ACToR / CPCat

Kathie Dionisio[1], Richard Judson [2]
[1]U.S. EPA, National Exposure Research Laboratory
[2]U.S. EPA, National Center for Computational Toxicology
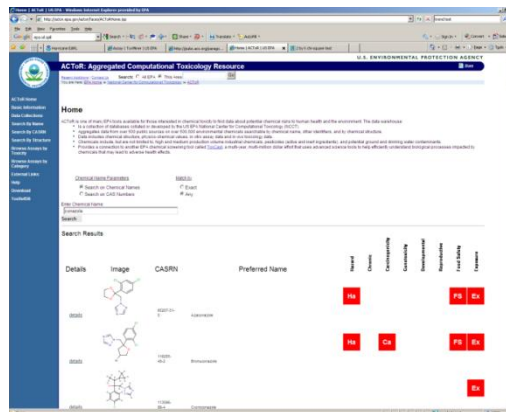*Office of Research and Development*

UNITED STATES ENVIRONMENTAL PROTECTION AGENCY

COMPUTATIONAL TOXICOLOGY
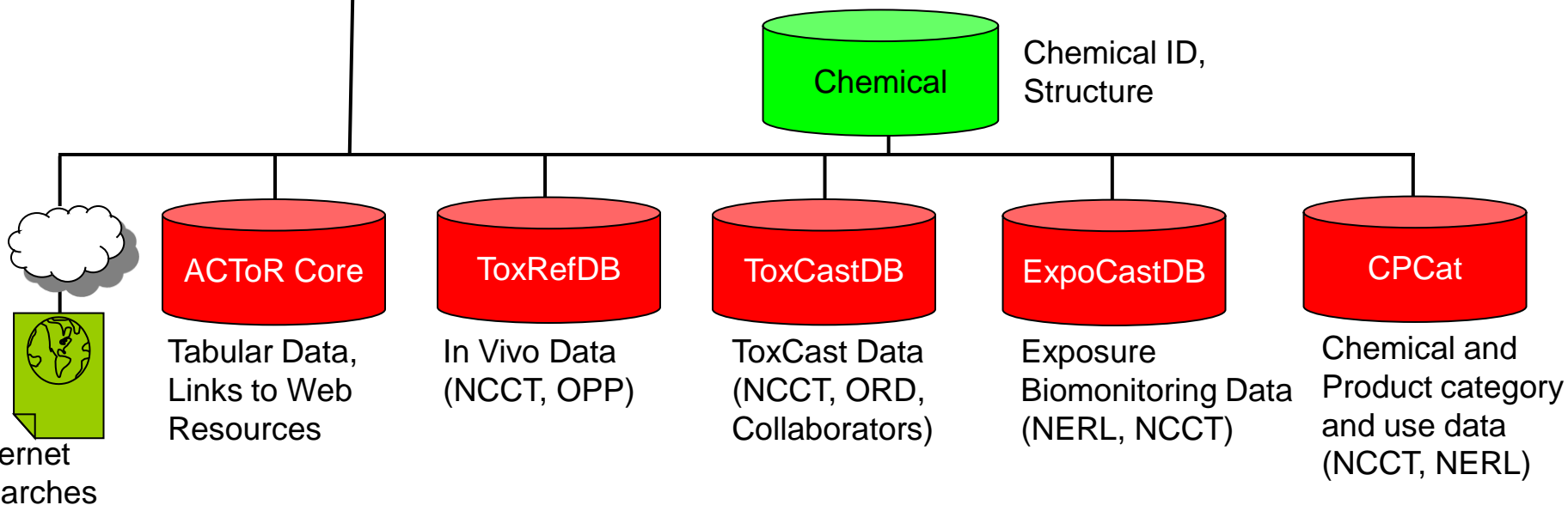
ExpoDat November 2013

