Utilizing automated and semi-automated data analytic tools for curating data in the ECOTOX Knowledgebase

Dale J. Hoff
U.S. EPA ORD/CCTE/GLTED
Duluth, MN

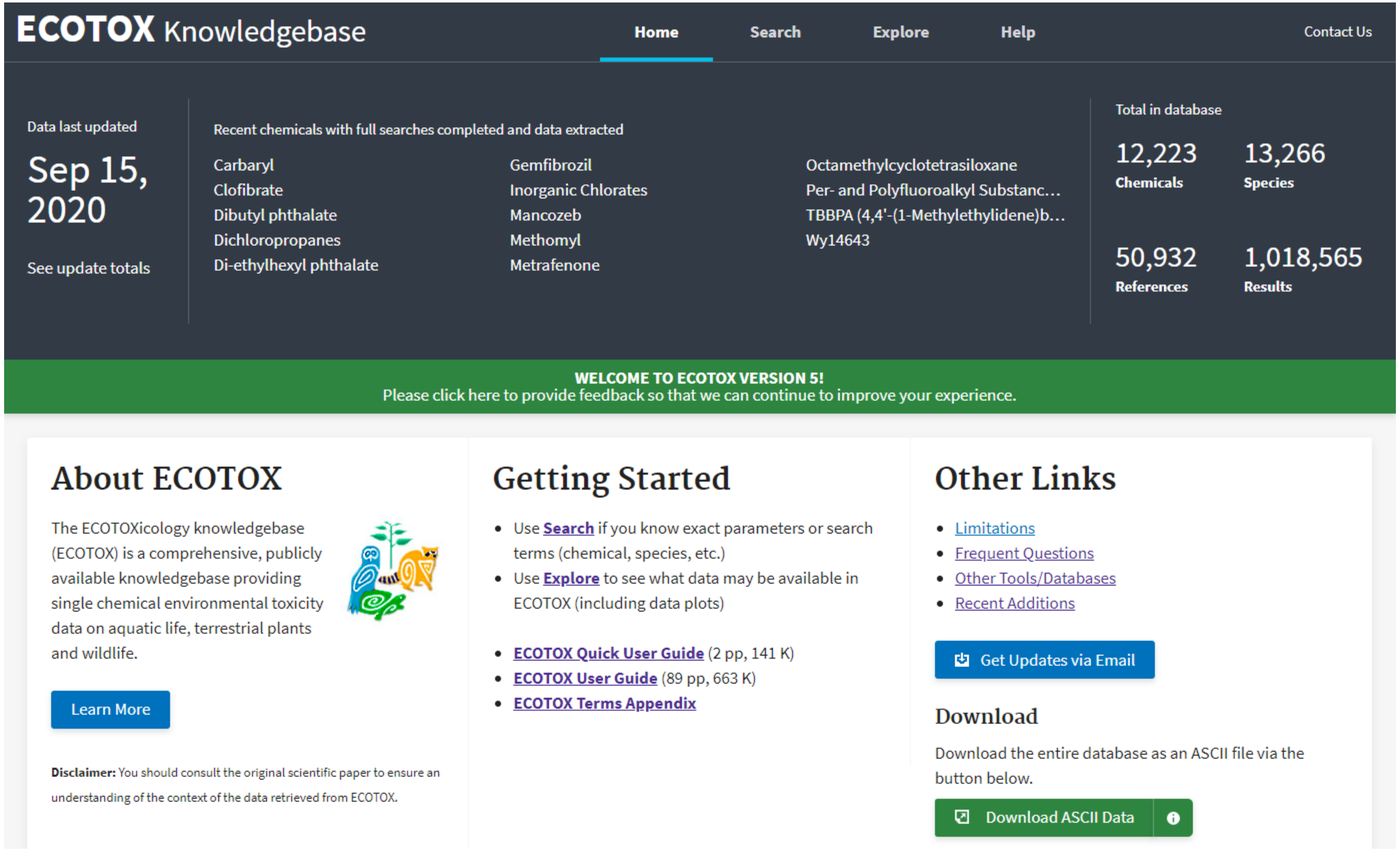
SETAC North America 41st Annual Meeting, SciCon_2
15-19 November 2020

Identifying Empirical Evidence: ECOTOX Knowledgebase

www.epa.gov/ecotox

Curated database providing single chemical environmental toxicity data for aquatic life, terrestrial plants and wildlife

- Comprehensive literature searches
- Literature review
- Data extraction



ECOTOX presentations/posters
J. Olker ID: 1.14.16
C. Elonen ID: 5.17.09

Uses of the ECOTOX Knowledgebase in Environmental Decision Making

United States Environmental Protection Agency | Office of Water 4304T | EPA 822-R-16-006 June 2016

Aquatic Life Ambient Water Quality Criterion for Selenium – Freshwater 2016

U.S. Environmental Protection Agency
Office of Water
Office of Science and Technology
Washington, D.C.

Used for Ambient Water Quality Criteria for Aquatic Life since 1985.

Used for Ecological Risk Assessment for Office of Pesticides for chemical registration and re-registration (FY19 – 30 chemicals).

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY
WASHINGTON D.C., 20460

MEMORANDUM March 26, 2008

Subject: Registration Review –Preliminary Problem Formulation for Ecological Risk and Environmental Fate, Endangered species and Drinking Water Assessments for Diazinon (PC Code 057801; DP Barcode D349527)

To: Jude Andreasen, Chemical Review Manager
Laura Parsons, Team Leader
Special Review Branch
Special Review and Reregistration Division (SRRD)

From: Kristina Garber, Biologist
Thomas Steeger, Senior Biologist
Environmental Risk Branch 4
Environmental Fate and Effects Division
Office of Pesticide Programs

Through: Elizabeth Behl, Chief
Environmental Risk Branch 4
Environmental Fate and Effects Division
Office of Pesticide Programs

The Environmental Fate and Effects Division (EFED) has completed the preliminary problem formulation (attached) for the ecological risk, environmental fate, endangered species, and drinking water assessments to be conducted as part of the Registration

Texas Commission on Environmental Quality

Home | Air | Land | Water | Licenses | Permits | Reporting

Data | Forms | Maps | Public Notices

Home / Remediation / Eco / Useful Links for Developing Ecological Risk Assessments

>> Questions or Comments: techsup@tceq.texas.gov

Useful Links for Developing Ecological Risk Assessments

Links to pages on this and other sites that have useful information for developing ecological risk

U.S. EPA

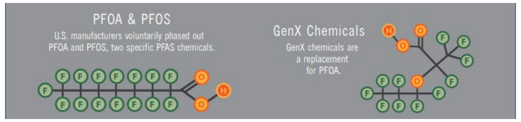
- Integrated Risk Information System (IRIS)
- ECOTOX Database (aquatic and terrestrial toxicological data)
- EPA Office of Water (links to sediment guidance, water quality standards, and other useful topics)
- EPA People Locator
- EPA Region 4 Ecological Screening Values
- Combustion Guidance for Human Health (some defaults in here used at times)
- Superfund Risk Assessment (variety of links)
- Superfund: Natural Resource Damages and Ecological Risk Assessments

Used by Regions and States for hazardous waste site assessments and in emergency response

Providing ecological hazard data for the prioritization and assessment of chemicals for TSCA/Lautenberg Act

Overview of TSCA Work Plan Methodology

Maria Doa
U.S. EPA, Office of Pollution Prevention and Toxics
December 11, 2017



Providing ecological toxicity data for PFAS to researchers, EPA ERA Forum, DoD Tri-Services ERA Work Group, and others

