



# Reference Toxicological Database (RefToxDB) for Pesticides and Other Environmental Chemicals

Matthew T. Martin<sup>1</sup>; Larry Proctor<sup>3</sup>; Elizabeth Mendez<sup>2</sup>; Elissa Reaves<sup>2</sup>; Vicki Dellarco<sup>2</sup>; Keith Houck<sup>1</sup>; Ann Richard<sup>1</sup>; and David J. Dix<sup>1</sup>

<sup>1</sup>U.S EPA, ORD, Computational Toxicology Research Program; <sup>2</sup>U.S EPA, OPP; <sup>3</sup>IBM Almaden Research Laboratory, San Jose, CA

research & development

Project ID: IIC-3

## Science Question

National Center for Computational Toxicology (NCCT) research using high-throughput screening (HTS) and toxicogenomic technologies for environmental chemical classification, prioritization, and toxicological profiling will require a reference toxicological database to characterize associations between toxicological properties and bioactivity profiles. EPA's Office of Pesticide Programs (OPP) evaluates submitted toxicological data in a standardized review process captured in Data Evaluation Records (DER) making it an ideal source for high quality toxicological data and transformation into a relational database system. The reference toxicological database will be used to analyze chemical relationships with biological outcomes, and link genomic and HTS data to the reference toxicological data. The proof of concept demonstration of ToxCast will establish associations between HTS bioactivity profiles and toxicological properties that might be useful in predicting toxicity for environmental chemicals where toxicological data is not available (e.g., HPVs, EDCs, anti-microbial pesticides). In addition, the reference toxicological database will be a component of the computational tools needed for the application of toxicogenomics to risk assessment.

## Research Goals

Acquisition of large pesticide toxicological dataset to be used as reference data for other environmental chemicals and various HTS and toxicogenomic technologies

Development of extraction process, including database development, automated and manual extraction and QA/QC

Develop standardized vocabulary for specified data fields to control data input including study type, animal information, units, results, and endpoints

Capturing results data uniformly utilizing tiered standardized result descriptors and normalized numeric fields

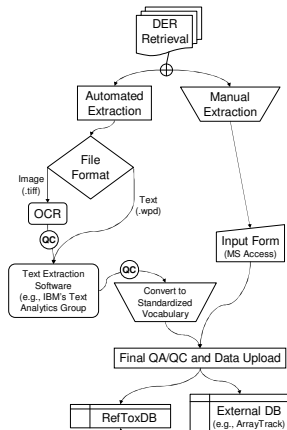
Permit querying of the reference data through chemical structure, toxicological endpoints, critical effects, relevant dose-response data

Development of computational tools to query the reference dataset to analyze for relationships across chemical and biological activity and linking such data with HTS and toxicogenomic data

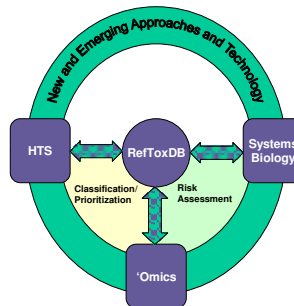
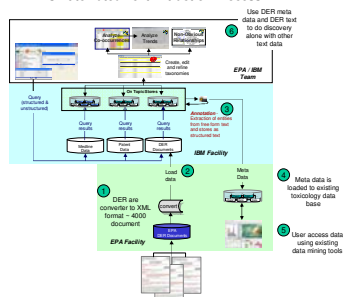
Targeted Toxicological Data Collection	
Source:	EPA's Office of Pesticide Programs (OPP)
Format:	Data Evaluation Record (DER)
Chemical:	Conventional Pesticide Active Ingredients (~800)
Purity:	Technical Grade (>90%)
Dosing:	Primarily Orally Administered (based on availability and use pattern of pesticide)
Study Type:	Subchronic Toxicity (Rodents and Non-Rodents) Prenatal Developmental Toxicity Reproduction and Fertility Effects (2-generation) Chronic Toxicity (rat, mouse, and dog) Carcinogenicity (rat and mouse) Developmental Neurotoxicity Immunotoxicity

\*\*\*Data Collection Results\*\*\*  
>4000 DER (2500 studies) for over 400 pesticides

