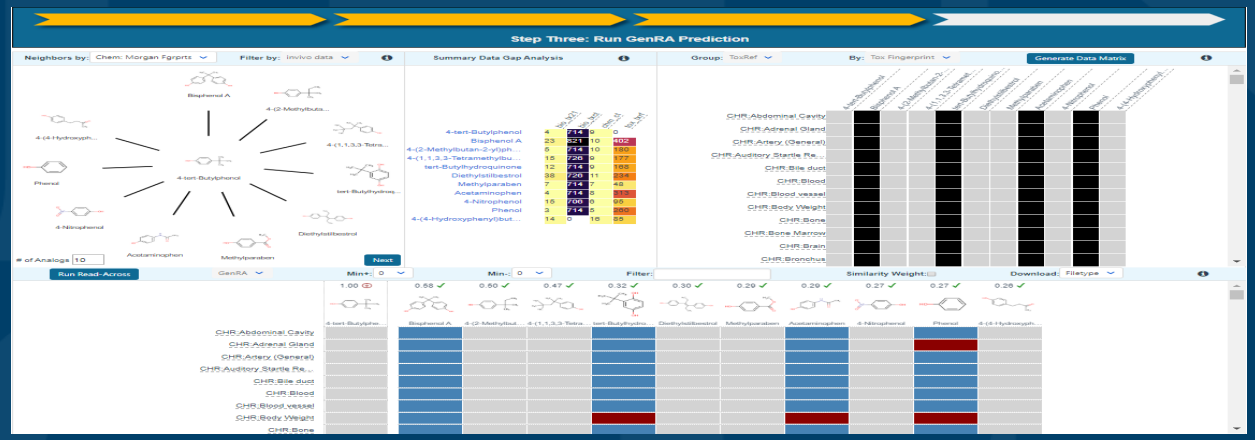
Navigating towards data-driven read-across approaches: Generalised Read Across (GenRA), a workflow module within the EPA CompTox Chemicals Dashboard



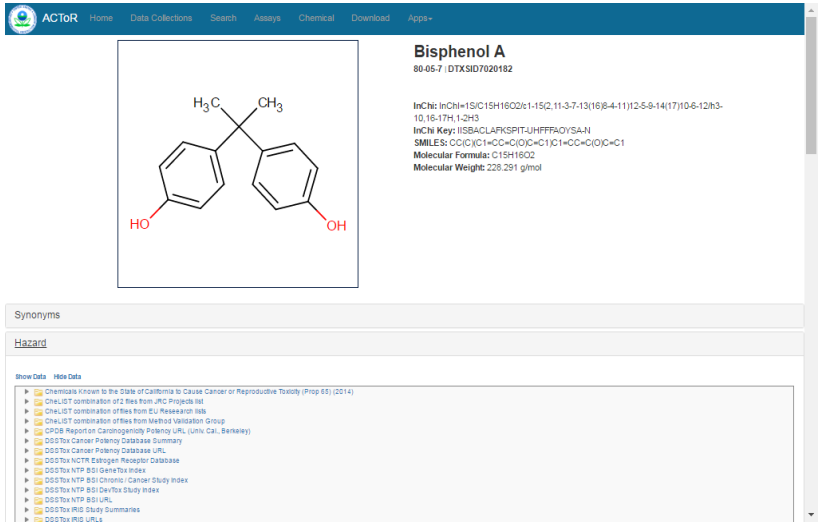
Grace Patlewicz
National Center for Computational Toxicology (NCCT), US EPA

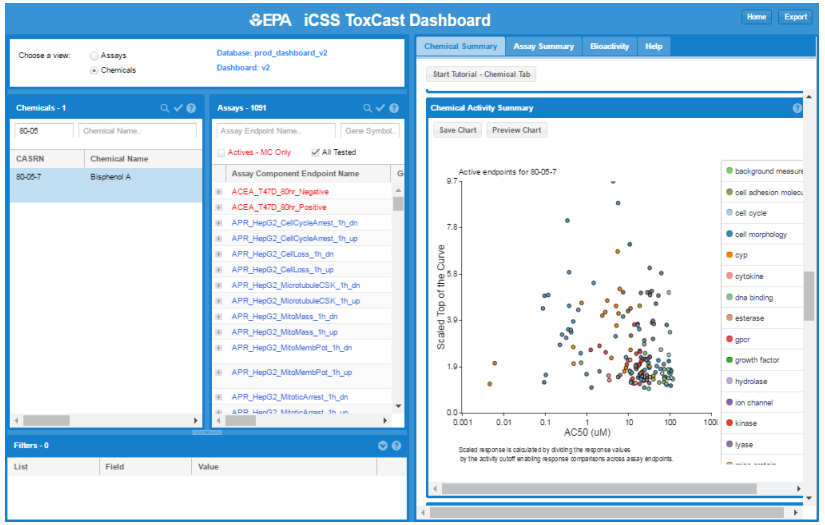
The views expressed in this presentation are those of the author and do not necessarily reflect the views or policies of the U.S. EPA

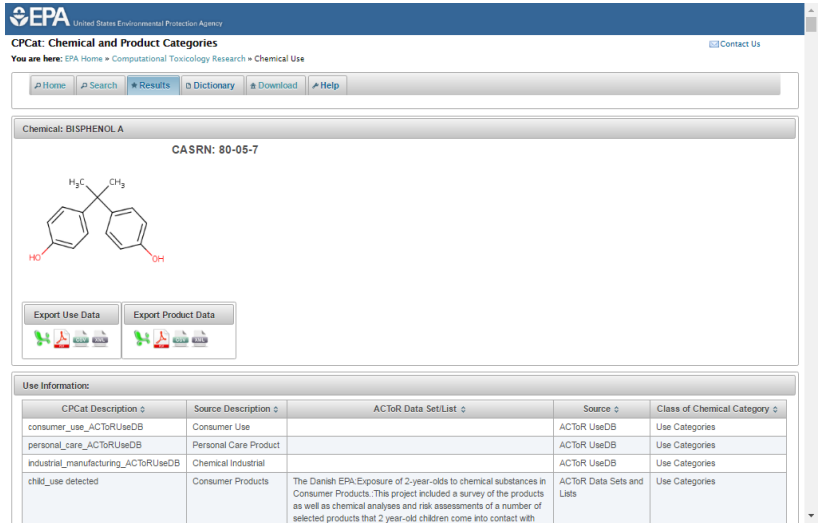
Outline

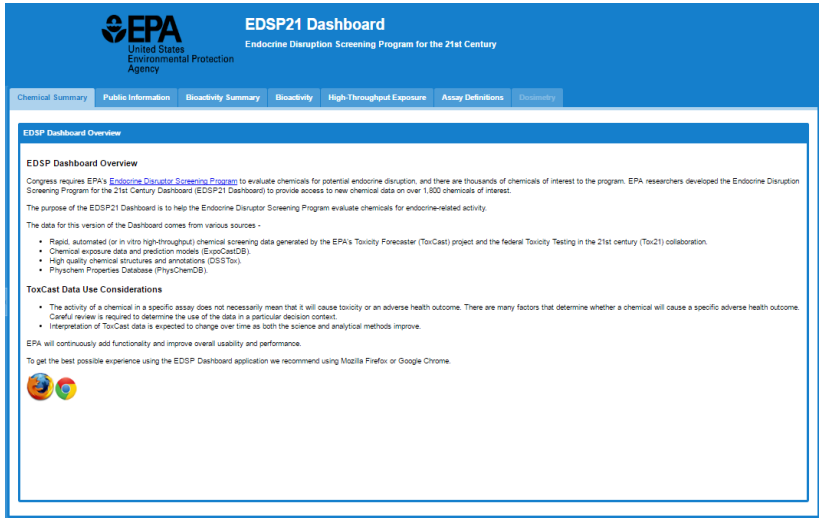
- Introduction to the EPA CompTox Chemicals Dashboard
- Read-across – existing resources
- Generalised Read-across (GenRA) approach
- GenRA implementation
- Practical application
- Ongoing research to enhance GenRA
- Acknowledgements