Ecological Hazard

Ecological hazard data are extracted from the EPA ToxValDB database where it had been compiled from the EPA ECOTOX database. Although data are available for a variety of species, only data for aquatic species are used in the current illustration. The data can come from any of the following study types: mortality:acute, mortality:chronic, reproductive:acute, reproductive:chronic, growth:acute, growth:chronic (all from ECOTOX). The types of effect levels are LDxx/LCxx/ECxx/EDxx where xx can range from 1% to 100%, and LOEL/NOEL/LOEC/NOEC. Values must be in units of mg/L. For each chemical, the lowest toxicity value was separately determined for acute and chronic studies, regardless of species. The

Need for high volume rapid data identification for assessment purposes

- The accelerated pace of chemical risk assessment for ecological receptors
- Discovery of “new” chemicals of ecotoxicological concern
- Increase in number and diversity of journals
- Changing landscape of toxicity data used for ERA
- Decreasing dependency of traditional whole animal testing for ERA
- Development of tools to expedite gathering of information to focus shortened time frames on data integration and professional judgement
- Development of tools that can sort rapidly through BROAD keyword searches (e.g. chemical name)
- Digital formats of journals allow for use of data analytic applications
- Shift from apical endpoints to NAMs requires transparent development of controlled vocabulary for systematic data curation
- Providing a means to take full advantage of existing data before conducting new toxicity studies

ECOTOX Data Curation Pipeline

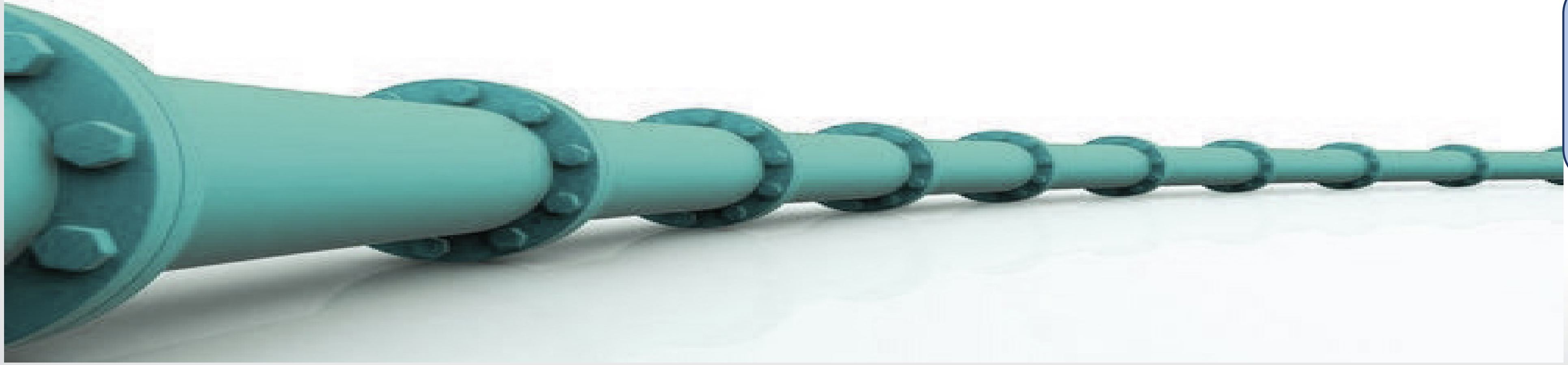
Chemical verification and development of search terms

Conduct literature searches

Identify and acquire potentially applicable studies

Review literature for applicability to ECOTOX

Extract data into ECOTOX Knowledgebase



Data Evaluation

Chemical verification and development of search terms

Ecotox Task 1 - Search Engine(s) Search

Provided: Sep. 4th, 2020

Save message and link number numerical sorting added: Sep. 23rd, 2020

Search terms: Diazinon [Search Selected Search Engines]

Google
Bing
Yahoo
Wikipedia

Documents to parse (max 100): 20

Fixed Flags set

fate
transport
degrad*
metaboli*
dissipat*
convert
conversion
route

Fate
Degrad*
Metab*

User added flags

CASRN

Directed search sites

California Department of Pesticide Regulation	https://www.cdpr.ca.gov/docs/chemical/monster2.htm
EPA Chemistry Dashboard	https://comptox.epa.gov/dashboard
University of Hertfordshire PPDB: Pesticide Properties DataBase	https://sitem.herts.ac.uk/aeru/ppdb/en/search.htm
RED documents	https://archive.epa.gov/pesticides/reregistration/web/html/status.html
Office of Pesticide Programs Pesticide Chemical Search	https://ofmpub.epa.gov/apex/pesticides/f?p=CHEMICALSEARCH:1
Food and Agriculture Organization of the United Nations	http://www.fao.org/home/en/
World Health Organization	http://apps.who.int/pesticide-residues-jmpr-database/
Cornell University Extoxnet	http://pmep.cce.cornell.edu/profiles/extoxnet/index.html
PAN Pesticides Database	http://www.pesticideinfo.org/Search_Chemicals.jsp

Computer Assisted Search Term Generation

Enter Chemical Name

Select Search engine (e.g. Google) for web scrubbing

Enter fixed/default or other optional “flags” or “tags to assist in finding synonyms or related terms.

Application-Programing Interface (API) to query established online chemical identification resources to return synonyms, product formulation names, alternate CASRN, etc.

Chemical verification and development of search terms

Text mining: highlights flags/tags within document

Users more rapidly find associated terms

In this example, Metabolites are identified

4.1.2 Distribution

(a) Humans

Poklis et al. (1980) detected diazinon in tissues (blood, bile, adipose, liver, brain, and kidney) after intentional oral ingestion of diazinon. No other data on tissue distribution of diazinon in humans were available to the Working Group.

(b) Experimental systems

In experimental animals, diazinon is widely distributed to tissues after absorption. The elimination half-life of diazinon in the blood of male Wistar rats given intraperitoneal doses of 20 mg/kg bw or 100 mg/kg bw was estimated to be 4 hours and 6 hours, respectively (Tomokuni et al., 1985). Similarly, immediately after administration of intravenous (10 mg/kg bw) and oral (80 mg/kg bw) doses in rats, plasma concentrations of diazinon indicated half-lives of 4.7 and 2.9 hours, respectively (Wu et al., 1996). Most diazinon in the plasma (89%) is bound non-covalently to albumin and other plasma proteins (Wu et al., 1996; Poet et al., 2004). By 8 hours after intravenous administration (20 mg/kg bw) to rats, the concentration of diazinon was significantly higher in the kidney than in the liver, or brain (Tomokuni et al., 1985). After intravenous dosing (1 or 10 mg/kg bw), diazinon was distributed and eliminated rapidly in male Sprague-Dawley rats, and concentrations of diazinon in saliva were comparable to plasma concentrations of non-protein-bound diazinon (Lu et al., 2003).

