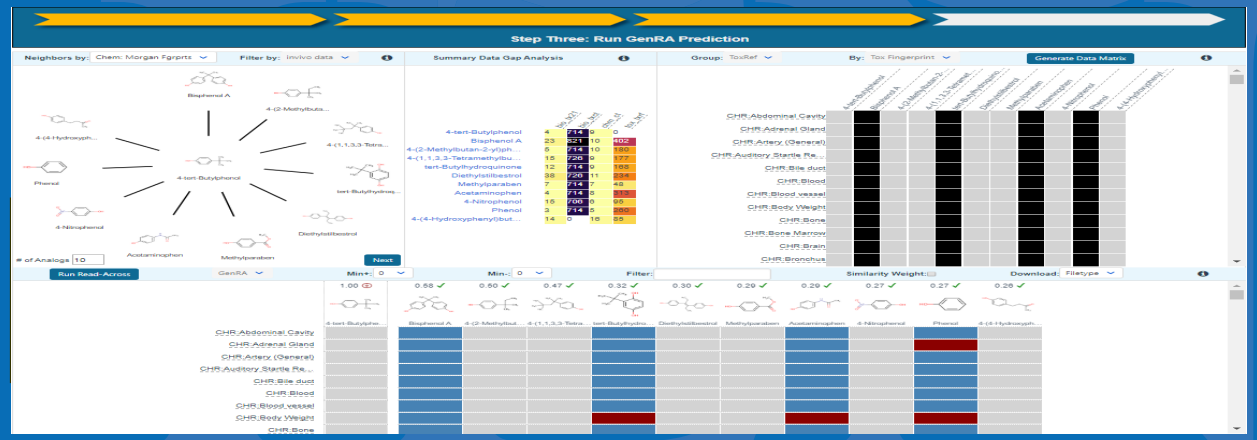
United States Environmental Protection Agency

Reaching for the summit of read-across: a brief journey through frameworks, tools and new approaches.



Grace Patlewicz
Center for Computational Toxicology & Exposure (CCTE), US EPA

EPA United States Environmental Protection Agency

Outline

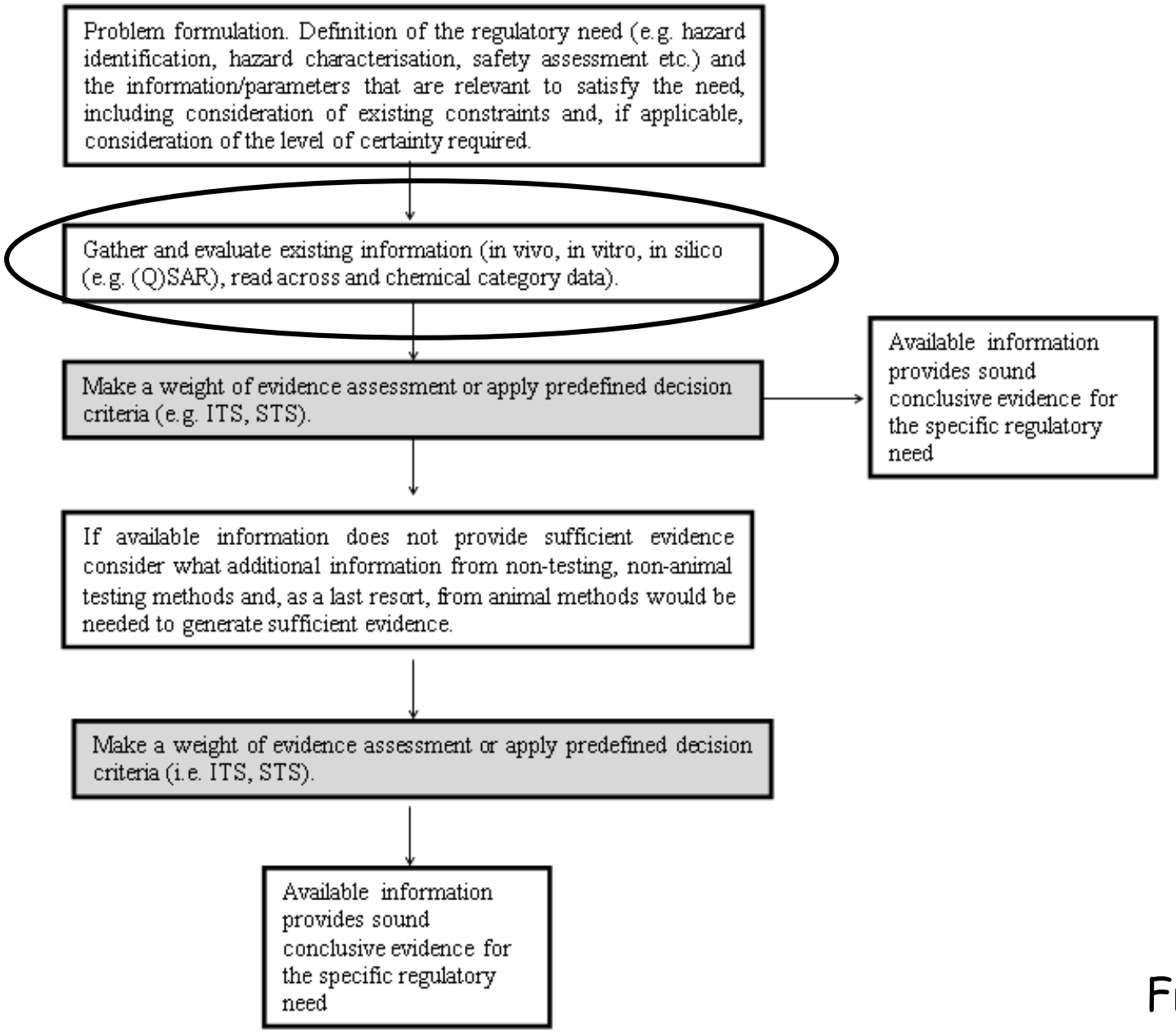
- Where does read-across fit in an integrated testing and assessment approach (IATA)
- What is read-across?
- Read-across tools and frameworks?
- How is read-across evolving? A GenRA perspective
- Summary Remarks
- Acknowledgements

