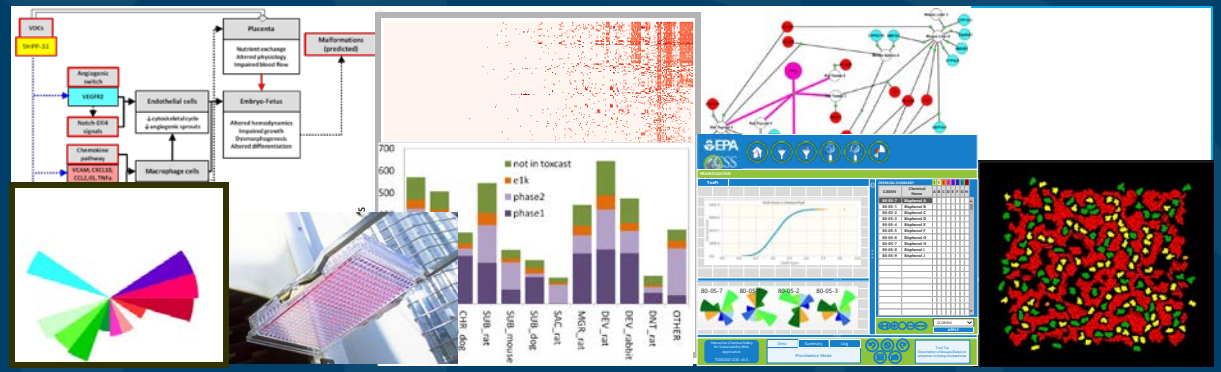
Navigating Through the Minefield of Read-Across Tools and Frameworks: An Update on Generalised Read-Across (GenRA)



Grace Patlewicz, George Helman*, Imran Shah

National Center for Computational Toxicology
***ORISE Fellow**

Conflict of Interest Statement

No conflict of interest declared.

Disclaimer:

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

Outline

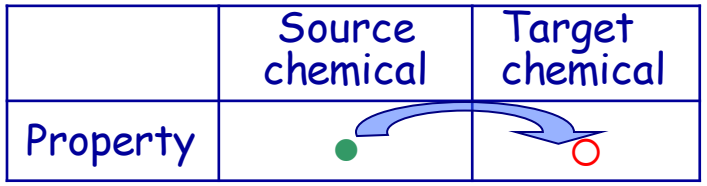
- Background and Definitions
- Workflow for category development and read-across
- Current tools and approaches
- Uncertainty assessment in read-across
- Quantifying uncertainties and Assessing Performance of read-across
- From research to implementation
- Summary

Background & definitions

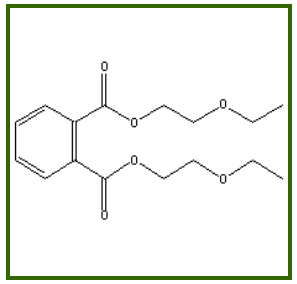
- Read-across describes one of the data gap filling techniques used within analogue and category approaches
- "Analogue approach" refers to grouping based on a very limited number of chemicals (e.g. target substance) + source substance)
- "Category approach" is used when grouping is based on a more extensive range of analogues (e.g. 3 or more members)

Definition: Read-across

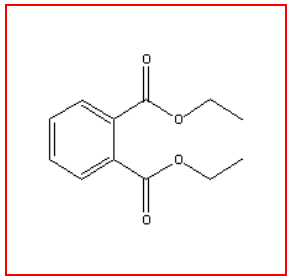
Known information on the property of a substance (source) is used to make a prediction of the same property for another substance (target) that is considered "similar" i.e. endpoint & often study specific



- Reliable data
- Missing data



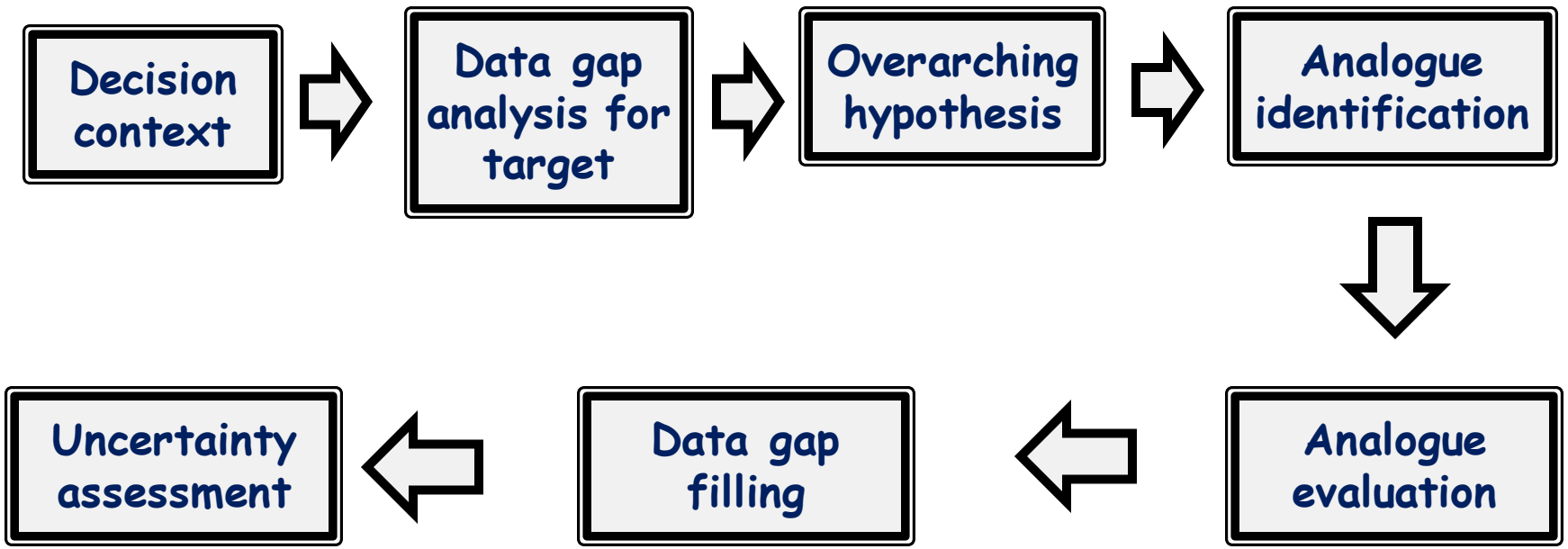
Acute oral toxicity?