4.1.3 Metabolism

(a) Overview of metabolism of diazinon

Organophosphate pesticides are subject to similar metabolic pathways in humans and experimental animals in vivo (Casida & Quistad, 2004); see also Section 4.1.3 of the *Monograph* on Malathion in the present volume. Biotransformation of organophosphates occurs primarily in the liver, and to a lesser extent in the small intestine, after oral exposure (Barr & Angerer, 2006). After absorption by the dermal or oral route, diazinon is rapidly biotransformed by several enzymes – including cytochrome P450 (CYP), paraoxonases, and carboxylesterases (CES) – to water-soluble metabolites that are rapidly eliminated (see Fig. 4.1). Both desulfuration and dearylation of diazinon are mediated by CYP. The bioactive diazoxon metabolite can be detoxified by paraoxonase (PON1)-catalysed reactions (Costa et al., 2013), yielding alcohol and diethylphosphate products. Alternatively, diazoxon can be subject to inhibition of CES function (Crow et al., 2012; Fig. 4.1). The oxon metabolite can escape detoxication by CES or PON1 in the liver and instead covalently modify (and inhibit) various serine hydrolase enzymes, including the B-esterase targets butyrylcholinesterase, acetylcholinesterase, and CES (Casida & Quistad, 2004; see Fig. 4.2). The bioactive oxon metabolite is generated by CYP-catalysed desulfuration (Buratti et al., 2005; Barr & Angerer, 2006). If the oxon is not degraded by hepatic paraoxonase or carboxylesterases, it can escape the liver and instead covalently modify (and inhibit) various serine hydrolase enzymes, including the B-esterase targets butyrylcholinesterase, acetylcholinesterase, and carboxylesterases (Casida & Quistad, 2004; see Fig. 4.2). Generation of the oxon metabolite is a bioactivation reaction, because the oxon is a much more potent inhibitor of B-esterases than the parent compound (Casida & Quistad, 2004). In general, analytical measurement of the oxons in blood is difficult due to the small quantities of metabolite that are formed and its relative instability (Timchalk et al., 2002). Nevertheless, the oxons are potent inhibitors of serine hydrolases, exhibiting bimolecular rate constants of inhibition varying from 10^3 to 10^7 $M^{-1}s^{-1}$, depending on the hydrolase and the specific oxon (Casida & Quistad, 2004; Crow et al., 2012). Most important with respect to the insecticidal and toxicological activity of the oxon is acetylcholinesterase, the

26	0
27	0
28	0
29	0
55	0
56	5
57	3
58	2
59	1

Angerer, 2006). After absorption by the dermal or oral route, diazinon is rapidly biotransformed by several enzymes – including cytochrome P450 (CYP), paraoxonases, and carboxylesterases (CES) – to water-soluble metabolites that are rapidly eliminated (see Fig. 4.1). Both desulfuration and dearylation of diazinon are mediated by CYP. The bioactive diazoxon metabolite can be detoxified by paraoxonase (PON1)-catalysed reactions (Costa et al., 2013), yielding alcohol and diethylphosphate products. Alternatively, diazoxon can be subject to inhibition of CES function (Crow et al., 2012; Fig. 4.1). The oxon metabolite can escape detoxication by CES or PON1 in the liver and instead covalently modify

Conduct literature searches

WR002 ECOTOX Literature Search for Diazinon 2019

Anita Pomplun, GDIT	**MEMORANDUM**	
Anne Pilli, GDIT	**To:**	Colleen Elonen, EPA
Arne Anderson, GDIT	**From:**	Brian Kinziger, GDIT
Brian Kinziger, GDIT	**CC:**	Anne Pilli, GDIT
Chris Suomi, GDIT	**CC:**	Arne Anderson, GDIT
Colleen Elonen, EPA	**Search Conducted by:**	Conductor
Gage Sachs, GDIT	**Document Drafted by:**	Drafter
Jennifer Olker, EPA		
Jennifer Olsen, GDIT		
John Frisch, GDIT	**Project:**	WR002 ECOTOX
Kara Tudor, GDIT	**Chemical:**	Diazinon
Katie Nehiba, GDIT	**Report Year:**	2019
Travis Karschnik, GDIT	**Chemicals of Concern:**	chem of concern
Tyler Gephart, GDIT	**UNIFY Title Search String:**	

3

Literature Search Terms

diazinon

Quota Check

Hero	Proquest	PubAg	PubMed	Science Direct FT	Science Direct TK	Scopus	Springer	Theses	USGS	Web of Science
☑	☑	☑	☑	☐	☑	☑	☑	☑	☑	☑

Send

Conducted in Abstract Sifter Plus*

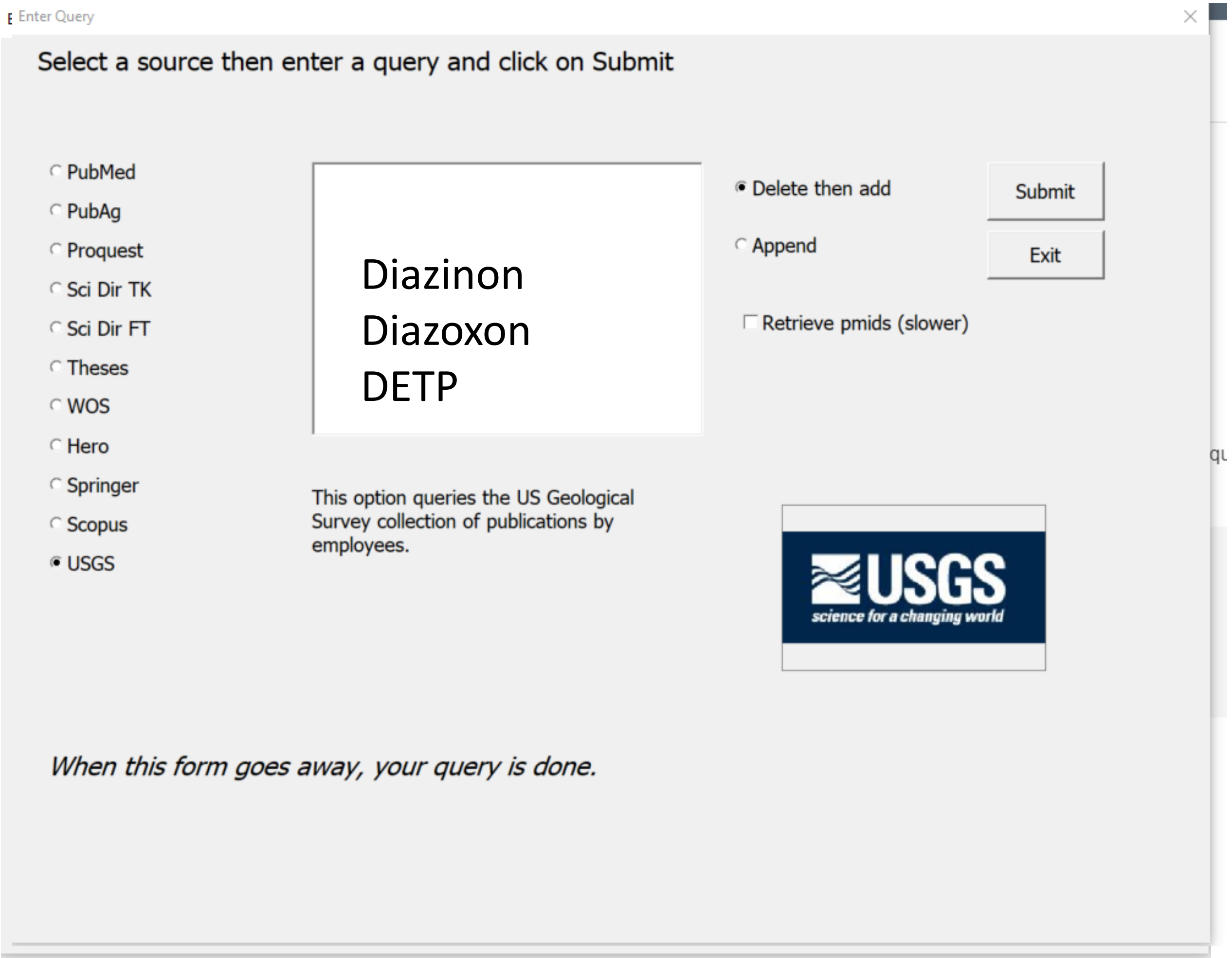
Center for Computational Toxicology and Exposure, EPA's PubMed Abstract Sifter. The United States Environmental Protection Agency's Center for Computational Toxicology Exposure. Software. https://doi.org/10.23645/epacomptox.10324379.v1

