

A semantic model leveraging pattern-based ontology terms to bridge environmental exposures and health outcomes

Lauren E. Chan¹, Nicole A. Vasilevsky², Anne Thessen^{1,2}, Nicolas Matentzoglou³, and William D. Duncan⁴, Christopher J. Mungall⁴, and Melissa A. Haendel^{1,2}

¹ Oregon State University, Corvallis, OR, 97331, USA

² University of Colorado Anschutz Medical Campus, Aurora, CO, 80045, USA

³ Semanticy Ltd, London, UK

⁴ Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

Abstract

Chemicals are a critical aspect of modern agriculture and residues of these chemicals are commonly consumed by humans. Consumption, inhalation, or topical exposure to agricultural chemicals can pose a risk for human health through a variety of mechanisms. Similarly, exposures to radiation, nutrient consumption, and many other environmental entities can impact health and thus a wide array of research has been pursued to better understand the mechanisms and impacts of environmental exposures. While extensive exposure research has been conducted and the data stored in environmental health databases, the ability to computationally assess these findings in the larger context of biomedical research to inform our knowledge for improved human health is still challenging. We developed an integrative exposure-disease model based on the Exposure Ontology (ExO) upper level ontology and established four Dead Simple OWL Design Patterns (DOSDP) for Mondo Disease Ontology. These patterns offer coordination of exposure event and exposure stimulus terms with disease terms, utilizing content from Open Biological Ontologies. Our model and pattern set can leverage logical axioms from integrated ontologies including the Food Ontology and the Environmental Conditions, Treatments, and Exposures Ontology (ECTO) for greater data and knowledge enrichment. Development of exposure event component terms and related logical axioms can facilitate the standardization needed for exposure modeling. Exposure content and our model can be utilized for the development of integrative knowledge graphs of exposure health data. Additionally, this model serves as a resource to aid the integration of common exposure data sources such as self-reported survey tools. Future work is needed to incorporate essential exposure data components into a comprehensive model, such as estimated or known exposure values, temporality of exposures, and biologically active exposure dosages that incur toxic effects.

Keywords 1

Ontology, knowledge graph, semantic model, environmental exposure, disease

1. Introduction

For decades, chemicals such as fertilizers, pesticides, herbicides, and insecticides have been used as an essential component to modern agriculture [1]. While the use of these agricultural chemicals is beneficial for promoting crop growth and controlling pests and diseases, they may also pose concerns to human health. Safety of various agricultural chemicals when ingested as residues on food and as inhaled or absorbed by humans applying the chemicals to crops continues to be a concern and research priority for toxicologists [2,3]. In addition to agricultural chemicals, humans experience hundreds if not thousands of environmental exposures daily (e.g., sun exposure, air pollution, and beauty products), each of which may pose health risks to the individual. In turn, environmental exposure characterization and documentation is essential to determining mechanisms of disease onset, understanding clinical sequelae, and recommending mitigating care strategies. Data from evaluations of model organisms, non-experimental exposures, and human exposures are maintained within environmental health databases. Unfortunately, limited computational standards are available for environmental health data [4,5]. This hinders the integration of environmental health findings to inform policy, health risk, and medical care [4,5]. Ontologies offer a unique opportunity to represent real life and experimental

exposures facing crops, model organisms, and humans. Additionally, ontologies can support integration and connection of heterogeneous research findings and modeled knowledge to facilitate inference and inform future research [6].

Previously, we developed the Environmental Conditions, Treatments, and Exposures Ontology (ECTO) to address these use cases. ECTO's terms represent a variety of stimuli and environmental conditions, including experimental and non-experimental exposures to humans, plants, and animals [7,8]. The Exposure Ontology (ExO) [9] is an upper ontology that models the relationship between 'exposure event', 'exposure stimulus', 'exposure receptor', and 'exposure outcome'. This foundation can be used to encode 'exposure event' terms that reference stimuli, mediums, and routes. ECTO classes utilize the ExO ontology and are at varying levels of granularity, allowing generalized querying or encoding of exposure to specific chemicals and other entities. In this paper, we expand prior work to establish an exposure-disease model that will enable inference regarding human exposures and their correlation with health concerns or disease states. Our model relies upon ontology term logical axioms and supports population of knowledge bases for mechanistic inquiry, including exposure events, genes, diseases, and pathways. We utilize exposure to agricultural chemicals as our primary example and describe four development patterns that are used to populate necessary classes in the model.