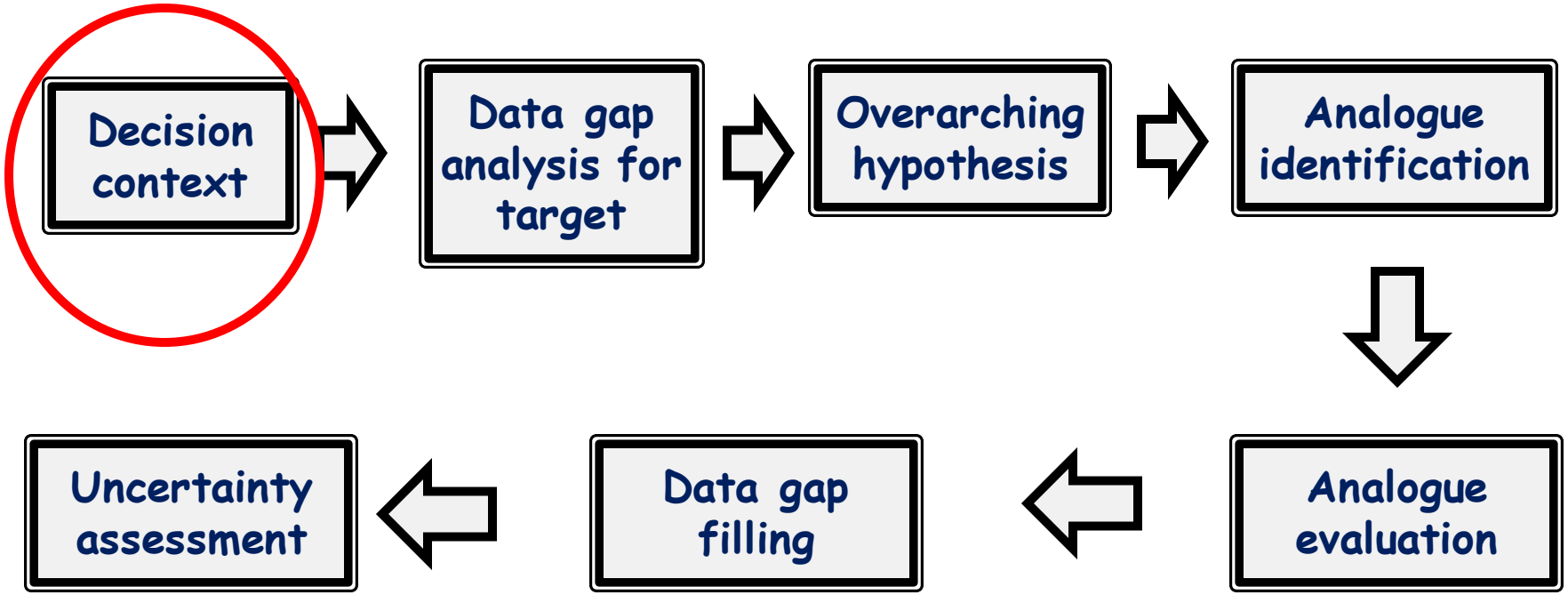
Known to be harmful

Predicted to be harmful

The Category Workflow



The Category Workflow

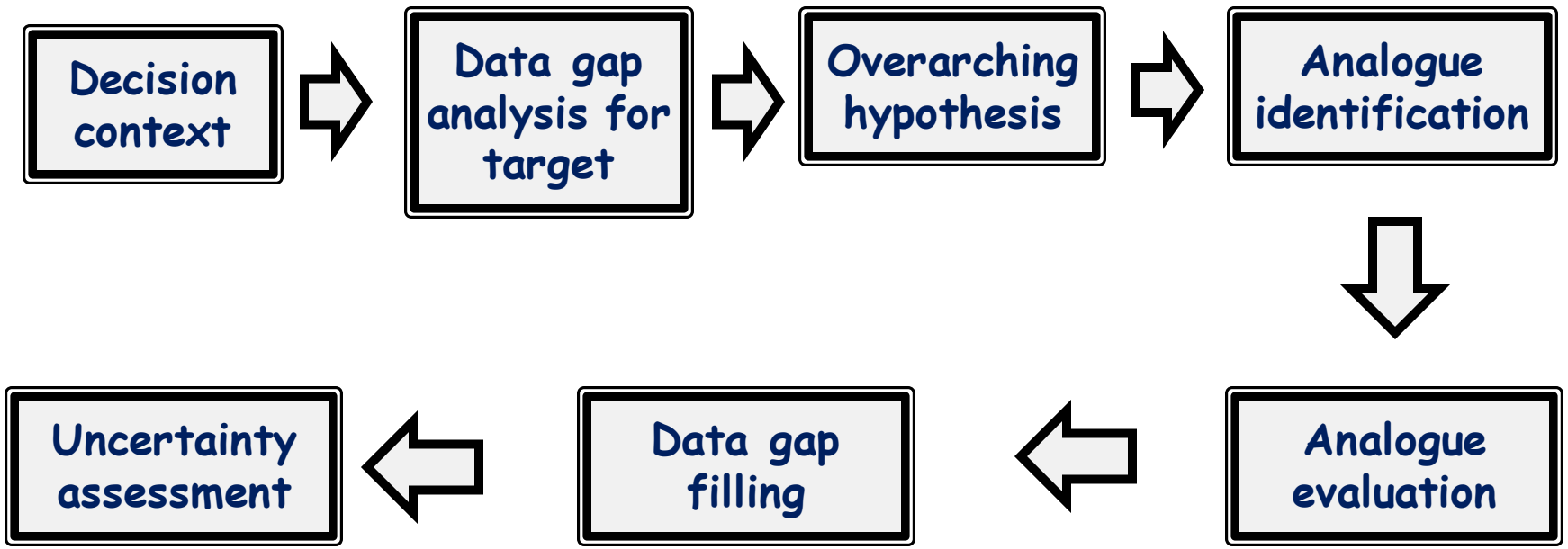


Decision Context

- Prioritisation, e.g. PMN
- Screening level hazard assessment
- Risk Assessment, e.g. PPRTV

- Different decision contexts will dictate the level of uncertainty that can be tolerated

