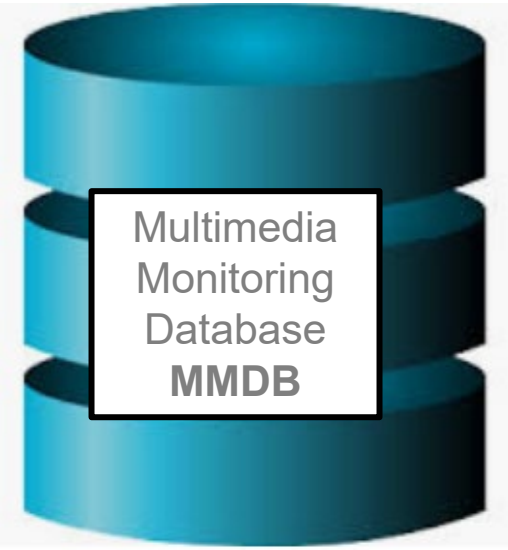
Final Score

Curation of Chemical Use Descriptors with Factotum

- We are using informatics approaches to obtain and curate chemical use descriptor information
- Public data sources: reports, consumer product ingredient data, etc.
- Utilizing standard curation/QA procedures and tools
- Currently supports EPA's Chemical and Products Database
- Integrates with ORD's chemical curation workflows
- Allowed us to curate many MN-specific documents for use in the workflow



Multimedia Monitoring Database (MMDB)



- ORD research database of measurements from over 20 public data sources
  - Includes data from several EPA programs, California state monitoring programs, the FDA, the Comparative Toxicogenomics Database, the EU's Information Platform for Chemical Monitoring Data (IPCHEM), the National Health and Nutrition Examination Survey (NHANES), the USDA, the International Council for the Exploration of the Sea (ICES), and the International Council of Chemical Associations' Long-Range Research Initiative (ICCA-LRI)
  - Harmonized to **chemical identifier and media** (e.g., drinking water, surface water, human blood or urine, soil, food, and ecological species).
- Developed in collaboration with OPPT
- Contains over **250 million** individual data records covering over **3200** unique chemicals
- Basis for future QSAR-like models for occurrence in different media
- Manuscript for submittal for peer-reviewed publication in internal EPA clearance