Where do read-across approaches fit within IATA?

- Integrated Approaches to Testing and Assessment (IATA)

- "A tiered approach to data gathering, testing, and assessment that integrates different types of data (including physicochemical and other chemical properties as well as *in vitro* and *in vivo* toxicity data). When combined with estimates of exposure in an appropriate manner, the IATA provides predictions of risk."

General framework of an IATA



From OECD

Definitions: Chemical grouping approaches

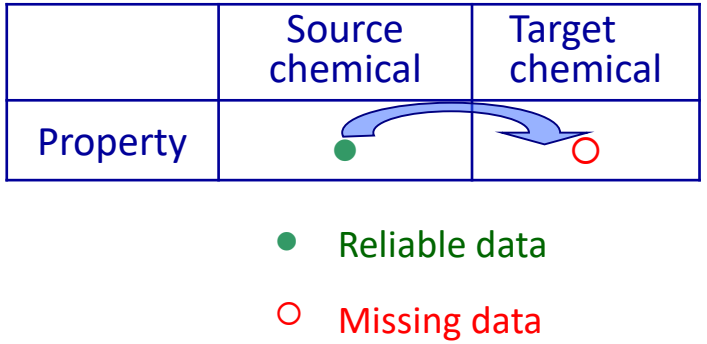
"Analogue approach" refers to grouping based on a very limited number of chemicals (e.g. target substance + source substance)

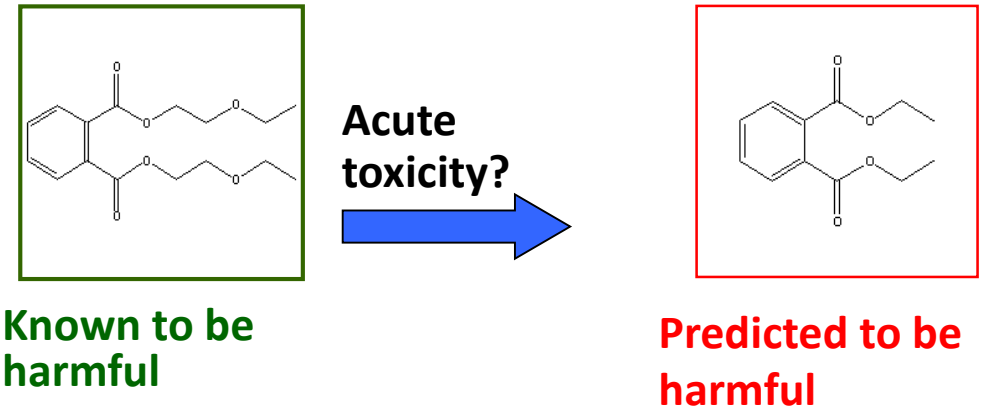
"Category approach" is used when grouping is based on a more extensive range of analogues (e.g. 3 or more members)

A chemical category is a group of chemicals whose physico-chemical and human health and/or environmental toxicological and/or environmental fate properties are likely to be similar or follow a regular pattern as a result of structural similarity (or other similarity characteristics).

What is 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.





