



Exploring consumer exposure pathways and patterns of use for chemicals in the environment



Kathie L. Dionisio^a, Alicia M. Frame^{b,1}, Michael-Rock Goldsmith^{a,2}, John F. Wambaugh^b, Alan Liddell^{c,3}, Tommy Cathey^d, Doris Smith^b, James Vail^b, Alexi S. Ernstoff^e, Peter Fantke^e, Olivier Jolliet^f, Richard S. Judson^{b,*}

^a U.S. Environmental Protection Agency, National Exposure Research Laboratory, 109 T.W. Alexander Drive, MC E205-02, Research Triangle Park, NC 27709, USA

^b U.S. Environmental Protection Agency, National Center for Computational Toxicology, MC B205-01, Research Triangle Park, NC 27709, USA

^c North Carolina State University, Department of Mathematics, Box 8205, Raleigh, NC 27695-8205, USA

^d Lockheed Martin, Research Triangle Park, NC, USA

^e Quantitative Sustainability Assessment, Department of Management Engineering, Technical University of Denmark, Produktionstorvet 424, 2800 Kgs. Lyngby, Denmark

^f Department of Environmental Health Sciences, School of Public Health, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA

ARTICLE INFO

Article history:

Received 1 December 2014
Received in revised form
17 December 2014
Accepted 18 December 2014
Available online 2 January 2015

Keywords:

Chemical exposure
Human exposure
High throughput
Exposure prioritization
Use category

ABSTRACT

Humans are exposed to thousands of chemicals in the workplace, home, and via air, water, food, and soil. A major challenge in estimating chemical exposures is to understand which chemicals are present in these media and microenvironments. Here we describe the Chemical/Product Categories Database (CPCat), a new, publically available (<http://actor.epa.gov/cpcat>) database of information on chemicals mapped to “use categories” describing the usage or function of the chemical. CPCat was created by combining multiple and diverse sources of data on consumer- and industrial-process based chemical uses from regulatory agencies, manufacturers, and retailers in various countries. The database uses a controlled vocabulary of 833 terms and a novel nomenclature to capture and streamline descriptors of chemical use for 43,596 chemicals from the various sources. Examples of potential applications of CPCat are provided, including identifying chemicals to which children may be exposed and to support prioritization of chemicals for

Abbreviations: ACToR, Aggregated Computational Toxicology Resource; AICS, Australian Inventory of Chemical Substances; CAS RN, Chemical Abstracts Service Registry Number; CDR, Chemical Data Reporting Rule; CPCat, Chemical/Product Categories Database; DCPS, Danish Consumer Product Survey; DfE, Design for the Environment; EDSP, Endocrine Disruptor Screening Program; EPA, Environmental Protection Agency; EWG, Environmental Working Group; GRAS, Generally Recognized as Safe; HTP, Human Toxome Project; IUR, Inventory Update Reporting Modifications Rule; MSDS, Material Safety Data Sheets; NICNAS, National Industrial Chemicals Notification and Assessment Scheme; RPC, Retail Product Categories Database; SDWA, Safe Drinking Water Act; SPIN, Substances in Preparation in Nordic Countries; TSCA, Toxic Substances Control Act.

* Corresponding author at: U.S. Environmental Protection Agency, 109 T.W. Alexander Drive (B205-01), Research Triangle Park, NC 27709, USA. Tel.: +1 919 541 3085; fax: +1 919 541 1194.

E-mail addresses: dionisio.kathie@epa.gov (K.L. Dionisio), frame.alicia@epa.gov (A.M. Frame), rocky.goldsmith@gmail.com (M.-R. Goldsmith), wambaugh.john@epa.gov (J.F. Wambaugh), liddell.10@nd.edu (A. Liddell), cathey.tommy@epa.gov (T. Cathey), smith.doris@epa.gov (D. Smith), vail.james@epa.gov (J. Vail), alexer@dtu.dk (A.S. Ernstoff), pefan@dtu.dk (P. Fantke), ojolliet@umich.edu (O. Jolliet), judson.richard@epa.gov (R.S. Judson).

¹ U.S. Environmental Protection Agency, Office of Solid Waste and Emergency Response, Arlington, VA 22202, USA.

² Chemical Computing Group, Inc., 1010 Sherbrooke Street West, Suite 910, Montreal, Quebec H3A 2R7, Canada.

³ University of Notre Dame, Department of Applied and Computational Mathematics and Statistics, 153 Hurley Hall, Notre Dame, IN 46556, USA.

<http://dx.doi.org/10.1016/j.toxrep.2014.12.009>

2214-7500/Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

toxicity screening. CPCat is expected to be a valuable resource for regulators, risk assessors, and exposure scientists to identify potential sources of human exposures and exposure pathways, particularly for use in high-throughput chemical exposure assessment.

Published by Elsevier Ireland Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

As high-throughput hazard screening approaches such as ToxCast and Tox21 [2,7,9,10,14] continue to evolve, there is a need to develop methods to obtain high-throughput exposure estimates so that chemical hazard screening approaches and exposure estimates together can allow for the complete development of high-throughput risk models. A major challenge in estimating the risk of chemical exposures to human health is the lack of consistent information describing how chemicals are used. A limited number of chemicals that are known to have biological targets, and with uses that suggest high exposures such as pharmaceuticals and some food use pesticides, have been well characterized on both the hazard and exposure axes. For the remaining majority of marketed chemicals, there is little publicly available information [3,6]. This information is critical since the presence of a chemical in specific products significantly influences the nature and extent of human exposures. While information on production volume of chemicals is currently available, a large, uniformly organized repository of information on how chemicals are used, product composition, and other properties (e.g. physicochemical form of the chemical within the product) currently does not exist. This paper describes an effort to characterize one component of the high-throughput exposure estimation process: categorizing the usage of chemicals.