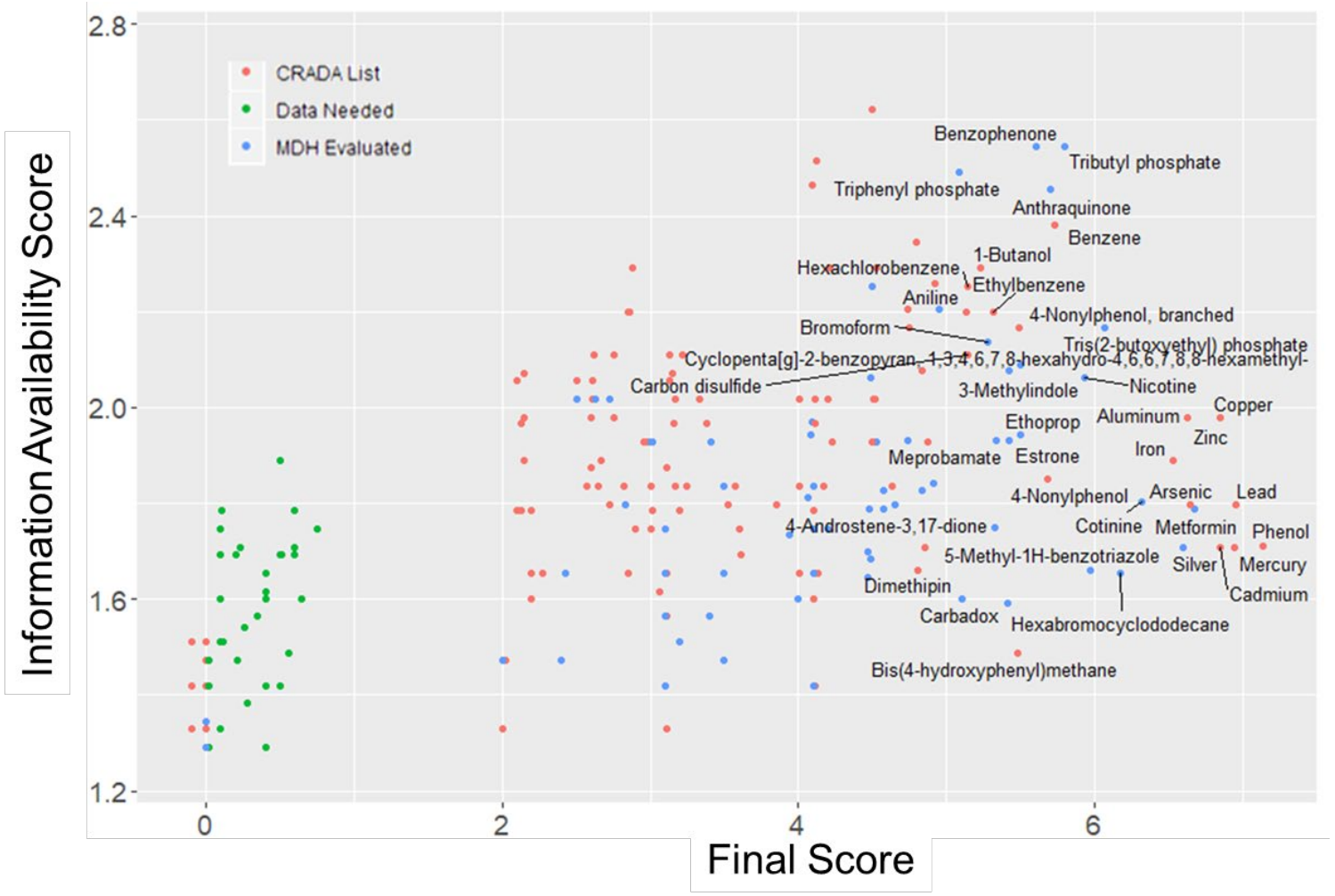
Data Source Summary

* Incorporate Exposure NAM data

Chemical Identity and Use	
Chemical Identifiers and Synonyms	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Uses	EPA-ORD's Chemicals and Products Database[1] (CPDat)
Uses	EPA's Chemical Data Reporting (CDR) Consumer, Commercial, Industrial uses
National Production Volume	EPA-ORD's CompTox Chemicals Dashboard (Underlying data)
Uses	EPA Safer Chemical Ingredients List
Chemical Properties	
Measured Properties	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Predicted Properties	EPA-ORD's CompTox Chemicals Dashboard (OPERA QSAR Models[4])
Predicted Wastewater Treatment Removal	EPA's Estimation Program Interface Suite (EPI-Suite)
Transformation Products	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Chemical Emissions and Disposal	
Pesticide Releases	National Agricultural Statistic Service
Chemical Releases	EPA's Toxics Release Inventory
Down-the-Drain Releases	EPA's SHEDS-HT model
Chemical Occurrence in Environment, Drinking Water, and Food	
Occurrence in Environmental Media, Including Drinking and Surface Water	EPA-ORD Multimedia Monitoring Database (MMDB)
Occurrence in US Water	US Geological Survey (USGS) Water Quality Portal data, via its application programming interface (API)
Occurrence in MN Water	Custom Database developed by USGS for MDH
Occurrence in MN Water	MN-specific reports, curated into EPA's chemical databases
Occurrence in Food	US Department of Agriculture (USDA) Pesticide Data Program
Occurrence in Food	US Food and Drug Administration (FDA) Substances Added to Food Database
Occurrence in Food	US Food and Drug Administration (FDA) Indirect Food Additives Database
Human Exposure	
Intake Exposures Inferred from Biomonitoring Data	EPA-ORD's CompTox Chemicals Dashboard/Underlying Databases
Biomonitoring Data	EPA-ORD Multimedia Modeling Database (MMDB)
Consumer Exposure Predictions	EPA-ORD's SHEDS-HT Model
General Population Exposures	EPA-ORD's Systematic Empirical Evaluation of Models (SEEM) Consensus Predictions
Presence on Biomonitoring Lists	Biomonitoring California