- ACToR (Aggregated Computational Toxicology Resource)
- What is the unique set of chemicals EPA is most concerned with?
  - Targets for the overall ToxCast Program
- How much is known about these chemicals?
- Where are the data gaps?
- Collaboration across EPA
  - Office of Research and Development (ORD)
  - Office of Pesticide Programs (OPP)
  - Office of Pollution Prevention and Toxics (OPPT)
  - Office of Water (OW)
  - Great Lakes National Program Office (GLNPO)
  - Endocrine Disruptor Screening Program (EDSP)
- Running this study required building a database

Office of Research and Development

ACToR
Aggregated Computational Toxicology Resource

http://actor.epa.gov/

ACToR API

Chemical — Chemical ID, Structure

Internet Searches

ACToR Core
Tabular Data, Links to Web Resources

ToxRefDB
In Vivo Data (NCCT, OPP)

ToxCastDB
ToxCast Data (NCCT, ORD, Collaborators)

ExpoCastDB
Exposure Biomonitoring Data (NERL, NCCT)

CPCat
Chemical and Product category and use data (NCCT, NERL)

ACToR Goals

- Compile all publicly available information on environmental chemicals
  - One-stop shop
  - Make previously hard-to-access data transparent (e.g. from OPP)
- Make data available for downloading, data mining
  - Entire DB can be downloaded and installed locally
- Make it easy to see data gaps
  - Provides resource for EPA testing programs
- Encourage wide-spread usage
  - over 10,000 regular users

ACToR Data Comes from Many Sources

- EPA (OPP, OPPT, National Center for Environmental Assessment (NCEA), National Exposure Research Laboratory (NERL))
- FDA, NIH, CDC, OSHA, USDA
- States and other countries
- Universities
- NGOs
- Companies

| Category	Count
Data Collections	1,359
Generic Chemicals	551,763
Generic Chemicals with Structure	437,732
Assays	3,180
Assay Results	45,830,875

- Substance
  - A chemical from one source
  - Name(s), CAS RN
  - Source-specific unique ID
  - Assay data
- Compound
  - Chemical structure from one source
  - Source-specific unique ID
- Generic chemical
  - CAS RN
  - Link to many substances (each with same CAS RN)
  - Link to one or more structures / compounds
  - Links to all assay data from substances with same CAS RN

- Assay
  - A collection of data on one or more substances
  - Comes from one data source
  - Can have several types of data included
  - Looks like an Excel spreadsheet

