Toward Safer Products:

Accelerating Change with a Chemical Hazard Data Commons

By Tom Lent
Contributors: Akos Kokai, Ann Blake, Bill Walsh, Larry Kilroy, Michel Dedeo

Originally published February 28, 2014
Revised June 23, 2014

This white paper is informed by the ideas and activities of many of the people and organizations convened in the *Building a Chemical Commons Conference* in October 2012 and a wide range of discussions in and out of the Data Commons working group formed out of that conference.

More information about the Chemical Hazard Data Commons Project is available at
https://commons.healthymaterials.net

Thank you to the Forsythia Foundation for their generous support for this project.

Table of Contents

The Vision

This white paper proposes a plan to create the Chemical Hazard Data Commons Project (hereinafter the Data Commons). The Data Commons is both a tool building project and a community building project. It envisions the creation of an Open Innovation platform that will enable the convening of a global community of people and organizations dedicated to sharing information about chemical hazards. The goal of this effort is to accelerate the development and commercialization of products and materials that eliminate hazards and reduce the overall demand for hazardous chemicals.

The Data Commons represents a breakthrough, using the capabilities of new information technology and the concepts of the "sharing economy," to solve one of the most significant impediments to reducing the use of hazardous substances: ineffective access to information. Ineffective access to chemical information is frustrating efforts to understand the impact of chemicals on humans and the environment, impeding research into alternatives, limiting public knowledge and understanding, and stifling innovation by commercial chemical users caught between the demands of an increasingly wary customer base and the resistance of their chemical suppliers. The end result is that society is generally unaware of potential adverse health impacts of the chemicals that are ubiquitous in 21st century life, or the opportunities to commercialize superior alternatives.

In so many fields new breakthroughs are transforming, almost overnight, information inefficiencies through sharing. Just as the power of sharing applications, such as Uber and Lyft, are solving the problem of "no taxi is available during a downpour," the Data Commons will help researchers, policy makers and private sector decision makers discover information about chemical composition and hazards that was not known to be available, but was, in fact, there all along. Just as websites such as TripAdvisor enhance expert information sources with "destination advisors" – informed community members with a demonstrated level of reliable knowledge – the Data Commons will invite robust, real-time, nuanced, multi-disciplinary contributions to an information field that is currently siloed and inaccessible. Just as the open source software movement completely reinvented and democratized access to information technology through online communities such as GitHub,[1] the Data Commons will reinvent and democratize access to chemical data. All of this will be done with the goal of eliminating human health hazards, reducing the overall demand for hazardous chemicals, and accelerating the transformation of the chemical industry to green chemistry.

The vision of the Data Commons is that anyone connected to the Internet will be able to find the most up-to-date information that has been made public about any chemical. They will also be able to join a global community of content experts who are dedicated to making this information actionable, filling data gaps, asking and answering questions and solving problems. The problems that can be solved with this approach can range from the most elemental, such as identifying substances that are misleadingly or inaccurately named, to the most advanced, such as accessing, interpreting and acting on emerging science, in advance of regulatory action, to protect human health.

The other side of this problem – and the one that is more difficult – is information that is confidential, largely due to the long-standing business practice of asserting broad claims of proprietary information. These practices are in rapid decline as leading businesses in many sectors view transparency as the "new normal," inevitable in the information age, and a leading indicator of sustainability. The Data Commons

[1] GitHub is a web-based hosting service for software development widely used to manage open source software development and other collaborative projects.

will accelerate the pace of this change, by supporting companies that share its values with a global team of experts eager to support efforts to reduce demand for hazardous chemicals.

Open Innovation Can Accelerate Innovation toward Reduced Hazard

It is now well documented that industry can, in fact, accelerate innovation radically faster by using an Open Innovation model[2] that is aligned with and driven by customer demands. In this sense the Data Commons will serve as a "virtual R&D" asset for any company committed to hazard reduction in its products and operations. No matter how big, chemical companies will not be able to capture and control the "virtual R&D" capacity of the Data Commons to find such solutions, any more than Microsoft was able to keep pace with open source software development. The Data Commons will provide the platform and tools necessary to drive and accelerate innovation, competition, and hazard reduction through the Open Innovation process.

To accomplish these goals, the Data Commons will begin by creating a portal to the most up-to-date, global information regarding chemical health hazards, and alternatives research and assessments. It will connect those who design chemicals and assess their hazards to those who make key decisions about products that incorporate these chemicals in commerce. It will shrink the time it takes to translate hazard data into actionable information for decision makers in both the private and public spheres. It will provide researchers and chemical designers guidance on where key social needs remain to be addressed. The result will be that decision makers will be better informed about the chemicals in the products we specify, manufacture and use. This will provide louder and clearer market signals to corporate planners, investors, innovators, entrepreneurs and ultimately policy makers. It will enable us to better align these activities with critical public health goals – reductions in cancer, asthma, reproductive disorders and impaired neural development.

This vision cannot be achieved by the efforts of any one agency or organization. The Data Commons will integrate and enhance the efforts of many organizations and individuals using proven collaborative strategies and methods adapted from open innovation business models and open source software development. With a practical goal of building the platforms and applications to enable this collaboration, the Data Commons project will provide a means of focusing and prioritizing the many useful proposals for improving data gathering, analysis, and sharing.

This proposal synthesizes the thinking and discussion among scores of individuals in organizations already committed to this path forward. It is based on a review of chemical information resources available on the internet, most of which are currently isolated, but which we envision as a network of linked *information nodes*. It identifies data types of interest and assesses critical challenges to accessing and using that data.

These consultations revealed a number of challenges, including differing perspectives within our community that require harmonization in order to realize the full potential of the Data Commons. Some of these issues will require resolution before we can proceed with constructing certain elements of the Data Commons infrastructure efficiently. Indeed, we envision that the Data Commons itself will facilitate discussion and resolution of these differences over time. Other differences, not being critical to the initial information systems architecture proposed for the first phase, can continue to be resolved while we begin constructing several fundamental components of the Data Commons.

[2] The Open Innovation model is discussed on p. 20

From our discussions of these issues we have distilled a series of recommended actions that can be taken now to dramatically improve our ability to share and utilize chemical information within one year from project launch, and establish an infrastructure that can support the collective vision of the Data Commons as it evolves for the long term.

Summary

In the remainder of this paper we outline in more detail how an increasing move toward transparency of product content is leading to a rising need for reliable, accessible chemical hazard information. We describe the steps of hazard-based assessment and the obstacles to obtaining the information needed to reliably guide development of safer alternatives. We describe how the proposed Data Commons can address each of the key obstacles and how collaboration and crowdsourcing efforts can help. We describe how a diverse range of potential participants can benefit from the work of the Data Commons. We describe the different data groups needed to make up the Data Commons and outline how the Data Commons proposes to tackle some specific challenges with targeted solutions. These include:

- **Hazard Assessment Exchange** to facilitate licensed access to hazard assessments in order to increase production and availability of these essential analyses
- **Hazard Data Portal** to aggregate the underlying data needed to do these assessments efficiently, consistently and effectively
- **Hazard based Substance Identification Registry** to improve accuracy of identification of chemicals and appropriate association with hazards
- **Functional Use Categorization Registry** to accelerate innovation by supporting searches for safer alternatives across industry boundaries.

The Data Commons vision is ambitious, but can be done in a phased six-year development plan outlined in detail. The major elements are:

- Phase 1 / Year 1
    - Build and pilot the Chemical Hazard Assessment Exchange
    - Build and pilot the Hazard Data Portal, providing single portal chemical hazard information search across select sources
    - Build and pilot a multi-generational process chemistry repository
    - Build and pilot the Hazard Identification Registry utilizing crowdsourcing tools
    - Establish four working groups to address priority challenges
- Phase 2 / Years 2 & 3
    - Expand the Hazard Data Portal to search more chemical hazard information sources
    - Develop a life cycle chemical repository to address multiple process chemistry paths and degradation products.
    - Build and pilot a functional use multi-parameter categorization registry to support searches for alternatives across industries.
    - Build facilitated discussion areas for chemicals
    - Expand the chemical identification crowdsourcing tools in the Identification Registry
    - Add more working groups
    - Develop governance & revenue models
- Phase 3 / Years 4 to 6
    - Develop full crowdsourced aggregation of data
    - Provide full alternatives assessment support, including functional design alternatives in addition to chemical and material replacement
    - Create integrated searches of science literature across open publishing platforms
    - Build data push tools for notifications of new science and assessments
    - Map data gaps & testing needs to inform and prioritize research agendas

This project is built upon detailed research underway assessing a wide range of information nodes. The

appendix lists the 160 web based resources that we have studied to date that contain useful chemical hazard information. Nine are proposed as the high priority nodes to incorporate in Phase 1 of the Chemical Data Commons and another 44 are targeted to add in Phase 2.