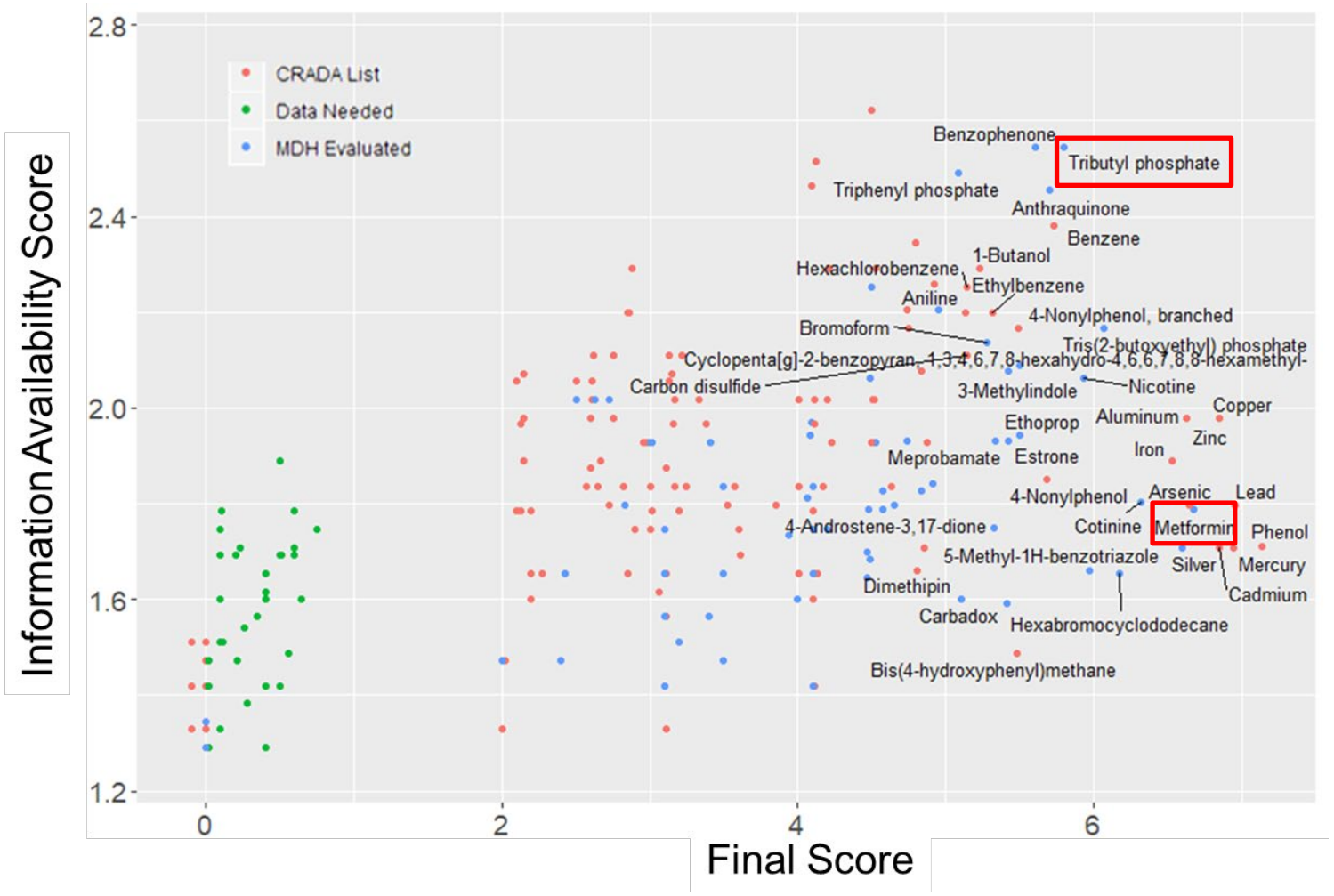
Results

- The automated workflow was applied to the 258 chemicals (87 evaluated by MDH previously, 171 on the current proof-of-concept list)
- Also defined an “Information Availability Score”
- All data collection, scoring, and report/table writing were completed in approximately 18 hours



