Chemical Hazard Data Commons White Paper Reviewer Commentary

Content comments submitted in review of the first draft of “Toward Safer Products: Accelerating Change with a Chemical Hazard Data Commons”

Feb 28, 2014

We circulated both an initial draft of the white paper and a survey to 76 people. We received 56 responses, nearly a 75% response rate. We received 16 responses through the survey instrument, 35 sent email commentaries and 23 submitted Word documents with track changes (The total of the commentaries adds up to more than 56 because some respondents submitted through two or all three methods). This document is a compilation of all of the content oriented commentary that we received. Names and organizations are withheld as we did not request permission to publicly circulate names and company affiliations in affiliation with this report. Some of these comments are reflected in the current version of the White Paper and the accompanying Working Papers while others remain to be addressed in future working group activities. Results from the Survey are compiled in a separate document

The White Paper, the Working Documents, the Survey results and more information about the Project are available along with this document in the Chemical Hazard Data Commons github repository at: **https://github.com/healthybuildingnetwork/chemical-data-commons**

Contents:

** These comments have already been pulled into the Chemical ID paper

*** These comments have already been pulled into the Functional Use paper

Overall comments (the fun stuff)

- The idea is extremely ambitious but (a) very much needed and (b) has great power for usefulness among many user groups. (NGO)
- This is an AMAZING document. (Industry)
- Overall, I think this is a really great proposal! I am very excited about the opportunities you present, and think that the underlying challenges and current state of technology is well laid out to present a compelling case for the Data Commons. (NGO)
- I got a chance to look at the doc and think it's fantastic...am just thrilled that it's happening. (NGO)
- Really great effort here... Looks comprehensive and is well written. No concerns or show stoppers (Govt.)
- It's a very interesting, worthwhile project! I especially liked the emphasis on developing both a tech infrastructure and a community around it (Academic)
- Overall, the document is very well written, particularly the Executive Summary. It's a quality piece of work, and it clearly reflects a tremendous amount of careful thought and deliberation. Editorially, it's a little uneven. The lifecycle sections are relatively weak. The unique identifier section is excellent and could practically stand alone — it nails the key issues. **T**he Data Commons concept is critically important and, from my perspective, central to the issues we're facing as a community. (NGO)
- This paper is quite impressive and I am excited about the picture that it paints! It will be an incredibly useful resource and forum for learning and sharing. (NGO)
- Kudos. This is an important paper. I love the overall direction. (Govt.)
- What an extraordinary amount of thought has gone into developing this! This is such an important next step and I'm really pleased to see this taking shape. (NGO)
- I've made it through the whole document and I felt like it read well overall, and also like it is a realistic and worthwhile project. (Academic)
- This is an amazing piece of work-- pulls together a very large and complex landscape of ideas into a well-organized and well articulated paper. I think it's effective. (Academic)
- You and others have put a ton of work into this document and it shows. I think that the ideas are sound, and I can't wait to see how the final product will evolve. I do worry that there is a little bit of promising everything to everyone, but as with any project, I imagine the final product will be influenced by those who actually show up to do the work. (Academic)
- This paper is a nice summary of the chemical data that is available publicly, as well as how it could be organized. The amount of data accessible to the public has greatly proliferated with the growth of the internet, and new organizing methods are always useful to discuss. (Govt.)
- I want to commend you on completing the first phase of an incredibly important piece of work. This was a monumental task and the paper does an excellent job of synthesizing the issues around data needs for decision making toward safer chemicals. You have done a terrific job in framing up the data commons. I hope that the funding community will understand how important it is to move this work forward.(Academic)
- I think this white paper lays out very nicely the intention, the approach, the partners and users, the pitfalls, and many exciting uses of such a commons... Aspects that really grabbed my attention: having a way to publicly share Greenscreens is really exciting; having an integrated database to identify alternatives by searching function is fantastic; recognizing that pitfalls will include the form of the material (you have the example of silica dioxide) is important; assessing mixtures – how exciting!. And, the ability to integrate different types of data including from models and non-animal studies is important. How wonderful that you are tackling this overwhelmingly big but extremely important task! As usual, HBN is not only ahead of the curve, but is helping to define the curve. (NGO)

- Overall, this is very exciting! And absolutely what is needed to drive transformational change in chemicals management. (Industry)
- Wow! I can tell a LOT of work went into researching and developing this proposal. It is incredibly comprehensive and well considered. Great work! Thanks for this wonderful contribution to the field! (Consultant)