To address the deficiency in chemical exposure estimates, previous efforts have utilized relatively simple high-throughput environmental and indoor fate and transport models that have been parameterized using widely available molecular descriptors, such as physicochemical properties, and simple binary descriptors of indoor and consumer use [11,18]. Specifically, researchers from our group have shown that the simple metric of presence or absence of a chemical in consumer products and associated indoor use is an indicator of a chemical being above the limit of detection for biomonitoring [18]. Although useful for specific applications, the uncertainty bounds on these models are relatively large and additional information on product and chemical use would help to refine these models. To fill this gap in knowledge, we present here the Chemical/Product Categories Database (CPCat), the result of a large-scale effort to catalog and consolidate relatively disparate data sources in order to make chemical use information publicly available, and in a form useful for high-throughput exposure modeling. This new database provides critical information for comparing between well-studied and novel chemicals with respect to use – a key factor driving human exposure to these chemicals.

Aggregating publicly available data sources which categorize chemicals using terms describing their usage, and merging these diverse sources into a single data set

with consistent chemical use categories, is the first step toward integrating chemical use data into high-throughput exposure models. We have compiled an extensive list of chemicals and their associated categories of chemical and product use. Unique use category taxonomies from each source are mapped onto a single common set of terms. We provide several examples of the application of the database that identify and enumerate chemical exposure pathways, including: (1) identifying all documented potential uses of a specific chemical; (2) cataloging all chemicals that meet an exposure scenario (e.g., exposure from children's products); and (3) examining the potential uses of the chemicals implicated with a specific adverse outcome pathway (AOP) [1] (e.g. for use in the U.S. Environmental Protection Agency (EPA) Endocrine Disruptor Screening Program (EDSP)). We anticipate that this open-source database will grow as relevant data continues to become available and is integrated into CPCat, and that this resource will be useful in chemical exposure research and to regulatory agencies.

2. Methods

Here we describe the methodology used to construct the relational CPCat database, available for public download, and as an online searchable website, at <http://actor.epa.gov/cpcat>. Our approach to developing CPCat involved collecting a variety of publicly available data on chemicals and associated categorical (use-categorization) groupings, annotating and curating these data, and harmonizing these categories into a single set of terms. CPCat integrates information from major national and international sources to provide categorical groupings for 43,596 unique chemicals.

2.1. Classes of chemical use categories

Chemical use categories as defined by the data sources can be grouped into 5 general classes. When a chemical has a variety of documented uses and functions, it may be associated with multiple classes, and/or multiple categories within each class (Table 1).

2.2. Data sources

Multiple data sources, including information provided by companies, trade associations, and regulatory agencies, were used to construct the CPCat database. Table 2 details the class of chemical use category (as provided by each source), the number of specific categories (provided by the source), and the number of chemicals associated with each source.

Table 1
Classes of chemical use categories.

Class	Definition
General-use	General categories for chemicals which do not fall into any of the more specific classes of chemical use categories defined below (e.g., lipstick)
Product-use	Categories taken from classifications used for retail products (e.g., children's toys)
Therapeutic-use	The chemical is used as an ingredient in a pharmaceutical, with categories defined by the type of ailment being treated (e.g., anti-acne)
Functional-use	Categories defined by the chemical's properties, which determine the chemical's use; does not specify the type of product in which the chemical is performing the function (e.g., a solvent)
Industrial sector-use	The chemical is used in an industrial sector, with categories defined by the type of industry (e.g., mining)

2.2.1. Aggregated Computational Toxicology Resource (ACToR) data sets and lists

The U.S. EPA's ACToR database is a compilation of publicly available data on chemical toxicity for more than 550,000 unique chemicals (<http://actor.epa.gov>) [5,6,8]. ACToR includes, but is not limited to, high and medium production volume industrial chemicals, pesticides (active and inert ingredients), and potential ground and drinking water contaminants. The ACToR database is organized around chemicals, data sets, and lists, where an ACToR data set refers to data linking chemicals to physicochemical properties, bioactivity, and hazard measurements and an ACToR list refers to chemicals meeting a given criteria. ACToR includes many sources which were subsequently included in CPCat, through both ACToR data sets and lists. Note the Danish Consumer Product Survey (DCPS; http://www.mst.dk/English/Chemicals/consumers_consumer_products/danish_surveys_consumer_products/) is included within this source. The DCPS analyzes consumer products with laboratory testing to determine if they may pose a threat by releasing chemicals to the air, or when in contact with the human body. The DCPS includes information on which chemicals were detected in experimental tests, and which chemicals were analyzed for but were not detected.

Table 2
Summary of data sources used to construct the CPCat database.

Original data source ^a	Class of categories ^b	Original categories ^c	CPCat cassettes ^d	Chemicals
ACToR data sets and lists	General-use	131	173	35,838 ^e
ACToR UseDB	General-use	15	15	31,622
CDR 2012:				
Consumer	General-use	34	36	3321
Industrial function	Functional-use	34	27	5023
Industrial sector	Industrial sector-use	42	43	5226
DfE	Functional-use	11	9	444
Dow	Functional-use	19	18	104
DrugBank	Therapeutic-use	582	460	1754
2006 IUR	General-use	19	24	1152
Keml	Functional-use	61	31	876
NICNAS	General-use	17	17	177
Retail Product Categories	Product-use	359	191	2778
SPIN:				
detpcat	General-use	781	284	6491
Industrial sector	Industrial sector-use	580	221	4603
NACE	Industrial sector-use	57	52	7745
UC62	General-use	61	59	9059
Toxome	Functional-use	16	16	442

^a Source names listed match source names used in the downloadable CPCat database.

^b Class of category used for chemical categorization in the original data source.

^c Number of unique chemical categories in the original data source.

^d The term "CPCat cassette" is defined below in Section 2.3.

