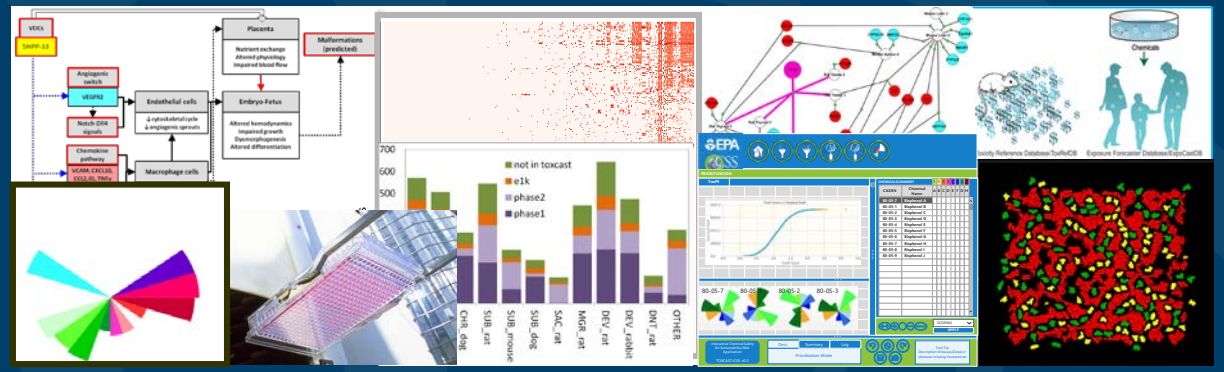
Quantifying uncertainty in read-across assessment – an algorithmic approach



SOT 2017 Workshop: Opportunities for read-across development and application using QSAR approaches
Tuesday 14th March 2017

Grace Patlewicz
National Center for Computational Toxicology

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

Conflict of Interest Statement

No conflict of interest declared.

Disclaimer:

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

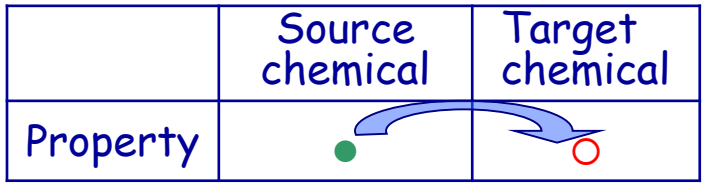
- Background and Definitions
- Workflow for category development and read-across
- Current tools and approaches
- Uncertainty assessment in read-across
- Quantifying uncertainties and Assessing Performance of read-across
- From research to implementation
- Summary

Background & definitions

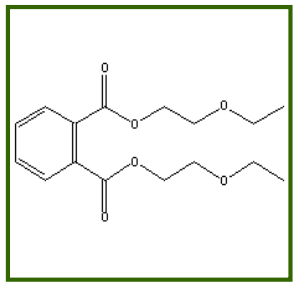
- Read-across describes one of the data gap filling techniques used within analogue and category approaches
- "Analogue approach" refers to grouping based on a very limited number of chemicals (e.g. target substance) + source substance)
- "Category approach" is used when grouping is based on a more extensive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

Definition: Read-across

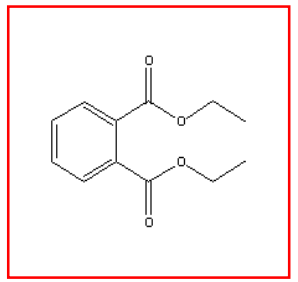
Known information on the property of a substance (source) is used to make a prediction of the same property for another substance (target) that is considered "similar" i.e. endpoint & often study specific



- Reliable data
- Missing data



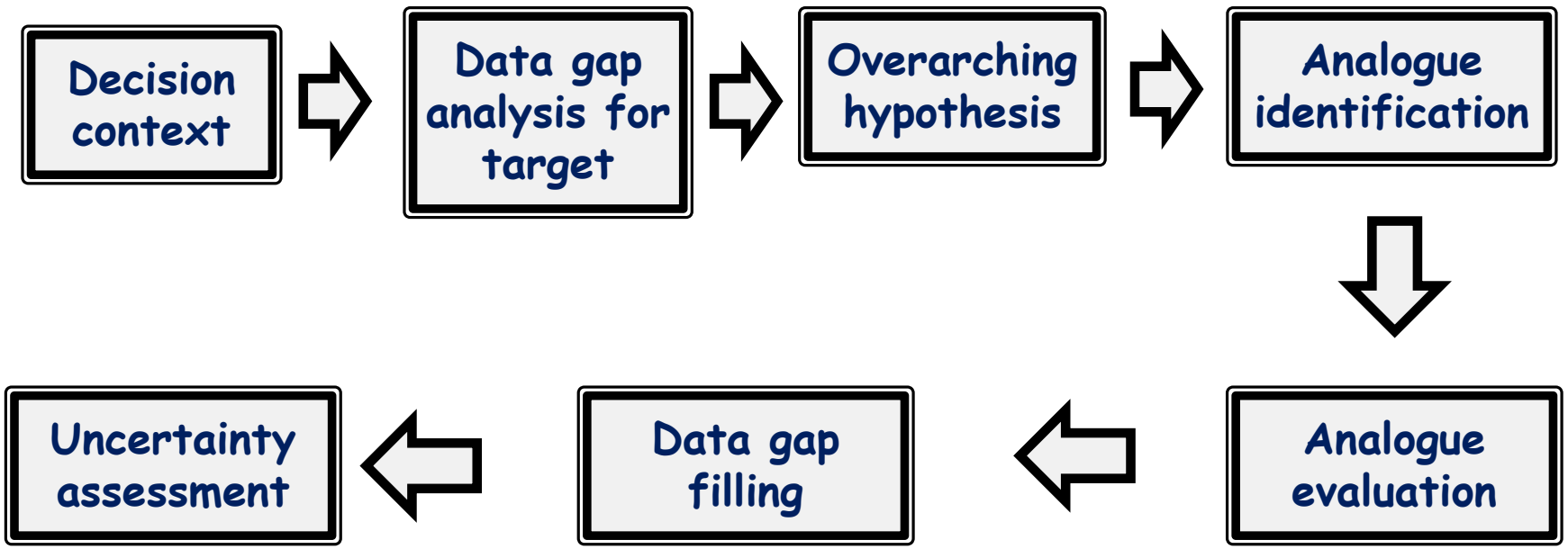
Acute fish toxicity?



