



Commentary

# Catalyzing Knowledge-Driven Discovery in Environmental Health Sciences through a Community-Driven Harmonized Language

Stephanie D. Holmgren <sup>1,\*</sup> , Rebecca R. Boyles <sup>2</sup>, Ryan D. Cronk <sup>3</sup>, Christopher G. Duncan <sup>4</sup> , Richard K. Kwok <sup>5,6</sup> , Ruth M. Lunn <sup>7</sup>, Kimberly C. Osborn <sup>8</sup>, Anne E. Thessen <sup>9</sup> and Charles P. Schmitt <sup>1</sup>

- <sup>1</sup> Office of Data Science, National Institute of Environmental Health Sciences (NIEHS), Durham, NC 27709, USA; charles.schmitt@nih.gov
- <sup>2</sup> Research Computing, RTI International, Durham, NC 27709, USA; rboyles@rti.org
- <sup>3</sup> Health Sciences, ICF, Durham, NC 27713, USA; ryan.cronk@icf.com
- <sup>4</sup> Genes, Environment, and Health Branch, Division of Extramural Research and Training, NIEHS, Durham, NC 27709, USA; christopher.duncan@nih.gov
- <sup>5</sup> Epidemiology Branch, Division of Intramural Research, NIEHS, Durham, NC 27709, USA; richard.kwok@nih.gov
- <sup>6</sup> Office of the Director, NIEHS, Bethesda, MD 20892, USA
- <sup>7</sup> Integrative Health Assessment Branch, Division of the National Toxicology Program, NIEHS, Durham, NC 27709, USA; lunn@niehs.nih.gov
- <sup>8</sup> Health Sciences, ICF, Fairfax, VA 22031, USA; Kim.Osborn@icf.com
- <sup>9</sup> Environmental and Molecular Toxicology Department, Oregon State University, Corvallis, OR 97331, USA; annethessen@gmail.com
- \* Correspondence: holmgren@niehs.nih.gov



**Citation:** Holmgren, S.D.; Boyles, R.R.; Cronk, R.D.; Duncan, C.G.; Kwok, R.K.; Lunn, R.M.; Osborn, K.C.; Thessen, A.E.; Schmitt, C.P. Catalyzing Knowledge-Driven Discovery in Environmental Health Sciences through a Community-Driven Harmonized Language. *Int. J. Environ. Res. Public Health* **2021**, *18*, 8985. <https://doi.org/10.3390/ijerph18178985>

**Academic Editors:**  
Alireza Daneshkhah,  
Amin Hosseinian-Far, Vasile Palade  
and Samer A. Kharroubi

Received: 7 July 2021  
Accepted: 19 August 2021  
Published: 26 August 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

**Abstract:** Harmonized language is critical for helping researchers to find data, collecting scientific data to facilitate comparison, and performing pooled and meta-analyses. Using standard terms to link data to knowledge systems facilitates knowledge-driven analysis, allows for the use of biomedical knowledge bases for scientific interpretation and hypothesis generation, and increasingly supports artificial intelligence (AI) and machine learning. Due to the breadth of environmental health sciences (EHS) research and the continuous evolution in scientific methods, the gaps in standard terminologies, vocabularies, ontologies, and related tools hamper the capabilities to address large-scale, complex EHS research questions that require the integration of disparate data and knowledge sources. The results of prior workshops to advance a harmonized environmental health language demonstrate that future efforts should be sustained and grounded in scientific need. We describe a community initiative whose mission was to advance integrative environmental health sciences research via the development and adoption of a harmonized language. The products, outcomes, and recommendations developed and endorsed by this community are expected to enhance data collection and management efforts for NIEHS and the EHS community, making data more findable and interoperable. This initiative will provide a community of practice space to exchange information and expertise, be a coordination hub for identifying and prioritizing activities, and a collaboration platform for the development and adoption of semantic solutions. We encourage anyone interested in advancing this mission to engage in this community.

**Keywords:** community of practice; community-driven; controlled vocabulary; metadata; taxonomy; ontology; standards; semantic; environmental health; toxicology

## 1. Introduction

The use of a harmonized language to describe scientific methods and discoveries is well recognized as being critical for a variety of needs, including searching the literature, integrating data and knowledge, and conducting comparative analyses. Given the diversity

of environmental health sciences research and practice, the use of a harmonized language spanning biomedical sub-disciplines and other fields (e.g., environment, climate, disaster, population health) is important. Developing and applying a common, harmonized language is an ongoing and challenging task, as it must keep pace with scientific advances that generate new methods and subsequently new data types and knowledge.

We define “environment” broadly as any external influences on health, including physical, chemical, biological, social and cultural factors, and related behaviors [1]. An EHS language covers diverse scientific sub-domains and technical and semantic fields that include: data elements collected during scientific inquiries (e.g., PhenX measures, NHANES data elements, Tox21 endpoints, Disaster Research Response (DR2) tools); terminologies and metadata that are used to describe data elements; ontologies that formalize knowledge representation, such as the Gene Ontology (GO) [2]; mappings that link data elements and metadata to ontologies; and knowledge bases that combine and harmonize terminologies and ontologies (e.g., Comparative Toxicogenomics Database [3], Monarch Initiative [4]). Language harmonization efforts should support and promote the tools that help to capture, represent, and apply vocabulary elements, such as chemical read-across models [5] and gene enrichment analysis [6]. The future of an environmental health sciences language is expected to incorporate machine-oriented language constructs that are built on linguistic-based representations [7], such as word- and concept-embeddings [8], that facilitate machine learning methods to conduct data integration and reasoning [9–11]. Work to integrate human-focused vocabularies with machine-focused language constructs is progressing [12–14] and shows promise for biomedical analysis [15,16].

We highlight and seek support for a community-based initiative to help define and address gaps in the EHS language, including terminologies, ontologies, mapping, knowledge bases, tools, and emerging means of representing knowledge. Our initial work describes the formation of community processes, which includes the development of use cases and competency questions to identify the gaps and challenge the value of using existing approaches; workshops and webinars to identify and focus activities on targeted issues; and engagement with communities that are developing and applying standardized vocabularies in subfields (e.g., earth and social sciences). We emphasize a more sustainable approach that leverages the existing governance structures and communication platforms of the Research Data Alliance (RDA) model [17] to address focused use cases that are relevant to the sharing and integration of EHS data across heterogeneous sources and modalities [17]. These improvements address challenges that have hindered similar efforts in the past.

## 2. Discussion

To frame this objective, we present representative challenge areas and recent advances, followed by efforts to lay the foundation for a sustainable EHS language community.

### 2.1. Representative Challenge Areas

*Disease Modeling:* Incorporating exposure data in disease models is challenging because of the complicated metadata that is needed to characterize the exposure [18]. Issues surrounding the temporality of exposure, such as the timing relative to a life cycle [19], duration of exposure [20], frequency of exposure [21], latency [22], route of exposure [23], and point of contact are critical metadata [24]. Another modeling concern is that environmental exposures do not occur in isolation and interactions between multiple exposures can be critical to exposure health impacts [25]. Evidence codes [26] and probability measures [27] are needed to correctly weigh a piece of evidence in an integrated data set. Metadata increases the size and complexity of the model but are essential for correctly interpreting the data. In addition to the complex cause-and-effect relationships, translating exposure data from a human-readable to a computable format is difficult because the laboratories producing these data are not using a community-wide standard [28]. Developing, maintaining, and sustaining these standards is challenging for any community and requires significant time and resources [29].

*Systems Toxicology:* The field of toxicology continues to transition from predominately animal testing to a spectrum of in vivo, in vitro, and in silico approaches, with each focused on gaining specialized knowledge on outcomes and bio-chemical mechanisms [30]. Despite often high levels of control over toxicology testing conditions, challenges persist regarding accurately and systematically describing key observations across labs [31–33], which can lead to significant costs to integrate observations, where quality and data loss issues occur when data are integrated. The evolving mix of testing paradigms, organisms, in vitro platforms, and assays are increasing the diversity, volume, and specificity of toxicology data. This makes integration at the observational level more challenging and increases the need to integrate and compare data and findings at the knowledge level. As such, constructing bridges between the collection of toxicology assays and descriptions of bio-chemical and bio-mechanistic processes (e.g., adverse outcome pathways (AOPs), GO-causal activity models) is increasingly important [34,35] and a challenge as newer assays continue to emerge.