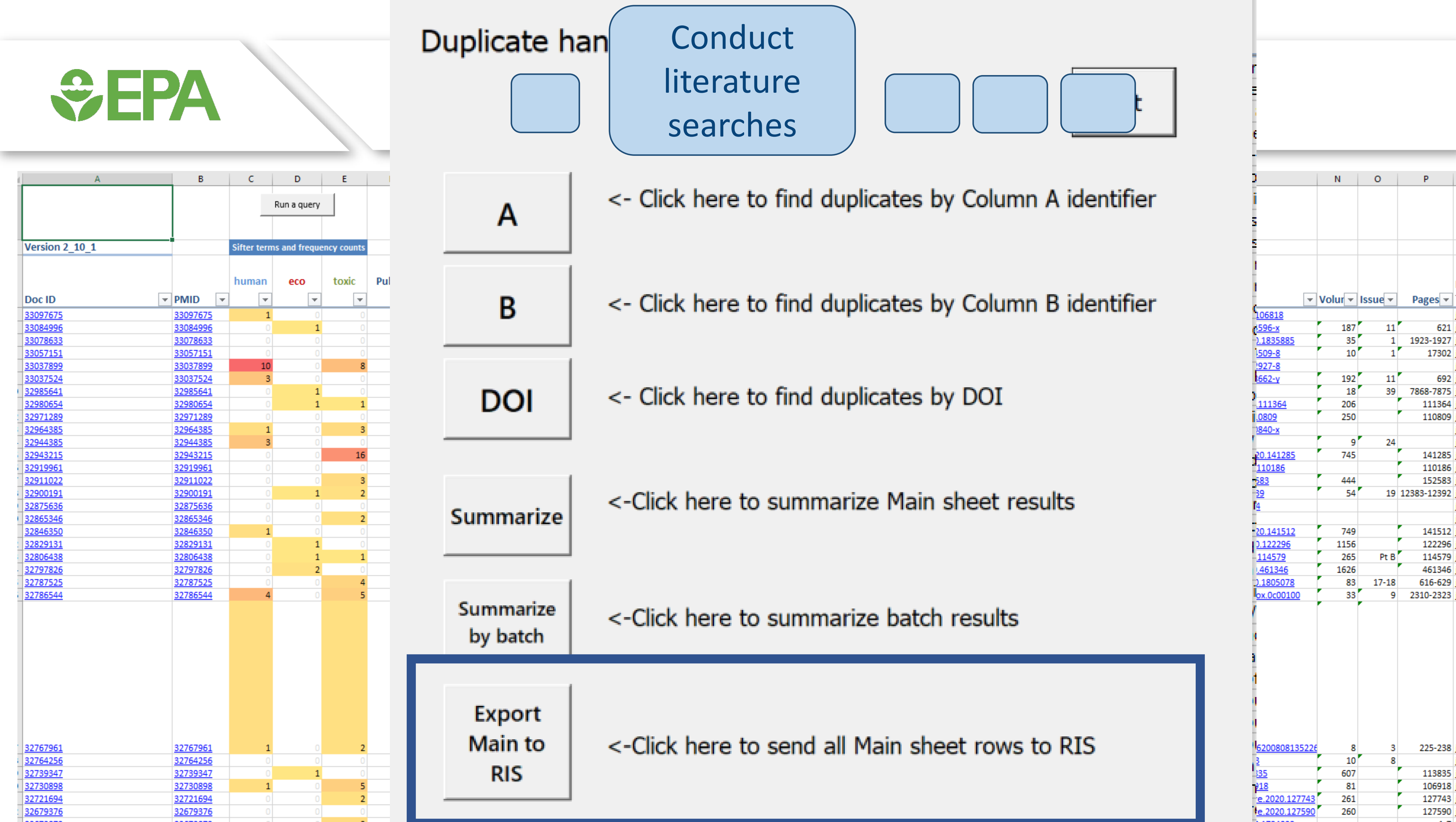
*Customized for Ecotoxicology

Conduct literature searches

Web Services Automated searches of several citation index databases

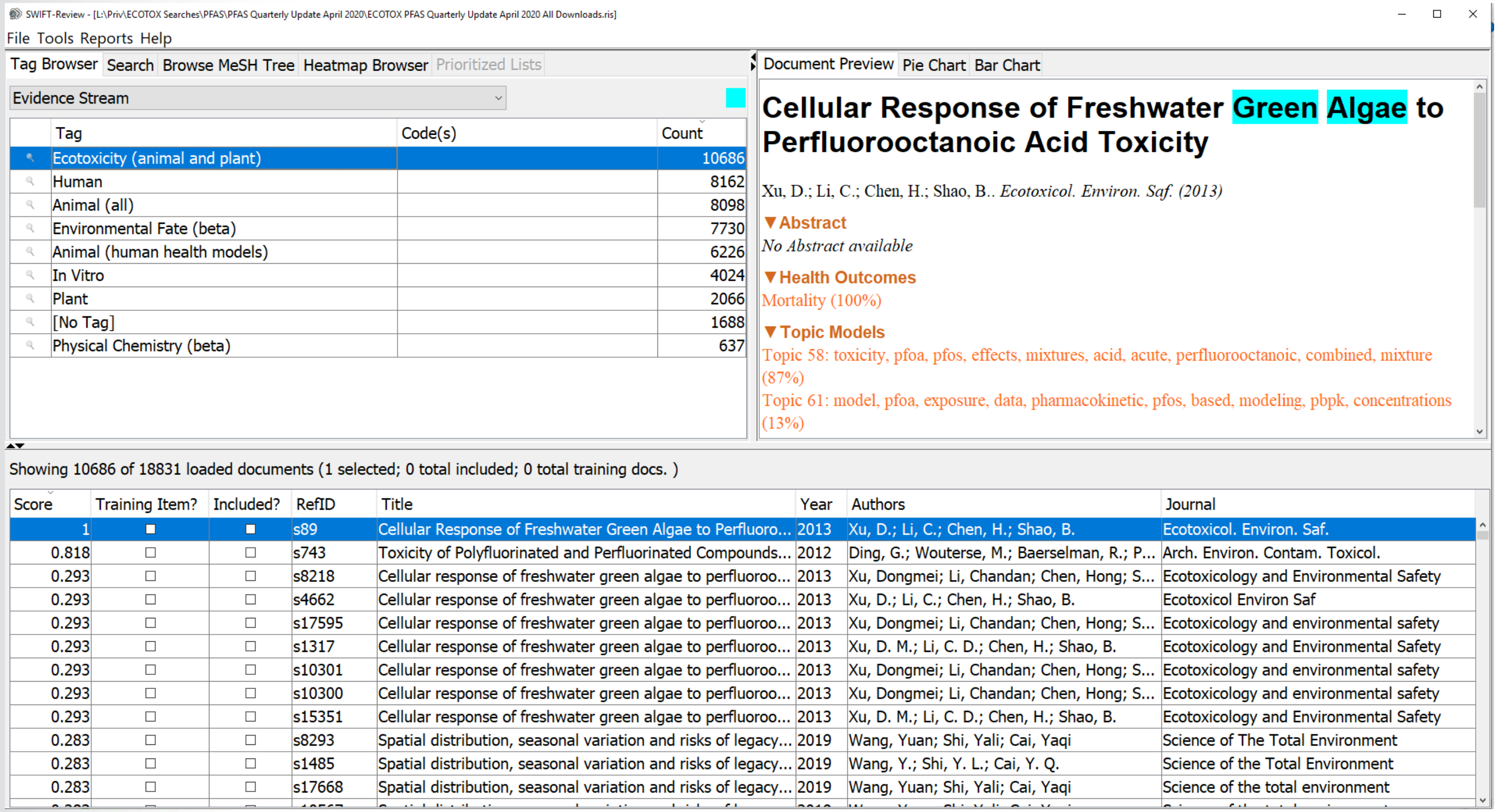
1.) PubMed
2.) PubAg - Agricola
3.) Proquest
4.) Science Direct –Title & Keywords
5.) Science Direct – Full Text
6.) Theses: PMC library of dissertations and theses
7.) Web of Science
8.) USEPA Hero database
9.) Full text of Springer Nature articles
10.) Scopus-Elsevier title abstract and Key words
11.) USGS collection of employee pubs

