The Environmental Conditions, Treatments, and Exposures Ontology (ECTO): Connecting Toxicology and Exposure to Human Health and Beyond

Lauren E. Chan[1], Anne E. Thessen[1,2], William D. Duncan[3], Nico Matentzoglu[4], Charles Schmitt[5], Cynthia Grondin[6], Nicole Vasilevsky[2], Julie McMurry[2], Peter Robinson[7], Christopher J. Mungall[3], and Melissa A. Haendel[1,2]

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

[2] University of Colorado Anschutz Medical Campus, Aurora, CO, 80054, USA

[3] Lawrence Berkeley National Laboratory, Berkeley, CA, 94720, USA

[4] Semanticly Ltd, London, UK

[5] National Institute of Environmental Health Sciences, Durham, NC 27709, USA

[6] North Carolina State University, Raleigh, NC, 27965, USA

[7] The Jackson Laboratory for Genomic Medicine, Farmington, CT 06032, USA

Background: In this manuscript we present the Environmental Conditions, Treatments, and Exposures Ontology (ECTO), a species-agnostic ontology focused on describing experimental and natural exposure processes such as dietary, workplace, or research contexts. ECTO is intended for use in harmonizing environmental health data resources to support cross-study integration and inference for mechanism discovery.

Evaluating the impact of environmental exposures on organism health is a key goal of modern biomedicine and critically important in an age of greater pollution and chemicals in our environment. Environmental health utilizes many different research methods and generates many types of data. However, to date, no comprehensive database represents the full spectrum of environmental health data. Due to a lack of interoperability between databases, tools for integrating these resources are needed.

Methods and Findings: ECTO is a terminology designed for describing organismal exposures such as toxicological research, environmental variables, dietary features, and patient-reported information from surveys. ECTO utilizes the base model established within the Exposure Ontology (ExO). ECTO is developed using a combination of manual curation and Dead Simple OWL Design Patterns (DOSDP), and contains over 2700 environmental exposure terms, with axioms to existing ontologies such as for chemicals or environments. ECTO is an Open Biological and Biomedical Ontology (OBO) Foundry ontology that is designed for interoperability, reuse, and axiomatization with other ontologies. ECTO terms have been utilized for logical axiomatization within the Mondo Disease Ontology to represent diseases caused or influenced by environmental factors, as well as for survey encoding for the Personalized Environment and Genes Study (PEGS).

Conclusions: We have constructed ECTO to meet Open Biological and Biomedical Ontology (OBO) Foundry principles to increase translation opportunities between environmental health and other areas of biology. ECTO has a growing community of contributors consisting of toxicologists, public health epidemiologists, and health care providers to provide the necessary expertise for areas that have been identified previously as gaps.

Kew words: Biomedical Ontology, environmental exposures, environmental health

Introduction

Environmental health is a branch of public health that encompasses the study of the inter-relationship between organisms (typically humans) and their environment and how that may impact health. Environmental health often relates to hazardous toxin exposures, but it can also encompass exposure to chemicals and environments that we may not always consider hazardous, such as vitamins, climate, and social stressors. Identification of stimuli in environmental substances is critical for disease prevention and management of adverse health outcomes, as well as to identify and evaluate mechanisms of action to develop clinical treatments. Environmental health has evolved alongside other fields including genomics, phenomics, nutrition, epidemiology, and crop sciences. Each of these interconnected disciplines are essential to understanding the full picture of how environments can prevent, cause, or ameliorate disease.

Toxicology is an important sub-field of environmental health. Existing toxicology-focused databases and data repositories including Chemical Effects in Biological Systems (CEBS)[1], Comparative Toxicogenomics Database (CTD)[2], National Health and Nutrition Examination Survey (NHANES), and Aggregated Computational Toxicology Resource (ACToR) databases [3] currently house a mix of structured, semi-structured, and unstructured information regarding environmental exposure impacts on a variety of species [4]. These resources offer unique features, including being repositories for raw data from toxicology studies, aggregating and inferring findings from the literature, or housing survey questions and results. For some of these data resources (e.g., NHANES surveys) Common Data Elements (CDEs) are utilized, which include standardized survey questions and responses intended to unify data from multiple resources using the same CDEs. While the attempts to align a variety of related but heterogeneous data resources using CDEs is meaningful, unfortunately, CDEs are often lacking in their computational encoding, making them challenging to use for making data interoperable[5].

Resources such as the Human Health Exposure Analysis Resource (HHEAR)[6], the Unified Medical Language System (UMLS)[7] and the Adverse Outcomes Pathway Knowledgebase (AOP)[8] have opted to include ontology terminology in their modeling. However, content regarding environmental exposures is still needed within AOP and UMLS, and analytical opportunities and widespread uptake are still limited using HHEAR.

Even the most comprehensive resources are still limited by their lack of standardized language, computational structure, or cross-study and cross-discipline data comparison capabilities. In efforts to support data integration within and beyond environmental health, a common standard for describing and coordinating these data is necessary.

Currently, ontologies related to environmental exposure are limited, with most ontologies focused on the description of environments, chemicals, or species specific exposure conditions. However, no species-agnostic ontology focused on the exposure process including the stimuli and media currently exists, limiting the opportunity to harmonize existing and future data regarding environmental exposures and related health outcomes.

A demand for integration of environmental health into interoperable data resources using ontologies is documented [9–11], with a variety of toxicologists, public health epidemiologists, and health care clinicians seeking established standards and resources. For this reason, we have created the Environmental Conditions, Treatments, and Exposures Ontology (ECTO) to satisfy the gaps seen within current ontology resources and to provide a translation tool for toxicology and biological data integration. ECTO's exposure event structure is a species-agnostic approach that can be used to align existing environmental health databases and resources. For example, CTD offers highly relevant data regarding exposure stimuli including potential biological ramifications of exposure and references to literature. While it is meaningful data, the data is structured in a format that does not create context for the exposure itself (e.g. multiple rows of data may relate to an exposure to chlorpyrifos and list some reported outcomes, but the outcomes are not coordinated with each other to provide an exposure phenotype profile or to compare to any known diseases and their common phenotypes). By utilizing the computable structure of ECTO, resources like CTD could be directly aligned with other ECTO compatible resources and could be leveraged for inference regarding exposures and human phenotype or disease outcomes across data sources. Additionally ECTO follows Findable, Accessible, Interoperable, and Reusable (FAIR) principles[12].

