

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

Lauren E. Chan¹, Nicole A. Vasilevsky², Anne Thessen¹, 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

Agricultural chemical use is a critical aspect of modern agriculture and residues of these chemicals are commonly consumed by humans. These chemicals and other environmental exposures pose risk for human health through a variety of mechanisms, prompting toxicological and public health research to better understand their impacts. While extensive exposure research has been conducted and the data stored in toxicological 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 quite challenging.

We developed an integrative model of exposure events that utilizes content from the Open Biological Ontologies to build a semantic framework of environmental exposures and health outcomes. Logical axioms included within the Mondo Disease Ontology; the Food Ontology; and the Environmental Conditions, Treatments, and Exposures Ontology (ECTO) all further enrich the proposed model. Further 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.

Further 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 disease, they may also pose concerns to human health. Safety of various agricultural chemicals when ingested as residuals 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, each of which may pose health risks to the individual.

Human diseases can be defined by a variety of factors including age of onset, heritability, genetic endowment, and environmental factors. However, the delineation of how environmental factors drive the course of human disease continues to fuel debates within the literature[4,5]. In turn, environmental exposure characterization and documentation is essential to determining mechanisms of disease onset, understanding clinical sequelae, and recommending mitigating care strategies.

A significant number of investigations using model organisms or evaluating non-experimental exposures for humans have been conducted and toxicological databases exist to house data regarding exposure-to-disease relationships. However, the inability to integrate these findings to inform policy, health risk, and care poses great concern due to the limited computational standards currently available for toxicological data[6,7]. 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[8]. 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[9,10]. ECTO utilizes the upper model provided by the Exposure Ontology (ExO)[11], which is a data model that focuses on the components of an exposure event.

Figure 1A illustrates the ExO model, which models the relationship between 'exposure event', 'stressor', 'receptor', and outcomes. ECTO uses the ExO model as the foundation to encode granular 'exposure event' terms that reference stressors, mediums, and routes. By modeling specific exposure events using classes from other ontologies, a semantic framework can be created that associates chemicals, environments, genes, diseases, pathways, and beyond.

2. Semantic Modeling Goals:

To facilitate modeling of agricultural chemical exposures and their impact on health, a semantic model was developed that includes exposures, food products, crop plants, mechanism of action, phenotypes, and disease (**Figure 1B**). This semantic framework was the outcome of multiple workshops and community coordination, which included ExO, ECTO, and Mondo developers[7]. 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)[12], Gene Ontology (GO)[13,14], National Center for Biotechnology Information Taxonomy (NCBI Taxon)[15,16], Food Ontology (FoodOn)[17,18], Human Phenotype Ontology (HPO)[19,20], Mondo Disease Ontology (Mondo)[21], and the OBO Relations Ontology (RO)[22,23].

The model in **Figure 1B** includes a food-based exposure, but this model can be altered to include a variety of exposure stressors, mediums, and routes due to the inherent inclusion of ExO upper-level modeling. Additional relationships can also be modified, for example to include classes such as known medical therapies or to modify outcomes for accuracy (e.g. including Zebrafish Phenotype Ontology terms for models of zebrafish exposure). Using this model approach, heterogeneous data related to exposure events can be coordinated and integrated using ontology terminology into the appropriate node.

From the foundations of ExO and ECTO modeling, **Figure 1C** displays a specific example of an 'exposure event' and the resulting model. We selected a practical use case of common agricultural chemical usage on apple trees. Chlorpyrifos, an organophosphorus insecticide, is a common agricultural chemical used for production of produce and other crops within the US and beyond[24]. 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[25]. Using our proposed model, one can evaluate the ontology terms and relationships that can be used to describe the risks posed by chlorpyrifos. 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, this model of chlorpyrifos exposure offers access points in which further information can be annotated. 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.