- Assay component
  - One column of an assay table

- Assay result
  - A data value for one substance and one assay component

ACToR is Widely Used

Unique users per month
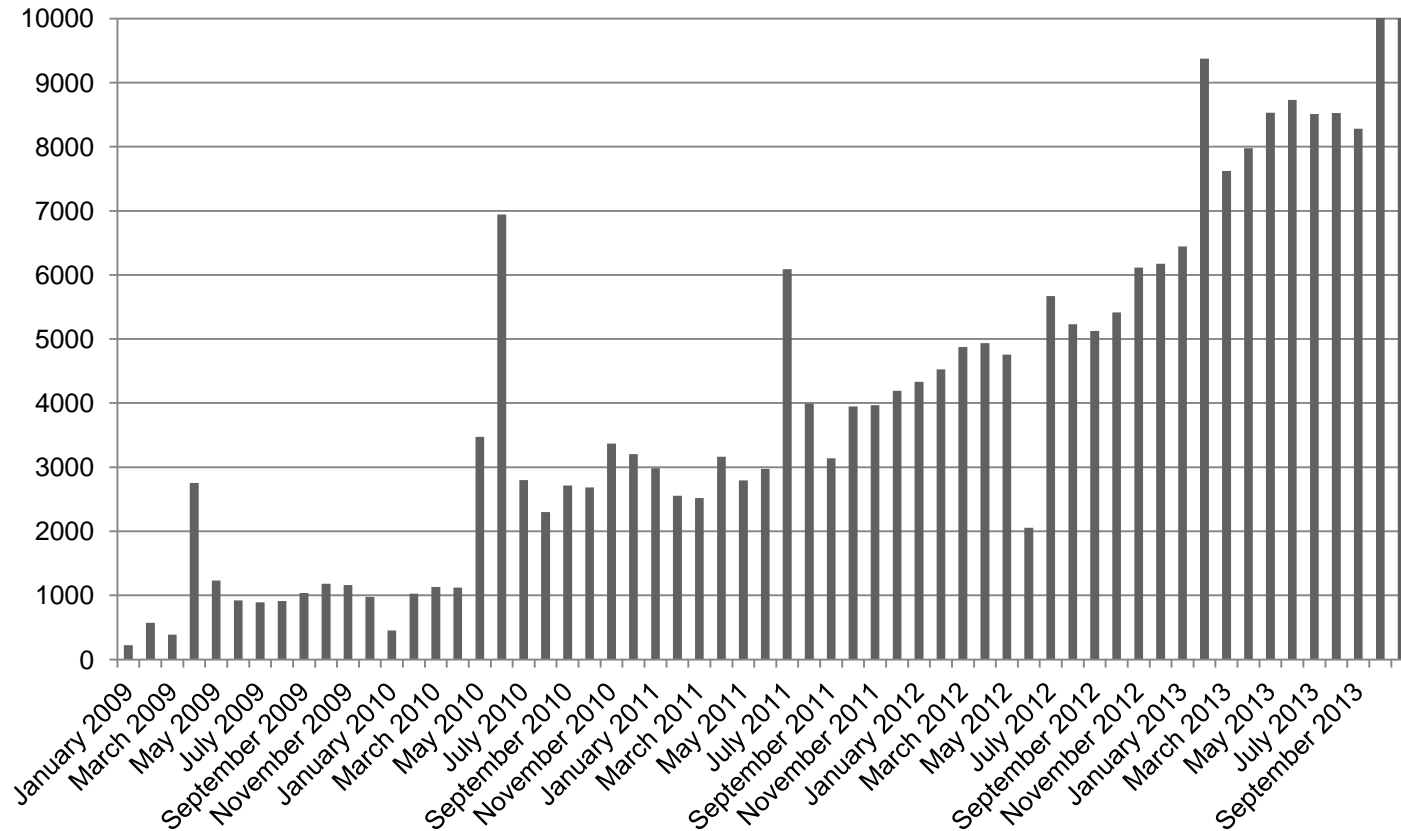
Now seeing 10,000 unique users per month

CPCat – Chemical and Product Category Database

- There is a need to catalog how chemicals are used
- One major input to exposure modeling
- Information exists but was widely dispersed
- Use ACToR to help bring this data together
- Create CPCat database
  - Chemical Use Categories
  - Product Use Categories
  - Mapping from Chemicals to Products
- ACToR UseDB one part of CPCat
- URL:
  - http://actor.epa.gov/cpcat

Data Sources – 41,866 total chemicals

Category Source	Type	Source Categories	CPCat Categories	Chemicals
ACToR Assays and Lists	Use Categories	137	102	35732
ACToR UseDB	**Use Categories**	**16**	**15**	**31622**
DfE	Use Categories	12	9	444
Dow	Functional Categories	20	18	104
DrugBank	Therapeutic Categories	586	460	1753
IUR 2006	Use Categories	21	20	1152
KemI	Functional Categories	62	19	876
NICNAS	Use Categories	19	17	152
Retail Product Categories	Product Categories	450	74	2778
SPIN detpcat	Use Categories	1210	153	6488
SPIN IndustrialSector	Industrial Sector Categories	1934	50	4602
SPIN NACE	Industrial Sector Categories	138	20	7731
SPIN UC62	Use Categories	63	40	9024
Toxome	Functional Categories	17	15	435

ACToR UseDB – source of categories currently released

- Used automated approach to mine ACToR

- Assign chemicals to a small number of high-level chemical use categories

- All chemicals in ACToR checked against the assays extracted from ACToR, and assigned to one or more of 15 UseDB use categories

  - WOE: the number of assays per use category which included a given chemical (the number of 'hits') was output

  - any chemicals with fewer than three hits for a given use category were manually checked to determine whether or not that use category was appropriately assigned.

