

OBO Foundry Food Ontology Interconnectivity

Damion Dooley^{1*}, Liliana Andrés-Hernández², Georgeta Bordea³, Leigh Carmody⁴, Duccio Cavalieri⁵, Lauren Chan⁶, Pol Castellano-Escuder⁷, Carl Lachat⁸, Fleur Mouglin³, Francesco Vitali⁹, Chen Yang⁸, Magalie Weber¹⁰, Matthew Lange¹¹

¹ Centre for Infectious Disease Genomics and One Health, Simon Fraser University, Burnaby, BC, Canada

² Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia

³ Bordeaux Population Health, University of Bordeaux, Bordeaux, France

⁴ The Jackson Laboratory for Genomic Medicine, Farmington, CT, USA

⁵ Department of Biology, University of Florence, Florence, Italy

⁶ College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, USA

⁷ Biomarkers and Nutritional & Food Metabolomics Research Group, University of Barcelona, Barcelona, Spain

⁸ Department of Food Technology, Safety and Health, Ghent University, Ghent, Belgium

⁹ Institute of Agricultural Biology and Biotechnology - National Research Council, Milano, Italy

¹⁰ INRAE, UR BIA, Biopolymères Interactions Assemblages, Nantes, France

¹¹ International Center for Food Ontology Operability Data and Semantics, Davis, California, USA

* corresponding author

Abstract

Since its creation in 2016, the FoodOn ontology has become an interconnected partner in various academic and government inter-agency ontology work spanning agricultural and public health domains. This paper examines existing and potential data interoperability capabilities arising from FoodOn and partner food-related ontologies belonging to the encyclopedic Open Biological and Biomedical Ontology Foundry (OBO) vocabulary platform, and how research organizations and industry might utilize them for their own operations or for data exchange. Projects are seeking standardized vocabulary across all direct food supply activities ranging from agricultural production, harvesting, preparation, food processing, marketing, distribution and consumption, as well as indirectly, within health, economic, food security and sustainability analysis and reporting tools. To satisfy this demand and provide data requires establishing domain specific ontologies whose curators coordinate closely to produce recommended patterns for food system vocabulary.

Keywords

Ontology, data harmonization, OBO Foundry, food systems, public health, epidemiology, multiontology framework, One Health

1. Introduction

Ontologists and semantic web advocates envision a future in which stakeholders in all sectors will be able to take advantage of a harmonious federated data landscape built on the interoperability prowess of ontologies. This data interconnectivity vision has been supported by academic and government research sectors, exemplified in curation consortia such as the open source inter-agency Open Biological and Biomedical Ontology Foundry (OBO) [1], which contains a collection of domain

IFOW 2021: 2nd Integrated Food Ontology Workshop, held at JOWO 2021: Episode VII The Bolzano Summer of Knowledge, September 11-18, 2021, Bolzano, Italy

EMAIL: damion_dooley@sfu.ca (D. Dooley)

ORCID: 0000 0002 8844 9165 (D. Dooley)

© 2021 Copyright for this paper by its authors.

Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).



CEUR Workshop Proceedings (CEUR-WS.org)

specific ontologies that facilitate interoperability by adhering to a set of curation and logic patterns. OBO promotes best practices to their member ontologies for ensuring such vocabulary is clear and accessible while allowing shared principles governing ontology development to evolve. OBO provides a web service for permanent links to term resolution (called purls), as well as guidelines for establishing term hierarchies, term deprecation, and logically tested, versioned quality control within an encyclopedic (domain specific) curation environment of experts. OBO's standard OWL ontology format is suited to expressing international standards in minute detail as data structures with context-sensitive terminology, synonymy, and categorical, numeric and textual variables. Tools such as the OBO Dashboard [2] are helping with the continuous improvement of ontology quality.

The OBO approach was presaged by a paper advocating for the separation of curated biomedical vocabulary from the content of clinical and research databases [3]. Relative to food, Lange et. al. explored requirements and a prototype for a multi-ontology framework for describing and guiding agriculture, food, diet and health [4]; a current review of farm-to-fork data harmonization approaches [5] reinforces how important ontologies are in this effort.

OBO houses many evolving ontologies that support a One Health [6] paradigm, including FoodOn [7], which provides a food-centric perspective which networks with many of the aforementioned ontologies and which can also be used in conjunction with food production and processing ontologies outside the OBO landscape. One key OBO Foundry curation principle, called The Minimum Information to Reference an External Ontology Term (MIREOT) [8], is the reuse of terms from other OBO ontologies to avoid the costly and confusing situation where term entities having similar or identical semantics but different identifiers exist in multiple ontologies. Reusability patterns include the wholesale import of large branches of ontologies, for example, the ENVO import of FoodOn food products; as well as selective term reuse; both approaches are described on the FoodOn [technical reuse page](#). This principle has been particularly important in the growth, expansion, and development of FoodOn and related ontologies listed below:

- **The Agronomy Ontology (AGRO)** covers agricultural management practices applied to matrices of crop plots [9].
- **The Chemical Entities of Biological Interest (ChEBI)** ontology organizes molecular entities - mainly 'small' chemical compounds (natural or synthetic) - pertinent to the processes of living organisms [10].
- **The Farm to Fork Food Ontology (FoodOn)** provides terms for generic (non-branded) food products available at any point in the global food supply chain, as well as facets of terms for food production processes and food characteristics. FoodOn can be reused in OBO wherever food product references occur [7].
- **The Compositional Dietary Nutrition Ontology (CDNO)** covers nutritional composition terms (vitamins, carbohydrates etc.) of organism anatomical parts like seeds, or fruit (which form the fundamental layer of food products in FoodOn) in a diet-and-nutrition community friendly ontology [11].
- **The Human Disease Ontology (DOID)** names food related allergies and their food product triggers [12].
- **The Medical Action Ontology (MAxO)** project is exploring the relation between nutrition deficiency and rare diseases [13]
- **The Health Surveillance Ontology (HSO)** describes animal health, public health and food safety surveillance systems, including proactive surveillance and reactive investigation methods and objectives [14].
- **Environmental Conditions, Treatments, and Exposures Ontology (ECTO)** provides language to describe experimental and environmental factors for public health and environmental monitoring objectives [15].

- **The Food-Biomarker Ontology (FOBI)** documents chemical biomarkers of consumed food products left in stool or urine [16].
- **The Food Interactions with Drugs Evidence Ontology (FIDEO)** focuses on vocabulary useful to identify research on the influence of food consumption on oral drug ingestion [17].
- **The Ontology for Nutritional Studies (ONS)** focuses on vocabulary for modeling nutritional studies, including diet and dietary pattern variations [18].
- **The Ontology for Nutritional Epidemiology (ONE)** focuses on detailing nutritional study document structure (e.g. research manuscripts, dietary surveys, food-based dietary guidelines, etc.) and dataset characteristics [19].
- **The Process and Observation Ontology (PO²)** details characteristics and sampling regimes of foods during their manufacturing process, as well as the processing steps they are subjected to [20].

2. Methods

The process of developing an ontology within the context of OBO can resemble a stand-alone ontology effort to start - both stand-alone and OBO infused efforts may involve reuse of parts from other ontologies, and with just a few curators representing one or more stakeholder groups. However, as the OBO community is engaged, the development methodology requires familiarization with many aspects of OBO interdependence:

- Adoption of OBO standardized term URL's that usually point to an ontology search engine term result like ontobee.org [21]. Along with the MIREOT principle, this fulfills the encyclopedic FAIR data vision of OBO.
- A service model whereby new term requests (NTR) are handled usually via Github requests or larger bulk review and import projects. This is critical for satisfaction of peer networked ontologies, otherwise bottlenecks occur.
- An ontology is expected to attract more than one user, thus encouraging other stakeholders to be identified and even encouraged to join the curation team. Eventually monthly, bi-weekly or weekly curation calls are needed depending on its growth curve and volunteer or funded development capacity. This greatly improves an ontology's standing as a de facto or more official standard.
- Specialized workgroups can focus on particular problems, for example the Food Process Ontology Workgroup (progress will be reported at IFOW 2021).
- Ontology scope evolution to allow branches of terms to be added or calved off into a new ontology if core skilled curation teams can support them.
- Other structural and logical constraints aimed at supporting data harmonization within OBO [22] including Basic Formal Ontology (BFO) [23] compatibility.

