The Future of EPA's Toxicity Reference Database, ToxRefDB

Madison Feshuk[1], Jason Brown[1], Amar Singh[1], Katie Paul Friedman[1]
[1] Center for Computational Toxicology and Exposure, ORD, US EPA, RTP, NC

Madison Feshuk | Feshuk.madison@epa.gov | 919-541-5570

Rationale

In vivo study results inform toxicity predictions as training data or may be used to build scientific confidence in the performance of new approach methodologies (NAMs). However, these efforts require NAM and animal study data to be computationally accessible and interoperable. An application-driven curation workflow and graphical user interface will be created to support expansion and use of the chemical and study data in the Toxicity Reference Database (ToxRefDB) for NAM development and evaluation.

Previous Versions

ToxRefDB contains guideline and guideline-like in vivo study data that serve as a resource for retrospective and predictive toxicology.

ToxRefDB v1.0 captured basic study design, treatments, and treatment-related effects (for only "positive" results).

ToxRefDB v2.0 improved v1.0 via a manual curation effort using Microsoft Excel and Access files. This database release included:

- Curated information from nearly 6000 in vivo studies;
- Chemical library of over 1,000 chemicals;
- **Standardized Terminology** for endpoints and effects;
- Study reliability evaluations;
- Guideline level information, and
- **Endpoint Observation Status**.

See Watford *et al.* (2019)
DOI: 10.1016/j.reprotox.2019.07.012

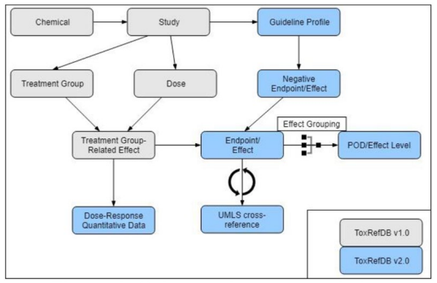

Figure 1: Schema changes made from ToxRefDB v1.0 to ToxRefDB v2.0

Endpoint Observation Status

Endpoint Observation Status distinguishes negative and missing (no tested) effects based on the study's specific guideline requirements.

Table 1: Endpoint Observation Status

Tested Status	Reported Status	Assumption
Yes	Yes	The study explicitly states the endpoint was measured, or data was presented in tables for the endpoint.
No	Yes	The study explicitly states the endpoint was not measured or data was not collected, even though the endpoint was required by the study guidelines.
Yes	No	The study does not state the endpoint was measured and data for the endpoint is not present. However, other evidence suggests that the endpoint was measured.
No	No	The endpoint was NOT specifically written in the text of the study document and is not required by the guideline, so it is assumed that the endpoint was not collected in this study.

Expansion using the Data Collection Tool (DCT)

The Data Collection Tool (DCT) was developed using low code platform Oracle APEX software for the curation of additional legacy documents with enhanced quality control and data provenance capabilities. The DCT was designed to replace this workflow for ToxRefDB and create a more sustainable process for loading curated information back to MySQL.

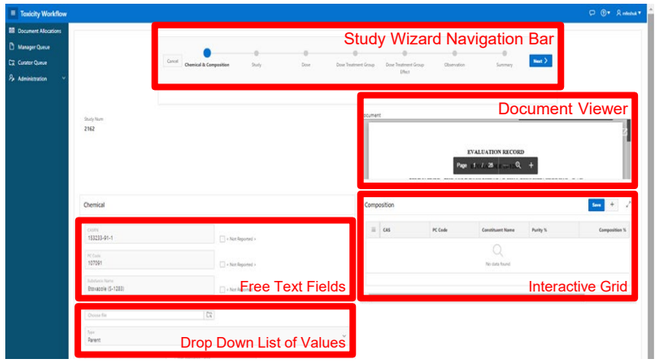

Figure 2: Within a DCT Curation: The study wizard navigation bar includes: Chemical & Composition, Study, Dose, Dose Treatment Group, Dose Treatment Group Effect, and Observation. The Chemical & Composition page (displayed) features a document viewer that allows curators to view documents on a new tab within their internet browser.