*Precision Medicine:* As precision medicine transitions to practice, a critical challenge is to identify both the genetic etiologies and environmental factors in disease in order to translate basic science into prospective interventions and advance healthcare [25]. If we include gene and phenotype relationships that are derived using model organisms, approximately 83% of known human coding genes are matched with available phenotype data [4], but the environmental component of these relationships is underrepresented in public databases and knowledge bases [36]. Without considering the effect of the environment, our ability to understand human disease and realize precision health is limited, even with the promise of modern genomics [37]. Achieving the integration of phenotype, genotype, and environmental information (including psychosocial stressors and cultural factors) requires an extensive translation of data into a computable form and the extension of the gene/phenotype data model, which are both done to ensure the discovery of extant data and to provide a structure that encourages new data discoveries and analyses. The informatics approaches that are needed to address these challenges have historically been focused on genomics, with less attention paid to additional types of data streams. Consequently, the types of algorithms that are used for genetic diagnostics are not accessible for diseases that have critical environmental components [38], such as a spectrum of environmental causes, exacerbations, compensatory mechanisms, repair, and potential therapeutic interventions.

## 2.2. Recent Efforts

Biomedical knowledge bases integrate information that is contained in terminologies, ontologies, and literature, providing great potential to find patterns in data that are hidden due to the volume, heterogeneity, and complexity of that data [39]. They are important resources for supporting the interpretation of findings and hypothesis generation by researchers. Efforts to integrate biomedical databases have resulted in the creation of several related, but often unconnected, biomedical knowledge graphs [40]. While these graphs are being used to better understand cancer [41], identify new drug candidates [42], and diagnose rare diseases [43], insights about the effect of environmental exposures are not forthcoming, largely because the environmental aspect of diseases is not included in these knowledge graphs. This is for two reasons: there are few curated data sources that associate environmental exposures to phenotypic outcomes in a structured manner [18,44], and the complexity of exposure science has not yet been modeled sufficiently using modern semantic structures to allow for large-scale data integration [18].

Development of the standards for associating environmental exposures to phenotypic outcomes and the associated metadata is a vast undertaking that requires substantial community engagement. The Comparative Toxicogenomics Database (CTD) has made substantial progress in aggregating and standardizing the associations between exposures, chemicals, genes, and diseases. Its exposure ontology (ExO) provides a semantic model of an exposure event and its outcome [3,45]. The Environmental Conditions, Treatments, and Exposures Ontology (ECTO) was developed based on the ExO to model exposure

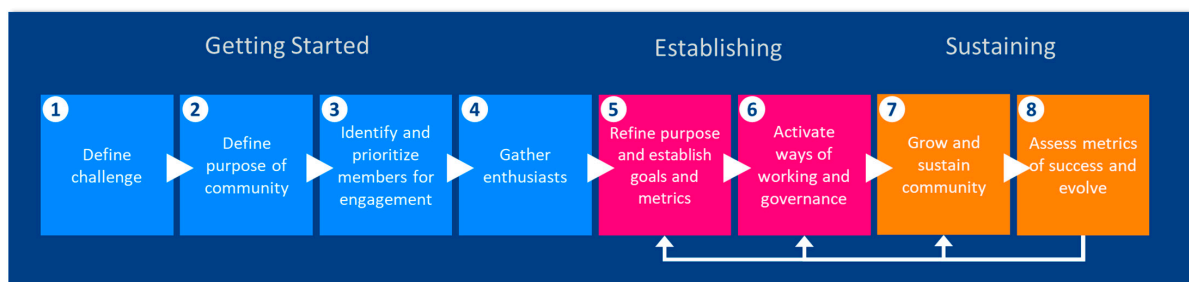
events for use in semantic models of disease, and the ENVO environmental ontology [46] provides linkages to environmental entities and processes. Several knowledge bases, such as the NCATS Translator [47], Monarch Initiative [48], and PheKnowLator [49], are incorporating environmental and chemical exposures from sources, such as the CTD and AOP Wiki [50], and linking that information to other knowledge constructs (e.g., diseases, phenotypes, genes, variants, therapeutics). A European strategy for exposure science has been proposed that includes substantial community building and data integration infrastructure development [51].

Relatedly, there have been promising advances in the development of data and meta-data standards and the mapping of data elements to ontologies. Efforts within the field of toxicology, including UMLS and Adverse Outcome Pathways (AOPs), map tox-related assays to standards [52]. The Children's Health Exposure Analysis Resource (CHEAR) and Human Health Exposure Analysis Resource (HHEAR) programs have developed an interdisciplinary CHEAR/HHEAR ontology that supports the analysis of the exposome through the harmonization of health and exposure data that is consistent across the program [53–55]. The NIEHS Superfund Basic Research and Training Program (SRP) has supported efforts to foster data sharing, interoperability, and reuse through the broader adoption of data standards and ontologies to support cross-center research collaborations [56]. These efforts help to advance cross-center standards and support for minimal information standards in the environmental health sciences domain, such as MIATE (Minimum Information about Animal Toxicology Experiments (in vivo) 2021). The National Institutes of Health (NIH) has increased focus on generating and adopting common data elements (CDEs), such as those in the PhenX Toolkit and the NLM Common Data Elements (CDE) Repository. Recent CDE development has been directed at supporting COVID-19 research through the Disaster Research Response (DR2) portal and their work with the RADx initiatives [57,58].

Despite these advances, more work is needed, particularly in implementing the sociological aspects of community-driven standards development. Several workshops have been held to mobilize the EHS community around standards development [59,60] and a strategy for toxicology ontology development was proposed [61]. The exposure science community is exploring the use of ontologies in research, both for data integration and gaining new insights [28,50,62–66], and a relatively new field of computational toxicology has gained popularity [67]. The collective recognition of research questions that require access to harmonized EHS data for analysis and the development of foundational semantic technology makes now the time to bring together the tools and the community.

### 3. Proposed EHS Community Model

We used the framework shown in Figure 1 to guide the development of the community approach. To initiate this process, the following steps were conducted: a review of prior EHS-related community-building efforts, a scan of research and guidance on collaboration and community building, interviews with successful communities, and working group discussions. From these actions, an approach that comprised a community organization model, a repository of use cases, and community events, was proposed. This approach is anticipated to evolve over the next year through feedback from planned community events.



**Figure 1.** Community development framework. Sources that influenced and guided this framework include [68–71].

*Previous workshops:* One of the first workshops to build an EHS language community was hosted by NIEHS in 2014 [59]. The intent of the workshop was to clarify research areas that would be advanced by using EHS language standards, identify stakeholders interested in creating a community to champion standards, and draft guidelines for the development of EHS standards. The workshop participants proposed eight guiding principles for establishing a community: engage a broad community, facilitate collaboration, enable the navigation of existing standards, support the citation and attribution of standards, adopt software development best practices, assist funding early-stage development, create a sustainable and flexible framework, and capitalize on opportunistic development. The formal workshop recommendations included federal funding to ensure the development, expansion, and adoption of standards, as well as a phased approach to development. The first phase addressed the need to identify relevant EHS research questions and use cases for the immediate application of semantic standards. Phase two suggested the development of a web-based standards toolkit to enable easier navigation of existing standards, the extraction of terms for specific project needs, and contribute terms to expand the standard. The community raised the need for a governance and sustainability plan.

A Computable Exposures Workshop was held in 2020 to “foster the development of data reporting standards and a computational model which will facilitate the inclusion of exposure data in computational analysis of human data” [60]. Four important gaps were identified at the workshop that prevented the adoption and use of computable exposures:

- A minimum reporting standard for exposure science and toxicology.
- Curated mappings across chemical authorities.
- A semantic model for exposure data.
- Ontological coverage.

Participants in this workshop developed use cases and competency questions to guide infrastructure development to fill these gaps. At the end of the workshop, a semantic model for exposure data was proposed. Attempts to make progress on these gaps are ongoing and will contribute to the proposed community effort.

The level of engagement and the well-defined next steps at these workshops indicate that support exists for the development of a sustained community to advance this work.