CPCat term assignment

- Each data source has its own category hierarchy
- These are harmonized in CPCat by mapping categories from data sources to a common set of "terms" (keywords)
- Each original source category may be associated with several CPCat terms
- CPCat uses 281 terms, excluding drugs (738 total)

CPCat term notation

- Comma (",") indicates a group of CPCat terms
  - baby, toy

- Underscore ("_") indicates a single CPCat term
  - arts_crafts

- Suffix "*_ActorUseDB" indicates the CPCat term comes from one of 15 broad ACToR UseDB use categories
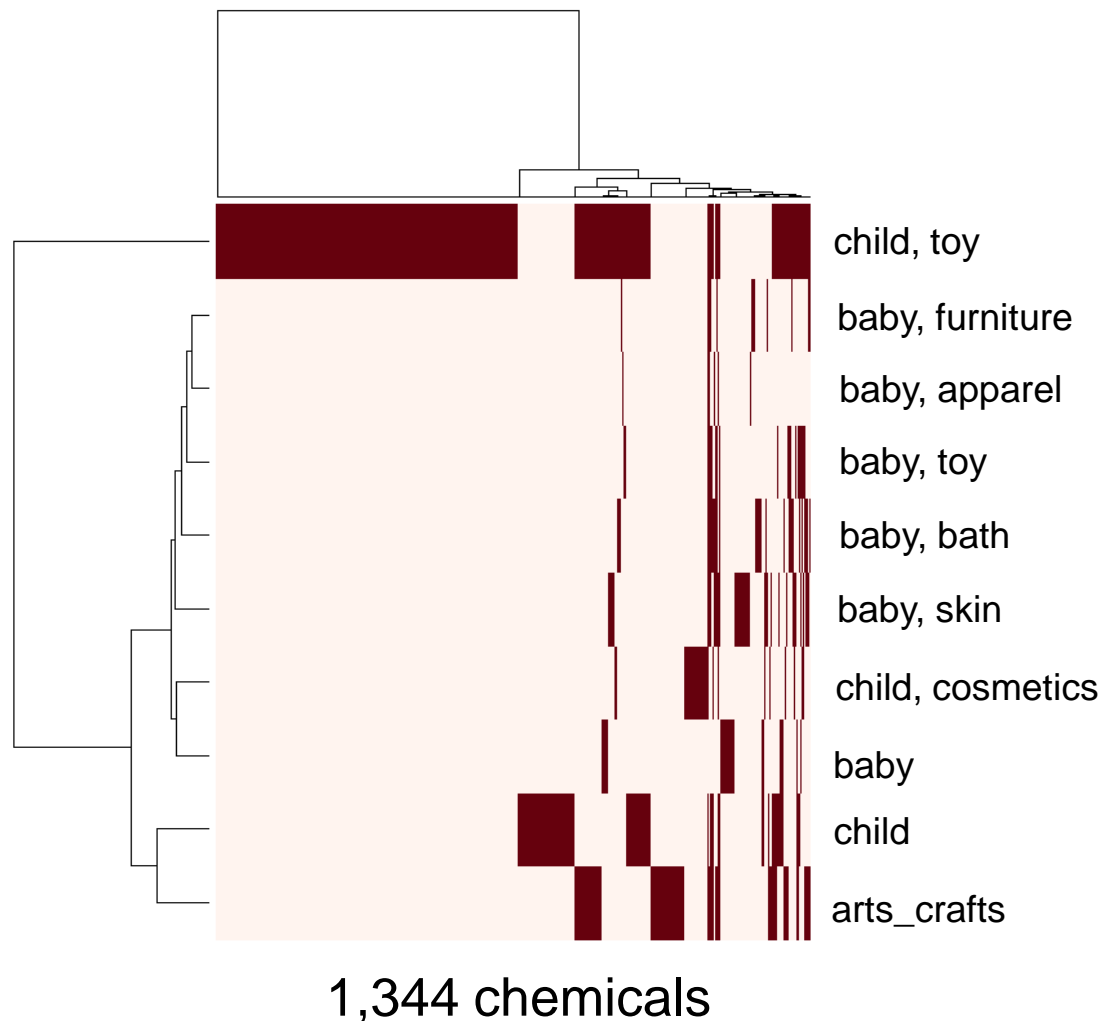 - food_additive_ActorUseDB