A workgroup called the Joint Food Ontology Workgroup [24] was launched in the spring of 2020 as an informal methodology for term development within the food systems domain. FoodOn had previously received batches of requests for diet and nutrition terms from a few other ontologies, but was aware that ONS was under development as a new entry into OBO, and could be a potential niche home for the requested terms. The workgroup convened once a month with representation from the NTR-requesting ontologies as well as USDA, FDA, and other academic and research agencies keen to help and to assess the appropriate reuse of ontologies within their operations. During that time a corpus of over 60 diet and dietary pattern terms was discussed, reviewed, approved and then handed to ONS for implementation. A similar discussion group called the Food Process Ontology Workgroup

is under way, tasked with reviewing process related ontologies and creating a generic food processing model, and closely related recipe and ingredient model.

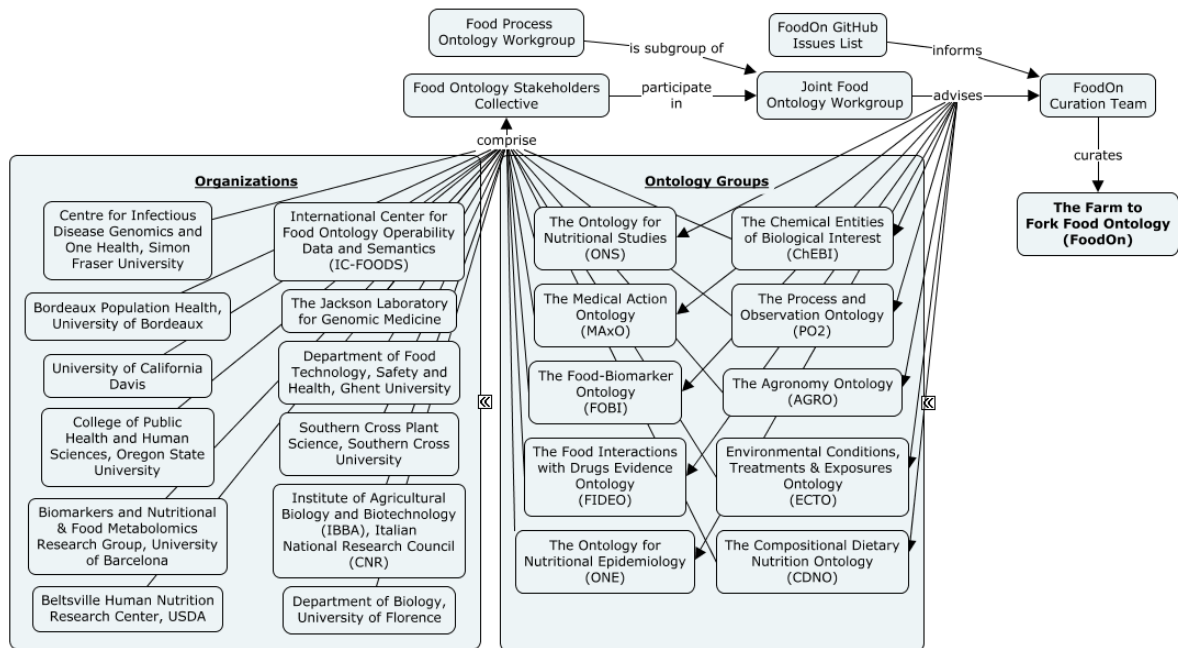


Figure 1: OBO Foundry food ontology design and curation workflow

Through this methodology, each additional project strengthens the capacity of OBO in general, and FoodOn in particular, to become the lingua franca for the unambiguous mapping / exchanging / synthesizing of agriculture, food, diet, and health -related data. The key value-add is that knowledge produced by this federated activity exists precisely because the unique combination of vocabularies that comprise FoodOn, and indeed the entire OBO Foundry, is not constrained by the narrower mandates of any particular organization.

2.1. FoodOn: A farm to fork food ontology

FoodOn entered the OBO Foundry in 2016, calving-off food terms from ENVO [25], and subsequently integrated with other OBO food chemistry and nutrition ontologies in a piecemeal fashion as research projects required it. FoodOn also inherited, and has since evolved, the basic structure of LanguaL [26], a popular food composition database (FCD) vocabulary originating in US FDA CFSAN in 1975. FoodOn’s mandate is to describe and provide precomposed terms for generic (non-branded) food products that a food producer, food manufacturer/processor, or consumer can find in the food supply chain, ranging from wild or farmed food, to processed, wholesale, retail, prepared, vendor, restaurant or home-cooked food. This includes extensive food description facets, such as applied cooking treatments, preservation methods, packaging, and food source organism taxonomy. FoodOn is being used or introduced collaboratively into a number of databases and standards, for example:

- **The USDA FoodData Central** website (<https://fdc.nal.usda.gov>) now provides FoodOn identifiers and categories for its Foundation Foods database entries, with plans to expand ontology capability in the future.
- **The FDA CFSAN GenomeTrakr database** ([Whole Genome Sequencing \(WGS\) Program | FDA, Poster](#)) which contains over 45,000 foodborne pathogen genomic sequences and their metadata, which are matched to FoodOn, NCBITaxon and ENVO ontology terms using a textual sample description to ontology term software called LexMapr. GenomeTrakr records

are then submitted to the NCBI Biosample sequence repository to assist in foodborne outbreak and antimicrobial resistance research.

- **WikiFCD** (<https://wikifcd.wiki.opencura.com>), a wikibase database of food composition and nutrient information that explores crowdsourcing curation.
- **The Genomic Standards Consortium (GSC) of minimum information standards (checklists) (MiXS)**(<https://genc.org/mixs/>) is adding a food package [27] for agriculture and industry-situated sampling metadata for pathogen and metagenomic analysis.
- The draft ISO/TC 34/SC 9 standard "**Microbiology of the Food Chain — Whole Genome Sequencing, Typing and Genomic Characterization of Foodborne Bacteria**"
- **The FDA Seafood Product List** is being worked on collaboratively by FoodOn and FDA staff to expand the mapping of common language fish names to precise scientific taxonomy names in order to improve food traceability and authentication.
- **FoodKG** (<https://foodkg.github.io/>), a knowledge graph launched in 2019 representing over 1 million recipes which is constructed with an ontology combining FoodOn, CHEBI and other resources like the USDA Nutrient Database, and the <http://im2recipe.csail.mit.edu/> photograph-recipe matching project, enables querying of recipes by ingredient, cook time, course type, and meal type.

FoodOn reuses CHEBI, ENVO, CDNO and ONS terms, among others. Not all OBO ontologies are integrated with FoodOn, for example, the Drug Ontology (DRON) has '[cucumber allergenic extract](#)' but no FoodOn, UBERON or NCBITaxon term references related to it, so partner ontology integration is an ongoing refinement. As well, new terms are being introduced into OBO via partnerships, such as an Institute for Food Safety at Cornell University list of over 350 food related equipment and tool terms.

Additionally FoodOn imports agency hierarchies to varying depths from European, North American and some international standards as they were represented in LanguaL - including the EFSA FoodEx2 Exposure hierarchy, the US Code of Federal Regulations (CFR) hierarchy, and GS1 food categories. An upcoming objective is to map to these branches more extensively by way of 'has member' relation to FoodOn's own food product hierarchy, to enable data exchange and harmonization to the deepest level of food product classes.

2.2. AGRO: The Agronomy Ontology

The Agronomy Ontology (AgrO) describes agronomic management practices, implements, and variables used during agronomic experiments. AgrO was started in 2017 in the context of the CGIAR Platform for Big Data in Agriculture, from traits and parameters identified by agronomists and crop modelers and from the Environment Ontology (ENVO). As an OBO Foundry ontology, AgrO reuses terms coming from several ontologies including ENVO, CheBI, UO, PATO, TO/CO and FoodOn. A main use case for AgrO is the Agronomy Field Information Management System (AgroFIMS). AgroFIMS enables the design of agronomic trials and the digital collection of agronomic data that is annotated from the start with agronomic terms coming from AgrO.