The Category Workflow

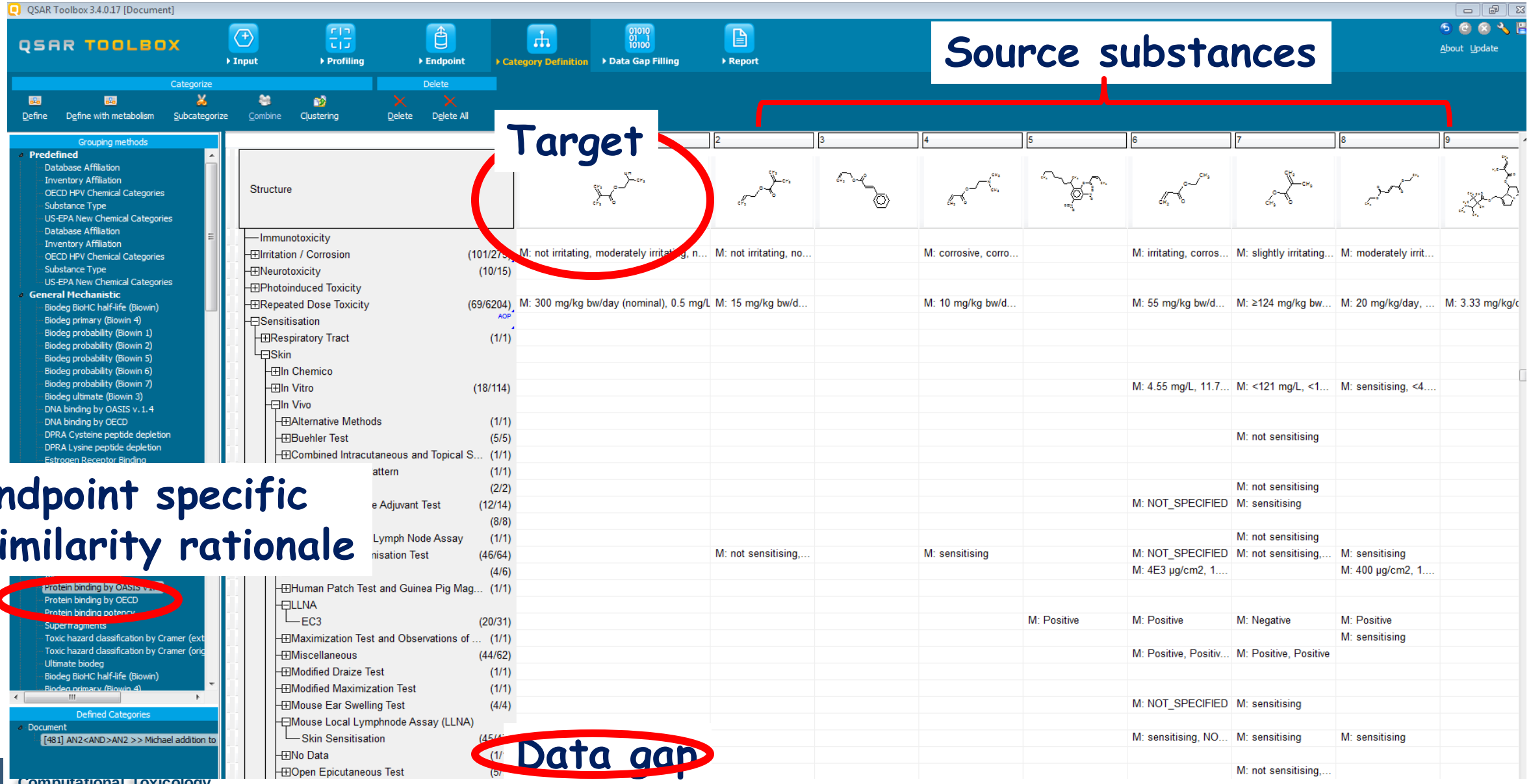


Selected Read-Across Tools

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

Analogue identification & evaluation within the OECD Toolbox

United States Environmental Protection Agency



Endpoint specific Similarity rationale

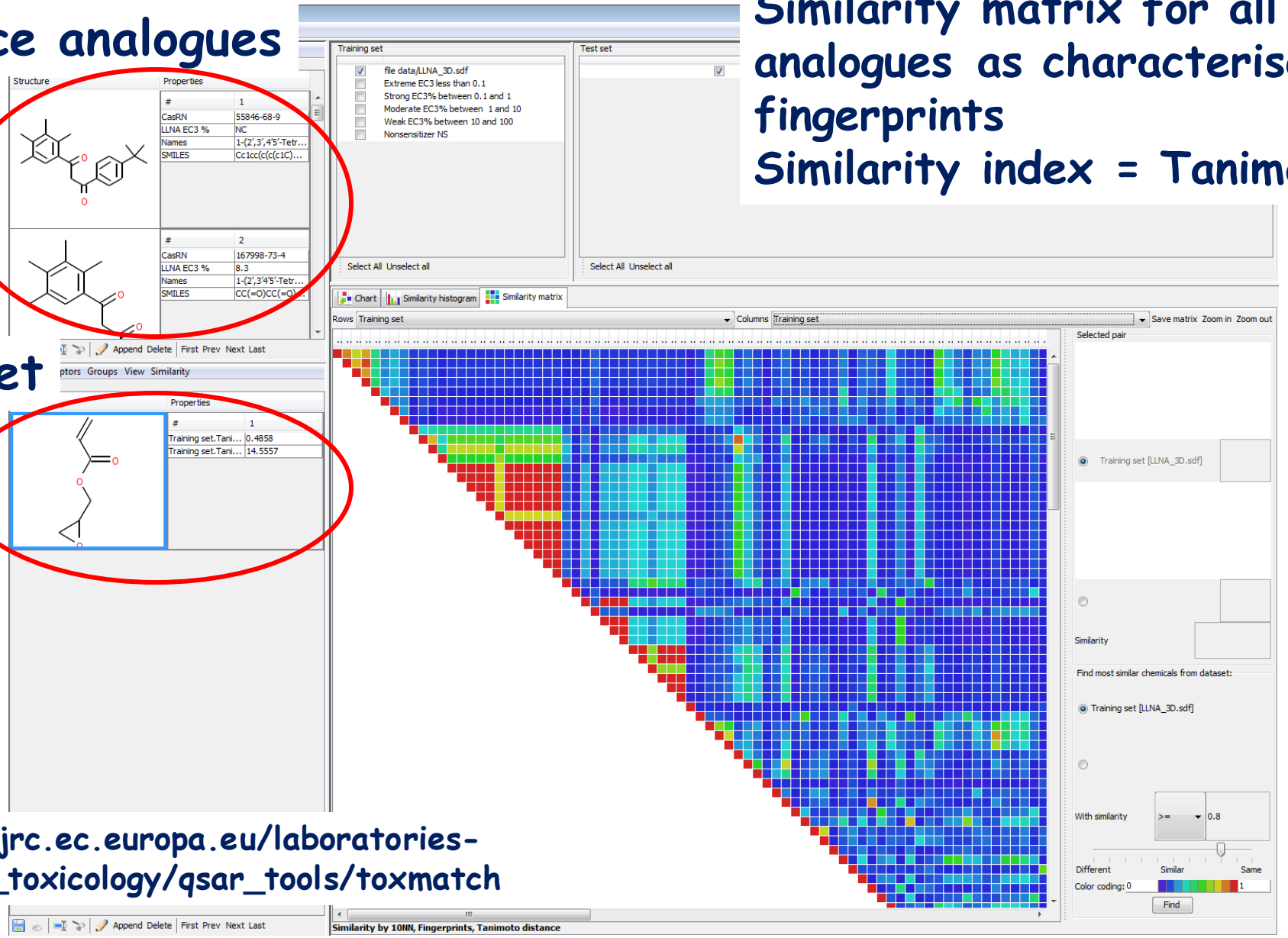
Analogue identification & evaluation within Toxmatch

Source analogues

Target

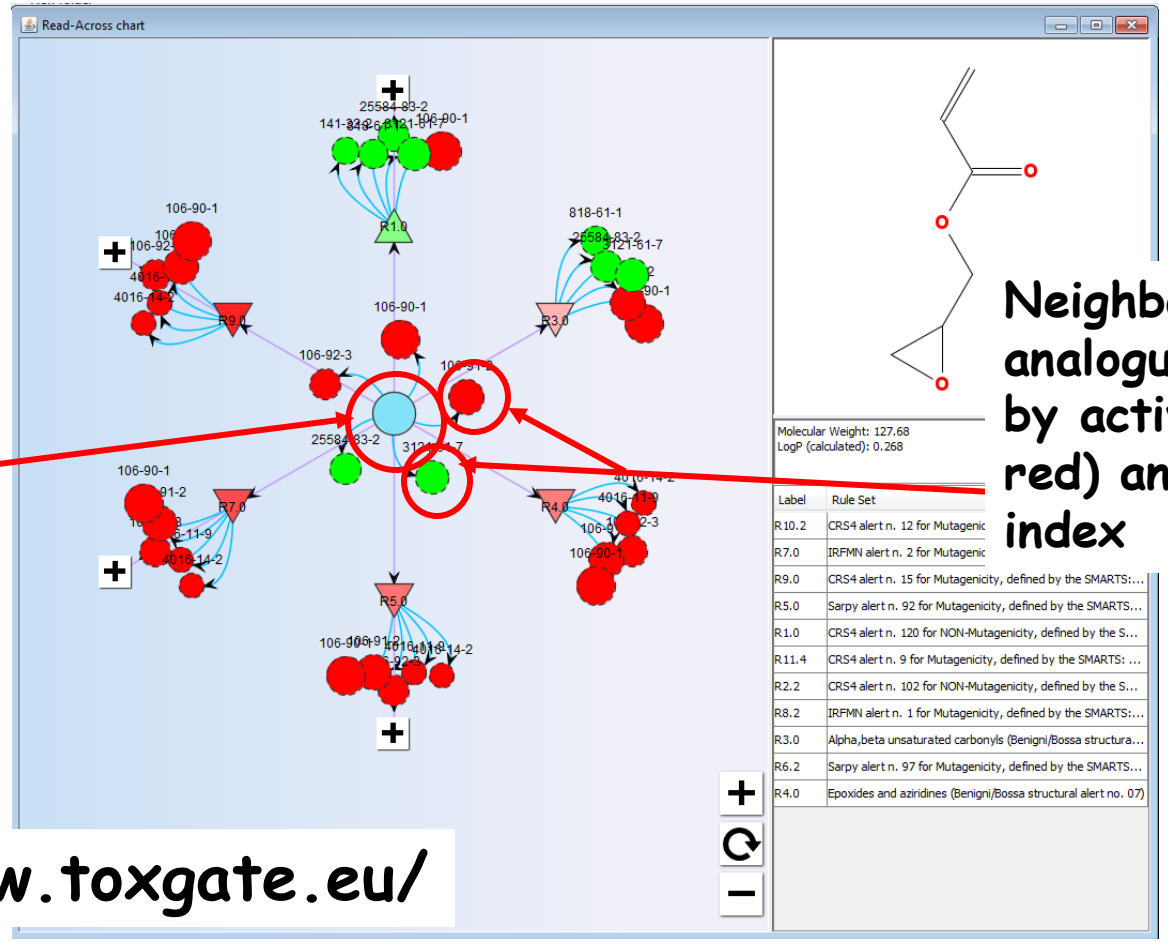
Similarity matrix for all source analogues as characterised by fingerprints

Similarity index = Tanimoto distance



https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxmatch

ToxRead



Target

Neighboring source analogues, colour coded by activity (positive = red) and by similarity index

http://www.toxgate.eu/

Selected Read-Across Tools – Review paper

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

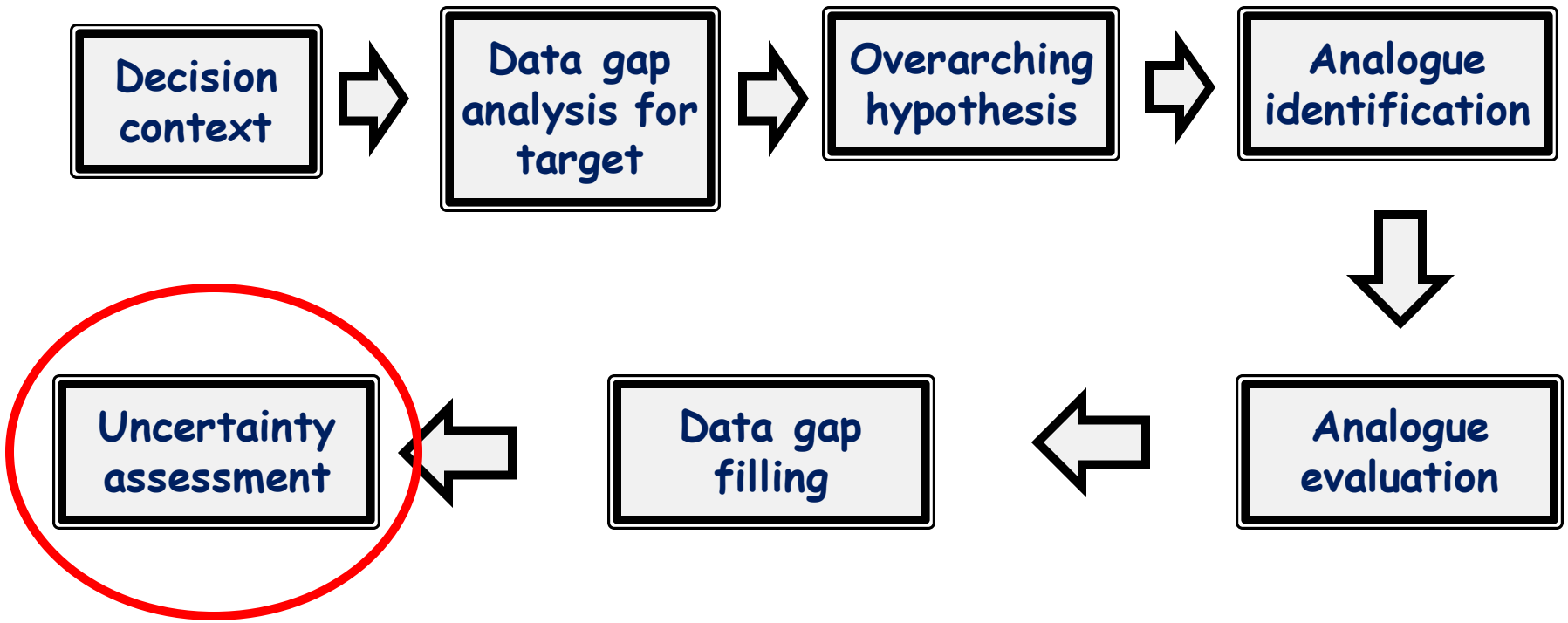
Keywords:
Category approach
Analogue approach
Data gap filling
Read-across
(Q)SAR
Trend analysis
Nearest neighbo

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.