The Environmental Conditions, Treatments, and Exposures Ontology (ECTO)

ECTO contains compositional classes which utilize content from existing biomedical ontologies (such as the Environment Ontology (ENVO)[13] and the Chemical Entities of Biological Interest (ChEBI)[14]) to create exposure classes. Examples of exposures represented in ECTO include: experimental treatments and interventions used in research (e.g. toxicological investigations), exposures experienced by humans or other organisms in daily life, natural and artificial stimuli experienced by organisms, and environmental conditions or ecosystems experienced by a single organism or population of organisms. By maintaining a general scope of terms, ECTO can provide a wide range of content that can be applied in research settings ranging from wet lab to clinical care. Included in ECTO's exposure content are an organism's internal and external exposures, mixtures of known and inferred exposures, and indication of the route and medium of exposure when available.

For example, acute and chronic dietary exposure to agricultural chemicals may pose a risk to human health, particularly for children and developing fetuses [15]. Chlorpyrifos was banned for household use in the US in 2000, but up until recently it has continued to be used in American agriculture, regardless of potential detrimental health effects[16]. **Figure 1** showcases ECTO's unification capacity, coordinating existing ontologies or data sources as part of an 'exposure to chlorpyrifos', which even in low doses may have resulting phenotypes such as a runny nose, tears, or drooling[17]. In this example, a person presented to a health care provider complaining of an ongoing runny nose, tears, and drooling in recent history with no known illness, and they report eating apples daily. Our existing knowledge from our exposure event structure includes 1) chlorpyrifos is sprayed on some apples, and 2) runny nose, tears, and drooling are all chlorpyrifos associated phenotypes. Having structured knowledge in a format that supports queries, we can more quickly identify the exposure concern and provide an intervention. This structure would also allow for alignment of heterogeneous databases and data sources such as electronic health record and survey based resources.

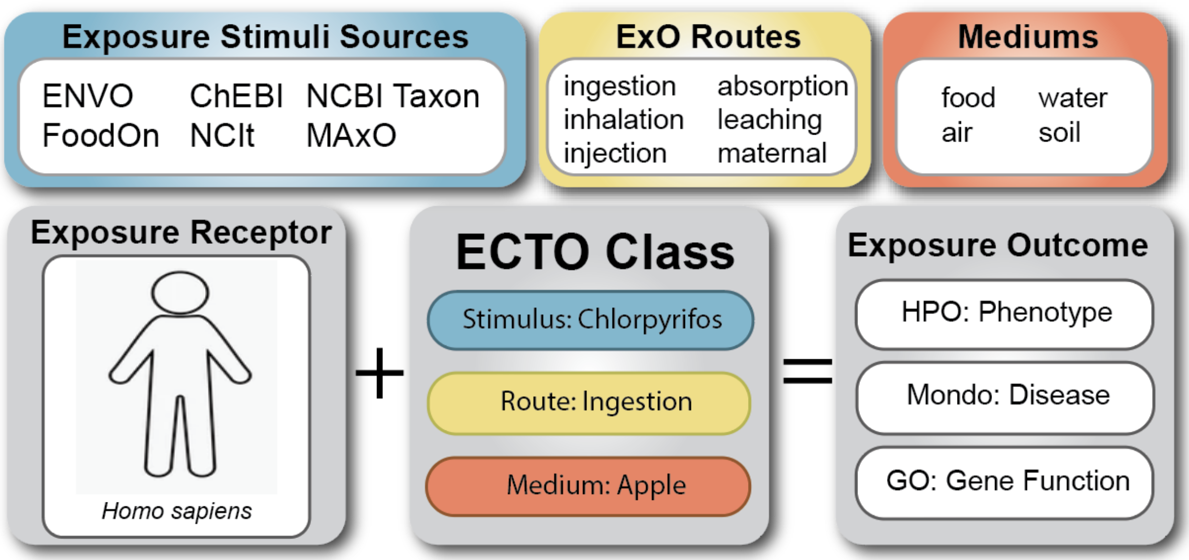

Figure 1: ECTO unifies exposure attributes. Existing ontologies contain terms describing common exposure stimuli, exposure routes, and potential exposure media, however unifying terms to describe the exposure process were not yet represented. Utilizing the schema of the Exposure Ontology (ExO), ECTO classes can coordinate the stimulus, route, and exposure components into a single process term for consistent exposure classes. Logical axioms can be used to define the relationships between exposure receptors, ECTO classes, and documented exposure outcomes from the literature. Instance level schemas can also be developed using individual data points as the exposure receptor or exposure outcomes. Abbreviations: ENVO, The Environment Ontology; ChEBI, Chemical Entities of Biological Interest; NCBI Taxon, National Center for Biotechnology Information Taxonomy; FoodOn, Food Ontology; NCIt, National Cancer Institute Thesaurus; MAxO, Medical Actions Ontology; HPO, Human Phenotype Ontology; Mondo, Mondo Disease Ontology; GO, Gene Ontology.

The primary audience for ECTO includes toxicologists, clinicians, integrative and/or computational biologists, and exposure researchers who are seeking a standard for documenting environment and exposure based interventions. Additionally, ECTO is intended to serve environmental epidemiologists whose experimental designs may focus on identifying environmental exposures impacting their subjects. In turn, researchers who are interested in any related areas of biology can then also capitalize on any indicated relationships between organism exposure and health outcomes. A variety of competency questions and use cases have been documented by stakeholders including toxicologists, public health epidemiologists, and clinicians[9]. Some examples can be seen in **Supplement Table 1**.

ECTO's Methodological Framework

ECTO is available on GitHub[18]. The ECTO life cycle is managed using the Ontology Development Kit (ODK), a standardized approach for initializing ontology creation via GitHub [19]. ODK offers initial Makefiles to support a release workflow, incorporates Travis-CI for any pull requests, and allows for a standardized documentation and layout for ease of navigation. ECTO is designed to include modular classes that are generated using existing ontology terms from other related ontologies using DOSDPs. This allows for class composition to work like piecing together building blocks to create the term of interest. Exposures modeled in ECTO are based on the upper level Exposure Ontology (ExO)[20], while offering specific content such as an exposure to a chemical, an environmental condition, or a mixture of components. Created by exposure science community researchers in 2012, ExO contains a high level toxicology schema to connect exposure stimuli, receptors, routes, and media as described in Figure 2A. The ExO schema describes the components of an environmental exposure event, which can be leveraged for depicting a specific environmental exposure such as those terms found in ECTO. Figure 2B showcases the detailed modeling that is achievable using ECTO exposure terms as well as the logical axioms that can be instantiated based on literature findings.