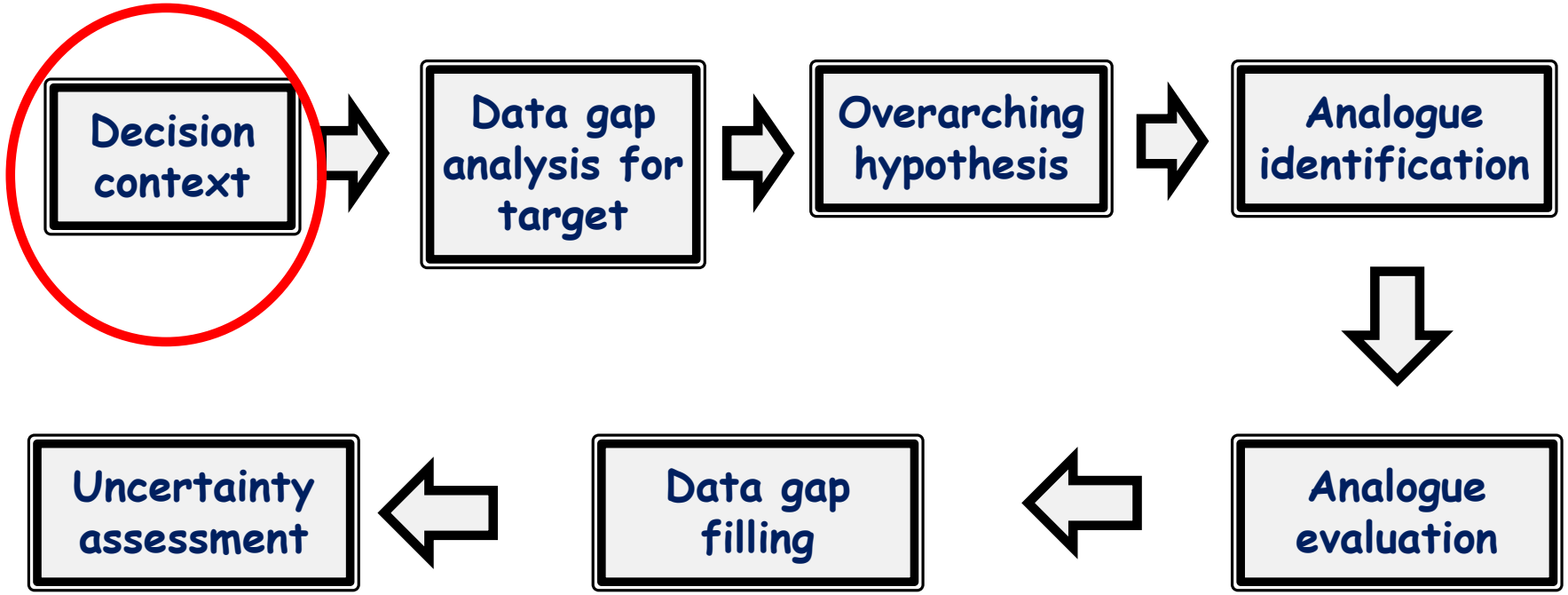
Known to be harmful

Predicted to be harmful

The Category Workflow



The Category Workflow

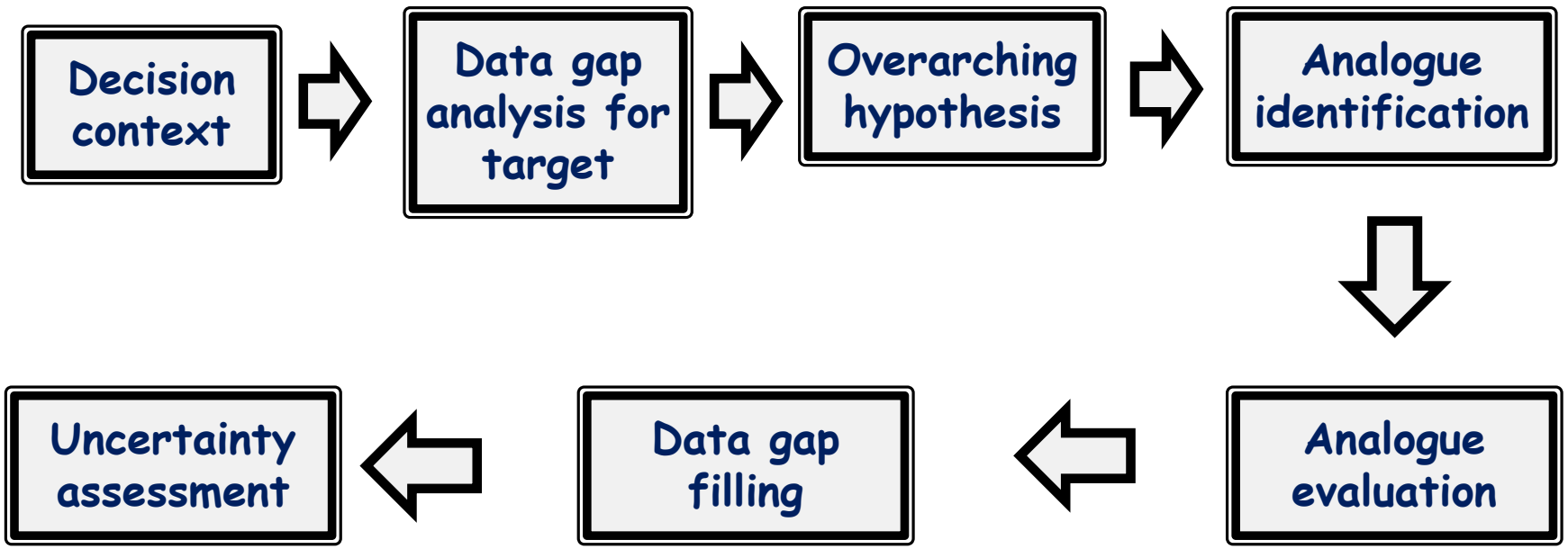


Decision Context

- 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

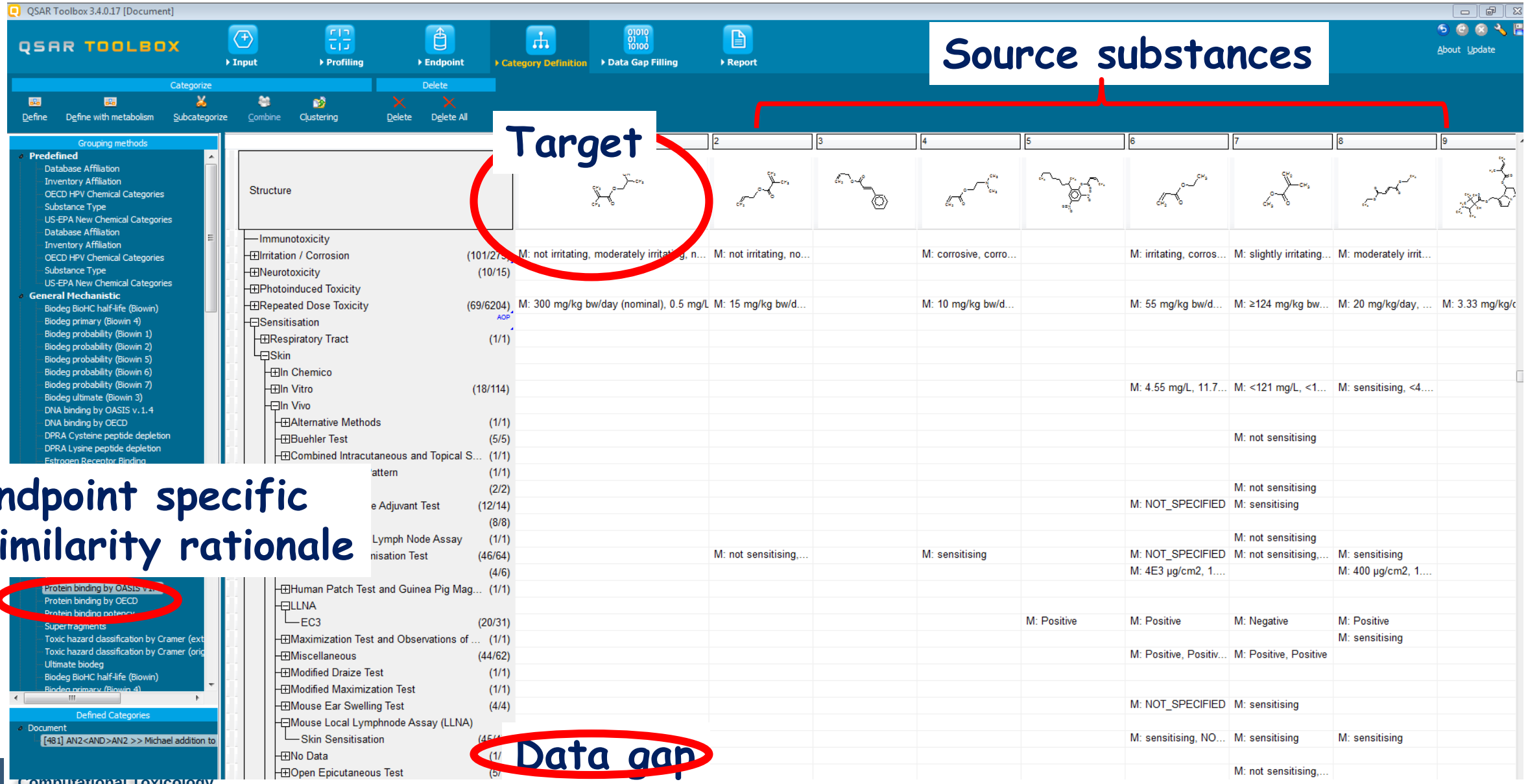


Selected Read-Across Tools

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

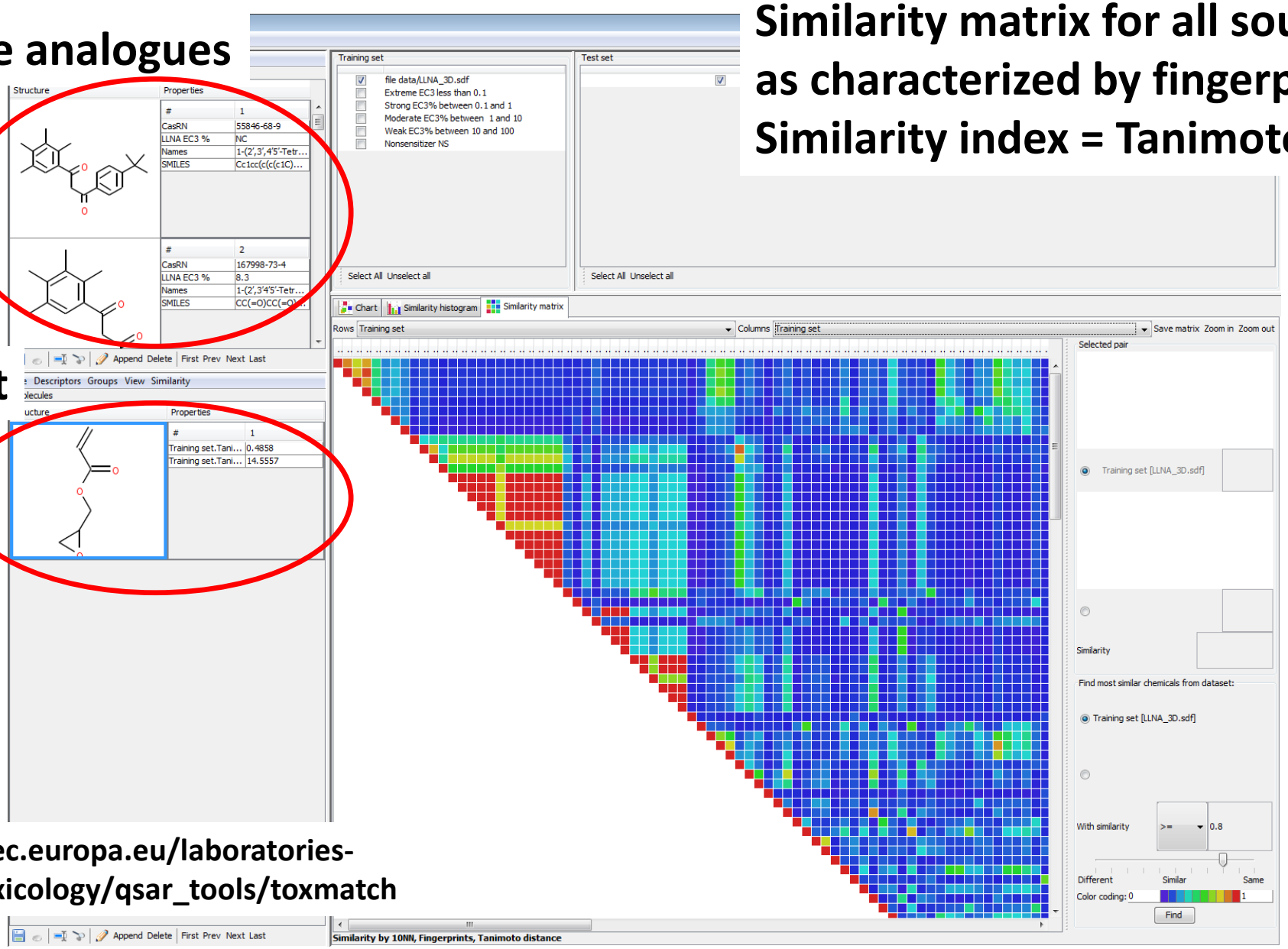
Analogue identification & evaluation within Toxmatch