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

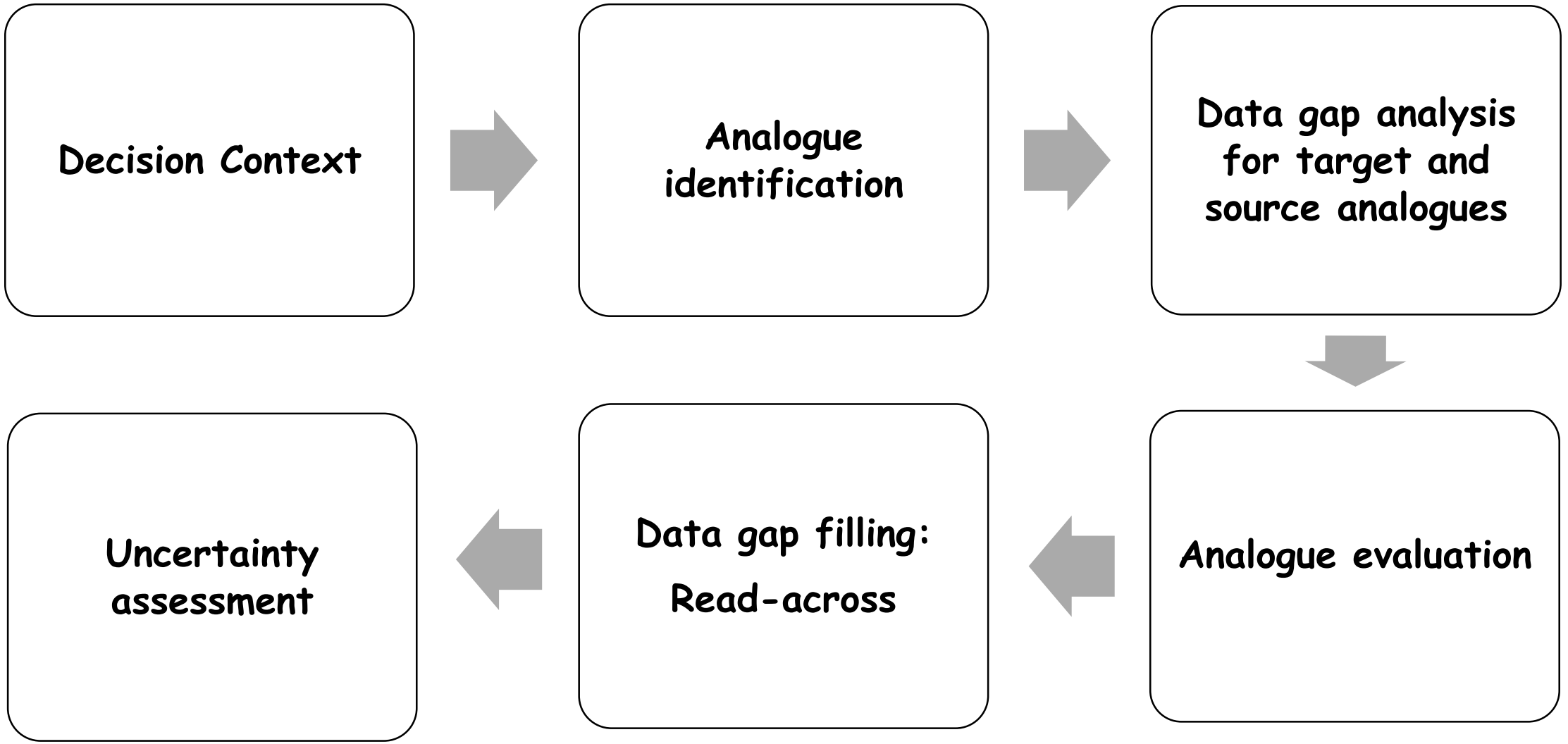
Keywords:
Category approach
Analogue approach
Data gap filling
Read-across
(Q)SAR
Trend analysis
Nearest neighb