Thanks to our Reviewers

This white paper is a high level summary of a series of much more detailed Working Papers written to more fully describe some of the key challenges that we see impeding the success of chemical hazard assessment and proposing approaches to their solution. Fifty-six people from a wide range of academic institutions, private companies, government agencies and nonprofit organizations generously provided review and commentary on the preliminary drafts of these documents.

Comments were submitted to the authors by email commentary, track changes on the draft documents and through a web survey. The comments and survey results reinforced the need for this effort and helped reframe and identify issues not initially addressed. Some of those comments are reflected in this draft of the White Paper and the accompanying Working Papers while others remain to be addressed in future working group activities. This is only the beginning of the collaborative process. We anticipate that these Working Papers and the comments they generated will help shape the first work of the working groups proposed in this document. To read the Working Papers and the comments generated so far, and to learn more about the future of the Chemical Hazard Data Commons, visit http://commons.healthymaterials.net

Successful industrial transformation towards the manufacture of products with inherently less hazardous ingredients will require much more efficient, affordable, effective, and consistent tools for assessment - tools that harness the collective expertise of many individuals and organizations.

The Opportunity: A Rising Demand for Chemical Hazard Information

Increasing public awareness of - and desire to avoid - the potential hazards of exposure to chemicals of concern through everyday products has triggered movements in numerous sectors to reduce these hazards. A variety of industry, NGO, and governmental collaboratives are working to identify chemicals of concern, and encourage reformulation to avoid their use in building products, electronics, textiles, clothing, and other consumer products.

These trends are converging to create, for the first time, a significant and rising demand for chemical hazard information – information which is vital to efforts to both eliminate and substitute chemical hazards. For reasons further outlined in this paper, access to such information has been inadequate to meet the needs of hazard avoidance research and development efforts. As demand for this information grows, the opportunity is arising to reconceive the way in which chemical hazard information is developed, maintained and accessed – on a global basis.

To develop the needed solution, it is important to understand how hazard avoidance based research is conducted.

1. **Ingredient Inventory** – Ingredients in a product (or process[3]) are identified.
2. **Chemical Hazard Assessment (CHA)** – Potential hazards and data gaps of the inventoried ingredients are identified and prioritized.[4]
3. **Alternatives Assessment (AA)** – Alternative ingredients or designs are identified that can perform the same function with lower inherent hazard.[5]

Ingredient Inventory

A number of tools have been developed to facilitate ingredient inventory. In the building industry, the Pharos Project supports public disclosure in product ingredients for an increasing number of products. The Health Product Declaration (HPD) has created a standard format for this product content and hazard communication, which is rapidly accelerating disclosure in the building industry. Good Guide, Skin Deep and the Household Products Database represent a growing wave of tools that catalog

[3] The language in this paper primarily focuses on product inventory and assessment but the data and analysis can be applied to assessment of intermediate processes as well.

[4] Prioritization may take into account relative hazard and relevant pathways of exposure although quantified risk assessment is avoided. Hazard assessment should also account for the synergistic effects of exposure to multiple chemicals, however little data exists to support mixture synergy assessment

[5] This is a dramatic simplification of the alternatives assessment process that may involve a substantially more complex trade off analysis accounting not only relevant pathways of exposure but also other non-hazard factors.

ingredients in consumer products. The International Material Data System (IMDS) tracks the content of the thousands of materials and products used in the automotive industry. Cradle to Cradle™ certification provides third party supply chain verification in many different product industries under non-disclosure agreements.

Progress in ingredient inventory is uneven across industries. Many sectors have yet to establish disclosure standards. Even where standards exist, lack of communication along the supply chain, from supplier to assembler to manufacturer and retailers, can hamper transparency. Inadequate understanding of process chemistry and residuals can lead to inventories that overlook the chemicals used in manufacture that can leave critical contaminants in the final product. Nonetheless, exciting inventory projects are beginning to successfully progress from niche application to larger industry adoption in some sectors. As these inventory barriers are addressed, the opportunity and demands for robust hazard and alternatives assessment rise.

This paper focuses on the task of building on these inventories to create effective hazard and alternatives assessments that can drive changes towards safer product formulation.

Chemical Assessments

Once products are inventoried, the assessment process for the ingredients can be daunting. The hazard information needed for assessment is distributed across disparate sources with varied data formats, protocols, and reliability. Significant data gaps exist. Assessment tools are being developed and improved, but research and interpretation of the data requires considerable expertise. Chemical hazard assessments that can translate this information for decision-making by the non-toxicologist are difficult to find and plagued by data gaps and inconsistencies. Assessors need better access to hazard information and support for a more collaborative approach to do assessments more thoroughly, consistently and cost effectively. Product decision makers need access to these assessments to make effective specification and purchase decisions.

Alternatives Assessment

After determining what chemicals have hazards that warrant their replacement, finding functional alternatives with sufficient data to demonstrate lower hazard is even more challenging. This in turn raises the risk of “regrettable substitution” – replacing a known chemical of concern with a similar chemical with little data indicating harm but which turns out to have significant hazard once studied. Product designers need support to assist their search for a broader set of functional alternative chemicals with proven lower hazard – and to find design alternatives that can avoid the need for the chemical function. These problems will not be solved by improved tools alone. Significant collaboration will be required to research, analyze and effectively communicate hazard data and utilize this information to find the most effective alternatives.

The Chemical Hazard Data Commons Project

This paper proposes a Chemical Hazard Data Commons Project (Data Commons) as the new, global means for developing, maintaining access to chemical hazard data and applying it to alternatives assessment. Inherently, as an open information platform, it will broaden collaboration among the community of researchers, and even more importantly, make possible collaboration where it is, today, impossible. The Data Commons will accelerate the development of ingredient inventory tools and avoid

inconsistent, overlapping and redundant tool development. Greater collaboration will also improve access to and the quality of the data required for actionable hazard assessments that can inform business decision-making and public policy development. The Data Commons will foster the expert, interdisciplinary collaboration necessary to conduct complex assessments and make them useful and actionable. Actionable assessments have greater business value, creating better economic incentives for companies to undertake assessments and add them to the growing library of assessment data.

The Chemical Hazard Data Commons Project comprises two major components:[6]

1. **Chemical Hazard Data Commons Platform** – a set of internet-based tools that facilitate access to the information in the Data Commons, and support collaboration to improve, link, manage, summarize, and discuss this information as a community. The Platform will not be the single repository for all chemical hazard information, but rather will serve as a portal linking resources that are housed and managed elsewhere, wherever possible, and only provide a primary repository for data where needed.
2. **Chemical Hazard Data Commons Community** – a community of information users, bound by common principles, who will collaborate to support the development, improvement, and assessment of chemical hazard information, and the improvement of the Platform, toward the goal of chemical hazard reduction, safer and informed substitutions, and over the long term, the widespread commercialization of green chemistry.

Throughout this paper the term "Platform" will refer to the specific online tools that comprise the Data Commons, "Community" will refer to the users who are engaged in working with the information and the tools, and "Data Commons" will refer to the entire project.

Addressing the Three Steps of Hazard Assessment

The Data Commons platform and community will improve key elements of the three step process for addressing hazard concerns:

1) Ingredient Inventory – Assessing product (or process) content requires accurate ingredient information. Ingredient information reported by manufacturers is often incomplete, particularly when a product contains ingredients that were not intentionally added or specified by the final manufacturer.