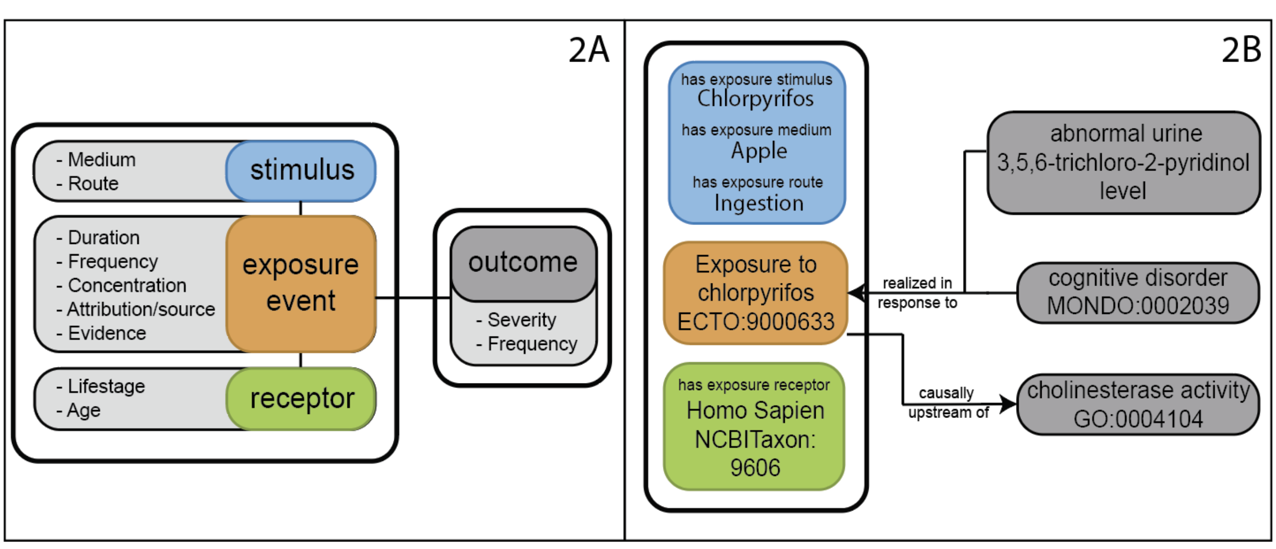

Figure 2. Overview of exposure schema. A. ExO Upper Level Schema: Modified from Mattingly et al. (2012)[20], the ExO schema for modeling exposure events forms the base infrastructure for ECTO. B. ECTO exposure schema. Utilizing the ExO schema, ECTO terms include detailed information regarding the stimulus, medium, and route of an exposure. Relationships from the OBO Relations Ontology also facilitate annotations of the exposure receptor and a variety of exposure outcomes. Further annotations regarding data specific information such as temporality or dose of the exposure can be included as annotations within a knowledge graph or other computational data structure.

ECTO treats exposures as events; in ontological terms, they are types of `occurrents` (e.g. an entity with temporal parts and that happens, unfolds or develops through time). As a subclass of `occurrent`, the `exposure event` includes interactions between a receptor (typically an organism, but could be a population of organisms or an organism part) and a `stimulus` (an agent or process that has a potential effect on the receptor). The `stimulus` may interact with the organism through some kind of environmental `medium` (e.g. air, water, soil), and may enter via some `route` (e.g. permeating the skin or analogous barrier). In turn, the exposure terms in ECTO range from somewhat broad terms (e.g. `exposure to lead)` to more specific (e.g. `exposure to lead in water via ingestion`). ECTO terms follow a standardized nomenclature of ‘`exposure to X`’ with ‘`X`’ referring to an ‘`exposure stimulus`’ term that is an existing ontology term, and the ability to add variable terms referring to the medium and route if required. By utilizing terms from existing ontologies, ECTO can additionally harmonize content from other databases annotated to the terms (e.g. CAS Registry Numbers for chemical terms). Inclusion of annotations like temporality of exposure or dose of exposure may be desired within a data model for analysis. Annotation models, analogous to those used for the Gene Ontology[21], can be used to annotate ECTO terms within a knowledge graph format to associate instance level data with the standardized exposure terms.

Content Creation for ECTO

Terms are developed for ECTO using both expert manual curation using the Protege ontology editor tool[22] and pattern based curation. To avoid ECTO becoming overly complex and subsequent maintenance challenges, a pattern-based annotation format can be used to describe unique features of the stimulus, receptor, exposure event, and outcome. Pattern-based curation is conducted using Dead simple OWL design patterns (DOSDPs)[23]. DOSDPs are easy to read, YAML based templates for generation of ontology content including labels, synonyms, text definitions, and logical axioms. DOSDPs are particularly useful in their ability to reference existing terms from other ontologies, which for ECTO is an essential component to term development. DOSDPs contain set classes and relationships that can be used to structure logical axioms, as well as variable fields which will differ for each class created with the template. Application of DOSDPs is further depicted in **Figure 3**.

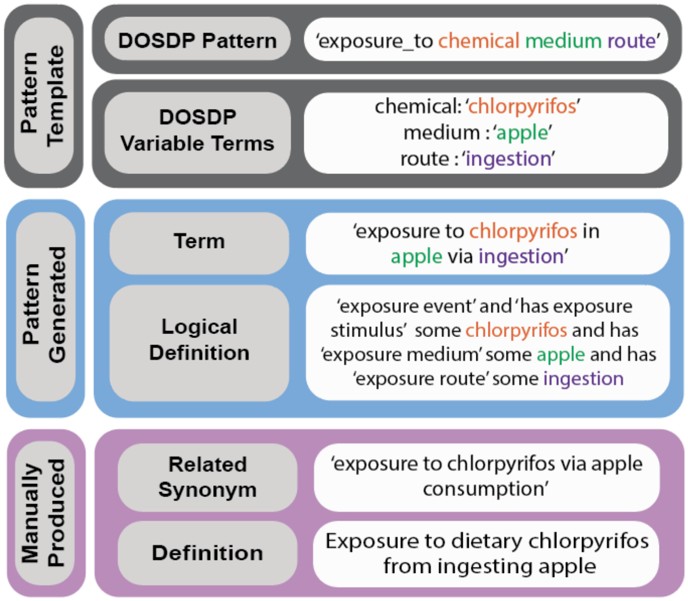

Figure 3. Chlorpyrifos Exposure in Apple Design Pattern. Represented is the 'exposure to chemical medium route' DOSDP, which can be used to create the term '`exposure to chlorpyrifos in apple via ingestion`' and its logical definition using existing 'chlorpyrifos', 'apple' and 'ingestion' terms and the DOSDP template developed by a curator. Synonyms and human readable definitions can be manually added to the term or can also be included within the DOSDP if appropriate.