Framing, visualizing & defining

- The document emphasizes the notion of the Data Commons "**platform**". There are lots of conceptual and operational definitions of "platform" as an issue in information technology and as a business model. I think it would strengthen the paper to explicitly **define "platform" and add a graphic** illustrating the core and peripheral components of the Data Commons platform. Similarly, the document frequently uses the term "**tool**" in a number of circumstances. I think these need to get defined and fleshed out with specific examples. (NGO)
- Page 7, 1st paragraph notes that the Data Commons project will "make these tools available". Many of the tools are available already - could be stronger to instead reinforce that the Commons offers comprehensive information (or context) by aggregating data from these tools. (NGO)
- Pages 8-10 "Potential Participants in the Data Commons" - It is very useful have these characteristics of uses/needs, but the lists limit the ability to understand the crossover/uniqueness of the supply/demand. Could be more powerful to track these commonalities and singular needs in a visual way. Suggest a matrix with the group name on one axis and characteristic on the other. Then use 2 different visual markers to document where the groups create this data vs has a need for it. (NGO)
- Would it be possible to include a few graphics showing the overall vision for the Chemical Commons? (Govt.)
- There should be a relatively comprehensive review of what's available today and why these solutions fail to meet the objectives. Certain sections do this analysis, but there is no equivalent at the top level to make the case that this highly detailed proposal is needed in the first place. It was also unclear who's speaking. Who's going to do all of this work? Who's responsible? Is this a business plan? Right now, it sounds like, "Hey, wouldn't it be cool if…" which is not actually that useful or meaningful. I wouldn't know what to do with this white paper if it were forwarded from a colleague who wasn't an author. Finally, this seems very ambitious, which can be okay as long as the level of detail is commensurate with the likelihood of adoption. In other words, you may not need 10 pages on proposing a particular schemas if that's something to be worked out in implementation. Maybe the schema could go into an appendix as an example of a preferred embodiment or as a starting point for discussion. (Industry)
- What is the basic scope of what you propose to cover? Would it mainly be building materials, or would it include other classes such as drugs and pesticides? How large would the database be? This is a very key question for resource allocation. (Govt.)
- You describe the need for/wanting to develop so many things related to chemical data, assessments, etc. As you read the paper it's quite overwhelming. I realize that this is a concept paper and not a proposal. But, is there some way to now step back and describe how you propose to build this thing? What first, what next....How the parts interrelate? A diagram would be really helpful to better understand how the modules connect and relate to eachother. Maybe that's phase 2 but it sure would be helpful now, as someone who has not been intimately involved, to get the big picture of what you are trying to create. I would imagine that would be helpful for getting the development funded. (Consultant)
- I have two key questions after reading this: 1) Who's going to build these tools? Just laying down the foundation will be an enormous amount of work. 2) What are the most obvious first steps to get

this started? (Consultant)

- I suggest that the executive summary be strengthened and tightened to clarify the next steps for building the data commons. The current draft of the document is lengthy and somewhat repetitive. I am concerned that potential funders will not read the whole document. A succinct and clear discussion of the short-term work needed to lay the groundwork for the commons is needed. (Academic)
- I'm wary of promising too much or making the work sound easier or cheaper than it is. The use of words like "rapidly," "easily" etc make me nervous, especially if this is something HBN is going to be shepherding. I think the proposal is strong enough without these promises, and I'd be more comfortable if we are straightforward about the work being proposed without speculating on how easy or fast it is. Just making this convergence and community etc happen would be a huge accomplishment – no matter what the cost or time frame. Related to this – I'm wary of making it sound like HBN currently offers more than it does, and I have a couple comments in there about the description of our Pharos API. (NGO)
- The impact the Data Commons can have on CHA and AA is clearer to me than product inventory activities. Product inventory is a critical aspect of making the Data Commons work. Might be good for HBN to visually identify groups and/or service providers that fit within each of these categories so their roles and adjacencies are better defined/understood. (NGO)

Assessment frameworks:

- I think GS is a GREAT tool, however, it is not a FW that works well with **all** chemicals. For example, it doesn't assess polymers or colorants as effectively as I would like to see. I would want to make sure that other FW's are considered as well. I think you have that captured but just want to make sure. I am a big fan of the DfE FW. It builds in an ability to assess polymers, colorants and even enzymes. I am trying to build a FW that works for LS&Co that builds on GS, DfE, GRAS (Industry)
- Will the project highlight alternative forms of risk/hazard assessment, or Health Impact Assessments (emphasizing "best practices"/most progressive), across different policy arenas? And, will it consider the relationships among "best practices" across different policy arenas (including their potential conflicts and tensions, in both practice and framing)? How will the project address alternative forms of risk/hazard assessment? (From my perspective, there's definitely a "politics of evidence" in the fields of environmental and public health; so basically, I'm wondering how the Data Commons project may address those issues?) (Academic)
- **Developing a shared vocabulary**...I thought this was being done by the GreenScreen folks (CPA)? I would suggest partnering/harmonizing with the GreenScreen on this point. This is why I felt, when the HPD was being developed, that there should only be one tool (GreenScreen) that should address the hazards and risks associated with chemicals, mixtures, polymers and products/articles. (Industry)
- Flagging information quality and applicability. All information is not created equal. Not all information about a chemical—and particularly not all existing evaluations—will be appropriate to use in all other evaluations. There needs to be some way to clarify the gaps or problems with base data and with existing AAs or hazard assessments for a chemical. This is a difficult thing to do, as it will attract chemical defenders, but is essential because best practice for AAs is still being defined. For example, one of the current best practices for hazard assessment (Green Screen) includes inherent weighting systems that might work well for certain classes of products, but that might not be the most desirable for all use patterns (e.g., products used outdoors and products discharged to the sewer system). Without some mechanism to flag potential misuse or

misinterpretation of existing documents, we might see Data Commons users inadvertently implement regrettable substitutions. Based on a current Green Screen assessment, a data commons user for a different type of product might elect to replace chemicals that pose human health hazards with chemicals that pollute water or harm wildlife. This is exactly what is happening with pesticides, through a similar though less explicit set of weightings in product regulatory decisions. A human-prioritized evaluation system makes sense at some levels, but its overuse on pesticides in particular has had dire consequences for the non-human environment, particularly for aquatic ecosystems. (Consultant)