AgrO relies on FoodOn as the source of terms for crop residues [AGRO:00000154] that derive from food products. In the near future, AgrO will have dependencies on FoodOn for food and nutrition terms important at the post harvest stage.

2.3. CDNO: Compositional Dietary Nutrition Ontology

CDNO, launched in 2020 and now a key part of FoodOn, provides terminology for nutritional attributes from crops, livestock, and fisheries that contribute to human diet and which are referenced in precision food commodity laboratory analytics.

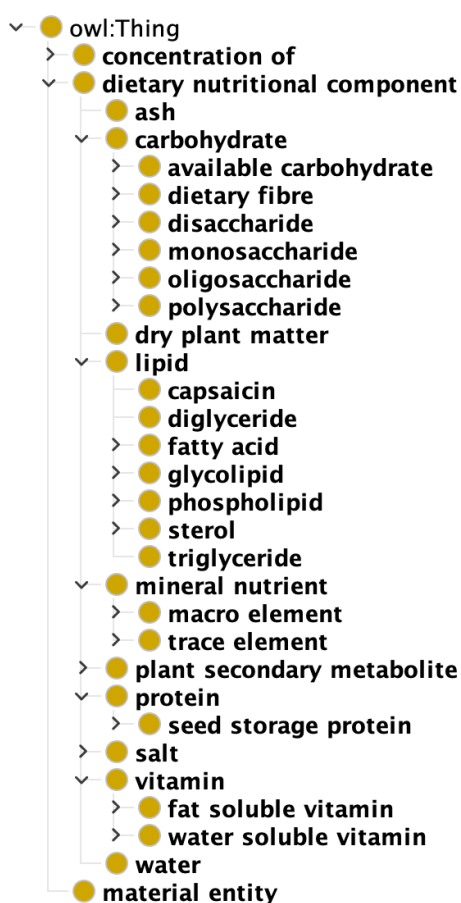


Figure 2: Visualization of the top level CDNO 'dietary nutritional component' class and its subclasses.

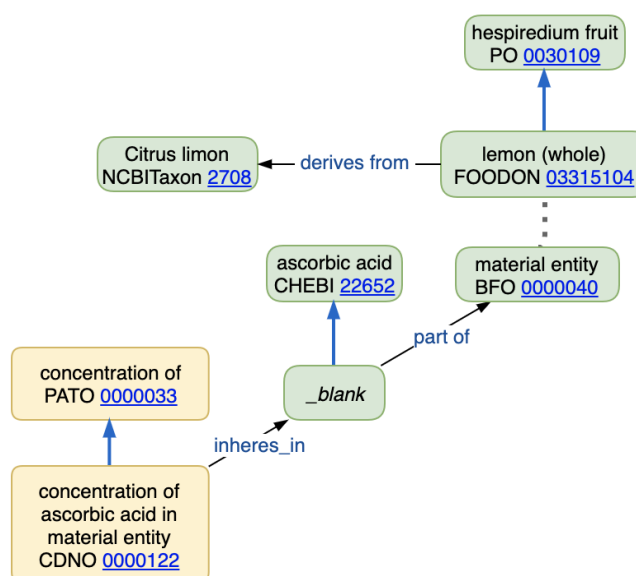


Figure 3: Visualization of the term 'concentration of ascorbic acid in material entity' [CDNO:0000122] from the 'nutritional component concentration' class.

CDNO defines a comprehensive nutrition-oriented hierarchy to organize ChEBI chemicals and associated concentrations by following a design pattern that involves the reuse of the term 'concentration of' from PATO and the term 'material entity' from the BFO (Figure 3). (This view was created because of ChEBI's inherent hierarchy of molecular entities and their roles is not easy for nutritionists to navigate [11]). The CDNO hierarchy is complemented by proposed classes for physical and functional attributes and dietary functional roles. One goal of this work is to allow harmonisation of the nutrient measures used for international standards such as the FAO sponsored International Network of Food Data Systems (INFOODS)[28] system for tagging nutrient measures in food composition databases as well as the USDA Nutrient DB codes. However, the nomenclature required to describe variation in analytic measures is a future mission.

CDNO was developed with the primary aim of adding value to datasets and their comparison, where terms from the 'nutritional component concentration' class are associated in the data curation process with specific food raw materials and associated metadata at any point in the supply chain, from cultivation/production in agriculture through to processing and consumption. CDNO, FoodOn

and Plant Ontology (PO) curators worked together in a collaborative effort to establish an initial set of over 58 food raw materials entities, defining the source of plant samples and crop production processes or stages from which the food raw material was taken. This process required analysis and attribution of existing terms from the PO and NCBI Taxon ID, in order to represent specific plant food products. For example “an apple fruit” is represented with the label ‘apple (whole)’ in FoodOn as a subclass of PO ‘pome fruit’ [PO:0030110] which ‘derives from’ [RO:0001000] the organism ‘Malus domestica’ [NCBITaxon:3750]. Species-specific datasets in tabular form can simply have columns for organism, anatomical part, and FoodOn term if desired, while graph database treatments can have structures created directly from CDNO and FoodOn OWL axioms that reflect specific material dietary nutritional component concentrations.

2.4. DO: The Human Disease Ontology

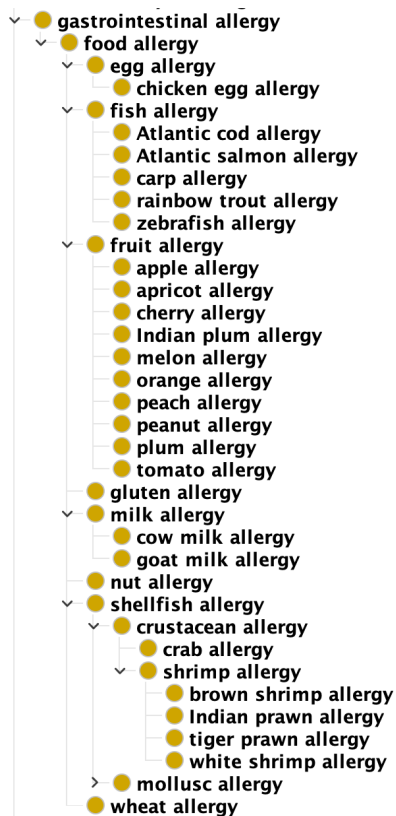


Figure 4: The DO gastrointestinal allergy hierarchy

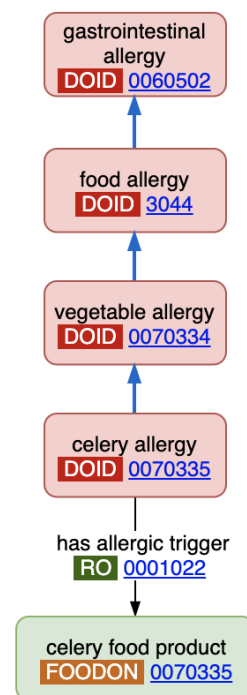


Figure 5: Food disease hierarchy enables inference.

The Human Disease Ontology (DO) uses Relation Ontology (RO) term “[has allergic trigger](#)” to attach an allergic disease to the food(s) that trigger it. This relation is used for other connections in addition to food, for example, to connect penicillin to penicillin allergy. Food allergies (a subclass of gastrointestinal allergy, defined as “An allergic disease that is located_in the gastrointestinal tract”) have fish and shellfish, fruit, milk, wheat and vegetable subclasses, and corresponding relations to FoodOn food products that cause them. This branch DOID also supports the Immune Epitope Database (IEDB) [29] by way of an IEDB “slim” export file of almost the entire food allergy branch. It appears that the vegetable allergy branch is accidentally omitted from this slim file (shown in Figure 4), and a Github issue has been raised to remedy the omission. There may be further opportunity for modelling disease here by linking to allergy symptoms.

