http://www.orcid.org/0000-0002-2668-4821

Development of a Tool for Integrating Traditional and New Approach Methodologies (NAMs) for Chemical Safety Decisions

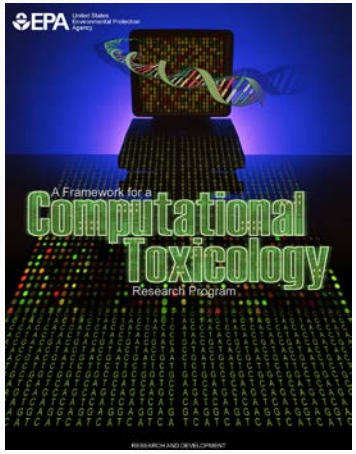
Antony Williams, Richard Judson, Chris Grulke and Rusty Thomas

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

March 2018

ACS Spring Meeting, New Orleans

National Center for Computational Toxicology



- National Center for Computational Toxicology (NCCT) established in 2005 to integrate:
  - High-throughput and high-content technologies
  - Molecular biology
  - Data mining and statistical modeling
  - Computational biology and chemistry
- Outputs: a lot of data, models, algorithms and software applications
- Open Data – we want scientists to interrogate it, learn from it, develop understanding

Background

- EPA must designate a set of high-priority chemicals for detailed risk assessment.

- NCCT is building a web-based tool as *one* approach to help guide that selection

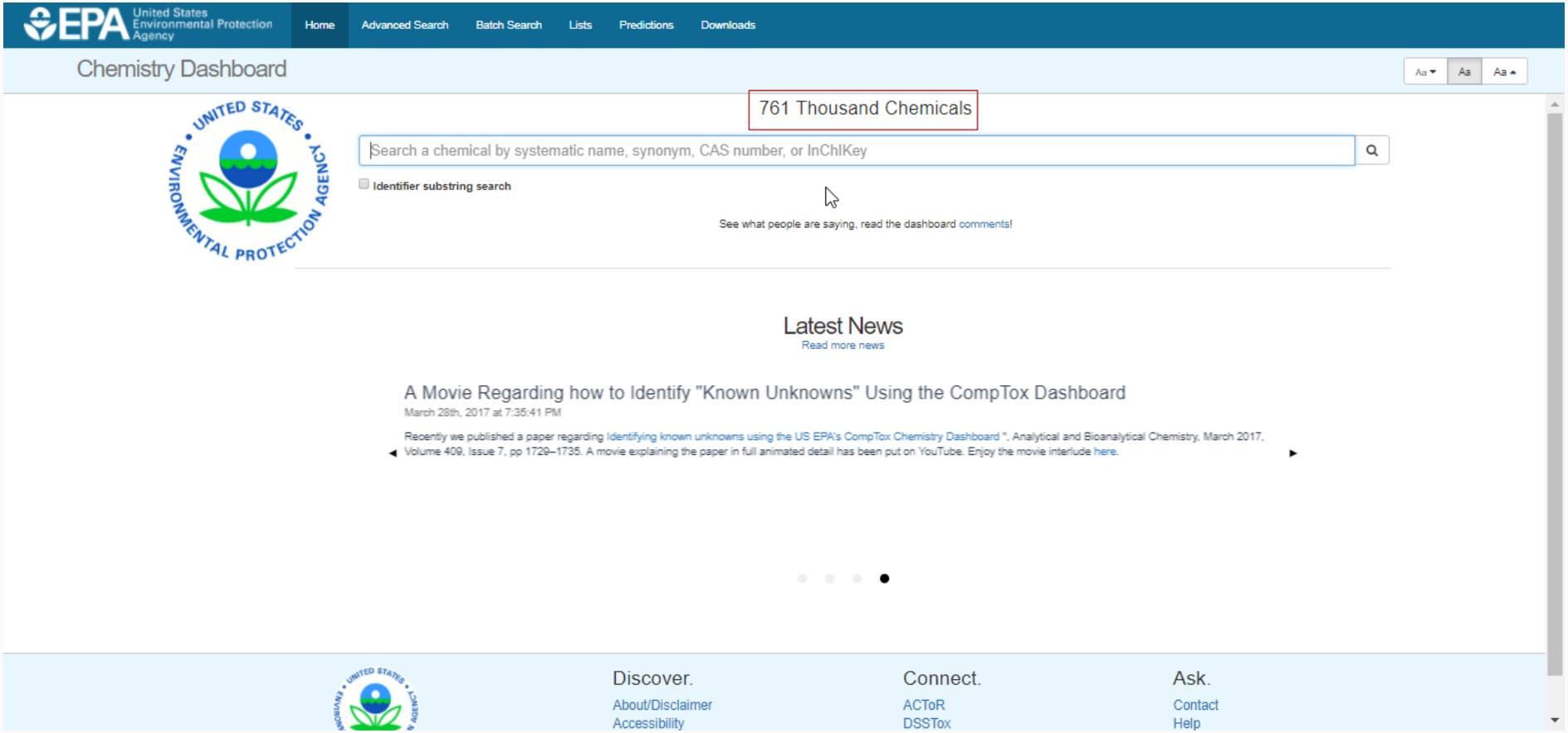
The CompTox Dashboard

- NCCT has lots of data to build relevant tools and approaches. The source data are available via the CompTox Dashboard