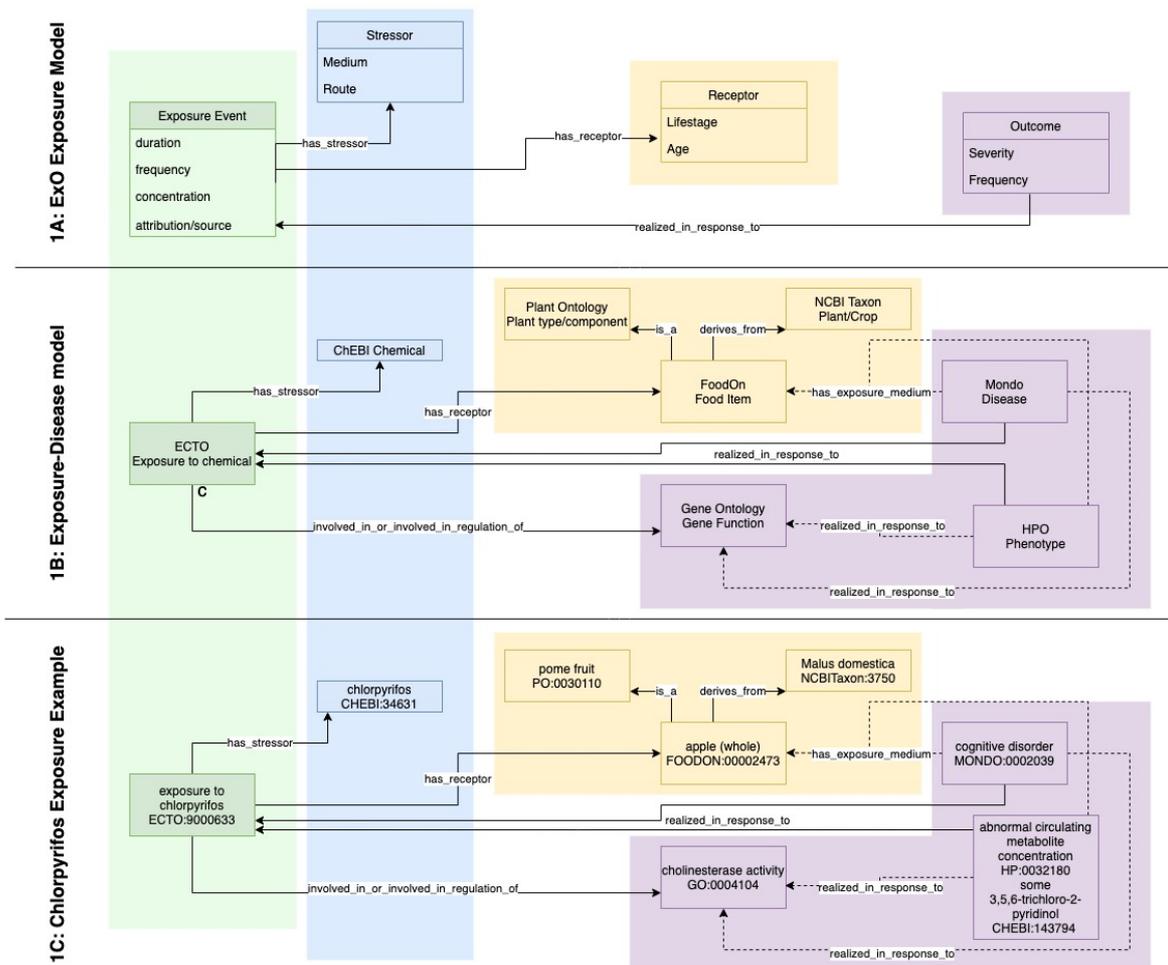


Figure 1. Defining and populating the exposure semantic framework. All figure panels contain consistent model variables: the exposure event in green, the entity stimulating the exposure in blue, the organism or entity being exposed in yellow, and the resulting outcomes in purple.

Figure 1A. ExO upper level modeling: The Exposure Ontology (ExO) upper level model[11] includes the central ‘exposure event’ as well as associated ‘stressor’, ‘receptor’, and ‘outcome’ elements. Each element can be annotated with associated metadata.

Figure 1B. ECTO specific exposures model: Utilizing ECTO ‘exposure event’ classes, a comprehensive semantic model can be developed to include a variety of key factors from exposure components to biological function to phenotype and disease outcomes. 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.

Figure 1C. Chlorpyrifos exposure example: The ‘exposure event’ of ‘exposure to chlorpyrifos’ can be coordinated with its component ‘stressor’ and ‘receptor’, while also being associated with resulting health outcomes. Inferred relationships between health outcomes and exposure variables and pathways can also be seen in dashed lines.

3. Existing Infrastructure to Support Semantic Modeling

To support the semantic modeling goals proposed above, we aimed to integrate new terms, patterns, and relationship axioms within existing ontology structures using the OBO orthogonality principle.

ECTO terms are developed as pre-composed classes with exposure terms inherently coordinated with the relevant ontology term for the chemical, environmental stimulus, or condition the ECTO term name refers to. In turn, each pre-composed ECTO class includes a reference to another ontology term. For example, with the ECTO term ‘exposure to fertilizer’ (ECTO:9000091) the logic

behind this term includes a reference to the ChEBI term ‘fertilizer’ CHEBI:33287. This logic means that all ECTO classes inherently include the ‘has_exposure_stimulus’ relationship proposed within this model.