2.5. HSO: The Health Surveillance Ontology

HSO is a knowledge model for data collection, collation and reporting of One Health surveillance activity. It recognizes that although detailed surveillance data may not be easily harmonizable and/or shareable due to specific national agency reporting requirements, top-down semantic harmonization of reporting elements is required to pool summary data for international / multi-agency contexts. Engaging multilaterally about what essential components of public health, animal health, and food safety surveillance data are, lead the HSO curation team to develop vocabulary and a model centred around a ‘surveillance activity’, a subclass of Ontology for Biomedical Investigations (OBI) [30] planned process’. HSO is an outcome of the One Health surveillance Initiative on harmOnization of data collection and interpretation (ORION) project and is currently interoperable with various catalogues in EFSA’s Standard Sample Description.

2.6. MxO: The Medical Action Ontology

MxO, launched in 2020, is a broad ontology that provides a structured vocabulary to medical procedures, interventions, therapies, treatments, or clinical recommendations, including nutritional recommendations. MxO was designed to provide a thorough resource for annotating diseases, in particular, rare diseases, where nutritional needs are often critical. In order to capture the relationship between treatments and diseases, the Phenotypic Observation Explication Tool (POET) was developed to establish a relationship between MxO, Human Phenotype Ontology (HPO), and Mondo Disease Ontology (Mondo) terms. This tool will allow researchers to actively participate in annotating diseases in their expertise. MxO annotations and the POET tool will be available on the HPO website (hpo.jax.org) by 2022.

nutrition intake avoidance

http://purl.obolibrary.org/obo/MAXO_0000056 Copy

Synonyms: food consumption avoidance

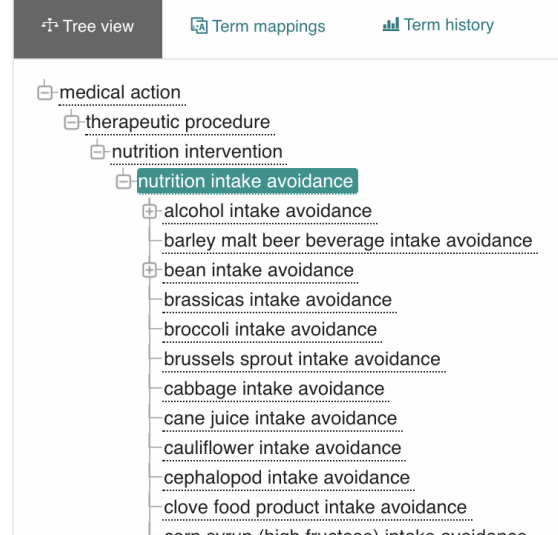


Figure 6: Nutrition intervention hierarchy within MxO

MxO provides a lexicon of 76 dietary intake avoidance behaviour terms, and an “[avoided food](#)” object property to detail which FoodOn products are being avoided. Moreover, MxO utilizes 5 FoodOn terms for nutritional supplementation recommendation. These terms will be used to annotate diseases that require nutrition therapy or management.

2.7. ECTO: The Environmental Conditions, Treatments, and Exposures Ontology

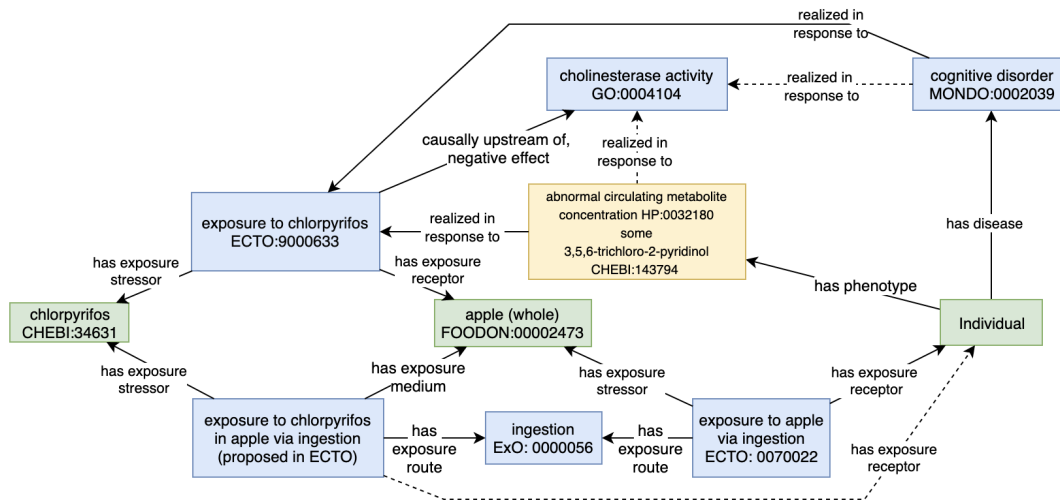


Figure 7: Model of a food-related pesticide exposure in ECTO

ECTO has been gradually evolving since 2016 with a focus on documenting precomposed experimental treatments, non-experimental exposures, and environmental conditions that may impact humans and other organisms. ECTO ranges widely and encompasses terms such as ‘exposure to arsenic’ or ‘exposure to increased temperature’ which can be meaningful for modeling experimental designs. Additionally, this broadly scoped ontology includes exposure terms related to food and nutrient exposures such as “ingestion of skim milk”. ECTO currently has 160+ food ingestion terms, 17 vitamin and mineral ingestion terms, and a developing Dead Simple OWL Design Pattern (DOSDP) [31] which will integrate exposures to specific diets that refer to terms found within the Ontology for Nutritional Studies. In turn, terms within ECTO can be utilized to describe and document research designs in toxicology and exposures, epidemiology, and nutrition in support of standardized language and data harmonization across the literature. ECTO terms can also be leveraged for modeling components of human disease, environmental toxin exposure, and alteration of biological function.

2.8. FOBI: The Food-Biomarker Ontology

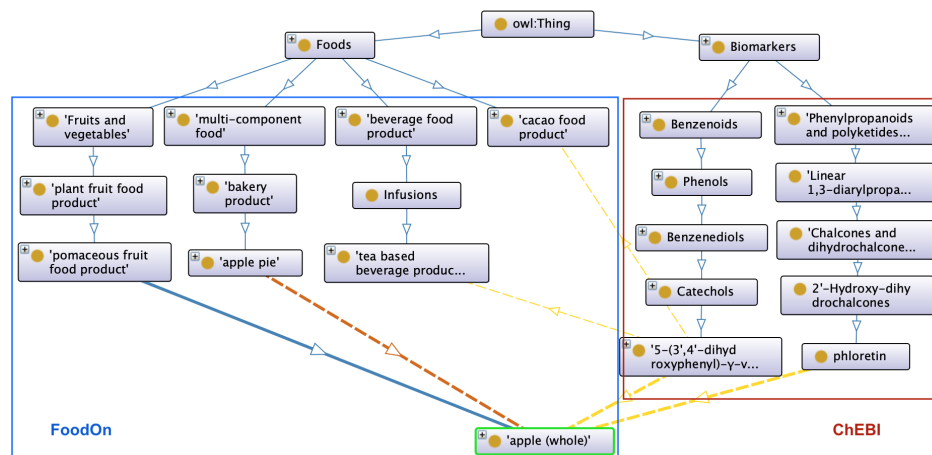


Figure 8: Linkages between FoodOn and ChEBI in FOBI

FOBI, launched in 2020, is aimed at describing the relationships between foods and food metabolome, that is, the collection of all metabolites in the body directly derived from the digestion and biotransformation of foods and their constituents. FOBI is composed of two interconnected branches: a “Foods” branch consisting of raw foods and multi-component foods; and a “Biomarkers” branch containing food intake biomarkers classified by their chemical classes. The food branch is composed mainly of FoodOn terms, while the biomarker branch is composed of both ChEBI terms and FOBI specific terms. At the moment, FOBI has a total of 1197 terms, containing 590 food biomarkers connected by a “BiomarkerOf” object property to 29 foods adopted from FoodOn, such as “[cacao food product](#)”.