The Category Workflow



EPA United States Environmental Protection Agency

Sources of Uncertainty

- Analogue or category approach? (# analogues)
- Completeness of the data matrix – no. of data gaps
- Data quality for the underlying analogues for the target and source analogues
- Consistency of data across the data matrix – concordance of effects and potency across analogues

Sources of Uncertainty (cont'd)

- Overarching hypothesis/similarity rationale – how to identify similar analogues and justify their similarity for the endpoint of interest
- Address the dissimilarities and whether these are significant from a toxicological standpoint e.g. ToxDelta
- Presence vs. absence of toxicity
- Toxicokinetics

Uncertainty Assessment

- A number of publications exist that can guide the construction and assessment of categories and use of read-across
  - Guidance and examples (OECD (2014), ECHA (2008), ECETOC (2012))
  - Frameworks for identifying analogues (e.g., Wu et al (2010), Patlewicz et al (2013))
  - Frameworks for assessing read-across (Blackburn and Stuard (2014), Patlewicz et al (2014), Patlewicz et al (2015), ECHA – RAAF (2015), Schultz et al (2015), Ball et al (2016))

Uncertainty assessment

- However read-across acceptance relies on a subjective expert assessment
- There is no objective measure of read-across performance
- Different approaches have been explored to characterise uncertainties both qualitatively and quantitatively
- E.g. Blackburn and Stuard (qualitative), Molecular Networks (quantitative), EPA NCCT (quantitative and generalisable)

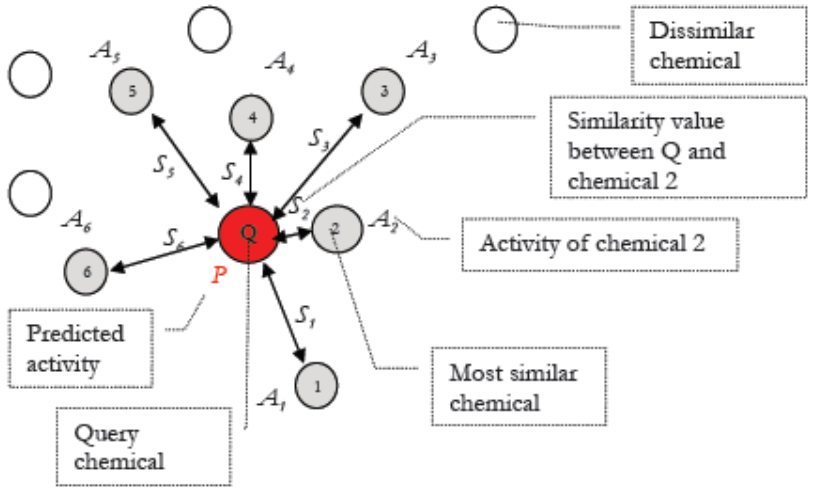
Quantifying Uncertainty & Assessing Performance of Read-Across

- GenRA (Generalised Read-Across) is a "local validity" approach
- Predicting toxicity as a similarity-weighted activity of nearest neighbors based on chemistry and bioactivity descriptors
- Generalised version of Chemical-Biological Read-Across (CBRA) developed by Low et al (2013)
- Systematically evaluates read-across performance and uncertainty using available data

$$y_i^{\beta,\alpha} = \frac{\sum_j^k s_{ij}^{\alpha} x_j^{\beta}}{\sum_j^k s_{ij}^{\alpha}}$$

Jaccard similarity:

$$s_{ij} = \frac{\sum_l (x_{il} \wedge x_{jl})}{\sum_l (x_{il} \vee x_{jl})}$$



GenRA - Approach

I. Data

1,778 Chemicals
3,239 Structure descriptors (chm)
820 Bioactivity assays (bio)
ToxCast
574 Apical outcomes (tox)
ToxRefDB

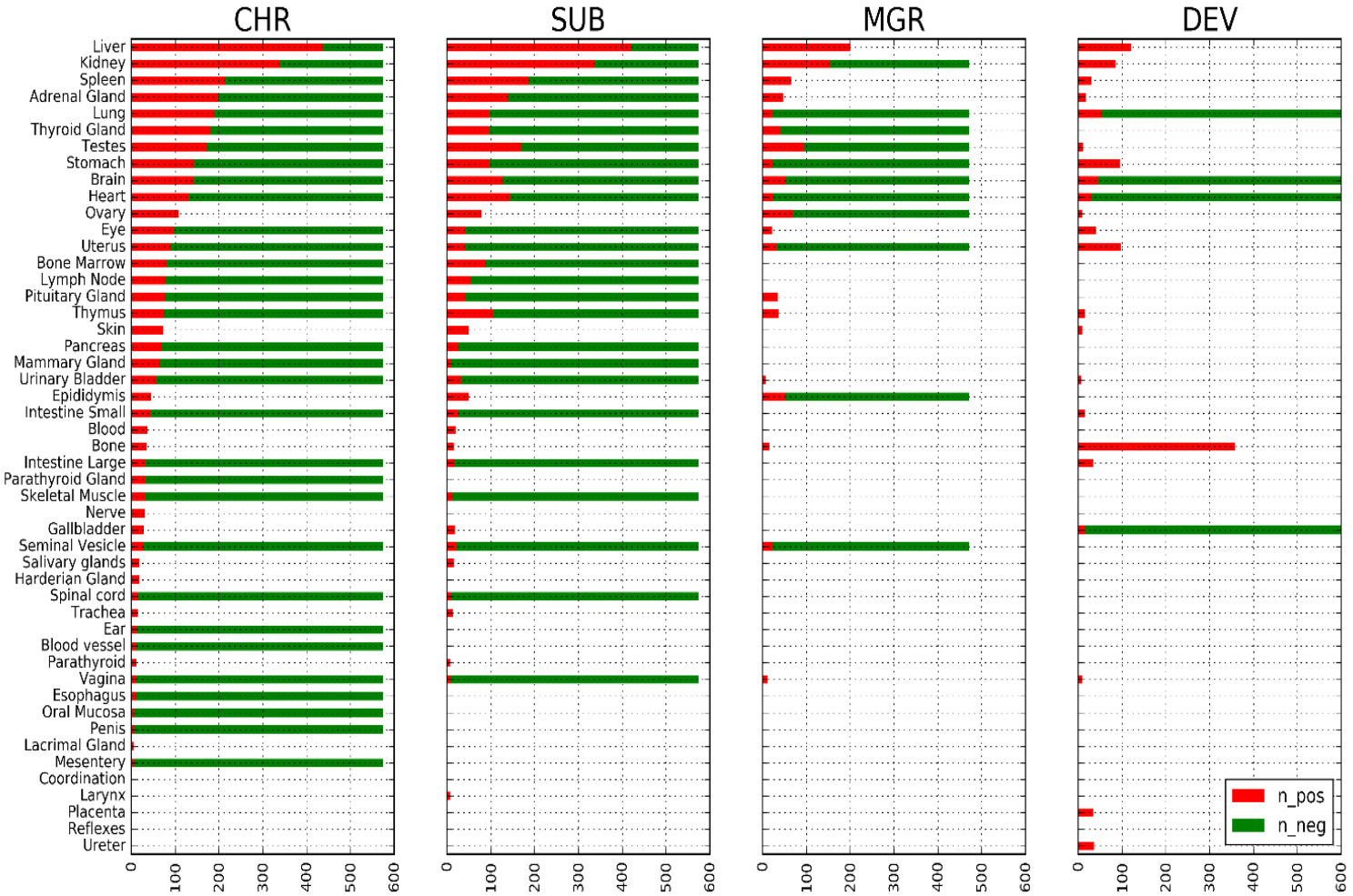
II. Define Local neighborhoods

Use K-means analysis to group chemicals by similarity
Use cluster stability analysis
~ 100 local neighborhoods