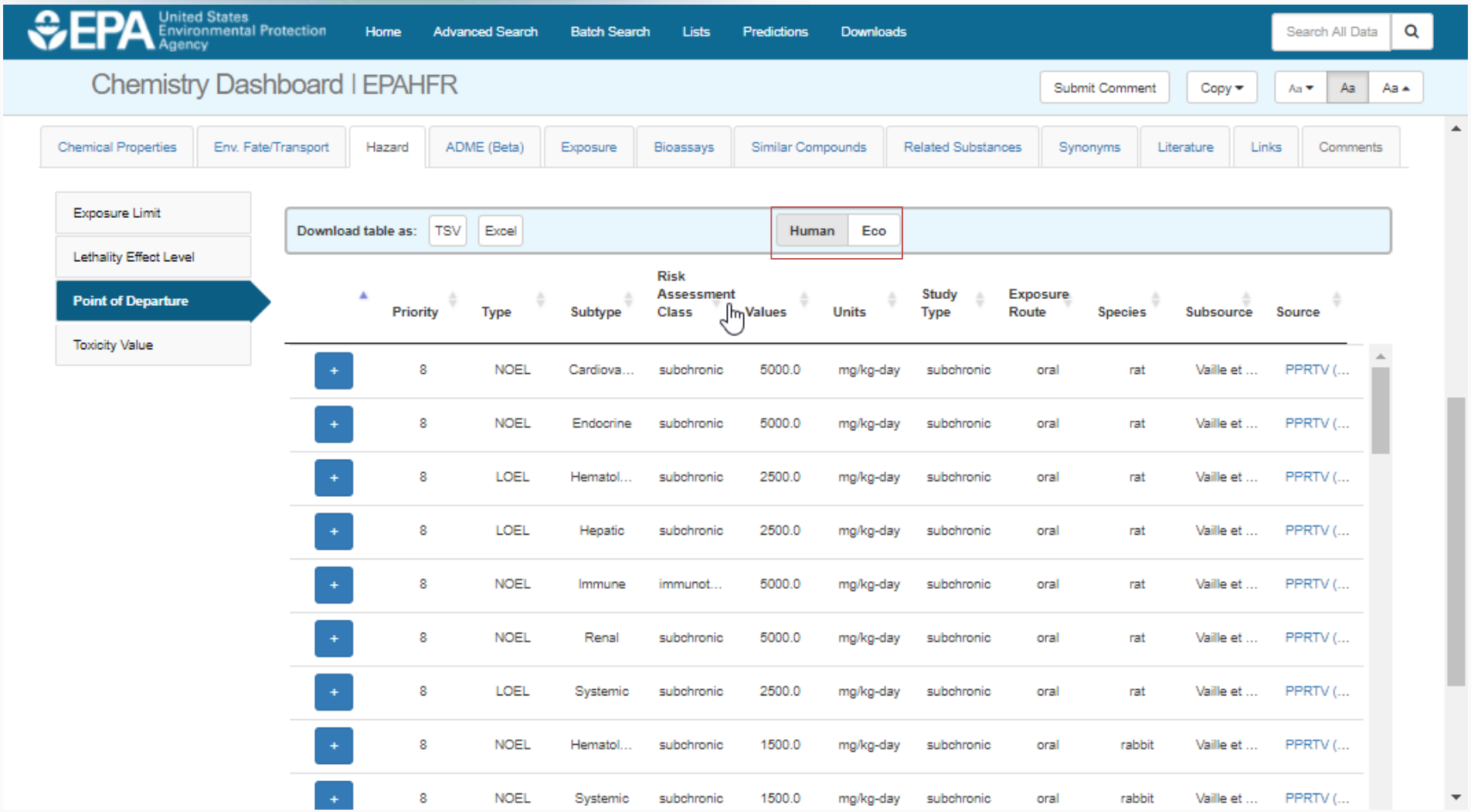
- A **publicly accessible website** delivering access:
  - ~760,000 chemicals with related property data
  - Experimental and predicted **physicochemical property data**
  - Integration to "**biological assay data**" for 1000s of chemicals
  - Data regarding **chemical exposure**
  - Links to other agency websites and public data resources
  - "Literature" searches for chemicals using public resources
  - "Batch searching" for thousands of chemicals

CompTox Dashboard

https://comptox.epa.gov/dashboard



Chemical Hazard Data



	Priority	Type	Subtype	Risk Assessment Class	Values	Units	Study Type	Exposure Route	Species	Subsource	Source
+	8	NOEL	Cardiova...	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Endocrine	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Hematol...	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Hepatic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Immune	immunot...	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Renal	subchronic	5000.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	LOEL	Systemic	subchronic	2500.0	mg/kg-day	subchronic	oral	rat	Vaille et ...	PPRTV (...
+	8	NOEL	Hematol...	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaille et ...	PPRTV (...
+	8	NOEL	Systemic	subchronic	1500.0	mg/kg-day	subchronic	oral	rabbit	Vaille et ...	PPRTV (...