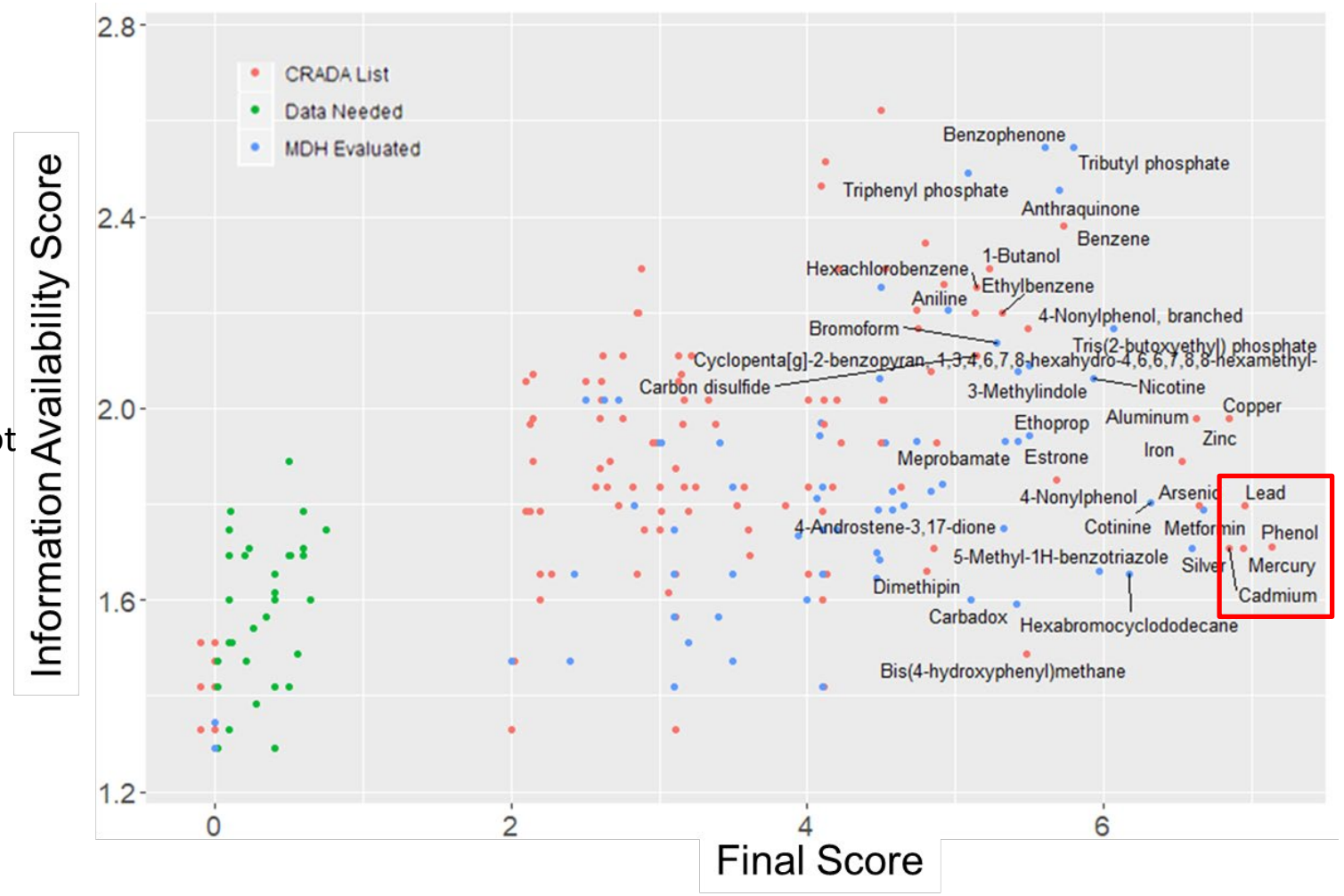
Results

- The automated workflow was applied to the 258 chemicals (87 evaluated by MDH previously, 171 on the current proof-of-concept list)
- Also defined an “Information Availability Score”
- All data collection, scoring, and report/table writing were completed in approximately 18 hours
- Many of the chemicals with the highest scores (>5) have already been screened by MDH.