*Community building and collaboration literature scan:* To guide the formation of an EHS language community, we reviewed case studies and literature reviews of community-building efforts in EHS and similar fields. A success factor that was stressed by Arnaud et al. in the development of an ontology CoP for the Consultative Group on International Agricultural Research (CGIAR) Platform for Big Data in Agriculture is the necessity of the regular engagement of CoP members across relevant networks in data curation for biological, food and agronomic, and socioeconomics research [72]. In a case study on developing a CoP for scientific programming for life scientists, Stevens et al. indicated that identifying a core group of dedicated individuals and identifying champions who can become leaders were factors contributing to the successful formation of CoPs [73]. Pyrko, Dorfler and Eden [70], in their study of CoPs, emphasized that CoPs foster “thinking together” and that “thinking together” is necessary for CoPs to thrive.

*Interviews with existing communities:* Because community building is challenging, we wanted to learn how sustainable and impactful communities have successfully overcome these challenges and what lessons they learned along the way. For this purpose, interviews were conducted with participants that were familiar with the origin story of several communities: Earth Sciences Information Partnership (ESIP), Adverse Outcome Pathway Wiki (AOP Wiki), RDA, and the Open Biological and Biomedical Ontologies (OBO) Foundry. The goals were to learn why and how these communities formed and to understand the essential components or actions that were needed to form, grow, and sustain an effective EHS language community. As a result of these interviews, we distilled several takeaways for creating and sustaining a successful community:

- Form the community around a defined/shared purpose. The community needs to identify its purpose and have a clear understanding of its goals.

- Start with a small circle of champions who can communicate the value of the community.
- Have committed/dedicated financial, technical, and labor resources. Successful communities have an infrastructure to support administrative operations.
- Create a sense of “I found my people” among the members.
- Target a specific action to undertake and grow from there.
- Identify the incentives that are needed to get people actively engaged. While the most likely incentive is that the community activities align with the person’s work-related tasks, some members are simply motivated to make a difference.
- Activate ways of working that meet the community’s culture (e.g., formal versus informal governance, preferred channels of communication).

### 3.1. Proposed Community Organization

This section describes the name, mission, goals, and structure of the proposed community. These are being put forward as starting points for community discussion at the upcoming pre-workshop and workshop events.

The proposed vision of the Environmental Health Language Collaborative is to leverage community-driven environmental health language standards to catalyze knowledge-driven discovery and improve public health.

The mission of the Collaborative would be to advance integrative environmental health sciences research by developing and promoting the adoption of a harmonized language.

To achieve this mission, the community would:

- Define use cases for applying knowledge organization systems in research.
- Foster community-based development of harmonized vocabularies, terminologies, and ontologies.
- Promote and develop methods and tools for applying harmonized language in research.
- Cultivate a vocabulary-aware environmental health community through training and education.
- Apply language standards and best practices for accurate environmental health data and knowledge representation

To implement this mission, the community would comprise three elements:

**Community of practice:** A community of practice provides a hub to exchange information, ideas, and expertise, as well as advance the appreciation for and adoption of semantic and language approaches through education and training.

**Forum for coordination and collaboration:** The community serves as a forum to coordinate harmonization activities and collaborate on defining use cases and gaps, prioritizing activities, and developing the language strategies or approaches that are needed for enabling data querying, sharing, and interoperability.

**Platform to develop and implement:** Based on the identified gaps in the use cases, the community serves to support and promote the development and application of harmonized language solutions to address the use cases’ needs.

Based on interviews with organizations and discussions within a community model working group, the community model shown in Figure 2 was proposed. A key aspect of any community is having an infrastructure for communications, hosting meetings, and other daily operational activities. One of the recurring messages from stakeholders and community interviews was to not “reinvent the wheel.” As such, the Research Data Alliance (RDA) was proposed to provide structure for the EHS community.

The RDA started in 2013 with funding from the European Commission, National Science Foundation, National Institute of Standards and Technology, and Australia’s Department of Innovation, with the mission to “build the social and technical bridges to enable open sharing and reuse of data to accelerate data-driven innovation.” It is a community-driven, grassroots organization with more than 11,000 members from 145 countries. Individual membership is free. Individuals come from a variety of disciplines and professions,

including researchers, IT architects, project managers, data scientists, publishers, and librarians. Through interest groups and working groups, members exchange knowledge and share discoveries, discuss barriers and potential solutions, define policies, and seek to harmonize standards to enhance/facilitate global data sharing and re-use. The goals and activities of the RDA align with the proposed EHS community mission, and participation within RDA provides an opportunity to tap into international expertise and perspectives from other related disciplines [74,75].

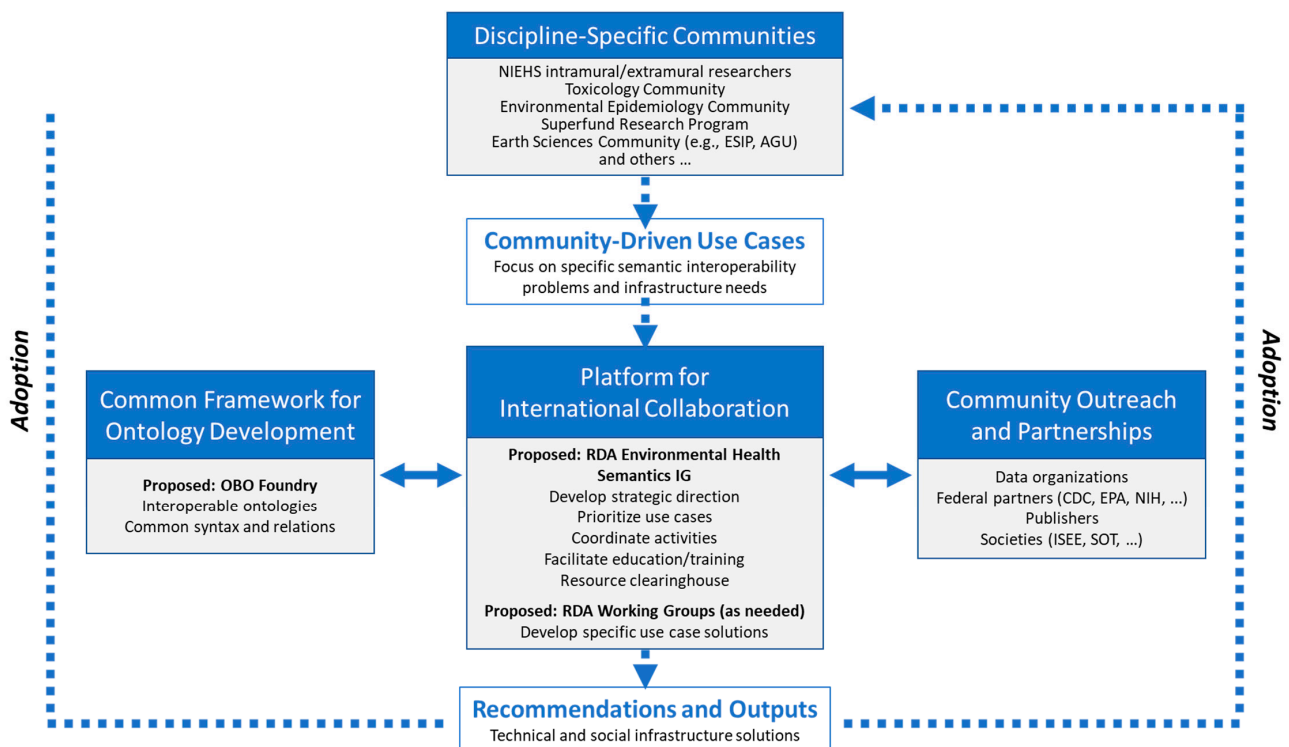


Figure 2. Proposed community model.

The proposed model begins with individuals and/or groups from discipline-specific communities that generate use cases based on research questions that are of interest to them. These use cases represent the need for harmonized language solutions that will enhance the findability, sharing, and interoperability of EHS data. The use cases will be brought to a proposed RDA Environmental Health Language Interest Group (IG). This IG will provide a platform for the overall coordination and collaboration between interested members. Its goal is to design a strategic direction for developing and adopting semantic solutions, identify and prioritize use cases, coordinate activities, and be a CoP for exchanging information, offering a resource clearinghouse, and fostering education/training. An RDA Working Group could be formed whenever a specific work product needs to be developed. If the product is an ontology, then ideally, its development would follow the OBO Foundry framework to be interoperable with other ontologies.

