TITLE: Development of a Consumer Product Ingredient Database for Chemical Exposure Screening and Prioritization

AUTHORS:

M.-R. Goldsmith, [1*] C. Grulke, [1] R. D. Brooks, [2] T. Transue, [3] Y. M. Tan, [1] A. Frame,[1,5] P.P. Egeghy, [1] R. Edwards, [4] D. T. Chang[1], R. Tornero-Velez,[1] K. Isaacs,[1] A. Wang,[1,5] J. Johnson,[1] K. Holm,[1] M. Reich,[6] J. Mitchell,[7] D. Vallero,[1] L. Phillips,[1] M. Phillips,[1] J. F. Wambaugh, [1] R. S. Judson, [1] T. Buckley, [1] C. C. Dary[1]

AFFILIATIONS:

1. United States Environmental Protection Agency (U.S. EPA), Office of Research & Development, RTP, NC 27711
2. Student Services Contractor at U.S. EPA, RTP NC
3. Lockheed-Martin Information Technology, RTP, NC 27711
4. North Carolina State University, 2200 Hillsborough St, Raleigh, NC 27695 and rdedward@ncsu.edu
5. Oak Ridge Institute for Science and Education Fellow
6. Fuquay-Varina High School, Fuquay-Varina, NC 27526 and madeline.reich@gmail.com
7. Biosystems and Agricultural Engineering, Michigan State University, E. Lansing, MI 48824

*** Correspondence addressed to**: goldsmith.rocky@epa.gov, 919-541-0497

Disclaimer: The United States Environmental Protection Agency has provided administrative review and has approved this manuscript for publication. The views expressed in this manuscript are those of the authors and do not necessarily reflect the views or policies of the Agency.

Abstract: Chemical exposure is a concern within modern society due to the more than 10,000 chemicals estimated to be in commerce. Consumer products are a primary source of chemical exposures, yet little information is available on the chemical ingredients of these products and the concentrations at which they are present. To address this data gap, we created a database of chemicals in consumer products using product Material Safety Data Sheets (MSDSs) publically provided by a large retailer. Information from the MSDSs was extracted using regular-expression parsing, Optical Character Recognition (OCR), chemical name and registration number identification software, and manual curation through a straightforward multi-screen graphical user interface (GUI). The database includes product formulations and percent

composition. In addition, methods were developed to annotate consumer product use categories. The resulting database represents 1569 unique chemicals mapped to 8921 consumer products and a hierarchy of 319 consumer product use categories within a total of 15 top-level categories. We discuss the utility of this database for the purpose of prioritizing chemicals, and discuss ways in which it will be integrated with appropriate datastreams and workflows in support of (i) exposure screening and prioritization, (ii) generic or framework formulations for several indoor/consumer product exposure modeling initiatives, (iii) candidate chemical selection for monitoring exposure from proximate sources, and (iv) as activity tracers or ubiquitous exposure sources using “chemical space” map analyses; Chemicals present at high concentrations and across multiple consumer products and use categories that hold high exposure potential are identified. This work establishes a methodology for creating and augmenting a database from publicly available MSDSs. Our database is publicly available to serve regulators, retailers, manufacturers, and the public for predictive screening of chemicals in new and existing consumer products on the basis of exposure and risk. This work highlights the need for data on the prevalence of consumer product use to reduce a key source of uncertainty in assessing chemical exposure.

About the Consumer Product Chemical Profiles database (CPCPdb) - Product information in the CPCPdb is derived from publicly accessible sources of Material Safety Data Sheets (MSDSs), and future updates will also include information from product labels and information

provided directly by manufacturers. The U.S. EPA does not endorse or guarantee the accuracy or completeness of this information. The CPCPdb should not be considered a substitute for obtaining ingredient information either from product labels or directly from the consumer product manufacturers. The Agency makes no expressed or implied warranties, representations or endorsements (including, without limitation, warranties of title or non-infringement, or the implied warranties of fitness of brands for a particular purpose) with regard to any information provided in the CPCPdb The user assumes full responsibility for using these data and all data available through EPA's ACToR (Aggregated Computational Toxicology Resource) site (http://actor.epa.gov) and understands and agrees that the Agency is neither responsible nor liable to anyone for any claim, loss or damage resulting from use of such data.

Keywords

Chemical exposure; Consumer products; Ingredients; Product formulation; Near field exposure

List of Abbreviations

ACToR	Aggregated Computational Toxicology Resource
ADME	Absorption, Distribution, Metabolism, and Elimination
CAS-RN	Chemical Abstracts Service Registry Number
CBI	Confidential Business Information
CFR	Code of Federal Regulations
CGI	Common Gateway Interface
CHAD	Consolidated Human Activity Database
CPCPdb	Consumer Product Chemical Profile Database
CPSC	Consumer Product Safety Commission
CTCP	Clinical Toxicology of Commercial Products
DSSTox	Distributed Structure-Searchable Toxicity Database

EFH	Exposure Factors Handbook
FIFRA	Federal Insecticide, Fungicide, and Rodenticide Act
HPDB	Household Product Database
HTML	HyperText Markup Language
IUPAC	International Union of Pure and Applied Chemistry
MSDS	Material Safety Data Sheet
NACE	Nomenclature des Activités Économiques dans la Communauté Européenne
NAICS	North American Industry Classification System
NEISS	National Electronic Injury Surveillance System
NIOSH	National Institute for Occupational Safety and Health
NLM	National Library of Medicine
OCR	Optical Character Recognition
OSHA	Occupational Safety and Health Administration
QSAR	Quantitative Structure-Activity Relationship
REACH	Registration, Evaluation, Authorisation and Restriction of Chemical substances
SPIN	Substances in Products in the Nordic Countries
TSCA	Toxic Substances Control Act
URL	Uniform Resource Locator

1 Introduction

The production and use of chemicals is a hallmark of a modern and consumer oriented society. However, there is also a growing awareness that there are thousands of chemicals to which humans are unavoidably exposed everyday within a modern society (Glegg and Richards, 2007; Sanderson et al., 2013; Weschler, 2009). Of particular interest are chemicals from anthropogenic sources that can be potentially controlled. However, this is a daunting task since it has been estimated that there are some 80,000 different chemicals in commerce (Egeghy et al., 2012). To evaluate the potential risk to human health associated with chemical exposure, there are two primary considerations: the chemical hazard and exposure. Because of the large number of chemicals and limited available information about hazard or exposure, there is a need for screening methods of evaluation that have minimal data requirements , which can accommodate hundreds or thousands of chemicals at time, i.e. high throughput manner. Fortunately, under a multi-agency initiative (NRC Tox Testing 21st Century), considerable progress has already been made with respect to the development and implementation of high-throughput hazard testing. Progress on toxicity testing has in part highlighted the need for complementary high throughput evaluation of exposure.

For the purpose of high throughput exposure evaluation, there is strong justification based on both measurement (Morgan, M.K., et al., 2005; Pellizzari, E.D., et.al., 1999; Wallace, L.A., et.al, 1987) and modeling (Lai et al 2000; Nazaroff et al. 2012) studies to consider first and foremost those chemicals that are found in consumer products and are brought into our homes where circumstances exist for relatively high levels of exposure depending on their prevalence and frequency use along with their degredation or removal rates (Hertwich 2005). It is within the indoor residential environment where these consumer products tend to be stored and used, where people spend a large majority of time, and where susceptible individuals (very young and elderly and those who are sick or disabled) tend to spend even more of their time (Klepeis et al.

2001). Classic historical examples demonstrating both the exposure and health risk potential of chemicals in consumer products that led to their subsequent mitigation include, lead in paint (Farfel et al., 1990) and chlorpyrifos in household use pesticides (Jaga et al., 2003). It is recognized that depending on their purpose, exposure to chemicals in consumer products can result through either direct or indirect routes of contact (Dodson et al., 2012; Schettler, 2006; Rudel et al., 2003; Weschler and Nazaroff, 2008)

Because consumer product use is an important determinant for human exposure to a broad range of chemicals, it follows that information about the chemicals and their concentrations within those products is also an important consideration. There are a couple of potential sources for such data. One is material safety data sheets or MSDSs. MSDSs are required under the Occupational Safety and Health Administration (OSHA) Hazard Communication Standard to include product ingredients with known toxicity. They are intended to inform workers and emergency response personnel of hazards and their safe management. All hazardous components in excess of 1% (0.1% for carcinogens) are required to be disclosed through product labeling and the MSDS. OSHA defines a hazardous chemical very broadly as one that could possibly cause any physical or health effect under expected conditions of use or reasonably anticipated conditions of misuse. OSHA does not require MSDSs to be provided to consumers. However, with greater public interest for this kind of information, many manufacturers and retailers are providing MSDS forms as a public service. For example, Walmart has made the MSDS inventory of their products available (http://msds.walmartstores.com) since at least 2002. A second source of such data is the Consumer Product Safety Commission (CPSC). The CPSC has been collecting data of this nature for internal use and regulatory purposes for the last four decades (Bracken and Weiss, 1977; Byer et al., 1976). A third source is the book *Clinical Toxicology of Commercial Products* (CTCP). This book was one of the first to aggregate consumer product ingredient

compendiums dating back three decades including 14,000 products and over 900 generic product formulations (Gosselin et al., 1984). More recent efforts by the National library of Medicine (NLM) provide similar data known as the Household Product DataBase (HPDB) in a web-accessible format indexed by consumer product category. The licensed form of this database is the Consumer Product Information Database (CPID). It includes the MSDS listing and chemical percent composition by weight when available and is soft-linked to multiple NLM informatics resources (http://whatsinproducts.com/). This database is unique in that it provides quantitative composition information which is critical for evaluating the exposure potential (Jayjock et al., 2009). A fourth effort is underway by the chemical trade associations that represent both manufacturers and formulators to develop a Consumer Product Ingredient Communication Initiative to provide consumers with information about ingredients in products (Egeghy et al., 2011).