In Vitro Bioassay Screening

ToxCast and Tox21

Sources of Exposure to Chemicals

- **Product & Use Categories**
- Chemical Weight Fraction
- Chemical Functional Use
- Monitoring Data
- Exposure Predictions
- Production Volume

osure | Bioassays | Similar Compounds | Related Substances | Synonyms | Literature | Links | Comments

Product & Use Categories (PUCs)

Categorization type	Number of Unique Products
PUC	288
PUC	206
PUC	117
PUC	107
PUC	107
PUC	101
PUC	101
PUC	90
PUC	89

Approach

- Use **available data** for *in vivo, in vitro*, exposure and chemical property data
- Develop **scoring schemes** to merge different types of data
- Develop methods to note or **fill data gaps**
- Make data, scores, prioritization ranking available in the **web-based tool**

Potential Data Sources

***In Vivo* Human Hazard:**

- Mammalian toxicity studies – guideline-like, use POD
- System-specific *in vivo* data (Cancer, developmental)
- Models (QSAR) to predict POD and organ-specific effects
- Genotoxicity
- *In vitro*-derived endocrine disruption and neurotoxicity models

***In Vivo* Eco Hazard**

- Aquatic *in vivo* studies – POD
- Models (QSAR) of POD

Potential Data Sources

EPA United States Environmental Protection Agency

Human Exposure

- Data on production volume and releases
- Quantitative biomonitoring data
- Predictions of oral and inhalation exposure

Eco Exposure

- Biomonitoring data
- Predictions of water concentrations

Physchem

- Persistence and Bioaccumulation models (OPERA)


Mansouri et al. J Cheminform (2018) 10:10
https://doi.org/10.1186/s13321-018-0263-1

Journal of Cheminformatics

RESEARCH ARTICLE Open Access

OPERA models for predicting physicochemical properties and environmental fate endpoints

Kamel Mansouri[1,2,3*], Chris M. Grulke[1], Richard S. Judson[1] and Antony J. Williams[1]

Aggregating Data

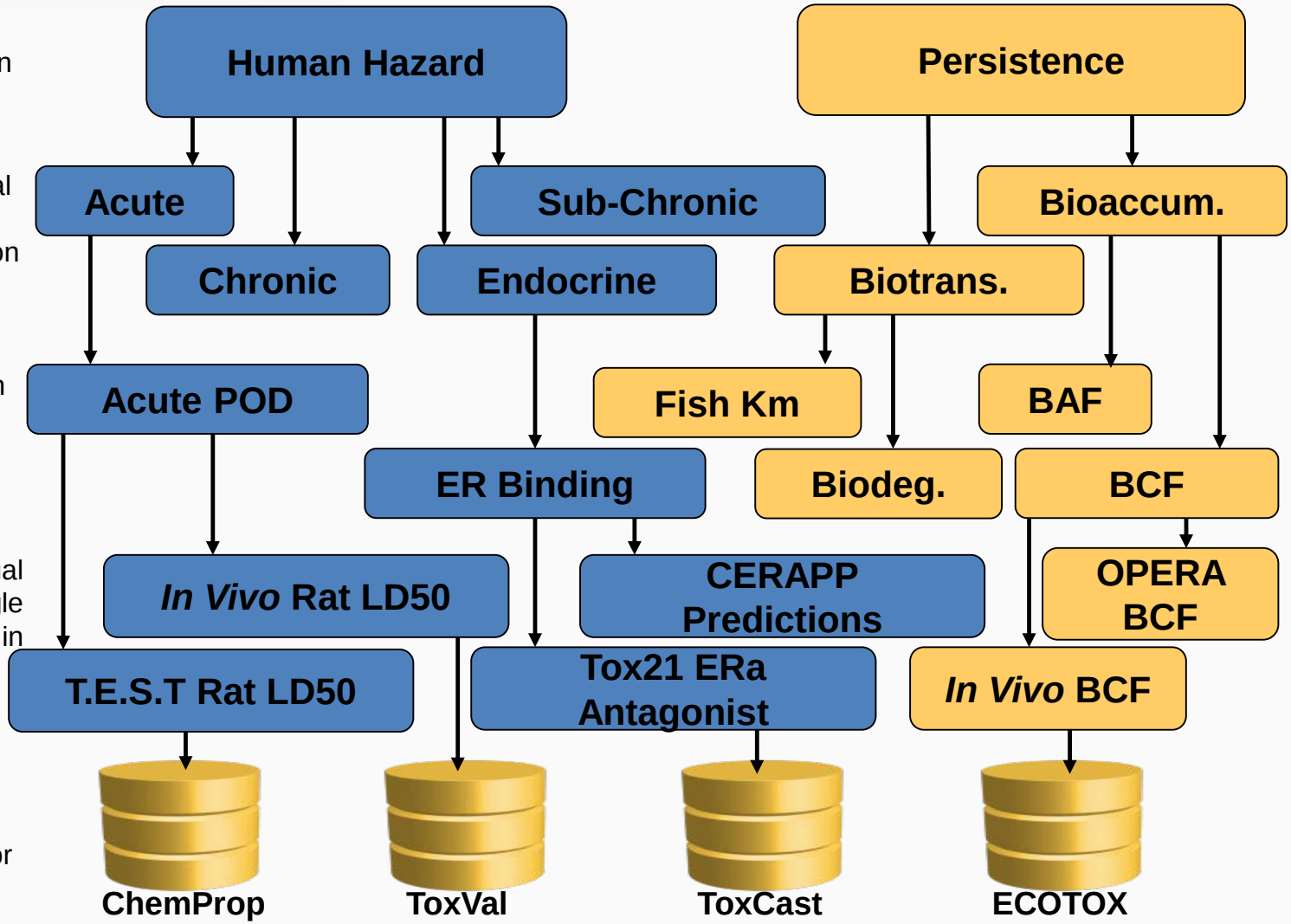
DOMAINS: Top level aggregation for reporting in the web-interface