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 workflow



A harmonised hybrid read-across workflow

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

A harmonised hybrid read-across workflow

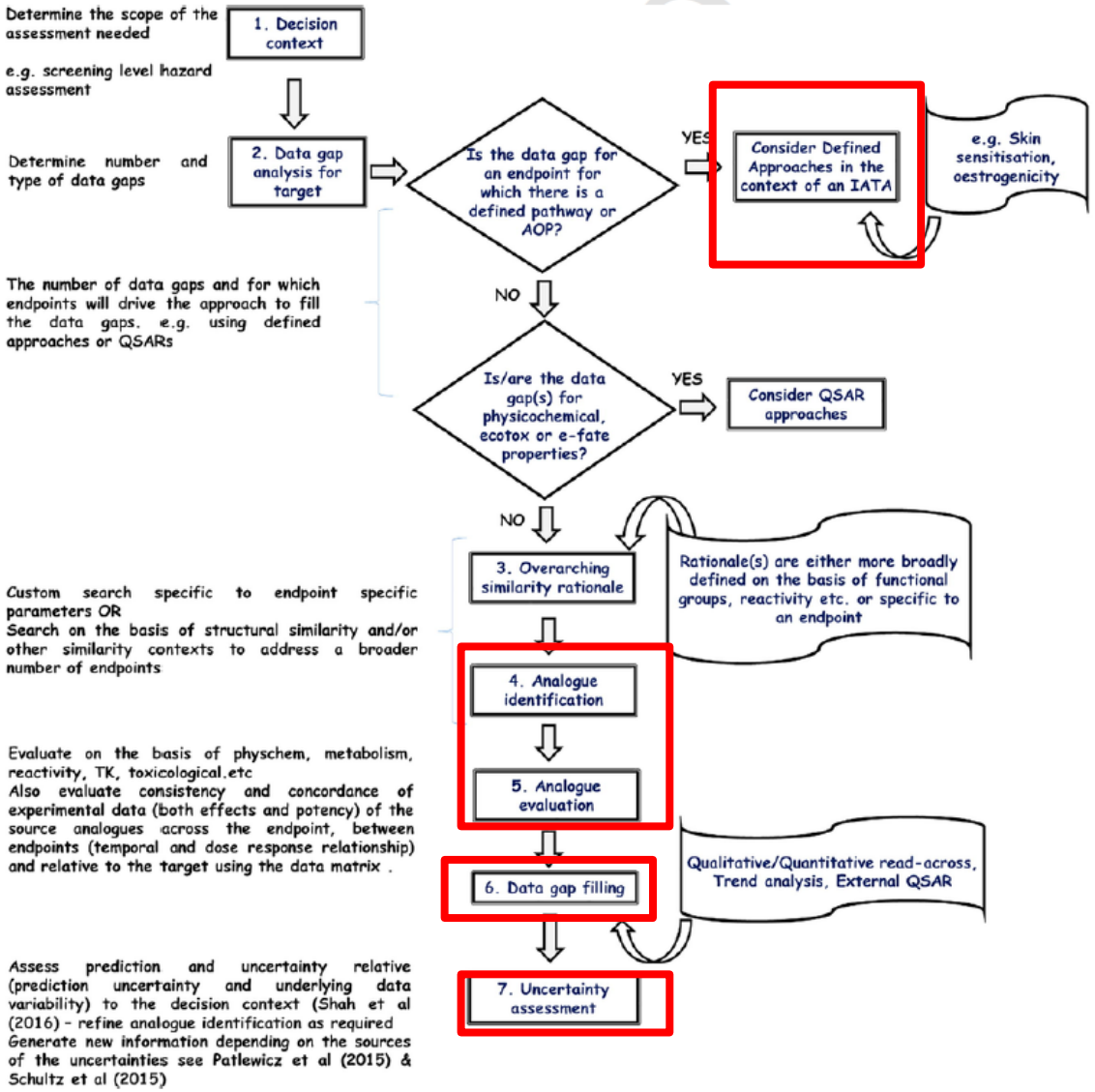


Fig. 9. A harmonised hybrid development and assessment framework.

Where do other new approach data streams fit? E.g. mechanistic data from ToxCast

How should we transition to data-driven approaches? moving away from subjective expert driven assessments.

What about characterising the uncertainty of the predictions made?

Patlewicz et al., 2018

GenRA (Generalised Read-Across)

- Predicting toxicity as a similarity-weighted activity of nearest neighbours based on chemistry and bioactivity descriptors (Shah et al, 2016)
- Generalised version of the Chemical-Biological Read-Across (CBRA) developed by Low et al (2013)
- Goal: To establish an objective performance baseline for read-across and quantify the uncertainty in the predictions made

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

Jaccard similarity:

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

$\alpha \in \{chm, bio, bc\}$

$\beta \in \{bio, tox\}$

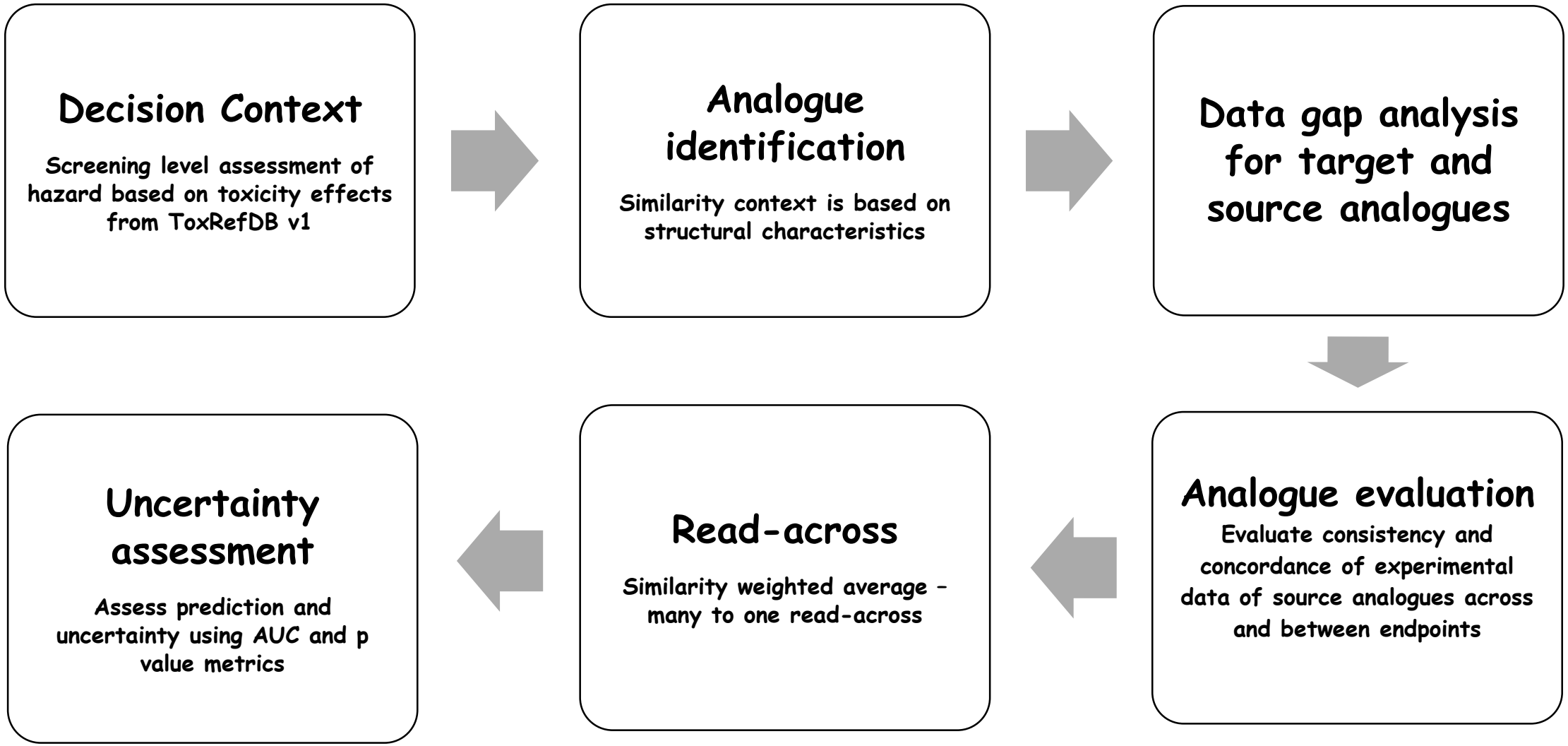
y_i = predicted activity of chemical (c_i)

x_j^{β} = activity of c_j in β

s_{ij}^{α} = Jacccard similarity between $\boldsymbol{x}_i^{\alpha}, \boldsymbol{x}_j^{\alpha}$