The IG and WG(s) will work in concert with other relevant communities or partner organizations toward the development and implementation of any recommendations and outputs. Those products will be communicated back to the discipline-specific communities with the anticipation of adoption.

*How would this model work in practice?* The intent of the model is to provide support to those developing and applying semantic approaches, as outlined in Figure 3. The example begins with an investigator (or someone else) who has a use case that can benefit from a semantic solution. At this stage, the investigator can work with the RDA Interest Group to raise awareness of the needs, tap into expertise, and identify potential collaborators to

work on a team. Use case leads may choose to form a working group outside of the RDA, but they can also decide that creating an RDA working group will assist in gaining broader community input and perspectives. Whether the activities are done within or outside an RDA WG, the IG can support the working group’s activities by offering time at the IG’s plenary sessions to do work and/or provide additional support in the form of workshop activities, presentation time, and webinars. Any developed product(s) from the working group would be brought to the RDA and shared with the broader community, as well as added to a resource clearinghouse. In addition, the RDA IG can assist with disseminating and promoting the adoption of the product if needed. The RDA IG will maintain the catalog of existing use cases, which will aid others in recognizing and prioritizing gaps and issues to which they can provide solutions.

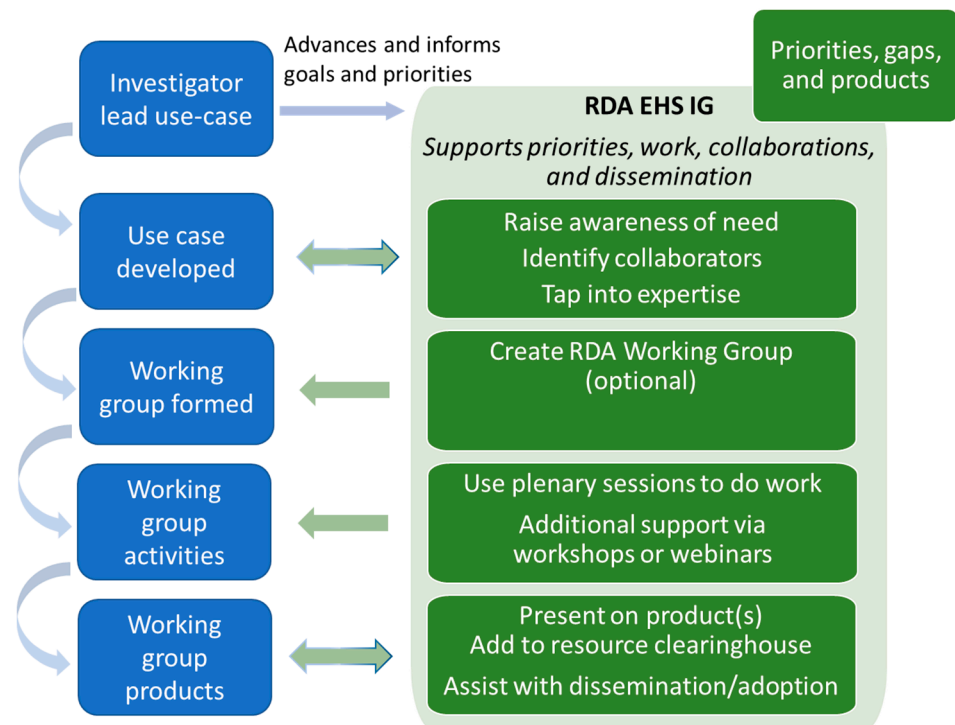


Figure 3. The community model in practice.

Sustaining the proposed community model requires three supporting players, as shown in Figure 4. The NIEHS proposes to engage by providing in-kind volunteer support to the IG and WGs and working to develop funding strategies for relevant efforts. The NIEHS will help support workshops or other events, such as codeathons, as well as develop policies and processes based on the RDA or other’s recommendations that would advance the community’s goals. In-kind volunteer support will be needed from discipline-specific communities, primarily through serving on the IG and WG. Finally, collaborating partners in academic, federal, and industry sectors will be identified and involved to provide both in-kind contributions, support for funding community activities, and promoting the adoption of outputs.



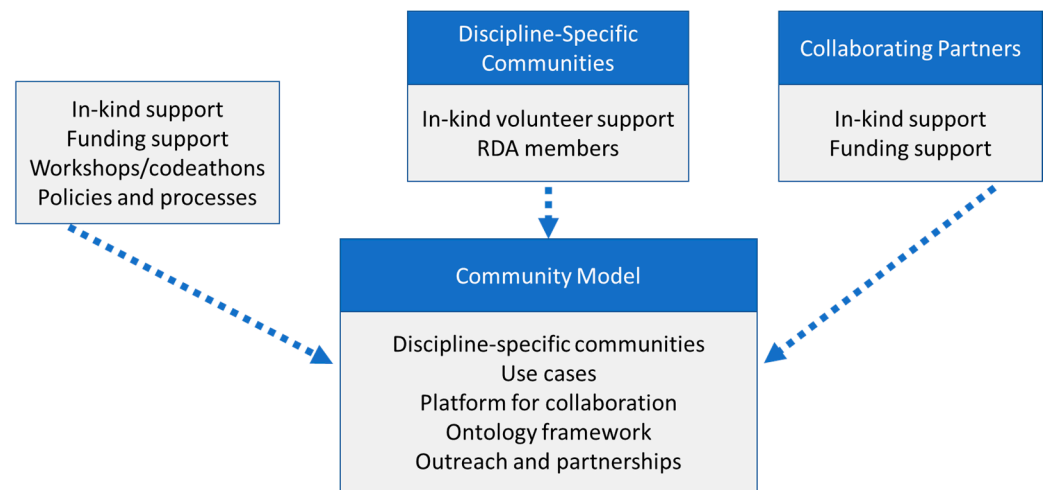


Figure 4. Sustaining the community.

### 3.2. Community Events

The community effort (Figure 5) will launch with a virtual workshop entitled *Catalyzing Knowledge-Driven Discovery in Environmental Health Sciences through a Harmonized Language* on 9–10 September 2021 and will consist of two tracks. The first track will focus on “Build a Sustainable Community” with the goal of achieving agreement on the community’s mission and goals and the development of a roadmap for governance, outreach, and partnerships. The second track, “Develop Solutions,” will be an interactive workathon that will be dedicated to initiating the development of semantic solutions for specific use cases. Participants in this session will define draft use cases in detail and propose and discuss solutions. The desired outputs from this session include (near) complete use case packages that describe the research question, available resources to address the use case, and specified gaps that require solutions. Some use cases may be at the stage of outlining the next steps for developing solutions. We aspire to have working groups form to continue work on the use cases post workshop. A workshop paper will be developed that details the workshop activities, outcomes, and next steps.

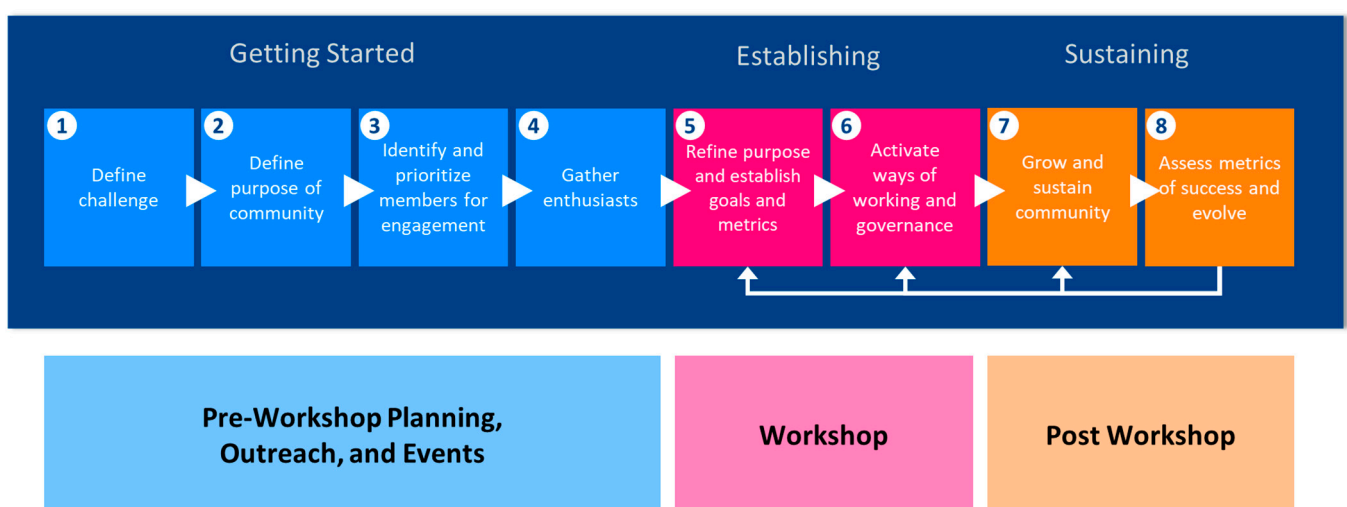


Figure 5. Components of building a community.