Earlier Dashboard Applications: Single architecture in development

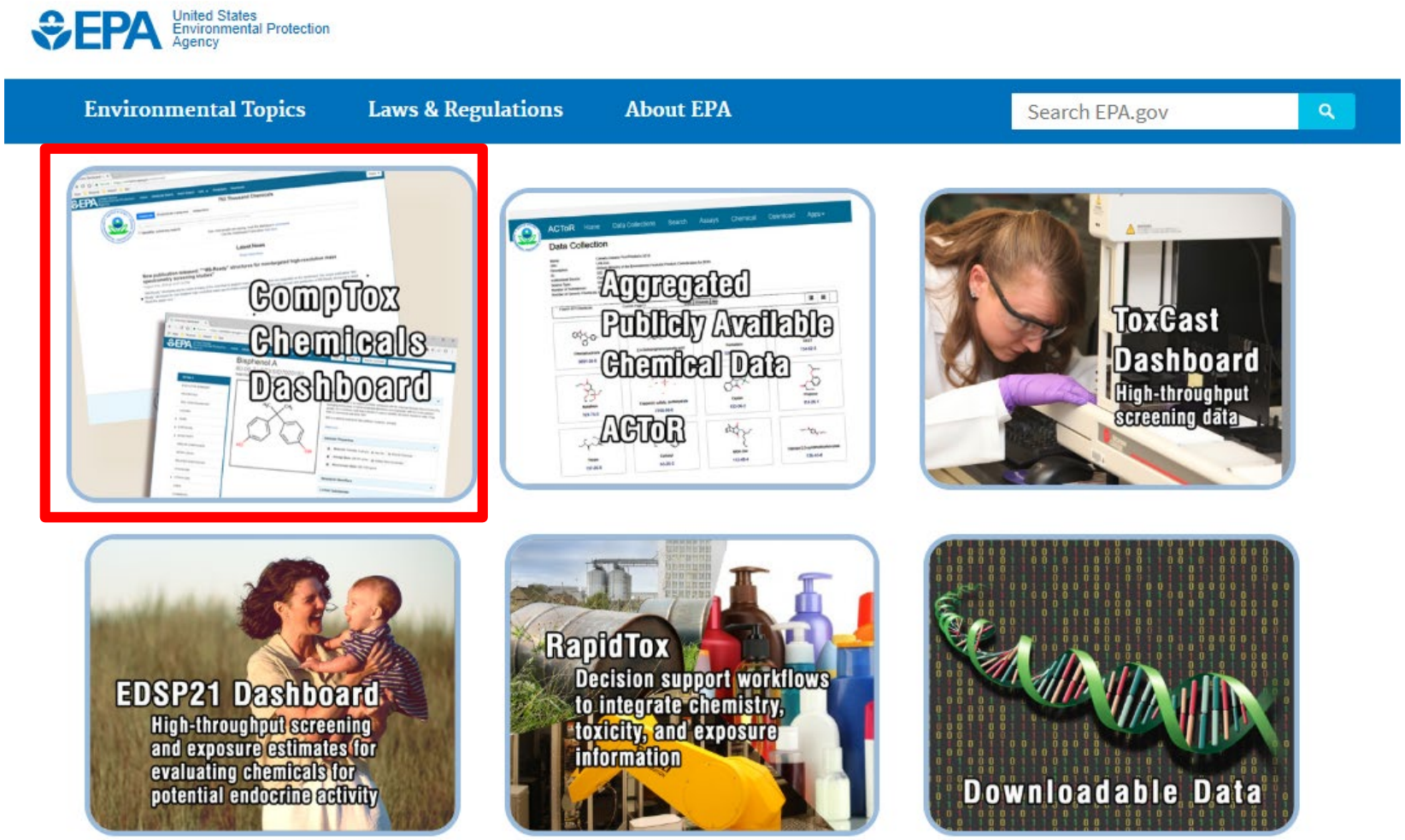








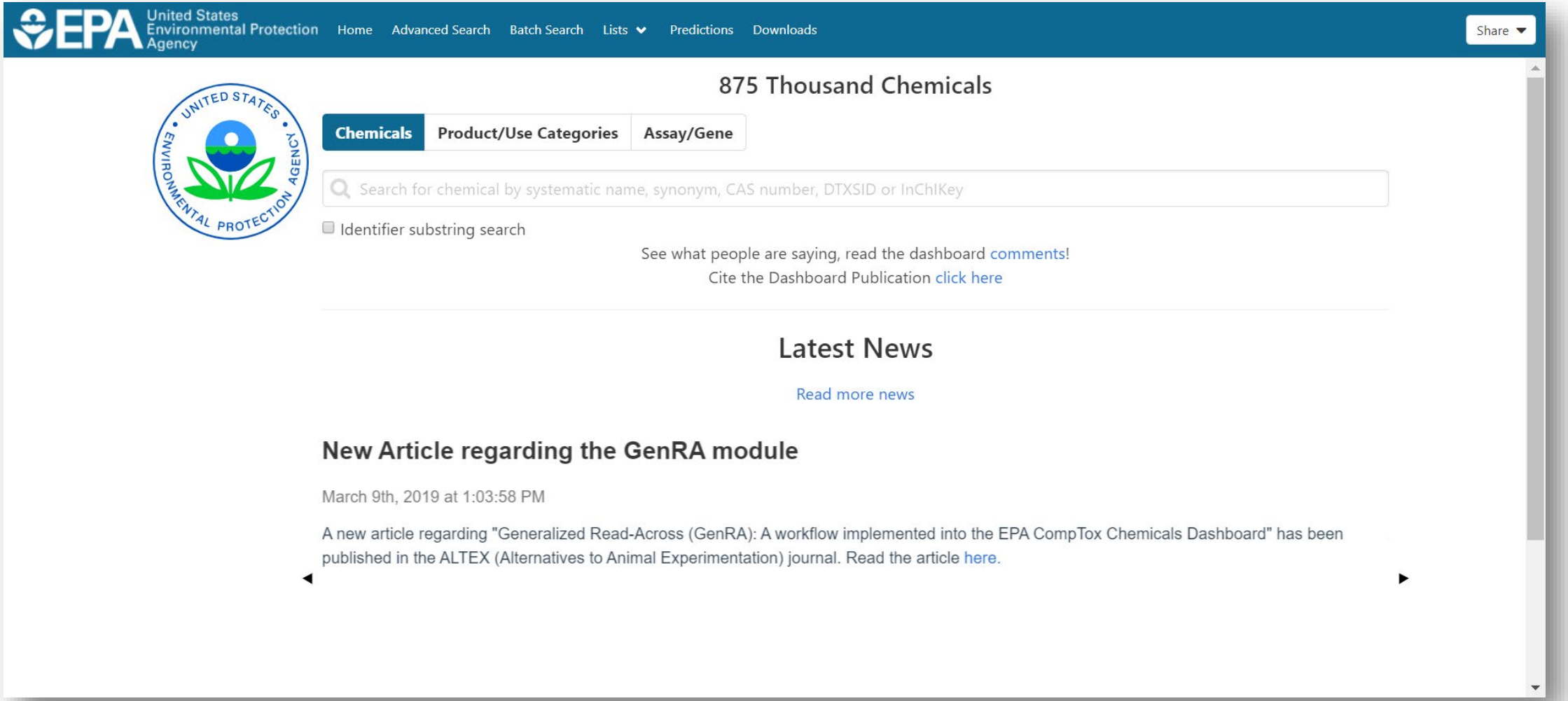
The CompTox Portal
https://comptox.epa.gov/



EPA CompTox Chemicals Dashboard

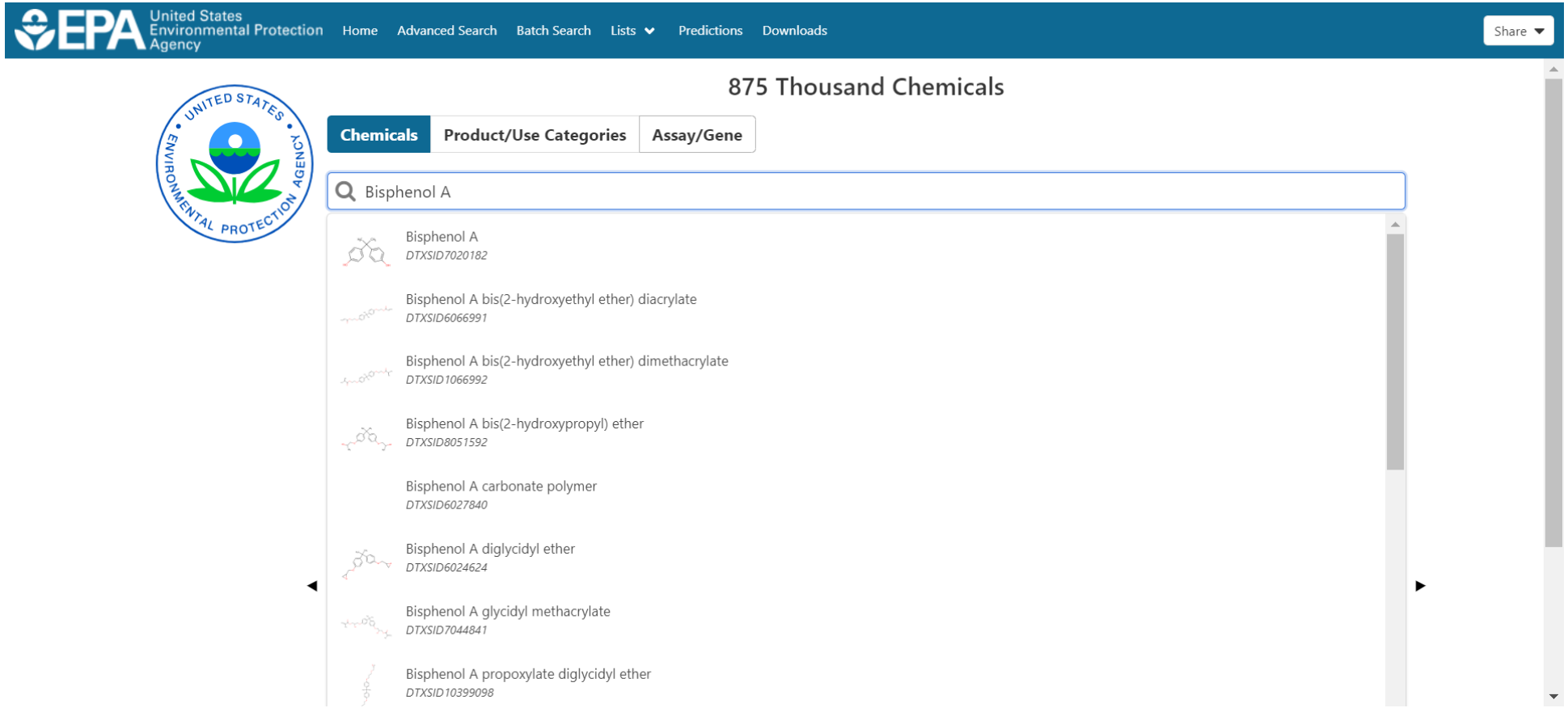
- A publicly accessible website delivering access:
  - ~875,000 chemicals with related property data
  - Experimental and predicted physicochemical property data
  - Integration to "biological assay data" for 1000s of chemicals
  - Information regarding consumer products containing chemicals
  - Links to other agency websites and public data resources
  - "Literature" searches for chemicals using public resources
  - "Batch searching" for thousands of chemicals
  - DOWNLOADABLE Open Data for reuse and repurposing