Source analogues

Target

Similarity matrix for all source analogues as characterized by fingerprints

Similarity index = Tanimoto distance

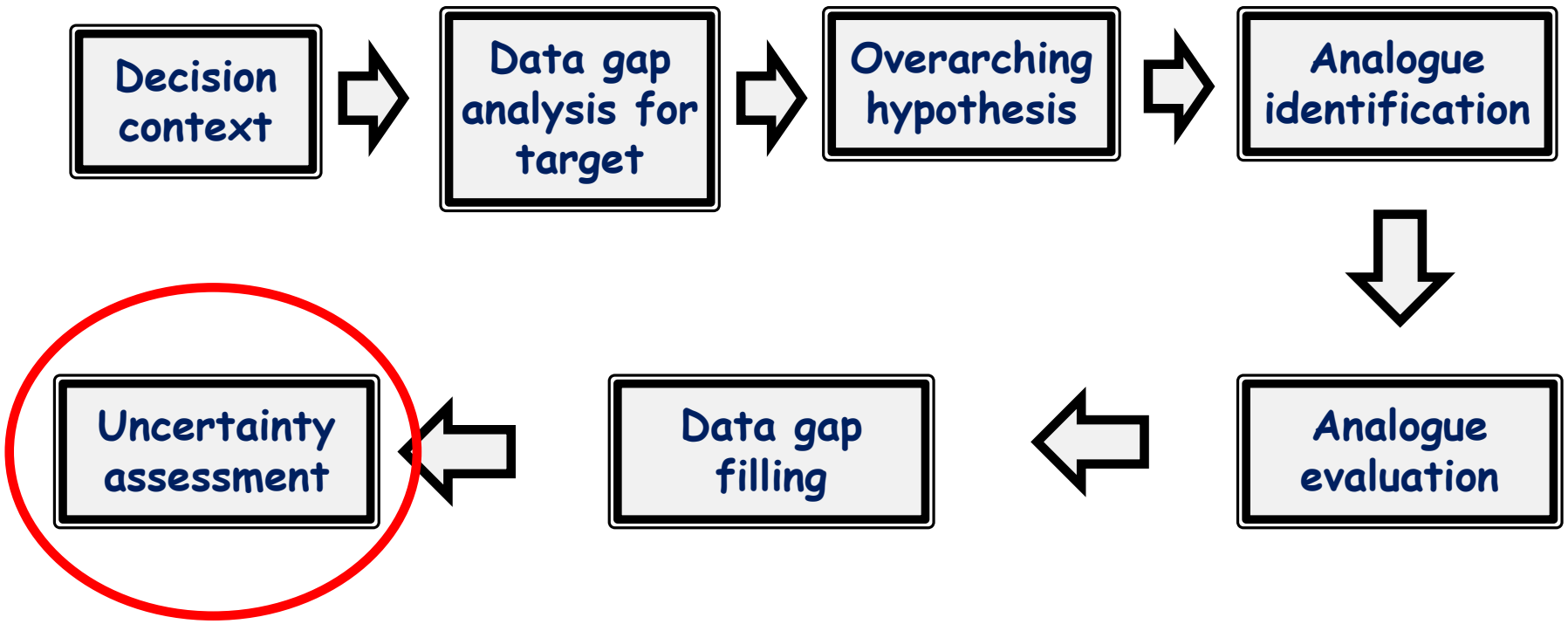


https://eurl-ecvam.jrc.ec.europa.eu/laboratories-research/predictive_toxicology/qsar_tools/toxmatch

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

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, Molecular Networks, EPA NCCT

Uncertainty assessment

- **Low** – degree of uncertainty is judged to be comparable to having direct data on a chemical

Regulatory Toxicology and Pharmacology 68 (2014) 353–362

Contents lists available at ScienceDirect

Regulatory Toxicology and Pharmacology

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

A framework to facilitate consistent characterization of read across uncertainty

Karen Blackburn *, Sharon B. Stuard

Central Product Safety Department, The Procter & Gamble Company, Mason Business Center, 8700 Mason Montgomery Road, Cincinnati, OH 45040, United States

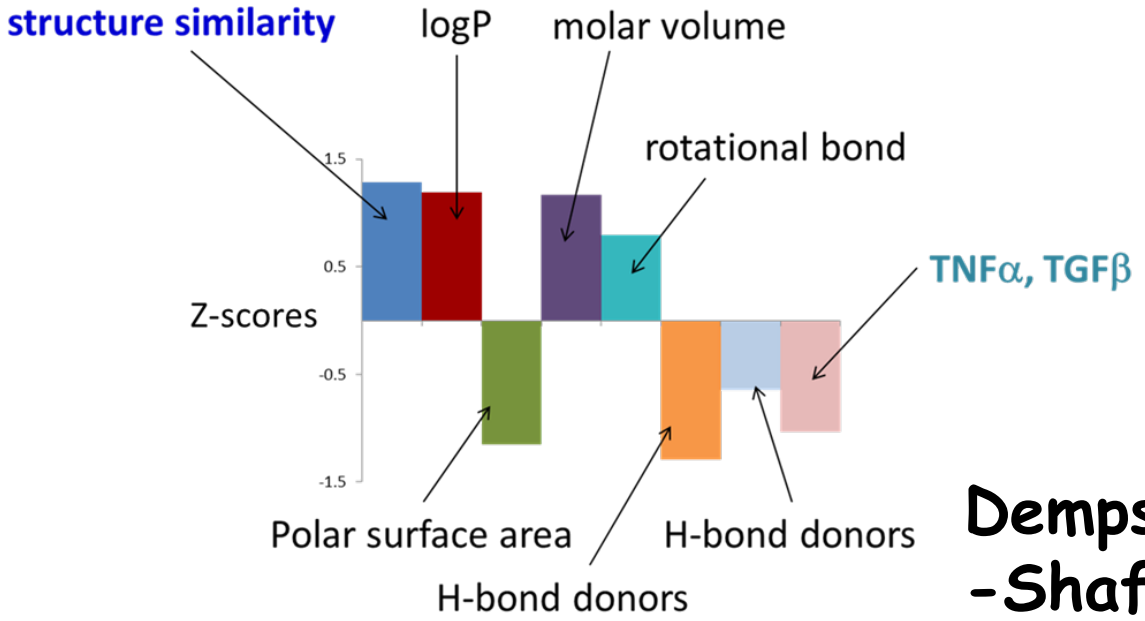
- **High** – degree of uncertainty is judged to be significantly greater than if direct data were used
  - Read across is not actionable without additional information

* Courtesy of Karen Blackburn and Sharon Stuard, P&G

Setting up the table of reliability metrics

INFORMATION TYPE: Analogs, In vivo toxicity data, Knowledgebase

Target/Analog	
Experimental data reliability	 reproductive developmental Klimisch score
Analog similarity:	 structure properties biological assays
Analog similarity:	 Skyline profile
Predictions alerts/QSAR): endpoints Reliability	 reproductive developmental PPV/NPV Odd ratio