Three distinct axiomatic patterns have been developed specifically for use within ECTO, including Exposure, Exposure + Route, and Exposure + Route + Medium pattern formats. Current patterns in ECTO can be viewed on GitHub[24].

In both precomposed ECTO terms and in postcomposed annotations, the Relation Ontology (RO) provides standardized relationships, including those seen in **Figure 4**.

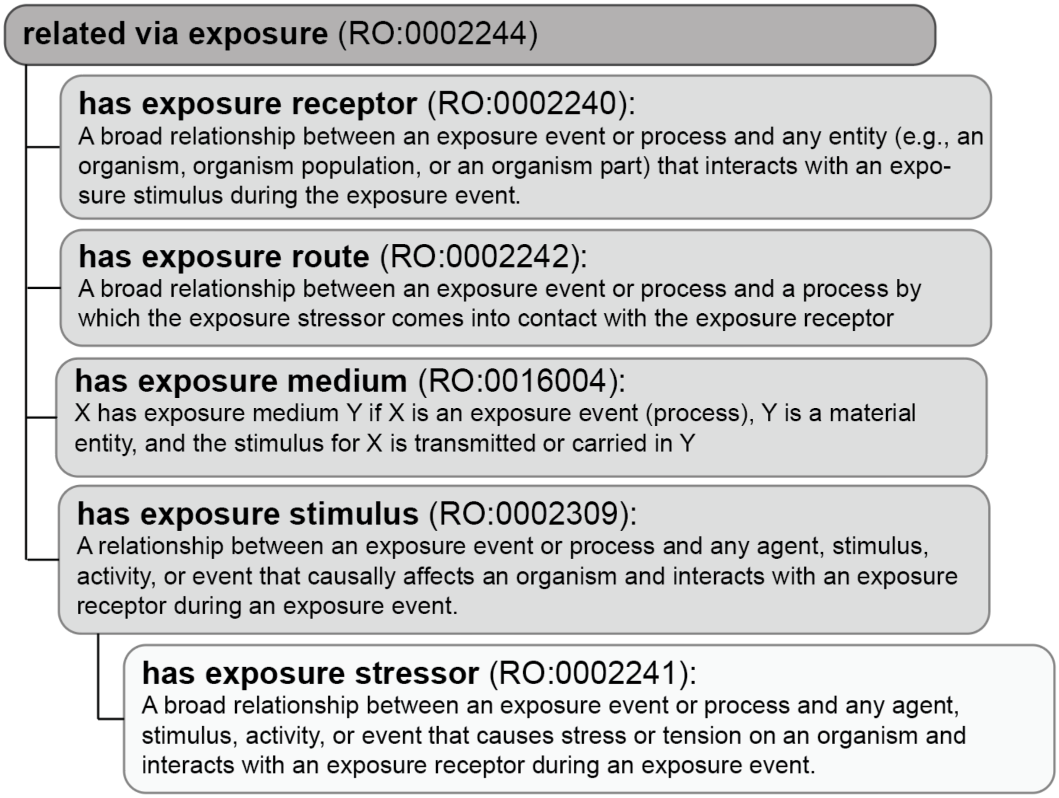

Figure 4. Exposure Based Relations. The OBO Relations Ontology contains a variety of relationship terms that fall under the superclass of ‘related via exposure’. Each of these relations can be used in conjunction with ECTO terms to create an exposure schema.

Using ECTO in Data Annotations

Annotations using ECTO are created by associating environmental exposures to a phenotype, disease, gene, or behavior in efforts to create a depiction of current exposure knowledge. Annotations can include a variety of information such as temporality, concentration/dose, and related evidence that connects the exposure to the term of interest.

ECTO annotations contain the following components:

- ECTO term (required)
- Associated phenotype/disease/behavior/gene (required)
- Reference (required if assertion is from literature)
- Evidence (required)

ECTO annotations are intended to be supported either directly or indirectly by relevant and accurate scientific literature. Information from databases and resources such as CTD or TOXNET can also be leveraged for annotations. For example, with the use of ECTO terms, annotations from CTD can be integrated into the larger Monarch knowledge graph [25]. Similarly, data from the National Toxicology Program (NTP)[26] could be structured using ECTO in combination with dosing and timing regimens along with outcomes encoded using uPheno or HPO as an annotation file format for use in downstream computation.

ECTO terms are used to axiomatize exposure-related diseases in the Mondo Disease Ontology (Mondo). Mondo integrates several underlying disease terminologies and ontologies into a merged resource that provides semantic mappings to source ontologies (https://mondo.monarchinitiative.org/). Mondo provides a library of DOSDPs, including a pattern for diseases where the cause of the disease is an exposure to an environmental stimulus[27]. Axiomizating Mondo using these standard exposure patterns allows for auto-classification of the hierarchy, and an overall more robust description of the disease term. This pattern is now used for 46 different exposure influenced disease terms in Mondo.

ECTO and Model Organism Research

Exposure modeling and annotations can be particularly useful for toxicology research using model organisms. Robust phenotype ontologies have been developed for model organisms, such as the Zebrafish Model Organism Network (ZFIN) which describes genetic, genomic, phenotypic, and developmental data for zebrafish[28], and the overarching Unified Phenotype Ontology (uPheno) which integrates multiple phenotype ontologies into a unified cross-species phenotype ontology[29].

Ontologies or standards for experimental conditions exist for some model organisms. For example, the Zebrafish Experimental Conditions Ontology (ZECO) describes experimental designs in zebrafish studies[30]. Planteome, a network of ontologies that integrate data from experiments on plants, offers a Plant Trait Ontology (TO) as well as a Plant Experimental Conditions Ontology (PECO)[31]. PECO terms describe common treatments, growing conditions, and/or study types used in plant biology experiments. Similarly to how uPheno has been developed for the unification of cross-species phenotype content, we hope ECTO can follow a similar approach to offer cross-species content regarding environmental conditions and treatments for any model organism or humans.