CompTox Chemicals Dashboard: Landing Page

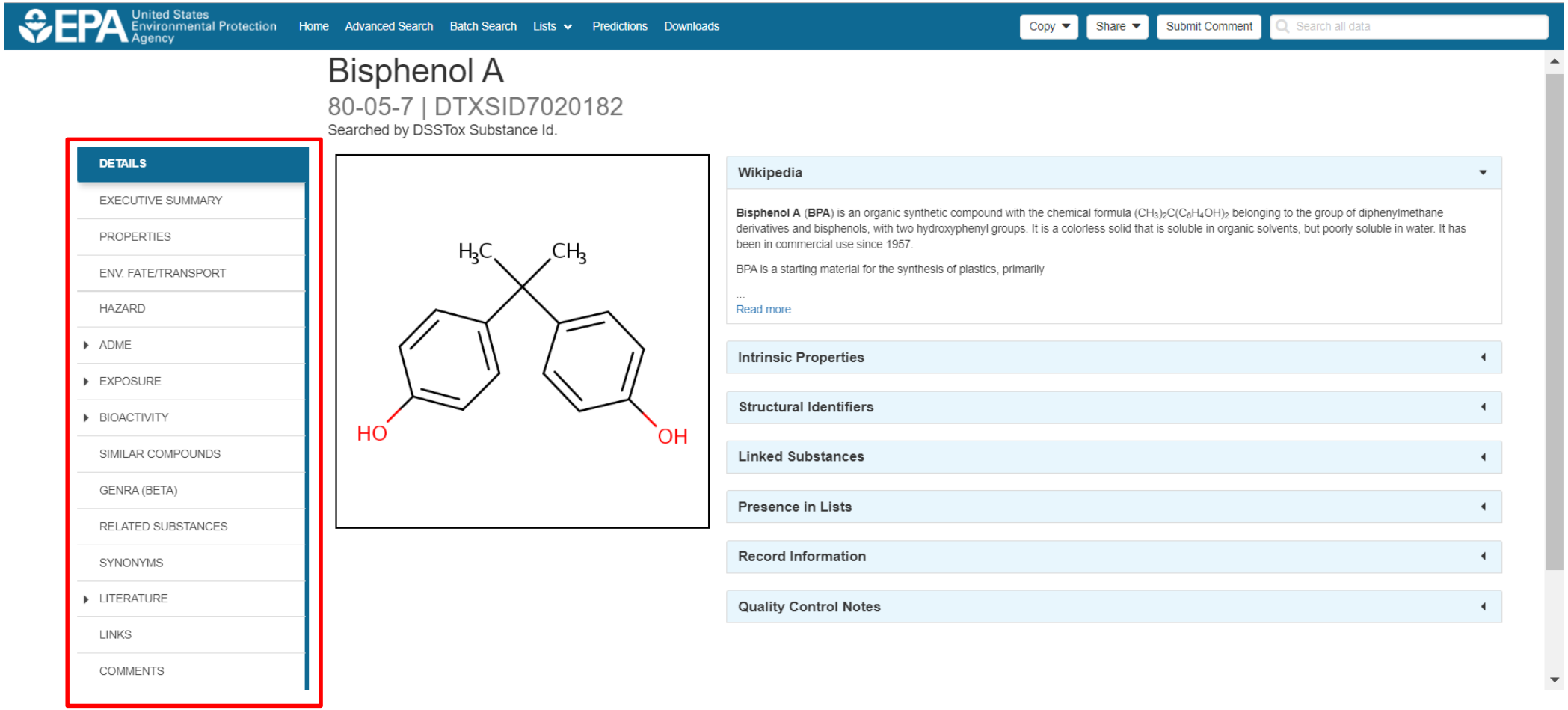


CompTox Chemicals Dashboard: Landing Page

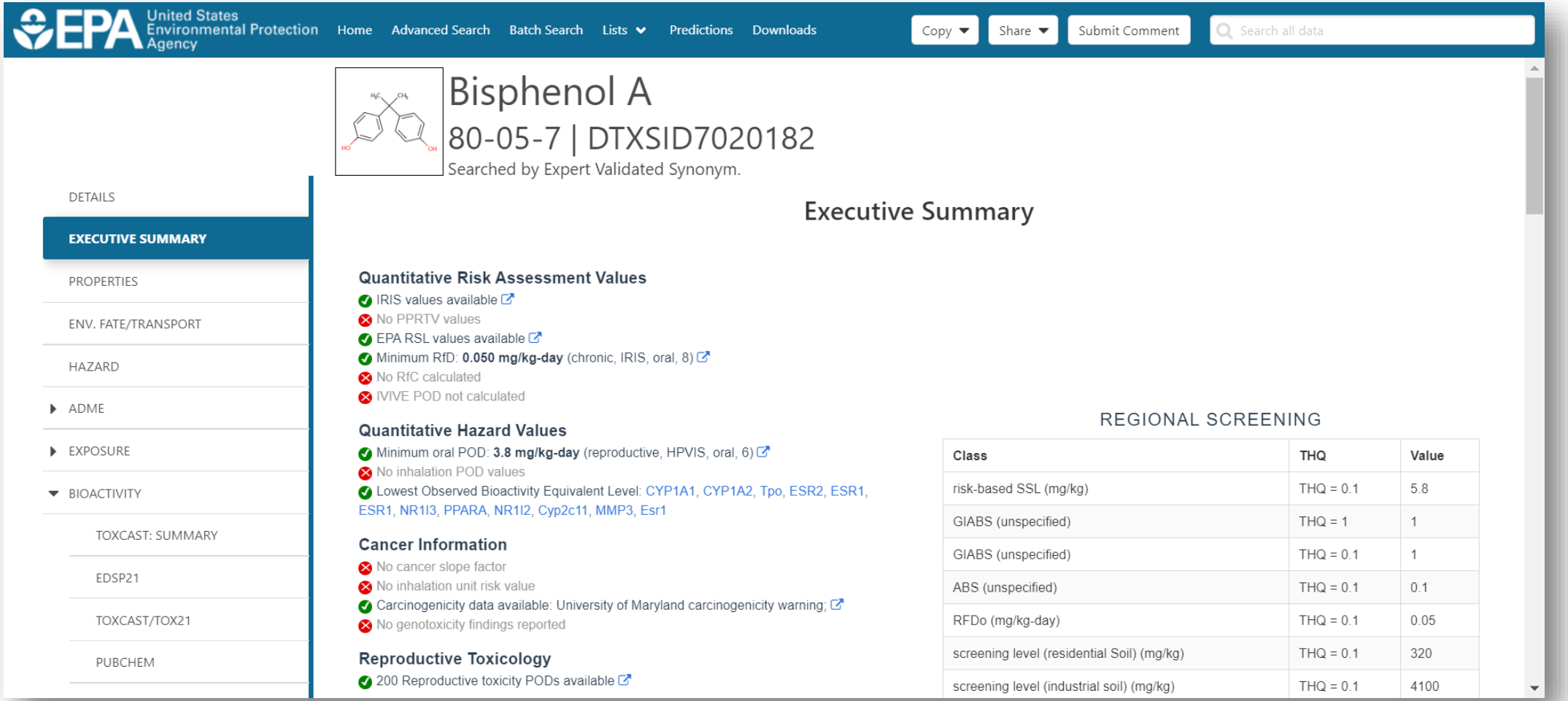
- Different entry points depending on domain of interest



CompTox Chemicals Dashboard: Landing Page for a specific chemical



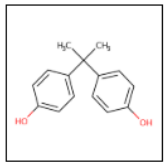
CompTox Chemicals Dashboard: Executive Summary



Generalised Read-Across (GenRA) as a workflow module

EPA United States Environmental Protection Agency

Home | Advanced Search | Batch Search | Lists | Predictions | Downloads | Copy | Share | Submit Comment | Search all data

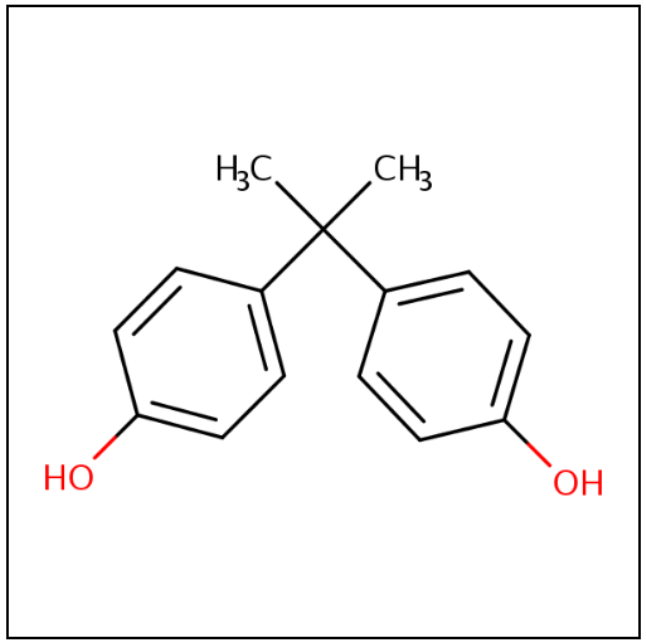


Bisphenol A

80-05-7 | DTXSID7020182

Searched by DSSTox Substance Id.

- DETAILS
- EXECUTIVE SUMMARY
- PROPERTIES
- ENV. FATE/TRANSPORT
- HAZARD
- ADME
- EXPOSURE
- BIOACTIVITY
- SIMILAR COMPOUNDS
- GENRA (BETA)
- RELATED SUBSTANCES
- SYNONYMS
- LITERATURE



Wikipedia

Bisphenol A (**BPA**) is an organic synthetic compound with the chemical formula $(CH_3)_2C(C_6H_4OH)_2$ belonging to the group of diphenylmethane derivatives and bisphenols, with two hydroxyphenyl groups. It is a colorless solid that is soluble in organic solvents, but poorly soluble in water (0.344 wt % at 83 °C).
BPA is a starting material for the synthesis of plastics, primarily certain polycarbonates
...
Read more

Intrinsic Properties

Molecular Formula: $C_{15}H_{16}O_2$ | Mol File | Find All Chemicals

Average Mass: 228.291 g/mol | Isotope Mass Distribution

Monoisotopic Mass: 228.11503 g/mol

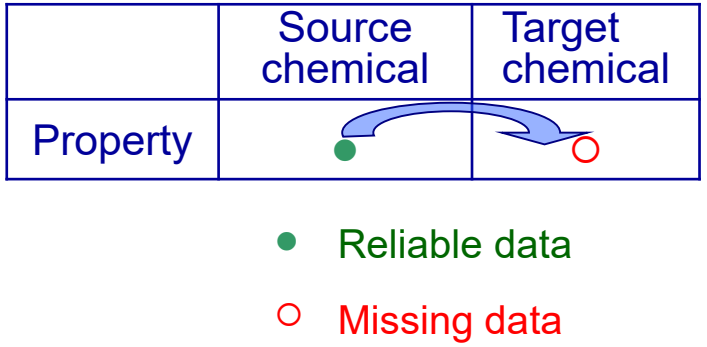
Structural Identifiers

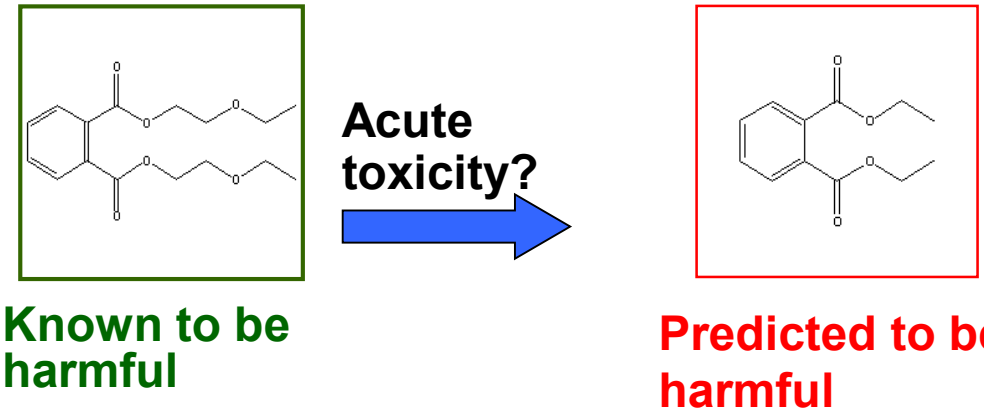
Linked Substances

Presence in Lists

Definitions: Read-across

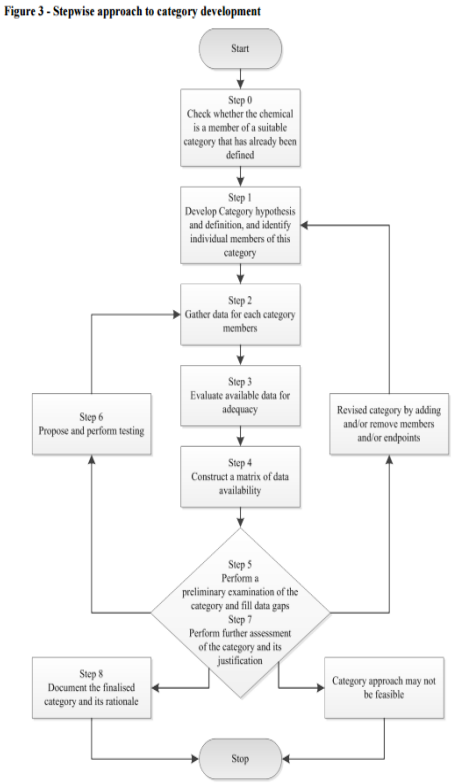
- Read-across describes the method of filling a data gap whereby a chemical with existing data values is used to make a prediction for a 'similar' chemical.
- A target chemical is a chemical which has a data gap that needs to be filled i.e. the subject of the read-across.
- A source analogue is a chemical that has been identified as an appropriate chemical for use in a read-across based on similarity to the target chemical and existence of relevant data.