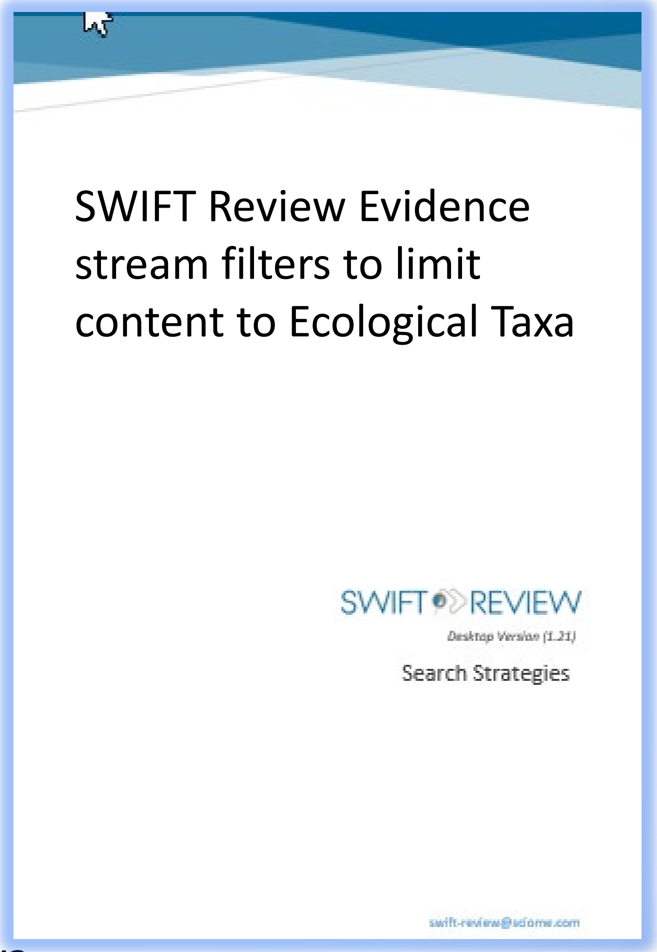




Duplicate han
Conduct literature searches
A
<- Click here to find duplicates by Column A identifier
B
<- Click here to find duplicates by Column B identifier
DOI
<- Click here to find duplicates by DOI
Summarize
<-Click here to summarize Main sheet results
Summarize by batch
<-Click here to summarize batch results
Export Main to RIS
<-Click here to send all Main sheet rows to RIS
Run a query
Version 2_10_1
Sifter terms and frequency counts
human
eco
toxic
Doc ID
PMID
Volur
Issue
Pages

Identify and acquire potentially applicable studies





All references loaded in Swift Review and Evidence Stream “Filter” applied.

Ecotoxicity evidence stream identifies titles and abstracts with Ecologically relevant Taxa

In this example, 18,631 References were reduced to 10,686 by applying the evidence stream filter

Identify and acquire potentially applicable studies


SWIFT REVIEW

Desktop Version (1.21)

Search Strategies

swift-review@sciome.com

Ecotoxicity (animal and plant)

tiab :("Norway Rat" OR "Rattus norvegicus" OR "Rainbow Trout" OR "Oncorhynchus mykiss" OR "Water Flea" OR "Daphnia magna" OR "Zebra Danio" OR "Danio rerio" OR "Fathead Minnow" OR "Pimephales promelas" OR "House Mouse" OR "Mus musculus" OR "Common Carp" OR "Cyprinus carpio" OR "Bluegill" OR "Lepomis macrochirus" OR "Domestic Chicken" OR "Gallus domesticus" OR "Japanese Medaka" OR "Oryzias latipes" OR "Mallard Duck" OR "Anas platyrhynchos" OR "Goldfish" OR "Carassius auratus" OR "Corn" OR "Zea mays" OR "African Clawed Frog" OR "Xenopus laevis" OR "Green Algae" OR "Pseudokirchneriella subcapitata" OR "Honey Bee" OR "Apis mellifera" OR "Bread Wheat" OR "Triticum aestivum" OR "Soybean" OR "Glycine max" OR "Northern Bobwhite Quail" OR "Colinus virginianus" OR "Water Flea" OR "Ceriodaphnia dubia" OR "Nile Tilapia" OR "Oreochromis niloticus" OR "Rice" OR "Oryza sativa" OR "Channel Catfish" OR "Ictalurus punctatus" OR "Yellow Fever Mosquito" OR "Aedes aegypti" OR "Earthworm" OR "Eisenia fetida" OR "Silver