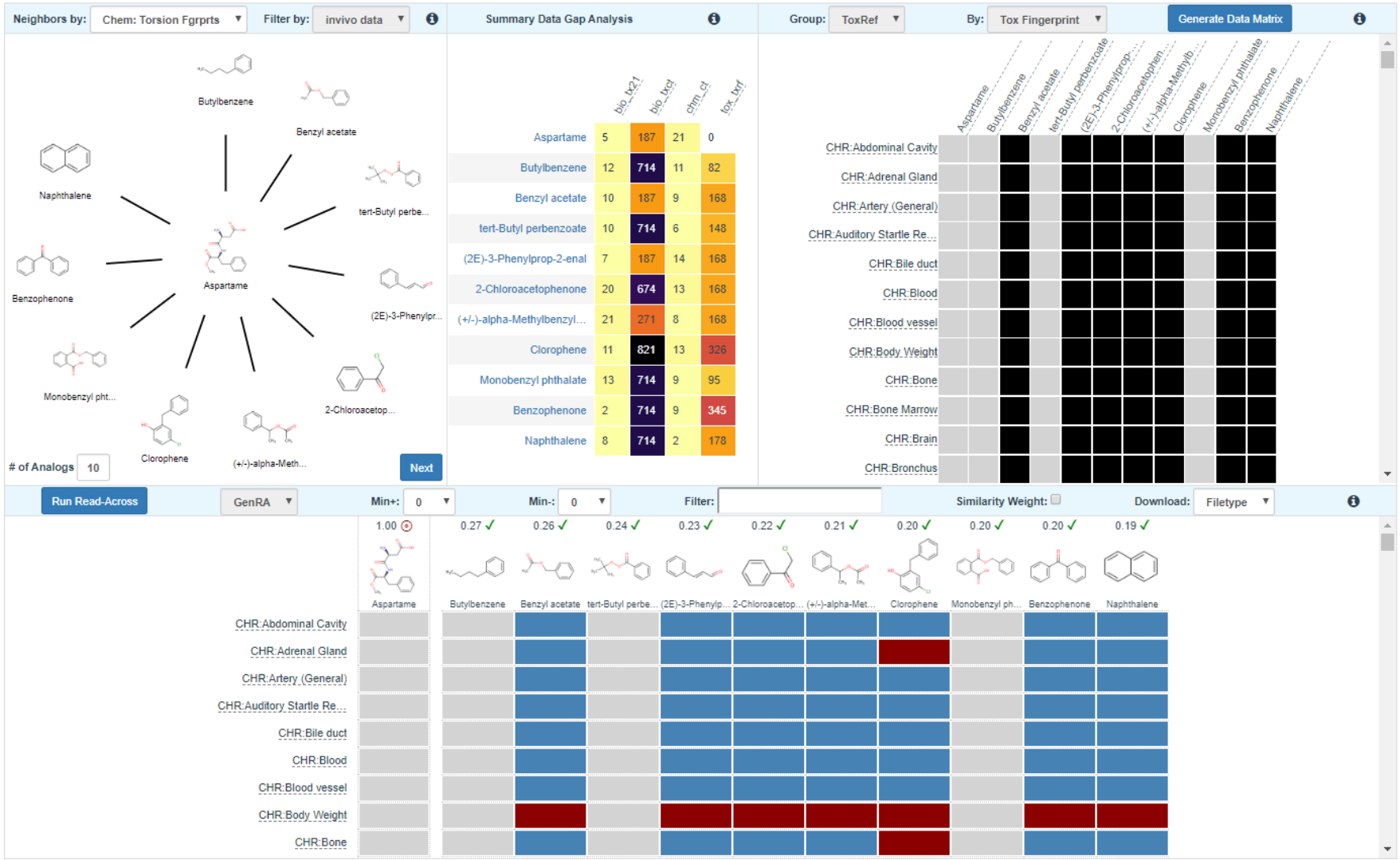
k = up to k nearest neighbours

Read-across workflow in GenRA v1.0



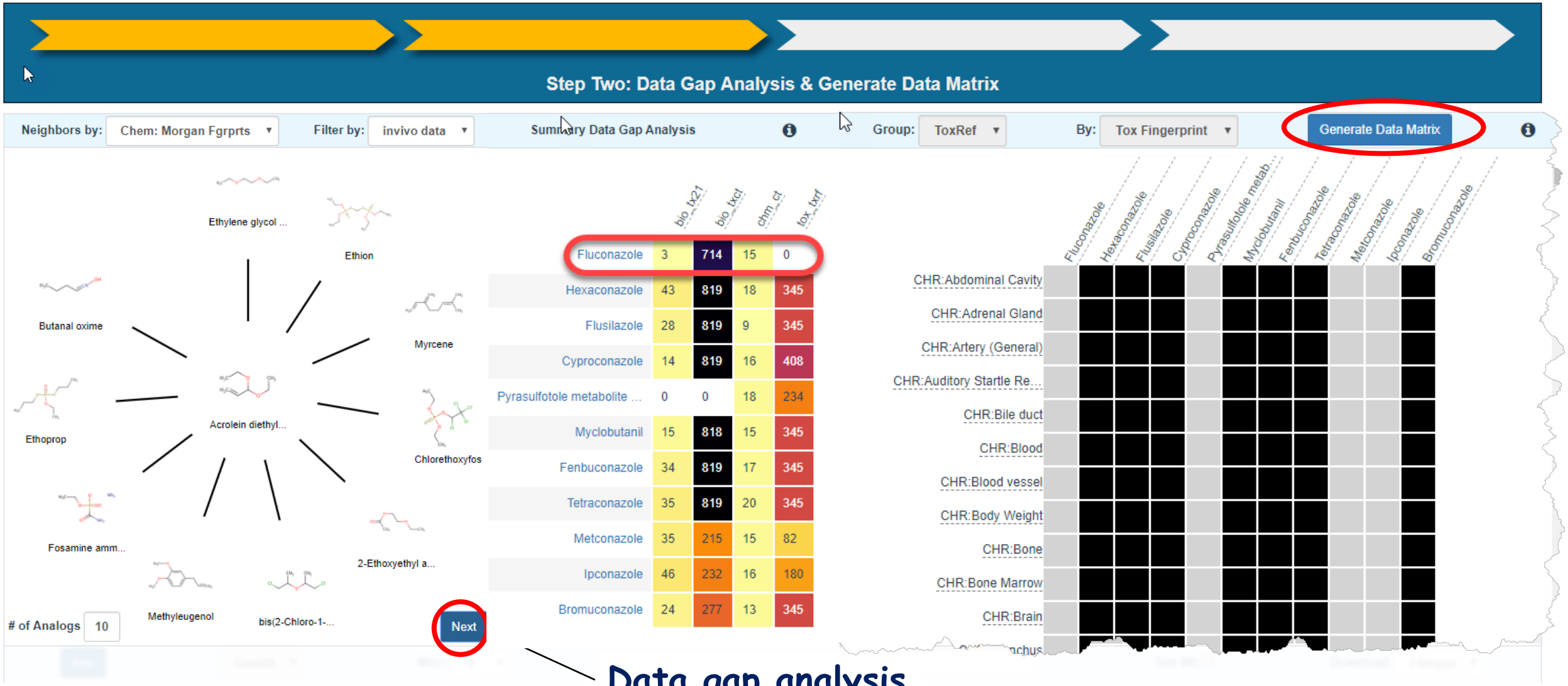
GenRA tool in reality

- Integrated into the EPA CompTox Chemicals dashboard



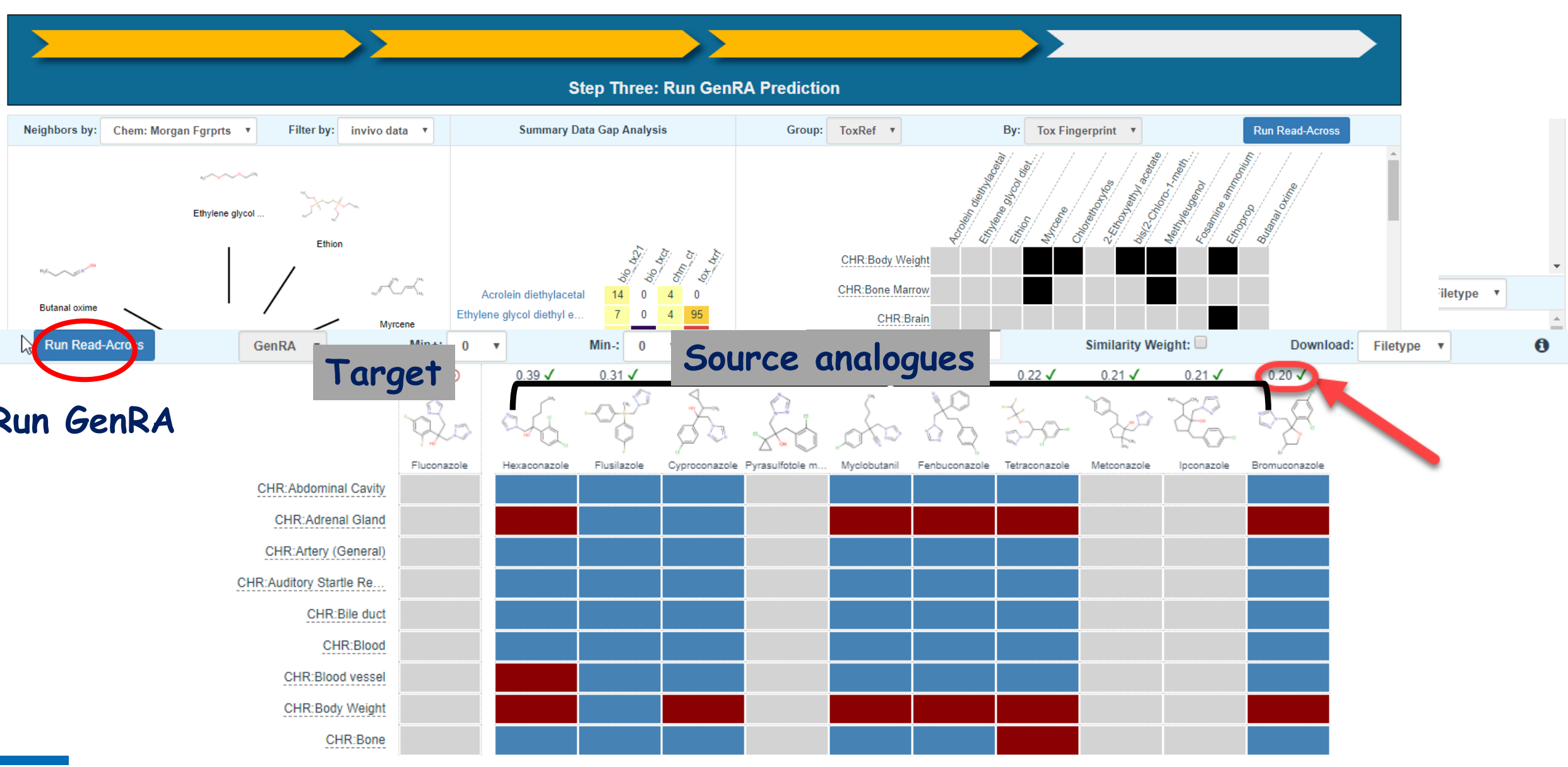
GenRA tool in practice

GenRA



Data gap analysis

GenRA tool in practice



Run GenRA

EPA United States Environmental Protection Agency

GenRA – Next Steps

- Ongoing research:
- Summarising and aggregating the toxicity effect predictions to guide end users – what effect predictions are we most confident about (digesting & interpreting the predictions more efficiently)
- Consideration of other information to define and refine the analogue selection & evaluation – e.g. physicochemical similarity, metabolic similarity, reactivity similarity, mechanistic similarity (transcriptomics similarity, phenotypic profiling similarity)
- Transitioning to quantitative predictions of toxicity e.g. LOAEL, LD50
- Read-across to predict other in vitro endpoints to supplement in vitro-in vivo extrapolations

GenRA Current Workflow (within Dashboard)

Workflow Priority	Decision Context (user need)	Analogue Identification	Data Gap Analysis	Analogue Evaluation	Read-across Prediction	Uncertainty Assessment	Comments
Workflow #1	Screening level assessment of hazard based on ToxRefDB	Similarity based on morgan fingerprint and Jaccard index	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity	ROC AUC and p-value based on cross-validation in local neighbourhood	Assessment of global and local landscape of structure clusters

GenRA Proposed Workflows

Workflow Priority	Decision Context (user need)	Landscape evaluation (global)	Analogue Identification (local)	Data Gap Analysis	Analogue Evaluation	Read-across Prediction	Uncertainty Assessment	Comments
