OCSPP-TSCA Inventory: Prioritization Proof of Concept

BOSC Meeting
February 3, 2021

The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

Prioritization and Pre-prioritization

- Many organizations face the problem that they have too many chemicals to evaluate given the available resources
- One solution is to use a data-driven approach to prioritize chemicals for detailed assessments
  - OCSPP: TSCA High and low priority chemicals
  - OCSPP: EDSP, potential endocrine disruptors
  - OW: Candidate Contaminant List (CCL)
  - OW: Chemicals in biosolids
  - Health Canada: Domestic Substances List (DSL)
  - Minnesota Department of Health: Chemicals of concern to children

The TSCA Prioritization Problem

- Under the Lautenberg Act, 2016 Amendment to TSCA (*):
  - EPA must establish a risk-based process to determine which chemicals it will **prioritize** for assessment, identifying them as either "high" or low" priority substances.
    - High priority – the chemical may present an unreasonable risk of injury to health or the environment due to potential hazard and route of exposure, including to susceptible subpopulations
    - Low priority – the chemical use does not meet the standard for high-priority
- Assessments for High Priority chemicals must be completed in 3 years, requiring a complete data package at the beginning
- The TSCA Active Inventory contains over 33,000 chemicals
- CompTox resources can provide key inputs to aid this prioritization process

(*) https://www.epa.gov/assessing-and-managing-chemicals-under-tsca/highlights-key-provisions-frank-r-lautenberg-chemical

The CompTox Opportunity

- CCTE staff have been developing resources with data on large numbers of chemicals covering hazard, exposure, toxicokinetics and physico-chemical properties
- Traditional Animal Toxicology: ToxRefDB, ToxValDB
- In Vitro Hazard: ToxCast, specific models for endocrine pathways
- Exposure: ExpoCast (SEEM), CPCat & CPDat, models of use
- Toxicokinetics: HTTK
- PhysChem: OPERA models of physchem and other properties
- Experience building large-scale integrative models

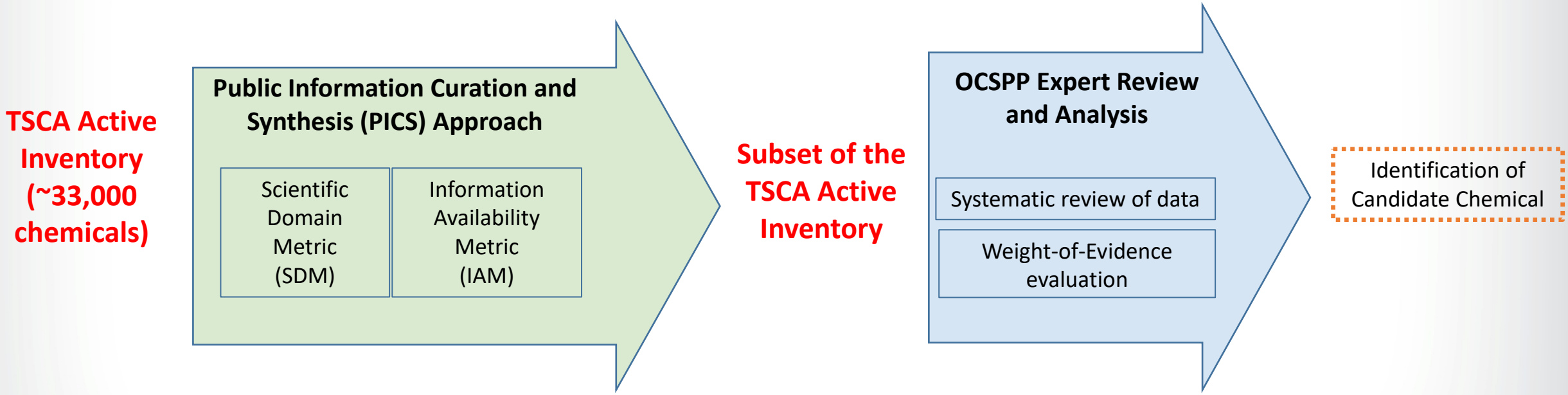
Implementation of the Proof-of-Concept Study

- **Operationalized long-term strategy through development of the <u>Public Information Curation and Synthesis</u> (PICS) approach**
  - Integrates information from a variety of sources to better understand the landscape of publicly available information for large numbers of chemical substances
  - Synthesizes information across key scientific domains used to evaluate chemical risks
  - Consistent with the *Strategic Plan to Promote the Development and Implementation of Alternative Test Methods Within the TSCA Program to* integrate NAMs to fill gaps when traditional testing data are not available

Defining Intended Application of PICS Approach

- **The PICS approach was intended to:**
  - Understand the landscape of publicly-available information on the over 33,000 substances on the active inventory
  - Provide a transparent and reproducible process for integrating available information and identifying potential information gaps
  - Increase efficiency and manage workload by focusing expert review on substances that may have a greater potential for selection as high- or low-priority candidates
  - Create a flexible and sustainable process that can adapt to scientific advances and continual generation of new safety-related information
  - Organize the process into modular workflows that can be readily updated or adapted to address prioritization needs under other mandates
- **The PICS approach was <u>not</u> intended to:**
  - Replace the formal TSCA prioritization or risk evaluation processes
  - Create a ranked list of substances
  - Signal that the EPA has concerns with particular substances or categories of substances
  - Supplant expert judgment and review
  - Utilize confidential business information
  - Incorporate systematic review of information to address study and data quality