^e Note that >550,000 chemicals are included in ACToR, but only ~36,000 could be mapped to one or more use categories.

2.2.2. ACToR UseDB

The ACToR UseDB is a database of chemicals assigned to a small number of broad chemical-use categories. The UseDB was created by the authors based on information extracted from the ACToR database. See supplemental text for a detailed description.

2.2.3. Design for the Environment (DfE)

The DfE program of the U.S. EPA (www.epa.gov/dfe/) evaluates human health and environmental concerns for chemicals used in a range of industries. The program partners with various groups in order to identify safer products and ways to reduce the use of chemicals of concern. The DfE's Safer Chemical Ingredients List categorizes chemicals by functional-use (e.g., colorants, fragrances, solvents, etc.).

2.2.4. Dow

The Dow Chemical Company has published functional-use categorizations for many of the chemicals they manufacture, which are primarily used in the industrial sector (<http://www.dow.com/productsafety/assess/finder.htm>).

2.2.5. DrugBank

DrugBank is a database of pharmaceutical ingredients compiled by the University of Alberta, Canada, which

categorizes chemicals by therapeutic-use (<http://www.drugbank.ca/>) [19–21].

2.2.6. U.S. EPA 2006 Inventory Update Reporting (IUR) Modifications Rule and the 2012 Chemical Data Reporting (CDR) Rule

The U.S. EPA IUR rule (now known as the CDR rule) allows the U.S. EPA to collect and publish information on the manufacturing, processing, and use of commercial substances and mixtures on the Toxic Substances Control Act (TSCA) Chemical Substance Inventory (<http://cfpub.epa.gov/iursearch/>). Data from both the 2006 IUR and the 2012 CDR are included here, covering primarily industrial chemicals and their corresponding use categories. Note the 2012 CDR includes three distinct data sources which categorize chemicals by general-use (for consumer products), and by functional- and industrial sector-use (for industrial chemicals).

2.2.7. Swedish Chemicals Agency (KemI)

The Swedish KemI is a government agency responsible for ensuring the safe use of chemicals, and maintains a product registration list and variety of databases for pesticides and other chemicals. This organization has published a list of chemicals categorized by functional-use (<http://www.kemi.se/en/>).

2.2.8. National Industrial Chemicals Notification and Assessment Scheme (NICNAS)

NICNAS (<http://www.nicnas.gov.au>) maintains the Australian Inventory of Chemical Substances (AICS) list, a listing of industrial chemicals in use in Australia since January 1, 1977. The list categorizes chemicals by general-use, with a small number of categories.

2.2.9. Retail Product Categories (RPC) database

Goldsmith et al. developed a database of chemical information extracted from publicly available Material Safety Data Sheets (MSDS) for products sold at Walmart [4]. In addition to extracting quantitative information on chemical composition of products from the MSDS, products and their ingredients were mapped to a hierarchy of product-use categories.

2.2.10. Substances in Preparation in Nordic Countries (SPIN) database

SPIN is a joint project of government environmental agencies in Norway, Sweden, Denmark, and Finland, and is comprised of data from the Product Registries of each of these countries [13]. Four separate SPIN databases which categorize chemicals in different ways are used in constructing CPCat: old Danish and Norwegian categories (detpcat), use/function categories for chemical substances and preparations (UC62), the Statistical classification of economic activities in the European Community (NACE), and industrial-use information (Industrial Sector). The first two databases categorize chemicals by general-use, the latter two categorize chemicals by industrial sector-use.

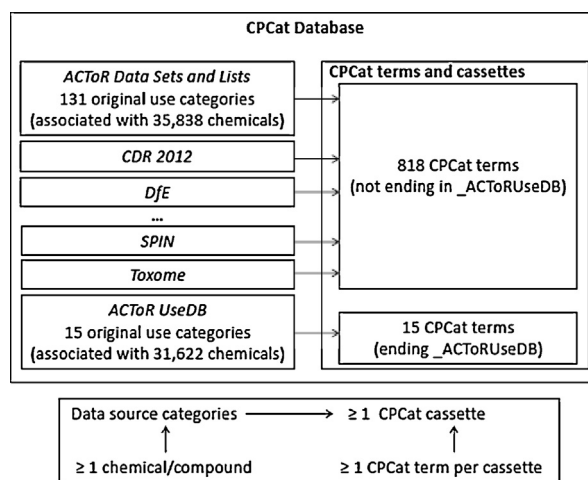


Fig. 1. CPCat database organization.

2.2.11. Human Toxome Project (HTP)

The Environmental Working Group (EWG) HTP collects biomarker data to help understand the scope of population-level exposure to industrial chemicals that enter the body through pollution or as ingredients in consumer products (<http://www.ewg.org/sites/humantoxome/>). Data from the HTP includes a small number of categories of functional-use which have an elevated toxicity risk.

2.3. Assigning CPCat terms and cassettes

The CPCat database consists of each of the chemicals for which one or more sources reported use data, and an associated set of CPCat terms describing usage. The terms are organized using a well-defined nomenclature to create 'cassettes.' Each of the data sources used to construct the CPCat database employed a unique set of chemical use categories (each falling into one of the five chemical classes described above) to meet a particular need. These tend to focus on one or a few types of uses or functional categories, or on particular classes of chemicals. No single categorization scheme included all of the categories covered in the global collection. To create CPCat, we manually mapped the chemical use categories and descriptions provided by each data source to CPCat terms and cassettes (Fig. 1). Mining the use category descriptions provided within each of the original data sources results in 2681 unique original source chemical use categories (noting that the same description/category can be used by more than one source), which were mapped to 833 unique CPCat terms (Fig. 1).

