

Title: FoodOn Evolution in the Food Ontology Ecosystem

Authors: Damion Dooley¹, Emma Griffiths¹, Rhiannon Cameron¹, Gurinder Gosal¹, Tarini Naravane², Ramona L. Walls³, Pier Buttigieg⁴, William Hsiao^{1,2}.

¹ Department of Pathology and Laboratory Medicine, University of British Columbia, Vancouver, BC, Canada.

² Biological and Agricultural Engineering, University of California , Davis, CA, USA

³ Bio5 Institute, the University of Arizona, Tucson, AZ, USA

⁴ Max Planck Research School for Marine Microbiology, Jacobs University, Bremen, Germany

⁵ BC Public Health Microbiology & Reference Laboratory, Vancouver, BC, Canada

Abstract: FoodOn ¹, published in 2018, is a recently developed ontology that aims to describe food and animal feed known in cultures throughout the world for human, companion animal, and agricultural animal consumption. FoodOn specializes in the provision of food terms that name plant and animal food sources, ingredients derived from them, and common multi-component foods. FoodOn also inherits a professional food composition vocabulary that can be traced back to the “[Factored Food Vocabulary](#)” introduced in 1975 by the Center for Food Safety and Applied Nutrition (CFSAN) of the United States Food and Drug Administration, and later adopted by the European LanguaL™ Technical Committee which supports [EuroFIR](#) partner national food nutrition databases. All 14 LanguaL™ [facets](#) of food description (heat treatment, preservation, packaging etc.) have been transformed and mapped into FoodOn ontology branches of vocabulary.

Initial FoodOn development involved coordination with interrelated ontologies in the Open Biological and Biomedical Ontology Foundry ([OBOFoundry](#)) ², which is supported by ontology developers committed to collaboration and adherence to shared principles, including logical conformance to the Basic Formal Ontology ([BFO](#)) ³, and orthogonal reuse of existing ontology terms and relations ⁴. To describe food components and origins, FoodOn reuses many [CHEBI](#) chemical terms, [UBERON](#) animal anatomy and [PO](#) plant anatomy terms, [NCBITaxon](#) taxa, and Relation Ontology ([RO](#)) relations. As well, the FoodOn curation team is incrementally adding basic process models of commodities and single ingredient food products derived from plants or animals, for example, [beans](#).

There are many specialized domains of knowledge in the farm-to-fork (or field-to-compost) food journey and this is reflected in the past year of development and interdependence of FoodOn and other ontology projects such as the Food Biomarker Ontology ([FOBI](#)) ⁵, the Ontology for Nutritional Studies ([ONS](#)) ⁶, the Ontology for Nutritional Epidemiology ([ONE](#)) ⁷, INRAE’s [MEATyL@b](#) project, the Food Interactions with Drugs Evidence Ontology ([FIDEO](#)) and Crop Dietary Nutrition Ontology ([CDNO](#)) ⁸ projects. Representatives of these ontologies have established a Joint Food Ontology Workgroup (<https://github.com/FoodOntology/joint-food-ontology-wg>) which has become fertile ground for feedback regarding diet, nutrition and agriculture related terminology, and has

led to consensus on semantics, including the opportunity to synchronize on ontological analysis paradigms such as a process model oriented approach that upper level ontologies like BFO have not delved into. At a peer-to-peer level this occasionally makes for a more complicated curation arrangement, but bears fruit in the ease of comprehension and compatibility of controlled vocabulary across scientific domains.

The development of FoodOn was initially motivated by the need to create a standard for reporting food biosample metadata related to public health foodborne pathogen surveillance and investigations. FoodOn is currently being implemented in public health agency vocabularies, platforms and surveillance networks like [NARMS](#), [GenomeTrakr](#), and the [IFSAC](#) epidemiology-focused food classification scheme used for categorizing foods implicated in outbreaks. Many new food terms, their synonyms, and their ontology identifiers have already been added in order to capture GenomeTrakr food descriptions in public genomic sequence repositories such as NCBI. The US FDA is implementing FoodOn facets in a new Minimum Information about any (x) Sequence Food Environmental Metadata Standard (MlxS) extension for food related data. FoodOn also aims to create mappings to EFSA FoodEx2's exposure hierarchy vocabulary, and to UN FAO INFOODS, AGROVOC and ASFIS resources, to facilitate data harmonization and to enrich datasets with additional knowledge annotations.

To expand FoodOn content, three major initiatives are underway.

- Expanding into the domain of nutrition, facets of FoodOn are being introduced by the USDA Agricultural Research Service into the in-house version of their Food Data Central website (<https://fdc.nal.usda.gov/>). To support this effort in the next year, FoodOn is revising its class structure to distinguish between single-source and multi-source foods which originated from the LanguaL™ indexed SIREN product hierarchy of over 9,600 food items. Generally, all single-source foods will be positioned under FoodOn's "[food by organism](#)" branch. The remainder exist under a "multi-component food" category, which is not intended to be exhaustive of every ingredient combination encountered, but does aim to provide terms for generic (non-branded) foods like lasagne, and some food composition information about them to the extent that generalizations can be made. This revision will be complemented with a data model for food composition in recipes, accompanied by units provided by the unit ontology (UO), an expanded food process branch as well as respective "[food product by quality](#)" entities.
- A Wikipedia FoodOn integration project is underway to augment FoodOn entries that lack definitions or image references with content from matching Wikipedia entries. As well, a Wikipedia food term report from the FIDEO ontology project will help guide FoodOn curation efforts to increase Wikipedia term coverage.

- FoodOn had previously drawn a distinction between NCBITaxon entities and FoodOn organism entities, but understanding that instances of NCBITaxon entities can be considered whole organism references (regardless of their life stage), FoodOn will be deprecating those of its food source plant and animal organisms that are covered by NCBITaxon entries, and adding its existing annotations and axioms to the latter.

Working with partner ontologies and agencies will take FoodOn to the next level, as a linchpin food item vocabulary connecting agricultural and wild food chains to consumer diet, nutrition and other health and ecosystem knowledge being amassed by researchers worldwide.

References

1. Dooley, D. M. *et al.* FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. *NPJ Sci Food* **2**, 23 (2018).
2. Smith, B. *et al.* The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nat. Biotechnol.* **25**, 1251–1255 (2007).
3. Arp, R., Smith, B. & Spear, A. D. *Building Ontologies with Basic Formal Ontology*. (MIT Press, 2015).
4. Courtot, M. *et al.* MIREOT: the Minimum Information to Reference an External Ontology Term. *Nature Precedings* (2009) doi:10.1038/npre.2009.3576.1.
5. Castellano-Escuder, P., González-Domínguez, R., Wishart, D. S., Andrés-Lacueva, C. & Sánchez-Pla, A. FOBI: an ontology to represent food intake data and associate it with metabolomic data. *Database* **2020**, (2020).
6. Vitali, F. *et al.* ONS: an ontology for a standardized description of interventions and observational studies in nutrition. *Genes Nutr.* **13**, 12 (2018).
7. Yang, C. *et al.* An Ontology to Standardize Research Output of Nutritional Epidemiology: From Paper-Based Standards to Linked Content. *Nutrients* **11**, (2019).
8. Azman Halimi, R., Barkla, B. J., Andrés-Hernández, L., Mayes, S. & King, G. J. Bridging

the food security gap: an information-led approach to connect dietary nutrition, food composition and crop production. *J. Sci. Food Agric.* **100**, 1495–1504 (2020).