In advance of the workshop, we held several pre-workshop events (see Table 1) to provide background on the initiative, allow researchers to learn more about ontologies

(e.g., what they are, how to find them, how to apply them), and begin work on pre-defined use cases that will inform workshop sessions.

**Table 1.** Events for the Environmental Health Language Collaborative.

Event	Date
The Value of Creating Language and Community in Catalyzing Knowledge-Driven Discovery in Environmental Health Research (virtual)	24 June 2021
A Primer on Using Terminologies, Vocabularies, and Ontologies for Knowledge Organization (virtual)	20 July 2021
Catalyzing Knowledge-Driven Discovery in Environmental Health Sciences through a Harmonized Language (virtual)	9–10 September 2021

### 3.3. Use Cases

The authors established a working group of EHS researchers and program officers in 2020 to develop an initial set of draft use cases. These initial use cases were placed into five high-level use case categories with the original use cases serving as example sub-cases. In several instances, the use cases require not only advances in standardized vocabularies but also in statistical and modeling approaches, which represent opportunities to engage with those communities. Although use cases overlap and some consolidation is possible, the use cases are provided in their near original form to avoid errors/simplifications that might result.

Each use case is being championed by a subject matter expert that is familiar with the research question. The champions will be holding small working group meetings to develop a draft use case package (Supplementary File S1) to provide a focus of discussion for the *Develop Solutions* track at the September workshop. The use case package will include a clearer definition of the use case research question, available datasets and ontologies/terminologies that can be used for developing solutions, existing gaps that need to be addressed, and other non-language-related challenges that need to be known.

#### **Use case #1: What data exists for a given chemical/endpoint/exposure scenario?**

This use case is targeted at finding published data related to a specific set of exposure scenarios. Example sub-cases include:

- What studies measuring endocrine systems perturbation are available?
- What chemicals are chemically similar to compound X and are there any 2-year cancer bioassay data available for these chemicals?
- What animal data exist that provides conclusions on endpoint X given different terms used to describe endpoint X?
- What other data are available for chemical X when it is found in a formulation?
- What assays were “active” for this chemical (where “active” may have different meanings across assays)?

#### **Use case #2: Combine data from multiple independent studies (e.g., heterogeneous study designs, data collection protocols)?**

This use case is targeted at the issue of integrating data from independent exposure studies. Example sub-cases include:

- Combine individual-level data from multiple independent studies (heterogeneous study designs and data collection protocols) to understand (with increased statistical power) how exposures X and Y impact health outcome Z.
- How can we describe model organism toxicological assays/data in a way that is interoperable and reusable to better understand the phenotypic/epigenomic/transcriptomic impact of exposures X and Y across species A and B?
- Integrate and compare data across labs to support more robust corroboration in the confidence of results from toxicological assessments.

- Given conclusive changes in endpoints to one or more exposures, what other data sources exist on the same exposures and endpoints that can confirm or contradict the findings, including across similar endpoints across different species?
- Given natural text mentions of concepts from scientific studies, what ontology(ies) do these mentions map to in order to normalize terminologies across 100–1000s of studies?

**Use case #3: Given the measures of biological responses to one or more exposures, what are the biological processes that might be related to the observed changes?**

This use case is targeted at supporting the mechanistic interpretations and hypothesis generation that arise given experimental data. Example sub-cases include:

- Given conclusive changes in endpoints to one or more exposures, what are biological processes that might lead to the observed changes?
- How can we use a knowledge graph to fill in the adverse outcome or adverse exposure pathways based on the start or end of the pathway?
- What other modes of action/adverse outcome pathways does this assay hit?
- What assays target this mode of action or key event?
- Given an association between exposure and outcome found in an epidemiological study, find the in vivo and in vitro studies that lend support to the association and that suggest involved bioprocesses, including associations that are dependent on developmental windows.
- Given the signatures of biological responses to exposures from multiple modalities (e.g., gene expression, pathology), can we link these signatures to known biological phenotypes and processes to characterize response signatures and to identify gaps in characterizations?
- Can we link a set of available assays (e.g., in PubChem) to known biological processes and phenotypes in order to better characterize chemical exposures?

**Use case #4: What are the biomarkers, phenotypes, and/or outcomes that can be measured and used as indicators of exposure?**

This use case is targeted at identifying known and hypothesized mechanistic markers of exposures. Example sub-cases include:

- What biomarkers can be used to examine exposure to a given chemical?
- Can we identify biomarkers for different classes of exposures (e.g., exposures to metals/metalloids in soil via dust inhalation, exposure to common pesticides via well water) that are contextualized by delivery route?
- Given conclusive changes in endpoints in response to one or more exposures, what other data sources exist on the same exposures and endpoints that can confirm or contradict the findings, including across similar endpoints across different species?

**Use case #5: What do my unique exposure conditions based on where I live and work (e.g., geographical location, occupation, regulations, hobbies) indicate about potential risks to my health?**

This use case focuses on the linkage of known environmental exposure data to personal health. Example sub-cases include:

- What is my biggest exposure risk based on my geographical location?
- What am I exposed to in my particular line of work? How might this impact my health?
- For what components of X industrial emission do we need more information on health outcomes?
- What levels of exposure to X will decrease the risk of health outcomes?
- What are the health and economic benefits from regulations or policies that reduce exposure to X?
- What are my biggest exposure risks based on work-life conditions, especially where I live and work (work, geography, hobbies)? What is the route of exposure that is most relevant to my specific conditions?

- How does the response to exposure change based on susceptibility (e.g., genetic, disease, SES backgrounds, differences between signatures of exposures, and differences of risk)?

### 3.4. Anticipated Outcomes

The broader biomedical research community is placing increased emphasis on data stewardship and data sharing. For instance, the new NIH Policy on Data Management and Sharing will require NIH-funded researchers to prospectively submit a plan that outlines how data will be managed and shared [76]. Data standards are a key element of these Data Management and Sharing Plans. To be interoperable, (meta)data should use vocabularies that follow the FAIR principles [77]. Community-endorsed data standards and vocabularies are key but are currently a major gap area for EHS. The proposed effort provides a much-needed vehicle for community engagement to address these documented needs.

Previous EHS experience and input from parallel communities emphasize the importance of sustained engagement to move forward. Leveraging the existing international infrastructure through RDA meets this need. Using the RDA model as a guide, some of the anticipated outcomes from these efforts include data and metadata models, roadmaps, extensions to existing EHS-relevant ontologies, and developing new EHS-relevant semantic approaches that will facilitate integration across heterogeneous data and nomenclature. This will not only foster data harmonization efforts but improve statistical power and the ability to generate hypotheses more efficiently. The products, outcomes, and recommendations endorsed by this community are expected to enhance data collection and management efforts for NIEHS and the EHS community, making data more findable and interoperable. These efforts can be fed into complementary efforts, such as the Canadian Urban Environmental Health Research Consortium (CANUE) project (<https://canue.ca/about/>, accessed on 23 August 2021), whose goals include providing a centralized repository for data and tools to study environmental health outcomes.

## 4. Contribute to the Community

This community is open to anyone interested in advancing environmental health sciences research through the development and adoption of a harmonized language. This community will benefit from diverse perspectives—it needs a range of subject matter experts, skill sets, and roles to be represented.

How can you become engaged in this effort?