Cassettes are comprised of one or more CPCat terms, separated by spaces; all CPCat terms within a cassette must be interpreted together to reflect the categorical information provided by the original data source. Because of the broad nature of the 15 original ACToR UseDB categories, these 15 categories were mapped directly to 15 corresponding CPCat terms (indicated by the suffix "_ACToRUseDB"); no categories from other sources were mapped to these "_ACToRUseDB" CPCat terms.

The full set of CPCat terms were selected by aggregating all categories provided by each data source, taking care to eliminate synonymous category names (e.g., drug and pharmaceutical), mistakes (e.g., spelling errors), and other redundancies or superfluous information. No attempt was made to extrapolate or fill in missing data on chemicals. Rather, CPCat incorporates only existing information on use categories for chemicals from each data source. An underscore between two words indicates a compound word (e.g., automotive_care, building_material) and should be considered the same as a single unique CPCat term. Any combination of CPCat terms can be combined to create a CPCat cassette; however there are some combinations of terms that are common, and others which never occur.

2.4. Interpreting CPCat terms and cassettes

A data dictionary including a list of all unique CPCat cassettes, and describing each unique CPCat term, is included with the release of the database at <http://actor.epa.gov/cpcat>. While a specific hierarchy was not defined, CPCat terms refer to different levels of detail, due to the varying levels of information available from each source regarding the usage of the chemical. When a source included specific information on the chemical usage, for example a specific type of beauty product such as lipstick, that information is reflected in the assigned CPCat cassette so that information is not lost. If more than one CPCat cassette was mapped to a single source category (separated by a comma), this indicates that the source reported more than one distinct usage for the chemical within one original category entry. In this situation, each cassette should be interpreted separately to reflect these multiple uses for the chemical.

Examples of CPCat cassettes include: (a) building_material, (b) manufacturing building_material wood, (c) building_material wood, and (d) furniture wood. Where (a) describes a chemical with a general use in building materials, but with no further information given in the original data source; (b) describes a chemical used when manufacturing wooden building materials; (c) describes a chemical contained in wooden building materials; and (d) describes a chemical used in wooden furniture. When a CPCat cassette is comprised of more than one term, the terms refer to increasing levels of specificity when reading from left to right. As an example of when multiple CPCat cassettes might be assigned to a single original data source category, if the original data source category described wooden furniture and housing materials, then this entry would have been assigned both (c) and (d) cassettes in order to reflect the multiple uses specified by the original data source entry.

Some data sources determined chemical content of a product through laboratory testing, rather than from listed ingredients. The CPCat term for this is 'detected.' Thus a chemical may appear in a use category due to unintentional inclusion of that chemical in a product (e.g., because of contamination). Any source which indicated chemicals were detected through laboratory testing (including all DCPS sources) include "detected" as a CPCat term within the

associated cassette(s). Note that the quantitative data from the laboratory testing is currently not included in CPCat, rather if the presence of a chemical is detected in laboratory testing, the information is included as such in CPCat. The "child_use" and "baby_use" CPCat terms are similarly unique in that they reference the class of consumer for which the product is intended. These terms were included due to the general interest in exposure of these demographics, and due to the number of products specifically marketed to these demographics.

CPCat terms associated with the 15 broad UseDB categories (Supplemental Table 1) are unique within CPCat. Because the 15 UseDB categories are quite broad, it was desired to distinguish these category assignments from the remainder of categorical assignments within CPCat. Then, if a user only wanted to analyze the 15 broadly defined UseDB categories and their associated chemicals, these could easily be extracted. Or, if a user wanted to exclude these broadly defined categories from their search, this could be done. The CPCat terms associated with the 15 UseDB categories include the suffix ".ACToRUseDB" to alert the user to these unique CPCat terms that indicate a potentially broad categorization of the chemical.

2.5. Data management and database availability

To aid data processing, chemical category taxonomies from each source were translated into a common format before entry into the CPCat database. For each chemical listing in the CPCat database, in addition to the assigned CPCat cassette(s), links to the underlying data source(s) and original taxonomy categories are maintained. In the database, each category is labeled by an alphanumeric ID, and a description. The top level of each source or taxonomy is always given the ID "Source_0000." When sources used an explicit ID for each category, they have been maintained in CPCat. This information is not included in the web interface.

The CPCat database is available in three formats. A.zip file containing a set of .txt and Microsoft Excel files is available for download (<http://actor.epa.gov/cpcat>), which includes R code for running the examples presented below. Alternatively, a MySQL database for download, and a searchable online version of the CPCat database, are available at the same location.

3. Results

3.1. Summary statistics

A total of 43,596 unique chemicals from the U.S. EPA's ACToR database mapped to at least one CPCat cassette. There are 1297 unique CPCat cassettes, including 473 related to drug uses and 824 related to other use categories. The cassettes are permutations of 833 unique CPCat terms, including 456 drug-related terms. Table 2 summarizes the sources with number of original categories, CPCat terms, and chemicals. See <http://actor.epa.gov/cpcat> for a list of all chemicals included in CPCat and the data dictionary for a list of all CPCat terms and cassettes.

Table 3
CPCat cassettes associated with ethylparaben^a

CPCat cassettes		
agricultural*	hunting	personal_care cosmetics*
arts_crafts*	industrial cleaning_washing	personal_care sanitizer hand
automotive_care	industrial_manufacturing_ACToRUseDB	personal_care sexual_wellness gel detected
child_use	inert_ACToRUseDB	personal_care shower gel
cleaning_washing*	manufacturing chemical	personal_care soap*
construction	manufacturing cleaning_washing polish	personal_care sunscreen*
consumer_use_ACToRUseDB	manufacturing detergent	personal_care wash*
detergent	manufacturing drug	personal_care_ACToRUseDB
drug*	manufacturing export	pesticide*
electronics batteries*	manufacturing metals	photographic
facility salon detected	manufacturing personal_care*	preservatives
fluid_property_modulator	manufacturing soap	raw_material personal_care cosmetics
food_additive*	paint	sports_equipment
food_additive_ACToRUseDB	paraben	surface_treatment
food_contact	personal_care	tools personal_care hair
fragrance consumer_use	personal_care bath	toys*

^a A * indicates multiple cassettes containing additional CPCat terms; see Supplemental Material for the full list.