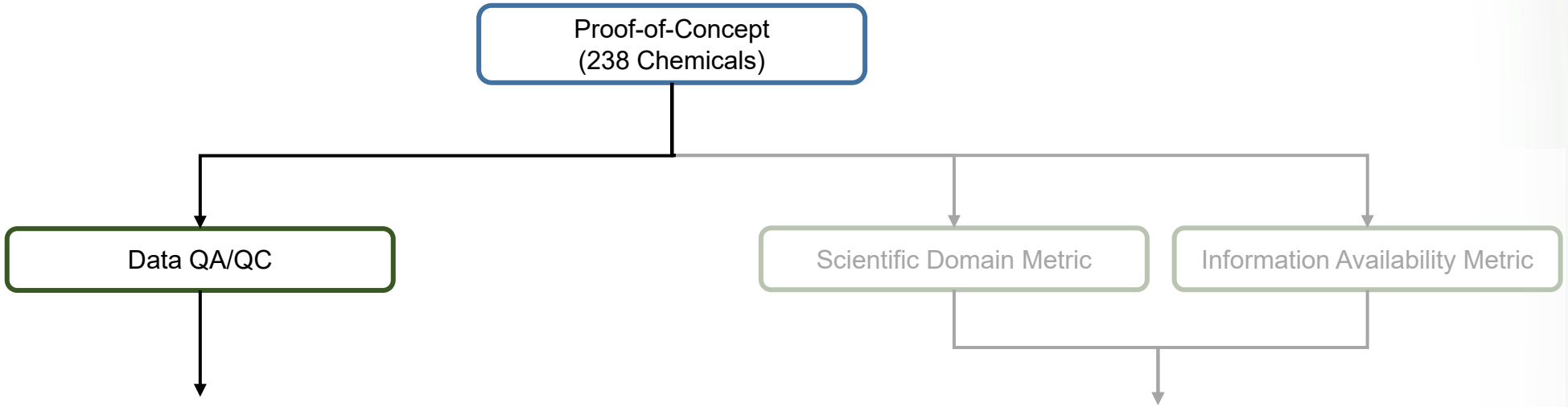
Schematic of PICS Approach Within the Candidate Selection Process



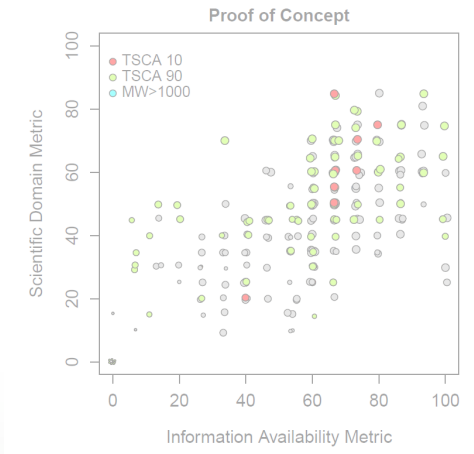
Proof-of-Concept Chemicals (POC 238)

- The process was carried out on the complete TSCA Active Inventory
- For illustration, a total of 238 substances selected from the curated, non-confidential active TSCA inventory
- Selection based on the following:
  - Proposed set of 20 high- and 20 low-priority candidate substances
  - Substances from the 2014 update to the TSCA Work Plan
  - Substances with known relevance to each of the scientific domains
  - Subset of chemical substances listed in the FDA's Substances Added to Food inventory and EPA's Safer Chemical Ingredients List (SCIL)