Sciome
Enabling Science via Analytical Informatics

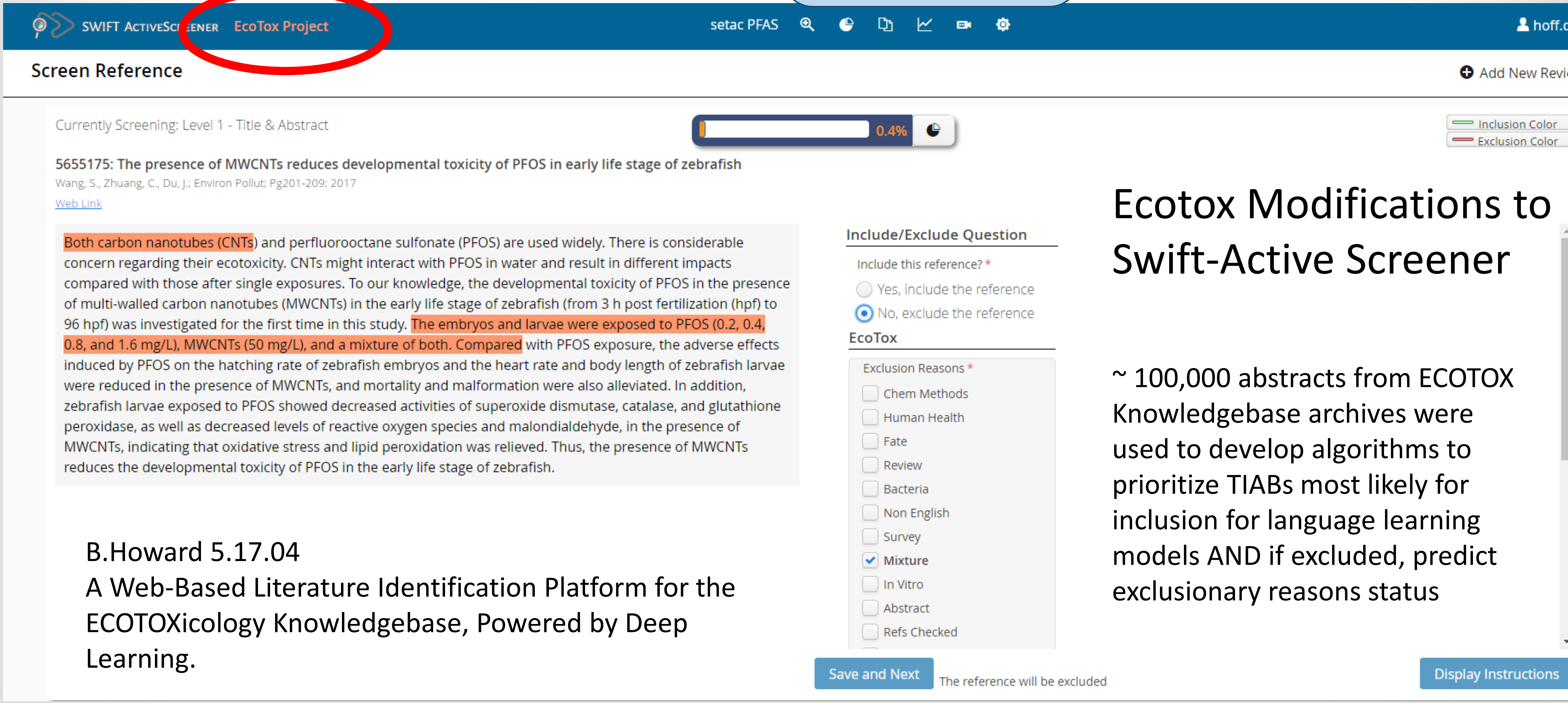
Page **81** of **210**

~25,000 Scientific & common names from all species with toxicity data identified in ECOTOX Knowledgebase + Generic species habitat tags (e.g. AQUATIC,AVIAN TERRESTRIAL, BENTHIC)

Identify and acquire potentially applicable studies

Ecotox Modifications to Swift-Active Screener

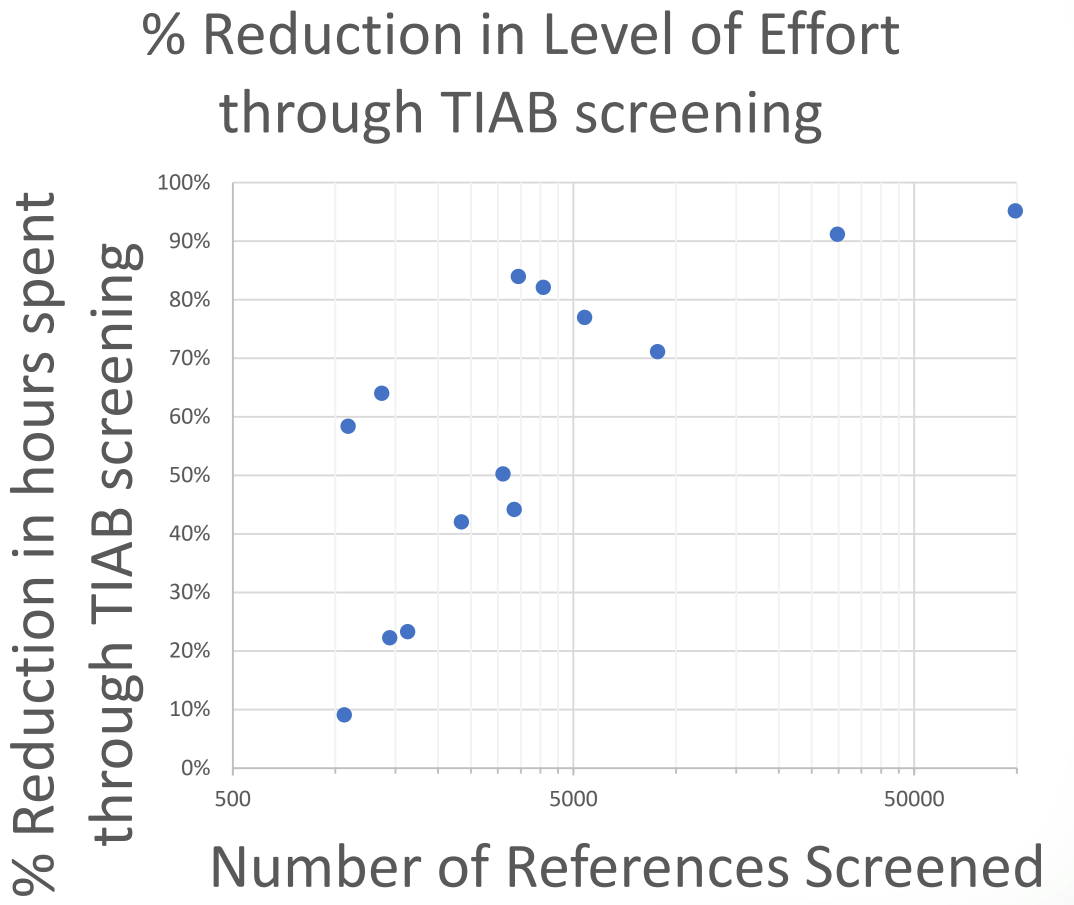
~ 100,000 abstracts from ECOTOX Knowledgebase archives were used to develop algorithms to prioritize TIABs most likely for inclusion for language learning models AND if excluded, predict exclusionary reasons status



B.Howard 5.17.04
A Web-Based Literature Identification Platform for the ECOTOXicology Knowledgebase, Powered by Deep Learning.