- We will need to define what is meant by "full hazard assessment." (Academic)
- Also for GS, it is important to speak to the fact that “reasonable toxicologists can disagree” when it comes to hazard assessment. The verification program can help to resolve disputes – at least for a period (because assessments are only good for 3 years max) – data will continue to emerge and new science means new understanding. You might speak to how the Commons will deal with constant change and different interpretations. Maybe just letting each system represented in the Platform have its own conflict resolution process is a way to go. But with transparency comes either consensus or disagreement! But since the transparency is key – and a key part of GS -- it might be good to address how you/The Community will deal with different interpretations. (NGO)
- How do you make decisions about what parts of lifecycle are addressed in an alternatives assessment? Is this focused on toxicity of product endpoints? Does process come into it? Related: (cradle to cradle may already have this tool) many chemicals (esp from petrochemical feedstocks) probably have similar cradle-to-chemical plant; can this information be fed into a whole bunch of different products to facilitate lightening the LCA burden? This could be especially important as we move into bio-based waste as a major chemical feedstock (Academic)

Data origins & data quality:

- I think that the document would benefit from explicit treatment of concept of “**metadata**” throughout. Many of the concepts in the document relate to “data-about-data” (provenance, quality, definitions, etc.). In other disciplines, these issues are addressed through **metadata standards**, and I think this term and idea are very relevant here. (NGO)
- Q**uality control of data**. You do not address that in any part of the paper. How would that be addressed through the data commons (Academic)
- Maintaining information quality is a fundamental challenge for the data commons. For example, funding source is known to affect outcome of scientific studies (that is why high-quality journals require authors to disclose their funding sources if that source has an economic interest in the outcome of the published research). In this case, the funders of the individuals participating in the commons could affect their contributions. Similarly, the sources of base data and/or resource reports (hazard assessments, alternatives assessments) may not be unbiased. (Consultant)
- For an example of how the federal government has gotten twisted in knots dealing with the data quality problems of existing data, see http://tinyurl.com/nxtl5jb (Consultant)
- one thing that struck me several times as I was reading about all of the ambitions and functionalities forecasted for the DC, is how important and difficult data quality and assurance will be for maintaining the integrity of the system. I know that this is a detail beyond the scope of this paper to introduce the DC, but it needs to be discussed soon in order to build confidence and trust that the system will provide scientifically accurate information. It also seems like it is going to take an enormous sum of money to build and maintain the system. (NGO)
- How does industry influence play into this? When industry helps get something banned they usually have motives (like a market edge on a competing product). How do we keep industry working in the best interests of the world supplying tox data while also in their own best interests?

(Academic)

Crowdsourcing & Curation & Governance & Community:

- just remain cognizant of the interest chemical manufacturers will have in influencing the assessments of their products. This raises questions about how to conduct the crowd-sourced elements of chemical information. (Academic)
- I wonder how the project might also address digital divides that affect who gets involved in internet-mediated projects in the first place, as well as participants' levels/types of engagement? (to me it's important to emphasize that crowdsourcing isn't synonymous with democratic participation…) (Academic)
- **I love the Github concept.** Larry had mentioned that a while ago, and I think it's brilliant, potentially game-changing. I think the document could elaborate on it. Why does Github work? Why is it likely to work for this community? Can the Github infrastructure we directly applied? Why not? I agree with the reference that suggests that need to start with our version of TCP/IP (I.e., a robust, scalable identification and communications protocol). (NGO)
- Page 24, "Discussing and Digesting…" It would be useful to provide details about how the Data Commons would become a trusted voice, particularly with so many contributors. Who decides what information is useful - how is the Commons curated? (Unnamed)
- Who would manage this – perhaps not necessary at this point, but I think it would be useful to think through. What you have proposed is a multi-person, multi-million dollar a year initiative. Have you thought through the logistics of that. I love big thinking and you bring in examples of where you have started this for example through Pharos, but it might be useful to think through how this would be done (Academic)
- By the close of the paper, I was impressed with what I had read, but aware of what I wished had been there. In particular, I missed some discussion on institutional structure, governance and resources. How would this Data Commons be organized? Is this a government function (I hope not)? Is there to be a new non-profit service organization? How would this be administered? Who "owns" it determines how much we trust it? Who funds it? (Academic)
- Who would be allowed to "crowdsource". There is a variety of skills out there, as we have seen in strange Wikipedia edits. Would expert curation be enough to assure quality? Would individuals in the community have the rights to change all data, including company supplied data? (Govt.)
- Address the organizational structure that will have the best chance of long term success. Can you point to similar endeavors that have worked or not worked and pull learning from that. Should this be for-profit or non-profit - why? Should this be academic or NGO? Is it critical to launch this with an endowment so the that the data is protected in perpetuity? Is it governed by a board or how to ensure quality and integrity? The plan you have written is rather compelling so my first thought is, for-profit.. I know there are reasons why this should be a non-profit so consider addressing it early. (NGO)
- The document clearly articulates the idea of an evolving professional community surrounding the Data Commons. I think this concept is important, and I think this would be stronger with some attention to the fundamental mechanisms for defining this community, notably **elements of education, technical training, and professional recognition** (e.g., the equivalent of a LEED-AP or –GA). I think this is both a business opportunity and an essential mission element, as members of the community need to share a common core of facts, terminology, and understandings. (NGO)
- I think you could emphasize the Community as a source for continual improvement of the methods and tools also. For GS, the "community" is key for continual improvement. For example, we are currently working on method improvements for polymers. Do you see any of these tools as open source? The community will also be key to learning how to integrate a broader set of predictive

toxicology tools and high throughput screening results for which the significance is not yet established. One of the keys to getting good hazard information is moving beyond the mean, slow and expensive animal testing to much faster but RELIABLE models and other predictive tools and assays. There is much work to be done here yet. (NGO)