Accordingly, the current research is motivated both by need and opportunity. Given the large and growing number of chemicals that are used in consumer products that result in considerable exposure potential to complex mixtures, there is a need for high throughput evaluation of exposure as a fundamental component of risk screening. At the same time, public information is increasingly available on many chemicals that are in products which can be used as a basis for exposure screening and prioritization. Heretofore, what is lacking is the methodology for capturing the available data in a form that supports high throughput exposure screening for rapid exposure and hazard based analysis. In this paper we rely on mining available data sources, to perform a large scale examination of product compositions that can be used to inform aggregate exposure and generic product formulations used in regulatory exposure assessment models (i.e. EFAST or RIVM's ConsExpo). To facilitate this effort, we have built the Consumer Product Chemical Profiles database (CPCPdb). This effort is well aligned with the NRC 2012 espousing the efficient development of data and the application of

computational methods to derive new models that better predict exposure and provide a scientifically credible means for screening chemicals based on their exposure potential.

2 Methods

The Consumer Product Chemical Profile database (CPCPdb) has been designed as an "evergreen" (i.e., living) database. Additional products and their chemical ingredients will be aggregated from other sources of MSDSs in future versions. The workflow developed for building the CPCPdb can be broken down into three major steps:

1. Building and curating a database for consumer product ingredients and percent compositions using available MSDSs;
2. Identifying and annotating product use categories for all products in the database;
3. Evaluating data quality

Descriptions of each of these efforts is contained below and captured in Figure 1.

2.1 Building and curating a database for consumer product ingredient and percent compositions using available MSDSs:

Building the database for consumer products with chemical composition data involves several complicated scripting techniques, which are followed by a manual, crowd-sourced effort. The work was broken down into the four functions outlined below (for additional details see supplemental information (S1).

2.1.1 Identification, retrieval, extraction

The first step necessary in the formation of the database was collecting consumer product MSDSs from publicly available sources. Although multiple sources were identified, we chose to extract data from a single source to reduce the complexity of the initial data retrieval effort while developing the procedure. In February of 2011 we aggregated and captured links to

each of the Wal-Mart product MSDS sheets, which were originated from manufacturers of the products they sold at the time. This specific retailer's inventory was chosen because **(a)** its consumer product portfolio is relatively diverse (e.g., personal care, automotive, arts & crafts, household pesticides, lawn/garden, cleaning, home maintenance, home improvement, office supplies, electronics); **(b)** Walmart's omnipresence in the consumer product marketplace is expected to provide a relatively accessible, high-market share product inventory with greater consumer coverage; **(c)** its database of MSDS was available in Adobe PDF format; and **(d)** the documents were not only publically accessible, but also available without any visible restrictions. MSDS files were downloaded and entries were input into a MySQL (*MySQL:the world's most popular open source database*., MySQL AB, 1995. Database documenting a sequential id, file location, retrieval URL, and document format.

2.1.2 Chemical ingredient data extraction and management

Information from the downloaded MSDSs, including product name, chemical name, and Chemical Abstract Services registry number (CAS-RN),, were extracted to the MYSQL database. Roughly 70% of the MSDSs (in Adobe PDF) were embedded with text that was relatively easy to parse using custom scripts. The remaining PDF documents contained scanned images and were piped through Tesseract-OCR [(a) Smith, 2007 and (b) https://code.google.com/p/tesseract-ocr/)] for Optical Character recognition (OCR) and subsequently mined with appropriate regular expression filters. MSDS entries in CPCPdb were updated with product name and manufacturer. Entries for each identified ingredient in a given MSDS were crafted to document the MSDS in which it was found with chemical name and/or chemical registry number.

2.1.3 Chemical ingredient and quantitative composition data entry: crowd-sourced curation

Programmatic extraction of data from freeform or marginally formatted text is a science in its adolescence. To ensure that the CPCPdb had a level of quality that is commensurate with

our need to support exposure assessment, a custom web-enabled interface was developed for manually curating the entries in our MySQL database (see **Figure 2**). The interface, developed in PHP and Perl, provided curators/annotators with a means of opening a given MSDS, verifying the chemical names and associated registry numbers, and manually entering the "Min" and "Max" percent composition when available on the MSDS. A detailed SOP was drafted and updated regularly to ensure that curators were following the same guidelines during mining and curation. An additional feature of this custom interface was its ability to identify products that contained identical chemical compositions or identical MSDSs (e.g., 32 ounce versus 100 ounce bottle of the same shampoo). The curator could then verify the level of congruence between the different MSDSs. If the MSDSs were identical, a "soft-link" between them was created. If chemical compositions in different MSDSs were similar, but not identical, the curator could then copy curated information to reduce the amount of manual input needed. This feature reduced the overall workload of curation considerably since many products proved quite similar.

2.1.4 Chemical Name verification

Chemical identifiers (e.g., common names, IUPAC names, registry numbers) selected either programmatically or via the manual curation interface were immediately cross-referenced in EPA's Aggregated Computational Toxicology Resource (ACToR) (Judson et al., 2012) (http://actor.epa.gov). This linking of extracted information directly with chemical structure was a major goal of the project so that secondary parameters could be derived that support a number of risk assessment objectives (i.e. QSAR based calculation of ADME properties, physico-chemical properties, etc...). To ensure that the link was made properly, the interface provided the ACToR name associated with the chemical entity, as well as the extracted information from the MSDS. The link was then verified by either expert judgement or using appropriate secondary chemical name datasources, such as NLM's PubChem (http://pubchem.ncbi.nlm.nih.gov/), the Royal Society of Chemistry's ChemSpider

(www.chemspider.com), or the DSSTox master list (http://www.epa.gov/ncct/dsstox/). For instance, if the MSDS listed “Ethyl alcohol” and the associated CAS-RN linked to an ACToR entry for “Ethanol”, expert judgment was likely sufficient to identify that these are the same chemical and the entry was verified; however if the MSDS listed “cetearyl alcohol”, but the ACToR entry noted “Alcohols, C16-18” a search of secondary sources might be warranted. As curators became more familiar with the chemical composition of products and their related naming conventions, secondary sources became less necessary.

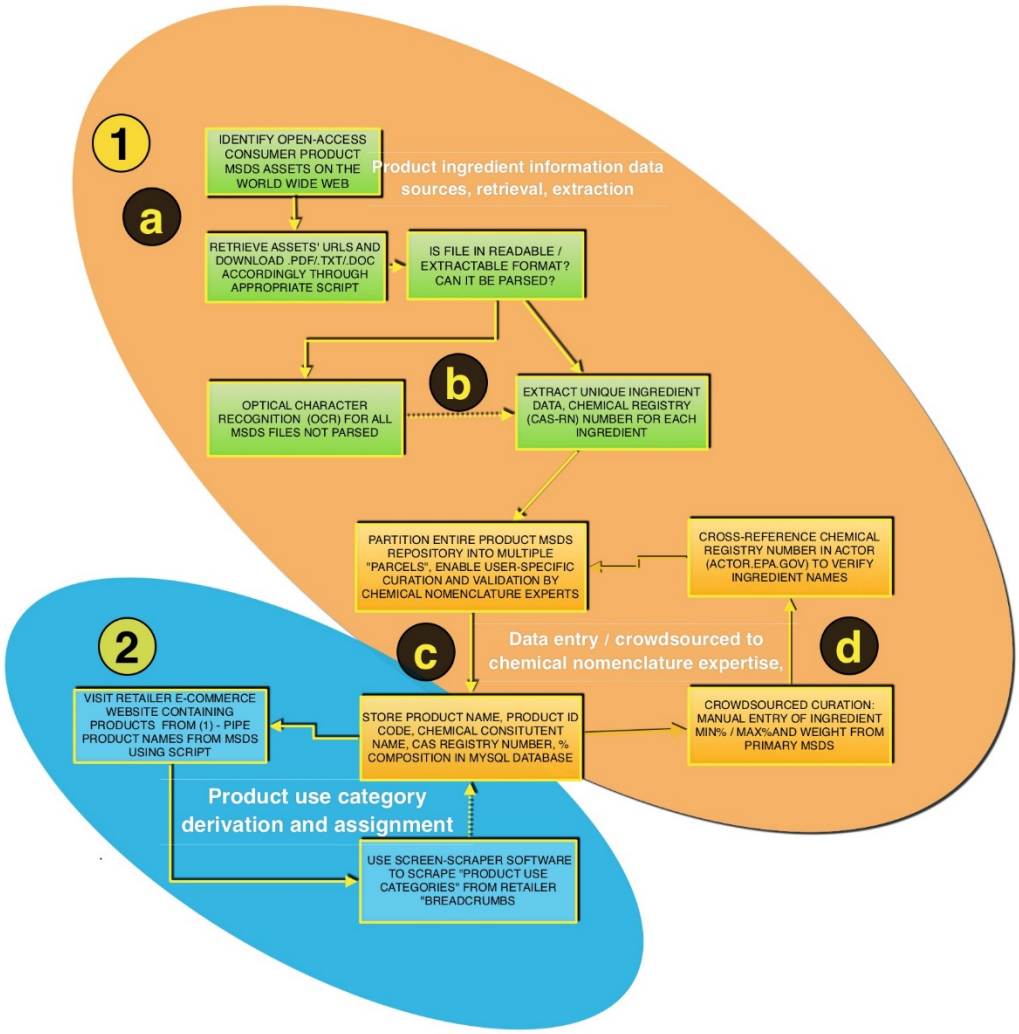

Figure 1: Consumer Product Chemical Profile **Data curation/annotation workflow:** A total of ~10K MSDS were identified, parsed, and manually annotated through a custom-built curation