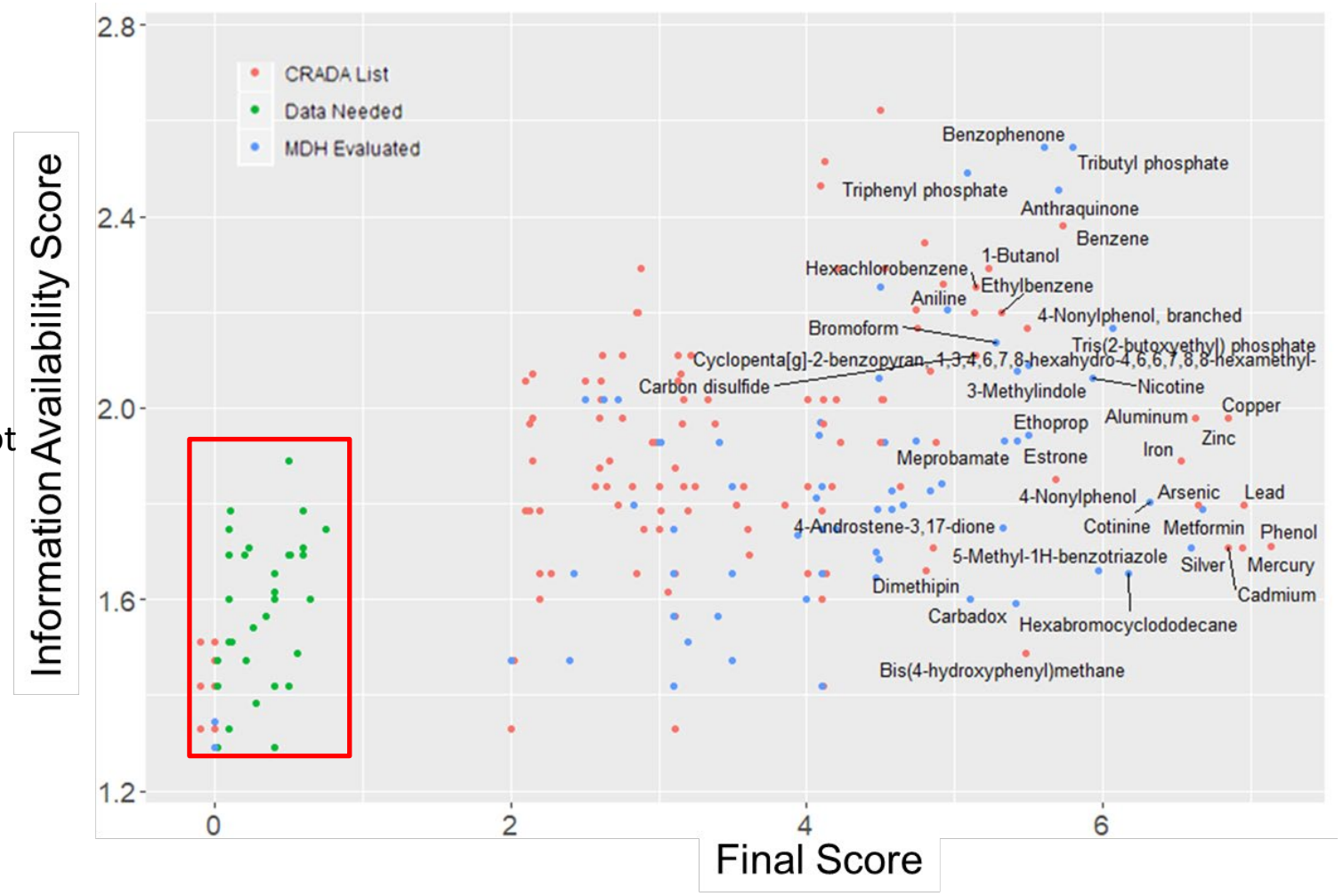
Results

- The automated workflow was applied to the 258 chemicals (87 evaluated by MDH previously, 171 on the current proof-of-concept list)
- Also defined an “Information Availability Score”
- All data collection, scoring, and report/table writing were completed in approximately 18 hours
- Many of the chemicals with the highest scores (>5) have already been screened by MDH.
  - Identified several other chemicals that have not undergone explicit exposure screening process by MDH but have been identified as priority to evaluate via assessments outside the CEC initiative



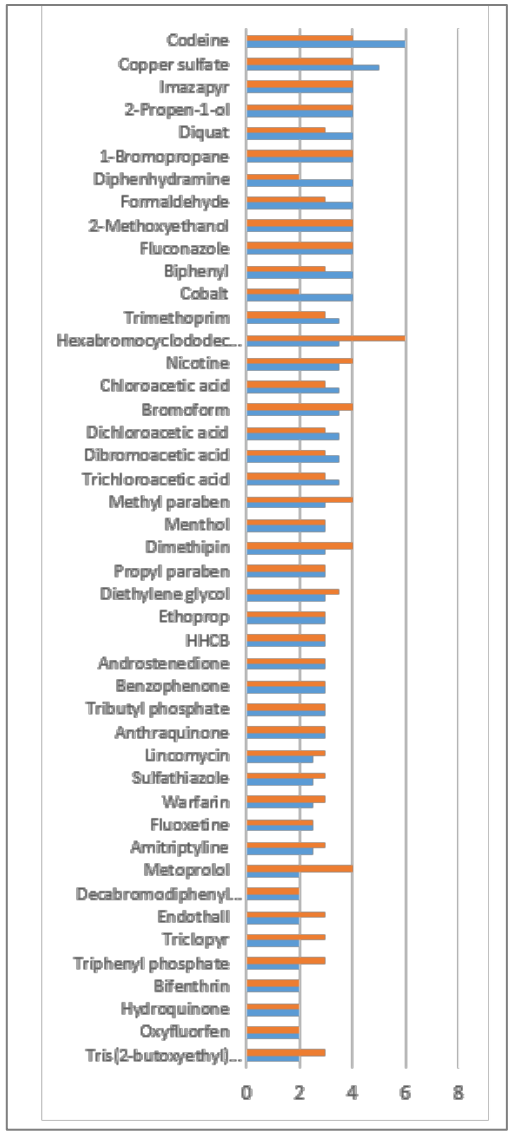
Results

- The automated workflow was applied to the 258 chemicals (87 evaluated by MDH previously, 171 on the current proof-of-concept list)
- Also defined an “Information Availability Score”
- All data collection, scoring, and report/table writing were completed in approximately 18 hours
- Many of the chemicals with the highest scores (>5) have already been screened by MDH.
  - Identified several other chemicals that have not undergone explicit exposure screening process by MDH but have been identified as priority to evaluate via assessments outside the CEC initiative
- There were 82 chemicals that did not have enough data for main unadjusted scores to be calculated
  - 36 had positive exposure scoring adjustment (might be priority for additional data collection/curation)



Initial Evaluation of Automated Workflow and Manual Results

- Excellent agreement between scores in **Persistence and Fate** and **Occurrence** domains