3.2. Example 1: CPCat cassettes associated with a single chemical

The CPCat database can be queried to produce a list of all CPCat terms and cassettes associated with a single chemical. As an example, ethylparaben (Chemical Abstracts Service (CAS) Registry Number (RN)= 120-47-8) is associated with a diverse group of CPCat cassettes (Table 3), most of which are consistent with the use of ethylparaben as a preservative in a variety of cosmetics, soaps, and shampoos. Different cassettes reflect the varying levels of detail present in the original sources categories, which may be important to understand exposure (e.g., “personal_care” vs. “personal_care cosmetics bath baby_use”). Users must also be aware of the “detected” term that may be contained within a CPCat cassette (e.g., “personal_care sexual_wellness gel detected”). This “detected” term indicates the chemical was detected in laboratory tests of the product. Thus, the association of a chemical with a specific product in the database can occur because it is a known ingredient, or because it was detected in laboratory measurements.

3.3. Example 2: Child exposure scenario

The CPCat database may be queried to identify all chemicals with reported data which fall under a specified exposure scenario. For example, CPCat can be queried for chemicals to which children could be exposed, beyond routine exposures from food, drinking water, dust, and ambient air. To identify such a list of chemicals, we selected CPCat cassettes which include the CPCat terms “baby_use” or “child_use,” excluding cassettes including the CPCat terms “food” or “manufacturing” (Table 4). For simplicity cassettes which linked to less than five chemicals were excluded.

Extracting the chemicals associated with these 16 cassettes results in 1074 chemicals mapped to 35 original categories in the RPC, ACToR Data Sets and Lists, and 2012 CDR Consumer database sources. Of these 1074 chemicals, 649 were associated with the chosen cassettes related to

Table 4
Selected CPCat cassettes for child exposure scenario.

CPCat cassettes
apparel baby_use diaper
arts_crafts child_use detected
baby_use detected
child_use
child_use detected
electronics toys child_use
personal_care cosmetics baby_use
personal_care cosmetics bath baby_use
personal_care cosmetics child_use detected
sports_equipment child_use
toys baby_use
toys child_use
toys child_use detected
toys fragrance child_use detected
toys lawn_garden child_use
toys mouthing baby_use

children’s exposure based on a single source within CPCat, 211 were associated with the chosen cassettes based on two different sources, and 214 chemicals were associated with the chosen cassettes based on three or more sources. Fig. 2 shows a heat map of the CPCat cassettes of interest and associated chemicals for the child scenario. Beyond chemicals associated with the generic “child_use” cassette where no additional descriptors are available, the majority of chemicals in this example are associated with the “toys child_use” CPCat cassette. This indicates that beyond routine exposures from food, drinking water, and ambient air, the largest fraction of chemicals identified were in children’s toys. Further, as seen in the gray highlighted bars in the heat map, 386 chemicals are associated with this child exposure scenario through cassettes that include the “detected” CPCat term, but not through any other cassettes. This indicates that if we were researching exposure to chemicals used in children’s products, we would be missing potential exposure to 386 chemicals that are not listed as product ingredients but nevertheless were detected in toys and other child-specific products. Information on detection of chemicals in laboratory testing comes from the DCPS source within ACToR Data Sets and Lists, as described above.

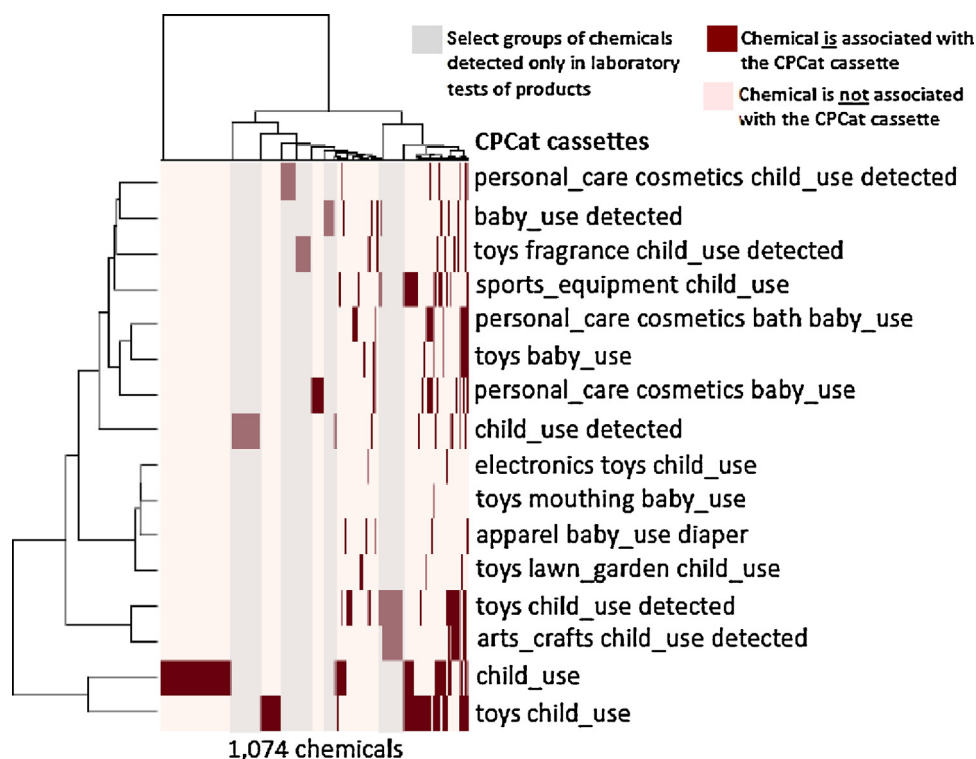


Fig. 2. Heat map of chemicals associated with CPCat cassettes from the child scenario. Individual chemicals are on the x-axis, and CPCat cassettes (i.e. use-category classifications) on the y-axis. There are a total of 1074 chemicals associated with at least one child scenario CPCat cassette.