- What about the chemical company -funded scientist looking to shape the crowdsourced assessment of a commercially valuable chemical or class of chemicals. (Academic)
- There is a lot of additional data needed to make actionable interpretation of this sort of data. It will be important to considering how much support, education, and filtering can be accomplished through online tools. I love the HSDB, but using and interpreting (even in a relative way, not an absolute greenscreen fashion) involves a lot of assumptions, data interpretation, and often subjective evidence weighing, and that is using an aggregated and curated source of data. I am comfortable getting an impression from this sort of ToxNet type data, but I don't really feel comfortable or productive going right to PubMed, which is what Meg will do, but that takes a whole lot more training. This issue of Toxicity data also brings up the issue of how to use the Chemical Commons as a tool to drive the production of actionable data within the commons for chemicals that lack data. (Academic)
- How do we address honesty and integrity of assessments done by non-independent bodies – i.e. Tyrone Hayes vs Syngenta on atrazine? This is a big challenge that I see because this document and any approaches to industry have to be made in the spirit of good science, trust, and transparency, but history seems to suggest that skepticism of data from vested parties is necessary. (Academic)
- Is chemical knowledge as broadly held and as easily verifiable as functional coding skills? This question isn't meant to sound elitist, but I spend most of my (not-at-work) life in a world where everyone and their dog can write functional code and teenagers do it as a hobby, but almost none of those people (including at work) could even tell you why acetone ought to be more persistent than ethyl acetate in a biological environment. (Academic)

Business models/funding

- I appreciate the intent of the Exchange and a Kickstarter-style business model. I agree that these considerations are essential and that the Data Common needs to research this issue. Personally, I think its worth giving some really critical thought to the distinctions between a Kickstarter model and, alternatively, something closer to an eBay-style marketplace. In my experience, Kickstarter is just what it's name implies: a kick-start to a self-sustaining business. I think **eBay-type markets** have proven more effective at scaling cottage industries — just what we need. Regardless, the document is certainly asking the right questions. (NGO)
- Funding: What is your source of funding for this? Seems like a big, costly project. (Industry)
- Page 41, 1st paragraph: there is mention of "system access fees". It could be useful to touch on this topic further, since the cost of access to other programs' data is mentioned as a hindrance. (NGO)
- I am a little unclear on the business model underlying this (e.g. how exactly it would incentivize data sharing/hazard assessment) but I love the direction it's headed. (Govt.)
- We think the concept of a data commons is good, but we still have numerous questions about the economic model that will support and grow the concept of shared assessments that are highly credible and based on up to date scientific data and assessment methods. That being said, I'm always open to a discussion of this issue to figure out how to increase access to information and assessments. (Assessor)
- How would funding work? If the scope was sufficiently large this would be a problem. The Government is low on new funding right now. Are you proposing "crowdfunding" (Govt.)

- The paper is ambiguous in how it deals with the big question of money. On the one hand there is lots of talk of crowdsourcing and things being freely available or affordable, but with not much detail on how the tremendous amount of work involved in building the system and convening/supporting the community would be funded (or who would be doing it, for that matter). On the other hand it gets quite specific in places such as the description of the kickstarter model (which I have some questions about) and the suggestion of incentives based on reduced system access fees. I think these specific and speculative comments are a bit distracting from the content. If the business model to support the DC still needs to be developed, or is in development, I think it's okay to say so, but I wouldn't provide potential specifics on certain pieces if the whole is not in place. (NGO)
- In general the idea of the data commons is a great one. In theory something like this is badly needed and could serve the public well. What you lay out regarding structure is clear and thorough. The design is very familiar to us here at SciVera mainly because it's very similar to what we have implemented with SciVera's Chemical Data Repository ("CDR") of hazard assessments. The CDR is the underlying data that supports our subscribers bill of materials and bill of substance data flowing in---generating material and product assessment results. SciVera's CDR is cloud-based, which offers important scalability, security, and reliability features. - Our questions on the white paper and the data commons concept focus mainly on the feasibility of implementing the labor- and expertise-intensive work of data collection and hazard assessment. As you know, it takes a great deal of effort and expertise to gather available hazard data for a chemical from a disperse set of sources, organize it, and then evaluate it to arrive at a hazard assessment by endpoint, then overall for each chemical. Throw in QSAR steps, modeling, weight of evidence, etc to close the myriad data gaps and you have lots of heavy lifting needed for the task of assessing a single CASRN. This is work we have already done and maintain on over 30,000 CASRNs in the SciVera CDR. - Your proposed approach of wiki-like effort sharing is interesting, and knowing what we do about the hazard assessment process, we are not clear on how it will work in the context of a sustainable financial model that will compensate those contributing content while maintaining data and assessment quality. You may have that figured out, it's just not presented in the white paper. As a critical element to the feasibility of a data commons effort. - As we discussed in Boston, in addition to our work directly with manufacturers and their suppliers, we have established a goal at SciVera of finding mutually beneficial ways of supporting chemical assessment more efficiently in other organizations by leveraging the investment we have made in the SciVera Lens technology and content. If we can develop a way to contribute SciVera's technology and/or content to the plans you have made for the data commons, we are interested in a collaboration. It also seems there is a huge opportunity to build with the momentum of other complimentary efforts like GreenBlue's MiQ program to help things move faster and more broadly. (Assessor)
- this is going to be the most difficult aspect of building the DC. This is going to take a lot of brainstorming with assessors at the table to figure out a financial model that incentivizes assessors to participate. At this juncture in our efforts to build a more transparent business culture, we are transitioning from pockets of knowledge that are privately held to emerging models that encourage knowledge to be shared more broadly (e.g., Pharos, HPD, MiQ, IC2). Assessors are critical partners in our efforts to build our sharing platforms. We need to fully leverage market-oriented approaches while we figure out how to build truly open systems like what is envisioned for the Data Commons. I will create that slide I was talking about that attempts to diagram the flow of data and $ that we can use to zero in on the econvoernanomic model of the Exchange. If you have already done a lot of deep thought on this and have a model to present, I would include it because the Exchange is critical to the success of building the DC, and discussion about how to actually structure the Exchange is absent, leaving big unanswered questions about how realistic the creation of the DC is. (NGO)