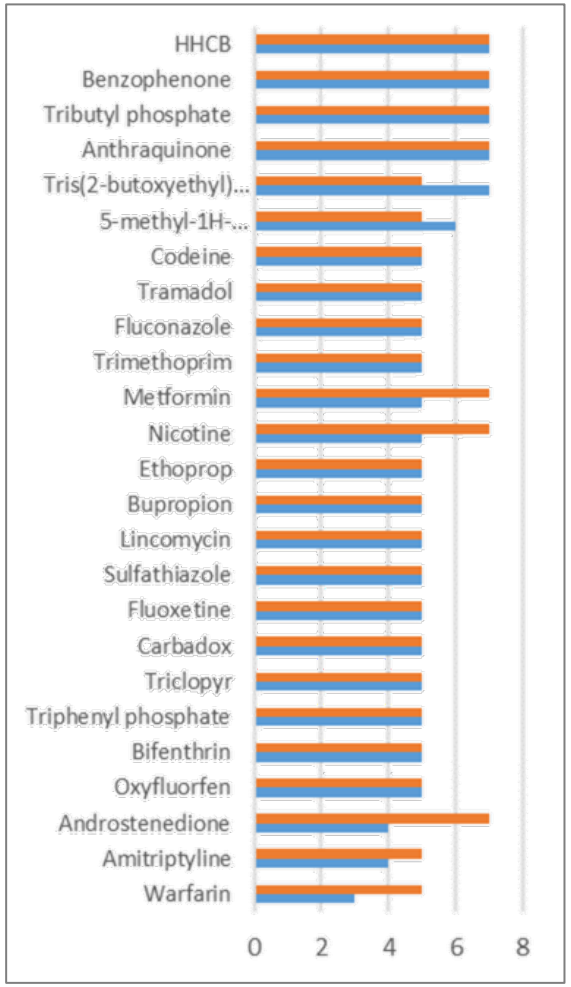
Persistence and Fate

Manual Score

Workflow Score

Initial Evaluation of Automated Workflow and Manual Results

- Excellent agreement between scores in **Persistence and Fate** and **Occurrence** domains



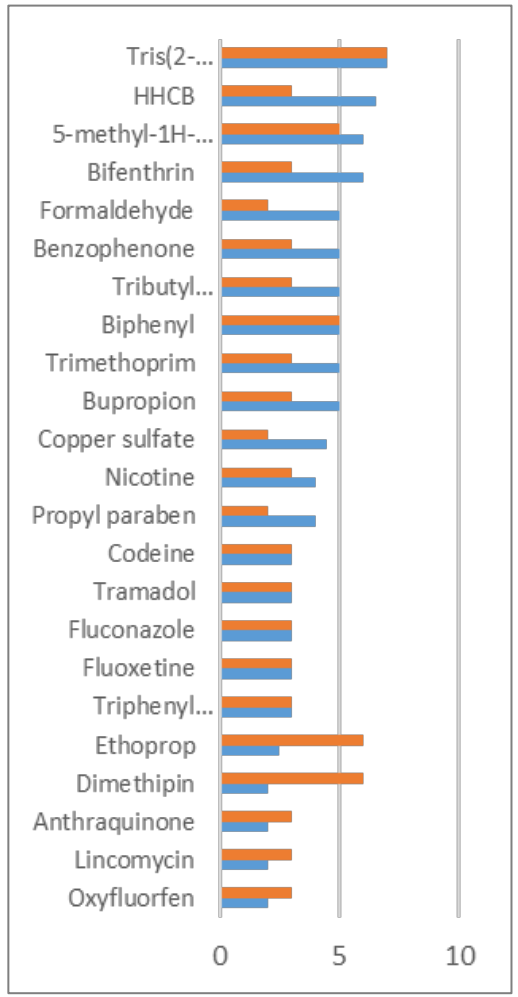
Occurrence

Manual Score

Workflow Score

Initial Evaluation of Automated Workflow and Manual Results

- Excellent agreement between scores in **Persistence and Fate** and **Occurrence** domains
- Somewhat poorer alignment in the **Release Potential** domain