system (see Figure 2). A unique set of roughly 9000 products resulted. Details of the workflow steps 1-3 are explained in the methods.

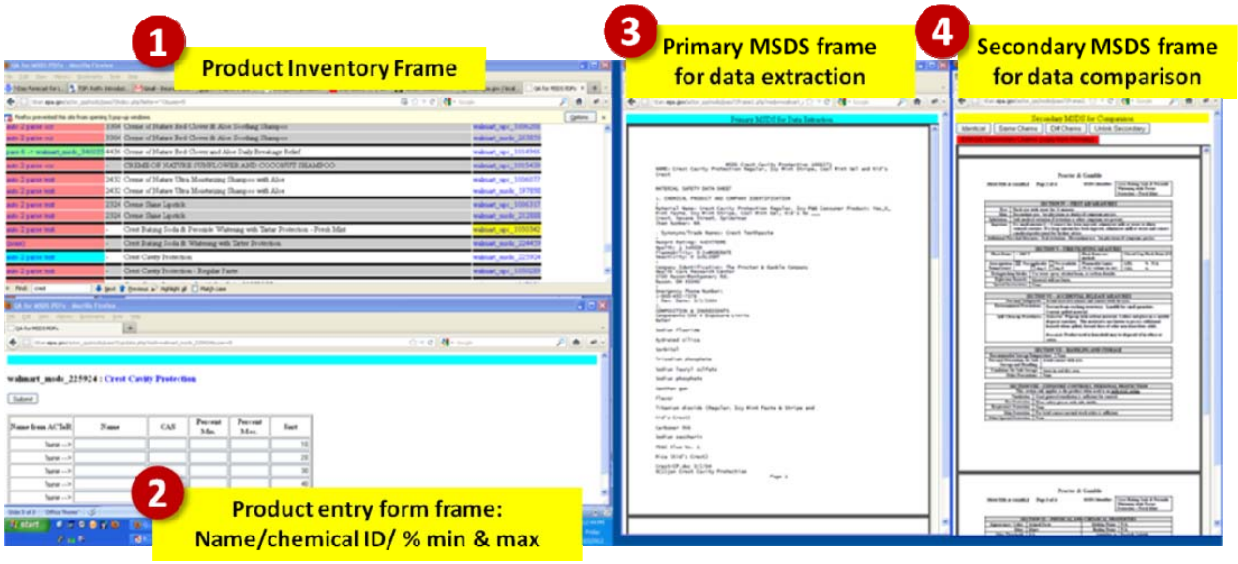

Figure 2: Chemical ingredient and quantitative composition data entry interface built for crowd-sourced curation, optimized for multi-screen end-user, built using a combination of MySQL, PERL (CGI), and HTML.

2.2 Identifying and annotating product categories for all products in the database

The classification of products informs manufacturers of the strategies needed for them to compete for market share. The classification also provides the basis for the society to monitor the health of the economy through observation of trends within the manufacturing sector(Meyer and Slick, 2001). Moreover, a system for product classification is valuable in assessing exposure to chemicals in those products during their manufacture, processing, and use (OECD, 2012). As a result, many categorization schemes are available for classifying consumer products (Table 3). The U.S. Environmental Protection Agency (EPA), Health Canada, and Environment Canada recognize that harmonized lists of “industrial function”, “consumer” and “commercial product” codes facilitates the transfer of information among different agencies; accordingly, they have developed harmonized codes with the intent of ensuring clear distinction, minimizing overlap between the codes, and ultimately improving exposure characterization. In

that same spirit, a crosswalk with the function and product codes in the Use Descriptor System developed by the European Chemicals Agency to support the implementation of Registration, Evaluation, Authorization and Restriction of Chemical substances (REACH) Regulation was developed under the auspices of the Organization for Economic Development (OECD, 2012). As these classification schemes are designed to both reduce burden on industry in chemical inventory reporting and to minimize the number of setting/process/use combinations that comprise the exposure scenarios used for evaluating chemicals, they broadly consolidate products into more general categories. The more detailed and hierarchical categorization schemes either tend to describe economic activities rather than products (e.g., NAICS, NACE), or share little consistency across applications (e.g., HPDB, NEISS, Simmons).

Abbrev.	*Name*	*Sponsor*	*Availability*
Product Categories			
HPDB	Household Products Database	U.S. National Library of Medicine	http://householdproducts.nlm.nih.gov/index.htm
C Series	U.S. - Canada Harmonized Consumer and Commercial Product Code	U.S. EPA, Health Canada, Environment Canada	http://www.epa.gov/cdr/tools/InstructionsManual.013112.pdf
EFH	Exposure Factors Handbook Consumer Products	U.S. Environmental Protection Agency	http://www.epa.gov/ncea/efh/pdfs/efh-chapter17.pdf
NEISS	National Electronic Injury Surveillance System	U.S. Consumer Product Safety Commission	http://www.cpsc.gov//PageFiles/106513/completemanual.pdf
Simmons	Simmons National Consumer Study	Experian Marketing Services	http://www.experian.com/simmons-research/consumer-study-details.html
REACH PC/AC	REACH Use Descriptor System Chem-ical Product and Article Categories	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
UCN	Use Code Nordic	Nordic Council of Ministers, Chemical group	http://90.184.2.100/DotNetNuke/LinkClick.aspx?link=DNNPortal-Download%2fFunktionskoder-eng+htm.htm&tabid=58&mid=448
SPIN/ UC62	Substances in Products in the Nordic Countries	Nordic Council of Ministers, Chemical group	http://90.184.2.100/DotNetNuke/LinkClick.aspx?link=UC62_ExplanatoryTxt.doc&tabid=58&mid=448
Function/Process Categories			
U Series	U.S. - Canada Harmonized Industrial Function Categories	U.S. EPA, Health Canada, and Environment Canada	http://search.oecd.org/officialdocuments/displaydocumentpdf/?cote=env/jm/mono%282012%295&doclanguage=en
PROC	REACH Use Descriptor System Process Category	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
Appendi	REACH Use	European Chemicals	http://echa.europa.eu/documents/10

x R.12-6	Descriptor System Functional Category	Agency	162/13632/information_requirements_r12_en.pdf
Economic Activities			
NAICS	North American Industry Classification System	U.S. Census Bureau, Statistics Canada, and Mexico's Instituto Nacional de Estadistica y Geografia	http://www.census.gov/eos/www/naics/2012NAICS/2012_Definition_File.pdf
IS	Industrial Sector	U.S. Environmental Protection Agency	http://www.epa.gov/cdr/tools/ReplacingNAICSwithIS.pdf
NACE	Nomenclature des Activités Économ-iques dans la Communauté Européenne	European Commission Directorate for Economic and Financial Affairs	http://ec.europa.eu/competition/mergers/cases/index/nace_all.html
SU	REACH Use Descriptor System Sector of Use	European Chemicals Agency	http://echa.europa.eu/documents/10162/13632/information_requirements_r12_en.pdf
Activity and Location			
CHAD	Consolidated Human Activity Database	U.S. Environmental Protection Agency	http://www.epa.gov/heasd/chad.html
ATUS	American Time Use Survey	Department of Labor, Bureau of Labor and Statistics	http://www.bls.gov/tus/

Table 3: Common schemes for the classification of consumer products, functional categories, economic sectors, and related activities and locations.

Lacking a standard classification scheme with sufficient level of detail and suitable parent/child hierarchies, we chose to use the categorization scheme that was already most tightly linked to the source of our consumer product MSDSs, the retail product categories from the retailer's online shopping website. Each product sold through the online commerce interface is mapped to a retail category, displayed at the top left of each product page (see **Figure 3**). These "breadcrumbs" consist of a top level category (e.g., "beauty", "grocery", "home improvement") and a subcategory to which the product belongs (e.g., "shampoos", "pest

control", "weed & fungus control"). Although the categories may seem at times odd or less intuitive, the main objective was to use the categories codified by the retailer to avoid the creation of yet another product use classification scheme. As this categorization scheme is successfully being used by a large retailer, it is likely to be acceptable and readily understandable by the public.