- Review the materials from previous workshop events at <https://www.niehs.nih.gov/research/programs/ehlc/resources/index.cfm> (accessed on 23 August 2021).
- Provide input on the proposed community initiative and use cases at <https://www.niehs.nih.gov/research/programs/ehlc/> (accessed on 23 August 2021)
- Sign up for our email distribution list to be informed of future events and join the community of researchers, systems developers, ontologists, and others interested in working together on language standards in the environmental health sciences.

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/ijerph18178985/s1>, File S1: Environmental Health Use Case Template.

**Author Contributions:** Conceptualization, S.D.H. and C.P.S.; writing—original draft preparation, S.D.H. and C.G.D.; writing—review and editing, S.D.H., R.R.B., R.D.C., C.G.D., R.K.K., R.M.L., K.C.O., A.E.T. and C.P.S. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by the National Institute of Environmental Health Sciences.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** We would like to thank the following people for sharing their time and insights about their communities: Alan Blatecky, RDA; Steve Edwards, AOP Wiki; Melissa Haendel, OBO Foundry; and Chris Lehnert, ESIP. We would also like to thank these individuals for sharing their insights on use cases: Michelle Angrish, EPA; Steve Edwards, RTI; Rachel Morello-Frosch, University of California at Berkeley; Carmen Marsit, Emory University; Chirag Patel, Harvard; Jeanette Stingone, Columbia University; and Robyn Tanguay, Oregon State University. The authors are members of the Program Planning Committee for the *Catalyzing Knowledge-Driven Discovery in Environmental Health Sciences* Workshop and would like to acknowledge other committee members who have contributed to the development of the approach: Shannon Bell, Camden Byrd, Jennifer Fostel, Vickie Walker, and Jessica Wignall. Non-NIEHS author contributions were provided as part of contract support.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Prüss-Üstün, A.; Corvalán, C. *Preventing Disease through Healthy Environments. Towards an Estimate of the Environmental Burden of Disease*; WHO: Geneva, Switzerland, 2006. Available online: [https://apps.who.int/iris/bitstream/handle/10665/43457/9241593822\\_eng.pdf?sequence=1&isAllowed=y](https://apps.who.int/iris/bitstream/handle/10665/43457/9241593822_eng.pdf?sequence=1&isAllowed=y) (accessed on 23 August 2021).
2. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene ontology: Tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [[CrossRef](#)] [[PubMed](#)]
3. Davis, A.P.; Grondin, C.J.; Johnson, R.J.; Sciaky, D.; McMorran, R.; Wieggers, J.; Wieggers, T.C.; Mattingly, C.J. The Comparative Toxicogenomics Database: Update 2019. *Nucleic. Acids. Res.* **2019**, *47*, D948–D954. [[CrossRef](#)]
4. Mungall, C.J.; McMurry, J.A.; Köhler, S.; Balhoff, J.P.; Borromeo, C.; Brush, M.; Carbon, S.; Conlin, T.; Dunn, N.; Engelstad, M.; et al. The Monarch Initiative: An integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic. Acids. Res.* **2017**, *45*, D712–D722. [[CrossRef](#)]
5. Patlewicz, G.; Cronin, M.T.D.; Helman, G.; Lambert, J.C.; Lizarraga, L.E.; Shah, I. Navigating through the minefield of read-across frameworks: A commentary perspective. *Comput. Toxicol.* **2018**, *6*, 39–54. [[CrossRef](#)]
6. Mi, H.; Muruganujan, A.; Ebert, D.; Huang, X.; Thomas, P.D. PANTHER version 14: More genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic. Acids. Res.* **2019**, *47*, D419–D426. [[CrossRef](#)] [[PubMed](#)]
7. Landauer, T.K.; Dumais, S.T. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction, and representation of knowledge. *Psych. Rev.* **1997**, *104*, 211–240. [[CrossRef](#)]
8. Chen, Q.; Lee, K.; Yan, S.; Kim, S.; Wei, C.H.; Lu, Z. BioConceptVec: Creating and evaluating literature-based biomedical concept embeddings on a large scale. *PLoS Comput. Biol.* **2020**, *16*, e1007617. [[CrossRef](#)] [[PubMed](#)]
9. Cappuzzo, R.; Papotti, P.; Thirumuruganathan, S. Creating Embeddings of Heterogeneous Relational Datasets for Data Integration Tasks. In Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, Portland, OR, USA, 14–19 June 2020; pp. 1335–1349.
10. Hudson, I.L. Data integration using advances in machine learning in drug discovery and molecular biology. In *Artificial Neural Networks. Methods in Molecular Biology*; Cartwright, H., Ed.; Humana Press: New York, NY, USA, 2021; pp. 167–184.
11. Jiang, S.; Wu, W.; Tomita, N.; Ganoe, C.; Hassanpour, S. Multi-Ontology Refined Embeddings (MORE): A hybrid multi-ontology and corpus-based semantic representation model for biomedical concepts. *J. Biomed. Inform.* **2020**, *111*, 103581. [[CrossRef](#)]
12. Alshahrani, M.; Khan, M.A.; Maddouri, O.; Kinjo, A.R.; Queralt-Rosinach, N.; Hoehndorf, R. Neuro-symbolic representation learning on biological knowledge graphs. *Bioinformatics* **2017**, *33*, 2723–2730. [[CrossRef](#)]
13. Arguello Casteleiro, M.; Demetriou, G.; Read, W.; Fernandez Prieto, M.J.; Maroto, N.; Maseda Fernandez, D.; Nenadic, G.; Klein, J.; Keane, J.; Stevens, R. Deep learning meets ontologies: Experiments to anchor the cardiovascular disease ontology in the biomedical literature. *J. Biomed. Semantics* **2018**, *9*, 13. [[CrossRef](#)]
14. Smaili, F.Z.; Gao, X.; Hoehndorf, R. Onto2Vec: Joint vector-based representation of biological entities and their ontology-based annotations. *Bioinformatics* **2018**, *34*, i52–i60. [[CrossRef](#)] [[PubMed](#)]
15. Alshahrani, M.; Hoehndorf, R. Semantic Disease Gene Embeddings (SmuDGE): Phenotype-based disease gene prioritization without phenotypes. *Bioinformatics* **2018**, *34*, i901–i907. [[CrossRef](#)]
16. Kulmanov, M.; Khan, M.A.; Hoehndorf, R.; Wren, J. DeepGO: Predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics* **2018**, *34*, 660–668. [[CrossRef](#)]
17. Treloar, A. The Research Data Alliance: Globally co-ordinated action against barriers to data publishing and sharing. *Learn. Pub.* **2014**, *27*, 9–13. [[CrossRef](#)]
18. Richarz, A.-N. Chapter 1: Big data in predictive toxicology: Challenges, opportunities and perspectives. In *Big Data in Predictive Toxicology*; Royal Society of Chemistry: London, UK, 2019; pp. 1–37.
19. Aaseth, J.; Wallace, D.R.; Vejrup, K.; Alexander, J. Methylmercury and developmental neurotoxicity: A global concern. *Curr. Opin. Toxicol.* **2020**, *19*, 80–87. [[CrossRef](#)]
20. Lee, W.V.; Steemers, K. Exposure duration in overheating assessments: A retrofit modelling study. *Build. Res. Inf.* **2017**, *45*, 60–82. [[CrossRef](#)]