Workflow #2	Screening level assessment of POD based on ToxRefDB v2	Distribution of analogues in "global" inventory	Similarity based on morgan and ct fingerprint and Jaccard index	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity	RMSE, R^2 and p-value based on cross-validation in local neighbourhood	Done
Workflow #3	Screening level assessment of LD50 based on AcuteTox data	Distribution of analogues in "global" inventory	Similarity based on morgan and ct fingerprint and Jaccard index	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity	RMSE, R^2 and p-value based on cross-validation in local neighbourhood	Done
Workflow #4	Screening level assessment of hazard based on ToxRefDB v2	Distribution of analogues in "global" inventory	Similarity based on morgan, ct, gene(ltea) fingerprint and Jaccard index	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity	ROC AUC and p-value based on cross-validation in global and local neighbourhood	In progress

GenRA Proposed Workflows (continued)

Workflow Priority	Decision Context (user need)	Landscape evaluation (global)	Analogue Identification (local)	Data Gap Analysis	Analogue Evaluation	Read-across Prediction	Uncertainty Assessment	Comments
Workflow #5	Screening level assessment of POD based on ToxRefDB v2	Distribution of analogues in "global" inventory	Similarity based on metabolites / metabolic pathways	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity based on metabolism	RMSE, R^2 and p-value based on cross-validation in local neighbourhood	In progress
Workflow #6	Screening level assessment of hazard based on ToxRefDB v2	Distribution of analogues in "global" inventory	Similarity based on morgan, fingerprint and p-chem properties	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity	ROC AUC and p-value based on cross-validation in global and local neighbourhood	Done