3.4. Example 3: Potential exposure pathways for chemicals subject to the Endocrine Disruptor Screening Program

CPCat can be queried to identify exposures to chemicals of concern for specific adverse health impacts. For example, the U.S. EPA's Endocrine Disruptor Screening Program (EDSP) is mandated to identify and analyze chemicals for their potential to interact with and disrupt specified endocrine pathways (estrogen, androgen, thyroid and steroidogenesis).

The two main classes of chemicals covered in the EDSP are pesticide ingredients (active and inert) and chemicals with the potential to be found in drinking water. This makes up a chemical universe of approximately 5000 chemicals. In this example, we focus on a set of 5251 Safe Drinking Water Act (SDWA) chemicals that are candidates for exposure and hazard determination under the EDSP [17].

While the CPCat cassettes do not provide any direct, quantitative measure of exposure, they can be used as one input to a prioritization scheme. The first step in exposure prioritization could be to rank the SDWA chemicals by their likely exposure potential, with exposure potential based on the number of consumer-use related CPCat cassettes the chemical is associated with (i.e., the number of consumer-use related "hits"). Theoretically, the more consumer-use related CPCat cassettes that a chemical is associated with would translate to a larger number of potential exposure pathways for an individual [15]. Of all unique CPCat cassettes, 234 were selected as being broadly

related to consumer exposure (including exposures from food; Table 5). These 234 consumer exposure related CPCat cassettes are associated with 19,552 unique chemicals.

Of the 5251 SDWA compounds, CPCat contains data on 4189, and 3514 map to at least one of the consumer-use related CPCat cassettes. Table 6 provides the number of different consumer-use related CPCat cassettes for each of the 22 SDWA chemicals with ≥ 60 hits. These chemicals could be placed higher on the priority list based on exposure potential, while those compounds which are associated with < 5 CPCat cassette hits (2441 compounds) could be given a lower priority for assessment. It is important to again note that while the number of "hits" should not be taken as a quantitative surrogate for exposure measurement, this data can be useful in prioritizing chemicals of interest. A larger number of hits (e.g. ≥ 60 hits versus chemicals with < 5 hits) translates to more confidence in the strength of the evidence that the chemical is included in a variety of consumer-use related products. If we do not have the ability to discriminate between consumer products with high or low exposure dose potential, the presence of the chemical in a large number of products may be a plausible surrogate for an increased probability of exposure. In addition, hits on specific groups of CPCat cassettes could be prioritized based on their exposure potential. For example, if chemicals with fewer hits are included in cassettes with a high exposure potential (e.g., food related CPCat cassettes), those chemicals could be prioritized over chemicals with more hits on cassettes with a lower exposure potential (e.g., cassettes related to apparel).

Table 5
Consumer-use related CPCat cassettes selected for EDSP example^a

CPCat cassettes		
adhesive consumer_use*	drinking_water_contaminant*	lubricant consumer_use*
air_fresheners consumer_use*	electronics*	personal_care*
air_treatment consumer_use	explosives consumer_use	personal_care ACToRUseDB
apparel*	extermination consumer_use	pesticide consumer_use
apparel_care*	fertilizer consumer_use	pet
appliance consumer_use*	flame_retardant	polish apparel_care footwear
arts_crafts*	food*	solvent consumer_use
automotive_care consumer_use	food_additive*	sports_equipment*
automotive_component consumer_use*	food_contact*	stoves consumer_use
baby_use detected*	food_residue*	surface_treatment consumer_use
batteries consumer_use	fragrance consumer_use	tea_coffee
beverage*	fuel automotive	textile consumer_use*
building_material consumer_use*	fuel consumer_use	toilets baby_use
child_use*	fungicide consumer_use	tools consumer_use*
cleaning_washing*	furniture*	tools_lawn_garden
colorant consumer_use detected	heating*	tools_personal_care*
consumer_use	hunting	toys*
consumer_use ACToRUseDB	impregnation consumer_use detected	water_treatment consumer_use
décor*	lawn_garden consumer_use	writing*
drinking_water*	leather consumer_use	

^a A * indicates multiple cassettes containing additional CPCat terms. See Supplemental Material for the full list.

We can further reduce the list of chemicals with a high exposure potential in Table 6 by eliminating chemicals that are common food substances (e.g., ethanol, sodium chloride, citric acid) or are otherwise widely used and considered safe (e.g., talc or other substances on the U.S. FDA's Generally Recognized as Safe (GRAS) list). However, we also see that prioritizing based on the number of consumer-use related CPCat cassette hits does highlight certain phenol

Table 6
Number of consumer-use related CPCat cassettes that EDSP/SDWA chemicals are associated with. Chemicals associated with less than 60 consumer related CPCat cassettes are omitted.