III. GenRA

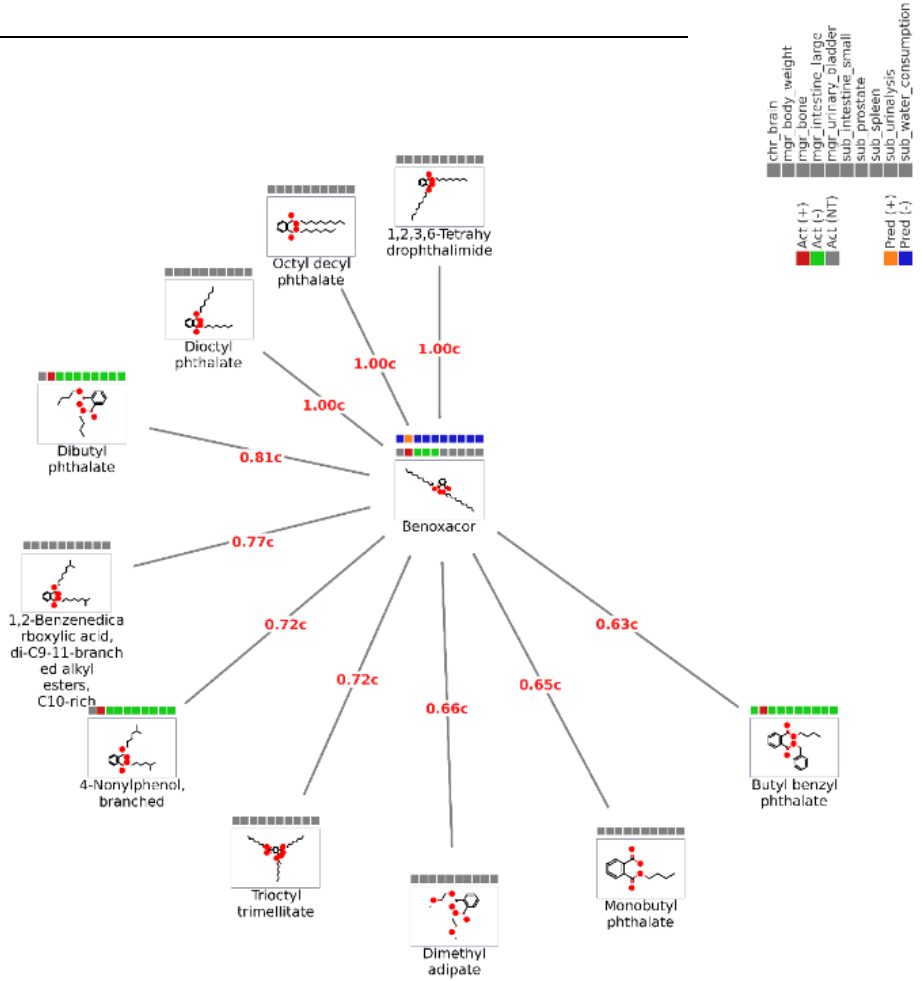
Use GenRA to predict apical outcomes in local neighbor hoods
Evaluate impact descriptors (chm, bio, bc) on prediction
Quantify uncertainty

GenRA - Toxicity Data from ToxRefDB

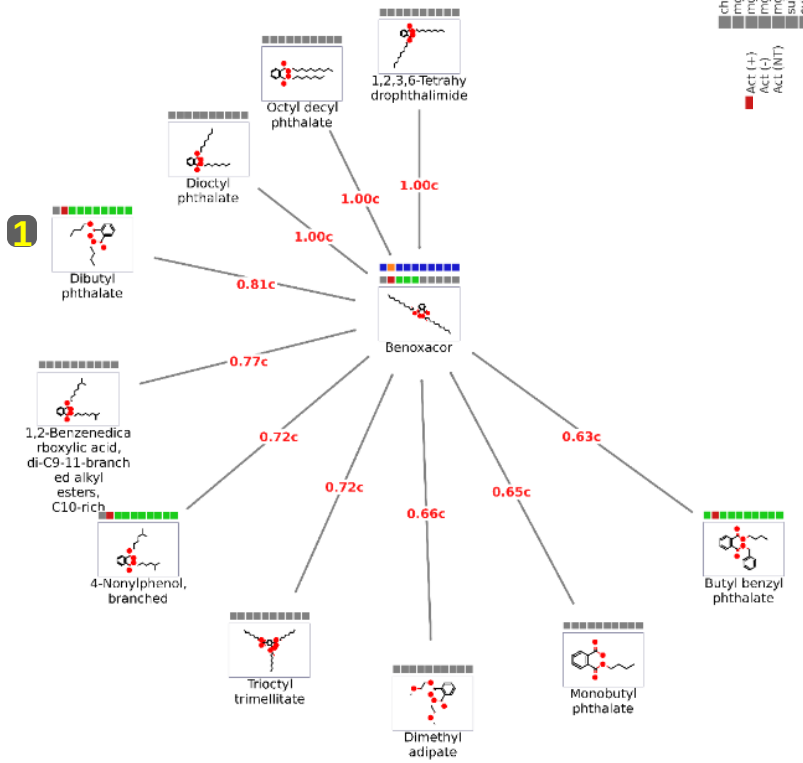


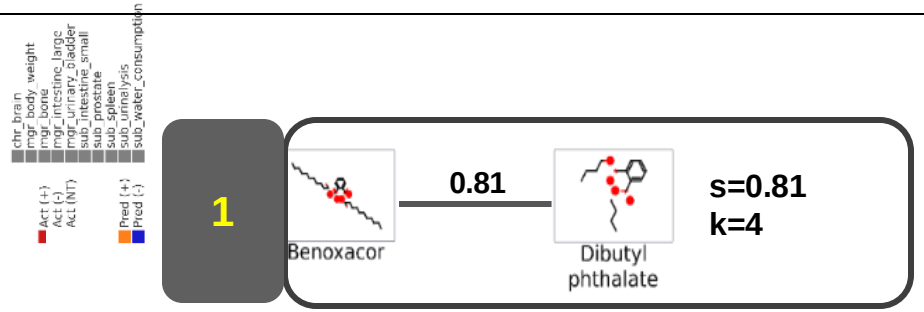
GenRA – Performance in Each Cluster

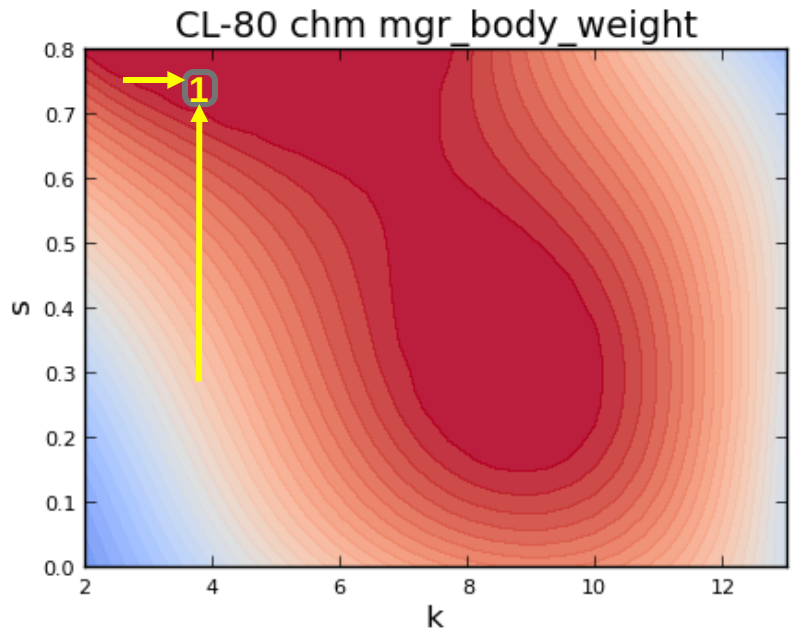
- Use GenRA to predict the similarity weighted toxicity scores for each
  - Toxicity type (β)
  - Descriptor ={chm,bio,bc} (α)
  - No. of nearest neighbors (k)
  - Similarity score threshold (s_{ij}^{α})
- Calculate performance by comparing predicted y^{tox} and true x^{tox} for all chemicals using area under ROC curve (AUC)
- Results: {cluster, α, β, k, s, AUC}



GenRA - Analysing Local Neighborhood of a Chemical







GenRA – Insights and Next Steps

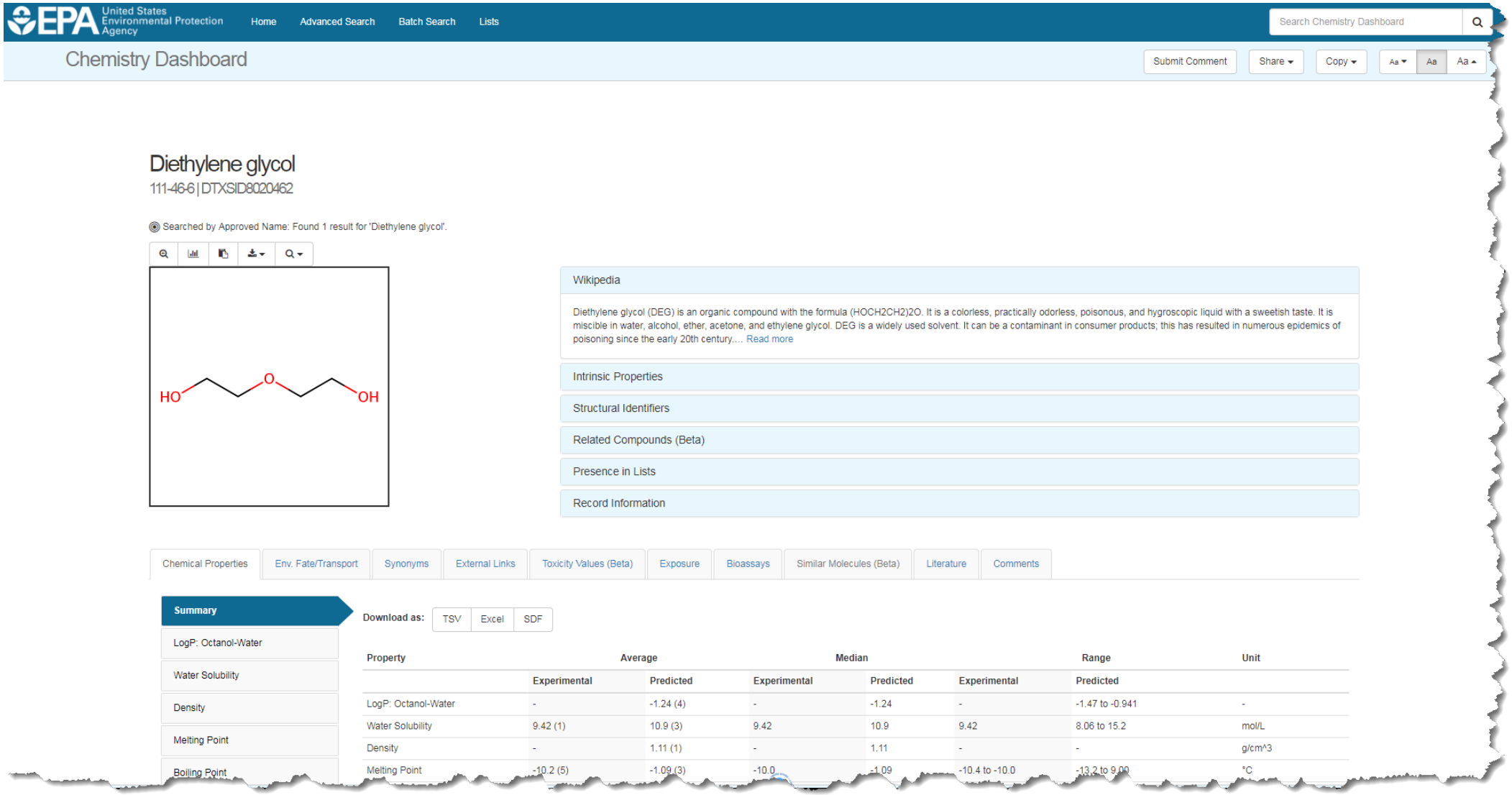
- The approach enabled a performance baseline for read-across predictions of specific study outcomes to be established but was still context dependent on the endpoint and the chemical