- Who is responsible for creating and maintaining this platform. How resilient will it be to organizational failure? (Consultant)
- I think this idea is very interesting and has a great deal of promise. I wonder if thought has been put to what the model might be if the party interested in initiating an assessment were a software provider who offers ingredient cataloguing to many companies and thus has data on the prevalence of use of many types of chemicals? That could be one way to prioritize chemicals needing assessments. Also, the scenario may require a different way of thinking through the pricing model if this one hypothetical company pays for an assessment to be done but intends to make the assessment results available to many manufacturers (probably for a price). (NGO)
- For those interested in specific chemicals, this (KickStarter Exchange) proposal makes sense. Will it be as attractive to a company that wants to see lots of chemical assessments? I'm also thinking that it is easier for us to make the case internally to support numerous chemical assessments that are relevant to our business, than to buy-in into one particular assessment. I could imagine saying to management here, look, if we pay $xx, we get access to assessments that represent xx% of the 4400 textile and footwear chemicals we might use in our products. (Industry)
- Does the Commons take a business share in product developments? Royalties? How do vested interests affect impartiality and focus of scope of database? (Academic)
- Differentiate data producer vs consumer use cases (including their motivations) for each participant. This essentially is the marketplace...so how do we balance supply with demand...where are the gaps? (Consultant)
- Suggest to conduct a deeper market study of what each participants needs are, and how a Commons could be best aligned with those incentives to ensure success. (Consultant)

Intellectual property

- How does the commons address IP sharing? Does it ask users to step outside of the commons to discuss proprietary formulations? (I know there is a vision of transparency that is somewhat orthogonal to this, but if there is no space for protected IP then will industry come to the table?) How do we make this global without our best work being scooped by China and putting the "good guys" who were willing to disclose out of business. (Michael semi-regularly gets email about his electronics hardware asking him to "sent me your codes") (Academic)

Potential participants

- I want to know more about the potential i**ntersections of the tech infrastructure and stakeholder community/ies**. - How might this resource be leveraged by the full range of **groups addressing the public health** dimensions of the built environment and city/regional planning? E.g. the white paper brought to mind this short piece by Howard Frumkin, "Health, Equity & the Built Environment": http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1257564/pdf/ehp0113-a00290.pdf ...because he writes about enviro health groups addressing the chemical environment and the built environment separately. It seems like your project and the Healthy Building Network could be uniquely poised to bridge these efforts (and see in data elements for EJ groups) (Academic)
- Because environmental hazards will be included in the Data Commons, it could be useful to include reference to the benefit for Environmental Product Declarations together with the other named programs. This could bring in additional support and relevance for some of the data from others in the materials ecosystem. (NGO)
- Suppliers should be included in the list of Potential Participants (page 9). This critical group and their contributions seems to be absent from the paper. For example, their engagement could also be a bulletpoint as a goal for the Data Commons project on page 11. Also, this could be an added example of the "Data Commons at Work" (p12) Can the goals for Data Commons include benefits

to manufactures and suppliers through Alternative Assessments or preferred alternatives? (p11) (NGO)

- Page 20 (top) - different stages of input are noted; would it be useful to note EPDs and HPDs as possible outputs for this “common language of summary level chemical hazard information”? (NGO)
- Participation in Data Commons Community likely to involve few base data generators, which may limit or preclude achievement of the vision. Data users (product designers & alternatives assessors) are for the most part not going to be the same people that generate the base data (research scientists). By “base data”, I mean human & environmental toxicity data, chemical characterization and environmental fate data, human & environmental monitoring data. The Data Commons Community is likely to be dominated by data users. Most base data are generated by either government & university research labs for primary purposes other than AAs, or by the manufacturers of the chemicals themselves. Only the latter group (manufacturers) is likely to have motivation and resources to participate in the Data Commons community. That tilt might not be helpful. I didn't see chemical manufacturers on your likely list of project participants. Unless there is some plan to exclude them, I would expect robust engagement from this sector. Most data centers are government based or designed to include only a subset of the groups listed in this document so as to maintain control over participation toward data quality protection and to precludes domination by one stakeholder/sector. (Consultant)
- I would like to press is compatibility and a push for the commons to facilitate rigorous chemical hazard information interfacing with other parts of the sustainability evaluation world. I would like to see the work of the commons make it easier for the people already doing sustainability assessments to incorporate hazard into their evaluations. This will help "mainstream" the collection and evaluation of this data. (Academic)
- Potential participants: Consider addressing contributors and users separately (although there is some overlap). It seems notable that there are relatively few contributors and that they are key to the success of the project (an opportunity and a threat). The number of potential users is impressive and you do a good job demonstrating the demand and evenue potential. (NGO)
- Unclear how “Designers” or “Retailers” are target audiences of the information contained in the Data Commons as described. The DC seems geared towards government, academia, NGOs, service providers, manufacturers (who have chemical level info) versus these other audiences. MiQ is designed to help Designers to select better options but via already assessed materials with results translated. Designers and Retailers don’t have the time or skills necessary to take hazard information about chemicals (which they probably do not have the identity of the chemicals anyway) and make judgments about the products they specify. (NGO)
- Process chemistry info will always be very hard to obtain and will have a high degree of uncertainty, unless there are concrete incentives or regulations driving this sort of disclosure. Another way to help identify some of these residuals would be to partner with well-established product testing laboratories to determine realistic diagnostic tests as well as common residuals by product category. While direct disclosure of process chemistry is preferable, I think it will be very important to think about other ways to drive this information into the data commons that will not rely on manufacture process disclosure. (Academic)