CAS RN	Name	CPCat cassette hits
57-55-6	1,2-Propanediol	121
64-17-5	Ethanol	114
56-81-5	Glycerol	110
67-63-0	Isopropyl alcohol	90
77-92-9	Citric acid	85
99-76-3	Methyl 4-hydroxybenzoate	85
1310-73-2	Sodium hydroxide	84
13463-67-7	Titanium dioxide	82
7647-14-5	Sodium chloride	80
102-71-6	2,2,2-Nitrioltriethanol	78
106-97-8	Butane	74
75-28-5	Isobutane	73
94-13-3	Propyl 4-hydroxybenzoate	72
128-37-0	2,6-Di-tert-butyl-p-cresol	72
3844-45-9	Brilliant Blue FCF ^a	65
122-99-6	Ethylene glycol monophenyl ether	64
1934-21-0	Acid Yellow 23 (Tartrazine) ^b	64
67-64-1	Acetone	63
2682-20-4	2-Methyl-4-isothiazolin-3-one	63
14807-96-6	Talc (Mg ₃ H ₂ (SiO ₃) ₄)	63
100-51-6	Benzyl alcohol	62
57-11-4	Stearic acid	60

^a Benzenemethanaminium, N-ethyl-N-[4-[[4-[ethyl[(3-sulphophenyl)methyl]amino]phenyl](2-sulphophenyl)methylene]-2,5-cyclohexadien-1-ylidene]-3-sulfo-, inner salt, disodium salt.

^b Trisodium 5-hydroxy-1-(4-sulphophenyl)-4-(4-sulphophenylazo)pyrazole-3-carboxylate.

compounds that, in their parent or metabolite form, may interact with the estrogen receptor (e.g., propyl 4-hydroxybenzoate, methyl 4-hydroxybenzoate).

4. Discussion

Here we have detailed the construction of the CPCat database, and provided examples of its utility for understanding potential sources of exposure for chemicals in the environment. CPCat contains use information (general-use, product-use, functional-use, therapeutic-use, industrial sector-use) on over 43,000 chemicals taken from major national and international data sources. Of particular note, we have identified a total of ~20,000 unique chemicals with consumer uses. CPCat provides information that one could use to prioritize further study of these chemicals for exposure potential.

There are a number of limitations of the CPCat database that should be taken into account with any use. First, though data from sources such as DrugBank, RPC, and DCPS were hand curated by their respective sources, as described in Methods, there was limited manual curation of data done by the authors, and detailed information about categorizations taken from the original sources was not always available. Besides ACToR, the largest contributor to CPCat is SPIN, and the origin of the data, including how it was identified and collected, is not always clear. Even with ACToR, we have taken data from a large number of smaller sources, again with limited manual curation. Therefore, it is best to take into account data quality and provenance as appropriate for a particular use, as errors and omissions in the original sources are carried forward in CPCat. However, by including multiple sources of information, one can gain confidence in a general category assignment, especially if the same use category arises from multiple sources.

Another limitation of the CPCat database is that certain category assignments may not equate with bioavailability or potential for exposure. For example, chemicals in CPCat

may be assigned to a fabric dye related cassette. It may be assumed by an investigator that individuals may have dermal exposure to these chemicals through clothing that is in direct contact with their skin; however these chemicals may be tightly bound to the fabric, and thus are likely not bioavailable. Nonetheless, being able to enumerate “all” potential use associations of a chemical has intrinsic value in prioritizing research geared toward elucidating relevant exposure routes, exposure points and exposure pathways from source to receptor.

Lastly, it is important to remember that the CPCat database contains only partial information on the quantities of chemicals in products (namely all information from the RPC Database [4]). As shown in Wambaugh et al., the presence or absence of a chemical in consumer products is often an indicator of detection of the chemical in biomonitoring of humans, however users should recognize that the presence of a chemical mapped to CPCat cassettes is a necessary step in identifying potential exposures, but it is likely insufficient for quantifying exposure.

We envision that a main use of CPCat will be for priority setting tasks, such as in Example 3. The CPCat database can be used to group chemicals by potential types of exposure sources (e.g., by selecting chemicals associated with consumer-use related CPCat cassettes), or by a large number of diverse potential sources (e.g., chemicals associated with a large number of unrelated CPCat cassettes). While CPCat cassettes and terms should not be used as a surrogate for exposure on their own, they can provide information that will aid investigators in identifying chemicals of interest for more detailed analysis. CPCat may also provide intrinsic value in other efforts including systems analysis of key input-output variables in exposure pathways, and ultimately life cycle assessment (LCA) analysis. In the case of LCA, CPCat could greatly assist with identifying inventory flows and processes from the technosphere (man-made world) and systems boundaries used in LCA in a logical, chemical-centric workflow [16].

An interesting potential use of this data comes in exposure modeling. There are existing exposure models that determine population level exposure to a chemical by aggregating doses from years of simulated individual human interactions with a variety of exposure pathways as they navigate the activities of daily life. These simulations require that there is sufficient data to determine what aspects of daily life may lead to exposure to a specific chemical [12]. For instance, one could add into such a model the sets of chemicals in the “child” scenario described above. Although there are a large number of chemicals, many of them are functional equivalents, so a given person would likely be exposed to one in the class, but not all. In modeling and simulation uses such as these, it will be important to define the functionally equivalent chemicals in a scenario, and perhaps run multiple simulations with different selections out of the equivalent sets.

5. Conclusions

In the absence of more detailed quantitative data on product composition, relevant dose from product use, and exposure routes, CPCat represents a major step forward

in characterizing human exposure by making available chemical-to-product use category information. We plan for CPCat to be a continually expanding resource for exposure research. Plans for future work include developing an ontology of exposure and relating the CPCat terms/cassettes to a set of delivery modes (e.g., exposure from a cleaning spray may come through dermal contact with the mixture during cleaning, subsequent ingestion from hand-to-mouth contact, or from inhalation when the mixture is sprayed on a surface) and eventually to exposure models. Other sources of exposure information such as further chemical-to-product mappings (i.e., linkages between retail products and the chemicals contained within the products) can be included, which would enhance the utility of the CPCat database by including quantitative information on the amount of chemicals included in various products. The CPCat database is easily extended, by adding new data, categories or cassettes. Other users could develop and implement their own set of terms or cassettes, which could be integrated into the current CPCat. We believe that this publicly available database will be a valuable resource for regulators, risk assessors and exposure scientists with a need to evaluate the safety of chemicals.