2. Semantic Modeling Goals

To facilitate encoding of environmental exposures and their impact on health, our adaptable exposure-disease model was developed to include exposures, food products, crop plants, mechanism of action, phenotypes, and disease. This model was the outcome of multiple workshops and community coordination, which included ExO, ECTO, and Mondo developers [5]. Within this proposed model, we have identified prospective ontologies from which to derive interoperable terms and relations including ECTO, Chemical Entities of Biological Interest (ChEBI) [10], Gene Ontology (GO) [11,12], National Center for Biotechnology Information Taxonomy (NCBI Taxon) [13,14], Food Ontology (FoodOn) [15,16], Human Phenotype Ontology (HPO) [17,18], Mondo Disease Ontology (Mondo) [19], and the OBO Relations Ontology (RO) [20,21].

Figure 1 depicts the three-part progression of our model including the upper ExO ontology (Figure 1A), our adaptable exposure-disease model (Figure 1B), and an application of the model using an instance level example of chlorpyrifos residue ingestion on an apple. Chlorpyrifos, an organophosphorus insecticide, is a common agricultural chemical used for production of produce and other crops within the US and beyond [22]. Chlorpyrifos has faced criticism previously for its potential impact on the human nervous system, and particularly for the risks it may pose to children's neurological development [23]. Based on reported literature of chlorpyrifos mechanisms, our model can be used to identify exposure sources, mechanisms, and associations with presenting disease and phenotypes.

As seen in Figure 1, ExO describes an 'exposure stimulus' as 'an agent, stimulus, activity, or event that causes stress or tension on an organism and interacts with an exposure receptor during an exposure event'. A 'exposure receptor' is defined as 'an entity (e.g., a human, human population, or a human organ) that interacts with an exposure stimulus during an exposure event'. An 'exposure outcome' is defined as 'entity that results from the interaction between an exposure receptor and an exposure stimulus during an exposure event' and represents the negative or positive outcomes of having been exposed to the stimulus. It is important to recognize that it is the axioms encoded within the model that connect exposure information to a variety of knowledge that then allows the potential inference of a candidate stimulus or a predicted outcome in response to one.