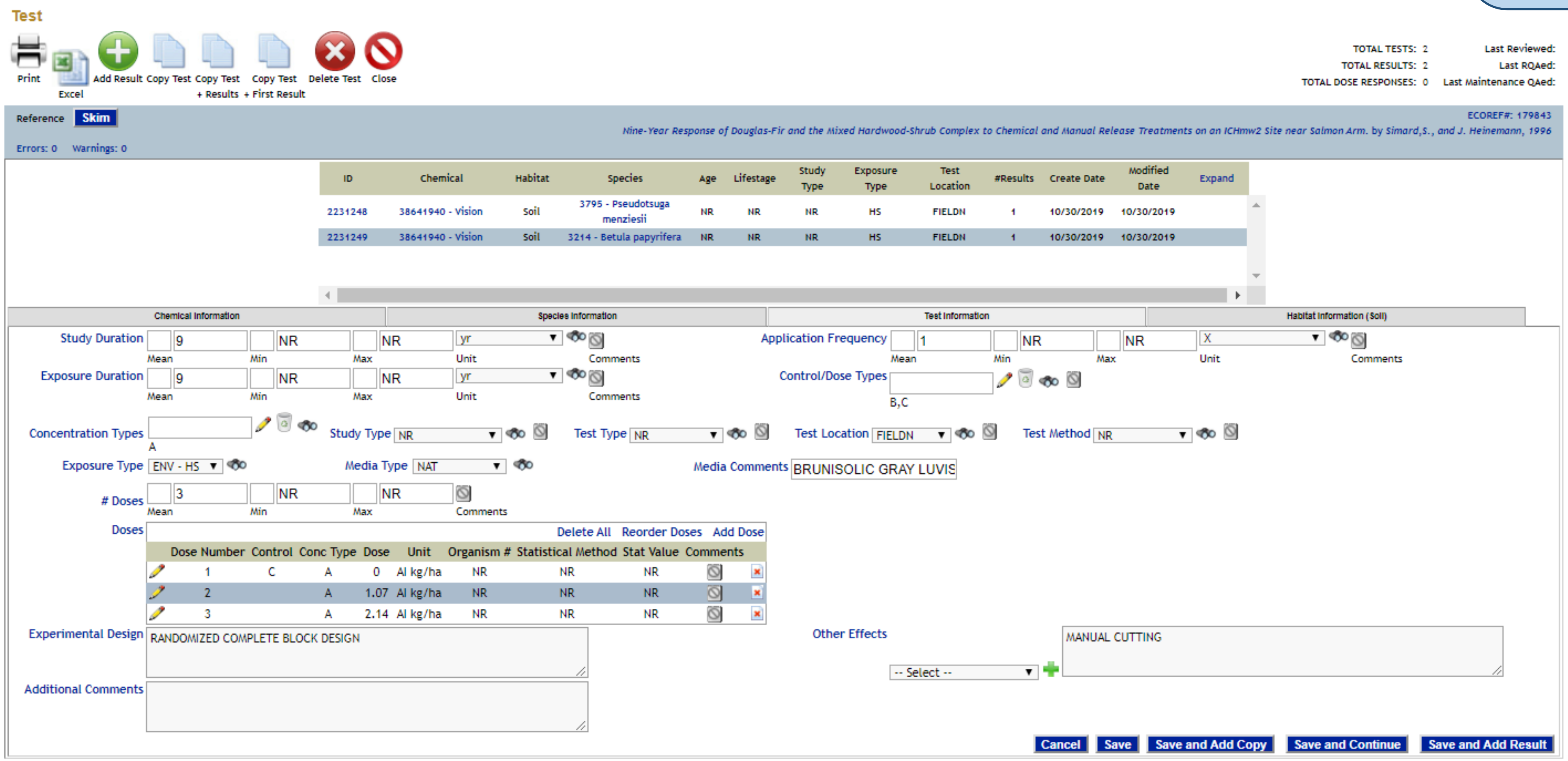
Measurable improvement in Efficiency using analytic tools

	# Refs	# of Hrs	Baseline normalized hrs/5000 refs	% change
Baseline	5000	127	127	0%
	Actual # Refs	Actual # of hrs		
PFAS	3443	14	87	-84%
Gemfibrozil	1369	13	35	-64%
Methomyl	1630	32	41	-23%
Clofibrate	5386	32	137	-77%
Chlorates	2345	35	60	-42%
Wy14643	1089	12	28	-58%
Carbaryl	8830	65	224	-71%
Dicamba	3106	39	79	-50%
Dicloran	1061	25	27	-9%
Clothianidin	1444	29	37	-22%
Thiamethoxam	3350	48	85	-44%
Phthalic acid	4076	19	104	-82%
Asbestos	29768	66	756	-91%
Nitrates	98902	120	2512	-95%



Average of 58% reduction in hours spent through TIAB screening / 5000 references

Extract data into ECOTOX Knowledgebase

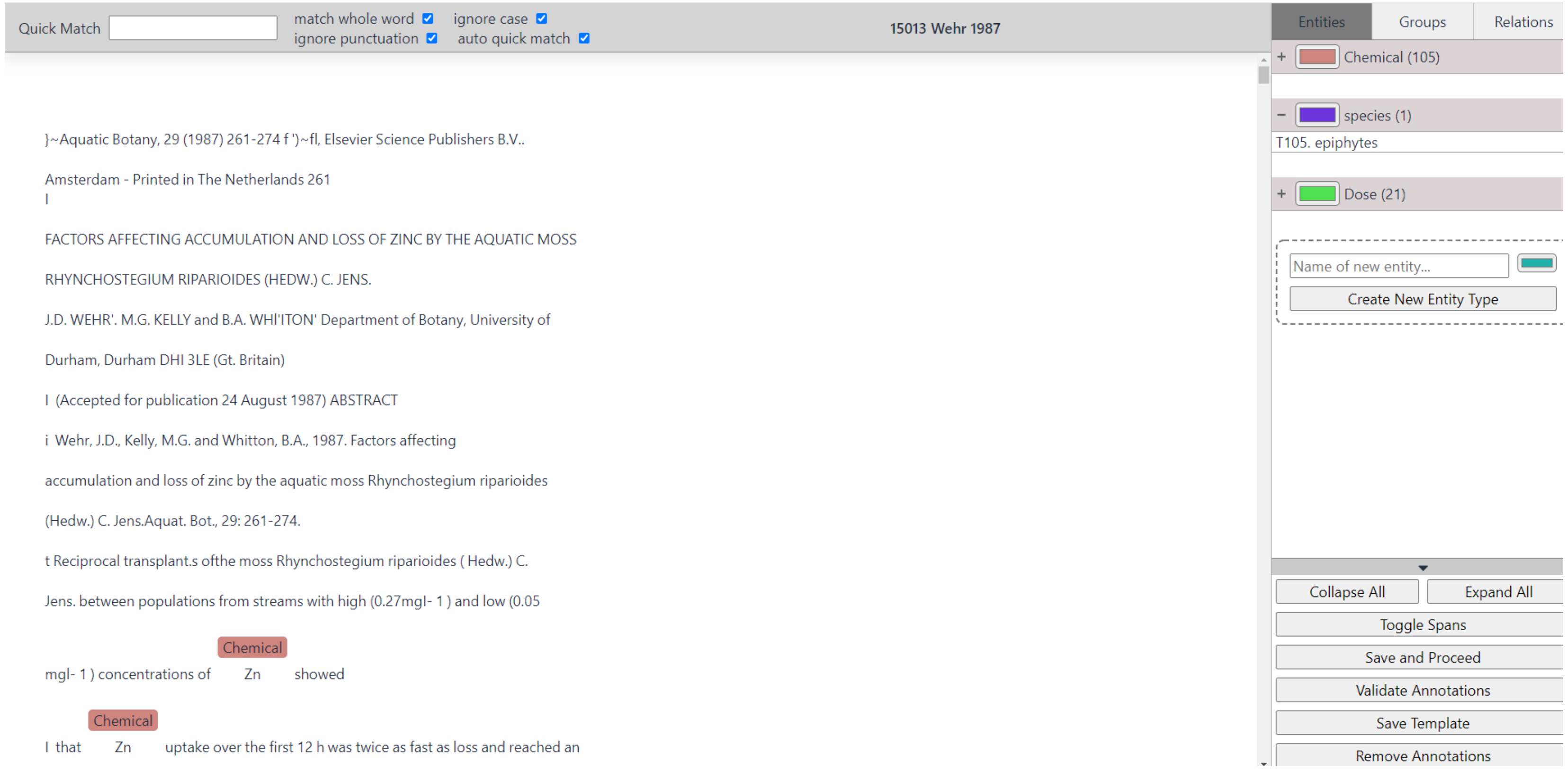


ECOTOX data curation in UNIFY

- Up to 250 fields of controlled vocabulary
- Computationally assisted
- Consistent extraction