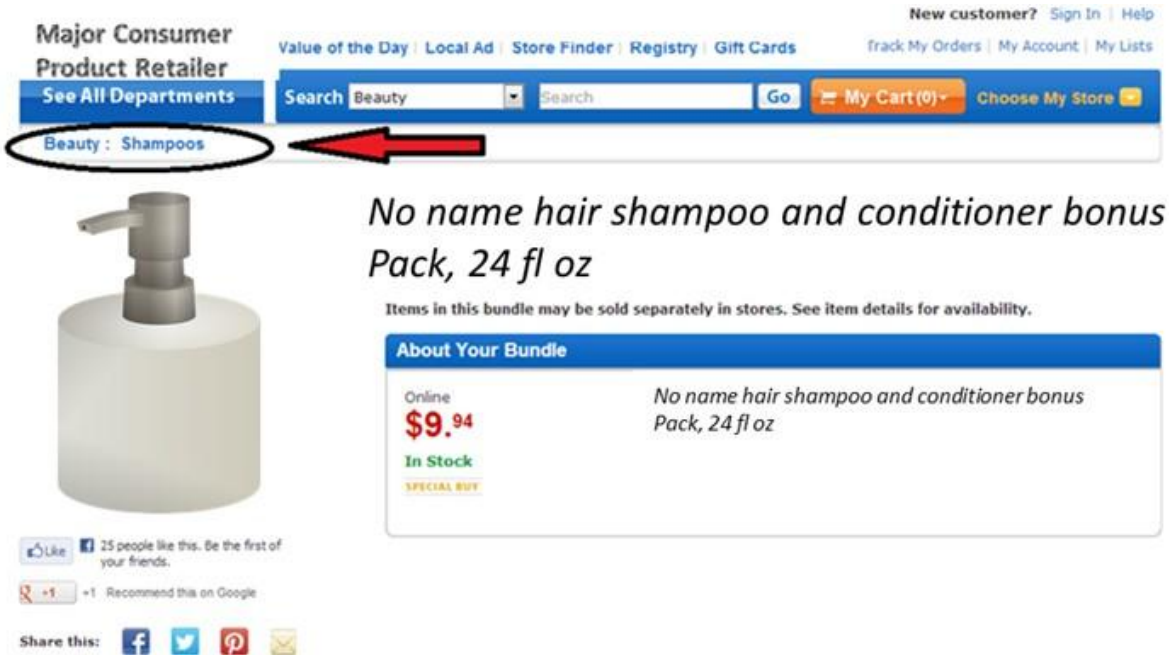

Figure 3: Product Use Category "Breadcrumbs" (circled in black with red-arrow) from retailer's e-commerce website, with the breadcrumb containing product use category (top level = Beauty, Sub-category = shampoos) for a fictitious commercial product.

A freely available data extraction program, ScreenScraper (http://www.screen-scraper.com/), was used to automatically retrieve these product categories from the retailer's website (www.walmart.com). A program was coded to search the website for each product by name, and if a record was found, the associated breadcrumbs were retrieved from that product page. Following the automated category retrieval, multiple manual checks were used to solve specific issues seen in the product classification. Misspellings and special characters in

category names were corrected (i.e. things like “P$rsn%ol Care” or “personal car” were corrected to “Personal Care”. Missing or clearly wrong categories were manually located via the retailers web interface. If the exact product was not found via the search, the expert curator made the decision whether the returned results were similar enough to be used as surrogates. If a similar product was not found, the category was eliminated. The scraping software only scraped the top returned record for a particular product search. Therefore, if a product was classified by the retailer in multiple categories, it was possible that very similar products could be assigned different categories by the automated search. To rectify this problem, curation experts evaluated every categorization to make sure that it was both consistent with assigned categories for similar products within our database and logical based on personal experience.

2.3 Evaluating data quality

A manufacturer only has to report ingredients that must be legally reported on a MSDS, which most often is a subset of all ingredients. This being said, many manufacturers are now reporting more of their product ingredients due to ever increasing public pressure for transparency. By having a large sample size with many brands of products for any given consumer product class, the belief is that while chemical information on certain individual products may be lacking, the aggregated data across a product class will effectively capture the majority of possible ingredients due to the same functionality. As we do not personally draft the MSDSs, nor does the provider of these documents, any lack of information is not our primary concern in quality metrics; our goal was to quantitatively evaluate our transcriptional accuracy.

Upon completion of the first pass of manual data curation and entry, we randomly selected approximately 130 MSDSs (sampled $n^{1/2}+1$, where n is approximately the 12K de-duplicated MSDSs of which 9K have relevant data) from our database to repeat our curation process. Entries of each randomly selected product were compared to those in the MSDS. Omissions and transcriptional errors within the database for names, CAS-RN, and %

composition were tabulated. Additionally, the stored categories for each of these products were evaluated based on both the results obtained by directly searching for the product on the retailer website and reasonable assumptions based on our expert's categorization knowledge.

3 Results

3.1 Data Extraction

In total, 22,401 MSDSs were retrieved from the selected data source. Initial extraction with natural language processing scripts yielded over 80,000 unique MSDS/CAS-RN pairing (with only about 5% of those being retrieved from image based PDF). Further improvement of extraction scripts netted an additional 15,000 putative MSDS/CAS-RN (with more than a third from OCRed sources). Upon completion of the scripted methods of data extraction, crowd-sourced manual curation by the authors was undertaken. Of the 22401 MSDS entries, 13514 duplicative entries (i.e. entries that have identical composition information as another MSDS entry) were identified, and they represented 5267 unique products. This process reduced the number of MSDSs for curation to 14294 and eliminated the need to curate nearly 27000 MSDS/CAS-RN pairings. Of the roughly 68,000 (80K + 15k - 27K) remaining programmatically identified pairings, 59K of the pairings were confirmed to be correct by manual analysis (a specificity of roughly 87%). 4500 additional pairings that were missed by the automated procedure were also identified (indicating a sensitivity of 93% in automated CAS-RN extraction). In total, ~13K MSDSs were found to have extractable chemical information (name or CAS-RN). Of the 3909 unique chemical entries remaining after manual extraction, only 2958 mapped to a generic chemical in the ACToR database.

3.2 Quality Assurance (QA)

This database was largely hand curated, which requires high attention to detail. In this project, 20 curators were involved with the possibility of 20 different interpretations of protocols for handling questionable cases (e.g., information formatted in a unique way or errors identified

in the MSDS such as mismatches between names and CAS-RNs). Based on our selected 130 product subset, it appears that while errors did occur, they were only occasional.

As CAS registry numbers were considered the most definitive identifier for a chemical during ingredient extraction, it was one of the key pieces of information included in our database. In the 130 product subset for quality assurance, only 7 of the products had a case where there was a listed CAS-RN in the MSDS that was not identified in our database. Only 1 entry had improperly transcribed CAS-RNs. In total, 97% (587/608) of the CAS-RNs in the 130 product sample were properly annotated. Secondary to CAS-RN extraction was the collection of chemical names listed in the MSDS. CAS-RNs were typically found in the tabulated ingredient section in the MSDS; however, names were found in nearly all sections of the MSDS making omission more likely. Upon detailed review of the 130 selected MSDSs, 88% (114/130) were found to have no missed chemical names. For chemical names, 824 of 914 names listed on the 130 MSDSs were captured. While the collection of names was more error prone than CAS-RN identification, the specificity of ingredient identification was errorless (i.e. no names were listed that did not occur in the MSDS, or conversely, all names listed were in fact in the MSDS). This high specificity is key since improperly labeling a product as containing a chemical is more problematic than missing an ingredient that does occur (Note that ingredient listings from MSDSs already have a relatively low sensitivity due to the limitations of legal requirements of MSDS reporting).

When examining the extraction of composition data from the MSDSs, we expected that there would be more omissions. The standard operating procedure (SOP) for manual extraction went through various upgrades during the curation process, and some of the earlier curation was focused only on name/CAS-RN extraction. As curation progressed, the importance of collecting the numeric compositions became more apparent and was added to the procedure. As such, we expected that some of products curated early in the process would be missing the

percent composition data. During QA, we found that more than 80% of compositions identified within the MSDS were annotated in the database. There were only a small number of cases (1 MSDS with 5 ingredients) where composition data stored in the database was incorrect. These errors were due to transcriptional errors in the proper linking of ingredient with quantitative composition; ingredient and max concentration values were mismatched due to ordering difference in the QA interface and the MSDS.