Dempster -Shafer Theory

* Courtesy of Chihae Yang

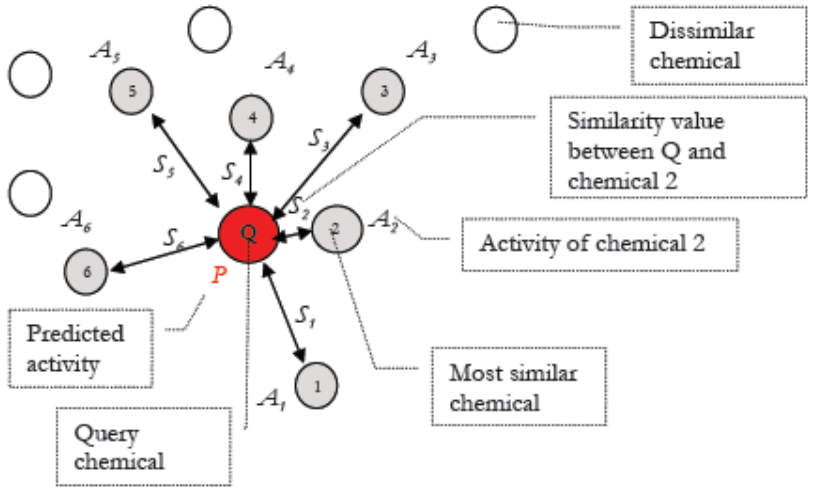
Quantifying Uncertainty & Assessing Performance of Read-Across

- 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
- Systematically evaluates read-across performance and uncertainty using available data

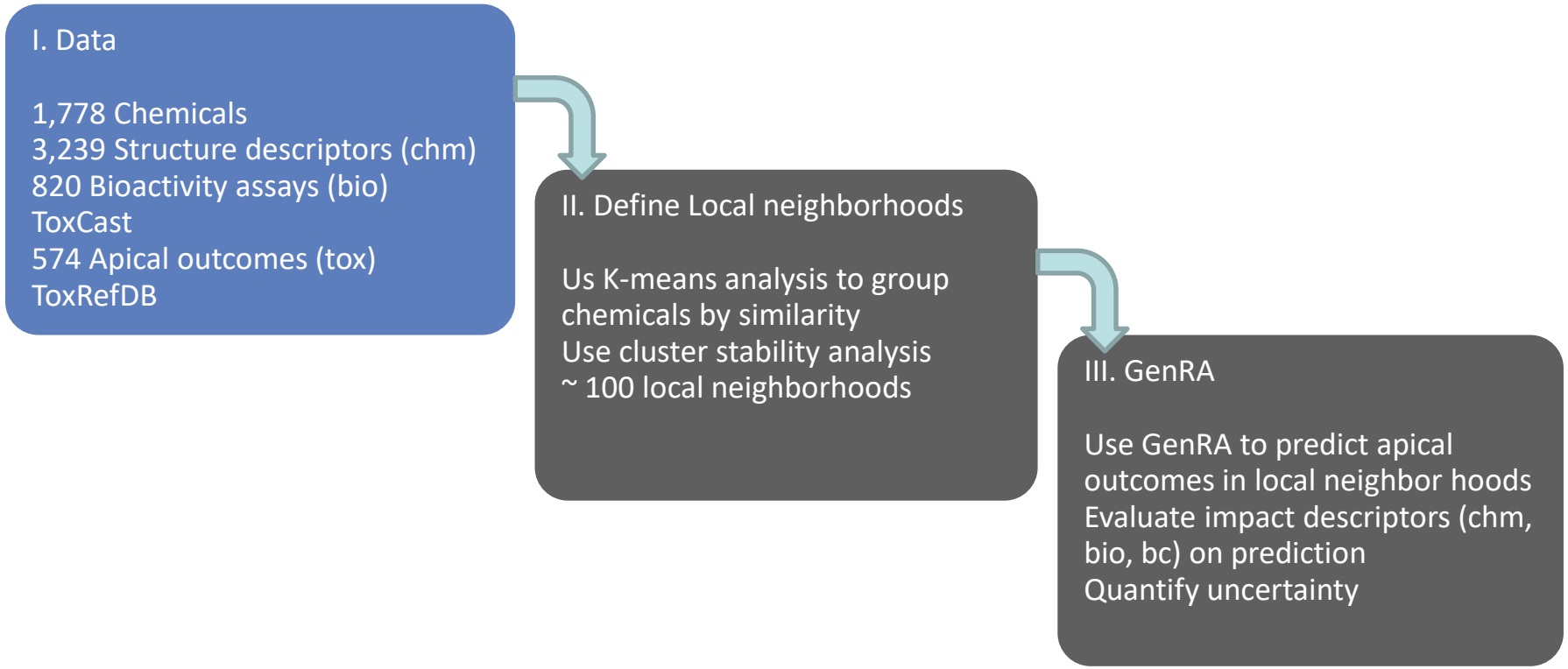
$$y_i^{\beta,\alpha} = \frac{\sum_j^k s_{ij}^{\alpha} x_j^{\beta}}{\sum_j^k s_{ij}^{\alpha}}$$

Jaccard similarity:

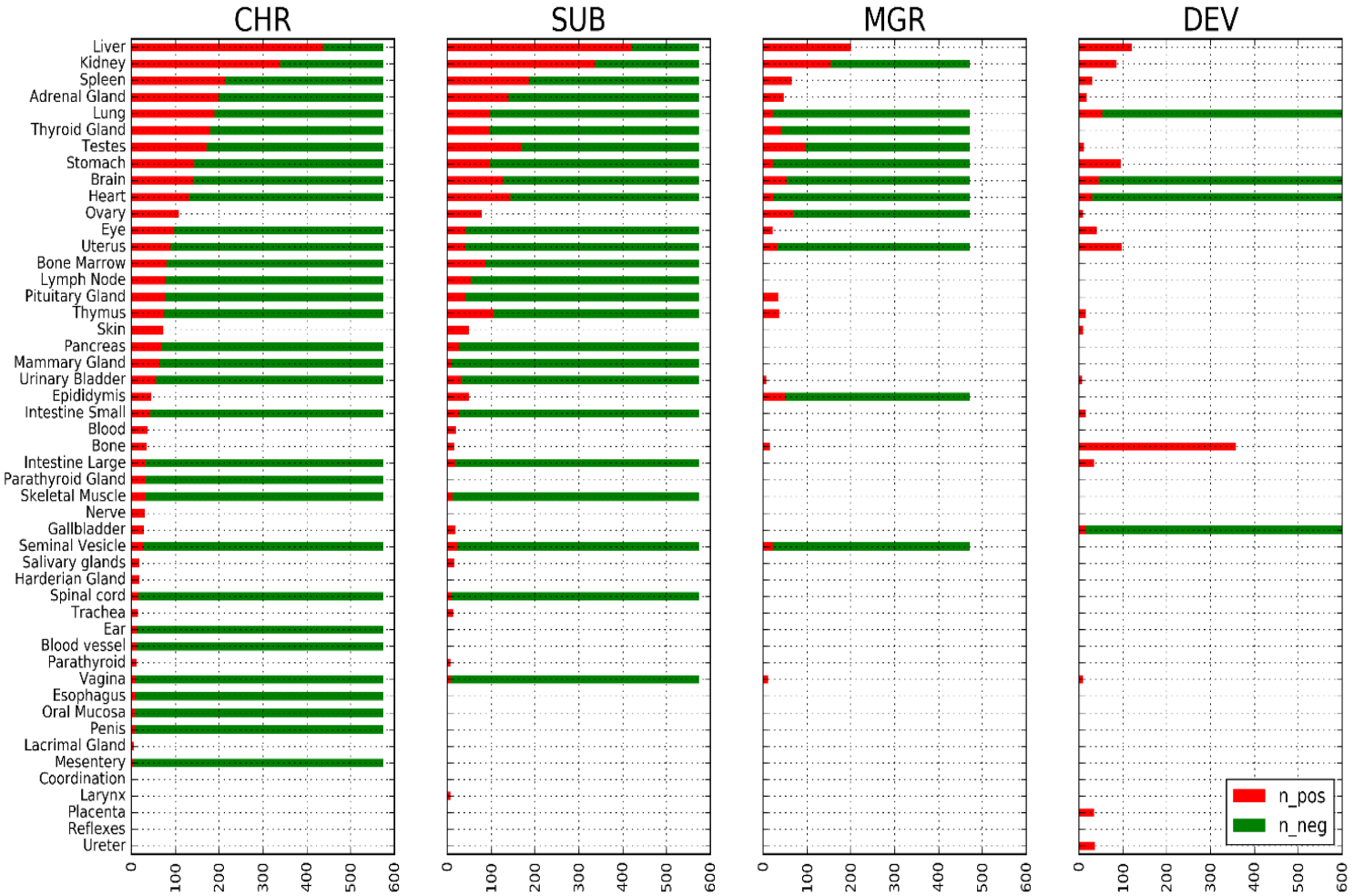
$$s_{ij} = \frac{\sum_l (x_{il} \wedge x_{jl})}{\sum_l (x_{il} \vee x_{jl})}$$