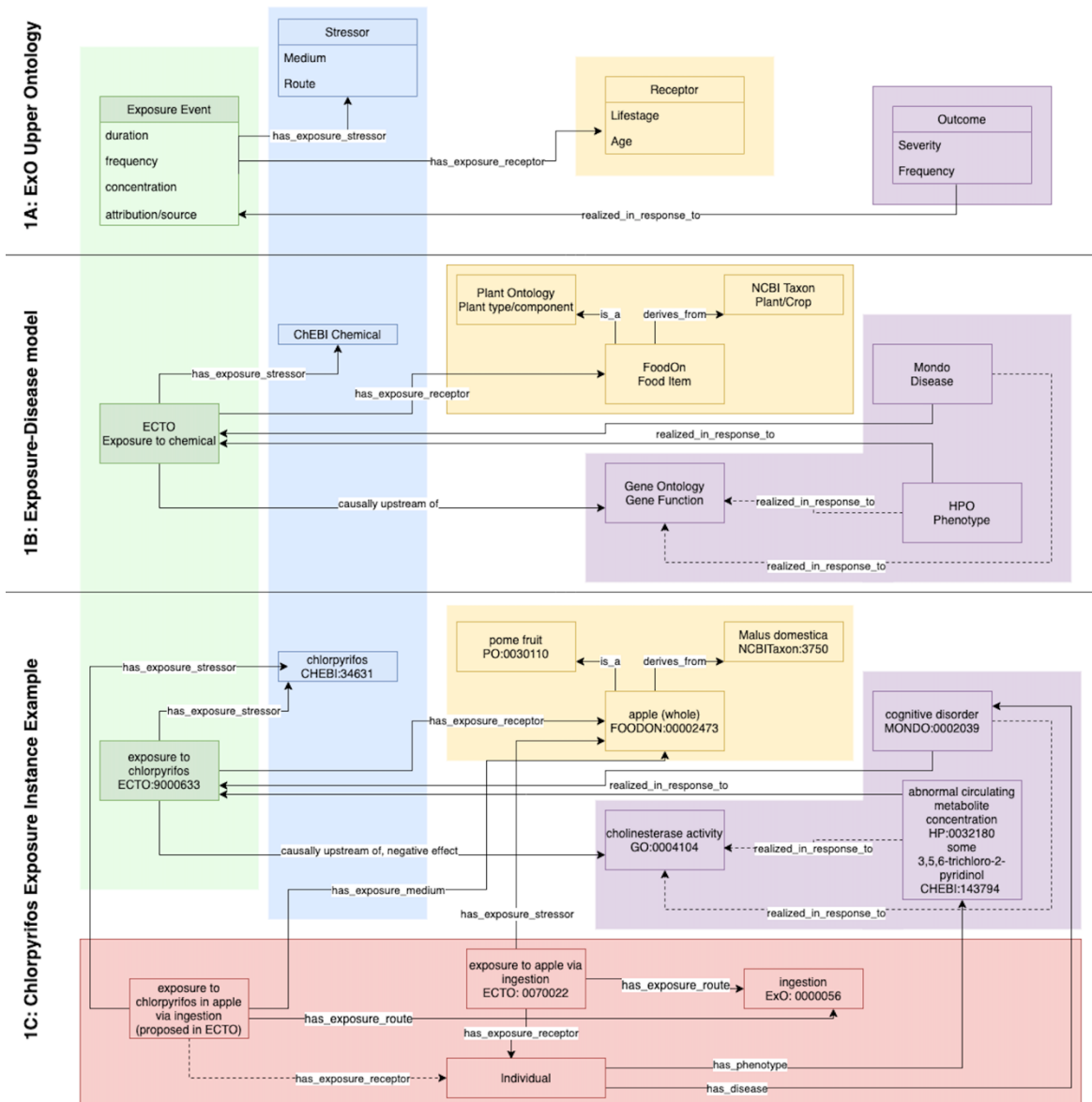


Figure 1. Defining and populating the exposure semantic framework. All figure panels contain consistent model variables: the exposure event in green, the entity stimulating the exposure in blue, the organism or entity being exposed in yellow, and the resulting outcomes in purple. Within Figure 1C, an instance level exposure is included in a red panel to display the integration of data.

Figure 1A. ExO upper ontology: ExO includes the central ‘exposure event’ as well as associated ‘exposure stimulus’, ‘exposure receptor’, and ‘exposure outcome’ elements. Granular exposures (e.g., exposure to chlorpyrifos) are modeled in ECTO, but leverage the upper ExO ontology. In ECTO, each element can be annotated with associated metadata.

Figure 1B. Exposure-disease model: Utilizing ECTO ‘exposure event’ classes, our model can include a variety of exposure stimuli, mediums, routes, and outcomes due to the inherent ExO upper level schema. Solid edges include direct relationships which can be modeled as a part of an exposure event, with dashed lines representing inferred relationships that are derived from the known direct relationships. This exposure-disease model offers a precomposed template for which to map documented relationships from the literature to support computational assessment of environmental health research findings.

Figure 1C. Chlorpyrifos exposure instance example: The adaptable exposure-disease model can be used to coordinate instance level data with ontology knowledge, resulting in a translatable schema for environmental exposures. This example provides a multilayer exposure process for an individual who ingests an apple after it is exposed to chlorpyrifos, coordinated with the known phenotype and disease

presentation in the individual. Documented relationships are seen with solid lines and inferred relationships are seen in dashed lines.