- Ongoing analysis:
- Consideration of other information to refine the analogue selection – e.g. TK similarity, metabolic similarity, reactivity similarity…

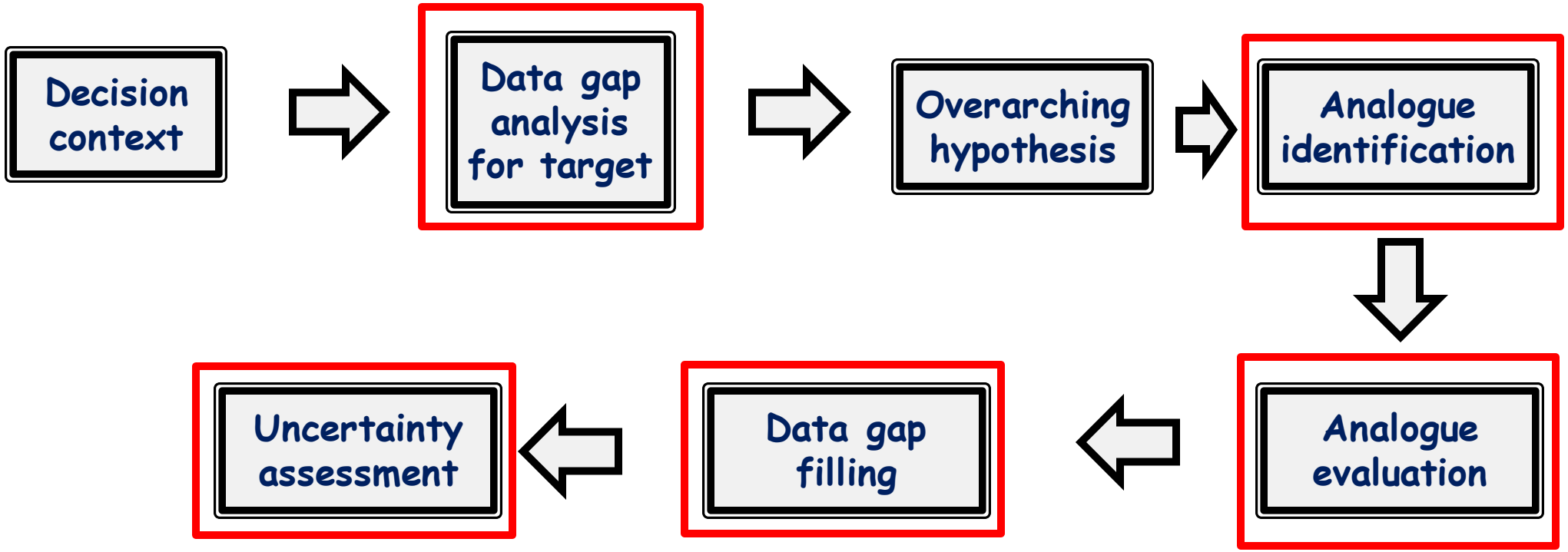
From research to implementation: GenRA prototype

- Intent is to integrate objective read-across functionality as part of ongoing dashboard efforts see https://comptox.epa.gov/dashboard
- A limited release of GenRA is currently undergoing internal beta testing
- A video tutorial and help manual has been created to explain the approach and how to use the tool