GenRA - Approach

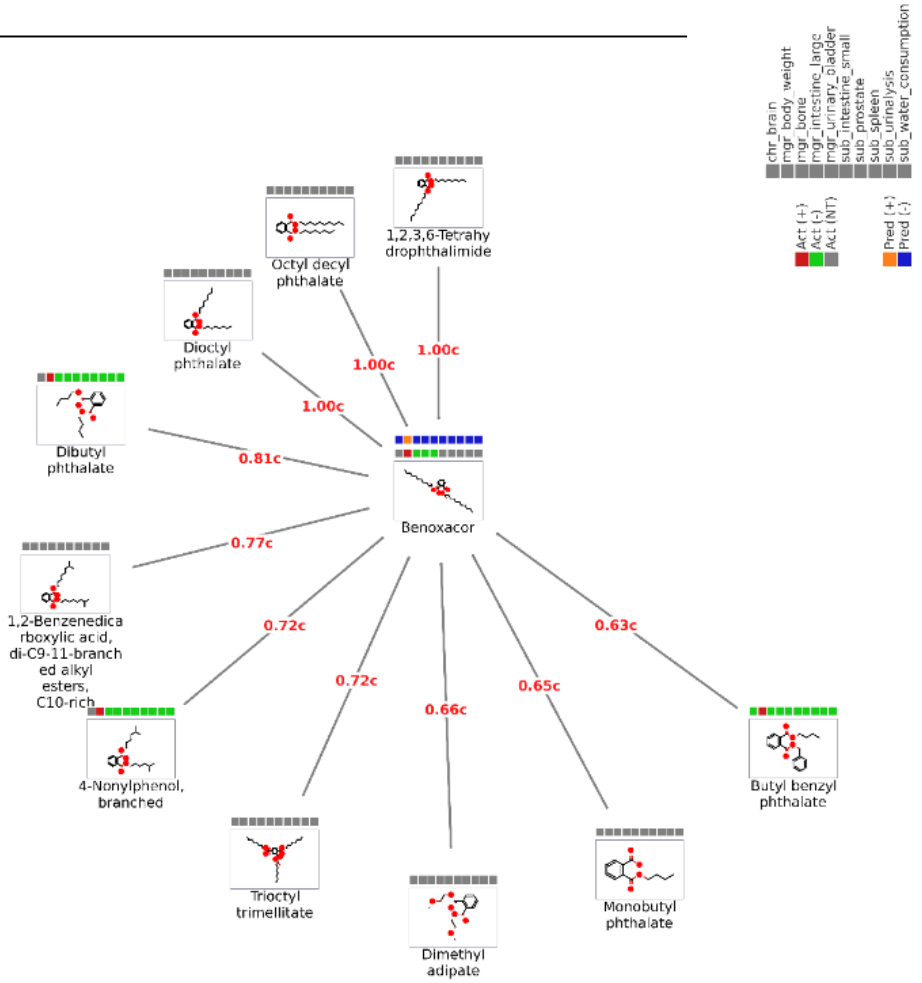


GenRA - Toxicity Data from ToxRefDB

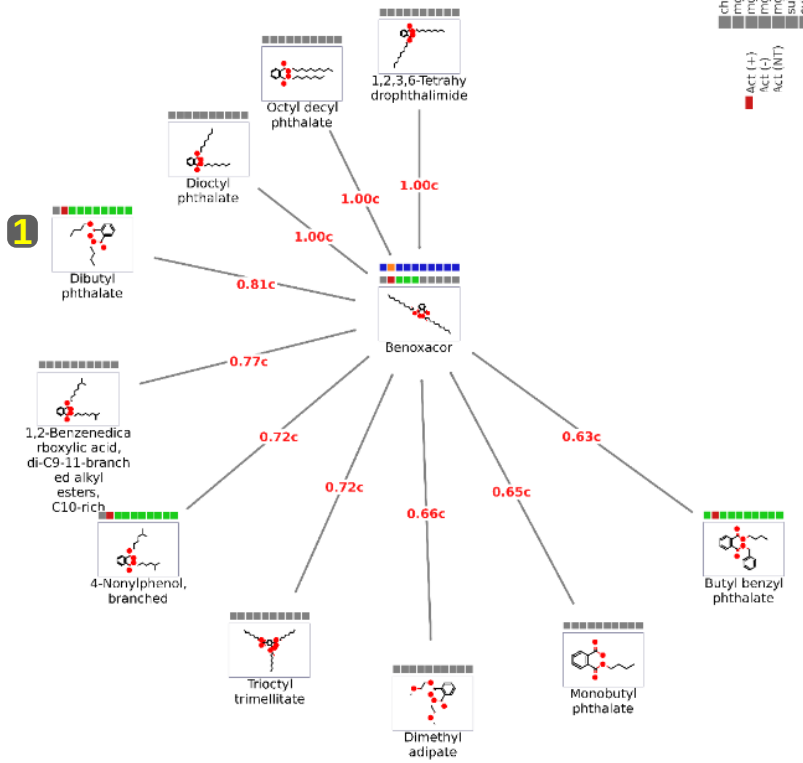


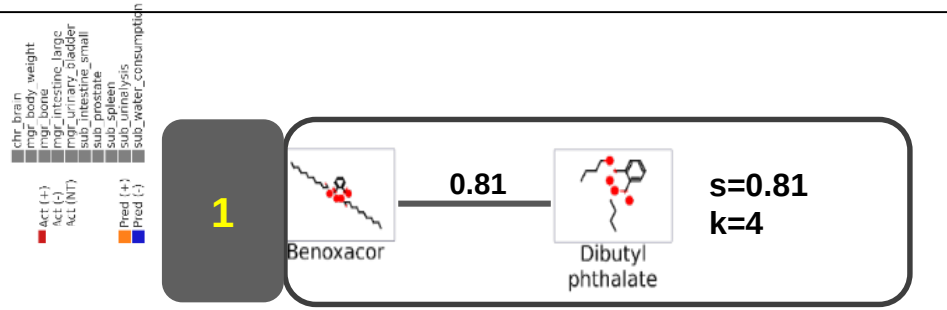
GenRA – Performance in Each Cluster

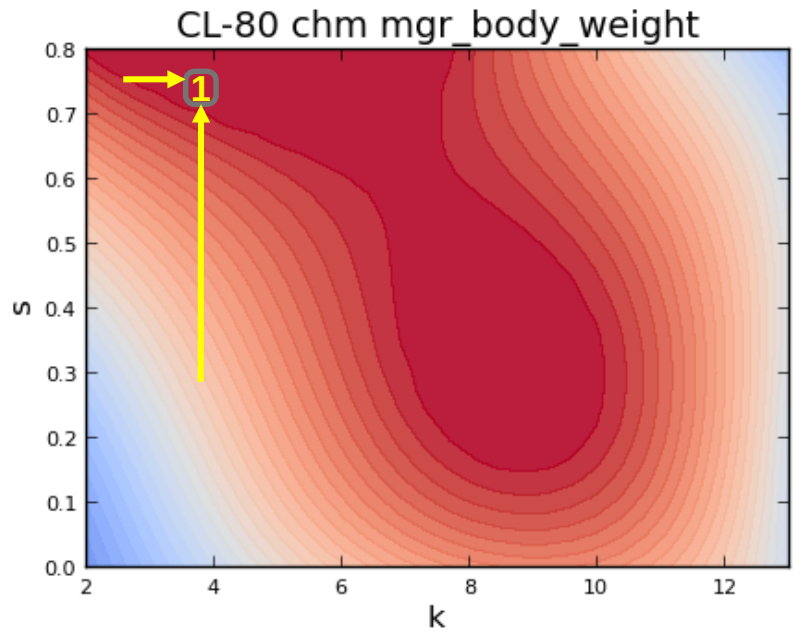
- Use GenRA to predict the similarity weighted toxicity scores for each
  - Toxicity type (β)
  - Descriptor ={chm,bio,bc} (α)
  - No. of nearest neighbors (k)
  - Similarity score threshold (s_{ij}^{α})
- Calculate performance by comparing predicted y^{tox} and true x^{tox} for all chemicals using area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