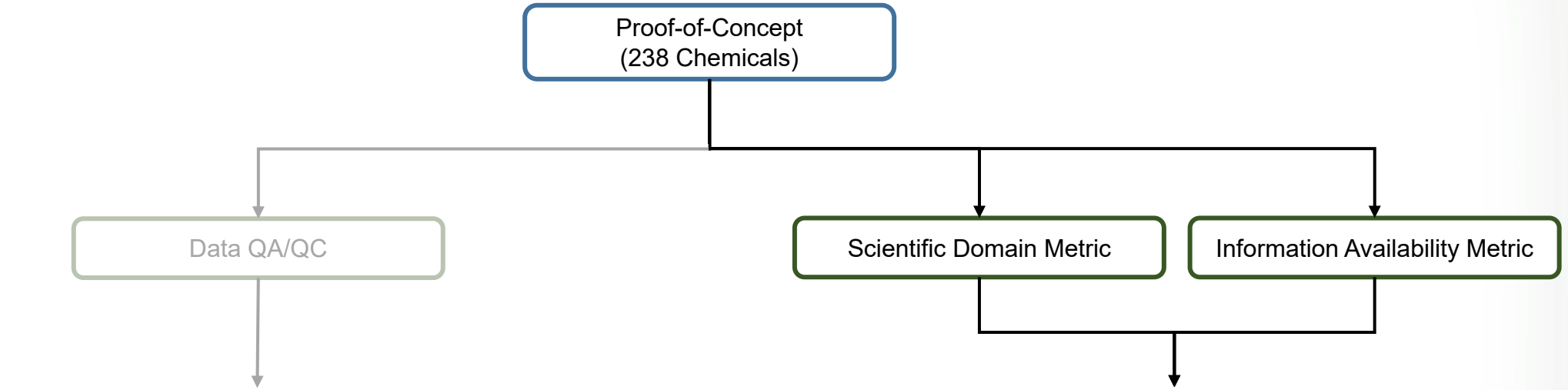
Proof-of-Concept: Data QA/QC



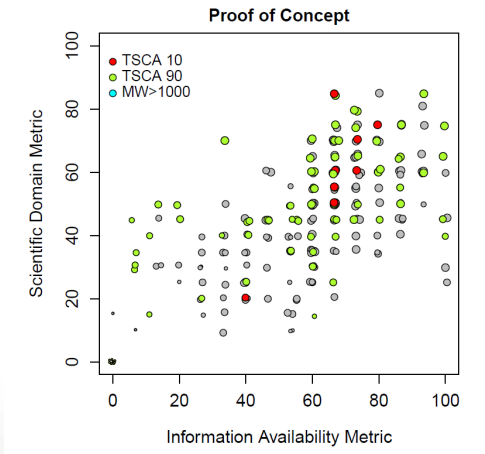
- Specific data domain and data source error rates
- Data QA plan for TSCA active inventory
- FTE estimates for data QC
- QA of massive amounts of data is an ongoing challenge



Proof-of-Concept: Metrics



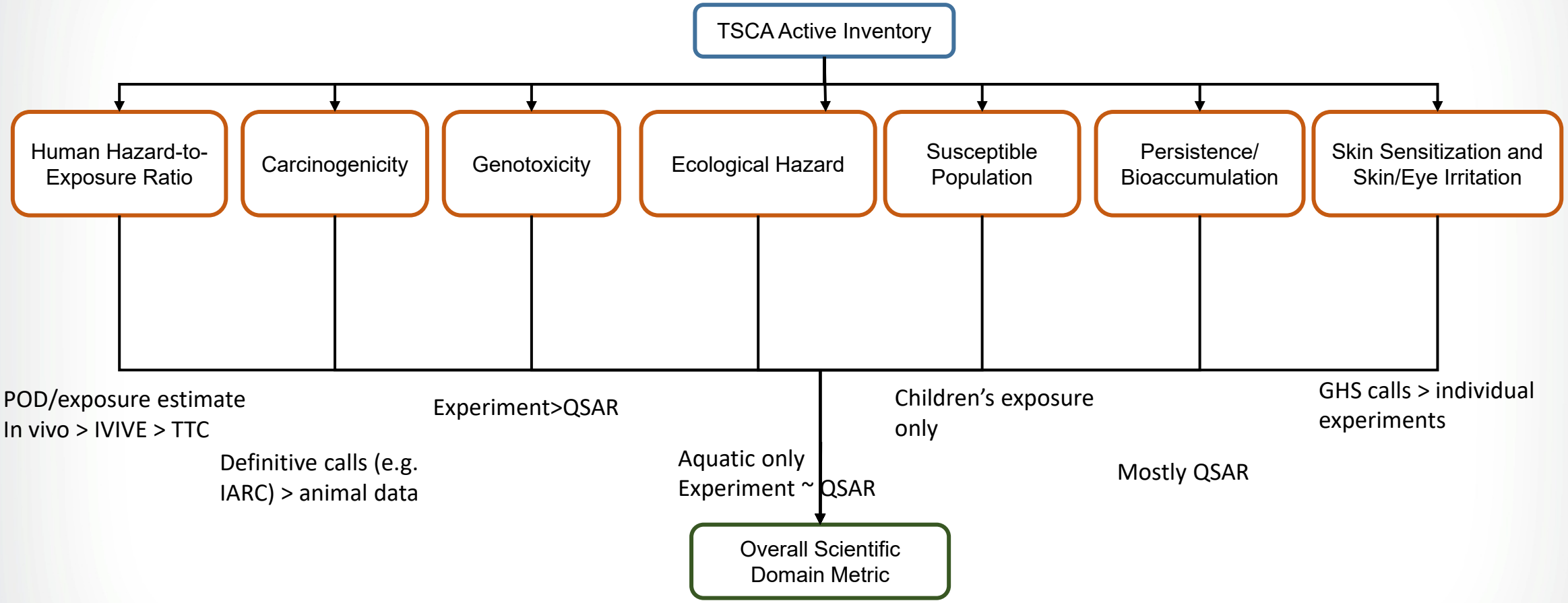
- Specific data domain and data source error rates
- Data QA plan for TSCA active inventory
- FTE estimates for data QC
- QC Tool (beta)



Scientific Domain Metric

- Seven scientific domains were selected based on:
  - Previous use in TSCA prioritization activities (i.e., TSCA workplan)
  - Statutory language in the amended TSCA
  - Consultation with OCSPP management and staff
- Tiered workflows for each scientific domain designed based on the current state of the science
- The overall scientific domain metric is determined by summing the results from the individual scientific domain workflows