QA was also completed on the categorization of the products listed within the CPCPdb. Based only on comparing to the category of the first record returned on the company website, 87.5% of the product categories matched at the top-level (15 categories), and 73% matched at the most detailed level of the classification tree. These differences were mainly caused by changes made by the retailer's website during the lag time between the original annotation (mid 2012) and the QA (Early 2013). Some categories that we had in the database no longer appear on the retailer's site. Some products were no longer found on the website. And, some products were found in more than one category, and the error here may be in the eye of the beholder. Take for example the product "Andis Cool Care Plus", a spray used to clean electric hair clippers, which is categorized under "Beauty; Hair Styling Tools". While this category may seem appropriate to the retailer and the consumers, an exposure scientist may consider changing the category to "cleaner". Another example is "Fizzies Fish 'n Splash Asst Holiday 2007" (http://goo.gl/74YWX). This package is no longer found on the retailer's website, but similar products are listed as bath toys. This package contains a toy (plastic fishing rod), bathing accessories (bath fizzies), and a body cleanser (moldable foam soap). The proper categorization for a combination product (or multiple component kit) of this type is difficult to discern either programmatically or during typical curation efforts. These examples highlight the need to generate a standard for categorization of consumer products in the realm of exposure

sciences. Determining the appropriate way to deal with the complexity of consumer product classifications is something upon which we will devote additional consideration in the future.

One of the goals of this project was to facilitate the identification of ingredients at the "product category" level. When considering the errors of omission identified in the ingredient database, we consider it unlikely that the same ingredient would be mistakenly omitted from all products contained within a single product category. For example, a hand soap (product #1528) falls under the "Health: Hand Soap and Sanitizers" retail product code category. This category also contains an additional 105 products spanning 100 unique chemicals. Each of these 100 chemicals are present in at least 2 (although mostly >>2) products within the category (See Figure 4). If a chemical or percent composition for a particular product was not captured in the curation process, the chemical would still be captured in other similar products within the same category. The MSDS for product #1528 listed a total of 13 ingredients, (http://goo.gl/74YWX) for which the composite had three ingredients that cannot map to CAS numbers as unique chemical entities. These three ingredients were "fragrance" (potentially 1 of > 3000 different possible fragrances used in commerce), "lanolin" (a mixture of several thousand chemical entities derived/extracted from sheep wool), and "polysorbate 20" (a complex mixture). If we rank order across all identified ingredients by "average % composition" we identify that the rank order of each of the chemicals is "identical" (literally) from highest to lowest % by weight. Furthermore, in the case of propane/isobutane (propellant) a % composition of 11% was listed non-specifically to both combined....our analyses provides a number of 4.6% (isobutane) and 4.4% (propane)

Using composite data as a surrogate product in a specific category supports an informatics-driven approach to rapidly summarize product-use categories into generic or framework formulations or chemical profiles; a requirement to reduce the chemical space

independent of product branding, and a data requirement for many consumer product exposure models today that require generic formulation to reduce dimensionality of exposure models.

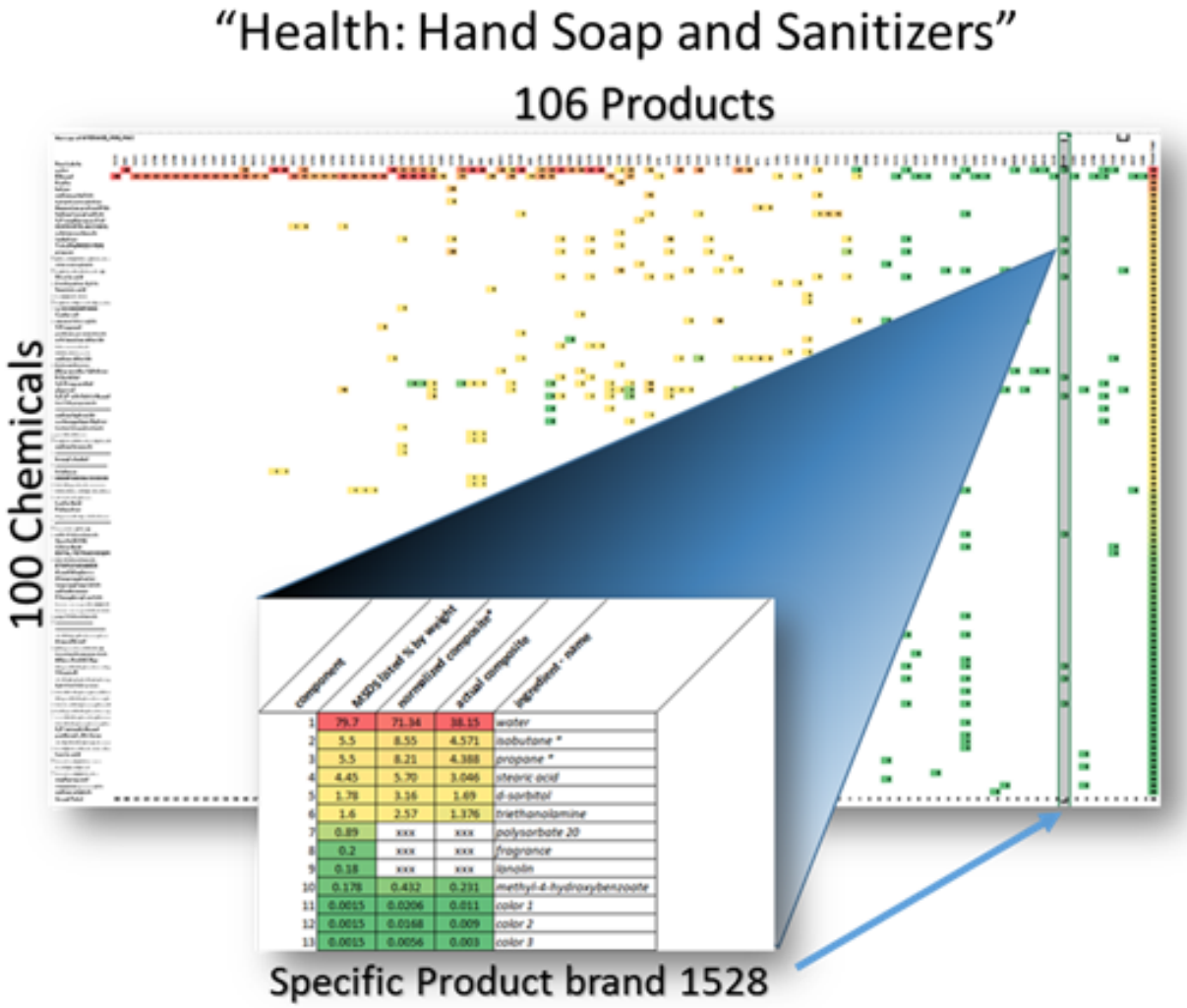

Figure 4: Example of how a unique product in the "Health:Hand Soap and Sanitizers" category with missing quantitative data (poor data curation with missing % min/max) can still be quantitatively estimated with only the list of ingredients present in the formulation using the composite data from other "same-in-class" products to generate a relevant ingredient profile. Whether verbatim aggregate estimates (column 3) or normalized estimates (column 2), the rank order of the ingredients derived from composites commensurates with actual ingredient rank order gleaned from the product's MSDS composition data (column 1).

3.3 Ingredient Diversity

One of the primary goals of this study was to document the diversity of chemical ingredients within the multitude of products commonly being purchased by consumers. Other data sources such as the NLM's Household Product database include on the order of 3000 chemicals, however many of the products listed there are now discontinued (nearly 30%). It is vital to keep the chemical space covered by consumer product ingredients current as products may have been reformulated or discontinued due to health concerns of certain chemicals, product efficacy optimization, environmental impact, or better/cheaper alternatives being discovered. If our database still includes such products, we will greatly overestimate the public exposure to chemicals in those products and identify "risks" that in reality have already been mitigated. Currently, our database contains on the order of 1800 chemicals, and there may be many more chemicals present in consumer products due to a number of reasons: 1) some chemicals are trade secrets, 2) chemicals of minute quantity may not be listed in the MSDS, and 3) our database was built using the inventory of a single retailer.

To visualize the variety of chemicals within the database, we populated a network diagram (seen in Figure 5) with the large nodes representing top-level product categories and the small nodes representing unique chemical species. Edges are drawn if a chemical is present within a product in that product category. In looking at Figure 5-1 (generated using Cytoscape: Shannon P, Markiel A, Ozier O, *et al.* (2003). "Cytoscape: a software environment for integrated models of biomolecular interaction networks"), it is easy to see three different classes: chemicals that are only present in a single product class (in green), those that are present in a small number of classes (in orange), and those that exist in a large number of product classes (in red).

The foremost class (containing compounds that are only present in a single product category) contains chemicals that may be used as *usage tracers* for a specific product use

category. Take for example lithium hydroxide, which is only found in beauty products--particularly hair conditioners and moisturizers. Measuring lithium hydroxide in different homes may provide an estimate of relative usage level of hair conditioners. Even if such quantitative comparison is not possible, if lithium hydroxide is found in a home, it is almost certain that a hair conditioning treatment has been used. Other sentinel chemical examples include 3,6,9-trioxaundecamethylene bis(2-ethylhexanoate) for “Home Improvement: Sealants, Fillers and Adhesives”, 2,5-thiophenediylbis (5-tert-butyl-1,3-benzoxazole) for “Toys: Drawing and Coloring”, Phosphordithioic acid for “Automotive and tires”, and ethofenprox for “Pets: Flea and Tick Control”. By knowing the degradation properties of sentinel chemicals, it may be possible to determine the likely time a product was used. This type of forensic examination may allow us to harness chemical markers to tell us about product usage in the home without relying on somewhat error-prone surveying mechanisms to monitor product application patterns. In addition, with the large number of ingredients identified to be potential tracers, scientists can choose the chemicals with the optimal properties for the detection media they prefer.