SUB-DOMAINS: Additional aggregation layer(s) to support flexible aggregation workflows

ENDPOINTS: Aggregation of data streams that are intended to estimate the same value

DATA STREAMS: Individual values coming from a single source that are of interest in prioritization.

DATABASES: Active database projects at US EPA can act as sources for data streams



Scoring approaches

- For each chemical, each domain receives a score of 1 (Low), 2 (Moderate), or 3 (High) concern
- **Hazard score** = maximum of human and eco hazard scores
- **Exposure score** = maximum of human and eco exposure scores
- **Total score** = hazard score + exposure score + physchem score
- If no data is available for a domain, it is given the "missing data score", currently 1 (Low)
- Scoring can include or exclude NAM

Implemented Scoring Methods

Method 1

- Maximum score from human and eco hazard: 1 – 3
- Maximum score from human and eco exposure: 1 – 3
- Maximum score from persistence/ bioaccumulation (P/B): 1 – 3
- No NAM
- Add hazard, exposure, and P/B
- Categorical bins
    - High: 7-9
    - Moderate: 5-6
    - Low: 3-4

Method 2: NAM Equal

- Same as Method 1 except NAM is incorporated with equal weighting in all domains
- Add hazard, exposure, and P/B
- Categorical bins
    - High: 7-9
    - Moderate: 5-6
    - Low: 3-4

Method 3: NAM Deferential

- Same as Method 1 except human hazard NAM is incorporated in the absence of traditional *in vivo* studies
- In other domains, NAM is given equal weight.
- Add hazard, exposure, and P/B
- Categorical bins
    - High: 7-9
    - Moderate: 5-6
    - Low: 3-4

Method 4: Sum of Scores

- Sum all components (incl. NAM) from human and eco hazard
- Sum all components (incl. NAM) from human and eco exposure
- Sum all components (incl. NAM) from persistence/ bioaccumulation
- Add hazard, exposure, and P/B
- Categorical bins
    - High: >30
    - Medium: 10-30
    - Low: ≤10

Method 5: H/BER*

- Ratio of the minimum effect level from *in vivo* toxicity studies or the quantitative human hazard NAM data divided by the maximum oral exposure
- Categorical bins
    - High: $\leq 10^4$
    - Medium: $10^4 – 10^6$
    - Low: $\geq 10^6$

*Hazard/Bioactivity Exposure Ratio

Overall Scoring Page

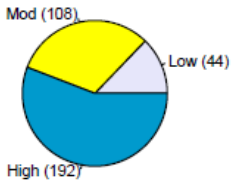
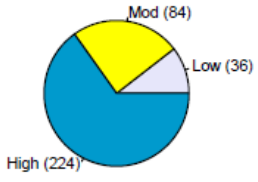
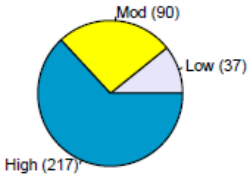
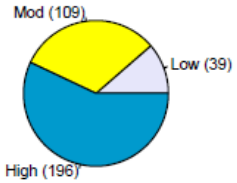
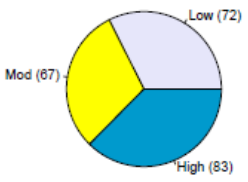
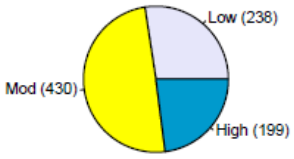
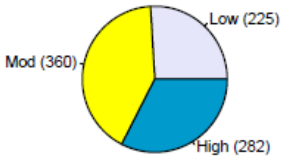
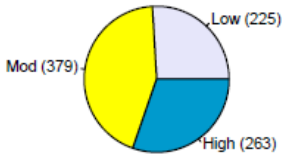
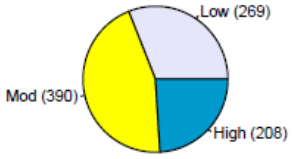
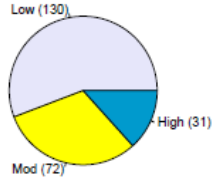
Cutoffs | Domain Help | Human Hazard Prioritization | Eco Hazard Prioritization | Human Exposure Prioritization | Eco Exposure Prioritization