2.9. FIDEO: Food Interactions with Drugs Evidence Ontology

The first version of FIDEO [32], released in 2020, represents interactions between foods and food supplements and drugs. Supporting evidence is equally represented to allow medical professionals to assess clinical significance of interactions. FoodOn terms and food categories are reused whenever possible, but other domain-specific food categories are locally defined. While initial efforts were focused on the design of the ontology based on the Basic Formal Ontology (BFO) and the OBO Foundry principles, more recent efforts are focused on a user-friendly visual interface that allows search and exploration of interactions [33] and on integrating food-drug interactions from various sources including compendia and existing databases to FIDEO using ROBOT [34].

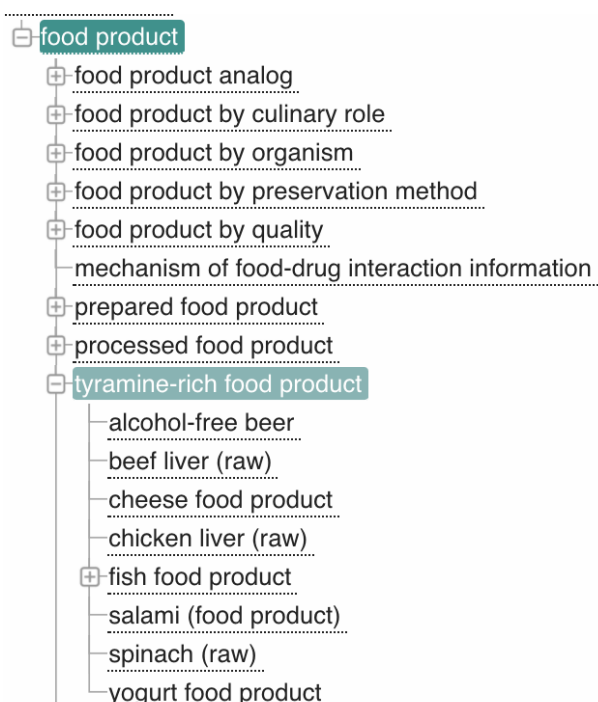


Figure 9: Food-Drug interactions example hierarchy in FIDEO

2.10. ONS: Ontology for Nutritional Studies

Since its first publication in 2018, the ONS has committed to describe nutritional studies in their multifaceted nature. A central concept in ONS is ‘diet’ [ONS:1000001], an ‘information content entity’ [IAO:0000030] defined as “the sum of food consumed by a person or other organism”. The diet concept is closely related to ‘dietary pattern’ [ONS:0000094], defined as “the quantity, proportion, variety and combination of different foods and drinks consumed in meals, and the frequency with which they are habitually consumed”. In ONS conceptualization, ‘dietary pattern’ denotes ‘diet’. Dietary pattern is intended to represent a ‘data item’ [IAO:0000027] typically resulting from assays in the context of nutritional epidemiology (i.e. ‘Food Frequency Questionnaire’ [ONE:0000007]), containing a specification of foods consumed and, as a result, denoting the type of diet to which a subject has adhered.

Curation and development revolving around the initial diet concept in ONS has greatly benefited from the collaboration with the Joint Food Ontology Workgroup. Thanks to this interaction, multiple different subclasses of the diet (and related dietary pattern) were defined. The annotation on food classes inclusion or exclusion for the various flavours of diet (and related dietary pattern) rely completely on the import and use of classes from FoodOn, at different granularity levels. As an example, the 'vegan diet' [ONS:1000021] and the 'vegan dietary pattern' [ONS:2000021] would be both annotated as characterized by eating vegetables ['eats' some 'plant food product'; RO:0002470 some FOODON:00001015] and by excluding the consumption of animal products [not('eats' some 'vertebrate animal food product'; not(RO:0002470 some FOODON:00001092)] (Figure 11).

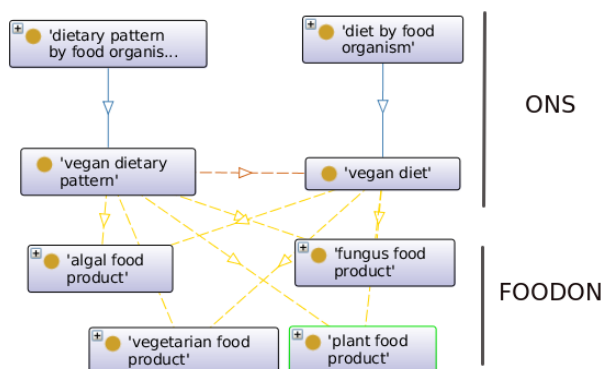


Figure 10: Modeling vegan diet and food products across ONS and FoodOn

2.11. ONE: Ontology for Nutritional Epidemiology

The ONE [19] details nutritional epidemiology manuscript and dataset characteristics. The ONE is structured according to a set of minimal requirements for the reporting of nutritional epidemiology research ([35]). The first version of ONE extends IAO document parts so they can cover research paper structure, and as well description of food surveys which form the underlying datasets for many studies. Abstract, discussion, ethics, methods, results and supplementary methods specific to dietary studies are defined. Classes of quality assessments of study designs and dietary recall methods are also defined (Figure 12).

vegan dietary pattern — ONS:2000021 — http:

Annotations Usage

Annotations: vegan dietary pattern

Annotations +

label [language: en]
vegan dietary pattern

DEFINITION [language: en]
A dietary pattern that define the quantity, variety or combination of different food in a vegan diet

'database cross reference'
<https://www.gs1.org/voc/DietTypeCode-VEGAN>

Description: vegan dietary pattern

Equivalent To +

SubClass Of +

- 'dietary pattern by food organism'
- eats some 'algal food product'
- eats some 'fungus food product'
- eats some 'plant food product'
- eats some 'vegetarian food product'
- not (eats some 'invertebrate animal food product')
- not (eats some 'part of animal')
- not (eats some 'seafood product')
- not (eats some 'vertebrate animal food product')

Figure 11: Example entry for vegan dietary pattern

The ONE was used previously to assess reporting completeness of manuscripts that present findings of nutritional epidemiology [36].

Ontology classes regarding food-based dietary guidelines (FBDGs) were added into ONE in 2021 to illustrate its potential applications for population dietary recommendations. Food-based dietary guidelines represent a wealth of accumulated diet knowledge summarized from nutrition studies [37,38].

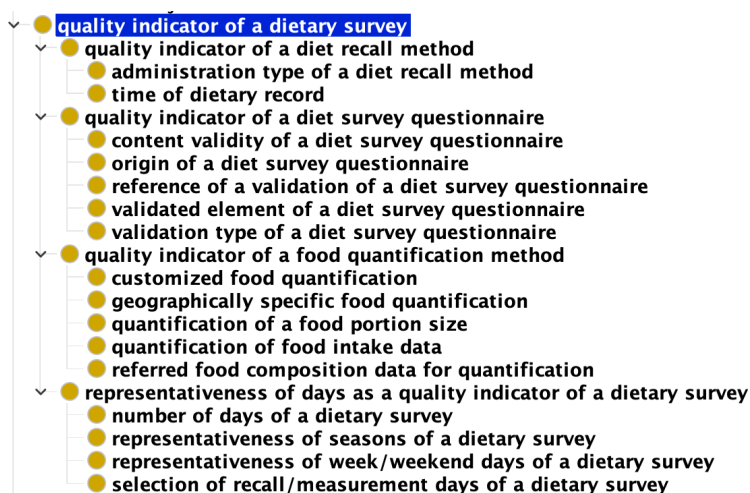


Figure 12: Hierarchy of a dietary survey quality indicators in ONE

FBDGs are important documents for policy makers, healthcare workers and educators, etc. to guide the general public, in order to help them build healthier eating habits. Applications of ontology will unlock information contained in the guidelines for automated modelling of trends to assess dietary habits.

The ONE curators are currently developing a Natural Language Processing (NLP)-SPARQL linkage to enable a natural language query of ONE, as well as dashboard development to visualize nutritional knowledge contained in research manuscripts and population based recommendations.