The second class (colored orange in Figure 5-1: containing compounds which are found in a **couple or multiple** product categories) contains chemicals that have lower value as a tracer than those in the previous class. These compounds can still narrow down products used in the home, but to truly understand product usage, more ingredients would have to be monitored. On the other hand, these compounds may be of greater interest to exposure scientists since being present in more products could be indicative of higher aggregate exposure potential.

The final class (colored red in Figure 5-1: containing compounds which are found in **several or most** product categories) contain chemical with no use as tracers, but have the greatest exposure potential. Many of these ubiquitous chemicals have been well studied, including water, ethanol, glycerol, titanium dioxide, 1,2-propanediol, methyl 4-hydroxybenzoate

(methyl parabens), propane, propyl 4-hydroxybenzoate (propyl parabens), isobutane, isopropyl alcohol. It is likely that most people are exposed to these chemicals, even though to greater or lesser degrees depending on their product use choices. Thus, when prioritizing chemicals based on the probability of exposure, chemicals in this class will be considered most likely to come into contact with a human receptor. In addition, since these chemicals are in many products, application amounts based on a single product or even a product category will be insufficient to accurately evaluate aggregate exposures.

3.4 Chemical Co-Occurrence and Product Use Category Similarity Clustering

A statistical visualization of 15 broad product use category types is shown in Figure 5-4 with an upper triangular matrix formulation of chemical co-occurrence in products (gradient color legend with red = high, yellow = low, blue = zero and grey = N/A). In Figure 5-1,2,3 or within row in Figure 5-4, products with similar ingredients cluster closer together (for instance “beauty”, “health” and “baby”, “grocery” are closer together than a satellite product node of “apparel” and “photo center”, yellow ovals in Figure 5-1,2,3; the remaining classes then cluster together

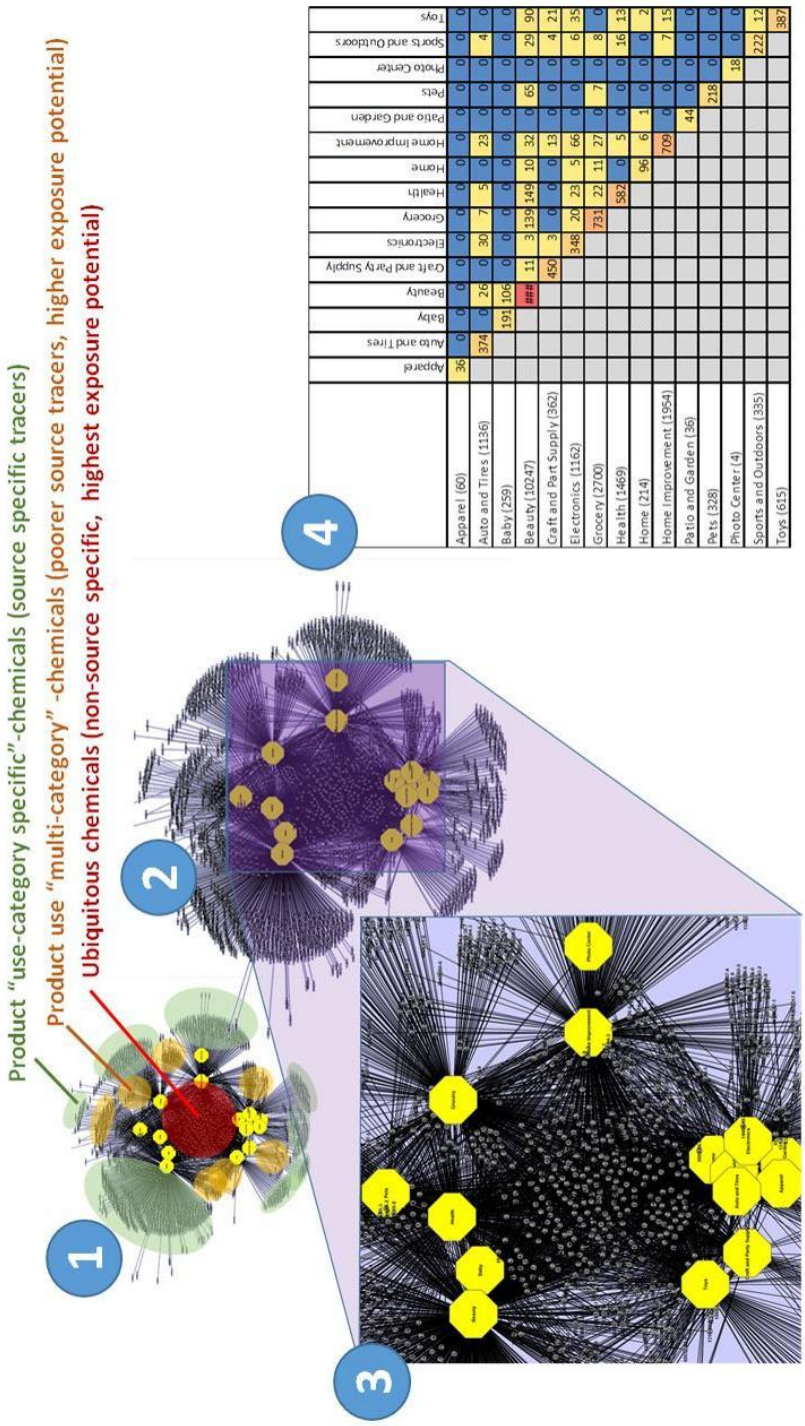

Figure 5: Network diagram of 15 top level product use categories (yellow) and summary table of the CPCPdb product and chemical landscape. (1) Network with various regions of interest, 2) broad overview, 3) detail view of ubiquitous high-exposure potential chemicals, and 4) upper triangular matrix formulation of chemical co-occurrence in products. In 1,2,3 or within row in 4, products with similar ingredients cluster closer together (for instance "beauty", "health" and "baby", "grocery" are closer together than a satellite product node of "apparel" and "photo center"; the remaining classes then cluster together.

3.5 Database Applications & Usage

Based on the results from examining our network diagram (Figure 5), the CPCPdb can help to **identify tracer chemicals and biomarkers of exposure**. By applying proper chemical property filters, marker substances with extended half-life and low vapor pressures can be identified and surmised to leave residues on indoor surfaces and dust if they are present in products used within the home. Based on marker substances, proper analytical schemes could be developed to allow for utilizing dust samples and residue swabs as fingerprints of personal activities. Such use-centric, as opposed to hazard-centric, selection of chemicals for study may lead to better understanding of general exposure mechanisms for those chemicals that are frequently encountered in daily life. In addition, knowing that a population has near field exposure to a specific chemical, large-scale biomonitoring studies such as the National Health and Nutrition Survey (NHANES) may consider monitoring this chemical in human bodies.

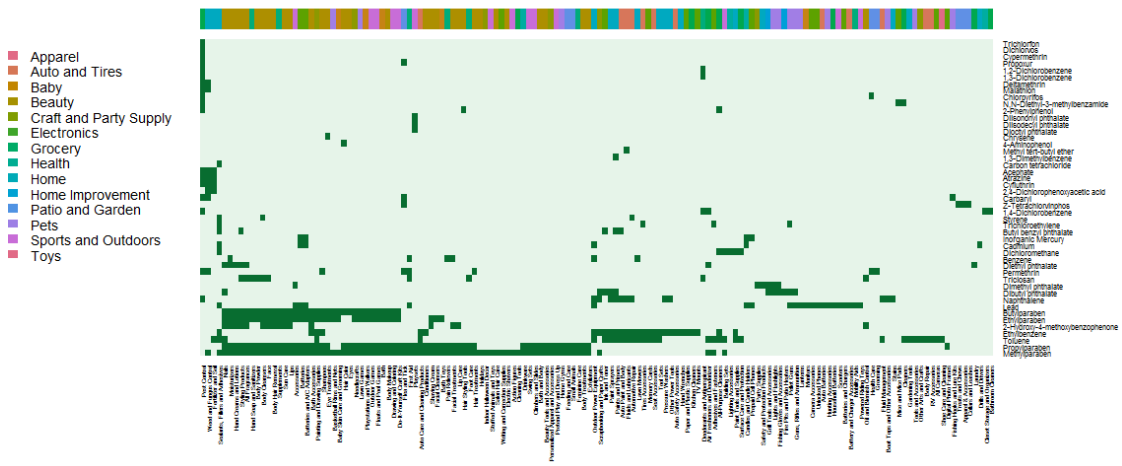

Figure 6: **Product categories for NHANES Chemicals that are in the CPCPdb.** Dark color indicates the presence of these chemicals in a specific product (lower) category. The ribbon at the top indicates the corresponding upper product category. Chemicals and products were clustered using Euclidean distance and complete.