Physchem Prioritization | Overall Prioritization | Data Coverage Summary | Distribution Graphs

Show 25 entries Search:

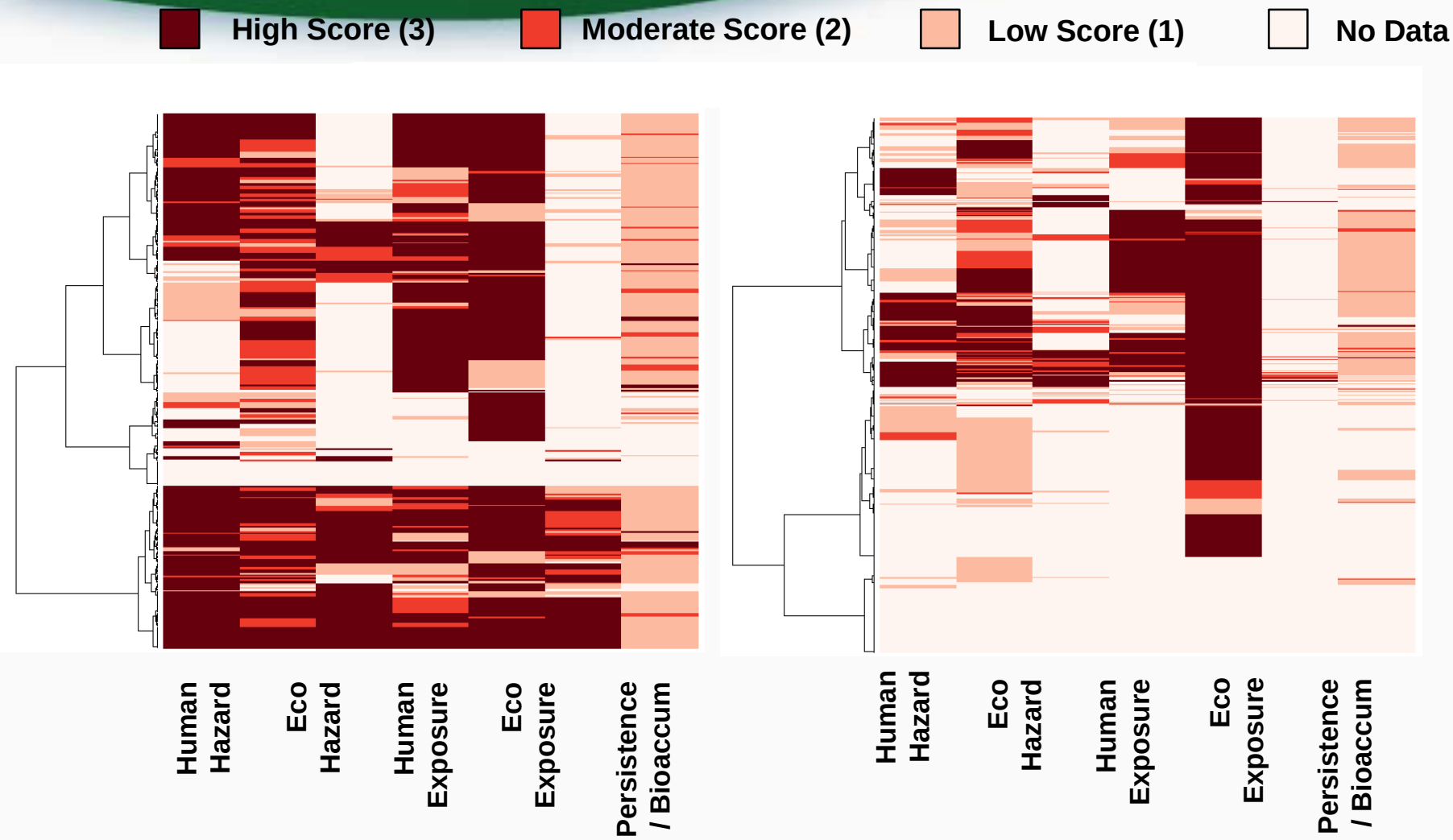
DSSTox_ID	CASRN	Name	Method 1 score	Method 1 bin	Method 2 score	Method 2 bin	Method 3 score	Method 3 bin	Method 4 score	Method 4 bin
DTXSID124356	1234-56-7	Name1	5	Moderate (ExeHzeHzh)	6	Moderate (ExeHzeHzh)	6	Moderate (ExeHzeHzh)	25	Moderate (ExeHzeHzh)
DTXSID124357	1234-56-8	Name2	5	Moderate (PbExeHzeHzh)	7	High (PbExeHzeHzh)	7	High (PbExeHzeHzh)	24	Moderate (PbExeHzeHzh)
DTXSID124358	1234-56-9	Name3	5	Moderate (ExeHzeHzh)	6	Moderate (ExeHzeHzh)	6	Moderate (ExeHzeHzh)	41	High (ExeHzeHzh)
DTXSID124359	1234-56-10	Name4	5	Moderate (ExeHzeHzh)	5	Moderate (ExeHzeHzh)	5	Moderate (ExeHzeHzh)	13	Moderate (ExeHzeHzh)
DTXSID124360	1234-56-11	Name5	5	Moderate (PbExeHzeHzh)	5	Moderate (PbExeHzeHzh)	5	Moderate (PbExeHzeHzh)	20	Moderate (PbExeHzeHzh)
DTXSID124361	1234-56-12	Name6	5	Moderate (ExeHzeHzh)	5	Moderate (ExeHzeHzh)	5	Moderate (ExeHzeHzh)	35	High (ExeHzeHzh)
DTXSID124362	1234-56-13	Name7	5	Moderate (PbExeHze)	5	Moderate (PbExeHze)	5	Moderate (PbExeHze)	21	Moderate (PbExeHze)
DTXSID124363	1234-56-14	Name8	6	Moderate (PbExeHze)	6	Moderate (PbExeHze)	6	Moderate (PbExeHze)	16	Moderate (PbExeHze)
DTXSID124364	1234-56-15	Name9	5	Moderate (PbExeHze)	5	Moderate (PbExeHze)	5	Moderate (PbExeHze)	26	Moderate (PbExeHze)
DTXSID124365	1234-56-16	Name10	6	Moderate (PbExeHze)	6	Moderate (PbExeHze)	6	Moderate (PbExeHze)	23	Moderate (PbExeHze)
DTXSID124366	1234-56-17	Name11	5	Moderate	5	Moderate	5	Moderate	17	Moderate