- **Ingredients not specified:** The complete ingredient list for a product (or process) can be hard for a manufacturer to determine when long supply chains mask the identity of those ingredients. Ingredients in a final assembly may include ingredients not specified for the product by the manufacturer but added intentionally by a supplier several tiers up the supply chain (example: FRs added to foam padding when it is not required or specified by the final product manufacturer). The Data Commons can help identify ingredients commonly added to materials and provide hazard information to projects that inventory supply chain ingredients.[7]

[6] Many of the concepts for organizing and managing this information that form the core of the Data Commons have been well characterized by Akos Kokai in "Open knowledge management for chemical hazard assessment & alternatives analysis" Version 0.3, 2014-01-31 http://www.ocf.berkeley.edu/~akokai/organizing.html .

[7] The Data Commons Community may include projects like Material IQ who inventory complex supply chains. It is

- **Ingredients not intentionally added: Residuals** such as unpolymerized monomers that remain in the product from the manufacturing process, and impurities or contaminants from the original resource extraction process, may pose significant health hazards, but may be unrecognized in the inventory of intentional ingredients. The Data Commons will host a library of process chemistry for chemicals and materials and diagnostic testing results to help point to the likely residuals resulting from different processes.

2) Chemical Hazard Assessment – Identifying and prioritizing the potential health hazards of the chemicals in a product can be split into two levels of assessment:

- **List screening** compares chemicals against authoritative hazard lists to identify the associations between a substance and health or environmental endpoints such as cancer, skin irritation, or aquatic toxicity. Aggregated searching can help speed up this process by combining dozens of lists for searching, reducing the cost of research and increasing consistency of results.[8]
- **Full assessments** of the scientific literature across a predefined broad range of health and environmental endpoints are necessary to fill the large gaps in health endpoints left by List Screening[9], to address life cycle issues such as potential transformation or breakdown products, and to identify exposure issues that may need to be assessed to prioritize between chemicals of comparable hazard and to affirmatively identify preferable chemicals.[10] The Platform can help facilitate the time-intensive research function for science literature and analogs and models to address data gaps. By providing a common base for this research it can also increase consistency of assessments. By facilitating peer review, the Community can in turn improve data quality and increase credibility. The Data Commons can also help identify data gaps that stymie current hazard assessment and provide a roadmap and prioritization for research.

3) Alternatives Assessment – In this process the assessor seeks alternative ingredients, products, or designs that have similar functions but with lower hazard assessments and identifies trade-offs. The Data Commons can help assessors find potential functional alternatives from different industries along with key hazard information.

Obstacles to Accessing Hazard Information

Designing and selecting healthy products is stymied by the difficulty and expense of obtaining reliable, consistent, and easily understandable information about the hazards that are associated with certain chemicals, not to mention the hazards of alternatives that may be available. Chemical hazard information is split among many public and private databases and individual documents. Appendix 1 provides a list of *information nodes* – resources that provide data that could be useful to chemical

---

not currently anticipated that the Portal, however, will be a repository of proprietary materials.

[8] Pharos currently provides this service for all lists in the GreenScreen List Translator plus additional hazard lists, restricted substance lists (RSLs) and endangered wood species lists. HBN will integrate this list screening with other Data Commons functions through Pharos.

[9] Only a relatively small number of the total quantity of chemicals in commerce have been evaluated by authoritative agencies for these listings and only for some of the possible health endpoints.

[10] List screening can only identify chemicals with a known high hazard in one or more endpoints. Confirming that a chemical is low in hazard requires a toxicological assessment of the scientific literature across all of the relevant health and environmental fate endpoints.

hazard analysis. A number of additional systemic factors present major obstacles:

- **Formats**: Many resources provide data in inconsistent formats that are not structured for easy access, either for a person doing a manual search or for machine processing or automated access.

- **Access policies:** Some databases provide open data that is freely or inexpensively accessible, while other others contain proprietary information and are only accessible to a limited group of people behind expensive paywalls[11], or are considered trade secrets. The expense of producing and curating this data often leads to expensive paywalls, leading wasteful redundancy as more sources put more effort into producing and curating the same data separately.

- **Data presentation and interaction:** Users have different skill levels for interpreting data and need different types of information and levels of aggregation; they therefore need different types of information presentation and interface tools.

- **Chemical Identification:** Ambiguities in naming and identifying chemicals make it difficult to reliably and consistently associate the right chemicals with the right hazard information. For example, a chemical might be referred to by over 50 different chemical names, trade names, or numeric IDs. Systems like the CASRN[12] system are proliferating identities, minting thousands of new identifiers for substances that may never see the outside of a laboratory.[13]

- **Data strength, quality and bias**: Different datasets may have different strengths and limitations that are not easily identified but critical to translating it into usable hazard information and/or identifying data gaps. These can include size of the samples, weight of the evidence and funding source bias.

- **Confidential business information (CBI)/Intellectual Property (IP)**: intentional obscuring of chemical identities in association with hazard assessment information to protect confidentiality can render listed hazard information useless or irretrievable.

All of this makes doing research on chemical hazards and their alternatives time- and labor-intensive, expensive, and inconsistent. Furthermore, updating previous research can be just as resource intensive as the initial research. Lack of interoperable systems that share data means that researchers and assessors do redundant work, further increasing costs and reducing efficiency.

High cost and inconsistent assessment results can discourage engagement and slow the innovation necessary to identify hazards and find safer alternatives. This environment is ripe for improvement with the application of modern data technologies and a collaborative information system design.

Overcoming these Obstacles

The Data Commons aims to overcome the obstacles described above by improving access to and

[11] A paywall is a system that prevents Internet users from accessing web content without a paid subscription.
[12] The Chemical Abstract Service Registration Number (CASRN) is the most widely used identifier for chemicals in industry.
[13] The CAS registry contains “more than 81 million organic and inorganic substances” and “Approximately 15,000 new substances are added each day” https://www.cas.org/content/chemical-substances/faqs#q6

consistency of chemical hazard and alternatives data, and making it faster and easier to obtain and compare. This would be achieved by using web-based tools to facilitate data sharing across platforms. The Commons Community will use these tools to help fill data gaps and make the data more widely available, consistent, understandable, and actionable. This collaborative effort will produce better results at less cost than individual research to:

- **Consistently identify chemicals and link them with hazards** despite use of different CAS numbers, names, or trade identifiers.
- Provide access to information about the hazards of chemicals in consistently **structured data formats**.[14] This facilitates both single chemical manual searches by an individual and automated bulk searches by a computerized system.
- Identify **related chemicals** throughout the chemical/material lifecycle: feedstock contaminants, process chemicals, manufacturing contaminants and transformation products.
- Assist the process of consistently assessing and prioritizing hazards by **supporting transparent, well defined, public protocols**.
- Help share assessment results to reduce redundant assessments.
- Provide opportunities to make the assessment process more transparent and robust by **facilitating peer review** and helping fill data gaps.
- Link chemicals to information about how they can be used ("**functional use**") to help the search for safer alternatives.
- Help users **track emerging science** or innovation to be informed of new developments that may change the identity, hazard, or functional use information used in their assessments.

Harnessing the Efforts of Many: A GitHub for Hazard Assessment

The ambitions of the Data Commons cannot be met by the efforts of any one agency or organization. Many resources exist that already serve parts of these needs, but they are not well synchronized to create the necessary whole. The goals of the Data Commons Project can only be met by integrating the efforts of many organizations and individuals to take advantage of existing resources and realize greater synergies. The Data Commons will improve and enhance the work of participating organizations, and extend their influence to a larger global audience.

Automated integration of data from many websites will be an important attribute of the Data Commons platform, but it will only go so far to meet the needs of decision-making based on hazard assessment. Expert judgment is also needed to make these resources useful. The Data Commons will harness the power of this scattered institutional knowledge through carefully curated crowdsourcing methods that are a proven strategy for building complex yet reliable systems.