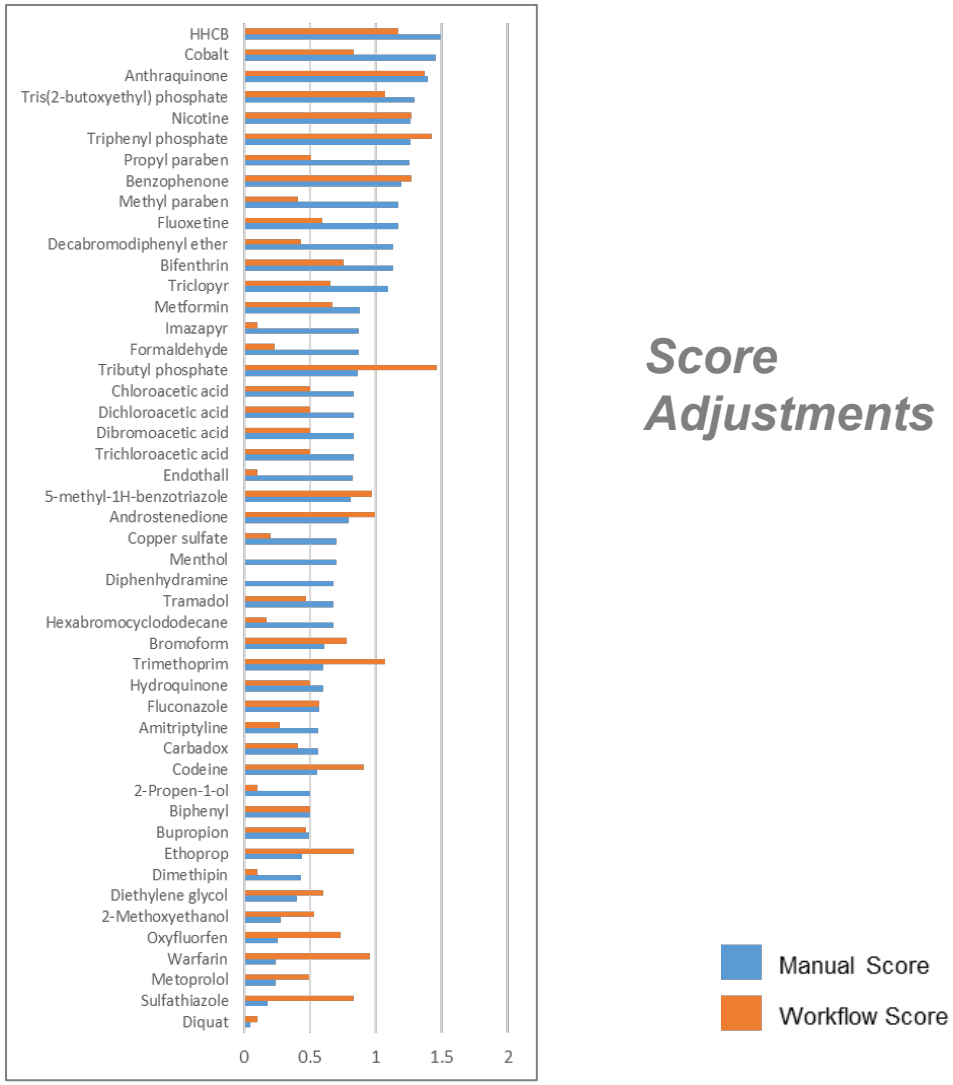
Release Potential

Manual Score

Workflow Score

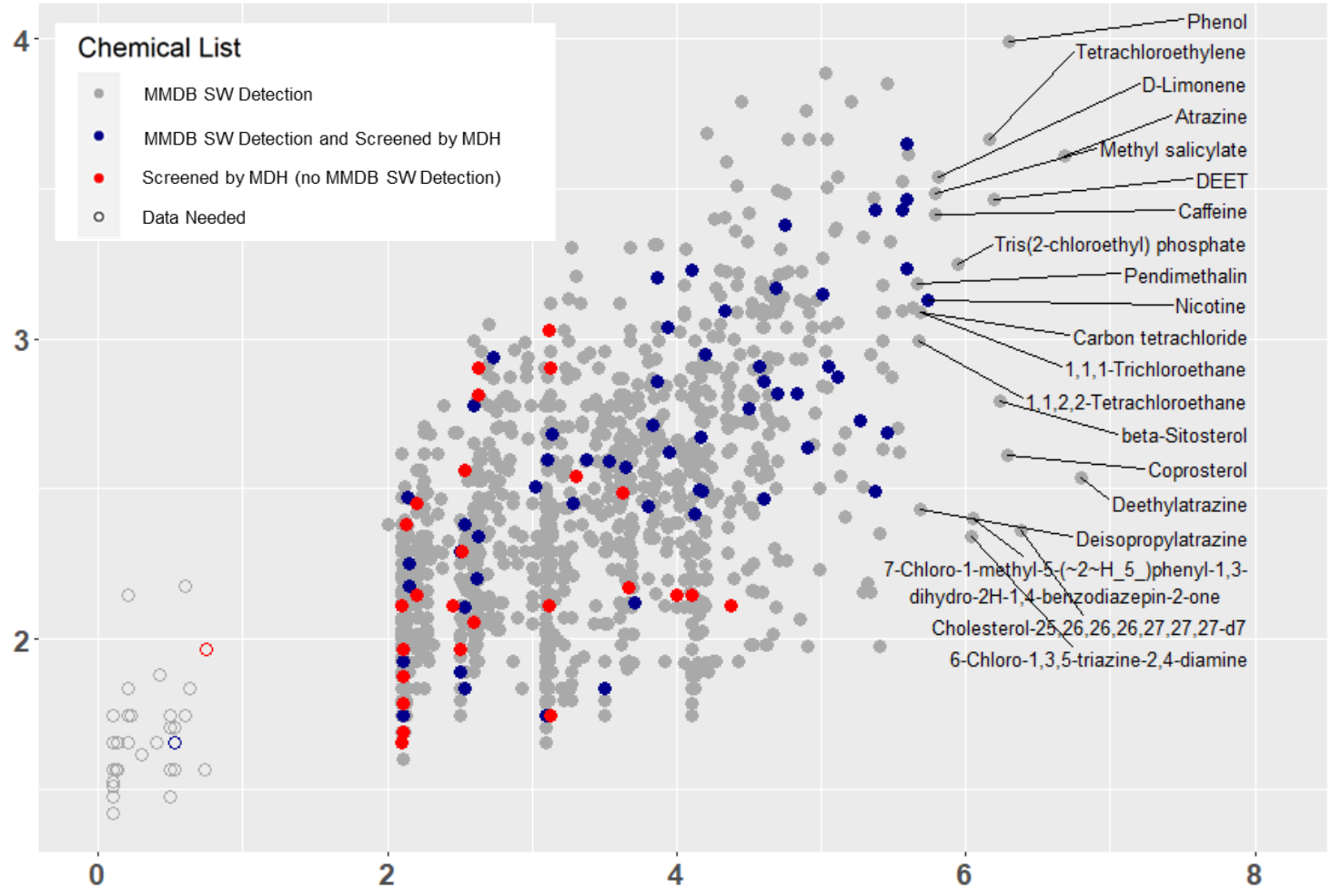
Initial Evaluation of Automated Workflow and Manual Results