Example screen shot: CPCat

Example screen shot

CPCat terms associated with ethyl paraben

agricultural
arts_crafts
automobile, care_product
automobile, maintenance
baby, bath
baby, furniture
baby, skin
battery
child
child, toy
cleaning_washing
cleaning_washing, consumer
cleaning_washing, industrial
cleaning_washing, manufacturing
construction_activity
consumer_product, fragrance
consumer_product_ActorUseDB
cosmetics
cosmetics, eye
cosmetics, fragrance
cosmetics, hair
cosmetics, hair, cleaning_washing
cosmetics, hair, colorant
cosmetics, lips
cosmetics, nails
cosmetics, preservatives
cosmetics, raw_material

cosmetics, skin
cosmetics, skin, cleaning_washing
cosmetics, skin, colorant
cosmetics, uv_blocker
drug
drug, inactive_ingredient
drug, veterinary
electronic_component
food, flavor
food_additive
food_additive_ActorUseDB
food_contact, disinfectant
fragrance, manufacturing
industrial, manufacturing
industrial_manufacturing_ActorUseDB
metal, manufacturing
paint
paraben
personal_care, cleaning_washing
personal_care, sexual_wellness
personal_care, sunscreen
personal_care_ActorUseDB
pesticide
pesticide, biocide
pesticide, inert_ingredient
pesticide, preservative
pesticide_ActorUseDB, inert
photographic
preservative
solvent
surface_treatment

child, toy

baby, furniture

baby, apparel

baby, toy

baby, bath

baby, skin

child, cosmetics

baby

child

arts_crafts

1,344 chemicals

Example:

children's exposure

■ Chemical is associated with the CPCat term(s)

□ Chemical is not associated with the CPCat term(s)

Relevant CPCat terms

Potential Use in EDSP: Exposure Routes for SDWA Chemicals

5,251 SDWA chemicals in EDSP

$\downarrow$

4,189 in CPCat

$\downarrow$

3,373 associated with at least one consumer-use related CPCat term

(92 consumer-use related CPCat terms total associated with these chemicals)

EDSP / SDWA chemicals with most consumer-related terms

CASRN	Name	CPCat term hits
57-55-6	1,2-Propanediol	63
64-17-5	ethanol	62
56-81-5	glycerol	60
7647-14-5	sodium chloride	58
67-63-0	Isopropyl alcohol	53
99-76-3	methyl 4-hydroxybenzoate	52
102-71-6	2,2,2-nitrilotriethanol	50
77-92-9	citric acid	49
1310-73-2	sodium hydroxide	49
50-00-0	formaldehyde	48
94-13-3	propyl 4-hydroxybenzoate	48
13463-67-7	titanium dioxide	48
128-37-0	2,6-di-tert-butyl-p-cresol	45
2682-20-4	2-methyl-3(2H)-isothiazolone	45
7631-86-9	silicon dioxide	44
122-99-6	2-phenoxyethanol	42
14807-96-6	Talc (Mg3H2(SiO3)4)	41
26172-55-4	5-chloro-2-methyl-3(2H)-isothiazolone	41

CASRN	Name	CPCat term hits
100-51-6	benzyl alcohol	40
67-64-1	acetone	39
471-34-1	calcium carbonate	39
1314-13-2	zinc oxide	39
57-11-4	stearic acid	38
64-02-8	EDTA, tetrasodium	37
75-28-5	isobutane	37
106-97-8	butane	37
141-78-6	ethyl acetate	37
3844-45-9	Acid blue 9	37
111-76-2	2-butoxyethanol	36
144-55-8	sodium bicarbonate	36
64-19-7	acetic acid	35
111-90-0	2-(2-ethoxyethoxy)ethanol	35
532-32-1	sodium benzoate	35
36653-82-4	cetyl alcohol	35

Recommendations

- Recommendations to refine CPCat database and/or ACToR UseDB use categories?

- Questions?

Acknowledgments

Thanks to Alicia Frame (former EPA NCCT trainee, currently at Dow Agrochemicals) for the work she contributed to this project.