Crowdsourcing has been used successfully in the open source software movement to organize and incentivize coders around the world to collaboratively build powerful tools such as the Android mobile phone operating system and the Firefox web browser tools. There is a growing literature about the applicability of these concepts to the development of scientific knowledge and how to curate it to insure credibility and quality.[15] We propose to build crowdsourcing capabilities into the Platform, both in

[14] "Structured format" means that instead of the information being embedded in open, unorganized text it is formatted in a consistent predefined way so that the user – or another computer system – knows exactly where to find each type of information and how it is defined.

[15] For example, Marcio von Muhlen, "We Need a GitHub of Science". There are also some interesting discussions in process on data transfer lessons as well. Cameron Neylon "GitHub for science? Shouldn't we perhaps build TCP/IP

GitHub-type project organizing and Wiki-type document development to support multi-organizational team efforts to collaboratively develop the Commons.

Potential Participants in the Data Commons

The Data Commons is designed to support a growing network of organizations and institutions that are engaged in chemical hazard assessment, actively seeking and assessing information in order to reduce chemical hazards in products.

Designers, developers and owners in the building industry are increasingly changing their **specifications** and other purchasing practices in an effort to design buildings and purchase products with disclosed ingredients and avoid chemicals of concern.[16] **Retailers** are establishing sustainable **product specifications for their purchasing programs,** including chemical hazard criteria for the consumer products they sell.[17] **Unions** and other worker support organizations are **researching chemicals** used in the workplace to understand potential hazardous exposures and engage in decision-making about which chemicals are used, what protective strategies are needed, and what alternatives are available.[18]

The Data Commons could provide these users the access they need to aggregated chemical hazard assessments and the tools to interpret them.

Non-profits are developing **voluntary screening programs** to help meet these user needs and to encourage disclosure of ingredients, hazards and selection of products that avoid chemicals of concern.[19] They are doing this through development of **protocols, selection tools, scorecards, and certifications** to help the groups listed above understand and digest the assessment information to encourage and support informed selection.

These programs provide tools that would be highly valuable to Commons community members, and conversely. The Data Commons could provide these non-profits with access to hazard assessments and the data that underlie them needed to make these tools more effective and accurate.

Manufacturers and brand owners are responding to all of this activity by beginning to **evaluate their products** and explore ways to **reformulate** them to avoid chemicals of concern.[20] **PLM/ERP software**

---

first?"

[16] The Healthy Building Network's Pharos Project, the rapid spread of usage of the Health Product Declaration and the USGBC's new credit in its LEED building rating program have played important roles in this change. Examples of organizations adopting this move include architecture firms such as HOK and Gensler, developers like The Durst Organization and large building owner/occupants like Google.

[17] Wal-Mart, Target, Whole Foods, Home Depot, and Staples are all prominent examples of this trend in retail.

[18] The BlueGreen Alliance has developed computer tools for its member organizations to use in workplace chemical use assessment.

[19] The Healthy Building Network's Pharos Project, Clean Production Action's GreenScreen, the USGBC's LEED system, Cradle to Cradle Product Innovation Institute C2C certification program, the BizNGO Plastics Scorecard and the International Living Future Institute Living Building Challenge are all examples of systems developed based on screening for hazard.

[20] Nike, SC Johnson and Levi Strauss are examples of brands with major hazardous ingredient reduction goals. Note that many of these companies are developing their own systems for categorizing hazard in the absence of consensus on a single system and common resources.

providers [21] are developing tools that give manufacturers the ability to collect ingredient info from their suppliers and perform analytics on such data (e.g. appearance on regulatory lists, carbon footprint info, etc.).

Through the Data Commons, manufacturer specifiers and their software providers can go well beyond the limits of regulatory compliance to preemptively identify emerging concerns about their ingredient lists, enabling them to prioritize the product inputs on which to focus substitution efforts ahead of regulatory action. Through the Data Commons manufacturers and their customers can have access to the same information and assessments, and work cooperatively on solutions. Functions or features built on the Data Commons platform could also provide alerts to let the manufacturer know when a new scientific study or a new authoritative hazard listing is published for a chemical in their product(s) that may raise the need for reassessment. Alternately it may help identify a new alternative for a functional use they are seeking to replace.

Product or material designers for those manufacturers are seeking to develop healthier products and formulations but need more ready access to hazard and alternatives information.

In the Data Commons they will be able to look up ingredients they are considering for a new product design, understand how well assessed those ingredients are, purchase assessments for unassessed chemicals, identify chemicals of concern, and search for alternative ingredients with similar functional use and better hazard profiles. The Commons will provide a place to share ideas and challenges with others trying to solve similar hazard and design problems. While focusing on chemical hazard, the Data Commons will also provide pathways to design strategies for approaching functional challenges as well, including other design partners like the Biomimicry Ask Nature database. Ideally, rather than being a standalone tool for the designer, the Data Commons would ultimately feed hazard information into other material selection tools that they already use[22], providing a richer, more complete and up-to-date data set for their toolset.

Assessment organizations are emerging to help manufacturers **inventory their supply chain** and do toxicological **assessment of their ingredient hazards** to support external disclosure demands and the internal need for alternatives assessment.[23] Assessments may be provided in the form of public third party certifications or internal consultation. Assessors currently spend too much time in research instead of assessment and optimization.

In the Data Commons, an assessor could look up the chemicals in the product being reviewed and quickly see which ones are on lists of chemicals of concern or have already been fully assessed. The Data Commons could also help with the search for scientific literature to fill in the gaps after an authoritative hazard list screening.[24] Chemicals that are not readily identified could be explored and better characterized through collaborative discussions. Commons tools and discussion areas could help link these chemicals to related chemicals that have been flagged for hazard. The assessor could then focus on the ingredients that are not yet well assessed, completing product assessment faster and more consistently. Interactive functions could support peer review of specific assessments and broader

[21] ERP (Enterprise Resource Planning) and PLM (Product Lifecycle Management) providers working on these issues include PTC, SAP, Rubali Dolphin, Wercs, and Material IQ.
[22] Granta Material Intelligence is one such tool.
[23] ToxServices, NSF, UL, PE International, MBDC, EPEA all provide hazard-based assessment services.
[24] Full searches in the open publishing literature, abstracts only in the subscription fee based journals

process discussions, improving the assessor experience. Ultimately applications built on the Data Commons platform could provide alerts to let the assessor know when a new scientific study or a new authoritative hazard listing is published that fills a data gap for a chemical that the assessor previously reviewed. We expect that these assessors will be some of the most active users and contributors to the Commons.

Government agencies also are developing both **voluntary and regulatory programs** to identify chemicals of concern and discourage their use in products.[25] While sharing the above needs, these entities may also need usage data and be important sources of authoritative hazard lists, compilations of emerging science literature and process chemistry descriptions. The functional use of chemicals in products will be very important to these users.

A policy analyst will be able to use the Data Commons to rapidly aggregate information from a diverse set of sources to identify lists of chemicals of concern to prioritize for policy action. The Data Commons can provide integrated access to hazard associations, functional use, and use volume data. Crowdsourced work on chemical IDs and processes can make sure that policies keep up with the proliferation of chemicals and synonyms and potential residuals. Applications built on the Data Commons platform could help keep the information used to implement the policies up-to-date and tuned to address emerging issues.

These **government agencies**[26] plus a**cademic institutions**,[27] **industry collaboratives**[28] and **NGOs** are also researching alternatives to targeted chemicals of concern in specific applications, and doing public health research. They will have similar needs to the agency users and also will in turn be important sources of hazard assessments for chemicals, functional use associations, emerging science studies, process chemistry descriptions and model development.

These public health researchers can use the Data Commons – both the data on hazard and functional use, and the interactive discussion elements – to identify key emerging issues and shape research agendas to be most relevant. Conversely, by plugging the results of their studies into the Data Commons and discussing them in the Commons interactive areas, researchers can ensure that their results immediately becomes available to and discussed by assessors and policy developers. Informing their health impact studies with functional use and use volume data can help give their research more relevance by emphasizing the potential impact of selecting or eliminating a specific chemical or range of chemicals.