			Product Categories			
			Lower Categories		Upper Categories	
Class	Number of Compounds	Number in CPCPdb	Number	Most Common	Number	Most Common
Carbamates	6	1	2	Flea and Tick	2	Grocery
DEET	1	1	3	Cleaners	2	Grocery
Dithiocarbamate Pesticides	8	0				
Environmental Phenols	4	2	21	All Fragrances	2	Beauty
Herbicides	7	3	4	Fertilizer and Soil	3	Home Improvement
Organochlorine Pesticides	21	1	3	All-Purpose Cleaners	2	Grocery
Organophosphate pesticides	2	1	2	Pest Control	2	Grocery
Organophosphorus Insecticides	39	7	2	Pest Control	2	Grocery
Other Pesticides	7	1	9	Pest Control	4	Pets
PAHs	12	3	5	Air Fresheners and Deodorizer	3	Grocery
Parabens	4	4	75	Accessories	12	Beauty
Perfluorinated Compounds	12	0				
Phthalates	10	7	23	Playsets	9	Beauty
Phytoestrogens	11	0				
Polybrominated Biphenyl	1	0				
Polybrominated Diphenyl Ethers	10	0				
Pyrethroid Pesticides	4	4	9	Pest Control	5	Pets
Sulfonyl Urea Herbicides	17	0				
Tobacco	1	0				
VOCs	33	12	54	Sealants, Fillers and Adhesives	11	Home Improvement
Total	210	47				

Table 4: The occurrence of chemicals within the CPCPdb of various NHANES chemical classes that might be in the home.

NHANES is a program designed to assess the health and nutritional status of adults and children in the U.S. (http://www.cdc.gov/nchs/nhanes.htm). NHANES chemicals are roughly grouped into chemical classes (e.g. environmental phenols, herbicides, perfluorinated

compounds); among those chemical classes likely to be in consumer products (listed in Table 4), the biomarkers monitored by NHANES imply exposure to at least 210 different chemicals. As is shown in Figure 6 and Table 4, of the 210 chemicals that might be found in consumer products, only 47 appear in the CPCPdb. Parabens, Phthalates, and Volatile Organic Compounds (VOCs) are in the largest number of product categories. Beyond these promiscuous chemicals, Figure 6 shows that most chemicals are in products with only a handful of uses, indicating that the presence of biomarkers for exposure to these chemicals may correlate with specific activities. Bisphenol A, Flame retardants (Polybrominated Biphenyl and Diphenyl Ethers) and Perfluorinated compounds are not found, though they are known to be either in packaging or used in the manufacture of many products in the home.

Combining biomonitoring results with knowledge of exposure routes and in-home chemical sampling data (e.g., dust) would provide a much more complete picture of the exposure paradigm.

4 Discussion

TheCPCPdb provides information as part of an ongoing effort to provide up-to-date knowledge on the chemicals people are exposed to from the products they use every day. These same data are vital quantitative components to be integrated into next-generation exposure prioritization frameworks.

One of the biggest challenges for new high-throughput exposure assessment methods is to infer near field exposure from the minimal data (e.g., production volume) available for most chemicals (Little et al., 2012; Nazaroff et al., 2012; Shin et al., 2012; Wenger and Jolliet, 2012). Recent evaluation of high-throughput exposure methods have demonstrated that near field use is highly predictive of exposure (Wambaugh et al. 2013). The CPCPdb will allow scientists and regulators to fill critical input gaps needed in quantitative models that predict near field

exposures. The information in the CPCPdb can significantly advance characterization of real-world exposures to complement toxicity screening in support of rapid risk assessment.

Near field exposure models such as ConsExpo (RIVM), CEPST/ComET (LifeLine) and E-FAST (U.S. E.P.A) require generic product formulations. Typically, product ingredients in the same functional use categories (e.g., solvent, surfactant, stabilizer, fragrance) have very similar chemical structures and/or physicochemical properties. As the chemical ingredients in the CPCPdb are annotated with their functional use categories, the generic formulations currently-used by these models can eventually be updated. Adding functional use categories to the CPCPdb will also allow *alternative* chemicals to be quickly identified. By recognizing the function a chemical has in a product, then if that chemical is found to be hazardous, potentially safer alternatives that maintain the product and formulation integrity with desired end characteristics could be identified.

For full, mechanistic exposure models additional data will be needed, such as “product use category”-specific exposure factors (e.g., duration, frequency, magnitude or amount and mode), physico-chemical uptake properties (e.g. Henry’s law constant, vapor pressure), and chemical-specific ADME factors (e.g., *in silico* predicted or *in vitro* measured properties such as route-specific absorption, biological half-life, metabolic clearance, etc…) (Blancato et al., 2006; Zartarian et al., 2012; Wambaugh et al., 2013).

The CPCPdb provides a lookup table for field technicians who might be documenting near field exposure in a study home. A barcode scanner app is currently under development that will allow a user to conduct a consumer product inventory around the household. The app links to CPCPdb to provide for each product scanned the chemical composition and any information contained within ACToR for those chemicals.

It is important to note that the CPCPdb consists of data reported on MSDS forms, many of which may be incomplete. To safeguard intellectual property and protect trade secrets, the

formulation of a product, functional components required to make a product, and the choice of a chemical ingredient to support the desired end-use functional characteristics of the product in finished form for consumers are typically classified under confidential business information (CBI); not all ingredients need to be reported. Obtaining information on fragrance and phthalate ingredients is known to be particularly problematic (Egeghy et al., 2011). For instance, there are well over 3060 unique materials used in fragrances, chemicals that have received considerable scrutiny due to their abundance and diversity (several thousand are used in various combinations in several thousands of products), associations with asthma and multiple chemical sensitivity, yet it is completely acceptable to include them in consumer products without listing a CAS-RN or a name (i.e., simply “trade secret”). We have not stored these components as it is impossible to identify them or even infer what they might be (e.g., one product contained twelve ingredients, each listed as “trade secret” with a percent composition listed as “1-100%”).

In addition to ingredients that are covered under trade secret protection, there are also chemicals that can be detected in consumer products that are not knowingly added such as accidental contaminants, products of degradation, by-products of manufacturing or chemical synthesis, residues from packaging or handling equipment, etc.. Since these chemicals are not intentionally added they are not documented in MSDS and are not included in our database, although some could potentially be inferred from product packaging material labeling or an RIC (resin identification code). For instance, if a product is purchased in a bottle that is either PVC or polycarbonate or other resin, chemicals used to produce the plastic resin may inadvertently become part of the formulation. Important examples include phthalates, which are used as plasticizers in a variety of otherwise brittle polymers, and bisphenol A (BPA), which is the monomer form of some polycarbonate polymers). It is highly unlikely that the manufacturer would list these chemicals, as they are a result of the production and packaging process and are not actual “ingredients.” Despite not being explicitly listed as present, phthalates are

commonly detected in bottled water because of the material used in the bottle or other packaging (Al-Saleh et al., 2011).

The Silent Spring Institute recently released a study that started with a list of 66 chemicals of interest because of their putative adverse health outcomes (Dodson et al., 2012). The study detected 55 of these chemicals in 213 different consumer products; however, several of these chemicals are not true ingredients, but are rather by-products, accidental contaminants, or packaging residues. However, the manufacturer of a consumer product does not necessarily also manufacture the packaging and therefore may be unaware of additional ingredients potentially introduced by packaging.

Finally, the current release of CPCPdb only contains “consumable” consumer products (e.g. liquids, solids, aerosols, powders) that are used up and replaced, and does not include articles (e.g. electronics, furniture, building materials). Identifying and characterizing the chemicals found in articles and their corresponding exposure potential (via emission into indoor air or via human contact) is an ongoing area of EPA research.

5 Conclusion

More than 80% of the chemicals listed in the EPA’s Toxic Substances Control Act (TSCA) workplace (http://www.epa.gov/oppt/existingchemicals/pubs/workplans.html) are present in consumer products. This effort is vital to ensuring public safety and fills in the data gaps in near field exposure models. By collecting consumer product ingredient information and product use category information in a centralized repository, we have developed a better understanding of near field chemical exposures. In parallel with the toxicoinformatic infrastructures required for computational toxicology, these data bring informatics-driven exposure prioritization one step closer to high-throughput risk assessment.

Acknowledgements

The United States Environmental Protection Agency through its Office of Research and Development funded and managed the research described here. The initial funding for this research came from a U.S. EPA Pathfinder Innovation Project award for “Systems Reality Modeling.” We are grateful to the Shaw University Research Internship program for providing students that assisted in this research. We thank **Charles Bevington** and **Cathy Fehrenbacher** of the U.S. EPA's Office of Chemical Safety and Pollution Prevention for collaboration and helpful discussions. We thank **Mike Uhl** (Lockheed Martin Information Technology Services), **Ravi Nair** and **Heidi Paulson** (both U.S. EPA’s Environmental Modeling and Visualization Laboratory) for project coordination on the interface design. We thank **Pertti Hakkinen** (National Library of Medicine) and **Henry Delima** (Henry Delima Associates) for providing input. We thank **Satori Marchitti** (NERL) for administrative review and feedback.