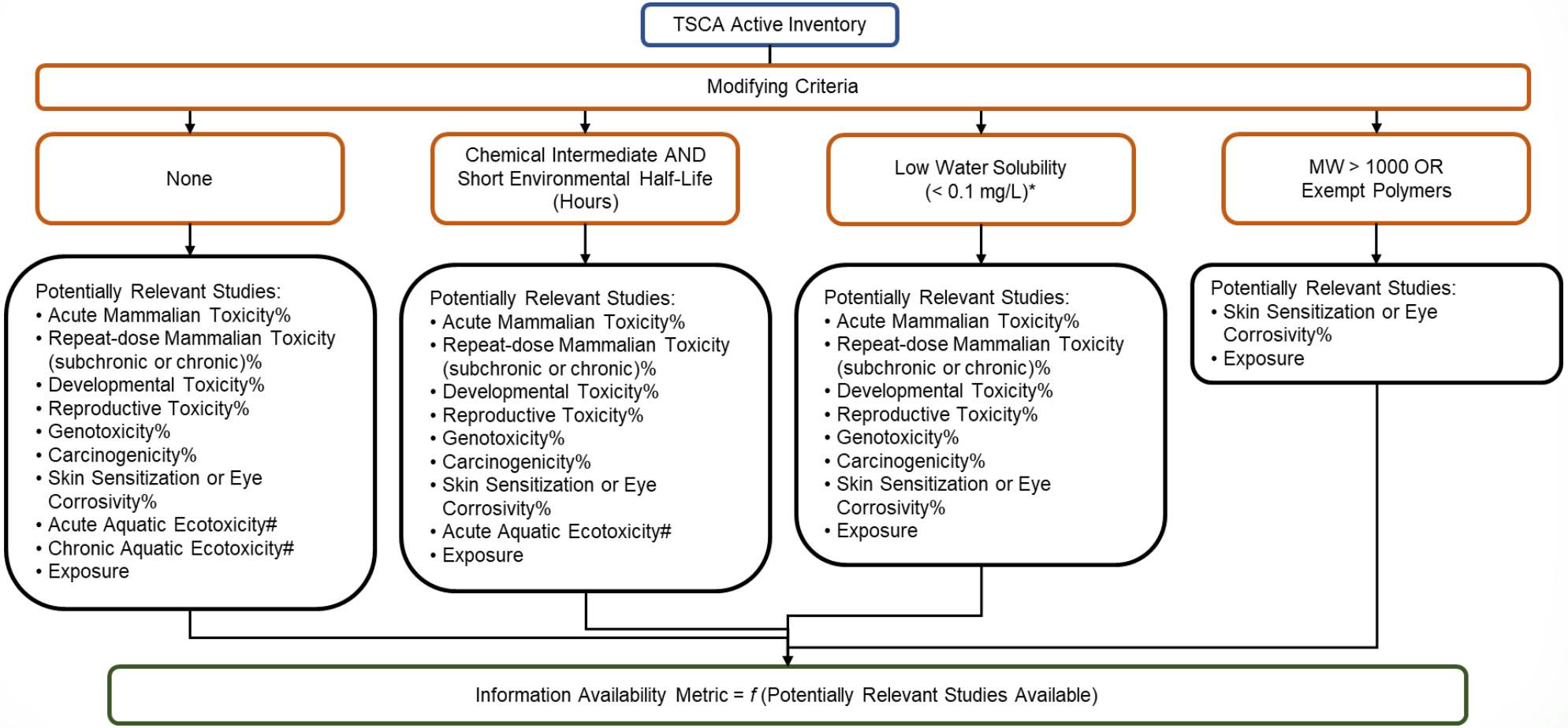
Overall Scientific Domain Metric



Information Availability Metric

- Included in PICS approach to evaluate the amount of information available for use in any future chemical substance risk evaluation
- Needed because detailed risk assessments cannot be carried out without sufficient data
- Based on the potentially relevant information for exposure, human health and ecological toxicity
- Modifying criteria (based on OPPT new chemicals program and consultation with OPPT technical staff) applied to make the score context-specific
- Incorporates “information gathering flags” to highlight data types used in specific scientific domain metrics as well as possible data gaps