Finally, academic or private **chemical researchers** need guidance from industry experience on the most

[25] The State of California, State of Washington, Massachusetts Toxic Use Reduction Institute, the US EPA, and the Northeast Waste Management Officials' Association all have developed programs to reduce use of hazardous chemicals.

[26] US Environmental Protection Agency's Design for the Environment (DfE) does explicit alternatives assessment. The National Institutes of Health and California's Office of Environmental Health Hazard Assessment, (OEHHA) are examples of national and state public agencies that are participating in public health studies of the impact of chemicals. This listing is very US centric. While the Data Commons is being initiated in the US, it will be important to engage and collaborate with partners in the EU and around the world who are doing important work in this space.

[27] For example, UMass Lowell and UC Berkeley

[28] The Green Chemistry and Commerce Council (GC3) and the Business-NGO Working Group for safer chemicals and sustainable materials (BizNGO) are two major collaboratives

fruitful high impact areas to study.

They will be able to use assessments and functional use information to identify product and chemical functions with significant uses, but for which all substitution options are hazardous. They can also use Data Commons discussion areas and assessment requests gathered through crowdsourcing to further clarify where research for new inherently safer chemical synthesis would be most useful.

Note on functions: Some functions, including most of the lookups described, are already available in some systems for more limited groupings of data, and could be aggregated to be more comprehensive through the Data Commons. Crowdsourcing functions could be overlaid for these lookups in a successive phase, while other functions, such as the alert functions proposed, are more aspirational and would become available in later generations of the Data Commons.

Data Groups Needed in the Data Commons

Each of the functions described above depends on some or all of a set of critical groupings of data:

- **Summary Information:** Policy makers and design decision-makers need rapid access to easy-to-understand summary information about chemicals and classes of chemicals. The Data Commons will help organize and facilitate development of and access to the following types of summary information:

    - **Authoritative Hazard Listings**: Governmental and professional organizations compile substance lists for which scientific literature has been assessed for associations between substances and one or more specific health or environmental endpoints. These lists establish specific threshold criteria for determining whether to list a chemical or a group of related chemicals. Screening against these authoritative hazard lists is a cost-effective prioritization strategy for targeting chemicals of concern for replacement, and is frequently the starting point for full assessments of chemical hazard. The GreenScreen for Safer Chemicals' List Translator[29] is an example of a screening protocol that addresses how to utilize a large number of authoritative hazard lists to do an initial prioritization screening. Automated tools such as those developed by Pharos to screen a chemical against dozens of lists at once can greatly speed up an assessment process.

    - **Chemical Hazard Assessments**: Chemical hazard assessments generally start from the authoritative list screening described above. The list screening, however only provides information about a small number of endpoints of the many of potential concern. This is a good starting point for prioritization and will identify relatively well known chemicals of concern but cannot confirm the safety of a chemical. Lack of an authoritative hazard listing does not equal low or no hazard. A full chemical hazard assessment requires doing deeper research into the scientific literature to find evidence for the level of hazard in other endpoints not addressed for this substance by the single endpoint authoritative hazard lists. Modeling data or study of close analogues may be used to fill in the data gaps where scientific literature specific to the substance is missing.

[29] The GreenScreen for Safer Chemistry is a protocol for assessing the scientific literature about a chemical on 18 human health and environmental health endpoints against set criteria and determining a hazard benchmark on a 4 level scale. www.greenscreenchemicals.org

These full assessments are critical to help manufacturers understand the hazards of substances in use and switch away from use of the highest hazard substances towards ingredients that are inherently safer. However, the assessments are hard to find and expensive to commission. The Data Commons Project is exploring several different models for a Chemical Hazard Exchange all aimed at creating a marketplace to distribute the cost of assessments across more manufacturers, lowering the cost to each manufacturer and helping assessors establish a business model that will cover their costs and incentivize the production of more assessments.

- **Data gaps** will continue to be a challenge for doing an assessment with certainty. Identifying data gaps found in full assessments can help the user understand the limits of the analysis while identifying agendas for research agendas to fill gaps. Meanwhile, developers of assessment protocols like the GreenScreen need to address the challenge of making informed decisions with insufficient data.

- **Alternatives Assessments**: Chemical hazard assessments may be followed by alternatives assessments in which a chemical or group of chemicals of concern are studied to find alternatives with a lower hazard profile[30]. This is often, though not necessarily, specific to the function that chemical or chemical group plays in certain product types. The alternatives reviewed may include everything from drop-in chemical substitutes to entirely different product designs that make the higher hazard chemical unnecessary. Sharing these assessments is critical to jumpstarting innovation and ending duplicative research.

- **Life cycle studies**: Surveys of health and environmental effects in the life cycle of a chemical or group of chemicals – from the extraction, manufacture, use, and disposal or recycling are also important contributors to policy efforts. These may include studies of occupational exposures, releases in fenceline communities,[31] biomonitoring, public health surveillance, and environmental monitoring. Cross referencing chemicals to these studies has been difficult and could be greatly facilitated by a Data Commons repository cross referenced with other hazard information.

- **Restricted Substance Listings**: Governmental agencies, for-profit companies, and non-governmental organizations each may publish lists of chemicals or groups of chemicals targeted for action based upon their assessment of a range of health, environmental and other factors. The action may be a regulation, purchasing specification, research agenda, reporting requirement or other. Regardless, access to these lists can be helpful for others attempting to serve this market or trying to establish complementary policies.

- **Assessment Summaries**: The Data Commons proposes to support decision makers and policy makers with both improved access to data of the types described above and to facilitate wiki style assessments and discussions of the policy issues raised. See the Data Commons Working Paper *Facilitating Decision Making with Assessment Summaries*[32] for further discussion of how the Data Commons can improve access to these assessments, including the proposals to

[30] Exposure differences may also be assessed where all options have similar hard profiles

[31] Fenceline communities refer to neighborhoods that adjoin industrial facilities (are at the "fenceline") and hence are likely to be the most immediately affected by to environmental releases to air or water from those facilities

[32] Lent, Tom, *Facilitating Decision Making with Assessment Summaries* is a Chemical Hazard Data Commons Working Paper in progress. Current draft is available at http://commons.healthymaterials.net/about

establish a Chemical Hazard Assessment Exchange

- **Key Underlying Data:** The assessments described above depend upon consistent and reliable access to a number of types of underlying data – and ways of addressing data gaps. The Data Commons will help organize and facilitate access to these kinds of data:

    - **Scientific studies**: Scientific studies (clinical, toxicological, epidemiological, environmental fate, modeling and analogue studies, etc.) underlie all chemical hazard assessments and lists. New studies are published constantly that may fill data gaps in assessments or provide new insight into previously listed substances. These studies, however, are scattered in many journals, most of which have high subscription fees, and few of which are indexed in a way that facilitates the searching needed for hazard assessments. Easily finding relevant studies is key to doing hazard assessments. The Data Commons can develop carefully curated crowdsourcing systems to help facilitate the indexing of emerging science relevant for searching.

    - **Estimation**: Modeling and analogs can be critical to filling data gaps when the scientific literature is lacking.

    - **Physical characteristics:** Physical characteristics are important for determining the applicability of alternatives. The hazard of a substance can change considerably with its physical form: a large solid versus small, inhalable particles, versus a nano scale form. Physical characteristics, such as vapor pressure or solubility, may in turn affect exposure to the hazard and be relevant for prioritization.

    - **Life cycle information**: Information about the chemicals and processes used to synthesize a substance, and potential contaminants generated at its extraction, is needed to anticipate toxicity issues for workers and fenceline communities. This information can also help predict what residuals might be present in the final product. It is also important to predict whether any degradation products might occur during use or disposal as end-of-life is a critical part of a hazard profile. Finally, capturing the results of field studies of environmental monitoring, surveillance and biomonitoring will help to complete the picture of how a chemical migrates from products into the environment and into people.

- **Hazard-Based Substance Identification**: Before linking a hazard is possible, a substance must first be correctly identified. Multiple schemas exist for chemical naming. CASRNs are used almost ubiquitously but do not always provide unique identifiers. Most chemicals that are actively used in industry have many synonyms and brand names and the field of synonyms can grow steadily. CASRNs also rarely address identification of groups of chemicals or different specific forms.