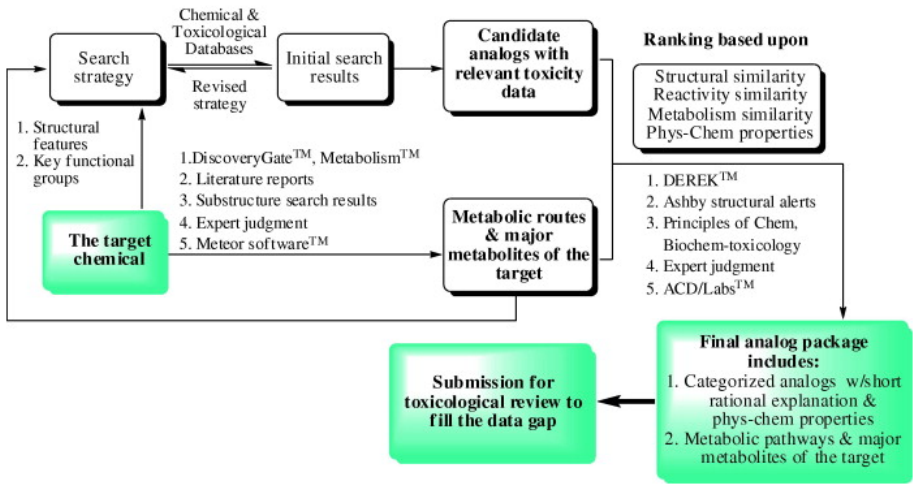




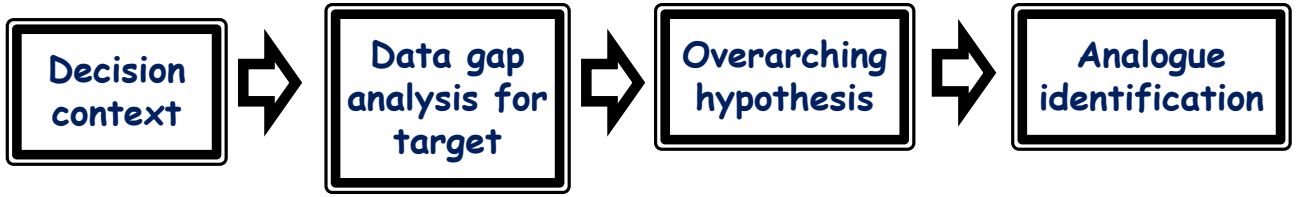
Frameworks for developing category/analogue approaches

OECD (2014)





Wu et al, 2010



Patlewicz et al, 2015

Frameworks for the assessment of read-across

READ ACROSS UNCERTAINTY EVALUATION QUESTIONNAIRE FOR:
Target chemical (SOI) = *(list CAS#)*

INSTRUCTIONS
Complete the Questionnaire. Answer the questions for each endpoint where SAR was conducted, and follow instructions listed in each section below. (In general, NO responses indicate potential areas of uncertainty in the proposed read across.)

Questions	Responses by Endpoint	
	Repeat Dose Toxicity	Reproductive Toxicity
Section I. Chemical similarity between source (analogue) and target (SOI)		
1. For each endpoint, list the CAS#s of the source (analogues) contributing the critical study for the read across fo		
	CAS# ___	
2. What is the 'suitability rating' of the analogue?		
	____ Suit ____ Suit (skip to se ____ Suit (continue of the inte	
3. Are any differences in functional groups and associ be more reactive than the target)?		
	____ YES ____ NO ____ UNKNOWN ____ No Differences	____ NO ____ UNKNOWN ____ No Differences
	NOTES, if any:	NOTES, if any:

Suitability of Analogs contributing data

Are all features of SOI covered or differences in conservative direction

Blackburn and Stuard (2014)

Table 2
Scientific confidence considerations in Read-across evaluation.

Data issues	Similarity rationale
Analogue/category approach	Similarity rationale/hypothesis that underpins the analogue/category approach – Metabolic transformation – Structural similarity
Completeness of data matrix – No of data gaps e.g. source analogue(s) have many data points to address, target substance has a handful of data gaps.	Analogue validity – Analogue similarity with respect to general and endpoint specific considerations – Rationalization of why structural differences do not impact the toxicity
Quality of data for source analogues – e.g. Klimisch scores of 1 or 2	Concordance of effects and potency (if relevant) per endpoint • Presence or absence of adverse effects • Type of read-across (Qualitative, Quantitative, Trend Analysis)
	Concordance of effects and potency (if relevant) across endpoints

Patlewicz et al (2015)

Read-across resources: Selected read-across tools

Computational Toxicology 3 (2017) 1–18

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com/locate/comtox

Navigating through the minefield of read-across tools: A review of in silico tools for grouping

Grace Patlewicz [a,*], George Helman [a,b], Prachi Pradeep [a,b], Imran Shah [a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA
[b] Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA

ARTICLE INFO

Article history:
Received 29 March 2017
Received in revised form 22 May 2017
Accepted 25 May 2017
Available online 29 May 2017

Keywords:
Category approach
Analogue approach
Data gap filling
Read-across
(Q)SAR
Trend analysis
Nearest neighb

ABSTRACT

Read-across is a popular data gap filling technique used within analogue and category approaches for regulatory purposes. In recent years there have been many efforts focused on the challenges involved in read-across development, its scientific justification and documentation. Tools have also been developed to facilitate read-across development and application. Here, we describe a number of publicly available read-across tools in the context of the category/analogue workflow and review their respective capabilities, strengths and weaknesses. No single tool addresses all aspects of the workflow. We highlight how the different tools complement each other and some of the opportunities for their further development to address the continued evolution of read-across.

Published by Elsevier B.V.

Read-across resources: Selected read-across frameworks

Contents lists available at ScienceDirect

Computational Toxicology

journal homepage: www.elsevier.com

Journal Cover Image

Navigating through the minefield of read-across frameworks: A commentary perspective

Grace Patlewicz[a, *], Mark T.D. Cronin[b], George Helman[a, c], Jason C. Lambert[d], Lucina E. Lizarraga[d], Imran Shah[a]

[a] National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency (US EPA), 109 TW Alexander Dr, Research Triangle Park (RTP), NC 27711, USA
[b] School of Pharmacy and Biomolecular Sciences, Liverpool John Moores University, Byrom Street, Liverpool L3 3AF, UK
[c] Oak Ridge Institute for Science and Education (ORISE), 1299 Bethel Valley Road, Oak Ridge, TN 37830, USA
[d] National Center for Evaluation Assessment (NCEA), US Environmental Protection Agency (US EPA), 26 West Martin Luther King Dr, Cincinnati, OH 45268, USA

PROOF

Read-across resources:
Selected read-across examples/decision contexts

Regulatory Toxicology and Pharmacology 106 (2019) 197–209

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Exploring current read-across applications and needs among selected U.S. Federal Agencies

Grace Patlewicz[a,*], Lucina E. Lizarraga[b], Diego Rua[c], David G. Allen[d], Amber B. Daniel[d], Suzanne C. Fitzpatrick[e], Natàlia Garcia-Reyero[f], John Gordon[g], Pertti Hakkinen[h], Angela S. Howard[d], Agnes Karmaus[d], Joanna Matheson[g], Moiz Mumtaz[i], Andrea-Nicole Richarz[j], Patricia Ruiz[i], Louis Scarano[k], Takashi Yamada[l], Nicole Kleinstreuer[m]

[a] National Center for Computational Toxicology, U.S. Environmental Protection Agency, 109 TW Alexander Dr, Research Triangle Park, NC, 27709, USA
[b] National Center for Environmental Assessment, U.S. Environmental Protection Agency, 26 West Martin Luther King Drive, Cincinnati, OH, 45268, USA
[c] Center for Devices and Radiological Health, U.S. Food and Drug Administration, 10903 New Hampshire Avenue, Silver Spring, MD, 20993, USA
[d] ILS, P.O. Box 13501, Research Triangle Park, NC, 27709, USA
[e] Center for Food Safety and Applied Nutrition, U.S. Food and Drug Administration, 5100 Paint Branch Parkway, College Park, MD, 20740, USA
[f] Environmental Laboratory, U.S. Army Engineer Research and Developmental Center, 3909 Halls Ferry Rd., Vicksburg, MS, 39180, USA
[g] U.S. Consumer Product Safety Commission, 5 Research Place, Rockville, MD, 20850, USA
[h] National Library of Medicine, 6707 Democracy Blvd., Bethesda, MD, 20892, USA
[i] Agency for Toxic Substances and Disease Registry, 1600 Clifton Rd., Chamblee, GA, 30341, USA
[j] European Commission, Joint Research Centre (JRC), Ispra, Italy
[k] Office of Pollution Prevention and Toxics, U.S. Environmental Protection Agency, 1200 Pennsylvania Ave. NW, Washington, DC, 20460, USA
[l] Division of Risk Assessment, Biological Safety Research Center, National Institute of Health Sciences, 3-25-26, Tonomachi, Kawasaki-ku, Kawasaki, Kanagawa, 210-9501, Japan
[m] National Toxicology Program Interagency Center for the Evaluation of Alternative Toxicological Methods, National Institute of Environmental Health Sciences, P.O. Box 12233, Research Triangle Park, NC, 27709, USA

More than just a 'REACH' regulatory context
ICCVAM Read-Across Workgroup

Regulatory Toxicology and Pharmacology 106 (2019) 278–291

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

journal homepage: www.elsevier.com/locate/yrtph

Predicting estrogen receptor activation by a group of substituted phenols: An integrated approach to testing and assessment case study