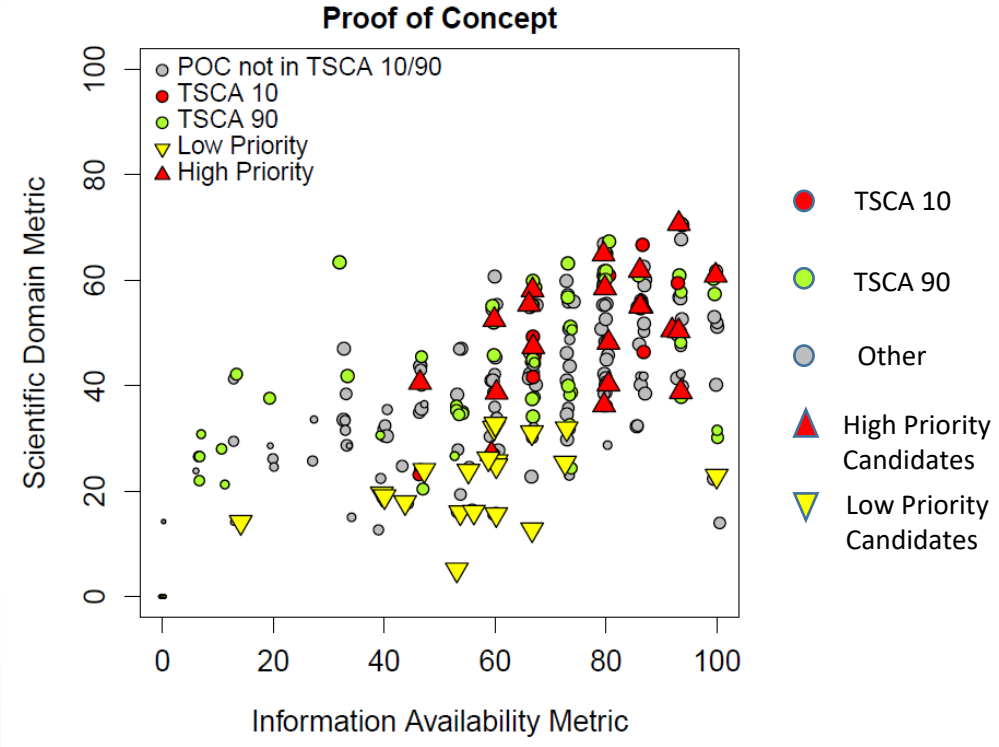
Information Availability Metric



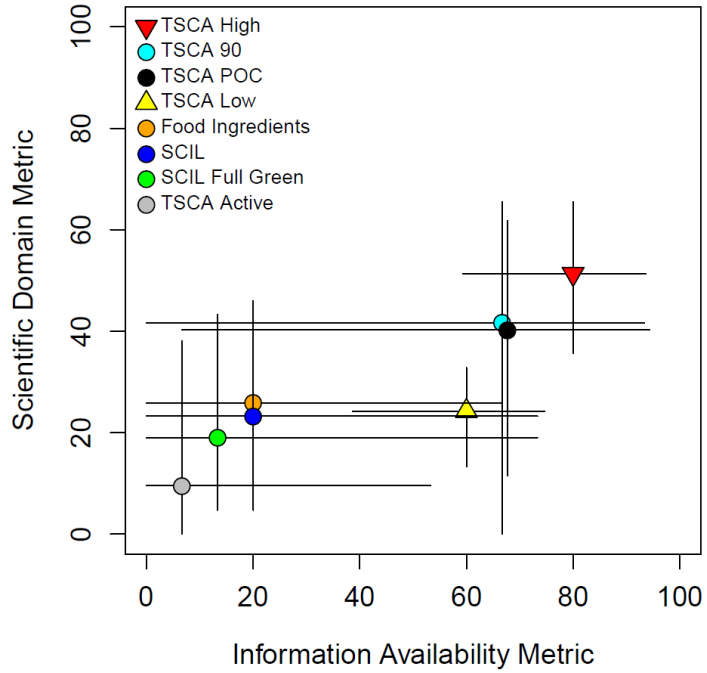
*Criteria based on Sustainable Futures Manual (EPA-748-B12-001); #includes multiple trophic level data; %Not required if chemical has an authoritative human hazard assessment

Proof-of-Concept Results

- High priority chemicals have larger scientific domain scores than the low priority
- “Safe” Chemical sets (e.g. food ingredients) tend to have low scientific domain scores
- The POC chemicals have larger than average information availability