2.12. PO²: Process and Observation Ontology

Based partly on the Sensor, Observation, Sample, and Actuator (SOSA) ontology [39], but also situated within the BFO hierarchy, and drawing upon OWL-Time, QUDT and IAO, the PO² ontology is designed to monitor industrial food processing, and describe food formulation. PO² can represent a food transformation process described by a set of experimental observations available at different scales and evolving in time through the different unit operations of a production process. The ontology contains a core layer dedicated to the generic modeling of both transformation and characterization processes, while domain specific sub-ontologies specialize the PO² core model for different projects. It has been implemented in databases covering dairy, meat, and biorefinery production, to represent the unique characteristics of foods during their manufacturing process. PO² has been tested in fit-for-purpose software for maintaining these databases. PO² curators participate in the Food Process Ontology Workgroup which is helping to build FoodOn's food transformation process branch. The ontology is available at <http://agroportal.lirmm.fr/ontologies/PO2>.

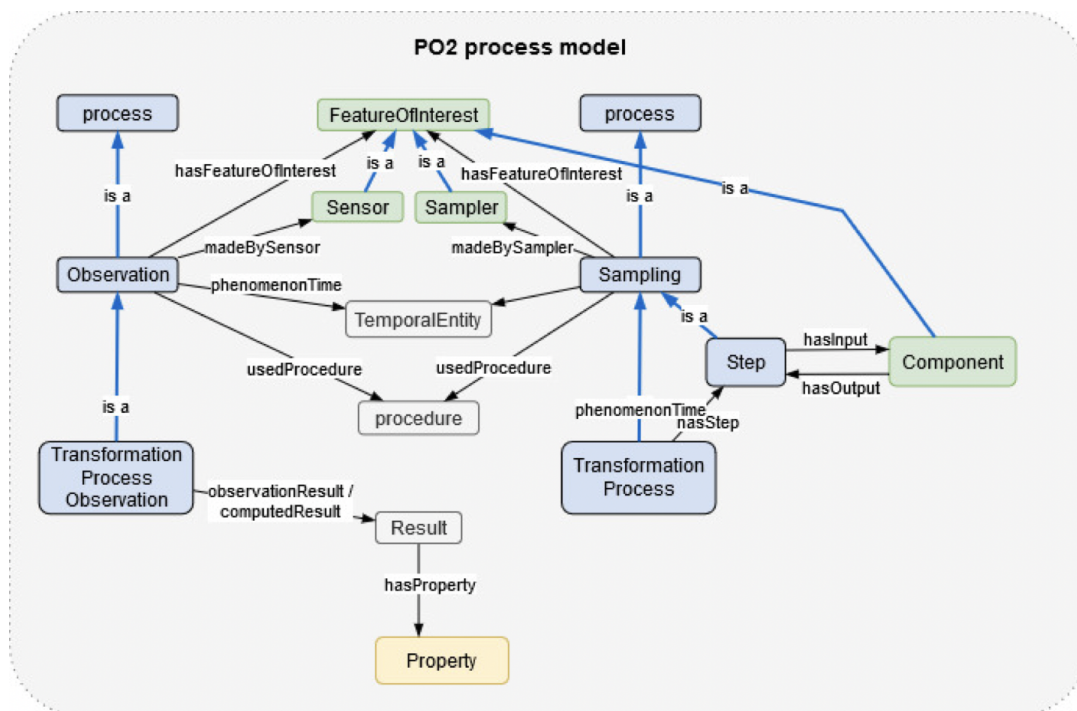


Figure 13: PO² uses similar components for both production line sampling and observation.

3. Discussion

The development of the OBO food related ontologies is occurring in a semi-autonomous parallel fashion, with interconnectivity issues arising on a weekly basis, and with the need to train new talent to help curate the growing volume of what is at the basic level a catch-up exercise to standardize and digitize vocabulary being used in research, policy and industry across the food spectrum. Knowledge is beginning to accrue within these OBO ontologies as they express at a class level the subject predicate object facts or assertions provided by the collective language that OBO can provide. These are much like “nanopublications” [40] that should be combinable to build larger and larger knowledge graphs about food. The FoodKG project demonstrates that factoid (“How much fat is in butter?”), comparison (“Which has more fat, butter or olive oil?”) and constraint (“Which dish has chicken, onion, and garlic?”) competency questions [41] can be applied to a knowledge graph composed of an import of triples from different sources. An overlap between ontology and knowledge base often occurs - axioms at the class level are expressing knowledge about class behaviour that instances conform to and inherit properties from. For example, FoodOn and DOID ontologies together hold the fact that *Apium graveolens* is the key organism of “celery food product” which “is allergic trigger of” “[celery allergy](#)”. Future additions of codified food allergies will enable hypothesis development and class-level associations that do not currently exist. For instance, potential cross allergenicities may be found to other foods originating from other organisms in the same taxonomic group, or grown/processed with similar methods. For example Figure 14 shows how parsley, dill and celery allergies each arise from different food products, yet each are also derived from organisms contained within the same sub-family, Tribe-level taxonomic hierarchy. Based on this information, one could hypothesize potential allergies for consumers who are allergic to any one of these products. Using the

same logic, one can imagine similar hypotheses being generated about foods that have been produced or processed with similar classes of exogenous chemicals.

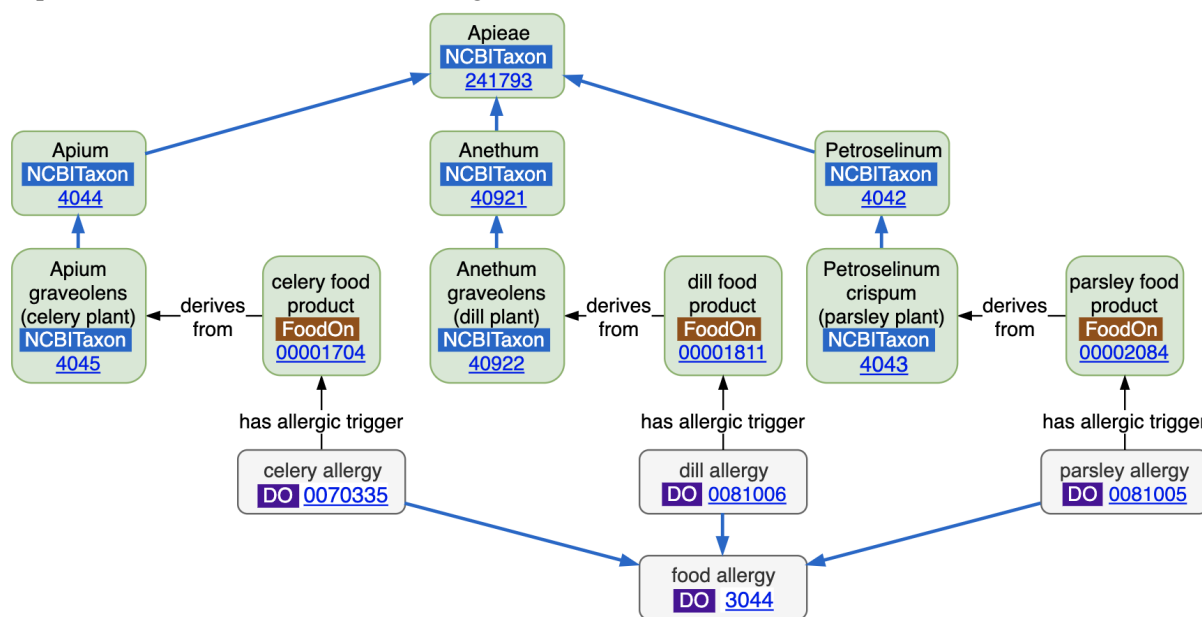


Figure 14: Advantage of class hierarchies across food related ontologies

The relative recency of the OWL standard has contributed to methodological growing pains stemming from its roots in formal logic and philosophy which are not easy to comprehend in terms of capability, computability, and ontology and database infrastructure especially for implementing term reuse and deprecation functionality. A new Core Ontology for Biology and Biomedicine (COB) is being lauded as a simpler development starting point that avoids a number of abstract BFO terms and contains many commonly reused OBO terms [42]. Tools like robot templates are enabling term curation of some kinds to be separated from the more intricate steps required for curation in a multiuser environment (see <https://foodon.org/design/robot-managed-vocabularies/>). As well, OBO is developing guidelines for multilingual labelling, and providing domain specific labels for terms shared between ontologies - via exact synonyms that are marked as belonging to a particular subset of ontology, which for example CDNO can use to provide nutritionist-friendly labels on ChEBI chemicals. Longstanding issues about the presentation of Vitamins as chemicals or as roles have recently been resolved between members of the Joint Food Ontology Workgroup and ChEBI; as well JFOW member dialogue has clarified the domain coverage and scope between ONE and ONS.