Francina Webster*, Matthew Gagné, Grace Patlewicz, Prachi Pradeep, Nicholas Trefiak, Richard S. Judson, Tara S. Barton-Maclaren

OECD IATA Case study

ARTICLE INFO

Keywords:
New approach methodology (NAM)
Integrated approach for testing and assessment (IATA)
Read-across
4-Tert-butylphenol
2,4-Di-tert-butylphenol
Octabenzone
Endocrine disruption
Estrogen
Administered equivalent dose (AED)
Bioactivity exposure ratio (BER)

ABSTRACT

Traditional approaches for chemical risk assessment cannot keep pace with the number of substances requiring assessment. Thus, in a global effort to expedite and modernize chemical risk assessment, New Approach Methodologies (NAMs) are being explored and developed. Included in this effort is the OECD Integrated Approaches for Testing and Assessment (IATA) program, which provides a forum for OECD member countries to develop and present case studies illustrating the application of NAM in various risk assessment contexts. Here, we present an IATA case study for the prediction of estrogenic potential of three target phenols: 4-tert-butylphenol, 2,4-di-tert-butylphenol and octabenzone. Key features of this IATA include the use of two computational approaches for analogue selection for read-across, data collected from traditional and NAM sources, and a workflow to generate predictions regarding the targets' ability to bind the estrogen receptor (ER). Endocrine disruption can occur when a chemical substance mimics the activity of natural estrogen by binding to the ER and, if potency and exposure are sufficient, alters the function of the endocrine system to cause adverse effects. The data indicated that of the three target substances that were considered herein, 4-tert-butylphenol is a potential endocrine disruptor. Further, this IATA illustrates that the NAM approach explored is health protective when compared to *in vivo* endpoints traditionally used for human health risk assessment.

A harmonised hybrid read-across workflow

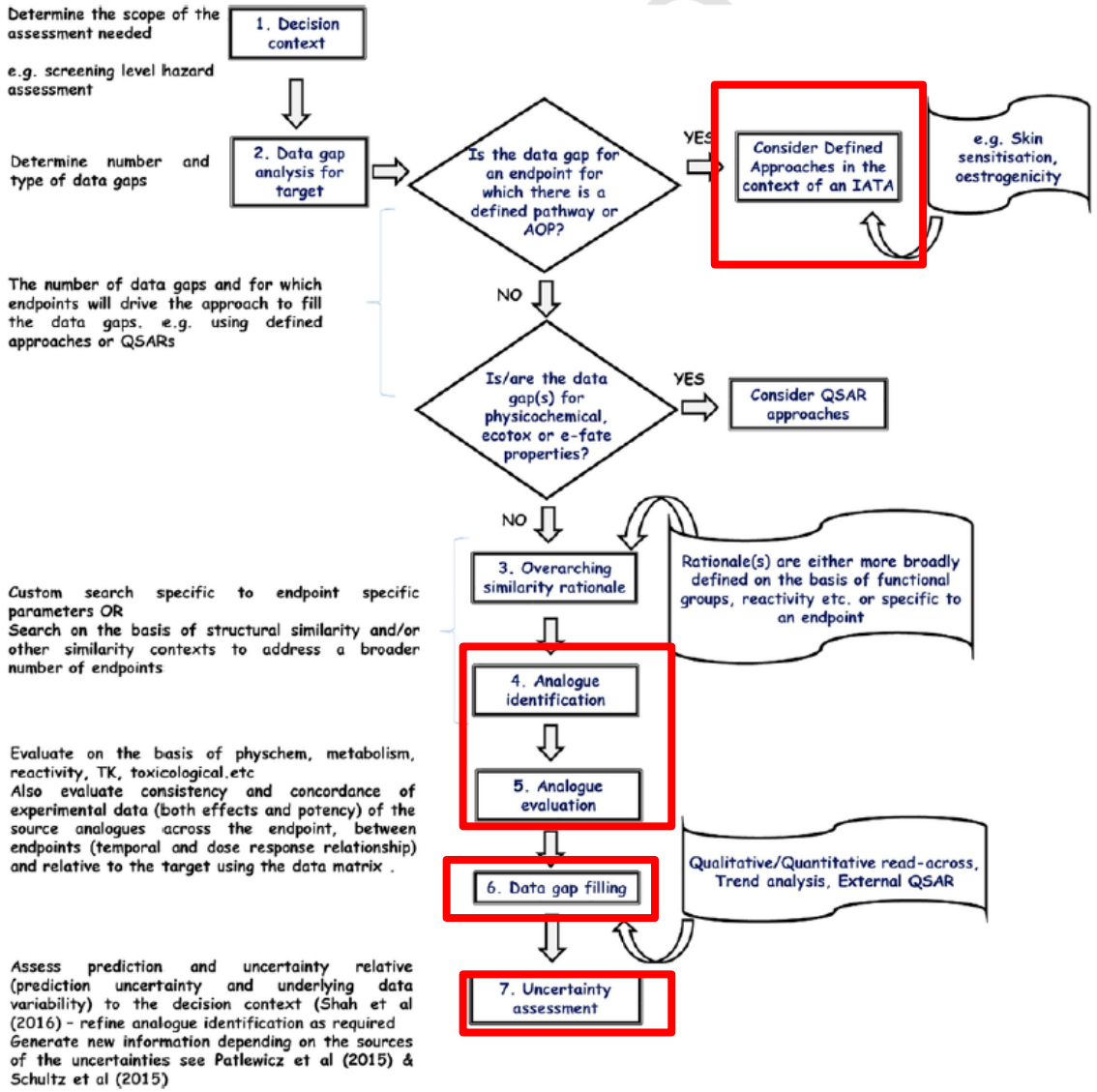


Fig. 9. A harmonised hybrid development and assessment framework.

- Where do other NAM fit?
- How should we transition to data-driven approaches? Limit subjectivity
- What about characterising the uncertainty of the predictions made?
- Generalisability/Scalability of read-across – coverage of read-across for inventories of chemicals?

Patlewicz et al., 2018

Selected read-across tools

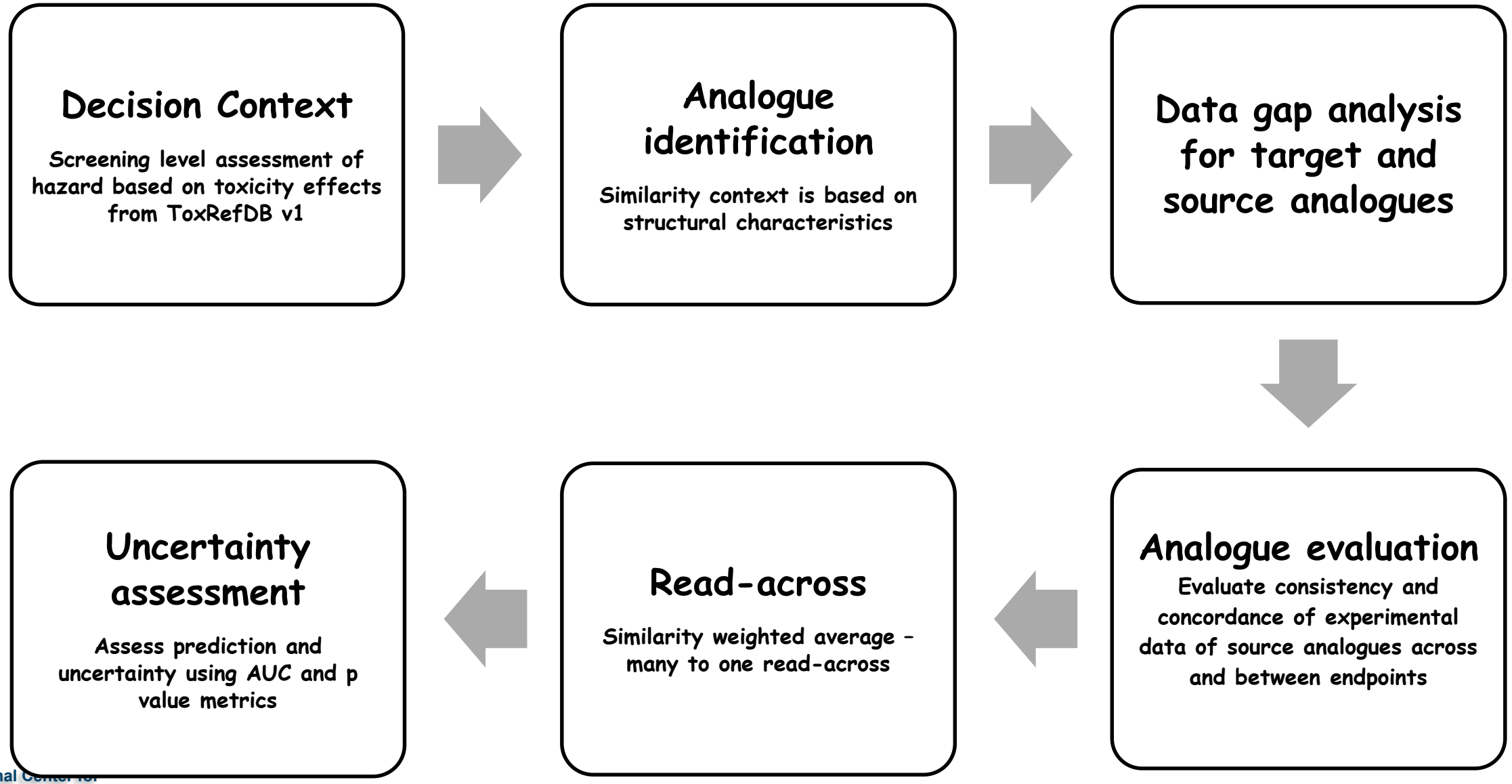
Tool	AIM	ToxMatch	AMBIT	OECD Toolbox	CBRA	ToxRead	GenRA
Analogue identification	X	X	X	X	X	X	X
Analogue Evaluation	NA	X	X by other tools available	X	X	X For Ames & BCF	NA
Data gap analysis	NA	X	X Data matrix can be exported	X Data matrix viewable	NA	NA	X Data matrix can be exported
Data gap filling	NA	X	User driven	X	X	X	X
Uncertainty assessment	NA	NA	NA	X	NA	NA	X
Availability	Free	Free	Free	Free	Free	Free	Free