By documenting not only food items that are the mediums for the exposure to chlorpyrifos, but also including the mechanism of action, known phenotypes, and disease states, our example schema of chlorpyrifos exposure offers access points in which further information can be inferred. For example, if another chemical served as an acetylcholinesterase inhibitor within humans, by inclusion of that chemical exposure and known regulatory activity, one could infer that the second chemical exposure may also be related to cognitive disorders, or that the chemicals composition may be similar to chlorpyrifos.

3. Exposure model axioms

To support the exposure-disease model described above, we utilize ontology relationship axioms and structures. Some examples include logical axioms in ECTO and FoodOn.

ECTO terms are developed as pre-composed classes. Exposure terms are inherently coordinated with the relevant ontology term for the chemical, environmental stimulus, or condition the ECTO term label refers to. Each pre-composed ECTO class includes a reference to another ontology term. For example, with the ECTO term 'exposure to fertilizer' (ECTO:9000091) the equivalence axiom for this term includes a reference to the ChEBI term 'fertilizer' CHEBI:33287. This logic provides the 'has exposure stimulus' relationship and aligns the exposure term with the detailed content of the referenced ChEBI term.

Class:

'exposure to fertilizer'

Equivalence axiom:

'exposure event'

and 'has exposure stimulus' some fertilizer"

Existing logic from FoodOn is also included in our model. Within FoodOn, the source ontology for food terminology, foods produced directly from a crop include a logical axiom. For example, the FoodOn term 'orange (whole)' (FOODON:03315106) has the logical axiom shown below that references the plant term 'hesperidium fruit' (PO:0030109) and the taxon 'Citrus sinensis' (NCBITaxon:2711).

Class:

orange (whole)

Logical axiom:

'hesperidium fruit and *derives from* some Citrus sinensis'

While these relationships and components of our model are already represented in ontology structures, the critical relationship between exposure and disease outcomes was not yet well defined.

4. Modeling Exposures as Disease Influencers

To facilitate the integration and modeling of environmental exposures and human disease, we have developed and implemented four patterns for disease terms with a known exposure basis for the Mondo Disease Ontology. Creation of these patterns established logical axioms within Mondo disease terms that coordinate with environmental exposure ECTO terms. This Mondo-ECTO term relationship can then be directly implemented into our exposure-disease model. Within Mondo, as well as other ontologies, Dead Simple OWL Design Patterns (DOSDP) are frequently used to develop new terms with logical axioms in a consistent and easily maintained manner [24]. Mondo is a significant resource for mapping disease knowledge across many disease information sources (e.g., MESH, ICD, and OMIM). We chose Mondo as the target of our modeling as it was relatively easy to extend the existing logic as well as supporting alignment of many disparate resources.

The disease 'radiodermatitis' (MONDO:0043771) conforms to the Mondo 'realized_in_response_to_environmental_exposure' design pattern (https://github.com/monarch-initiative/mondo/blob/master/src/patterns/dosdp-patterns/realized_in_response_to_environmental_exposure.yaml). This pattern uses the relation 'realized in response to' to link diseases to the exposures (represented by ECTO classes) causing the disease. The logical axiom utilized for this pattern is:

```
'%s and ('realized in response to' some %s)'
```

Vars:

- Disease
- Exposure

Within this logical axiom template are the variable (*vars*) fields, represented by '%s'. For each *vars*, a variable term is required to complete the axiom statement. In this instance, the *vars* are 'Disease' and 'Exposure'. These variable terms will be identified from Mondo and ECTO and will be used to fill in the first and second fields respectively. For example, the logical axiom for 'radiodermatitis' is represented as:

```
radiodermatitis
and ('realized in response to' some 'exposure to electromagnetic radiation')
```

For the variety of diseases that may be caused by or initiated via an environmental exposure or external entity, we have created multiple DOSDPs for Mondo that support general and specific exposure-based disease terms. Their content and applications are described in Table 1. At this time, over 390 terms have been implemented using these patterns, with 46 terms including logical axioms referencing 17 unique ECTO exposure terms.