Disclaimer

Although this work was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency policy. EPA does not endorse the purchase of any commercial products or services mentioned in this publication. The authors declare they have no actual or potential competing financial interests.

Transparency document

The [Transparency document](#) associated with this article can be found in the online version.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.toxrep.2014.12.009](https://doi.org/10.1016/j.toxrep.2014.12.009).

References

- [1] G. Ankley, R. Bennett, R. Erickson, D. Hoff, M. Hornung, R. Johnson, et al., Adverse outcome pathways: a conceptual framework to support ecotoxicology research and risk assessment, *Environ. Toxicol. Chem.* 29 (2010) 730–741.
- [2] D.J. Dix, K.A. Houck, M.T. Martin, A.M. Richard, R.W. Setzer, R.J. Kavlock, The ToxCast program for prioritizing toxicity testing of environmental chemicals, *Toxicol. Sci.* 95 (2007) 5–12.
- [3] P.P. Egeghy, R. Judson, S. Gangwal, S. Mosher, D. Smith, J. Vail, et al., The exposure data landscape for manufactured chemicals, *Sci. Total Environ.* 414 (2012) 159–166.
- [4] M.-R. Goldsmith, C.M. Grulke, R.D. Brooks, T.R. Transue, Y.M. Tan, A. Frame, et al., Development of a consumer product ingredient database for chemical exposure screening and prioritization, *Food Chem. Toxicol.* 65 (2014) 269–279.
- [5] R. Judson, A. Richard, D. Dix, K. Houck, F. Elloumi, M. Martin, et al., ACToR – Aggregated Computational Toxicology Resource, *Toxicol. Appl. Pharmacol.* 233 (2008) 7–13.

- [6] R. Judson, A. Richard, D.J. Dix, K. Houck, M. Martin, R. Kavlock, et al., The toxicity data landscape for environmental chemicals, *Environ. Health Perspect.* 117 (2009) 685–695.
- [7] R.S. Judson, K.A. Houck, R.J. Kavlock, T.B. Knudsen, M.T. Martin, H.M. Mortensen, et al., In vitro screening of environmental chemicals for targeted testing prioritization: the toxcast project, *Environ. Health Perspect.* 118 (2010) 485–492.
- [8] R.S. Judson, M.T. Martin, P. Egeghy, S. Gangwal, D.M. Reif, P. Kothiya, et al., Aggregating data for computational toxicology applications: the U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) system, *Int. J. Mol. Sci.* 13 (2012) 1805–1831.
- [9] R. Kavlock, D. Dix, Computational toxicology as implemented by the U.S. EPA: providing high throughput decision support tools for screening and assessing chemical exposure, hazard and risk, *J. Toxicol. Environ. Health Part B: Crit. Rev.* 13 (2010) 197–217.
- [10] R. Kavlock, K. Chandler, K. Houck, S. Hunter, R. Judson, N. Kleinstreuer, et al., Update on EPA's toxcast program: providing high throughput decision support tools for chemical risk management, *Chem. Res. Toxicol.* 25 (2012) 1287–1302.
- [11] J. Mitchell, J.A. Arnot, O. Jolliet, P.G. Georgopoulos, S. Isukapalli, S. Dasgupta, et al., Comparison of modeling approaches to prioritize chemicals based on estimates of exposure and exposure potential, *Sci. Total Environ.* 458–460 (2013) 555–567.
- [12] P.S. Price, C.F. Chaisson, A conceptual framework for modeling aggregate and cumulative exposures to chemicals, *J. Expo. Anal. Environ. Epidemiol.* 15 (2005) 473–481.
- [13] SPIN, Spin substances in preparations in nordic countries, 2013, Available: <http://www.spin2000.net> (accessed 9.07.13).
- [14] R.R. Tice, C.P. Austin, R.J. Kavlock, J.R. Bucher, Improving the human hazard characterization of chemicals: a Tox21 update, *Environ. Health Perspect.* 121 (2013) 756–765.
- [15] U.S. Department of Health and Human Services, Public Health Assessment: Guidance Manual (update), U.S. Department of Health and Human Services, Public Health Service, Agency for Toxic Substances and Disease Registry, Atlanta, Georgia, 2005.
- [16] U.S. EPA, Life Cycle Assessment: Principles and Practice, National Risk Management Research Laboratory, Office of Research and Development, U.S. Environmental Protection Agency, Cincinnati, Ohio, 2006.
- [17] U.S. EPA, U.S. Environmental Protection Agency Endocrine Disruptor Screening Program: Universe of Chemicals and General Validation Principles, U.S. EPA, Washington, DC, 2012.
- [18] J.F. Wambaugh, R.W. Setzer, D.M. Reif, S. Gangwal, J. Mitchell-Blackwood, J.A. Arnot, et al., High-throughput models for exposure-based chemical prioritization in the expocast project, *Environ. Sci. Technol.* 47 (2013) 8479–8488.
- [19] D.S. Wishart, C. Knox, A.C. Guo, S. Shrivastava, M. Hassanali, P. Stothard, et al., Drugbank: a comprehensive resource for *in silico* drug discovery and exploration, *Nucl. Acids Res.* 34 (2006) D668–D672.
- [20] D.S. Wishart, Drugbank and its relevance to pharmacogenomics, *Pharmacogenomics* 9 (2008), 1166–1162.
- [21] D.S. Wishart, C. Knox, A.C. guo, D. Cheng, S. Shrivastava, D. Tzur, et al., Drugbank: a knowledgebase for drugs, drug actions and drug targets, *Nucl. Acids Res.* 36 (2008) D901–D906.