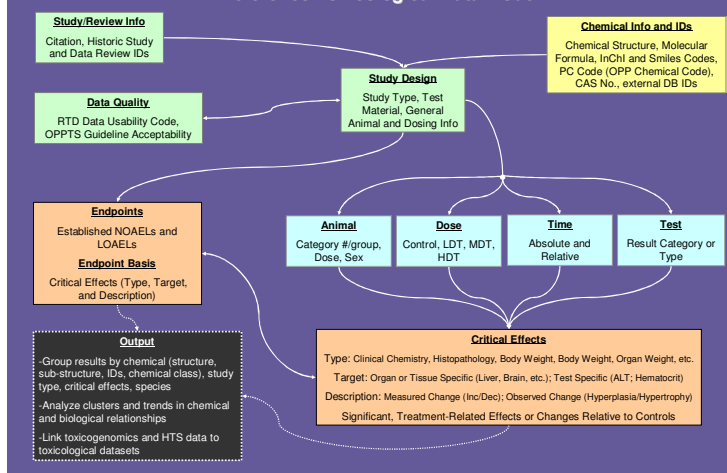
## Data Extraction Process



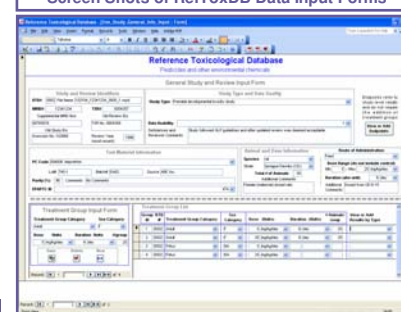
## IBM's Automated Text Extraction Process



## Reference Toxicological Data Model

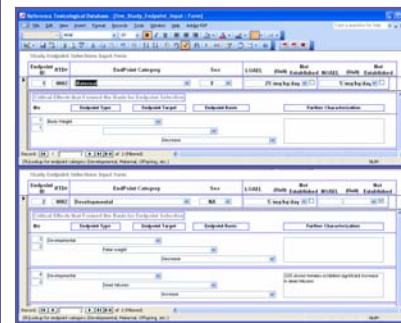


## Screen Shots of RefToxDB Data Input Forms



## Development of the RefToxDB Data Model and Lexicon

- ToxML compatibility and interoperability
- Standardized fields and vocabulary
  - Study Type (OPPTS/OECD Test Guidelines)
  - Data Usability (Data Quality) Code
  - Animal Info (Species, Strain, Sex Category)
  - Treatment Group Category (Adult, Offspring, etc.)
  - Endpoint Category (Systemic, Maternal, etc.)
  - Effect Descriptors (Type, Target, and Description Vocabulary)



## Results/Conclusions

Following retrieval and extraction of reference toxicological data, the uniformly populated database will serve three main functions:

- Demonstrate toxicological diversity among reference set of chemicals in various research efforts
- Provide qualitative associations to gene, pathway, protein, and cellular effects at the tissue, organ, and whole animal level
- Derive quantitative relationships across chemical and biological space both known and unknown

## Impact and Outcomes

RefToxDB will supply multiple ORD, Program Office, and Regional Office collaborations with the ability to obtain supporting toxicological data relative to specific research efforts using specific query and analytic tools.

With traditional toxicological testing currently driving toxicological profiling and risk assessment, RefToxDB provides a qualitative and quantitative link between traditional testing methods and new bioactivity and toxicological profiling technologies along the source-to-outcome continuum.

## Future Directions

RefToxDB will initially provide access to toxicological data for internal Agency research efforts, including ToxCast, toxicogenomic applications, and retrospective analyses of traditional toxicological testing. As RefToxDB develops, in concert with other computational tools within the NCCT, the eventual goal is to make this a public resource.

## Acknowledgements

Tina Levine and Jess Rowland (OPP) supported retrieval of toxicological data and guidance in handling the information. Brenda Groszkinsky (Region 7) provided helpful advice on issues relevant to Regional and Program Offices. Chihao Yang (Leadscope Inc.) continues to support the development of a standard toxicological lexicon and facilitated the development of ToxML.

Long-term goal: Develop computational framework for comparison across diverse toxicological data



This poster does not necessarily reflect EPA policy. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.