  The Data Commons will develop a Registry for a stable, open, unique ID system for chemicals (the Substance Hazard Identification Number) to facilitate the effective association of chemical identity and hazard information. The Registry will facilitate cross indexing between the various schemas and address some of the specific needs of hazard association such as identifying chemical groups and chemical forms and resolving some of the other gaps and ambiguities of the CASRN system. Carefully curated crowdsourcing can facilitate the constant updating needed to manage synonyms.

See the Data Commons Working Paper *Getting Hazard Right through Managing Chemical IDs*[33] for further discussion of types of chemical substance identifiers, factors that complicate substance identification for hazard assessment and the proposal for an Identifier Registry for the Data Commons.

- **Functional Use Categorization**: Information about hazards will help industries identify chemicals to target for replacement, but finding the right alternative can be tricky. The Data Commons can help product designers search for useful alternatives to hazardous ingredients by linking chemicals to both hazard assessments and the functions they perform in products (surfactant, carrier, etc.). Functional use characterization schemas have, to date, tended to be very industry specific one- or two-parameter lists. The Data Commons will develop a highly flexible multi-parameter system, supporting targeted searches to support the search for potential alternatives across industries and accelerate innovation. The Data Commons functional use parameter framework will at minimum include parameters at the chemical, material, application, and product levels. See the Data Commons Working Paper *Finding Alternatives through Categorizing Functional Use*[34] for further discussion of user needs from a functional use categorization schema, proposed parameters for alternatives assessment and an examination of current schema.

The inclusion of these elements into the Data Commons does not necessarily require creation of a single master database of all types of data for all substances. Most of these types of data are already housed and curated somewhere on the web for many substances. They are, however, scattered in a large variety of different databases in different formats with different types of search functionality and widely varying access criteria, with no way for integrated searching across the different systems.

The Data Commons project is studying the means by which different sites offer access to their data and how they can be made more consistent. We propose a phased process to consolidating a steadily increasing number of different data sources into one integrated search. See Appendix A for a survey of priority Information Nodes that we are studying for inclusion in the Data Commons functions. Ultimately truly meeting the needs of hazard assessment will require integrating not only data sources but also the scientific publishing industry. See the Data Commons Working Paper, "Speeding & Improving Assessment through Data Sharing"[35] for further discussion of types of chemical hazard data and how the Data Commons proposes to address them.

Data Commons Development Phases

Phase 1 / Year 1

Though we estimate that it will take at least four years to realize the vision of the Data Commons outlined in this white paper, there are at least four steps that can be taken right now that will, within the first year, significantly improve our ability to share and analyze chemical information, increase the

[33] Kokai, Akos, et al *Getting Hazard Right through Managing Chemical ID* is a Chemical Hazard Data Commons Working Paper in progress. Current draft is available at http://commons.healthymaterials.net/about
[34] Blake, Ann, *Finding Alternatives through Categorizing Functional Use* is a Chemical Hazard Data Commons Working Paper in progress. Current draft is available at http://commons.healthymaterials.net/about
[35] Lent, Tom, *Speeding & Improving Assessment through Data Sharing* is a Chemical Hazard Data Commons Working Paper in progress. Current draft is available at http://commons.healthymaterials.net/about

number and availability of chemical hazard assessments, build key elements of the Data Commons information technology infrastructure, and encourage participation in the Data Commons community. We are prioritizing these steps in order to leverage the increased market demand for access to full chemical hazard assessments that is being driven by the Health Product Declaration (HPD) and the new USGBC LEED material ingredient credit. These components of the Data Commons would be built on the existing Pharos Project Chemical and Materials Library (CML). These components will be opened as Pilot Projects (which may involve limited features or limits on participants) within one year of funding, with additional features and expanded participation in year two.

Additionally, we recommend funding for Data Commons community working groups that will begin to address critical substantive issues that need to be resolved in order to complete the information technology infrastructure.

1. Build and pilot the Chemical Hazard Assessment Exchange: The Chemical Hazard Assessment Exchange (hereinafter the Exchange) is a website that hosts full chemical hazard assessments. Centralized access to chemical hazards assessments will increase the demand for assessments that make it easier to complete HPDs, and make the LEED credit more easily attainable. Centralized access to chemical hazards assessment will also address market inefficiencies that are impeding their widespread adoption and use.

The site will be comprehensive, including both assessments that have been funded for open public access and privately commissioned assessments available for sale. Currently companies that commission assessments must fully fund the research of the assessors, and are therefore unwilling to share this information. By creating the Exchange, assessors will be able to develop a business model that reduces the price of assessments made available on the Exchange, and cover costs by reselling that information to multiple users.

Manufacturers who are seeking to characterize their products will be able to search for a chemical in a single Exchange portal provided by Pharos and find a range of useful results, including:

- GreenScreen List Translator results from Pharos
- Public GreenScreen full assessment results from Pharos and/or NEWMOA
- Public Design for the Environment (DfE) assessment results from EPA ChemView
- Private assessments for license from GreenScreen Profilers, Cradle to Cradle Assessors and possibly other related hazard profilers, such as SciVera offering its Dossiers.

We expect substantial impact from the Exchange on the new LEED credit rewarding use of products whose ingredients are GreenScreen assessed and avoid benchmark 1 chemicals. Attaining this credit is currently very difficult as there are very few products assessed for design teams to use. GreenScreen assessments licensed here can be immediately used to complete HPDs and to fulfill the LEED credit which we expect will dramatically accelerate uptake of the credit.

<u>Use of the Exchange, however, will not be limited to manufacturers, nor to building product related chemicals.</u> Assessors and profilers will be encouraged to post assessments for any chemicals they have assessed and all users will be invited to make requests for assessments for any chemical through the Exchange.

2. Build and pilot the Hazard Data Portal providing single portal chemical hazard information search:

This new tool will allow assessors to look up a chemical in Pharos and access information from multiple information nodes – web site collections of scientific literature and other hazard information that are required to be reviewed by GreenScreen or other assessment protocols. Currently, assessors must use a search engine and identify priority sources of information. With this search tool, only the desired priority information sources will be called up in the search. We will start with an integrated display of information from at least two nodes that provide API[36] access to their data, such as the Hazardous Substance Data Bank (HSDB) and TOXNET. For at least six other priority nodes for which API access is not yet available, but CASRN based URI[37] navigation is, the Portal will provide a targeted link to the searched chemical's page. These include the Aggregated Computational Toxicology Resource (ACToR), ChemIDPlus, and the NIOSH Pocket Guide to Chemical Hazards. This will dramatically facilitate assessments, putting a wealth of information at the researcher's fingertips, reducing costs of assessments and improving consistency and depth of the assessments.

3. Build and pilot a multi-generational process chemistry tool: Many chemical hazards associated with products and materials are hidden in their production processes. For example, a "formaldehyde free" product, may in fact involve the use of formaldehyde compound in its production process, exposing workers and fenceline communities to those hazards. Additionally, process chemistry leaves hazardous "residuals" that are not traditionally listed as ingredients, but which are present in the final product. Identifying these chemical processes is a tedious and complex research task. This tool will allow users to look up process chemical profiles in Pharos that reflect the complex multi-stage relationships of chemicals synthesis pathways to better understand manufacturing hazards and potential sources of residuals in final products. Users will also be able to contribute and share their research by building process chemistry profiles and adding them to the Pharos library.

4. Build and pilot the Hazard Identification Registry utilizing crowdsourcing tools: Perhaps the most exciting element of the Data Commons is the platform that enables community members to collaborate on solving challenges to understanding and using chemical information. One vexing challenge is the proper identification of chemicals which is necessary to insure proper assignment of hazards. This is of special concern to materials evaluation programs such as Pharos, GreenScreen, C2C, Declare and Material IQ because the use of unrecognized chemical names can mean that hazards that should be assigned to a material are not. Currently, few researchers have the incentive or resources to resolve these identification challenges. This tool will utilize crowdsourcing tools and techniques to reduce the burden of this research, while capturing and sharing the new information widely. Proper security, administration and curation of crowdsourcing platforms are essential to ensure the quality of the information. Participation during the Pilot Program will be limited so as not to create undue security and administrative burdens during the initial phase of use.