For each document, up to two curators will independently review and extract basic study design metadata, dosing information, significant treatment-related and critical effects, and endpoint observation testing status using free text fields, drop down list of values, or interactive grids. A manager role reviews after one or two curators to reconcile conflicts and ensure data quality.

The DCT provides document allocation, curation and workflow management among users, and management review with data conflict resolution, resulting in a record that directly links quality-controlled curation to source documents. The DCT offers flexibility via its modular workflow for curating the heterogeneous and complex in vivo study designs, that may include chemical mixtures, adjusted doses by sex, multiple generations, and/or additional treatment groupings (interim, recovery, satellite) under examination.

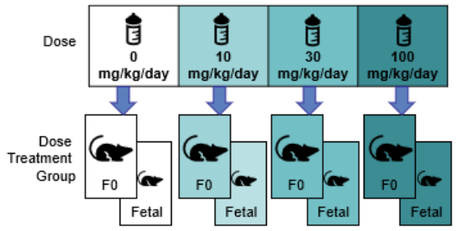

Figure 3: Dose Treatment Group Example As a prenatal developmental study design example, the number of doses administered (4) is multiplied by the number of treatment groups (2, dams and their litters) to determine the number of dose treatment groups (8) examined for effects at sacrifice.

Reviewers at EPA's Office of Pesticide Programs (OPP) create data evaluation records (DERs) for each in vivo study submitted in accordance with EPA testing guidelines for pesticide registration. Phase I DCT extraction will focus on developmental toxicity DERs from pesticide submissions. Current curation efforts focus on DERs published since 2008 and not currently in ToxRefDB.

Standardized Terminology

The DCT employs controlled terminology standardized to better reflect both the OCSPP Health Effects 870 series guidelines and DER summary reporting. This hierarchical relationship of effects and endpoints was adapted from ToxRefDB. Novel values can be added when found during a curation.

Observation
endpoint category, endpoint type, and endpoint target.

Observation
reproductive | reproductive performance | postimplantation loss

Effect
specific condition associated endpoint target

Effect Description
postimplantation loss

Treatment Group Effect
Lifestage, location, verbatim text

Life Stage	Target Site	Effect Description Free
adult pregnancy	uterus	postimplantation site loss: mean

Figure 4: Example of hierarchical terminology used for the effect of "postimplantation loss"

Data Visualization

ToxRefDB v2.0 is available for download as a SQL database with data stored across 28 tables. A graphical user interface will be used to visualize legacy in-vivo study data and serve as a prototype for future public interfaces.

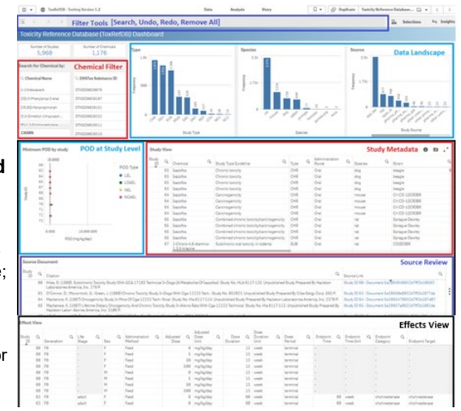

Figure 5: ToxRefDB Dashboard
This prototype dashboard includes:

- A filter to select studies by chemical;
- An overview of the data landscape by study type, species, and source;
- Calculated point of departure (POD) values by study;
- Study metadata;
- Hyperlinks to source documents for additional reference; and
- An extended table with dose-response, treatment-related and critical effects information.

Future Directions

ToxRefDB expansion with its new application-driven curation workflow and the creation of a ToxRefDB dashboard will increase its utility. These efforts will be necessary to promote development of NAMs and data interoperability for computational and regulatory toxicology applications to ultimately achieve reductions in animal testing.

Innovative Research for a Sustainable Future