21. Spear, L.P. Timing Eclipses Amount: The Critical Importance of Intermittency in Alcohol Exposure Effects. *Alcohol Clin. Exp. Res.* **2020**, *44*, 806–813. [CrossRef] [PubMed]
22. Gwinn, M.R.; DeVoney, D.; Jarabek, A.M.; Sonawane, B.; Wheeler, J.; Weissman, D.N.; Masten, S.; Thompson, C. Meeting report: Mode(s) of action of asbestos and related mineral fibers. *Environ. Health Perspect.* **2011**, *119*, 1806–1810. [CrossRef]
23. Custer, K.W.; Hammerschmidt, C.R.; Burton, G.A., Jr. Nickel toxicity to benthic organisms: The role of dissolved organic carbon, suspended solids, and route of exposure. *Environ. Pollut.* **2016**, *208*, 309–317. [CrossRef]
24. U.S. EPA (U.S. Environmental Protection Agency). *Guidelines for Human Exposure Assessment*; (EPA/100/B-19/001); Risk Assessment Forum, U.S. EPA: Washington, DC, USA, 2019. Available online: [https://www.epa.gov/sites/default/files/2020-01/documents/guidelines\\_for\\_human\\_exposure\\_assessment\\_final2019.pdf](https://www.epa.gov/sites/default/files/2020-01/documents/guidelines_for_human_exposure_assessment_final2019.pdf) (accessed on 23 August 2021).
25. Vineis, P.; Russo, F.E. Epigenetics and the exposome: Environmental exposure in disease etiology. In *Oxford Research Encyclopedia of Environmental Science*; Oxford University Press: Oxford, UK, 2018.
26. Rogers, M.F.; Ben-Hur, A. The use of gene ontology evidence codes in preventing classifier assessment bias. *Bioinformatics* **2009**, *25*, 1173–1177. [CrossRef] [PubMed]
27. Kluxen, F.M. "New statistics" in regulatory toxicology? *Regul. Toxicol. Pharmacol.* **2020**, *117*, 104763. [CrossRef]
28. Boyles, R.R.; Thessen, A.E.; Waldrop, A.; Haendel, M.A. Ontology-based data integration for advancing toxicological knowledge. *Curr. Opin. Toxicol.* **2019**, *16*, 67–74. [CrossRef]
29. Hankin, S.; Blower, J.D.; Carval, T.; Casey, K.S.; Donlon, C.; Lauret, O.; Loubrieu, T.; Srinivasan, A.; Trinanes, J.; Godøy, Ø.; et al. NetCDF-CF-OPeNDAP: Standards for Ocean Data Interoperability and Object Lessons for Community Data Standards Processes. In Proceedings of the OceanObs 2009, Venezia, Italy, 21–25 September 2009.
30. Knudsen, T.B. Roadmap for animal-free reproductive toxicity testing: Predictive toxicology and computational embryology. Presented at Scientific Roadmap for the Future of Animal-free Systemic Toxicity Testing Workshop, College Park, MD, USA, 30–31 May 2013.
31. Alghamdi, S.M.; Sundberg, B.A.; Sundberg, J.P.; Schofield, P.N.; Hoehndorf, R. Quantitative evaluation of ontology design patterns for combining pathology and anatomy ontologies. *Sci. Rep.* **2019**, *9*, 4025. [CrossRef] [PubMed]
32. Aulbach, A.; Vitsky, A.; Arndt, T.; Ramaiah, L.; Logan, M.; Siska, W.; Cregar, L.; Tripathi, N.; Adedeji, A.; Provencher, A.; et al. Overview and considerations for the reporting of clinical pathology interpretations in nonclinical toxicology studies. *Vet. Clin. Pathol.* **2019**, *48*, 389–399. [CrossRef]
33. Viant, M.R.; Ebbels, T.M.D.; Beger, R.D.; Ekman, D.R.; Epps, D.J.T.; Kamp, H.; Leonards, P.E.G.; Loizou, G.D.; MacRae, J.I.; van Ravenzwaay, B.; et al. Use cases, best practice and reporting standards for metabolomics in regulatory toxicology. *Nat. Commun.* **2019**, *10*, 3041. [CrossRef]
34. Escher, B.I.; Hackermuller, J.; Polte, T.; Scholz, S.; Aigner, A.; Altenburger, R.; Bohme, A.; Bopp, S.K.; Brack, W.; Busch, W.; et al. From the exposome to mechanistic understanding of chemical-induced adverse effects. *Environ. Int.* **2017**, *99*, 97–106. [CrossRef]
35. Vermeulen, R.; Schymanski, E.L.; Barabasi, A.L.; Miller, G.W. The exposome and health: Where chemistry meets biology. *Science* **2020**, *367*, 392–396. [CrossRef]
36. Vineis, P. A self-fulfilling prophecy: Are we underestimating the role of the environment in gene-environment interaction research? *Int. J. Epidemiol.* **2004**, *33*, 945–946. [CrossRef]
37. Wesseling, C.; Glaser, J.; Rodriguez-Guzman, J.; Weiss, I.; Lucas, R.; Peraza, S.; da Silva, A.S.; Hansson, E.; Johnson, R.J.; Hogstedt, C.; et al. Chronic kidney disease of non-traditional origin in Mesoamerica: A disease primarily driven by occupational heat stress. *Rev. Panam. Salud Publica* **2020**, *44*, e15. [CrossRef] [PubMed]
38. Smedley, D.; Jacobsen, J.O.; Jager, M.; Kohler, S.; Holtgrewe, M.; Schubach, M.; Siragusa, E.; Zemojtel, T.; Buske, O.J.; Washington, N.L.; et al. Next-generation diagnostics and disease-gene discovery with the Exomiser. *Nat. Protoc.* **2015**, *10*, 2004–2015. [CrossRef] [PubMed]
39. Bakal, G.; Talari, P.; Kakani, E.V.; Kavuluru, R. Exploiting semantic patterns over biomedical knowledge graphs for predicting treatment and causative relations. *J. Biomed. Inform.* **2018**, *82*, 189–199. [CrossRef] [PubMed]
40. Nicholson, D.N.; Greene, C.S. Constructing knowledge graphs and their biomedical applications. *Comput. Struct. Biotechnol. J.* **2020**, *18*, 1414–1428. [CrossRef] [PubMed]
41. Hasan, S.M.S.; Rivera, D.; Wu, X.C.; Durbin, E.B.; Christian, J.B.; Tourassi, G. Knowledge Graph-Enabled Cancer Data Analytics. *IEEE J. Biomed. Health Inform.* **2020**, *24*, 1952–1967. [CrossRef]
42. Reese, J.T.; Unni, D.; Callahan, T.J.; Cappelletti, L.; Ravanmehr, V.; Carbon, S.; Shefchek, K.A.; Good, B.M.; Balhoff, J.P.; Fontana, T.; et al. KG-COVID-19: A Framework to Produce Customized Knowledge Graphs for COVID-19 Response. *Patterns* **2021**, *2*, 100155. [CrossRef]
43. Zemojtel, T.; Kohler, S.; Mackenroth, L.; Jager, M.; Hecht, J.; Krawitz, P.; Graul-Neumann, L.; Doelken, S.; Ehmke, N.; Spielmann, M.; et al. Effective diagnosis of genetic disease by computational phenotype analysis of the disease-associated genome. *Sci. Transl. Med.* **2014**, *6*, 252ra123. [CrossRef] [PubMed]
44. Davis, A.P.; Wiegers, J.; Wiegers, T.C.; Mattingly, C.J. Public data sources to support systems toxicology applications. *Curr. Opin. Toxicol.* **2019**, *16*, 17–24. [CrossRef] [PubMed]
45. Mattingly, C.J.; McKone, T.E.; Callahan, M.A.; Blake, J.A.; Hubal, E.A. Providing the missing link: The exposure science ontology ExO. *Environ. Sci. Technol.* **2012**, *46*, 3046–3053. [CrossRef] [PubMed]