Modeling goals and development strategies (e.g. DOSDP) are aligned for ECTO as well as ZECO and other specific environmental condition ontologies. The similar construction offers the opportunity for a higher level unification (such as seen within phenotype ontologies and uPheno) through OWL Axiomatization and OWL Reasoning. If all experimental condition ontologies document the semantic axioms within their individual content, a reasoner (such as ELK or HermiT) can evaluate multiple ontology terms and find the overarching classes being referenced.

For example, the PECO term ‘formaldehyde exposure’ contains the logical axiom:

plant exposure and *has exposure stimulus* some formaldehyde

And a related ECTO term ‘exposure to formaldehyde’ has the logical axiom:

exposure event and *has exposure stimulus* some formaldehyde

These two similar logical axioms contain the same relationship of ‘has exposure stimulus’ and the stimulus of ‘formaldehyde’, so while the exposures (plant exposure vs exposure event) may differ, their logical axioms still allow for adequate association of the terms and a link between two related but distinct exposures.

While ECTO has similar modeling structures to related ontologies, ECTO is distinct in its descriptions of species agnostic exposures to environmental entities, chemicals, and other stimuli. Distinctions between ECTO and related ontologies are described in **Supplemental Table 2**.

Established Use Cases For ECTO

PEGS Use Case

An initial use case for ECTO was provided by the National Institute of Environmental Health Sciences (NIEHS) and their Personalized Environment and Genes (PEGS) research group [32]. PEGS researchers are focused on a variety of ways in which environmental exposures impact organism health. This use case is intended to identify methods for parsing environmental exposure and health data collected via self-reported survey and evaluate associations between singular or combined exposures that are associated with an adverse health outcome.

PEGS has developed three surveys for self-reported data collection including Health and Exposures, Internal Exposome, and External Exposome surveys. Each survey was developed to include some CDEs, however as previously noted CDEs are limited in their computational capacity and often inadequately aligned across survey tools for integrated data analysis. Our use case focused on ontological encoding of each survey question using ECTO to provide standardized language and computational structure to each survey item (**Figure 5**). In turn this methodology is a template for mapping preexisting data from heterogeneous surveys to align and compare findings. Additionally, these methods will support future survey development to enhance immediate data interoperability.

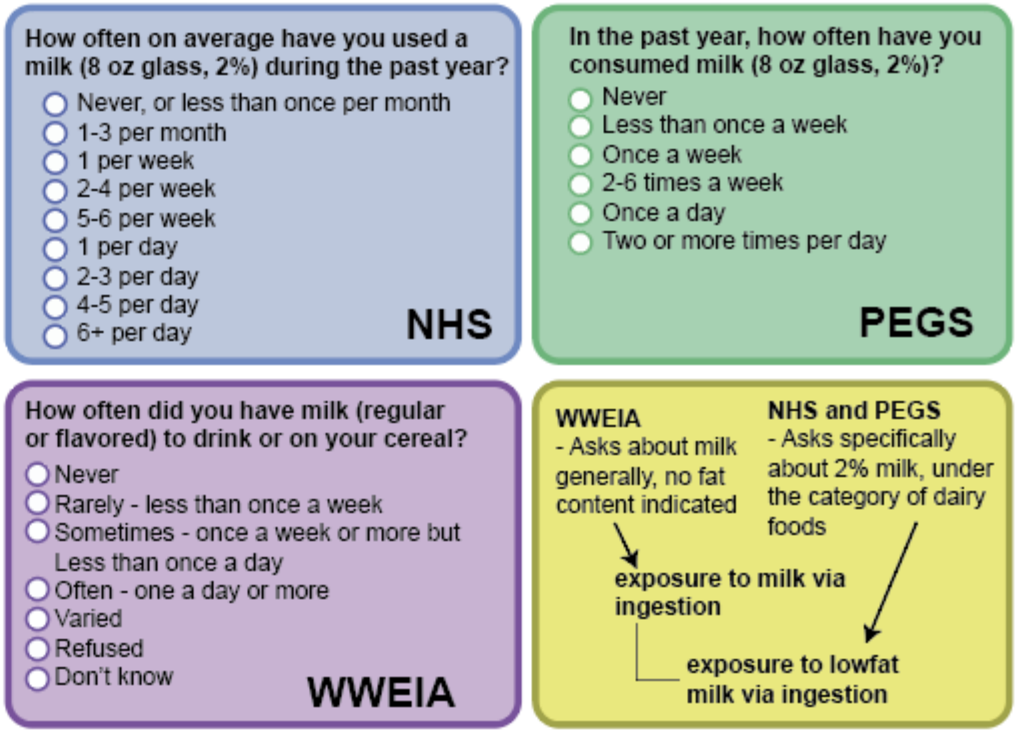

Figure 5. Making Common Data Elements (CDEs) interoperable. Surveys can include question and response CDEs. A variety of CDE registries exist, but not all of them are interoperable and some CDEs may be duplicative in their content. Three different surveys and CDEs can be seen in blue, purple, and green, which compare similar questions about milk from the Nurses' Health Study (NHS), What We Eat in America (WWEIA) and the PEGS surveys. While each question asks about milk, the responses elicited from each are not directly compatible and can be difficult to computationally assess. An ontology centric approach assesses each question and the resulting responses for the common exposure feature which can be classified using the ontology hierarchy and annotated in a knowledge graph to encompass a variety of potential responses for harmonization.

The content of the PEGS surveys was utilized as a primary resource for common workplace, home, hobby, and activity based exposures that informed ECTO's initial exposure classes. Of particular interest to PEGS researchers was the creation of mixture exposure terms that include metadata regarding each component of the mixture. For mixture compounds, our use cases required metadata for the components of the mixture within the exposure. To create mixture terms, we worked with the ENVO team to template all necessary elements for each term including information about the mixture components (e.g., methyl cellulose paste is composed of methyl cellulose and water), and created the subsequent exposure term (e.g., `exposure to methyl cellulose paste`).

Using this structure, the survey data can be classified based on exposure to the mixtures or exposure to the components within the mixture. Additionally, the exposure terms can also be classified based on the inherent relationships within the ontology (e.g. an `exposure to sulfuric acid` can be classified with other exposures to acids), further supporting higher powered assessment.