- Excellent agreement between scores in **Persistence and Fate** and **Occurrence** domains
- Somewhat poorer alignment in the **Release Potential** domain
- Poor agreement in score adjustments (i.e., detection frequency, human exposure potential)
  - Difference in estimates of detection frequencies in MMDB and MN sources
  - New exposure NAMs from ExpoCast



Scaling up: A Priori Identification of Candidates from Larger Libraries

- CRADA case study chemicals plus 1762 chemicals detected in surface water in MMDB
- All data collection, scoring, and report/table writing were completed in approximately 86 hours
- Examining of highly-scored chemicals could also be used to refine current screening algorithms (e.g., elucidation of additional criteria for refining chemical scores)
- Further ranking with bioactivity-to-exposure NAMs



Next Steps

- Continue evaluations
  - Closer look at differences across the data domains
  - Are there priority data sources to be added?
- Incorporation of additional data streams into workflow
  - Integration into workflow of MN-specific water measurement database
  - Additional exposure NAMs, including machine-learning models for media occurrence built using the MMDB monitoring descriptors
- Manuscript development
  - Integrating Tox21/ToxCast potency data, httk, and HT exposure predictions for chemicals that rank high in exposure score
- ORD toxicologists are working with MN to gather hazard data (including data from NAMs) for data-poor nominated CECs and those identified as having high exposure potential

Impact

- This workflow allows MDH health scientists to accelerate and expand exposure screening evaluations, freeing resources to complete the more complex aspects of exposure assessment
- Large libraries of chemicals relevant to MDH can be rapidly screened for *a priori* identification of new potential nominees (something that has never been feasible)
- The implemented workflow has formed a basis for exposure screening under another MDH regulatory program, the **Toxic Free Kids** initiative (implementation now underway, MDH concurrently developing screening algorithm in collaboration with ORD)
- ORD has had initial conversation with Office of Water to discuss potential use of a similar automated workflow approach for future CCL phases

CRADA Team
(Exposure Forecasting)

EPA-ORD

Kathie Dionisio
Jill Franzosa
Kristin Isaacs
Jason Lambert
Monica Linnenbrink
Katie Paul-Friedman
Amar Singh
Jonathan Taylor Wall
Antony Williams

MDH

Christopher Greene
Helen Goeden
David Bell
Sarah Johnson
James Jacobus

ExpoCast Project (Exposure Forecasting)

CCTE

Linda Adams
Miyuki Breen*
Alex Chao*
Dan Dawson*
Mike Devito
Kathie Dionisio
Christopher Ecklund
Marina Evans
Peter Egeghy
Michael-Rock Goldsmith
Chris Grulke
Mike Hughes
Kristin Isaacs
Richard Judson
Jen Korol-Bexell*
Anna Kreutz*
Charles Lowe*
Seth Newton
Katherine Phillips
Paul Price
Tom Purucker
Ann Richard
Caroline Ring
Marci Smeltz*
Jon Sobus
Risa Sayre*
Mark Sfeir*
Mark Strynar
Zach Stanfield*
Rusty Thomas
Mike Tornero-Velez
Elin Ulrich
Dan Vallero
John Wambaugh
Barbara Wetmore
Antony Williams

CEMM

Xiaoyu Liu

CPHEA

Jane Ellen Simmons

CESER

David Meyer
Gerardo Ruiz-Mercado
Wes Ingwersen

***Trainees**

Collaborators

Arnot Research and Consulting
Jon Arnot
Johnny Westgate
Institut National de l'Environnement et des Risques (INERIS)
Frederic Bois
Integrated Laboratory Systems
Kamel Mansouri
National Toxicology Program
Steve Ferguson
Nisha Sipes
Ramboll
Harvey Clewell
ScitoVation
Chantel Nicolas
Silent Spring Institute
Robin Dodson
Southwest Research Institute
Alice Yau
Kristin Favela
Summit Toxicology
Lesa Aylward
Technical University of Denmark
Peter Fantke
Tox Strategies
Miyoung Yoon
Unilever
Beate Nicol
Cecilie Rendal
Ian Sorrell
United States Air Force
Heather Pangburn
Matt Linakis
University of California, Davis
Deborah Bennett
University of Michigan
Olivier Jolliet
University of Texas, Arlington
Hyeong-Moo Shin