Fraction of chemicals in each bin

	General	NAM Equal	NAM Differential	Sum of Scores	H/BER
Set 1 (344)					 (222/344)
Set 2 (867)					 (233/867)

High Pre-Priority Moderate Pre-Priority Low Pre-Priority

Heat Maps



Heatmaps showing the domain-specific scores for chemical sets.

Registration and Curation of Chemicals

- Consolidation and registration of the original chemical lists into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs

Names to CASRN Mappings

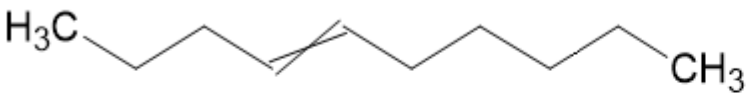
Substance Mapping

(1 of 66) 1 2 3 4 5 6 7 8 9 10 25

Source Casrn	Source Name	Hit Substance_ID	Hit Casrn	Hit Name	
19398-89-1	4-Decene	DTXSID50876156	19689-18-0	4-Decene	Remove Validation
112926-00-8	silica gel, cryst. -free	DTXSID9029851	112926-00-8	Hydrated silica	Remove Validation
124-28-7	1-Octadecanamine, N,N-dimethyl-	DTXSID4027026	124-28-7	N,N-Dimethyl-1-octadecanamine	Remove Validation
1330-43-4	Boron sodium oxide	DTXSID2034388	1330-43-4	Sodium tetraborate	Remove Validation
13492-26-7	Mono- and di-potassium salts of phosphorous acid	DTXSID9035961	13492-26-7	Phosphonic acid, potassium salt (1:2)	Remove Validation
135-37-5	Glycine, N-(carboxymethyl)-N-(2-hydroxyethyl)-, disodium salt	DTXSID8042008	135-37-5	Ethanoldiglycine disodium salt	Remove Validation

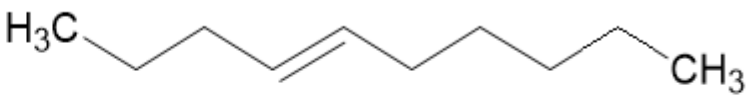
Subtleties

“4-Decene”



19398-89-1

E/Z-stereochemistry



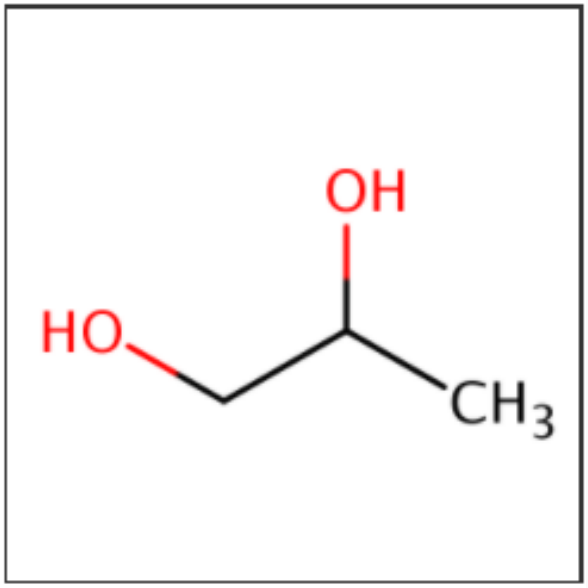
19689-18-0

E-stereochemistry

Registration and Curation

- Registration of the chemical list into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs
  - CASRN checking – Active, Alternate and Deleted