Class:

```
`exposure to fertilizer`
```

Equivalence axiom:

```
'exposure event'  
and 'has exposure stimulus' some fertilizer"
```

Other existing logic within ontologies includes the relationship exhibited between food terms and the source that produces the food product. Within FoodOn, the source ontology for food terminology, foods produced directly from a crop include a logical axiom to preserve this relationship. 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 some aspects of this model are already represented in ontology structures, the critical relationship between exposure and health outcome required enhanced modeling.

4. Modeling Exposures as Disease Influencers

The integration and modeling of environmental exposures and human disease is of great interest for developing inferences relating toxicology and public health outcomes - inferences that could influence diagnostics, risk assessment, and policy. Towards these ends, we implemented patterns for diseases with a known exposure basis within the Mondo disease ontology. 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[26]. Mondo is a significant resource for mapping disease knowledge across many disease sources and leverages DOSDPs for its development. We therefore 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_environmental_exposure’ 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’ being 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, multiple DOSDPs have been developed for use within Mondo. Their content and applications are described in Table 1. At this time, over 390 terms have been implemented using these patterns.

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 condition, chemical, or mixture. Include s reference terms from Mondo and ECTO.	Diseases that are not a direct result of an environmental exposure. Diseases caused by an infectious agent	'%s and ("realized in response to" some %s)' Vars: disease, exposure	alcoholic cardiomyopathy (MONDO:0006643)
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' Vars: agent	Toxoplasmosis (MONDO:0005989)

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[27,28]. We are modelling structures that can be used for chemical, nutrient, and other environmental condition exposures and their impact on phenotypes, disease, and gene function. Our 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, further variables can also be included 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.

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 incorporate 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.

Using this semantic framework, we will be able to populate a knowledge graph that will leverage knowledge found in numerous biomedical ontologies alongside instance level data from surveys, clinical data, and omics data. 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.

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 tetrahydrophthalimide in fruit growers with dermal exposure to captan. *Am J Ind Med*. 28 (1995): 245–256.
- [4] Dardiotis E, Xiomerisiou G, Hadjichristodoulou C, Tsatsakis AM, Wilks MF, Hadjigeorgiou GM. The interplay between environmental and genetic factors in Parkinson's disease susceptibility: the evidence for pesticides. *Toxicology*. 307 (2013): 17–23.
- [5] Antonelli A, Ferrari SM, Ragusa F, Elia G, Paparo SR, Ruffilli I, et al. Graves' disease: Epidemiology, genetic and environmental risk factors and viruses. *Best Pract Res Clin Endocrinol Metab*. 34 (2020): 101387.
- [6] 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>
- [7] 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.
- [8] Hoehndorf R, Dumontier M, Gkoutos GV. Evaluation of research in biomedical ontologies. *Brief Bioinform*. 14 (2013): 696–712.

- [9] Environmental Conditions, Treatments, and Exposures Ontology (ECTO), 2020. URL: <http://www.obofoundry.org/ontology/ecto.html>
- [10] GitHub Repository - Environmental Conditions, Treatments, and Exposures Ontology (ECTO), 2020. URL: <https://github.com/EnvironmentOntology/environmental-exposure-ontology>
- [11] 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.
- [12] 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.
- [13] Gene Ontology Consortium. The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Res.* 49 (2021): D325–D334.
- [14] Gene Ontology Resource, 2020. URL: <http://geneontology.org/>
- [15] Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res.* 40 (2012): D136–43.
- [16] NCBI – Taxonomy, 2021. URL: <https://www.ncbi.nlm.nih.gov/taxonomy>
- [17] FoodOn: A farm to fork ontology, 2020. URL: <https://foodon.org/>
- [18] 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.
- [19] Human Phenotype Ontology, 2020. URL: <https://hpo.jax.org/app/>
- [20] 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.
- [21] Mondo Disease Ontology, 2021. URL: <http://mondo.monarchinitiative.org>
- [22] OBO Relation Ontology, 2021. URL: <https://oborel.github.io/>
- [23] Guardia GDA, Vêncio RZN, de Farias CRG. A UML profile for the OBO relation ontology. *BMC Genomics.* 13 (2012): Suppl 5: S3.
- [24] Chlorpyrifos Facts, 2021. URL: <https://www.panna.org/resources/chlorpyrifos-facts>
- [25] 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.
- [26] Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. Dead simple OWL design patterns. *J Biomed Semantics.* 8 (2017): 18.
- [27] 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
- [28] 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.