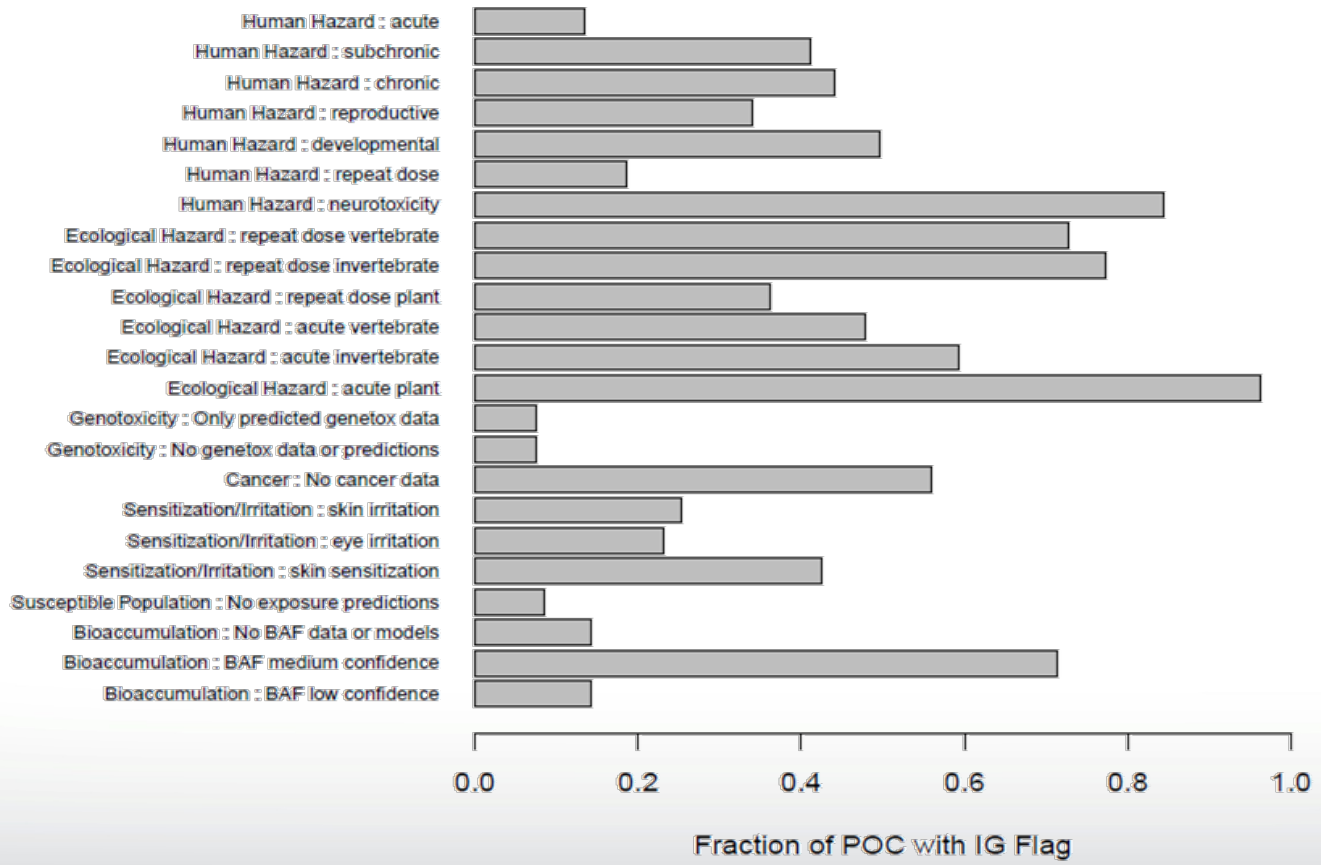
Information availability vs. scientific domain metrics for the POC238 set of chemical substances. Positions of points are staggered for ease of visualization.



Distributions of metric scores for selected chemical substance lists. For each list, the point shows the median scientific domain and information availability metrics. The whiskers span 90% of the distributions. Data here is taken from the lists across the TSCA Active Inventory. Uses data from the complete TSCA active inventory.

Proof-of-Concept Results

- The larger the value, the fewer the number of chemicals with that type of information
- Ecotoxicology, neurotoxicology BAF medium confidence have largest amount of missing data



Example: Compare Two Chemicals

CASRN	4435-53-4	71-43-2
Name	**3-Methoxybutyl acetate**	**Benzene**
Scientific Domain Metric	15.9	70.5
Information Availability Metric	60	93
IG flag human hazard (missing mammalian hazard data)	subchronic, chronic, developmental	developmental, reproductive
IG flag ecological hazard (missing eco hazard data)	acute plant, repeat dose invertebrate, repeat dose vertebrate	acute plant, acute invertebrate
Human hazard-to-exposure ratio metric	2.3	2.7
Ecological hazard metric	2.0	2.0
Carcinogenicity metric	0 (no data)	4
Genotoxicity metric	1	4
Susceptible population metric	2	4
Persistence bioaccumulation metric	1	2
Sensitization / irritation metric	1	3
HER repeat dose	13253000	11374
POD in vivo oral repeat dose	100 mg/kg-day	0.015 mg/kg-day
Human exposure (SEEM3)	0.0000075 mg/kg-day	0.0000013 mg/kg-day
Ecological min POD	0.71 mg/L	0.49 mg/L
Genotoxicity call	non-genotoxic	genotoxic
Carcinogenicity call		Group I: carcinogenic to humans
Skin sensitization metric		L
Eye irritation metric	L	H
Skin irritation metric	L	H
Volatile	No	Yes

Challenges

- Data sources are limited
  - Many chemicals do not have data in any source
  - Only public data was used, i.e., no CBI data
  - Largely only use data from other compilations, i.e., do not carry out targeted literature search and data extraction
- Manual data QA/QC is time and resource intensive for thousands of chemicals
  - CCTE is developing automated pipelines and web-based manual QC tools
- Apples and oranges tradeoffs
  - How to weigh relative concerns of hazard, exposure, physchem properties?
  - This is finally a policy decision

Summary

- The PICS approach was developed to better understand the landscape of publicly available information for large numbers of chemical substances
- It combines results from domain-specific workflows that reflect the overall degree of potential concern related to human health and the environment with the amount of relevant information
- It is intended to focus expert review on substances that may have a greater potential for selection as high- or low-priority candidates
- The proof-of-concept case study demonstrated that the PICS approach generally resulted in higher metrics for the high-priority candidates as compared to the low-priority candidates and identified areas for potential information gathering
- The method and software are flexible and can be customized for other prioritization applications