CAS Registry Numbers



1,2-Propylene glycol	Propylenglycol
Propane-1,2-diol	Sentry Propylene Glycol
1,2-Propanediol	Trimethyl glycol
57-55-6 Active CAS-RN	Ucar 35
alpha-Propylene glycol	a-Propylene glycol
(+/-) 1,2-Propanediol	alpha-propyleneglycol
(RS)-1,2-Propanediol	methyl glycol
dl-Propylene glycol	methylethyl glycol
3-01-00-02142 Beilstein Registry Number	1194046-20-2 Deleted CAS-RN
1,2-Propanediol	190913-75-8 Deleted CAS-RN
(.+-.)-1,2-Propanediol	4254-16-4 Deleted CAS-RN
(.+-.)-Propylene glycol	63625-56-9 Deleted CAS-RN

Registration and Curation

- Registration of the chemical list into the underlying database (DSSTox)
- Careful (and time-consuming) curation
  - Confirming mappings of chemical names and CASRNs
  - CASRN checking – Active, Alternate and Deleted
  - Misspellings, alternative synonyms, misassociations

Alternative Synonyms

Substance Mapping

(1 of 66) 1 2 3 4 5 6 7 8 9 10 25

Source Casrn	Source Name	Hit Substance_ID	Hit Casrn	Hit Name	
19398-89-1	4-Decene	DTXSID50876156	19689-18-0	4-Decene	Remove Validation
112926-00-8	silica gel, cryst. - free	DTXSID9029851	112926-00-8	Hydrated silica	Remove Validation
124-28-7	1-Octadecanamine, N,N-dimethyl-	DTXSID4027026	124-28-7	N,N-Dimethyl-1-octadecanamine	Remove Validation
1330-43-4	Boron sodium oxide	DTXSID2034388	1330-43-4	Sodium tetraborate	Remove Validation
13492-26-7	Mono- and di-potassium salts of phosphorous acid	DTXSID9035961	13492-26-7	Phosphonic acid, potassium salt (1:2)	Remove Validation
135-37-5	Glycine, N-(carboxymethyl)-N-(2-hydroxyethyl)-, disodium salt	DTXSID8042008	135-37-5	Ethanoldiglycine disodium salt	Remove Validation

UVCB Chemicals

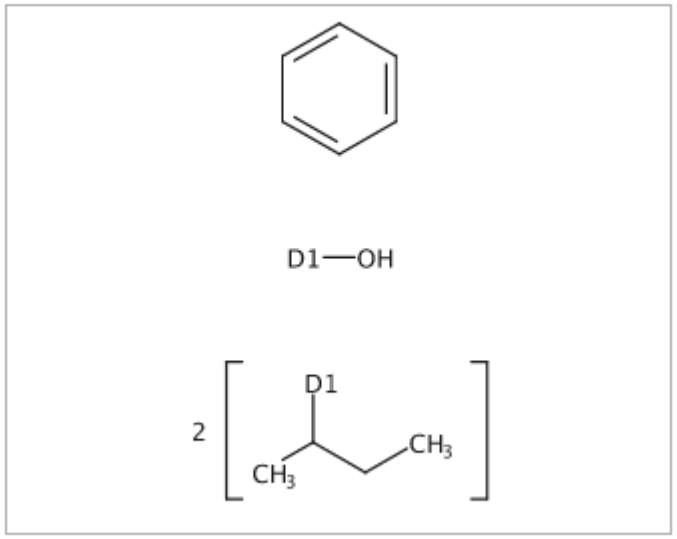
- UVCBs are chemical substances of unknown or variable composition, complex reaction products and biological materials
  - Surfactants (C11-14 linear alkyl sulfonates)
  - Reaction mass of p-t-butylphenyldiphenyl phosphate and bis(p-t-butylphenyl)phenyl phosphate and triphenyl phosphate
  - Almond Oil