Zebrafish Use Case

We have also used ECTO for the annotation of toxicology studies, such as exposure investigations in zebrafish (*Danio rerio*). Zebrafish are a commonly used toxicological model organism due to a variety of features such as low cost, quick breeding cycle, and transparent embryos [33]. However, it has been challenging to compare results of studies performed in different laboratories because the way in which the exposure chemicals, methods and parameters, and resulting phenotypes are encoded is laboratory-specific. Further, in some cases the chemicals themselves are obfuscated due to partnerships with commercial entities. However, it is still possible to classify such chemicals into higher level categories such as "`exposure to aldehydes`". Without the computational mappings of ontology terms and logical axioms to this instance level data, researchers would have to integrate manually. By enriching these data with ontology terms, we have empowered researchers to efficiently integrate heterogeneous data across labs at scale for more powerful statistical and meta-analyses (Fig. 6).

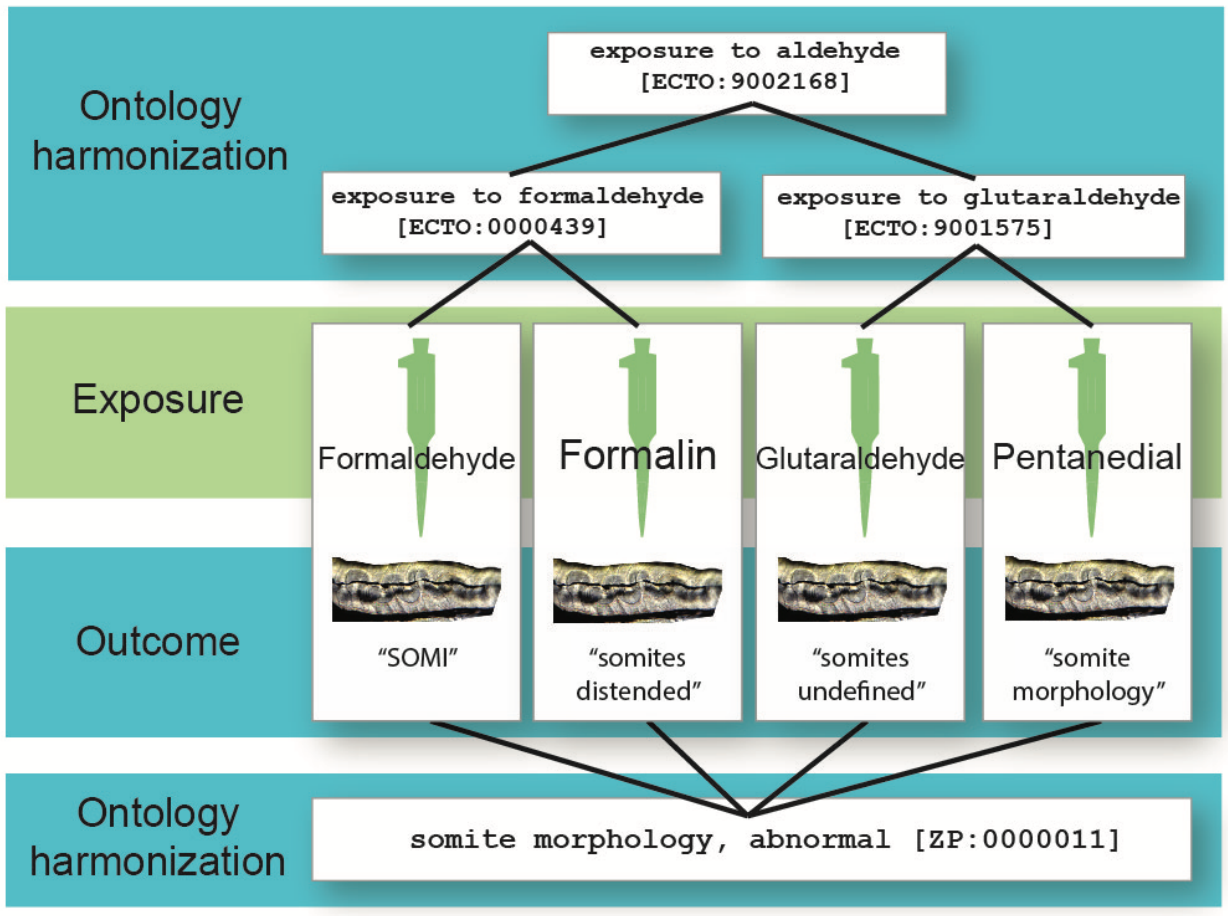

Figure 6: Standardizing exposures and outcomes in zebrafish. In this study, zebrafish embryos were exposed to an aldehyde by four different labs (formaldehyde CHEBI:16842 and glutaraldehyde CHEBI:64276). Once hatched, the zebrafish were observed for irregular phenotypes and all displayed an abnormality of their somites. The names used to describe the stimulus and the outcome was different in each of the four labs. Using ontologies like the Zebrafish Phenotype Ontology (ZP), the Environmental Conditions, Treatments, and Exposures Ontology (ECTO), and the Chemicals of Biological Interest Ontology (ChEBI) we can integrate data from four different labs even though they use different terms and different stimuli.

Limitations of ECTO

ECTO currently describes a wide array of chemical and natural or built environmental exposures, but it does not yet include exposures to infectious agents, many foods, nutrients, social environments (e.g. education, crime, and access), as well as more unique or complex multi-layered exposures (e.g. exposure to UV radiation while wearing SPF 30 sunscreen).

Another limitation of ECTO is its reliance on existing ontology content for the development of exposure terms. While many stimuli are represented in robust ontologies like ChEBI and ENVO, our team is consistently pursuing content requests in other ontologies to create ECTO terms for our use cases. This is sometimes hindered by non-OBO terminologies with relevant content not otherwise available within OBO, such as the National Cancer Institute thesaurus (NCIt).

Conclusions

ECTO is a comprehensive computational ontology designed to support any type of exposure event to any type of organism. It can be utilized to harmonize data from across sources and data modalities, such as surveys, literature annotations, toxicological studies, and in clinical research. ECTO will continue developing content for exposures to allergens, foods and nutrients, hobby and occupational exposures, and geographic location based exposures. We are particularly interested in coordinating dietary survey information from specific geographical regions with agricultural chemical usage data to ask questions such as “If a person ate an apple grown in Washington, are they likely to be exposed to chlorpyrifos?”, “what if the apple is washed?”, “was it an organic apple?” and other layers of questions to infer dietary exposures. We hope that by asking questions such as these within knowledge graphs and other instance level data visualizations, we can infer what and how exposures may be occurring, potentially assert some general quantification information on the exposure, and if available coordinate exposure findings with documented health outcomes in the respondent. With continued development of ECTO and its use in logical axioms like in Mondo disease ontology, we plan to integrate environmental exposures and coordinated health outcomes into the diagnostic tools The Monarch Initiative currently supports.