A larger challenge is to see how OBO might evolve to support domains that are outside of the life sciences. OBO would seem to have a natural counterpart in the Industrial Ontology Foundation (IOF), launched in 2016 (and now part of [OAGi](https://www.oagi.org/)), but as IOF is yet to launch its core ontology (scheduled for late summer 2021) there is not yet an ability to assess IOF ontology domains to trade terms with. Consequently ENVO remains the home for “manufactured product”, and a large influx of food industry and other equipment useful for laboratory and food process modelling, and biosample site description, will be listed under that class.

OBO-allied ontologies and SKOS or basic RDF based vocabularies in business operations or strategy and sustainability policy sectors can potentially add value together in a knowledge graph. Some, like OWL-Time fit well under BFO zero- and one-dimensional temporal regions. Others like the GS1 Global Product Classification for Food/Beverage/Tobacco [43] require much more work to link to OBO food related ontologies for describing nutrients, allergens, ingredients and serving sizes for example. This is critical for supporting traceability requirements of many blockchain projects

under development since GS1 standardized vocabulary is pervasive in business. The W3C PROV-O [44] provenance ontology is often resorted to for process modelling, but fails to satisfy various functional needs, and so the Food Process Ontology Workgroup is determining what extra functionality OBO needs - for equipment description, operating conditions, and extensions to PATO or FoodOn food characteristics - to cover this space.

On the horizon, the European Food and Safety Administration (EFSA) will be creating a pan-european food composition database presumably supported by the EFSA FoodEx2 vocabulary thesaurus, and this will need a gateway to OBO and other ontology based knowledge by way of terminology mapping. At first glance agency in-house food-related vocabularies would seem to benefit from conversion into globally accessible vocabularies that OBO contains, but technical, resource development, trust, versioning and regulatory issues suggest a more complex incremental alignment will occur. Some organizations may prefer a cautious medium-term use of open-source ontologies as a lingua franca hub of data exchange vocabulary with external partners.

4. Conclusion

Taken together, the above work represents an explosion of knowledge and data harmonization about food systems within OBO Foundry, and is emerging as the language for a federated database model. The successful reuse of terms is demonstrated, and the methodology of inter-agency curation points the way to faster de-facto standardization of vocabulary. The semantic web way of thinking about vocabulary through OWL ontologies that easily generalize and specialize about food related data in the world is proving to be a success in managing the complexity of life science knowledge, and a promising model for describing activity in health and sustainability policy and business domains as well.

5. Acknowledgements

This work is primarily supported by the USDA Non-Assistance Cooperative Agreement 58-8040-8-014-F and Genome Canada Grant 286GET to W. Hsiao.

6. References

- [1] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, OBI Consortium, Leontis N, Rocca-Serra P, Ruttenberg A, Sansone S-A, Scheuermann RH, Shah N, Whetzel PL, Lewis S. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat Biotechnol.* 2007 Nov;25(11):1251–5.
- [2] Jackson RC, Matentzoglou N, Overton JA, Vita R, Balhoff JP, Buttigieg PL, Carbon S, Courtot M, Diehl AD, Dooley D, Duncan W, Harris NL, Haendel MA, Lewis SE, Natale DA, Osumi-Sutherland D, Ruttenberg A, Schriml LM, Smith B, Stoeckert CJ, Vasilevsky NA, Walls RL, Zheng J, Mungall CJ, Peters B. OBO Foundry in 2021: Operationalizing Open Data Principles to Evaluate Ontologies [Internet]. *bioRxiv.* 2021 [cited 2021 Jun 8]. p. 2021.06.01.446587. Available from: <https://www.biorxiv.org/content/10.1101/2021.06.01.446587v1>
- [3] Cimino JJ. Desiderata for controlled medical vocabularies in the twenty-first century. *Methods Inf Med.* 1998 Nov;37(4-5):394–403.
- [4] Lange MC, Lemay DG, German JB. A multi-ontology framework to guide agriculture and food towards diet and health. *J Sci Food Agric.* 2007;87(8):1427–34.
- [5] Zeb A, Soinenen J-P, Sozer N. Data harmonisation as a key to enable digitalisation of the food sector: A review. *Food Bioprod Process.* 2021 May 1;127:360–70.
- [6] Avma. One Health: A new professional imperative [Internet]. American Veterinary Medical Association Schaumburg, IL; 2008. Available from: https://www.avma.org/sites/default/files/resources/onehealth_final.pdf

- [7] Dooley DM, Griffiths EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Lange MC, Schriml LM, Brinkman FSL, Hsiao WWL. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *NPJ Sci Food*. 2018 Dec 18;2:23.
- [8] He Y, Xiang Z, Zheng J, Lin Y, Overton JA, Ong E. The eXtensible ontology development (XOD) principles and tool implementation to support ontology interoperability. *J Biomed Semantics*. 2018 Jan 12;9(1):3.
- [9] Aubert C, Buttigieg PL, Laporte MA, Devare M, Arnaud E. CGIAR Agronomy Ontology [Internet]. CGIAR Agronomy Ontology. 2017 [cited 2017 Oct 17]. Available from: <https://github.com/AgriculturalSemantics/agro>
- [10] de Matos P, Alcántara R, Dekker A, Ennis M, Hastings J, Haug K, Spiteri I, Turner S, Steinbeck C. Chemical Entities of Biological Interest: an update. *Nucleic Acids Res*. 2010 Jan;38(Database issue):D249–54.
- [11] Andrés-Hernández L, Baten A, Azman Halimi R, Walls R, King GJ. Knowledge representation and data sharing to unlock crop variation for nutritional food security. *Crop Sci*. 2020 Mar 16;60(2):516–29.
- [12] Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, Felix V, Jeng L, Bearer C, Lichenstein R, Bisordi K, Campion N, Hyman B, Kurland D, Oates CP, Kibbey S, Sreekumar P, Le C, Giglio M, Greene C. Human Disease Ontology 2018 update: classification, content and workflow expansion. *Nucleic Acids Res* [Internet]. 2018 Nov 8; Available from: <http://dx.doi.org/10.1093/nar/gky1032>
- [13] MAxO [Internet]. Medical Action Ontology (MAxO). [cited 2021 Jun 24]. Available from: <https://github.com/monarch-initiative/MAxO>
- [14] Filter M, Buschhardt T, Dórea F, Lopez de Abechuco E, Günther T, Sundermann EM, Gethmann J, Dups-Bergmann J, Lagesen K, Ellis-Iversen J. One Health Surveillance Codex: promoting the adoption of One Health solutions within and across European countries. *One Health*. 2021 Jun;12:100233.
- [15] Thessen AE, Grondin CJ, Kulkarni RD, Brander S, Truong L, Vasilevsky NA, Callahan TJ, Chan LE, Westra B, Willis M, Rothenberg SE, Jarabek AM, Burgoon L, Korrick SA, Haendel MA. Community Approaches for Integrating Environmental Exposures into Human Models of Disease. *Environ Health Perspect*. 2020 Dec;128(12):125002.
- [16] Castellano-Escuder P, González-Domínguez R, Wishart DS, Andrés-Lacueva C, Sánchez-Pla A. FOBI: an ontology to represent food intake data and associate it with metabolomic data. *Database* [Internet]. 2020 Jan 1;2020. Available from: <http://dx.doi.org/10.1093/databa/baaa033>
- [17] Bordea G, Nikiema J, Griffier R, Hamon T, Mougín F. FIDEO: Food Interactions with Drugs Evidence Ontology. In: 11th International Conference on Biomedical Ontologies [Internet]. 2020. Available from: <https://hal.archives-ouvertes.fr/hal-03185166/>
- [18] Vitali F, Lombardo R, Rivero D, Mattivi F, Franceschi P, Bordoni A, Trimigno A, Capozzi F, Felici G, Taglino F, Miglietta F, De Cock N, Lachat C, De Baets B, De Tré G, Pinart M, Nimptsch K, Pischon T, Bouwman J, Cavalieri D, ENPADASI consortium. ONS: an ontology for a standardized description of interventions and observational studies in nutrition. *Genes Nutr*. 2018 Apr 30;13:12.
- [19] Yang C, Ambayo H, Baets BD, Kolsteren P, Thanintorn N, Hawwash D, Bouwman J, Bronselaer A, Pattyn F, Lachat C. An Ontology to Standardize Research Output of Nutritional Epidemiology: From Paper-Based Standards to Linked Content. *Nutrients* [Internet]. 2019 Jun 8;11(6). Available from: <http://dx.doi.org/10.3390/nu11061300>
- [20] Ibanescu L, Dibie J, Dervaux S, Guichard E, Raad J. PO² - A Process and Observation Ontology in Food Science. Application to Dairy Gels. In: *Metadata and Semantics Research*. Springer International Publishing; 2016. p. 155–65.
- [21] Xiang Z, Mungall C, Ruttenberg A, He Y. Ontobee: A linked data server and browser for ontology terms. In: ICBO [Internet]. 2011. Available from: <https://www.academia.edu/download/39270744/0deec5393d9ae86155000000.pdf>
- [22] OBO Principles [Internet]. OBO Foundry. [cited 2021 Jun 24]. Available from: <http://www.obofoundry.org/principles/fp-000-summary.html>
- [23] Arp R, Smith B, Spear AD. *Building Ontologies with Basic Formal Ontology*. MIT Press; 2015. 248 p.
- [24] JFOW [Internet]. Joint Food Ontology Workgroup. [cited 2021 Jun 24]. Available from: <https://github.com/FoodOntology/joint-food-ontology-wg>
- [25] Buttigieg PL, Morrison N, Smith B, Mungall CJ, Lewis SE, ENVO Consortium. The environment ontology: contextualising biological and biomedical entities. *J Biomed Semantics*. 2013 Dec 11;4(1):43.
- [26] Ireland JD, Møller A. LanguaL food description: a learning process. *Eur J Clin Nutr*. 2010 Nov;64 Suppl 3:S44–8.
- [27] Grim CJ, Windsor AM, Kocurek B, Leonard SR, Richter TKS, Gopinath G, Balkey M, Ramachandran P, Ottesen A, Jarvis K, Timme R. Development of a MIxS (Minimum Information about any (x) Sequence) Food Environmental Metadata Standard. In 2020 [cited 2021 Jun 22]. Available from: https://github.com/FoodOntology/joint-food-ontology-wg/blob/master/presentation/IFOW_2020_sept_30_i