Table 1: Exposure Related Mondo Patterns. All exposure patterns can be found on the Mondo GitHub page (<https://github.com/monarch-initiative/mondo/tree/master/src/patterns/dosdp-patterns>).

Pattern Name	Included	Not included	Logical axioms	Example disease
Poisoning.yaml	Diseases caused by exposure to a chemical or mixture that meets the threshold to cause poisoning or intoxication.	Diseases that include exposure to a chemical or mixture but that do not reach the threshold of poisoning or intoxication.	""poisoning" and "realized in response to stimulus" some %s' Vars: stimulus	colchicine poisoning (MONDO:0017859)
Substance_abuse.yaml	Behavioral diseases that include the abuse of a chemical substance	Diseases that do not include a behavioral substance abuse component	""substance abuse" and "realized in response to stimulus" some %s' Vars: stimulus	amphetamine abuse (MONDO:0003969)
Realized_in_response_to_environmental_exposure.yaml	Disease states that are directly realized due to exposure to an environmental	Diseases that are not a direct result of an environmental exposure.	'%s and ("realized in response to" some %s)'	alcoholic cardiomyopathy (MONDO:0006643)

	condition, chemical, or mixture. Includes reference terms from Mondo and ECTO.	Diseases caused by an infectious agent	Vars: disease, exposure	
Infectious_disease_by_agent.yaml	Diseases caused directly by an infectious agent.	Diseases not caused by exposure to an infectious agent (organism, virus, viroid etc.)	"infectious disease" and "disease has infectious agent" some %s'	Toxoplasmosis (MONDO:0005989)
			Vars: agent	

5. Future Directions

Building models for exposure risk and disease causality has been challenging due to the heterogeneity and lack of interoperability across agricultural, toxicological, and clinical data [25,26]. The model outlined in here is a preliminary foundation for how exposure influenced diseases can be described in a computable fashion. The four patterns presented here can be used to establish exposure based disease classes for Mondo Disease Ontology, and similar approaches could likely be translated into other ontologies.

Our modeling structure can be used for chemical, nutrient, and other environmental exposures and their impact on phenotypes, disease, and gene function. This modular approach supports adaptation of exposure source and types while also allowing for multiple different exposures to be integrated for a comprehensive mapping of exposures to outcomes. In addition to the proposed model, work to include variables for comprehensive exposure to health modeling such as estimated or known exposure values (e.g. residual agricultural chemicals consumed in average diet), estimated or known temporality of exposures, and biologically active exposure dosages for toxic effects are needed. We plan to utilize this semantic framework for integrating a wide range of dietary and other exposures for predictive analytics, inference of causality, and to inform mitigation of exposures. The goal is to be able to integrate clinical data and biomarkers of exposure with data collected via self-reported surveys, which are commonly used for dietary data collection and estimation tools for personal environmental exposures. Additionally, harmonization of this model with other existing resources for describing adverse outcome pathways and ecotoxicology such those presented by Myklebust et al. [27] would offer substantial data integration for inference development.

Using this semantic framework, we will be able to populate a knowledge graph that would leverage content found in numerous biomedical ontologies alongside instance level data from surveys, clinical data, and more. Future efforts will be focused on improving the accuracy with which exposure events can be documented to include temporality, dosage, and resulting environmental and health outcomes. In turn, these efforts are intended to support methods for risk estimations of disease and phenotype outcomes given predicted or known environmental exposures.