GenRA - Analysing Local Neighborhood of a Chemical







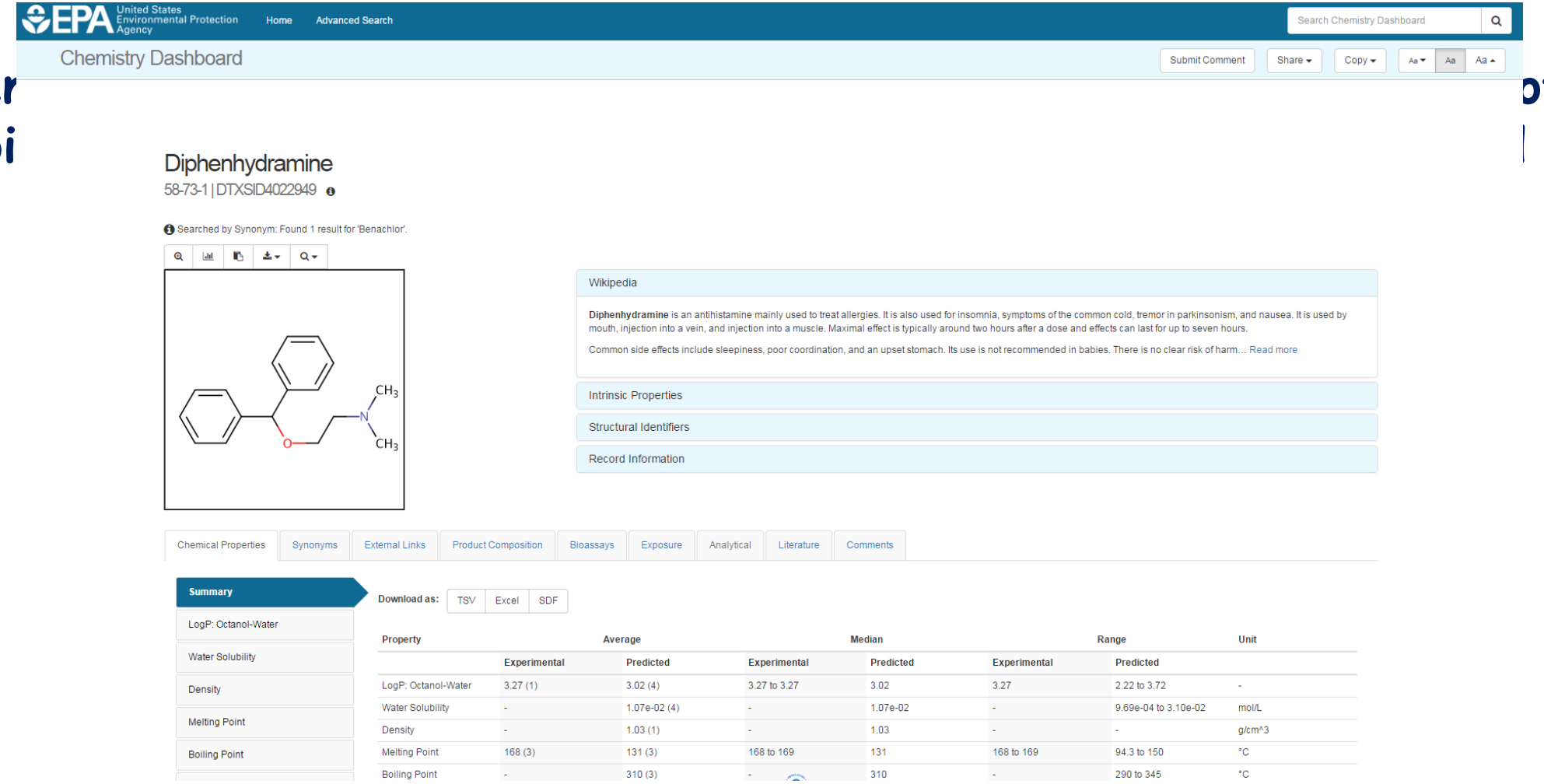
GenRA – Insights and Next Steps

- 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

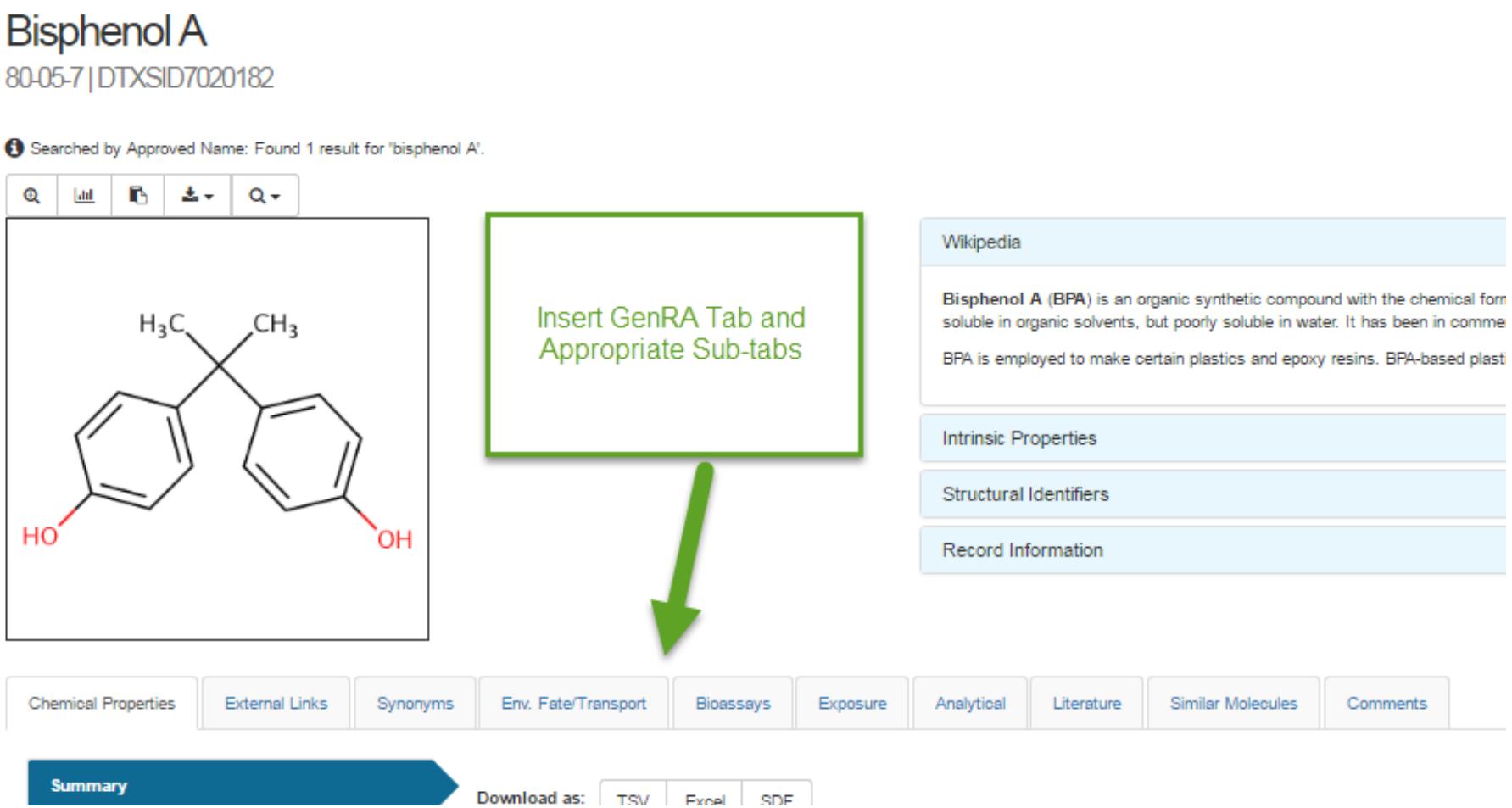
- Ongoing analysis:
- Consideration of other information to refine the analogue selection – e.g. TK similarity, metabolic similarity, reactivity similarity…