5. Establish working groups to address priority challenges: The respondents to the white paper have identified a series of issues that need to be addressed in order to realize the potential of the Data Commons. We propose to establish working groups so that the Community can work together to develop proposals addressing these issues. In the first year, we propose to initiate at least four working

[36] API stands for Application Programming Interface – a set of protocols that specify how two programs or computers can interact with each other. Within the Data Commons, APIs would allow computers to collect and process hazard information from multiple web resources at the same time.

[37] URI stands for Uniform Resource Identifier. Within the Data Commons, it means that a web resource's page addressing is organized to allow so that users can quickly and easily be directed to the information they are looking for.

groups in the first year, with more groups to be added to expand the issues addressed in future years. Membership and leadership in these working groups will be drawn from the leading academic institutions, NGOs and assessment organizations engaged in the process.

Priority issues for working groups include:

- **Chemical identifications**: refining unique chemical identification to address different form, population of compound groups, synonyms, etc.
- **Functional use parameters**: assessing, aggregating and extending functional use characterizations of substances
- **Life cycle chemistry**: developing repository of process chemistry paths and degradation products
- **Interoperability standards**: system to system communication, including improving authoritative hazard list access and substance identification
- **Curation of crowdsourced workflow**: developing collaborative information gathering and processing management and quality control
- **Hazard Assessments**: organizing, evaluating, and aggregating hazard assessments, both public and private and including setting up a marketplace for private licensed assessments
- **Emerging Data:** improving search and summarization of existing and emerging scientific studies
- **Governance & revenue:** developing methods of decision making and management and business plans.

Phase 2 / Years 2 & 3

During Phase 2 of the Data Commons project, our emphasis will be on refining and expanding the tools piloted during Phase 1; creating more tools to help assessors meet the demand for more assessment, rapidly, thoroughly and consistently; expanding and governing the Data Commons community; and developing revenue models to ensure the long term viability of the Data Commons. Key elements of Phase 2 will include:

- ***Expand the Hazard Data Portal to search more chemical hazard information sources:*** The single portal chemical search for scientific literature and other hazard information piloted in the first year targeting priority information nodes will be extended in the second year. Assessors and other users will be able to search all priority nodes that provide API or URI access options from one portal. We will strive for access to all nodes required by the assessment protocols.

- ***Develop a life cycle chemical* repository *to address multiple process chemistry paths and degradation products*:** The process chemistry tool piloted in the first year will be extended to allow users to use and help develop a repository of chemical inventories to understand the full impact of a chemical throughout its life cycle. This will include multiple process chemistry paths and degradation products as well as links to related environmental monitoring studies such as dust and biomonitoring studies. This data set will be built through a variety of means, including harvesting from assessments in the Exchange, funded studies, academic programs and crowdsourcing.

- ***Build and pilot a functional use multi-parameter categorization registry to support searches for alternatives across industries:*** Both chemical hazard analysis and alternatives assessment can vary with the functional use of the chemical in question. Functional use data, though

limited, does exist, however it is difficult to find, access and utilize. The functional use parameter repository will allow community members to research alternatives by searching by functional use as the Community tests the cross industry parameter set developed in the first year by the working group. The initial tagging of chemicals will build on existing functional use lists and be extended through connections with the HPD and other disclosure tools and through curated work areas for Community members.

- ***Build facilitated discussion areas for chemicals:*** Community members will be able to collaboratively develop resources to share and improve the information available in the Commons, both publicly and in some other cases limited to the members of the Community. Some examples of the types of activities that could be facilitated through discussion areas include:
    - Identifying related emerging science
    - Peer reviewing full hazard assessments
    - Contributing options to alternatives assessments
    - Framing policy

- ***Expand the chemical identification crowdsourcing tools in the Identification Registry:*** Security, administration and curation procedures will be established in order to enable full Community collaboration in a curated version of the chemical identification program piloted in the first year.

- ***Develop governance & revenue models:*** Working groups will develop a governance and revenue model to be submitted for ratification by the Community sometime during year three of the project.

Phase 3 / Years 4 through 6

In the fourth year of the project we will become truly a Commons. Members of the Community will participate in an extensive set of curated collaborative work spaces. The Community will work with others in the scientific open publishing community to extend integrated searching to more literature. The Portal will have full API capabilities that support independent application development to extend functions to meet many other user audiences and needs. Some of the key activities in this period may include work to:

- Develop full crowdsourced aggregation of data
- Provide full alternatives assessment support, including functional design alternatives in addition to chemical replacement
- Create integrated searches of science literature across open publishing platforms
- Build data push tools for notifications of new science and assessments
- Map data gaps and testing needs to inform and prioritize research agendas

Data Commons Governance

In order to achieve the vision of the Data Commons described in this white-paper, it will be necessary to establish a governance structure to address such issues as: admission standards and a code of conduct for membership; technical standards for evolving the technology platform; maintaining security and fostering interoperability with other information systems; and, managing collaborative relationships and

potentially revenue streams required to support these efforts.

The Open, User-led Innovation Community

We envision the Data Commons as an open, user-led innovation community in which the free flow of reliable information about chemical hazards and alternatives inspires and supports manufacturer innovation of products and processes that reduce the demand for hazardous substances. "Open innovation" is a paradigm in which manufacturers embrace external ideas and collaboration to advance their product design, rather than rely solely upon in-house resources for their R&D.[38] The related concept of "user-led innovation" emphasizes the role of customers and end users, who play a lead role in product innovation. This concept was originally described by Eric von Hippel of MIT in his seminal book, *The Democratization of Innovation,* who wrote:

> *. . . innovation is being democratized . . . User-centered innovation processes offer great advantages over the manufacturer-centric innovation development systems that have been the mainstay of commerce for hundreds of years. Users that innovate can develop exactly what they want, rather than relying on manufacturers to act as their (often very imperfect) agents. . . . The trend toward democratization of innovation applies to information products such as software and also to physical products.*[39]

The Data Commons will be a convening of lead users of a product – in this case, chemical hazard information. Their aim is to transform the products that they use to better fit their needs – in this case – innovation that drives widespread hazard reduction associated with chemicals use. **That is to say, the mission of the Data Commons is not information for information's sake, but rather to make that information actionable in service to chemical hazard reduction, non-regrettable substitutions of chemicals, and over the long-term, the widespread commercialization of green chemistry.**

Many examples of successful user innovation initiatives in industry have now been documented.[40] The Healthy Building Network successfully adapted this model recently to create a new "user innovation community" operating in the public interest: the Health Product Declaration Collaborative (HPD). The HPD Collaborative was organized to market, manage and evolve the Health Product Declaration Open Standard format. The HPD Collaborative is led and governed by the principal "users" of the HPD to evaluate, select and purchase building products, including architects, designers, building owners, and product evaluators. It also includes well over 100 building product manufacturers who participate in this forum through various means including a "Manufacturers Guild," a "Manufacturers Advisory Committee" and one seat reserved for a manufacturer representative on the 16-member Board of Directors. Both end users and manufacturers support the HPD Collaborative financially through various means, including donations and sponsorships.

The Data Commons will pick up where the HPD Collaborative leaves off, the principal "users" of Data Commons information being those who design chemicals and use them in products; those who assess

---

[38] This paradigm was developed by Professor Henry Chesbrough, founder of the Center for Open Innovation at the University of California, Berkeley. http://openinnovation.berkeley.edu/what_is_oi.html

[39] http://web.mit.edu/evhippel/www/books/DI/Chapter1.pdf

[40] See e.g., von Hippel, Eric A. "The Dominant Role of the User in Semiconductor and Electronic Subassembly Process Innovation." *IEEE Transactions on Engineering Management* EM-24, no. 2 (May 1977): 60–71. [http://evhippel.mit.edu/papers/section-1/]

and compare alternatives; and those who create chemical assessment policies and protocols.