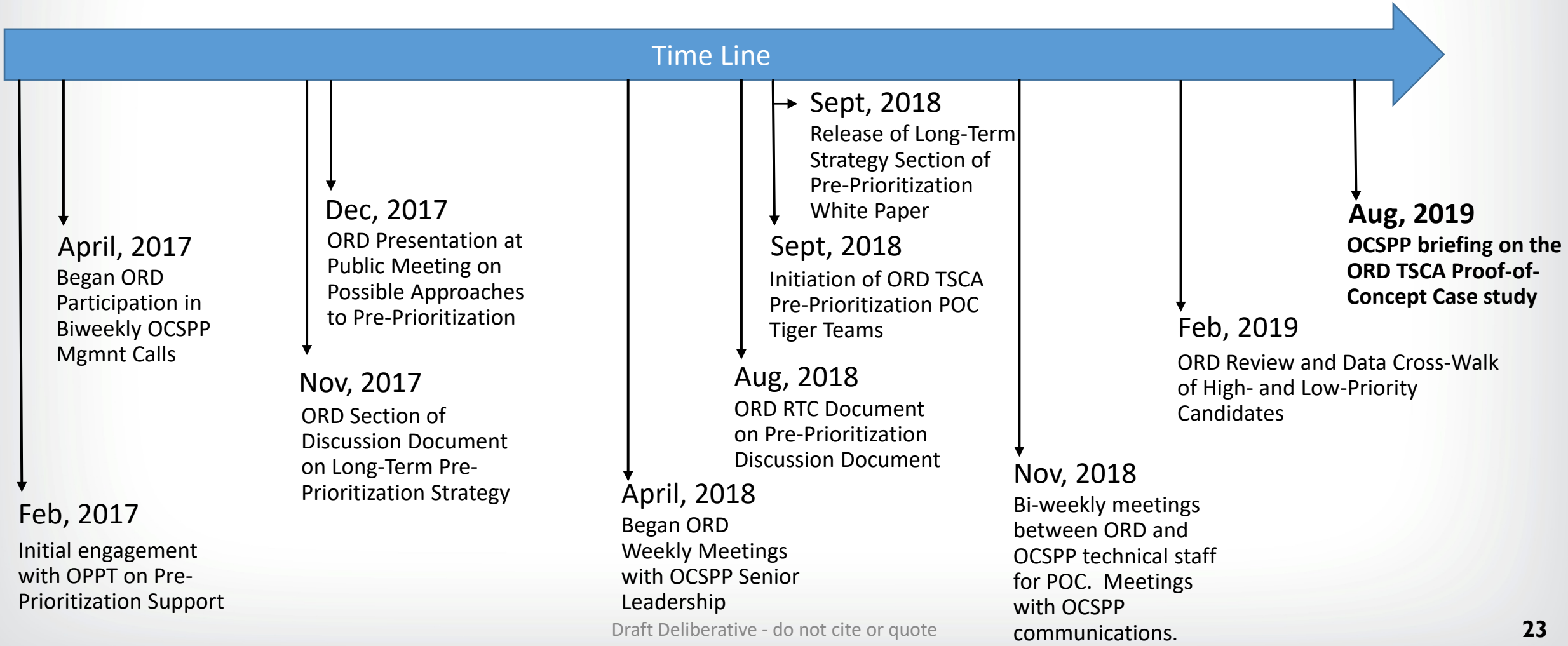
Data Curation and QC Tiger Team

- **General –** John Cowden (NCCT), Richard Judson (NCCT), Amar Singh (NCCT)
- **QC Data Integration and QA Automation Workgroup** - Richard Judson (NCCT), Jeremy Dunne (NCCT), Amar Singh (NCCT), Chris Grulke (NCCT)
- **Human Health Hazard/Risk Assessment Workgroup** - Johanna Congleton (NCEA), Urmila Kodavanti (NHEERL), Chris Lau (NHEERL), Mary Gilbert (NHEERL), Yu-Sheng Lin (NCEA), Dan Vallero (NHEERL), Kelly Garcia (NCEA), Carolyn Gigot (NCEA), Andrew Greenhalgh (NCEA), Allison Eames (NERL)
- **Ecological Toxicity Data Workgroup** - Dale Hoff (NHEERL), Colleen Elonen (NHEERL), Leslie Hughes (NHEERL), Anita Pascocello (NHEERL)
- **Exposure Data Workgroup** - Katherine Phillips (NERL), Janet Burke (NERL), Abhishek Komandur (NERL), Ashley Jackson (NERL), Lauren Koval (NERL)
- **Genotoxicity Data Workgroup** - David DeMarini (NHEERL), Maureen Gwinn (NCCT), Catherine Gibbons (NCEA), Sarah Warren (NHEERL), Jeff Dean (NCEA), Anita Simha (NCCT), Nagu Keshava (NCEA)
- **Chemistry Data Workgroup** - Kent Thomas (NHEERL), Michael Gonzalez (NRMRL), Doug Young (NRMRL), Chris Grulke (NCCT)