46. Buttigieg, P.L.; Morrison, N.; Smith, B.; Mungall, C.J.; Lewis, S.E.; Consortium, E. The environment ontology: Contextualising biological and biomedical entities. *J. Biomed. Semant.* **2013**, *4*, 43. [CrossRef]
47. Biomedical Data Translator, C. Toward A Universal Biomedical Data Translator. *Clin. Transl. Sci.* **2019**, *12*, 86–90. [CrossRef] [PubMed]
48. Shefchek, K.A.; Harris, N.L.; Gargano, M.; Matentzoglou, N.; Unni, D.; Brush, M.; Keith, D.; Conlin, T.; Vasilevsky, N.; Zhang, X.A.; et al. The Monarch Initiative in 2019: An integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic. Acids. Res.* **2020**, *48*, D704–D715. [CrossRef]
49. Callahan, T.J.; Tripodi, I.J.; Hunter, L.E.; Baumgartner, W.A. A framework for automated construction of heterogeneous large-scale biomedical knowledge graphs. *bioRxiv* **2020**. [CrossRef]
50. Ives, C.; Campia, I.; Wang, R.L.; Wittwehr, C.; Edwards, S. Creating a Structured AOP Knowledgebase via Ontology-Based Annotations. *Appl. In Vitro Toxicol.* **2017**, *3*, 298–311. [CrossRef]
51. Fantke, P.; von Goetz, N.; Schluter, U.; Bessems, J.; Connolly, A.; Dudzina, T.; Ahrens, A.; Bridges, J.; Coggins, M.A.; Conrad, A.; et al. Building a European exposure science strategy. *J. Expo. Sci. Environ. Epidemiol.* **2020**, *30*, 917–924. [CrossRef] [PubMed]
52. Wittwehr, C.; Aladjov, H.; Ankley, G.; Byrne, H.J.; de Knecht, J.; Heinzle, E.; Klambauer, G.; Landesmann, B.; Luijten, M.; MacKay, C.; et al. How Adverse Outcome Pathways Can Aid the Development and Use of Computational Prediction Models for Regulatory Toxicology. *Toxicol. Sci.* **2017**, *155*, 326–336. [CrossRef]
53. Balshaw, D.M.; Collman, G.W.; Gray, K.A.; Thompson, C.L. The Children’s Health Exposure Analysis Resource: Enabling research into the environmental influences on children’s health outcomes. *Curr. Opin. Pediatr.* **2017**, *29*, 385–389. [CrossRef] [PubMed]
54. FAIRsharing.org. CHEAR; Children’s Health Exposure Analysis Resource. Available online: <https://www.niehs.nih.gov/news/events/pastmtg/2016/chear/index.cfm> (accessed on 23 August 2021).
55. McCusker, J.P.; Rashid, S.M.; Liang, Z.; Liu, Y.; Chastain, K.; Pinheiro, P.; Stingone, J.A.; McGuinness, D.L. *Broad, Interdisciplinary Science In Tela: An Exposure and Child Health Ontology*; Association for Computing Machinery: Troy, NY, USA, 2017; pp. 349–357.
56. Heacock, M.L.; Amolegbe, S.M.; Skalla, L.A.; Trottier, B.A.; Carlin, D.J.; Henry, H.F.; Lopez, A.R.; Duncan, C.G.; Lawler, C.P.; Balshaw, D.M.; et al. Sharing SRP data to reduce environmentally associated disease and promote transdisciplinary research. *Rev. Environ. Health* **2020**, *35*, 111–122. [CrossRef]
57. Hamilton, C.M.; Strader, L.C.; Pratt, J.G.; Maiese, D.; Hendershot, T.; Kwok, R.K.; Hammond, J.A.; Huggins, W.; Jackman, D.; Pan, H.; et al. The PhenX Toolkit: Get the most from your measures. *Am. J. Epidemiol.* **2011**, *174*, 253–260. [CrossRef]
58. Miller, A.; Yeskey, K.; Garantziotis, S.; Arnesen, S.; Bennett, A.; O’Fallon, L.; Thompson, C.; Reinlib, L.; Masten, S.; Remington, J.; et al. Integrating Health Research into Disaster Response: The New NIH Disaster Research Response Program. *Int. J. Environ. Res. Public Health* **2016**, *13*, 676. [CrossRef]
59. Mattingly, C.J.; Boyles, R.; Lawler, C.P.; Haugen, A.C.; Dearry, A.; Haendel, M. Laying a Community-Based Foundation for Data-Driven Semantic Standards in Environmental Health Sciences. *Environ. Health. Perspect.* **2016**, *124*, 1136–1140. [CrossRef]
60. Thessen, A.E.; Grondin, C.J.; Kulkarni, R.D.; Brander, S.; Truong, L.; Vasilevsky, N.A.; Callahan, T.J.; Chan, L.E.; Westra, B.; Willis, M.; et al. Community Approaches for Integrating Environmental Exposures into Human Models of Disease. *Environ. Health. Perspect.* **2020**, *128*, 125002. [CrossRef]
61. Hardy, B.; Apic, G.; Carthew, P.; Clark, D.; Cook, D.; Dix, I.; Escher, S.; Hastings, J.; Heard, D.J.; Jeliakova, N.; et al. A toxicology ontology roadmap. *ALTEX* **2012**, *29*, 129–137. [CrossRef] [PubMed]
62. Costigan, S.L.; Werner, J.; Ouellet, J.D.; Hill, L.G.; Law, R.D. Expression profiling and gene ontology analysis in fathead minnow (*Pimephales promelas*) liver following exposure to pulp and paper mill effluents. *Aquat. Toxicol.* **2012**, *122–123*, 44–55. [CrossRef] [PubMed]
63. Currie, R.A.; Bombail, V.; Oliver, J.D.; Moore, D.J.; Lim, F.L.; Gwilliam, V.; Kimber, I.; Chipman, K.; Moggs, J.G.; Orphanides, G. Gene ontology mapping as an unbiased method for identifying molecular pathways and processes affected by toxicant exposure: Application to acute effects caused by the rodent non-genotoxic carcinogen diethylhexylphthalate. *Toxicol. Sci.* **2005**, *86*, 453–469. [CrossRef] [PubMed]
64. Egeghy, P.P.; Sheldon, L.S.; Isaacs, K.K.; Ozkaynak, H.; Goldsmith, M.R.; Wambaugh, J.F.; Judson, R.S.; Buckley, T.J. Computational Exposure Science: An Emerging Discipline to Support 21st-Century Risk Assessment. *Environ. Health Perspect.* **2016**, *124*, 697–702. [CrossRef] [PubMed]
65. Jarabek, A.M.; Hines, D.E. Mechanistic integration of exposure and effects: Advances to apply systems toxicology in support of regulatory decision-making. *Curr. Opin. Toxicol.* **2019**, *16*, 83–92. [CrossRef]
66. Meyer, D.E.; Bailin, S.C.; Vallero, D.; Egeghy, P.P.; Liu, S.V.; Cohen Hubal, E.A. Enhancing life cycle chemical exposure assessment through ontology modeling. *Sci. Total Environ.* **2020**, *712*, 136263. [CrossRef]
67. Kavlock, R.J.; Ankley, G.; Blancato, J.; Breen, M.; Conolly, R.; Dix, D.; Houck, K.; Hubal, E.; Judson, R.; Rabinowitz, J.; et al. Computational toxicology—a state of the science mini review. *Toxicol. Sci.* **2008**, *103*, 14–27. [CrossRef] [PubMed]
68. International Organization for Standardization (ISO). *Collaborative Business Relationship Management Systems—Requirements and Framework*; ISO Standard NO 44001:2017; ISO: Geneva, Switzerland, 2017.
69. Janke, K.K.; Seaba, H.H.; Welage, L.S.; Scott, S.A.; Rabi, S.M.; Kelley, K.A.; Mason, H.L. Building a multi-institutional community of practice to foster assessment. *Am. J. Pharm. Educ.* **2012**, *76*, 58. [CrossRef]
70. Pyrko, I.; Dorfler, V.; Eden, C. Thinking together: What makes Communities of Practice work? *Hum. Relat.* **2017**, *70*, 389–409. [CrossRef]

71. Valdes-Dapena, C. *Lessons from Mars: How One Global Company Cracked the Code on High Performance Collaboration and Teamwork*; John Hunt Publishing: Hampshire, UK, 2018.
72. Arnaud, E.; Laporte, M.A.; Kim, S.; Aubert, C.; Leonelli, S.; Miro, B.; Cooper, L.; Jaiswal, P.; Kruseman, G.; Shrestha, R.; et al. The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems. *Patterns* **2020**, *1*, 100105. [[CrossRef](#)]
73. Stevens, S.L.R.; Kuzak, M.; Martinez, C.; Moser, A.; Bleeker, P.; Galland, M. Building a local community of practice in scientific programming for life scientists. *PLoS Biol.* **2018**, *16*, e2005561. [[CrossRef](#)] [[PubMed](#)]
74. Berman, F.; Wilkinson, R.; Wood, J. Guest editorial: Building global infrastructure for data sharing and exchange through the Research Data Alliance. *D-Lib. Mag.* **2014**, *20*, 1–4. [[CrossRef](#)]
75. Berman, F. The Research Data Alliance—The First Five Years. 2019. Available online: <https://www.rd-alliance.org/research-data-alliance-%E2%80%93-first-five-years> (accessed on 23 August 2021).
76. Office of The Director National Institutes of Health. Final NIH Policy for Data Management and Sharing. 2020. Available online: <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html> (accessed on 6 July 2021).
77. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)] [[PubMed](#)]