This manuscript introduces the Environmental Conditions, Treatments, and Exposures Ontology (ECTO) as described using the minimum information for the reporting of an ontology (MIRO) guidelines[34].

Ontology Owner: The Monarch Initiative

Contact: Anne Thessen, annethessen@gmail.com

License: CC BY 3.0

Ontology URL: http://www.obofoundry.org/ontology/ecto.html

Ontology Repository: https://github.com/EnvironmentOntology/environmental-exposure-ontology

We welcome user requests for new terms and other contributions via our issue tracker (https://github.com/EnvironmentOntology/environmental-exposure-ontology/issues).

References

1. Lea IA, Gong H, Paleja A, Rashid A, Fostel J. CEBS: a comprehensive annotated database of toxicological data. *Nucleic Acids Res*. 2017;45(D1):D964-D971.
2. Davis AP, Grondin CJ, Johnson RJ, et al. The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res*. 2017;45(D1):D972-D978.
3. Judson RS, Martin MT, Egeghy P, et al. Aggregating data for computational toxicology applications: The U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System. *Int J Mol Sci*. 2012;13(2):1805-1831.
4. Davis AP, Wiegers TC, Grondin CJ, et al. Leveraging the Comparative Toxicogenomics Database to Fill in Knowledge Gaps for Environmental Health: A Test Case for Air Pollution-induced Cardiovascular Disease. *Toxicol Sci*. 2020;177(2):392-404.
5. Haendel M, Eddy J, Walden A, Volchenboum S. *Response to Request for Information (RFI): Use of Common Data Elements (CDEs) in NIH-Funded Research: NOT-LM-21-005.*; 2021. doi:10.5281/zenodo.4903509
6. Viet SM, Falman JC, Merrill LS, et al. Human Health Exposure Analysis Resource (HHEAR): A model for incorporating the exposome into health studies. *Int J Hyg Environ Health*. 2021;235:113768.
7. Bodenreider O. The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Res*. 2004;32(Database issue):D267-D270.
8. Ives C, Campia I, Wang RL, Wittwehr C, Edwards S. Creating a Structured AOP Knowledgebase via Ontology-Based Annotations. *Appl In Vitro Toxicol*. 2017;3(4):298-311.
9. Thessen AE, Grondin CJ, Kulkarni RD, et al. Community Approaches for Integrating Environmental Exposures into Human Models of Disease. *Environ Health Perspect*. 2020;128(12):125002.
10. National Academies of Sciences Engineering, Medicine. *Informing Environmental Health Decisions Through Data Integration: Proceedings of a Workshop—in Brief*. (Alper J, ed.). The National Academies Press; 2018.
11. Manrai AK, Cui Y, Bushel PR, et al. Informatics and Data Analytics to Support Exposome-Based Discovery for Public Health. *Annu Rev Public Health*. 2017;38:279-294.
12. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016;3(1):160018.
13. Buttigieg PL, Pafilis E, Lewis SE, Schildhauer MP, Walls RL, Mungall CJ. The environment ontology in 2016: bridging domains with increased scope, semantic density, and interoperation. *J Biomed Semantics*. 2016;7(1):57.
14. Hastings J, Owen G, Dekker A, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites. *Nucleic Acids Res*. 2016;44(D1):D1214-D1219.
15. Lozowicka B. Health risk for children and adults consuming apples with pesticide residue. *Sci Total Environ*. 2015;502:184-198.
16. Hites RA. The Rise and Fall of Chlorpyrifos in the United States. *Environ Sci Technol*. 2021;55(3):1354-1358.
17. Chlorpyrifos. Accessed January 26, 2022. http://npic.orst.edu/factsheets/chlorpgen.html
18. *GitHub - Environmental Conditions, Treatments, and Exposures Ontology (ECTO)*. The Environment Ontology; 2020. https://github.com/EnvironmentOntology/environmental-exposure-ontology
19. Matentzoglu N, Mungall C, Goutte-Gattat D. *Ontology Development Kit*.; 2021. doi:10.5281/zenodo.5762512
20. Mattingly CJ, McKone TE, Callahan MA, Blake JA, Hubal EAC. Providing the missing link: the exposure science ontology ExO. *Environ Sci Technol*. 2012;46(6):3046-3053.
21. Gene Ontology Consortium, Blake JA, Dolan M, et al. Gene Ontology annotations and resources. *Nucleic Acids Res*. 2013;41(Database issue):D530-D535.
22. Musen MA, Protégé Team. The Protégé Project: A Look Back and a Look Forward. *AI*