Policy

- it seems like direct hazard regulatory policy is more straightforward (in some ways) than e.g. chemical lifecycle policy, or a policy orientation that emphasizes "Health in All Policies," including built environments/city planning (http://www.phi.org/resources/?resource=hiapguide). Would it be possible to highlight these different policy arenas, orientations, and connections among them in the Data Commons project design? (This could include highlighting the policy arenas that the Data

Commons is most focused on or designed for addressing, versus those it's less focused on/designed for.) (Academic)

Access to science literature

- The document makes very important points about access to **scientific literature** — the raw material for assessments. I think it can make this point more strongly and with a little more depth. **Access to published scientific literature is a fundamental barrier for non-academic institutions**. It's too expensive, too fragmented, and too difficult to systematically find relevant information. Each of these can be tackled with specific strategies, including **Open Access publishing** (such as PLoS as is mentioned), **better indexing** (in coordination with groups like ISI), and **hazard-specific metadata standards** for relevant publications (coupled to metadata collection tools). (Unnamed)
- We need a repository for "all" chemical/material toxicology dossiers and assessments. Need to know much quicker where there are data gaps and if data is too old to be relevant. (Industry)
- Page 14, "Key Underlying Data, Scientific Studies" With the mention of journals and research, is there a value in bringing in papers from EHP? How would these be translated? Noticed a reference to templates. May be useful to connect this with a burgeoning effort under the NIH HiBR (Health in Buildings Roundtable) - if this could be relevant, let's talk further… (NGO)
- Once the commons is off the ground is there a good way to get a session at Tox (or other) conferences that take full papers, to facilitate contributors getting their data into the public sphere while still living up to the "publish or perish" of academia? How does this vary for fields like chemistry (where a conference is not a publication) vs CS (where a conference is the only publication)? (Academic)

APIs, Open data & data exchange

- I think there are also interesting emerging models for offering APIs as a service as part of the **Open Data movement**. Specifically, I'd recommend considering the implications of **Socrata**, one of the companies behind Data.gov and a leader in providing Open Data solutions to municipal governments. T**hey're helping turn fragmented spreadsheets into API-accessible information.** It's a useful model. (Unnamed)

Data elements:

- Will there be **geocoding** of data or a GIS mapping component included in the Data Commons (e.g. of chemical lifecycles, fenceline communities, etc.)? This would be a great resource for analyzing the **environmental justice** dimensions of these chemicals and material economies. It would enable collaboration with the many place-based strategies and initiatives on behalf of public health. (Academic)
- Life Cycle Hazard Information: This one should also partner/harmonize with the GreenScreen as they have thought about this and put together some ideas and policies around the idea of life cycle impacts. (Industry)
- **Environmental fate data** (hydrolysis, photolysis, sediment half-life, etc.). For example, if a chemical decomposes quickly when it is released during use, the hazard assessment should look at the chemical(s) it decomposes into. There is a major thrust in certain types of chemical design that involves seeking chemicals that are not persistent (this is responding to regulatory focus on persistence). Consequently, these new chemicals decompose quickly—sometimes into chemicals that are more hazardous than the parent compound. Without fate data, its not possible to determine the proper relevant factors and scope of AA and LCA work. (Consultant)

- **Environmental monitoring data**. Examples of environmental monitoring data for surface water are: CEDEN www.ceden.org , USGS http://cida.usgs.gov/nawqa_public/apex/f?p=136:1:0 , STORET http://www.epa.gov/storet/ , EPA Clean Water Act 303(d) list of impaired waters and the pollutants causing the impairments http://iaspub.epa.gov/waters10/attains_nation_cy.control?p_report_type=T (Consultant)
- **Exposure**: I agree with the three step process you describe but of course exposure potential becomes relevant, especially if there is not a drop in substitute. I think the section on functional use helps to address this issue. Although you are creating a chemical **hazard** data commons, data relevant to exposure will often be needed in decision making. Will the chemical hazard data commons include data on exposure or will users need to go elsewhere for these data? (Unnamed)
- Where will you find Life Cycle Information data at the chemical level to allow DC to include this attribute in its analyses? and what is the scope of the Data Commons? It seems that it is to provide robust information about the hazard characteristics of chemicals. But when you start talking about residuals in the "life cycle information" section, it gets confusing. Do you mean contaminants found in substances? (NGO)
- **Toxicology studies** which are often the early indicators of potential harm. These are harder to search and exist at a level of detail below the lists that GreenScreen is built upon. This is the type of information that is most helpful to me as a chemist and it is very far away from the information that is currently compiled. The only reason that this should be considered is that, ideal something like the "data commons" would drive more research in the area of toxicity testing and interpretation, but right now it is not clear that this is even part of the scope. (Academic)
- Perhaps there's a role the Commons can play in funding some of this research? (Consultant)
- IHow about knowing hazard "score" at different points in the supply chain or in the production process (e.g. chemical to material to component to subsystem)? (NGO)

Chemical ID:

- Does not appear clear as to how one would index/access a specific chemical or chemical group. Is there to be a new numbering system that replaces CAS/EC-type numbering? This should be made clear…(Industry) *(A: a new numbering system is being proposed. You will continue to be able search and link to other data systems by CAS/EC and various other identifiers, but the new system will support identifying levels of grouping above and form type detail below that CAS/EC do not support.*
- The creation of a new identification number is an interesting concept and one worth pursing. I have concerns about incorporating the word "Hazard" into its registry name because it seems to assume a negative impact rather than a neutral naming convention. Imagine that preferred chemicals would also be assigned an ID number, too, as well as alternative assessments and green chemistry. (NGO)
- I'm having trouble with the idea of creating an entirely new tracking ID for every chemical. I think that a new system is unlikely to be widely adopted. It seems better to create new tracking numbers only when needed to fill gaps in existing systems (e.g., Chemical Abstracts Services ID Numbers). I understand the challenges (particularly the proprietary nature) of the CAS system, but on a practical basis, I don't think that the data commons project can supplant it. I have witnessed many examples of wonderful new standard systems not being adopted because they are just too inconvenient for the majority of potential users.
  (On the other hand?) I like the idea of looking at chemical families. This could be strengthened to support "read-across" first approximations for data for closely related structures and structure based approximations (e.g., SARs) and other computational methods as means to obtain a first estimate of chemical properties, environmental fate, and/or human and eco toxicity. (Consultant)

- How would assigning your Substance Hazard Identification Number work? Would it depend on voluntary submissions by various organizations to a central source of numbers? I am glad you mention keeping track of chemicals without a structure, since many people are structure centric. (Govt.)
- I think it is a good idea to try to cross reference existing resources. We have traded identifiers and links with several organizations, including EPA SRS, and the FDA SRS, and this is relatively straight forward. If you set up a small prototype of your "Substance Hazard Identification Number (SHIN) Registry", this would serve as a test bed and proof of concept for some of the ambitious ideas that you have delineated in your paper. (Govt.)
- As far as I have been pushing for better & open IDs, I think we should be clear about *why* in the paper. I thought we were, but it needs more clarification on the following point: **We don't need to "convert" everyone who uses CASRN** to our new/weird/other system. A purpose-built open chemical ID system will simply be *necessary and helpful* for us to do what we are describing. Open ID will be an important element for the purposes of internal organization & data alignment, transparency, and effective sharing. In my edits of the white paper, I removed several paragraphs which I'd originally written with the intent of convincing people that we need an alternative to CASRN, specifically. I think that original writing helped create the impression that we are trying to take CAS down and universally replace it, but really, we're only seeking a better alternative for practical purposes. (Unnamed)
- *Agreed. This ID system will first and foremost provide us a way within the commons to translate between CASRNs and several other external and internal systems in use as well as to resolve ambiguities in the CASRN IDs and fill gaps in its addressing of the groupings and forms important to hazard. The usefulness of the ID outside of the Data Commons can be left to future discovery and is not necessary to our meeting the immediate needs.* (Unnamed)
- I would list the classes of substances that have consistently been a challenge to characterize due to the problem. It will help to focus more attention on the issue with technical folks who can help us to resolve the issue. Is the list of compounds on P36 a complete one or just representative? (NGO)
- If this is a proposal, can you speak to any traction you have in developing the SHIN. CAS numbers are pretty well institutionalized. Do you have any partners in government or American Chemical Society (to which Green Chemistry Institute belongs) who might support this work ? (I am on the board of the Green Chemistry Institute and would be glad to help if possible) (NGO)
- What is your strategy for building the SHIN? I think it would be helpful to engage good chemists in this process and some government people. You might even want to consider doing it as an ANSI standard because I don't think it is terribly controversial (famous last words?) and if you do it as an ANSI std then you will get government buy in. (NGO)
- Wikipedia actually does a nice job of listing large numbers of synonyms. (Academic)
- There seems to be an effort to pick a fight with the CAS in the nomenclature section. This is not necessary. Just state the limits and move on. (Academic)

Mixtures & Materials:

- **Mixtures**. There is one thoughtful paragraphs on mixtures. I wish this had been expanded, as this is a sleeping giant in any taxonomy like this. We humans think that chemicals exist in some pure form so that we can test them. However, chemicals in products are most often mixtures and chemicals are always reacting and concentrating and diffusing in any dynamic environment. There will always be surprising trace substances in any product and our single identifier/single chemical is a potentially misleading concept. (Academic)
- The paper stays mostly at the level of chemicals. However, most designers, retailers, consumers, etc. that you list as potential users are confronted with questions about materials and products. It

would be helpful to offer some means by which users of the Data Commons can get to the level of materials and products. This is not, as you know, a simple aggregation of chemicals, but some thought should go into this. (Academic)

- I agree that tracking down ingredients in commercial mixtures will be daunting. We have a lot of trouble finding this data. (Academic)
- This is probably a second-generation thing (don't sacrifice the good waiting for the perfect!) – how to integrate constructive/destructive interference on toxicity with chemical mixtures? (Academic)
- "Somewhere out there" I feel like a pretty compelling statistic exists that will tell you that XX% of all building/auto industry/other bulk or fine chemical user products are comprised of only N chemicals. This may not belong in this paper but certainly to bolster presentations. Meg and Mike would be good to check in with about this sort of number. (Their automotive industry surfactants work - Journal of Occupational and Environmental Hygiene, 2007, 4: 301–310) has had some pretty powerful impacts. (Academic)

Functional use & Alternatives assessment

- Functional Use. You note the Lowell Center functional use categories, but do not do much with them. It would be helpful to expand on this. The text notes that there is an on-going study of 16 functional use classifications. It would be helpful to describe this in more detail or (when completed) provide a summary table (probably in the Appendix). (Academic)
- Product Use Pattern. It would be easier for data users if there was some type of product "use pattern" classification developed. At a minimum, I would see Use Pattern tagging of existing reports (e.g., AAs). This way end users could easily tell if that existing document is relevant for the class of products that they manufacture. The reason that I think that use pattern matters so much is that it will affect the selection of relevant factors and the weightings of hazard categories in a well done AA. For example, an AA for copper in electronic assembled parts should have different "relevant factors" and different decision criteria than an AA for copper in vehicle brake pads or for copper in clothing fibers.