Evolution of the Data Commons Leadership

We recommend integrating the development of governance systems into the 3-Phase evolution of the Data Commons described earlier.

- In year one, HBN will manage the development and deployment of key products and features that will create the initial Data Commons information hubs and collaboration platforms. We will employ the Agile Project Management method that we have successfully used for the development of the Pharos Project, ChemHat, the HPD Builder, and Google Healthy Materials Tool. This control is essential in the short-term to ensure security, consistency and quality control of deliverables. The year one feature set will be made available as a pilot program to organizations currently participating in the Chemical Commons, Data Commons and Community of Practice projects (and perhaps a few other strategically selected organizations.) Leaders from these groups will also be enlisted by HBN to participate in working groups that will convene to address key issues requiring resolution in order to expand the products and feature set in years 2 and 3. These working groups will provide members to a Steering Committee to guide the Data Commons community while it continues to operate as an informal network, gaining experience in working together as a user innovation community.
- In Phase Two (years 2 and 3 of the project) HBN will continue to manage development and deployment of the information technology infrastructure of the Data Commons as it expands its features and capabilities. In year 2, after the group has had some experience working as a user innovation group, we recommend the Steering Committee convene a Governance Working Group to develop and recommend for ratification a specific governance model to be implemented in Year 3.

Governing Principles

In order to achieve the goal of building an information system that makes chemical hazard data widely available, the Data Commons must welcome widespread engagement by diverse stakeholders, while avoiding institutional paralysis, and discourage self-serving behaviors that undermine its purpose. The first step in striking this balance is a set of governing principles to which all participants in the data commons will subscribe.

Many organizations currently involved in the Data Commons project have governing principles that can be incorporated or adapted to the needs of the Data Commons community. We recommend that the Governance Committee convened in year 2 consider these sources, as well as governing principles of user-led innovation projects and open source software projects, when proposing a set of governing principles for the Data Commons.

Data Commons Revenue Streams

The Data Commons is capable of achieving long-term financial stability by evolving diverse revenue streams. New information streams will have business value for which many organizations will pay a reasonable price. The positive brand association with this new transparency initiative will attract offers of corporate sponsorship and advertising. Even the on-line behavior of Data Commons community members will have commercial value.

The advantages and challenges of various revenue options will need to be discussed and evaluated by the Data Commons community in order to strike an appropriate balance between providing as much information as possible to the widest possible audience free of charge, with the need to financially support operating and development expenses over the long term. This paper lays out (7) revenue options for consideration.

1. Donation Model

   In the long run it may be possible to generate financial support for the Data Commons using the donation model. Wikipedia uses this model to support free access to all of the information they have available. Donation models work best when the user community is large, and the average contributions can be small. In the short-term, 3-5 years, this Data Commons community is likely to number in the hundreds, and is unlikely to provide revenues at the scale needed to continue development of the Data Commons.

2. The Marketplace

   The Data Commons could provide access to community members who sell reliable goods and services related to chemical hazard reduction, while remitting a fee to the Data Commons for business that originates there. This approach might contribute to the solution for one current problem that is impeding the uptake of green screen analyses. Because the cost of a GreenScreen assessment is high, companies who commission them are reluctant to share them. With an easy to access marketplace for GreenScreen assessments, assessors will be able to reduce the initial price and resell assessments to other interested clients. This could have the advantage of attracting both assessors and clients to the Data Commons community.

3. Kickstarter

   There have been some objections to the Marketplace approach to revenue based on reservations that it creates a less than totally open and free information environment. An alternative approach has been suggested which would utilize a Kickstarter fundraising drive to increase the number of full hazard assessments available. Under this model, for each unassessed chemical the system would set one threshold for getting it assessed for a license (say $1,000) and another threshold for putting it in the public domain (say $2,000). A manufacturer can order an assessment right away (by putting in at least enough to meet the first threshold goal) or put in a smaller contribution toward an assessment that is not charged until it reaches the first threshold and is assessed. A manufacturer could buy a license to an assessed chemical immediately for a fixed minimum charge. When enough do that to push it over the second threshold goal the assessment becomes public. Any community member could contribute enough to the Kickstarter campaign to fund the assessment or make it public at any time.

4. Sponsorships

   We anticipate a significant number of businesses will both be willing to subscribe to the governing principles of the Data Commons, and provide sponsorship support for the website. There are now several useful models which use conditional sponsorship opportunities to both raise revenue and to drive and highlight positive corporate behavior. These include the Health Product Declaration Collaborative, and the International Living Future Institute.

5. Membership or Subscription Fees

   The Data Commons could charge a membership fee to offset operating costs. This model might also discriminate among users based upon their volume of usage or email address, providing discounts for .gov, .edu, and .org users while charging market rates for access by .com users. The challenge of this model is that it creates a barrier to widespread participation. Most information-sharing communities do not charge membership fees.

6. Feature Levels

   Many information technology products offer limited free access, while charging premiums for greater usage or more valuable information. Some examples of this model include Adobe, which provides a free "reader" but which charges for its many other document services, or on-line news services such as the New York Times, which offer a limited number of page views for free, and charge thereafter. This model might also discriminate among users based upon their email address, providing discounts for .gov, .edu, and .org users while charging market rates for access by .com users.

7. Data Mining

   Although data mining of online users often has a negative connotation, transparent, "opt-in" participation in a program that tracks user activity within the Data Commons might provide data that could be marshaled into valuable reports that promote the mission of the Data Commons.

Conclusion: Taking the Next Step

A revolution is underway. Awareness of the connection between our material environment and health is growing, as is awareness of the need for transparency and assessment. A concurrent revolution is reshaping the way we access and use information, creating a new "sharing economy" that is accelerating knowledge transfer and catalyzing new collaborative approaches to innovation. The time is right to tackle the challenge of organizing the world's chemical hazard information in service towards the manufacture of products with inherently less hazardous ingredients. This paper proposes a practical approach to this challenge.

We recommend as a next step towards creating a self-governing Data Commons community, that the Forsythia Fund convene a follow-up meeting of groups that are currently participating in the Chemical Commons, Data Commons and Community of Practice. At that meeting, HBN will facilitate discussion and amendments of this White Paper, with the goal of ratification by the end of the meeting. Once ratified by the current "founding Data Commons community" development of the Data Commons according to the plan agreed to in this white paper, as amended, can begin.

Glossary

API (Application Programming Interface) – a set of protocols that specify how two programs or computers can interact with each other. Within the Data Commons, APIs would allow computers to collect and process hazard information from multiple web resources at the same time.

CASRN (Chemical Abstract Service Registration Numbers) – the most widely used identifier for chemicals in industry.

CBI – Confidential business information

ERP – Enterprise Resource Planning – integration of multiple database applications to track and manage all stages of a business operation from product plan, through manufacture, marketing, sales, inventory, shipping, billing and payment.

Fenceline communities – neighborhoods that adjoin industrial facilities (are at the "fenceline") and hence are likely to be the most immediately affected by to environmental releases to air or water from those facilities

GitHub – a web-based hosting service for software development widely used to manage open source software development and other collaborative projects

Green chemistry – the design of chemical products and processes that reduce or eliminate the generation of hazardous substances

GreenScreen -- a protocol for assessing the scientific literature about a chemical on 18 human health and environmental health endpoints against set criteria and determining a hazard benchmark on a 4 level scale.

IP – Intellectual Property

Paywall – a system that prevents Internet users from accessing web content without a paid subscription

PLM – Product Lifecycle Management – tracking g and management of a manufactured good through its lifecycle from design through manufacture, use and disposal.

Siloed – Information is considered "siloed" when it's grouped into several separate, insular systems incapable of easy inter-communication. The Data Commons seeks to bring different siloed chemical hazard resources into an online space where they can be easily accessed together.

URI (Uniform Resource Identifier) – within the Data Commons, it means that a web resource is organized so that users can quickly and easily be directed to the information they are looking for.

User Innovation – User innovation is a process by which customers and end users play a lead role in product innovation.