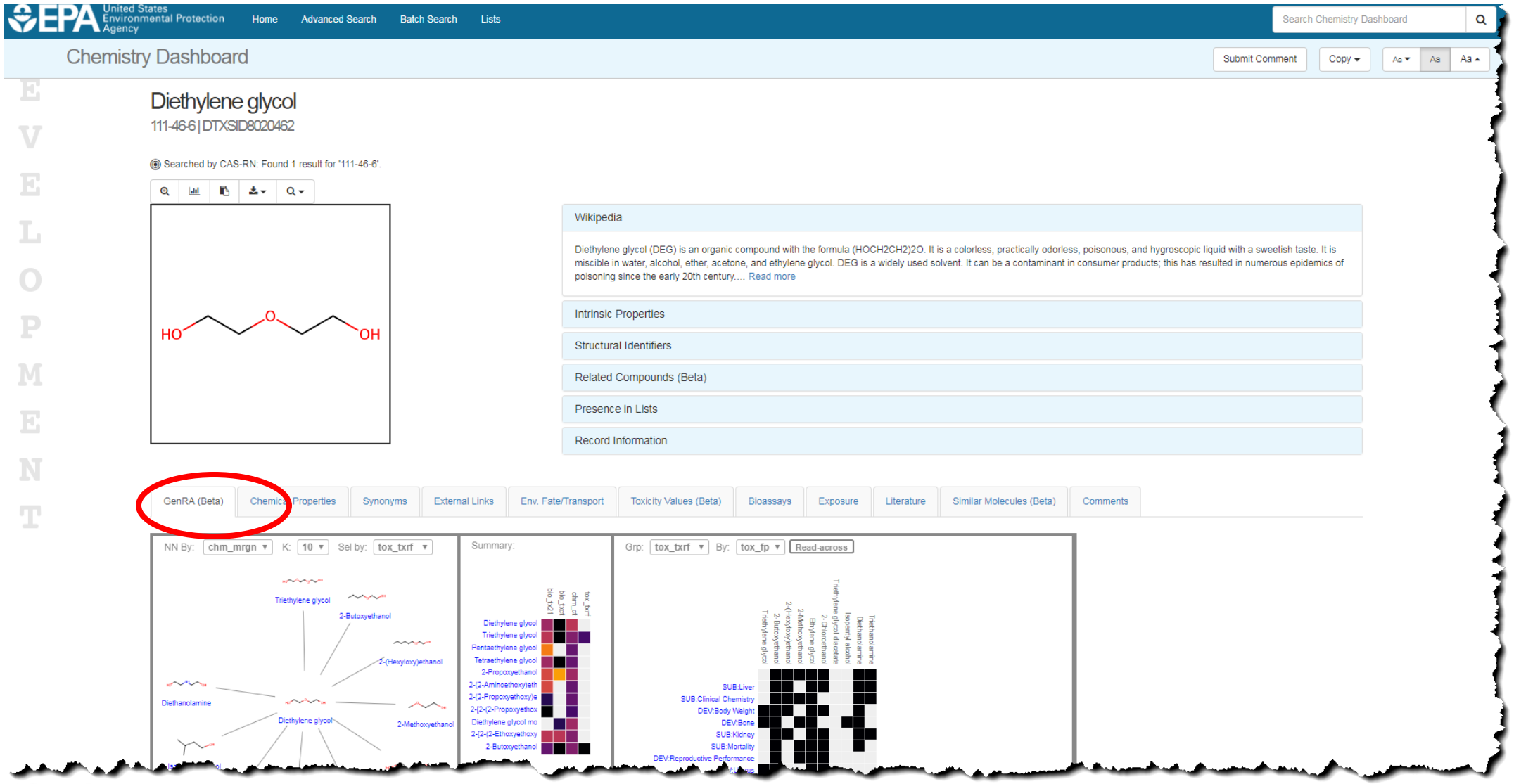
From research to implementation



GenRA prototype development



Basic Integration via GenRA tab

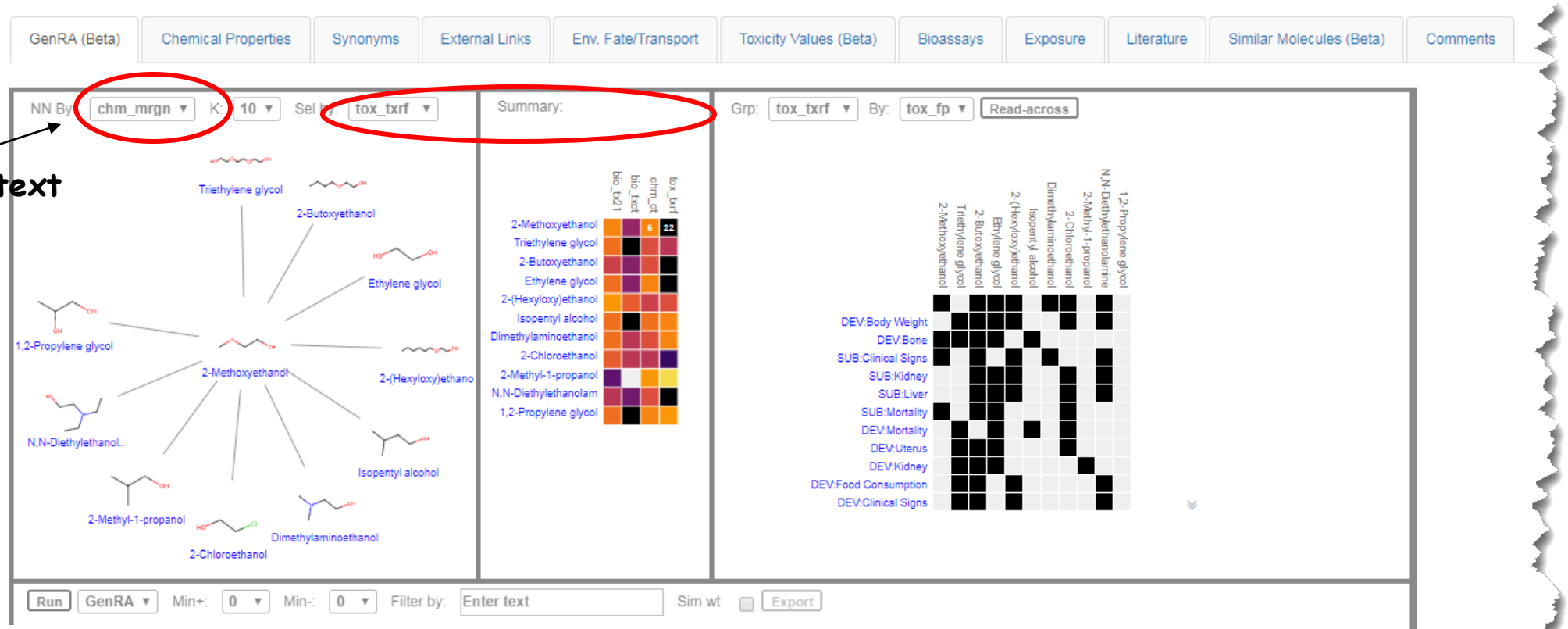


Selected Read-Across Tools

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	Beta for Internal testing

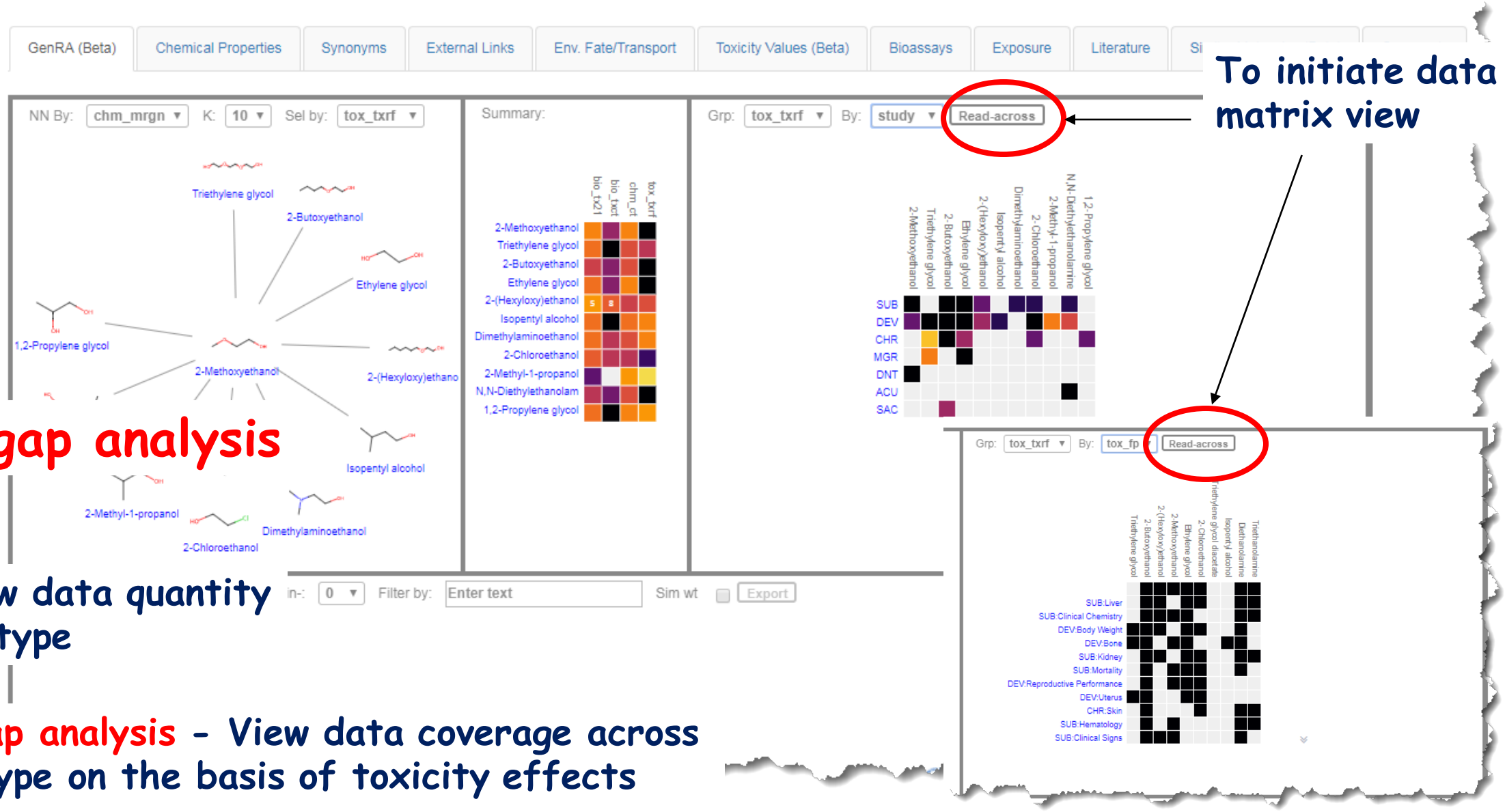
Working interface

Grid interface where windows are dynamically updated in subsequent windows



Similarity context

Analogue identification: Search for source analogues on the basis of chemical fingerprints, filtered by availability of in vivo data



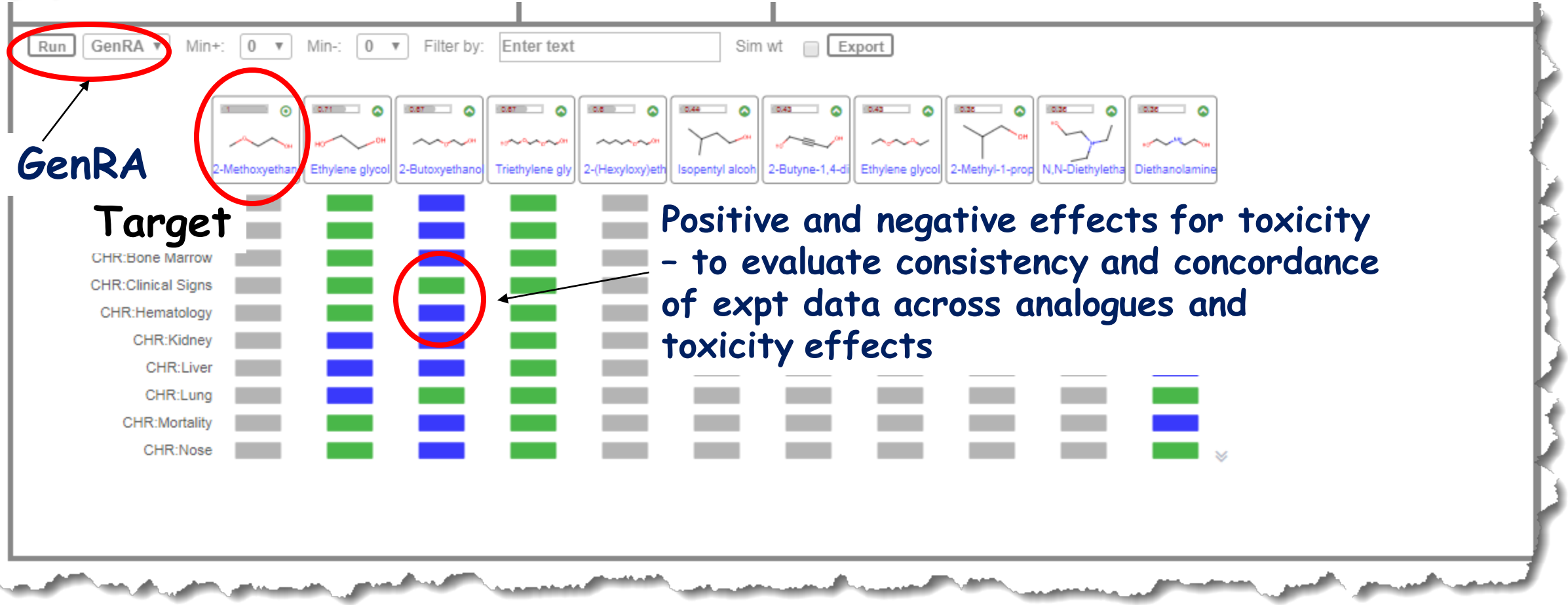
To initiate data matrix view

Data gap analysis

View data quantity by type

Data gap analysis - View data coverage across study type on the basis of toxicity effects