  In my view, the "functional use" definition proposed in the white paper is too broad. It can't really cover section C, the relationship to the environment, in any useful way. The process/environment listing examples in Section C are too diffuse to classify, and could be misleading (are automotive electronics really meaningfully different from other electronics?). And that approach isn't really the information that's needed. A product "use pattern" classification would be simpler and more useful. (It will also cross into AA guidance and help there if done correctly). (Consultant)
- This seems very focused on one-chemical replacements, despite introductory language about larger assessments. Is there anything that can be done to facilitate evaluation of alternatives that are not chemicals? Or products that entail wholesale redesign or reformulation? (Consultant)
- Although the paper is focused on building a "chemical" data commons, it seems important to point out the limitation of the "drop in" substitute approach. Many times there will not be a drop in substitute and there will be a requirement for a design or process change. Hopefully, the alternatives assessments that are collected in the data commons will not only be chemical hazard assessments but will offer other types of solutions to making safer products. (Academic)

Info Nodes

- Appendix. The table is very helpful. (Academic)
- *Note there have been a large number of suggestions from various reviewer for additional info nodes*

that are not noted here but have been directly incorporated in the Nodes chart.

Models to check out

- EU Footprint Project – Pesticide Database. This came into being largely to tackle the incompleteness of available data resources and the quality problems with some existing resources. http://www.eu-footprint.org/ppdb.html
- Pesticide Action Network - Database created by a respected scientist (Dr. Susan Kegley) with partial funding from government agencies. Its interface and completeness have generated a lot of fans for this system among a wide array of users (some of which are no fans of the host organization's policies) http://www.pesticideinfo.org/
- EPA has two separate environmental toxicity databases due to internal arguments about which type of data should be included. Neither is complete. They have different standards for data acceptance. The OPP database relies primarily on industry-sponsored studies, but because of their review process, the data are not necessarily all of lower quality.

  General EPA environmental toxicity database, ECOTOX: http://cfpub.epa.gov/ecotox/help.cfm?help_id=FAQ&help_type=define

  EPA Pesticides environmental toxicity database, also known as Ecotox (OPP): http://www.ipmcenters.org/Ecotox/index.cfm For an example of how the federal government has gotten twisted in knots dealing with the data quality problems of existing data, see http://tinyurl.com/nxtl5jb (Consultant)

Offers to help/requests to participate

- It is an exciting project, even knowing the challenges. I'd love to help out in any way I can. Recently I've been doing a lot documentation-type materials for Lauren, to faciliate the presentation and communication of complex information. I would also interested in research and compilation of the various data sets you will need for the tool (i.e., functional use data). (Consultant)
- The taxonomy work that you describe is really interesting to me, it is something I have been thinking a lot about recently, and I think it is a very necessary focus for this common knowledge platform. Let me know how I might help out as this thing moves along. (NGO)
- Biomimicry needs to align its position with chemical hazard assessment, and I have been looking for ways to do so. What we are now doing is aggregating a small number of companies -- of course they are in the HBN/HPD space -- to define our next Design Challenge. We are also procuring venture funding and academic partners to help incubate and mentor the winning teams. I'd love to figure out how to draw a strong nexus between the judging criteria and the entries. Of course the participants will need access to a database for material assessment, which may be a sticking point. Anyway, noodle on it as you're riding your bike from point A to point B. We're talking to BIFMA tomorrow about bringing on some of their members to help define the first challenge (ideas range from adhesives to particleboard alternatives). (NGO)
- Suggest chemical list users group to discuss issues with using authoritative hazard lists. Suggested Alex Stone also (NGO)
- I'd like to chat again about the intersections of our projects. I'm especially interested in how data projects like yours intersect with place-oriented health initiatives, particularly in the Bay Area, given all the progressive precedents unfolding here. (Academic)
- I'd love to discuss once you have a lull in the storm (Academic)
- I'm always open to a discussion of this issue to figure out how to increase access to information and assessments. (Assessor)
- it would be great to connect with you to discuss some options for collaboration that could

dramatically speed up the task of accomplishing much of what you propose in the doc. (Assessor)

Suggestions for others to participate

- I would recommend getting review and comment from **David Cebon** with **Granta Design** in the UK. I think he'd have a valuable perspective. His contact information is below. (Unnamed)
- Prof. Michael Frenklach & PRiME & is building a database of soot compositions (Academic)
- Something really great that Phil Jessop does at the ACS Green Chemistry summer school is to have students do cradle-to-gate on three routes to a particular substance. While at the time we were tired and cranky and annoyed that after a 12 hour day at school we had a homework assignment that was taking forever, the annoyingness of going through that lesson is very compelling to help me (and I suspect others) realize how important assembling this data is if we expect industry and government to do ANYTHING useful with alternatives assessment. (Academic)