Matters. 2015;1(4):4-12.
23. Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. Dead simple OWL design patterns. *J Biomed Semantics*. 2017;8(1):18.
24. *ECTO - Dead Simple OWL Design Patterns Repository*. Github Accessed February 23, 2022. https://github.com/EnvironmentOntology/environmental-exposure-ontology
25. Shefchek KA, Harris NL, Gargano M, et al. The Monarch Initiative in 2019: an integrative data and analytic platform connecting phenotypes to genotypes across species. *Nucleic Acids Res*. 2020;48(D1):D704-D715.
26. Home - National Toxicology Program. Accessed March 9, 2022. https://ntp.niehs.nih.gov/
27. *DOSDP - Realized_in_response_to_environmental_exposure.yaml*. Github Accessed February 23, 2022. https://github.com/monarch-initiative/mondo
28. Howe DG, Bradford YM, Eagle A, et al. The Zebrafish Model Organism Database: new support for human disease models, mutation details, gene expression phenotypes and searching. *Nucleic Acids Res*. 2017;45(D1):D758-D768.
29. Matentzoglu N, Anangnostopoulos A, Balhoff J, et al. uPheno 2.0: Unifying phenotype representation and analysis across species. https://docs.google.com/document/u/1/d/14J4AxwAhUfKgnHfFaUJ0AoMwQ8UKUI4kyuTpDOFnUjI/edit?usp=gmail&usp=embed_facebook
30. Bradford YM, Van Slyke CE, Toro S, Ramachandran S. The zebrafish experimental conditions ontology. Accessed July 7, 2021. http://ceur-ws.org/Vol-1747/IP25_ICBO2016.pdf
31. Cooper L, Meier A, Laporte MA, et al. The Planteome database: an integrated resource for reference ontologies, plant genomics and phenomics. *Nucleic Acids Res*. 2018;46(D1):D1168-D1180.
32. PEGS: Personalized Environment and Genes Study. Accessed November 10, 2021. https://www.niehs.nih.gov/research/clinical/studies/pegs/index.cfm
33. Dai YJ, Jia YF, Chen N, et al. Zebrafish as a model system to study toxicology. *Environ Toxicol Chem*. 2014;33(1):11-17.
34. Matentzoglu N, Malone J, Mungall C, Stevens R. MIRO: guidelines for minimum information for the reporting of an ontology. *J Biomed Semantics*. 2018;9(1):6.
35. Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, ENVO Consortium. The environment ontology: contextualising biological and biomedical entities. *J Biomed Semantics*. 2013;4(1):43.
36. Dooley DM, Griffiths EJ, Gosal GS, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *npj Science of Food*. 2018;2(1):1-10.
37. Golbeck J, Fragoso G, Hartel F, Hendler J, Oberthaler J, Parsia B. The National Cancer Institute's Thesaurus and Ontology. Published online 2003. doi:10.2139/ssrn.3199007
38. Federhen S. The NCBI Taxonomy database. *Nucleic Acids Res*. 2012;40(Database issue):D136-D143.

Supplement Table 1. ECTO Applications: A variety of use cases have been identified for ECTO based on previous workshops with stakeholders in this field. We foresee three primary theme areas in which ECTO may be utilized (stated in the column headers) and provide examples of what each theme entails.

	Bench Science/Translation	**Clinical Research**	**Epidemiology/Ecosystem**
General Use Case	Standardizing descriptions for toxicological experimental methods to offer coordination of data from same model studies as well as from different/higher level model studies for translational science.	Describing standardized exposures occurring in humans in coordination with clinical health outcomes, including genotype, phenotype, and disease.	Establishing structured concepts for single or composite exposures impacting large populations of humans or ecosystem environments.
Example Specific Use Case	Translating toxicological endpoint terms, like neurodevelopmental toxicity, into phenotype terms from HPO (Human Phenotype Ontology)	Coordinating an individual's exposure to air pollutants at their home and work place with their clinical phenotypes and health outcomes.	Predict what a disease cluster would look like based on the structure and/or composition of a chemical or mixture that is similar to the chemical/mixture in question.
Example Competency Questions	What genetic variants are associated with greater sensitivity to altered neurologic function from exposure to polychlorinated biphenyls? What are the developmental effects in animals and humans exposed prenatally to benzene over short or long periods of time? What are the methylation tag pattern changes related to prenatal exposure to benzene in animals?	What levels of benzene exposure in the air over what period of time are likely to cause blood cancers? Which blood cancers? What is my biggest exposure risk based on my geographical location? How can I prevent or minimize that exposure? What am I exposed to in my particular line of work? How might this impact my health?	Which chemicals in stormwater run-off have aquatic toxicity data available from assays done at different salinities? What is the effectiveness of environmental regulations for reducing health impacts from air and water contaminants? What sources of chemical exposures exist in my community?

Supplemental Table 2. Comparison of ECTO scope to other ontologies: The ontologies and resources indicated within this table are utilized as imports within ECTO to support its precomposed exposure classes. In addition to the listed ontologies, ECTO also imports Gene Ontology, Information Artifact Ontology (IAO), Medical Actions Ontology (MAxO), Neuro Behavior Ontology (NBO), NanoParticle Ontology (NPO), Phenotype and Trait Ontology (PATO), Relation Ontology (RO), and Uber-anatomy Ontology (UBERON).

Ontology	Status	No. of Classes	Scope	Description	URL or Reference
Environmental Conditions, Treatments, and Exposures Ontology	Active	2763	Exposures to environments, substances, behaviors, and scenarios.	A structured vocabulary for experimental, clinical, and natural world exposures.	https://github.com/EnvironmentOntology/environmental-exposure-ontology

Ontology	Status	No. of Class references in ECTO	Scope	Description	URL or Reference
Environment Ontology (ENVO)	Active	333	Environmental features and habitats	Concise, controlled description of environments.	[35] https://github.com/EnvironmentOntology/envo
Exposure Science Ontology (ExO)	Active	57	Higher level environmental structure	Vocabularies for describing exposure data to inform understanding of environmental health.	[20] https://github.com/CTDbase/exposure-ontology
Chemical Entities of Biological Interest (ChEBI)	Active	2170	Chemicals, molecular entities	A structured classification of molecular entities of biological interest focusing on 'small' chemical compounds.	[14] https://github.com/ebi-chebi/ChEBI

Food Ontology (FoodOn)	Active	169	Food material, food processing	A controlled vocabulary of food products and entities bearing a "food role"	[36] https://github.com/FoodOntology/foodon
National Cancer Institute Thesaurus (NCIT)	Active	48	Broad coverage of the cancer domain	A thesaurus of cancer related diseases, findings and abnormalities	[37]
Experimental Conditions Ontology (XCO)	Active	0	Internal and external environmental conditions	A vocabulary of conditions under which physiological and morphological measurements are made.	https://jbiomedsem.biomedcentral.com/articles/10.1186/2041-1480-4-26 http://www.obofoundry.org/ontology/xco.html
National Center for Biotechnology Information (NCBI) Organismal Classification (NCBI taxon)	Active	11	Organism taxonomy	A classification and nomenclature for all of the organisms in the public sequence databases.	[38]
Ecocore	Active	0	ecological entities, such as ecological functions (for predators, prey, etc), food webs, and ecological interactions.	Concise and controlled description of ecological traits of organisms.	https://github.com/EcologicalSemantics/ecocore
Zebrafish Experimental Conditions Ontology (ZECO)	Active	0	Experimental conditions for zebrafish	Descriptions of experimental conditions applied to zebrafish.	[30] https://github.com/ybradford/zebrafish-experimental-

					conditions-ontology
Plant Experimental Conditions Ontology (PECO)	Active	0	Experimental conditions for plants	Descriptions of experimental conditions applied to plants.	[31] https://github.com/Planteome/plant-experimental-conditions-ontology