Extract data into ECOTOX Knowledgebase

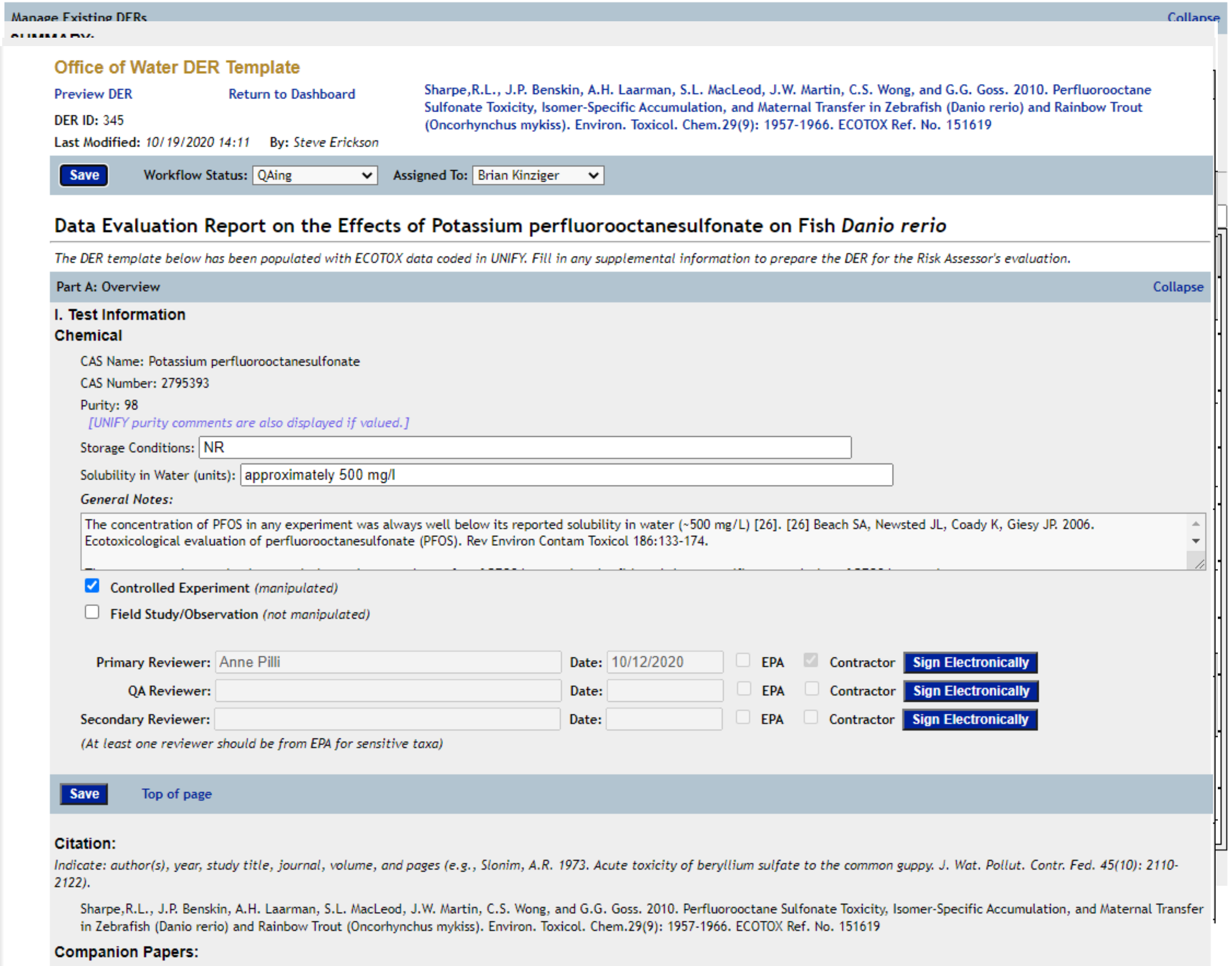
Automated/Semi-Automated Data Extraction



- Next priority is to develop automated/semi-automated capabilities for data extraction from identified .PDF documents
- Existing Controlled Vocabulary ready made for “Entity Definitions”

Developing Data Evaluation Tools

Collaboration with USEPA's Office of Water



Manage Existing DERs Collapse

Office of Water DER Template

Preview DER Return to Dashboard

Sharpe,R.L., J.P. Benskin, A.H. Laarman, S.L. MacLeod, J.W. Martin, C.S. Wong, and G.G. Goss. 2010. Perfluorooctane Sulfonate Toxicity, Isomer-Specific Accumulation, and Maternal Transfer in Zebrafish (Danio rerio) and Rainbow Trout (Oncorhynchus mykiss). Environ. Toxicol. Chem.29(9): 1957-1966. ECOTOX Ref. No. 151619

DER ID: 345

Last Modified: *10/19/2020 14:11* By: *Steve Erickson*

Save Workflow Status: QAing Assigned To: Brian Kinziger

Data Evaluation Report on the Effects of Potassium perfluorooctanesulfonate on Fish *Danio rerio*

The DER template below has been populated with ECOTOX data coded in UNIFY. Fill in any supplemental information to prepare the DER for the Risk Assessor's evaluation.

Part A: Overview Collapse

I. Test Information

Chemical

CAS Name: Potassium perfluorooctanesulfonate

CAS Number: 2795393

Purity: 98

[UNIFY purity comments are also displayed if valued.]

Storage Conditions: NR

Solubility in Water (units): approximately 500 mg/l

General Notes:

The concentration of PFOS in any experiment was always well below its reported solubility in water (~500 mg/L) [26]. [26] Beach SA, Newsted JL, Coady K, Giesy JP. 2006. Ecotoxicological evaluation of perfluorooctanesulfonate (PFOS). Rev Environ Contam Toxicol 186:133-174.

☑ Controlled Experiment *(manipulated)*

☐ Field Study/Observation *(not manipulated)*

Primary Reviewer: Anne Pilli Date: 10/12/2020 ☐ EPA ☑ Contractor Sign Electronically

QA Reviewer: Date: ☐ EPA ☐ Contractor Sign Electronically

Secondary Reviewer: Date: ☐ EPA ☐ Contractor Sign Electronically

(At least one reviewer should be from EPA for sensitive taxa)

Save Top of page

Citation:

Indicate: author(s), year, study title, journal, volume, and pages (e.g., Slonim, A.R. 1973. Acute toxicity of beryllium sulfate to the common guppy. J. Wat. Pollut. Contr. Fed. 45(10): 2110-2122).

Sharpe,R.L., J.P. Benskin, A.H. Laarman, S.L. MacLeod, J.W. Martin, C.S. Wong, and G.G. Goss. 2010. Perfluorooctane Sulfonate Toxicity, Isomer-Specific Accumulation, and Maternal Transfer in Zebrafish (Danio rerio) and Rainbow Trout (Oncorhynchus mykiss). Environ. Toxicol. Chem.29(9): 1957-1966. ECOTOX Ref. No. 151619

Companion Papers:

- Application pulls objective data extracted from existing records
- UNIFY platform now has ability for curators to add subjective observations
- Allows multiple reviewers
  - Primary
  - QA
  - Secondary

Acknowledgements - Coauthors

U.S. EPA ORD, CCTE

Great Lakes Toxicology and Ecology Division

Colleen Elonen

Jennifer Olker

General Dynamics Information Technology

Arne Anderson

Anne Pilli

Christina Suomi

Katie Nehiba

Tyler Gephart

Thank you!

hoff.dale@epa.gov

*Mention of software packages platforms is not an endorsement.