Disclaimer

The views expressed in this paper are those of the authors and do not necessarily reflect the views or policies of the U.S. Environmental Protection Agency.

References

Altman RG, Morello-Frosch R, Brody JG, Rudel R, Brown P, Averick A, Pollution Comes Home and Gets Personal: Women's Experience of Household Chemical Exposure. J. Health and Soc. Behav. 2008, 48, 417-435

Al-Saleh I, Shinwari N, Alsabbaheen A. Phthalates residues in plastic bottled waters. J Toxicol Sci. 2011 Aug;36(4):469-78. PubMed PMID: 21804311.

Anastas P, Teichman K, Cohen Hubal, EA, Ensuring the safety of chemicals. *J Expos Sci Environ Epidemiol* 2010, *20*, (5), 395-396.

Blancato et al., Exposure Related Dose Estimating Model (ERDEM): A Physiologically-Based Pharmacokinetic and Pharmacodynamic (PBPK/PD) Model for Assessing Human Exposure and Risk, June 2006, **EPA Report EPA/600/R-06/061**

Bracken M.C.; Weiss, I. J.; Database Development in a Regulatory Agency. J Chem Inf Comp Sci, 17(4), 202-205, 1977.

Byer, W. L., Landau, H. B.; Neufeld, M. L.; Rosenthal, H.; Building a Chemical Ingredient Data Base for Industrial and Consumer Products. J Chem Inf Comp Sci, 16(3), 137-141, 1976.

Dodson, RE, M Nishioka, LJ Standley, LJ Perovich, J Green Brody, RA Rudel. (2012) Endocrine disruptors and asthma-associated chemicals in consumer products. Environ Health Perspect; 120:935-943.

Egeghy PP, Vallero DA, Cohen Hubal EA. Exposure-based prioritization of chemicals for risk assessment. Environ Sci Pol. 2011; 14:950-964.

Egeghy PP, Judson R, Gangwal S, Mosher S, Smith D, Vail J, Cohen Hubal EA. The exposure data landscape for manufactured chemicals. Sci. Total Environ. 2012 414:159-166.

Farfel M R and Chisolm, Jr. J J Health and environmental outcomes of traditional and modified practices for abatement of residential lead-based paint. American Journal of Public Health October 1990: Vol. 80, No. 10, pp. 1240-1245.

Pediatrics. 2012 Dec;130(6):e1757-63. doi: 10.1542/peds.2012-2757. Epub 2012 Nov 26

Glegg GA, Richards JP (2007) Chemicals in household products: problems with solutions. Environmental Management; 40(6):889-901.

Goldsmith et. al. Summary Report for *"Personal Chemical Exposure Informatics: Visualization and Exploratory Research in Simulations and Systems (PerCEIVERS)"* June 26,27- 2012, **EPA report EPA/600/R13/041**

Gosselin R. E., Smith, R. P., Hodge, H. C. The Clinical Toxicology of Commercial Products. Wilkins & Wilkins, 1984.

Hertwich EG. Life cycle approaches to sustainable consumption: a critical review. Environ Sci Technol. 2005 Jul 1;39(13):4673-84.

Jaga K, Dharmani C., Sources of exposure to and public health implications of organophosphate pesticides, Rev Panam Salud Publica. 2003 Sep;14(3):171-85.

Jayjock MA, Chaisson CF, Franklin CA, Arnold S, Price PS. (2009) Using publicly available information to create exposure and risk-based ranking of chemicals used in the workplace and consumer products. J Expo Sci Environ Epidemiol.;19(5):515-24.

Judson RS, Martin MT, Egeghy P, Gangwal S, Reif DM, Kothiya P, Wolf M, Cathey T, Transue T, Smith D, Vail J, Frame A, Mosher S, Cohen Hubal EA, Richard AM. (2012) Aggregating Data for Computational Toxicology Applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System. Int J Mol Sci.; 13(2):1805-31.

Klepeis NE, Nelson WC, Ott WR, Robinson JP, Tsang AM, Switzer P, Behar JV, Hern SC, Engelmann WH. The National Human Activity Pattern Survey (NHAPS): a resource for assessing exposure to environmental pollutants. J Expo Anal Environ Epidemiol. 2001 May-Jun;11(3):231-52.

Lai AC, Thatcher TL, Nazaroff WW. Inhalation transfer factors for air pollution health risk assessment. J Air Waste Manag Assoc. 2000 Sep;50(9):1688-99.

Little JC, Weschler CH, Nazaroff WW, Liu Z, and Cohen Hubal EA (2012) Rapid Methods to Estimate Potential Exposure to Semivolatile Organic Compounds in the Indoor Environment. Environmental Sci. and Tech. 46L 11171-11178.

Meyer, Earl C. and Sharon K. Slick (2001) Consumer and industrial goods. In: Kaliski, Burton S. (Ed.), Encyclopedia of Business and Finance. New York: Macmillan Reference USA.

Morgan, M. K., L. S. Sheldon, C. Croghan, J. C. Chuang, C. Lyu, N. K. Wilson, R. A. Lordo, M. C. Brinkman, N. Morse, Y. L. Chou, C. Hamiltin, J. K. Finegold, K. Hand, AND S. M. Gordon. A PILOT STUDY OF CHILDREN'S TOTAL EXPOSURE TO PERSISTENT PESTICIDES AND OTHER PERSISTENT ORGANIC POLLUTANTS (CTEPP). U.S. Environmental Protection Agency, Washington, DC, EPA/600/R-04/193 (NTIS PB2006-101286), 2005.

Nazaroff WW, Weschler CJ, Little JC, Cohen Hubal EA (2012) Intake to Production Ratio: A Measure of Exposure Intimacy for Manufactured Chemicals. Env. Health Persp. 120: 1678-1683.

NRC (National Research Council) 2012. Exposure Science in the 21st Century: A Vision And A Strategy. Washington, DC: National Research Council. Available: http://www.nap.edu/catalog.php?record_id=13507

Organisation for Economic Co-operation and Development (OECD) (2012). Crosswalk of Harmonized U.S. - Canada Industrial Function and Consumer and Commercial Product Categories with EU Chemical Product and Article Categories. Environment Directorate, Series on Testing and Assessment No. 167. ENV/JM/MONO(2012)5.

Pellizzari, E. D., Perritt, R. L., & Clayton, C. A. (1999). National human exposure assessment survey (NHEXAS): exploratory survey of exposure among population subgroups in EPA Region V. Journal of exposure analysis and environmental epidemiology, 9(1), 49.

Price, Paul S., John S. Young, and Christine F. Chaisson, 2001, Assessing aggregate and cumulative pesticide risks using a probabilistic model. The Annals of Occupational Hygiene 45, Supplement 1(0):S131-S142

Rudel, R. A.; Camann, D. E.; Spengler, J. D.; Korn, L. R.; Brody, J. G. Phthalates, alkylphenols, pesticides, polybrominated diphenyl ethers, and other endocrine-disrupting compounds in indoor air and dust. Environ. Sci. Technol. 2003, 37 (29), 4543-4553.

Sanderson H, Greggs W, Cowan-Ellsberry C, DeLeo P, Sedlak R. Collection and Dissemination of Exposure Data Throughout the Chemical Value Chain: A Case Study from a Global Consumer Product Industry; Human and Ecological Risk Assessment: 19(4) 2013.

Schettler T. Human exposure to phthalates via consumer products. Int J Androl. 2006 Feb;29(1):134-9.

Shin H-M, McKone TE, Bennett DH (2012) Intake Fraction for the Indoor Environment: A Tool for Prioritizing Indoor Chemical Sources. Environ. Sci. Technol. 46: 10063-10072.

Smith, Ray. "An overview of the Tesseract OCR engine." In Document Analysis and Recognition, 2007. ICDAR 2007. Ninth International Conference on, vol. 2, pp. 629-633. IEEE, 2007.

Wallace, L., The Total Exposure Assessment Methodology (TEAM) Study: Summary and Analysis: Volume I, EPA/600/6-87/602a.

Wambaugh J.F., Setzer R.W., Reif D.M., Gangwal S., Mitchell-Blackwood J., Arnot J.A., Joliet O., Frame A., Rabinowitz J.R., Knudsen T.B., Judson R.S., Egeghy P., Vallero D.A., and Cohen Hubal E.A., (2013) High Throughput Models for Exposure-Based Chemical Prioritization in the ExpoCast Project, Environ. Sci. Tech, *in press*

Wenger Y, Li D, Jolliet O (2012) Indoor intake fraction considering surface sorption of air organic compounds for life cycle assessment. The International Journal of Life Cycle Assessment 17 (7): 919-931.

Weschler C.J., Nazaroff W.W. (2008) Semivolatile organic compounds in indoor environments. Atmospheric Environment 42 9018–9040

Weschler, C. J., Changes in indoor pollutants since the 1950s, Atmos. Environ., 43, 156-172, 2009.

Zartarian, V. G., et al., Quantifying Children's Aggregate (Dietary and Residential) Exposure and Dose to Permethin: Application and Evaluation of EPA's Probabilistic SHED-Multimedia Model. Journal of Exposure Science and Environmental Epidemiology 22(3):267-273, 2012.