GenRA

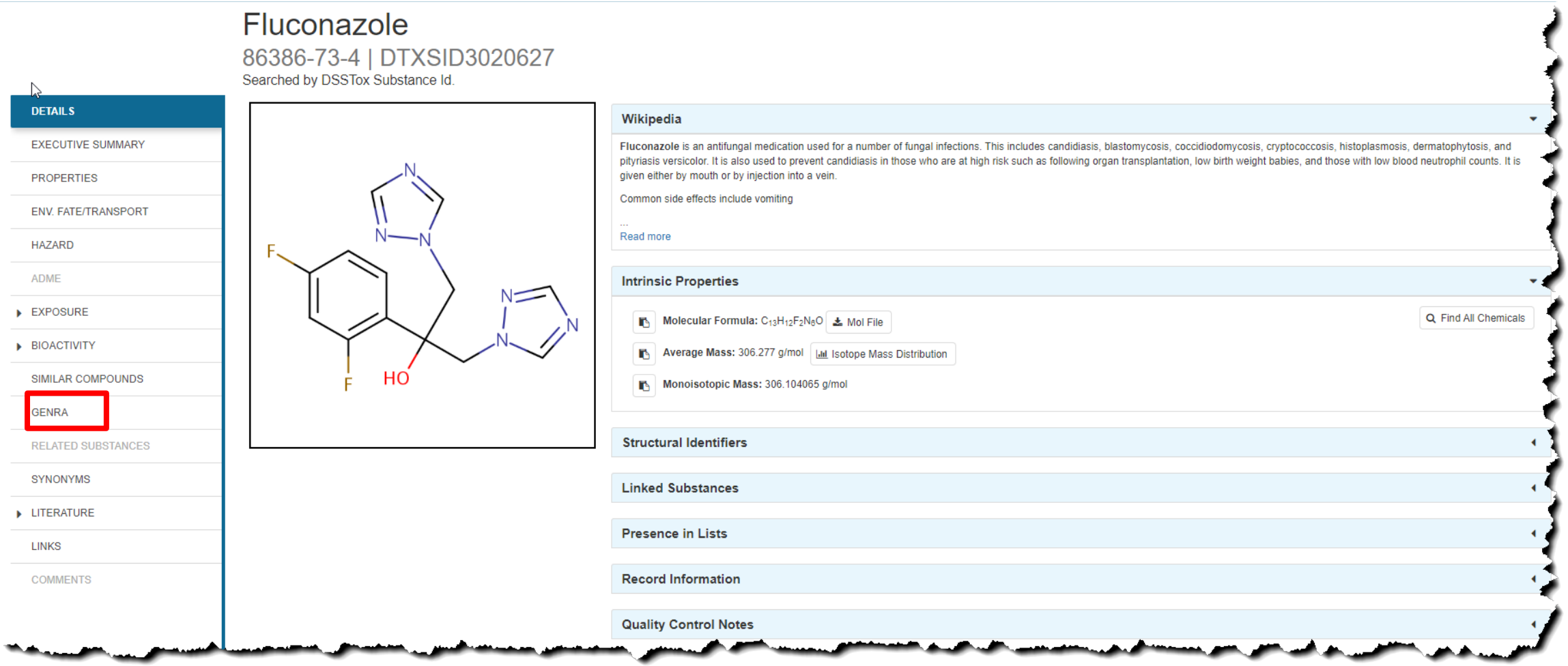
- GenRA (Generalised Read-Across)
- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors
- Systematically evaluates read-across performance and uncertainty using available data

Jaccard similarity:

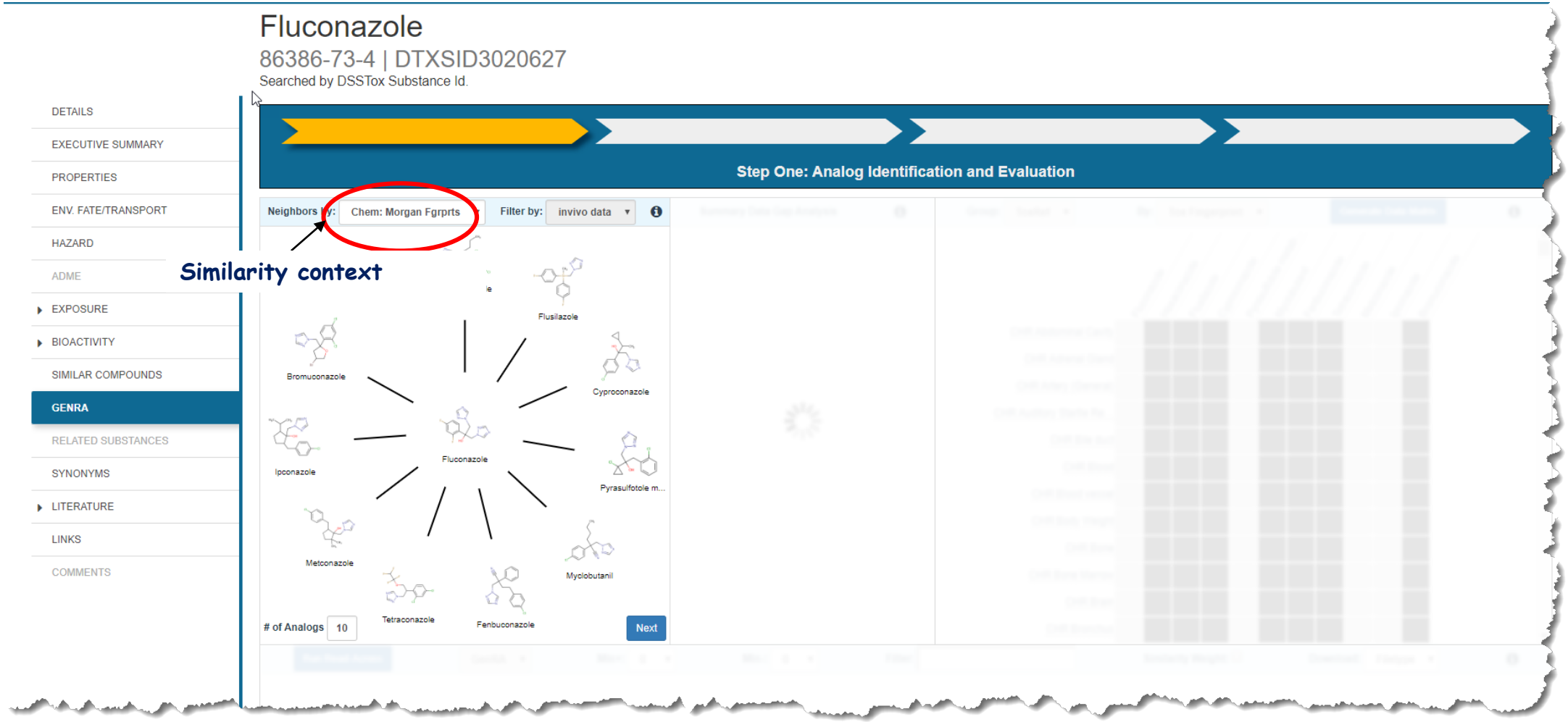
Read-across workflow in GenRA



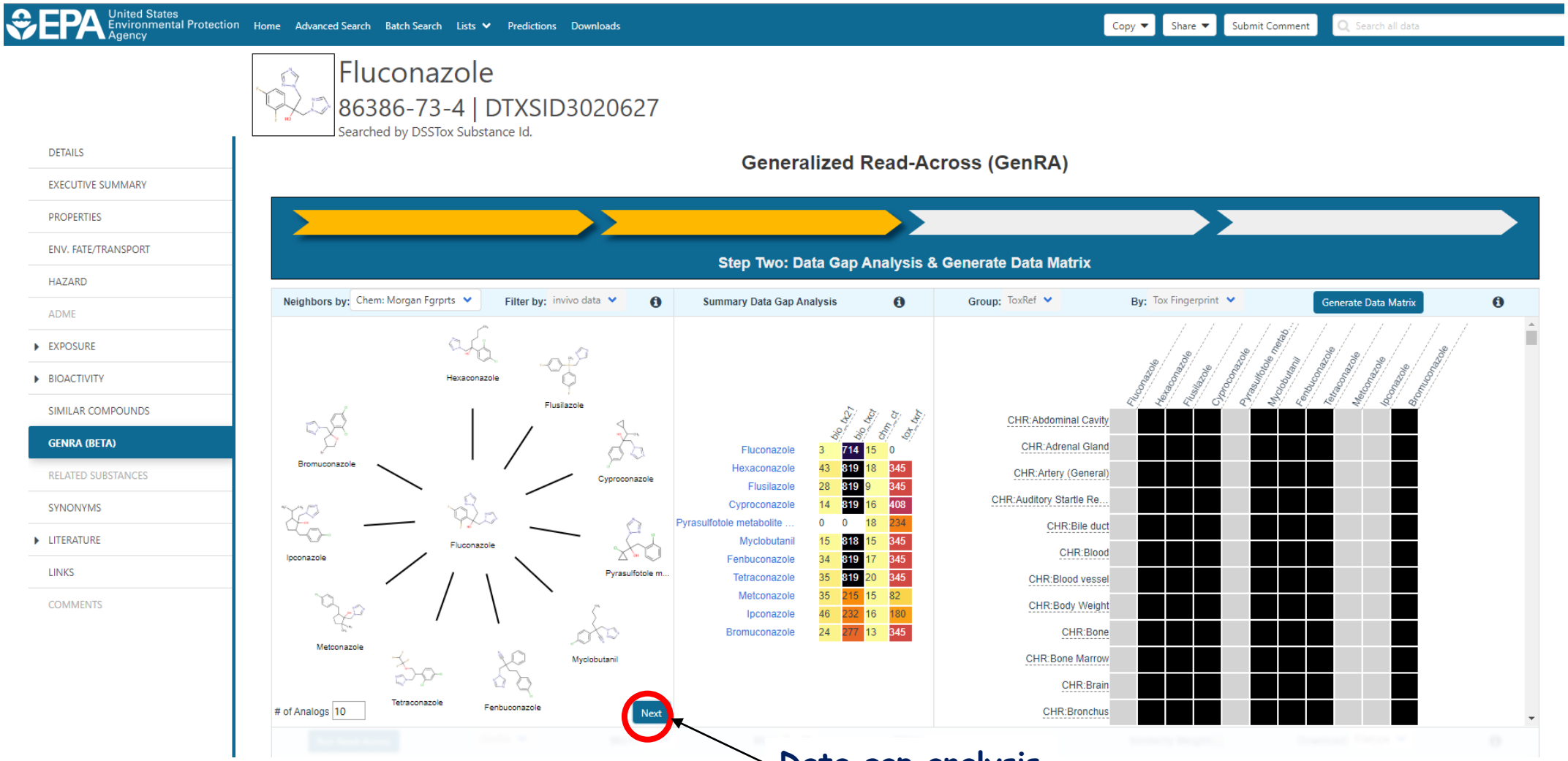
GenRA tool in practice



GenRA tool in practice

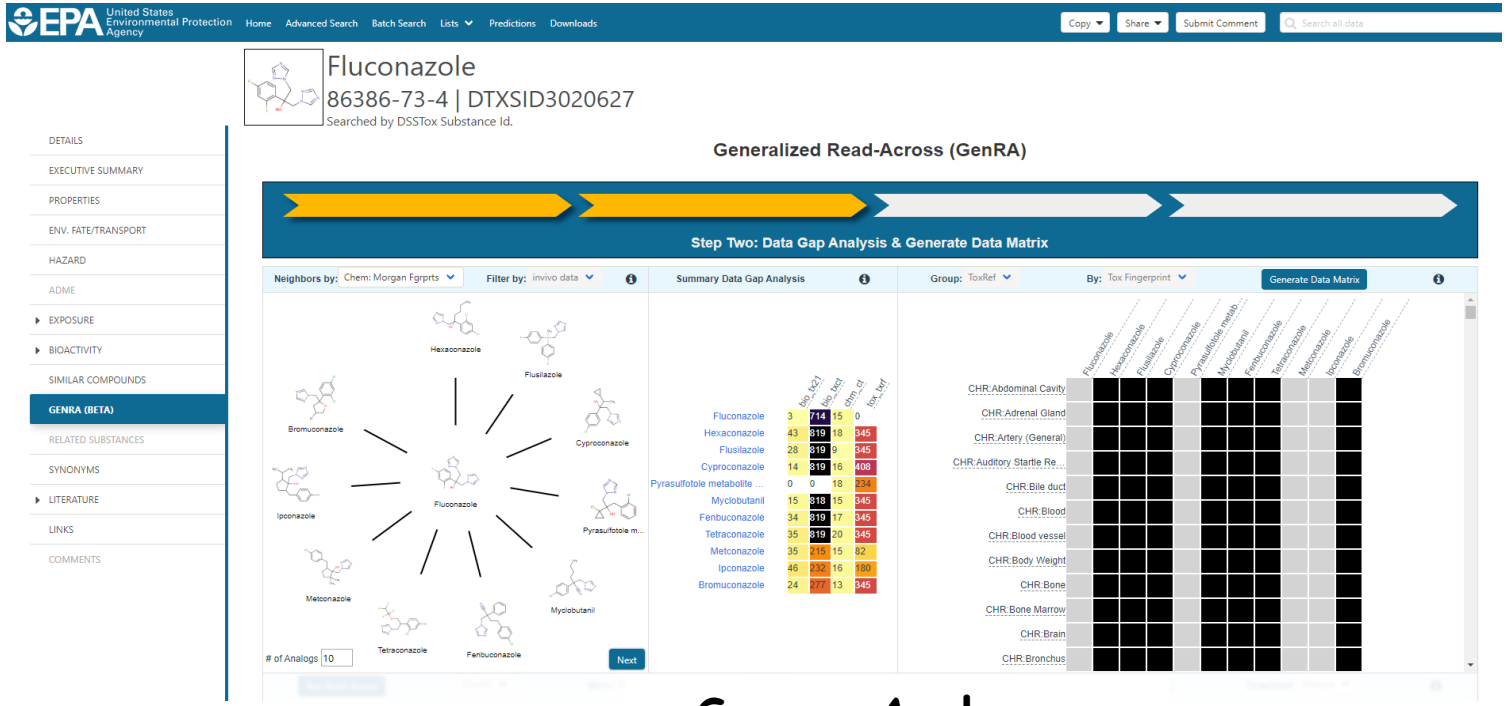


GenRA tool in practice

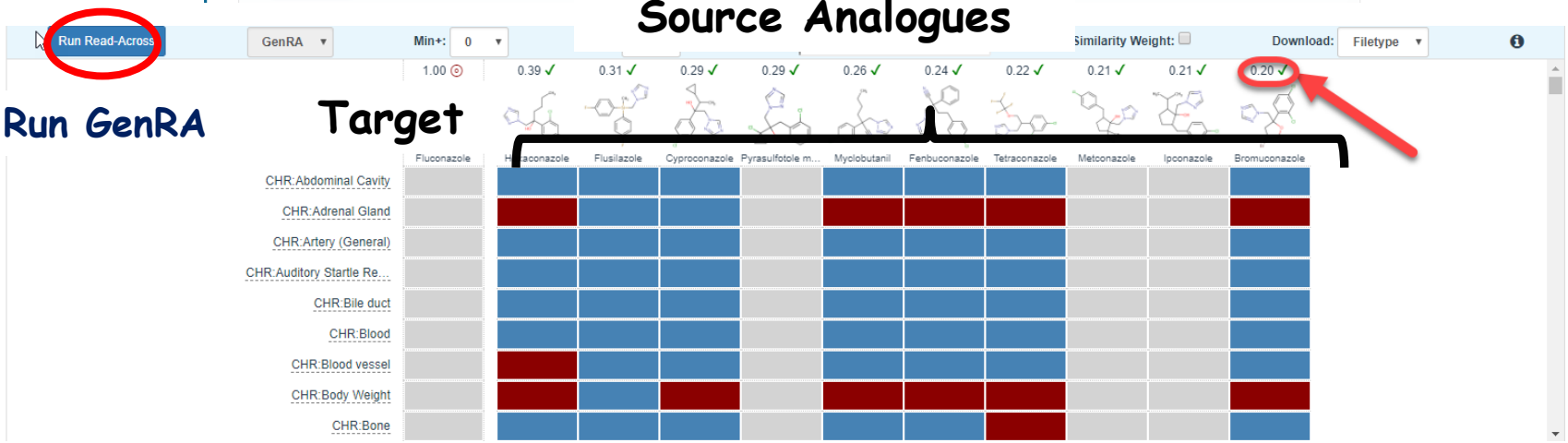


Data gap analysis

GenRA tool in practice



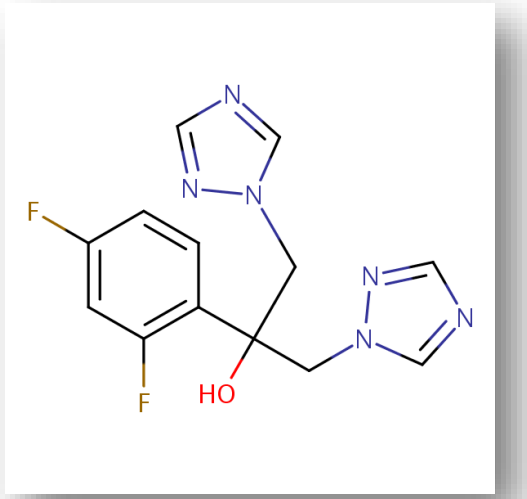
Source Analogues



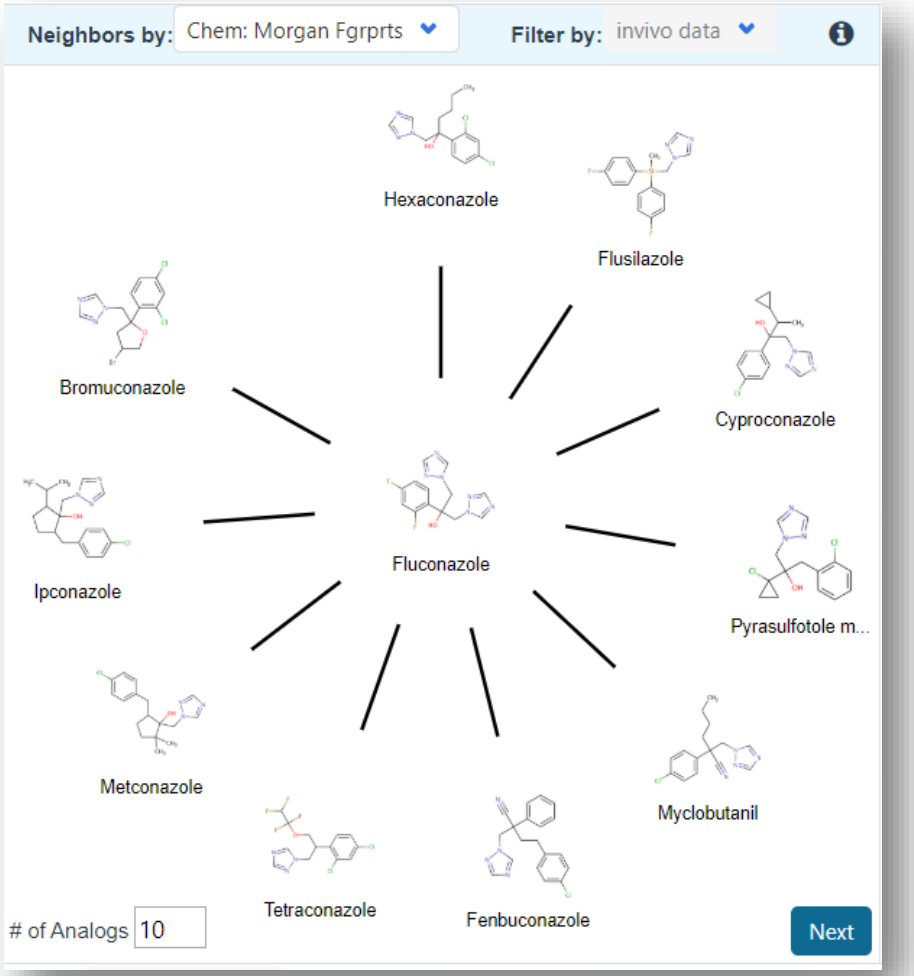
Run GenRA

Target

GenRA in practice – step by step

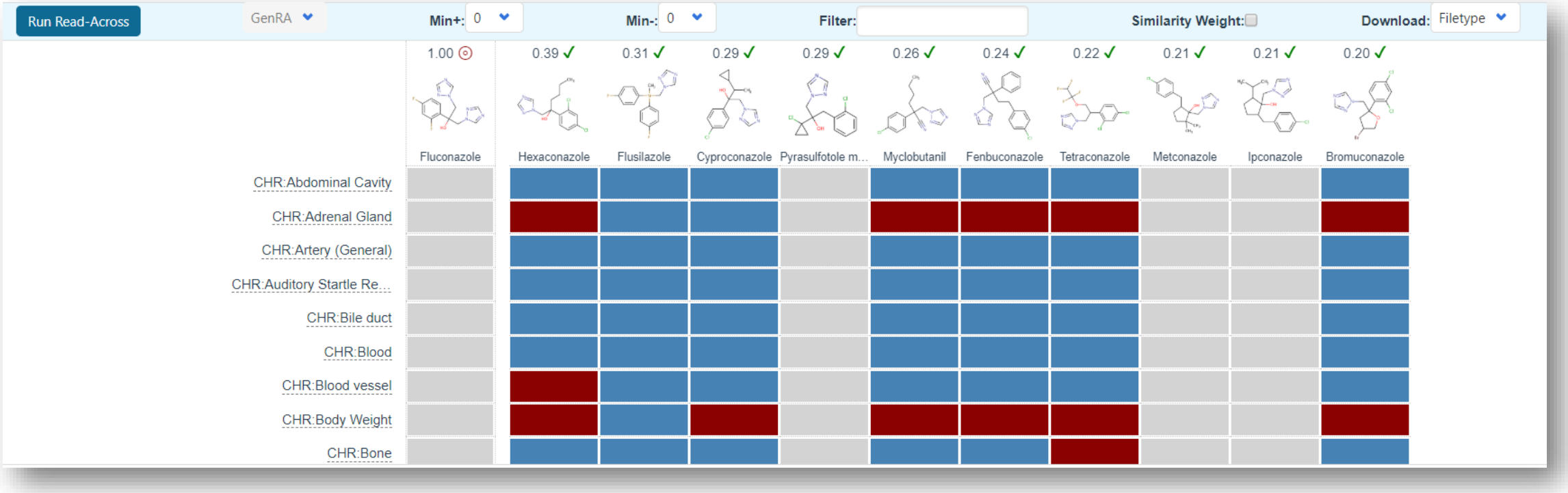


- Analogue identification:
- Similarity based on Morgan chemical fingerprints and selecting a default of 10 source analogues