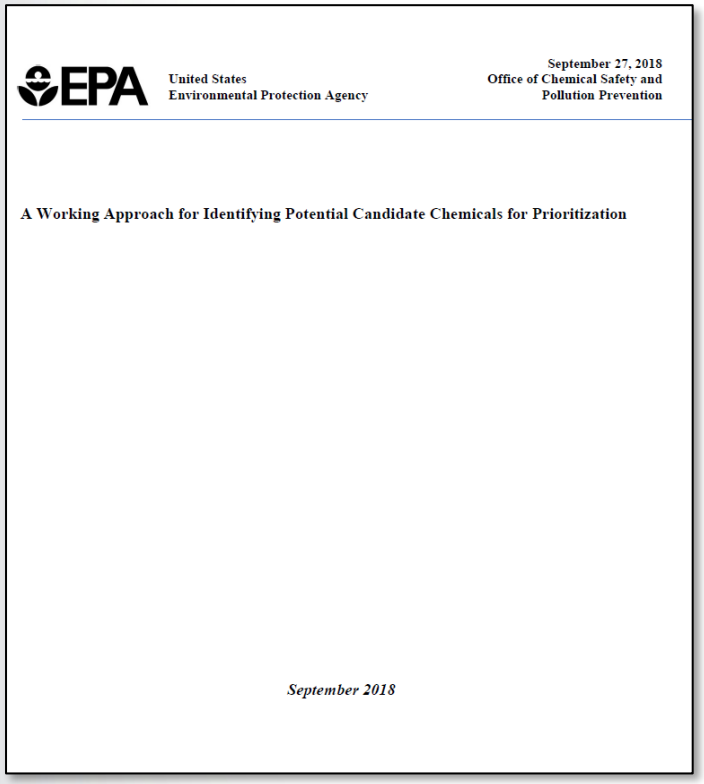
Proof-of-Concept Tiger Team

- **General** - Maureen Gwinn (NCCT), Richard Judson (NCCT), Amar Singh (NCCT)
- **Information availability** - Tony Williams (NCCT), Jeremy Dunne (NCCT), Jason Lambert (NCCT)
- **Human Hazard-to-Exposure Ratio** - Katie Paul-Friedman (NCCT), John Wambaugh (NCCT), Elaina Kenyon (NHEERL), Kristin Isaacs (NERL), Jason Lambert (NCCT)
- **Susceptible Population Exposure** - Kathie Dionisio (NERL), Kristin Isaacs (NERL), John Wambaugh (NCCT)
- **Carcinogenicity/Genotoxicity** - Grace Patlewicz (NCCT), David DeMarini (NHEERL), Catherine Gibbons (NCEA), Jeffry Dean (NCEA), Anita Simha (NCCT), Nagu Keshava (NCEA), Todd Martin (NRMRL), Sarah Warren (NHEERL)
- **Eco Hazard** - Dan Villeneuve (NHEERL), Carlie LaLone (NHEERL), Todd Martin (NRMRL)
- **Persistence/bioaccumulation** - John Nichols (NHEERL), Lawrence Burkhard (NHEERL), Eric Weber (NERL)
- **Skin sensitization/irritation and Eye irritation** - Todd Martin (NRMRL), Leora Vegosen (NRMRL)

History of Pre-Prioritization Support



Working Approach Document to Identify Candidate Chemicals for Prioritization


EPA United States Environmental Protection Agency

September 27, 2018
Office of Chemical Safety and Pollution Prevention

A Working Approach for Identifying Potential Candidate Chemicals for Prioritization

September 2018

- **Near-Term Strategy**
  - High-priority candidates selected from TSCA workplan based on priorities, quality and quantity of information, and workload
  - Low-priority candidates selected from EPA SCIL, ChAMP, and OECD SIDS based on quality and quantity of information for hazard and exposure for each condition of use
- **Long-Term Strategy**
  - Bin chemicals based on a combination of risk-related scoring and information availability
  - *Committed to subsequent release of proof-of-concept with a small number of substances that provides operational details on the data integration, scoring, and identification of information gaps*

Suggested Next Steps

- Obtain OCSPP feedback and comments on the proof-of-concept white paper
- Perform external peer-review
- Release proof-of-concept white paper for review and public comment

Data Extraction and Quality Control

- **Data extracted from "Type 1" data sources**
  - Type 1 data sources are publicly available and readily searchable, enabling data extraction in structured form
  - Consistent with approach outlined in the Near-term Strategy
- **Quality control (QC) was performed on the data for the chemical substance subset in order to:**
  - Estimate the accuracy of the data used in this case study
  - Inform the development of formal quality assurance (QA) plan
  - Obtain information on the scope and resources needed to perform QC for the entire active TSCA inventory or for other sets of chemicals

QC Rates and Time Requirements

- **Source Traceability and Error Rates:**
  - Transcription error rates were typically <1%.
  - Lack of primary and secondary sources was ~6%.
  - Lack of primary source was higher (5 – 60%)
- **Time Investment:**
  - QC review time ranged between 1 – 10 min/data point
  - For human health data, there are >2,200,000 data points for TSCA actives requiring ~100 person years to review
  - For eco data, there are >2,700,000 data points for TSCA actives requiring 25 person years
  - Development of data QC tool is expected to decrease these time frames