From research to implementation

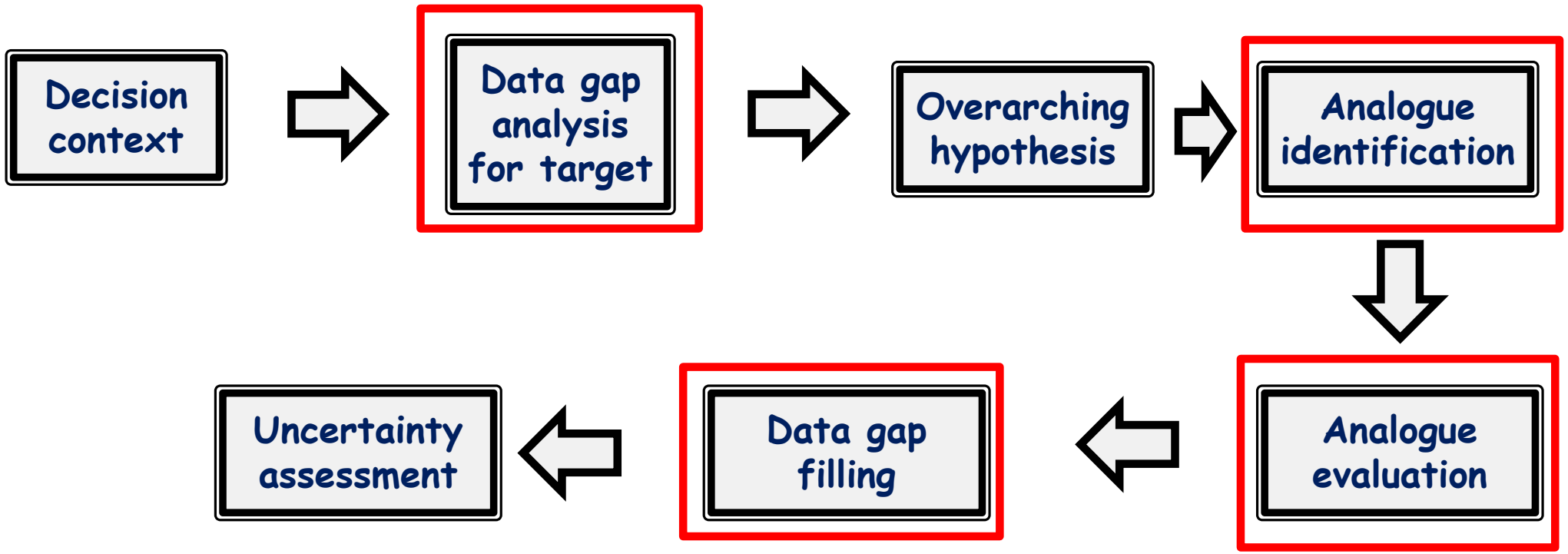
- Inter … of ongoi …



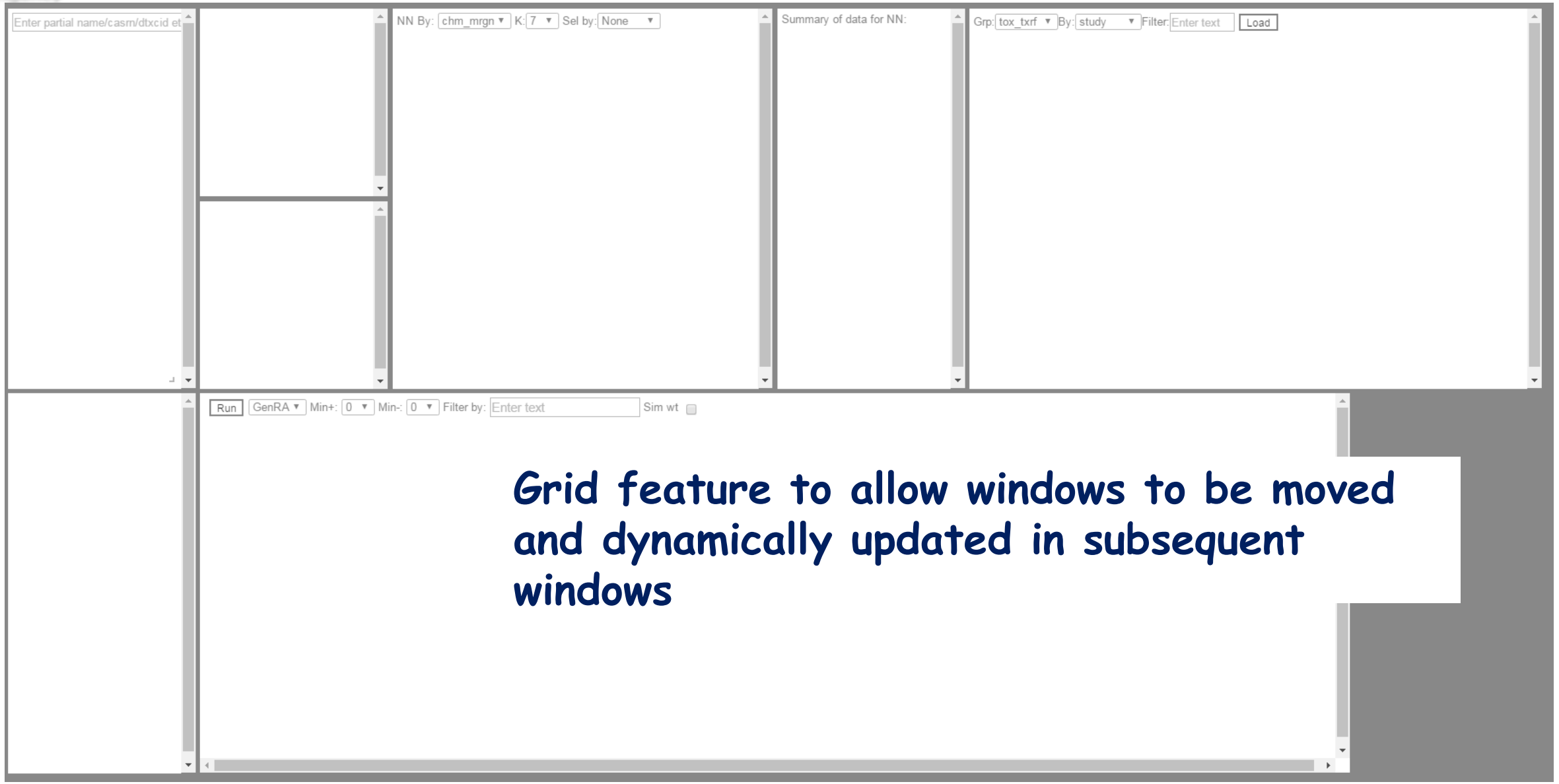
Basic Integration via GenRA tab



GenRA prototype development

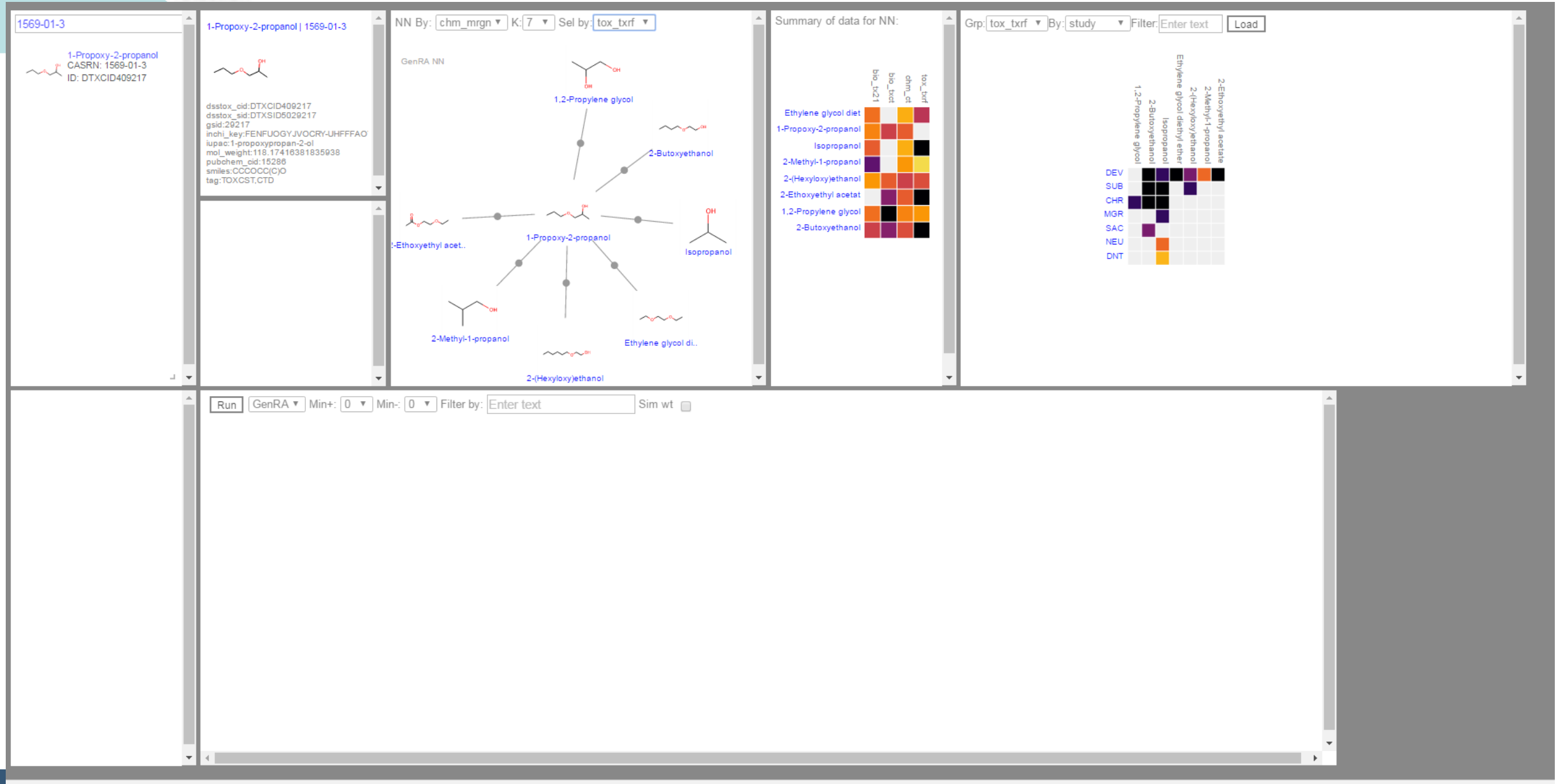


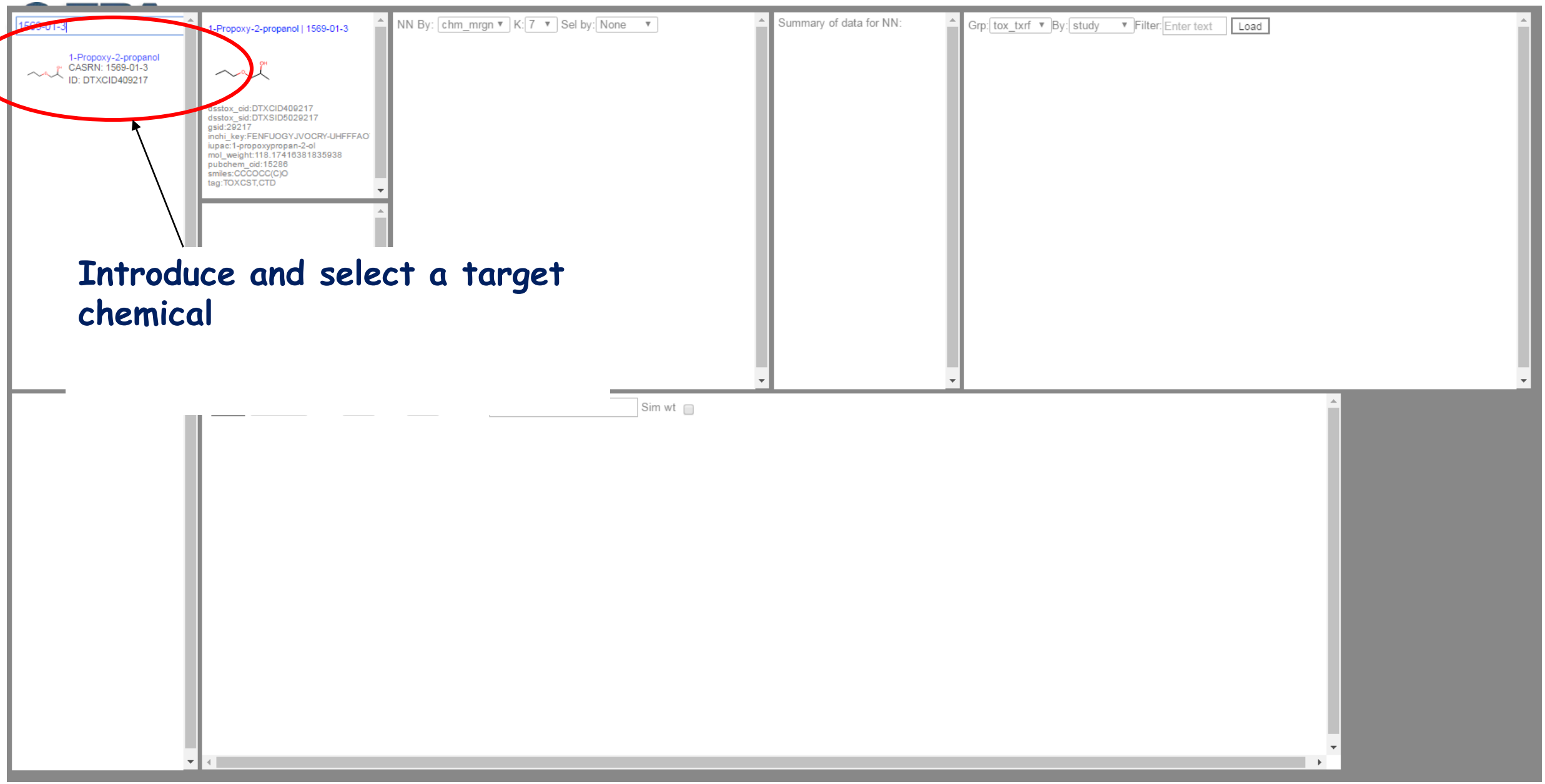
Initial interface



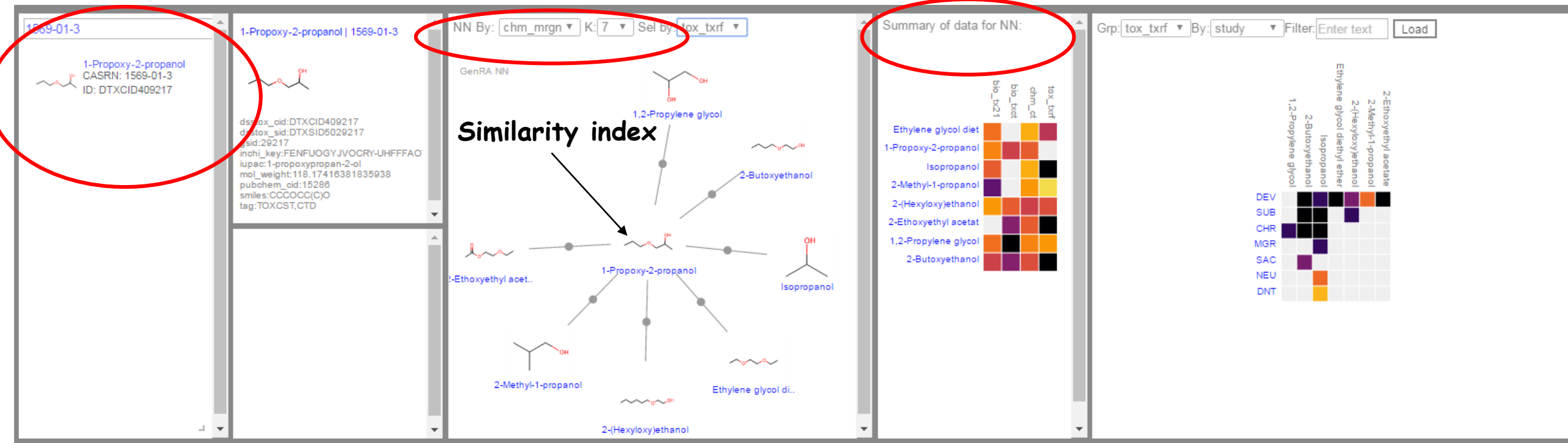
Grid feature to allow windows to be moved and dynamically updated in subsequent windows

Working interface

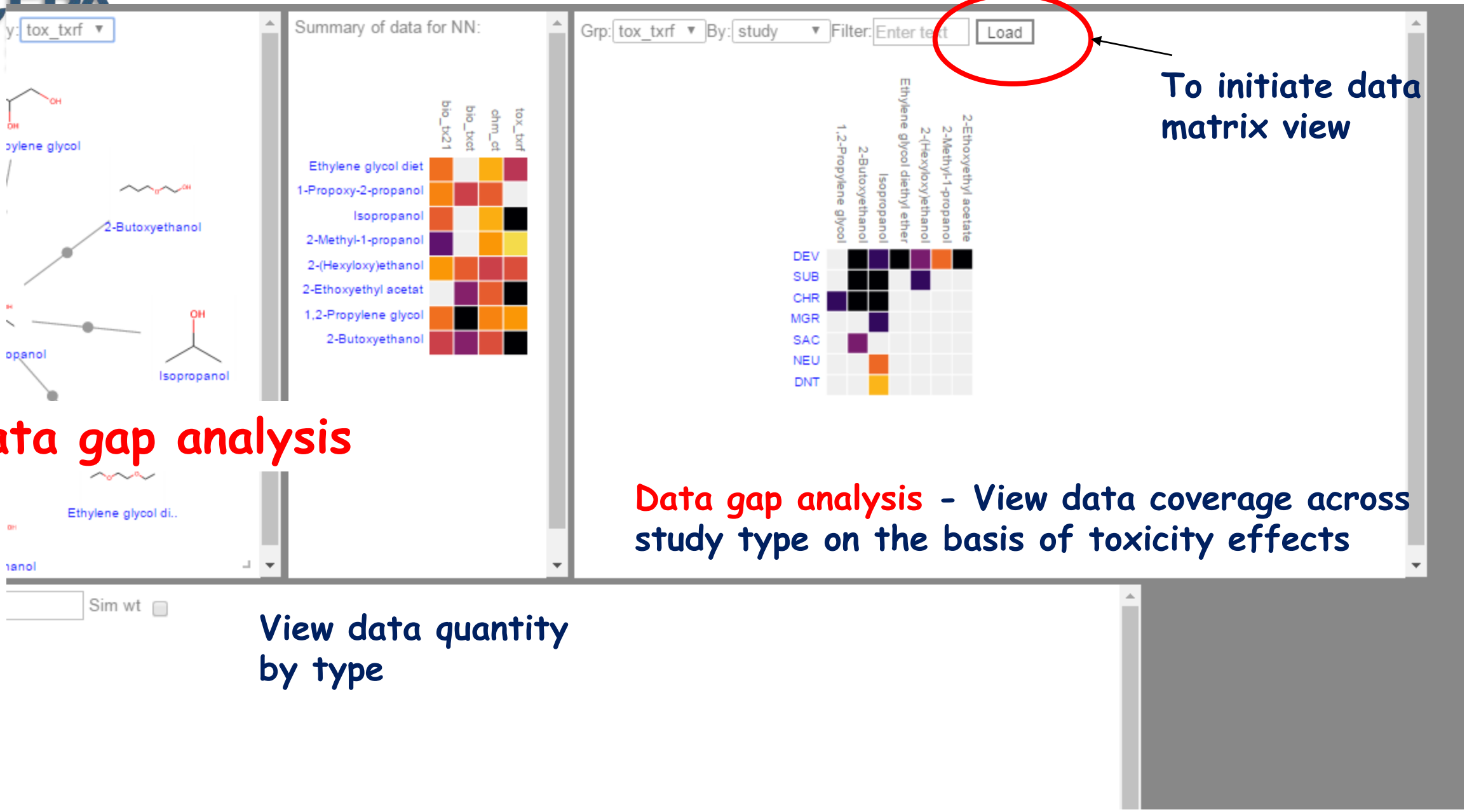




Introduce and select a target chemical



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



Data gap analysis

To initiate data matrix view

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

View data quantity by type

Analogue evaluation using data matrix view

Data gap filling using GenRA within data matrix

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
- To see more – stop by the EPA booth in the ToxExpo for live demonstrations of the CompTox dashboard, GenRA (Wed 1pm) and more..

Acknowledgements

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

- Karen Blackburn P&G
- Sharon Stuard P&G
- Chihae Yang Altamira LLC & Molecular Networks