The ECTO repository: <https://github.com/EnvironmentOntology/environmental-exposure-ontology>

The Exposure-Disease wiki: <https://github.com/EnvironmentOntology/environmental-exposure-ontology/wiki/Exposure-disease-model>

6. References

- [1] Rhoades WC. "The History and Use of Agricultural Chemicals." *Fla Entomol* 46.4 (1963): 275–277.
- [2] Calvert GM, Alarcon WA, Chelminski A, Crowley MS, Barrett R, Correa A, et al. Case report: three farmworkers who gave birth to infants with birth defects closely grouped in time and place—Florida and North Carolina, 2004–2005. *Environ Health Perspect.* 115 (2007): 787–791.
- [3] de Cock J, Heederik D, Hoek F, Boleij J, Kromhout H. Urinary excretion of tetrahydrophtalimide in fruit growers with dermal exposure to captan. *Am J Ind Med.* 28 (1995): 245–256.
- [4] Boyles RR, Thessen AE, Waldrop A. Ontology-based data integration for advancing toxicological knowledge. *Current Opinion in Toxicology.* 16 (2019): 67–74. <https://doi.org/10.1016/j.cotox.2019.05.005>
- [5] Thessen AE, Grondin CJ, Kulkarni RD, Brander S, Truong L, Vasilevsky NA, et al. Community Approaches for Integrating Environmental Exposures into Human Models of Disease. *Environ Health Perspect.* 128 (2020): 125002.
- [6] Hoehndorf R, Dumontier M, Gkoutos GV. Evaluation of research in biomedical ontologies. *Brief Bioinform.* 14 (2013): 696–712.
- [7] Environmental Conditions, Treatments, and Exposures Ontology (ECTO), 2020. URL: <http://www.obofoundry.org/ontology/ecto.html>
- [8] GitHub Repository - Environmental Conditions, Treatments, and Exposures Ontology (ECTO), 2020. URL: <https://github.com/EnvironmentOntology/environmental-exposure-ontology>
- [9] Mattingly CJ, McKone TE, Callahan MA, Blake JA, Hubal EAC. Providing the missing link: the exposure science ontology ExO. *Environ Sci Technol.* 46 (2012): 3046–3053.
- [10] Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res.* 44 (2016): D1214–D1219.
- [11] Gene Ontology Consortium. The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Res.* 49 (2021): D325–D334.
- [12] Gene Ontology Resource, 2020. URL: <http://geneontology.org/>
- [13] Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res.* 40 (2012): D136–43.
- [14] NCBI – Taxonomy, 2021. URL: <https://www.ncbi.nlm.nih.gov/taxonomy>
- [15] FoodOn: A farm to fork ontology, 2020. URL: <https://foodon.org/>
- [16] Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food.* 2 (2018): 1–10.
- [17] Human Phenotype Ontology, 2020. URL: <https://hpo.jax.org/app/>
- [18] Köhler S, Gargano M, Matentzoglou N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. *Nucleic Acids Res.* 49 (2021): D1207–D1217.

- [19] Mondo Disease Ontology, 2021. URL: <http://mondo.monarchinitiative.org>
- [20] OBO Relation Ontology, 2021. URL: <https://oborel.github.io/>
- [21] Guardia GDA, Vêncio RZN, de Farias CRG. A UML profile for the OBO relation ontology. *BMC Genomics*. 13 (2012): Suppl 5: S3.
- [22] Chlorpyrifos Facts, 2021. URL: <https://www.panna.org/resources/chlorpyrifos-facts>
- [23] Burke RD, Todd SW, Lumsden E, Mullins RJ, Mamczarz J, Fawcett WP, et al. Developmental neurotoxicity of the organophosphorus insecticide chlorpyrifos: from clinical findings to preclinical models and potential mechanisms. *J Neurochem*. 142 (2017) Suppl 2: 162–177.
- [24] Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. Dead simple OWL design patterns. *J Biomed Semantics*. 8 (2017): 18.
- [25] Chan L, Vasilevsky N, Thessen A, McMurry J, Haendel M. The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research. *Database*. (2021). doi:10.1093/database/baab003
- [26] Hartung T, FitzGerald RE, Jennings P, Mirams GR, Peitsch MC, Rostami-Hodjegan A, et al. Systems Toxicology: Real World Applications and Opportunities. *Chem Res Toxicol*. 30 (2017): 870–882.
- [27] Myklebust EB, Jimenez-Ruiz E, Chen J, Wolf R. Tera: the toxicological effect and risk assessment knowledge graph. *arXiv preprint arXiv*. (2019). URL: <https://arxiv.org/abs/1908.10128>