GenRA Tentative PFAS Workflow

Workflow Priority	Decision Context (user need)	Landscape evaluation (global)	Analogue Identification (local)	Data Gap Analysis	Analogue Evaluation	Read-across Prediction	Uncertainty Assessment	Comments
PFAS Workflow #1	Screening level assessment of POD based on ToxVal	Distribution of analogues in PFAS universe and/or *PFAS Categories*	Similarity based on chemical and/or bioactivity …. (research ongoing)	Evaluate availability of data for analogues (i.e. data matrix summary)	Evaluate consistency of source analogues (e.g. to deselect analogues)	Similarity weighted activity	RMSE, R^2 and p-value based on cross-validation in local neighbourhood	Revise PFAS categories based on rules derived from analogue identification and evaluation steps

GenRA – Overall goal

- Quantify the contribution that different similarity contexts play in toxicity prediction and how that differs depending on the toxicity endpoint of interest and the chemical of interest
- Quantify level of confidence for prediction made

=> objective, reproducible read-across assessments

EPA United States Environmental Protection Agency

Summary

- Defined where read-across fits within an IATA
- What read-across is in practice
- Read-across tools (selection) and where they fit within a read-across framework
- Evolving the framework for read-across to provide opportunities for NAM data
- How this is being developed within the context of Generalised Read-across (GenRA) can been applied in practice
- Highlight ongoing research in extending the GenRA approach by investigating how other contexts of similarity can be implemented and with what data streams

EPA United States Environmental Protection Agency

Acknowledgements

- Imran Shah
- George Helman (former student)
- Tia Tate
- Tony Williams
- Lucina Lizarraga
- Jason Lambert