Di-sec-butylphenol

CAS Representation

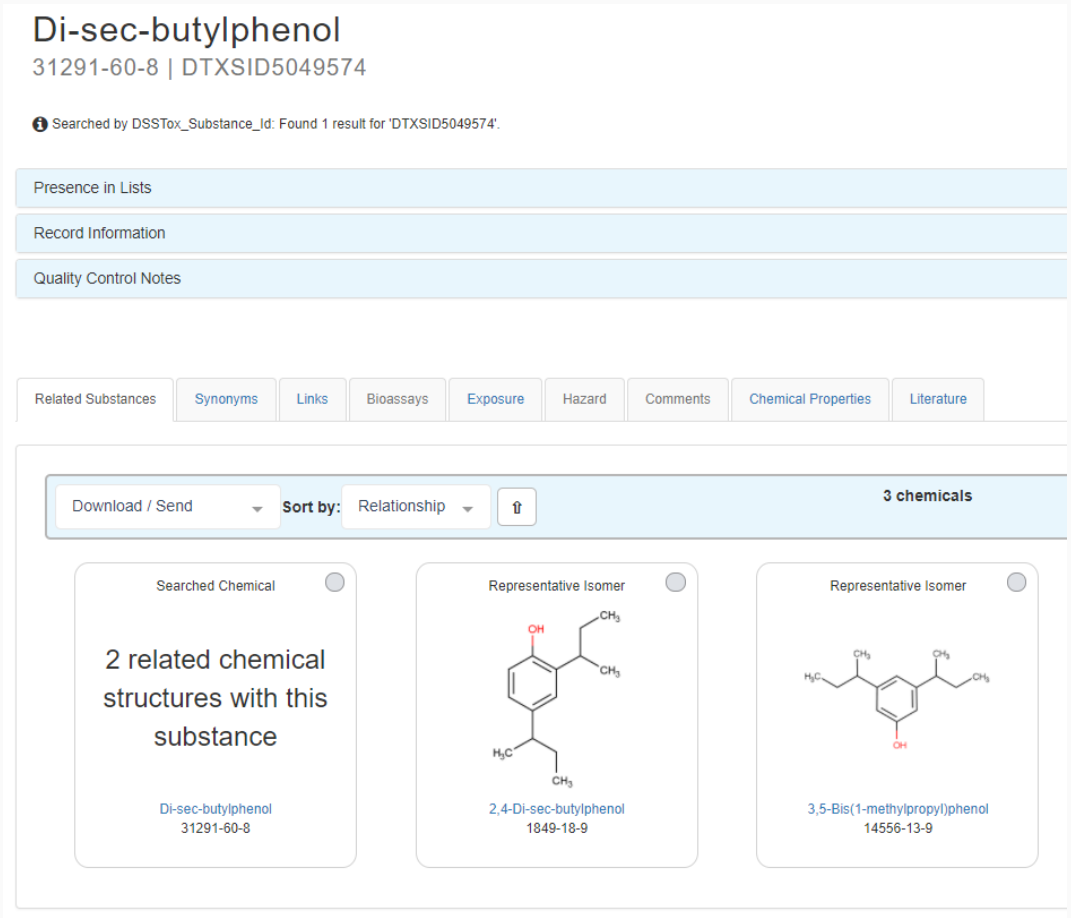
1. 31291-60-8

~31 ~4



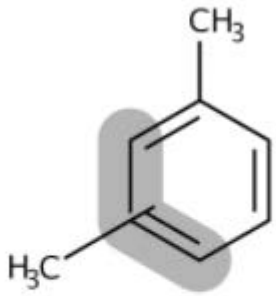
$C_{14}H_{22}O$
Phenol, bis(1-methylpropyl)-

Dashboard Representation

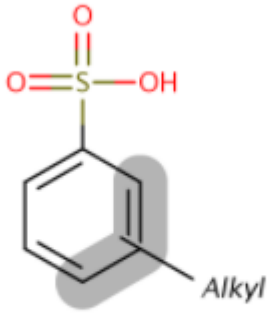


“Markush Structures”

https://en.wikipedia.org/wiki/Markush_structure

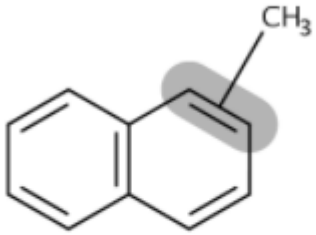


Xylenes
1330-20-7

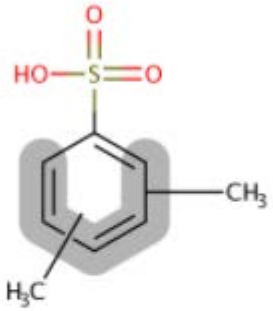


(C10-C16) Alkylbenzenesulfonic acid
68584-22-5

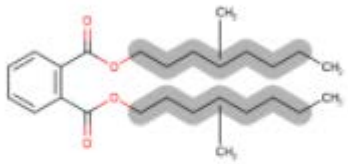
n-Nonylphenol
25154-52-3



Methylnaphthalene
1321-94-4



Sodium xylenesulfonate
1300-72-7



Diisononyl phthalate
28553-12-0

Summary

- RapidTox workflow enabled identification of data that contributed most to candidate selection
- Allowed flexible exploration of prioritization methods
- **Incorporation of NAM data** does result in changes for chemical test sets, either by adding data or by changing the overall bin (Low, Moderate, High)

Acknowledgments

- The NCCT CompTox Dashboard Development Team
- Kamel Mansouri – OPERA models
- Todd Martin – TEST predictions

Contact

Antony Williams
US EPA Office of Research and Development
National Center for Computational Toxicology (NCCT)
Williams.Antony@epa.gov
ORCID: https://orcid.org/0000-0002-2668-4821