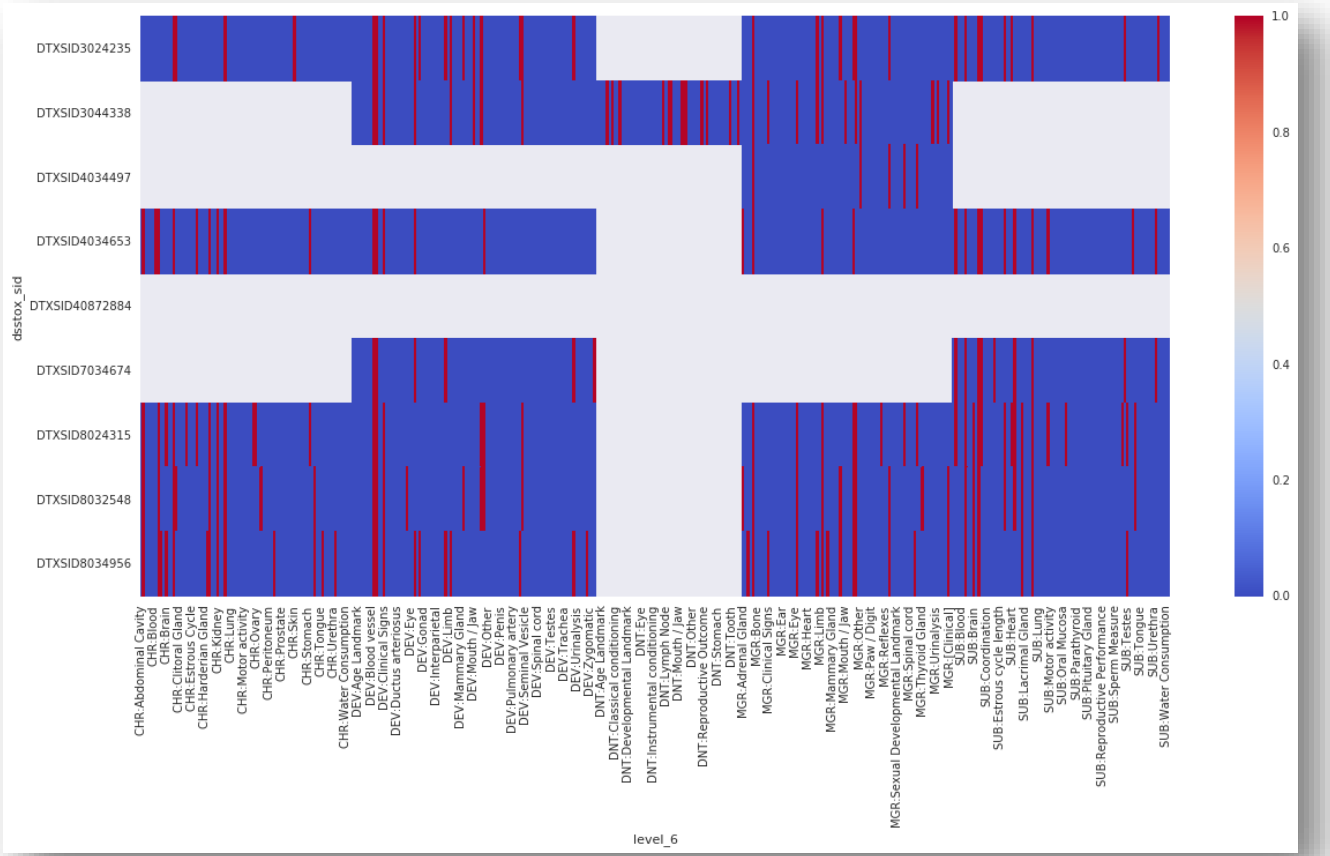


GenRA in practice – step by step

- Data matrix view of source analogues relative to target chemical



GenRA in practice – step by step

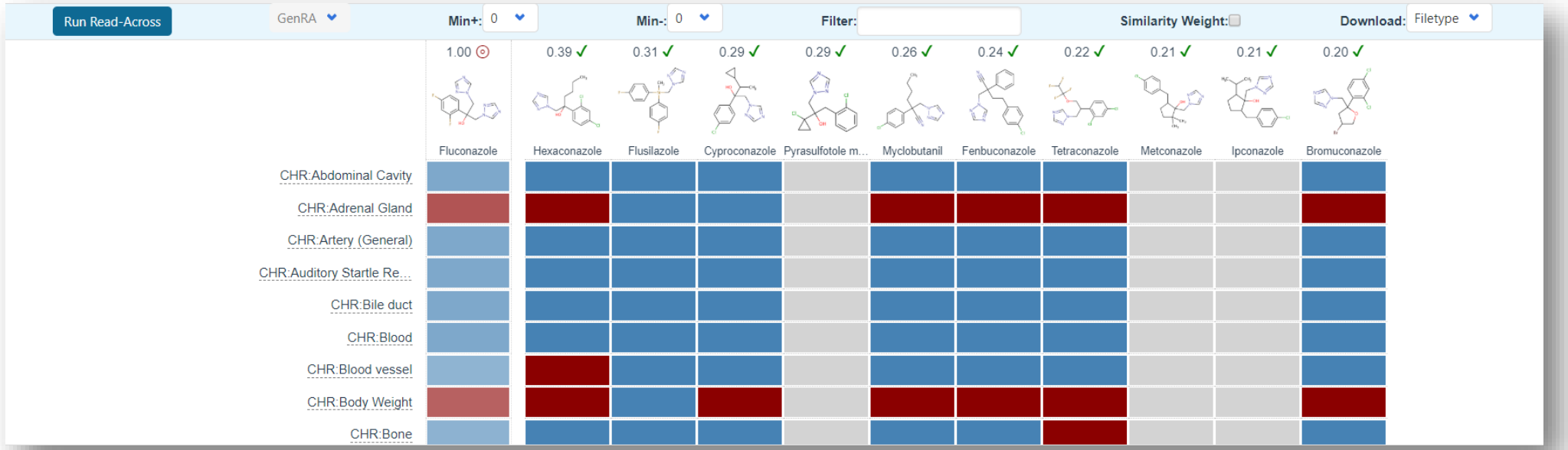


Look for commonality in profile across target effects

What are the most common effects across analogues

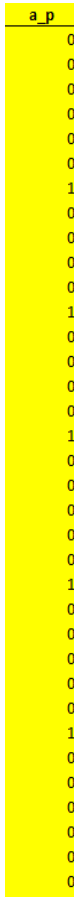
GenRA in practice – step by step

- Updated Data matrix view with GenRA predictions for target chemical



- Predictions are binary (yes/no) for toxicity effects within ToxRefDB v1 studies.
- Predictions summarised on a study level basis. Red: "positive" and Blue: "negative".

GenRA in practice: Approach



ALTEX preprint
published February 4, 2019
doi:10.14573/altex.1811292

Short Communication

Generalized Read-Across (GenRA): A workflow implemented into the EPA CompTox Chemicals Dashboard

George Helman[1,2], Imran Shah[2], Antony J. Williams[2], Jeff Edwards[2], Jeremy Dunne[2] and Grace Patlewicz[2]*

[1]Oak Ridge Institute for Science and Education (ORISE), Oak Ridge, TN, USA; [2]National Center for Computational Toxicology (NCCT), Office of Research and Development, US Environmental Protection Agency, Research Triangle Park (RTP), NC, USA

Abstract

Generalized Read-Across (GenRA) is a data driven approach which makes read-across predictions on the basis of a similarity weighted activity of source analogues (nearest neighbors). GenRA has been described in more detail in the literature (Shah et al., 2016; Helman et al., 2018). Here we present its implementation within the EPA's CompTox Chemicals Dashboard to provide public access to a GenRA module structured as a read-across workflow. GenRA assists researchers in identifying source analogues, evaluating their validity and making predictions of *in vivo* toxicity effects for a target substance. Predictions are presented as binary outcomes reflecting presence or absence of toxicity together with quantitative measures of uncertainty. The approach allows users to identify analogues in different ways, quickly assess the availability of relevant *in vivo* data for those analogues and visualize these in a data matrix to evaluate the consistency and concordance of the available experimental data for those analogues before making a GenRA prediction. Predictions can be exported into a tab-separated value (TSV) or Excel file for additional review and analysis (e.g., doses of analogues associated with production of toxic effects). GenRA offers a new capability of making reproducible read-across predictions in an easy-to use-interface.

GenRA – Ongoing research

- Summarising and aggregating the toxicity effect predictions to guide end users
- Consideration of other information to define and refine the analogue selection – e.g. physicochemical similarity, metabolic similarity, reactivity similarity…
- EPA New Chemical Categories
- Quantifying the impact of physicochemical similarity on read-across performance
- Quantifying the impact of reactivity similarity on read-across performance

GenRA – Ongoing research

- Dose response information to refine scope of prediction beyond binary outcomes

- Transitioning from qualitative to quantitative predictions – how to apply and interpret GenRA in screening level hazard assessment (e.g. effect level or point-departure [NOAEL, LOAEL, etc.] predictions)

- Using quantitative data from acute rat oral toxicity, ToxRefDB v2 [1 manuscript submitted, 1 in internal clearance]

Take home messages

- Harmonised framework for read-across provides opportunities for expanded integration of NAM data
- GenRA developed is aligned with this framework
- Initial GenRA (baseline) considers structural similarity but current work has evaluated the quantitative impact of physicochemical similarity (as it relates to bioavailability) and transitioned to quantitative predictions of effect levels or PODs
- Illustrated how GenRA baseline can been applied in practice

Acknowledgements

- Many but in particular..
- NCCT
- Imran Shah
- George Helman
- Tony Williams
- Jeff Edwards
- Jason Lambert

- NCEA
- Lucy Lizarraga