Analogue evaluation using data matrix view

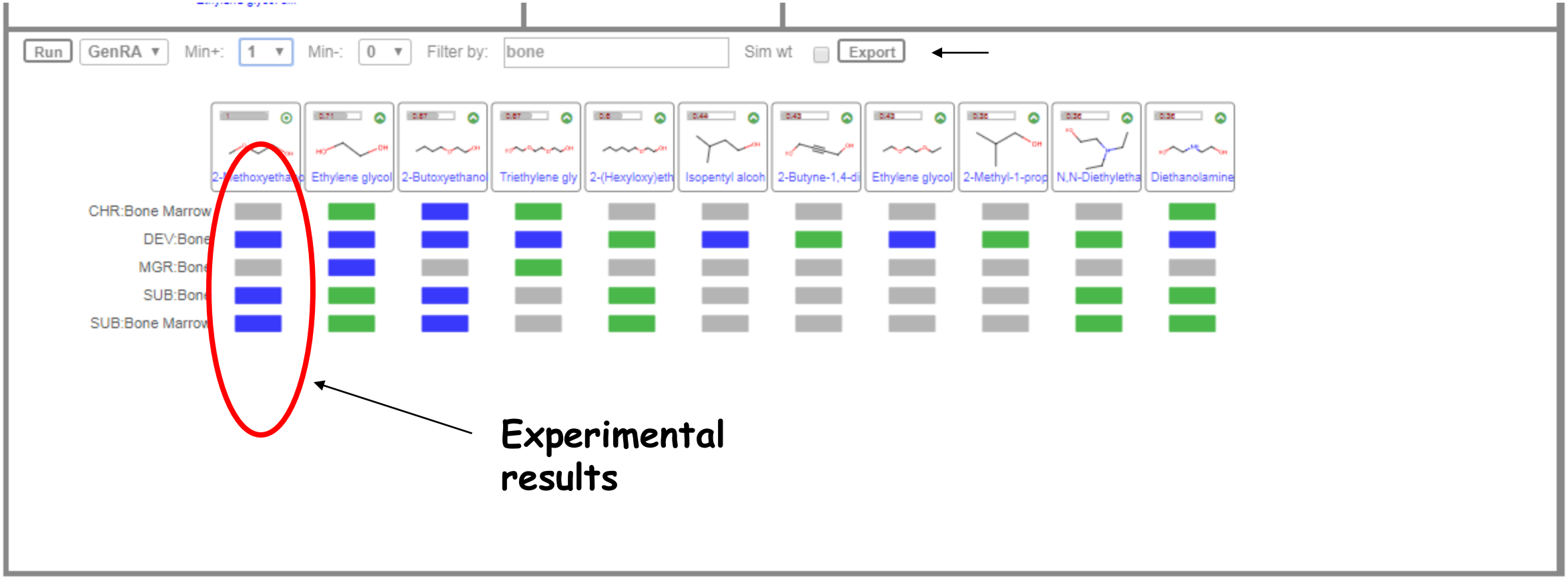


Run GenRA

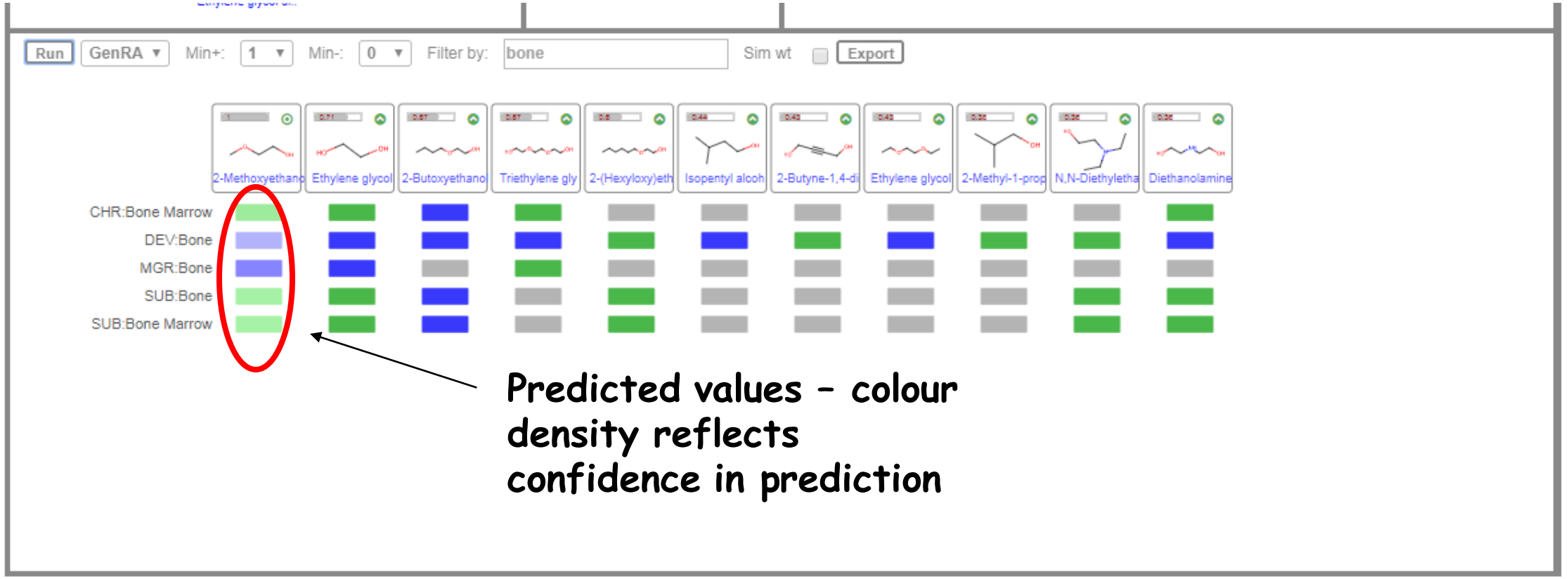
Target

Positive and negative effects for toxicity – to evaluate consistency and concordance of expt data across analogues and toxicity effects

Data gap filling using GenRA within data matrix



Data gap filling using GenRA within data matrix



Exported results using GenRA

A	B	C	D	E	F	G	H	I	J	K
cls	target	analog	analog	analog	analog	analog	analog	analog	analog	
label	2-Methoxyethanol	Ethylene glycol	2-Butoxyethanol	Triethyler	2-(Hexylo	Isopentyl	2-Butyne-	Ethylene g	2-Methyl-1-propanol	
dsstox_cid	DTXCID804182	DTXCID40597	DTXCID904097	DTXCID60	DTXCID60	DTXCID70	DTXCID90	DTXCID30	DTXCID601759	
casrn	109-86-4	107-21-1	111-76-2	112-27-6	112-25-4	123-51-3	110-65-6	629-14-1	78-83-1	
jaccard	1	0.714285714	0.666666667	0.666667	0.6	0.444444	0.428571	0.428571	0.375	
CHR:Bone Marrow	GenRA Neg Act=0 (0.326) AUC=0 p=0.685	no_effect	125.000 ppm	no_effect	no_data	no_data	no_data	no_data	no_data	
DEV:Bone	GenRA TP Act=1 (1) AUC=0 p=1(50.000 ppm)	750.000 mg/kg/day	100.000 ppm	5630.000	no_effect	0.500 p	no_effect	100.000	no_effect	
MGR:Bone	GenRA Pos Act=1 (0.517) AUC=0 p=0.51	1333.330 mg/kg/day	no_data	no_effect	no_data	no_data	no_data	no_data	no_data	
SUB:Bone	GenRA FN Act=0 (0.483) AUC=0 p=0.66(546.000 mg/kg/day)	no_effect	500.000 ppm	no_data	no_effect	no_data	no_data	no_data	no_data	
SUB:Bone Marrow	GenRA FN Act=0 (0.483) AUC=0 p=0.65(297.000 mg/kg/day)	no_effect	62.500 ppm	no_data	no_effect	no_data	no_data	no_data	no_data	

Demo

Summary

- Still many challenges remain in read-across – what information is relevant to integrate and ways in which that integration can be performed
- Quantifying the uncertainty of read-across prediction is a critical issue
- Have illustrated the research directions being taken within NCCT and work to implement these into practical tools

Acknowledgements

- Imran Shah
- George Helman
- Tony Williams
- Jeff Edwards
- Richard Judson
- Prachi Pradeep
- Chris Grulke