- n_the_agency/IFOW_2020_3_Development_of_a_MiX_S_Food_Environmental_Metadata_Standard.pdf
- [28] Murphy SP, Charrondiere UR, Burlingame B. Thirty years of progress in harmonizing and compiling food data as a result of the establishment of INFOODS. *Food Chem.* 2016 Feb 15;193:2–5.
- [29] Vita R, Mahajan S, Overton JA, Dhanda SK, Martini S, Cantrell JR, Wheeler DK, Sette A, Peters B. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.* 2019 Jan 8;47(D1):D339–43.
- [30] Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, Clancy K, Courtot M, Derom D, Dumontier M, Fan L, Fostel J, Fragoso G, Gibson F, Gonzalez-Beltran A, Haendel MA, He Y, Heiskanen M, Hernandez-Boussard T, Jensen M, Lin Y, Lister AL, Lord P, Malone J, Manduchi E, McGee M, Morrison N, Overton JA, Parkinson H, Peters B, Rocca-Serra P, Ruttenberg A, Sansone S-A, Scheuermann RH, Schober D, Smith B, Soldatova LN, Stoeckert CJ Jr, Taylor CF, Torniai C, Turner JA, Vita R, Whetzel PL, Zheng J. The Ontology for Biomedical Investigations. *PLoS One.* 2016 Apr 29;11(4):e0154556.
- [31] Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. Dead simple OWL design patterns. *J Biomed Semantics.* 2017 Jun 5;8(1):18.
- [32] FIDEO [Internet]. Repository for the Food Interactions with Drugs Evidence Ontology (FIDEO). 2018 [cited 2021 Jun 24]. Available from: <https://github.u-bordeaux.fr/erias/fideo>
- [33] Lalanne F, Bedouch P, Simonnet C, Depras V, Bordea G, Bourqui R, Hamon T, Thiessard F, Mougin F. Visualizing Food-Drug Interactions in the Thériaque Database. *Stud Health Technol Inform.* 2021 May 27;281:253–7.
- [34] Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA. ROBOT: A Tool for Automating Ontology Workflows. *BMC Bioinformatics.* 2019 Jul 29;20(1):407.
- [35] Lachat C, Hawwash D, Ocké MC, Berg C, Forsum E, Hörnell A, Larsson C, Sonestedt E, Wirfält E, Åkesson A, Kolsteren P, Byrnes G, De Keyzer W, Van Camp J, Cade JE, Slimani N, Cevallos M, Egger M, Huybrechts I. Strengthening the Reporting of Observational Studies in Epidemiology—Nutritional Epidemiology (STROBE-nut): An Extension of the STROBE Statement [Internet]. Vol. 13, *PLOS Medicine.* 2016. p. e1002036. Available from: <http://dx.doi.org/10.1371/journal.pmed.1002036>
- [36] Yang C, Hawwash D, De Baets B, Bouwman J, Lachat C. Perspective: Towards Automated Tracking of Content and Evidence Appraisal of Nutrition Research. *Adv Nutr.* 2020 Jun 6;11(5):1079–88.
- [37] Teicholz N. The scientific report guiding the US dietary guidelines: is it scientific? *BMJ.* 2015 Sep 23;351:h4962.
- [38] Montagnese C, Santarpia L, Buonifacio M, Nardelli A, Caldara AR, Silvestri E, Contaldo F, Pasanisi F. European food-based dietary guidelines: a comparison and update. *Nutrition.* 2015 Jul;31(7-8):908–15.
- [39] Janowicz K, Haller A, Cox SJD, Le Phuoc D, Lefrançois M. SOSA: A lightweight ontology for sensors, observations, samples, and actuators. *Journal of Web Semantics.* 2019 May 1;56:1–10.
- [40] Mina E, Thompson M, Kaliyaperumal R, Zhao J, der Horst van E, Tatum Z, Hettne KM, Schultes EA, Mons B, Roos M. Nanopublications for exposing experimental data in the life-sciences: a Huntington’s Disease case study. *J Biomed Semantics.* 2015 Feb 9;6:5.
- [41] Haussmann S, Seneviratne O, Chen Y, Ne’eman Y, Codella J, Chen C-H, McGuinness DL, Zaki MJ. FoodKG: A Semantics-Driven Knowledge Graph for Food Recommendation [Internet]. *Lecture Notes in Computer Science.* 2019. p. 146–62. Available from: http://dx.doi.org/10.1007/978-3-030-30796-7_10
- [42] Core Ontology for Biology and Biomedicine (COB) [Internet]. OBOFoundry; [cited 2021 Jun 25]. Available from: <https://github.com/OBOFoundry/COB>
- [43] GS1 Food [Internet]. GS1 Global Product Classification for Food/Beverage/Tobacco. 2018 [cited 2021 Jun 25]. Available from: <https://www.gs1.org/voc/FoodBeverageTobaccoProduct>
- [44] Lebo T, Sahoo S, McGuinness D, Belhajjame K, Cheney J, Corsar D, Garijo D, Soiland-Reyes S, Zednik S, Zhao J. PROV-O [Internet]. PROV-O: The PROV Ontology. 2013, W3C Recommendation. [cited 2021 Jun 25]. Available from: <http://www.w3.org/TR/2013/REC-prov-o-20130430/>