



# Establishing a Common Nutritional Vocabulary - From Food Production to Diet

Liliana Andrés-Hernández<sup>1</sup>, Kai Blumberg<sup>2</sup>, Ramona L. Walls<sup>2,3</sup>, Damion Dooley<sup>4</sup>, Ramil Mauleon<sup>1</sup>, Matthew Lange<sup>5</sup>, Magalie Weber<sup>6</sup>, Lauren Chan<sup>7</sup>, Adnan Malik<sup>8</sup>, Anders Møller<sup>9</sup>, Jayne Ireland<sup>9</sup>, Lucia Segovia<sup>10</sup>, Xuhuiqun Zhang<sup>11</sup>, Britt Burton-Freeman<sup>11</sup>, Paul Magelli<sup>12</sup>, Andrew Schriever<sup>12</sup>, Shawawn M. Forester<sup>13</sup>, Lei Liu<sup>1</sup> and Graham J. King<sup>1,14\*</sup>

<sup>1</sup> Southern Cross Plant Science, Southern Cross University, Lismore, NSW, Australia, <sup>2</sup> Department of Biosystems Engineering, University of Arizona, Tucson, AZ, United States, <sup>3</sup> Data Collaboration Center at the Critical Path Institute, Tucson, AZ, United States, <sup>4</sup> Faculty of Health Sciences, Simon Fraser University, Burnaby, BC, Canada, <sup>5</sup> International Center for Food Ontology Operability Data & Semantics (IC-FOODS), Davis, CA, United States, <sup>6</sup> INRAE, UR BIA, Nantes, France, <sup>7</sup> Nutrition Department, College of Public Health and Human Sciences, Oregon State University, Corvallis, OR, United States, <sup>8</sup> European Bioinformatics Institute, European Molecular Biology Laboratory (EMBL-EBI), Hinxton, United Kingdom, <sup>9</sup> Danish Food Informatics, Roskilde, Denmark, <sup>10</sup> London School of Hygiene and Tropical Medicine, University of London, London, United Kingdom, <sup>11</sup> Illinois Institute of Technology, Chicago, IL, United States, <sup>12</sup> WISEcode LLC, Reno, NV, United States, <sup>13</sup> Nutrient Institute LLC, a Non-profit, Reno, NV, United States, <sup>14</sup> School of Biosciences, University of Nottingham, Sutton Bonington, United Kingdom

## OPEN ACCESS

### Edited by:

Hayriye Sebnem Harsa,  
Izmir Institute of Technology, Turkey

### Reviewed by:

Dennis Mans,  
Anton de Kom University of  
Suriname, Suriname

### \*Correspondence:

Graham J. King  
graham.king@scu.edu.au

### Specialty section:

This article was submitted to  
Nutrition and Food Science  
Technology,  
a section of the journal  
Frontiers in Nutrition

Received: 26 April 2022

Accepted: 17 May 2022

Published: 21 June 2022

### Citation:

Andrés-Hernández L, Blumberg K, Walls RL, Dooley D, Mauleon R, Lange M, Weber M, Chan L, Malik A, Møller A, Ireland J, Segovia L, Zhang X, Burton-Freeman B, Magelli P, Schriever A, Forester SM, Liu L and King GJ (2022) Establishing a Common Nutritional Vocabulary - From Food Production to Diet. *Front. Nutr.* 9:928837. doi: 10.3389/fnut.2022.928837

Informed policy and decision-making for food systems, nutritional security, and global health would benefit from standardization and comparison of food composition data, spanning production to consumption. To address this challenge, we present a formal controlled vocabulary of terms, definitions, and relationships within the Compositional Dietary Nutrition Ontology (CDNO, [www.cdno.info](http://www.cdno.info)) that enables description of nutritional attributes for material entities contributing to the human diet. We demonstrate how ongoing community development of CDNO classes can harmonize trans-disciplinary approaches for describing nutritional components from food production to diet.

**Keywords:** dietary composition, food composition, ontologies, nutritional security, FAIR data, knowledge representation, human health

## INTRODUCTION

Food production and supply systems affect human nutrition and health in personalized and global contexts (1). However, nutrition-based decisions and data are seldom integrated along the production and supply chain. This information may affect selection of cultivars and conservation of genetic resources, the management of food supply, processing and distribution, and analysis of dietary consumption patterns segmented by various demographics (2). Although various conventions exist for naming individual chemicals and physical attributes of dietary components, comparison of data and feedback within food systems is often constrained by divergence in formal definitions and classifications (3). The exchange of knowledge and operational data between domains would benefit from a consistent framework that defines nutritional and phytochemical composition, as well as other attributes of food, including their dietary role and physiological function.

Knowledge representation underpins communication, and is particularly important for sharing complex data and information within and between diverse domains such as crop biodiversity, food supply, and nutrition (4). Defining and classifying commonly understood terminology facilitates data acquisition, exchange and interoperability, where formal systems of domain-specific controlled vocabularies such as ontologies contribute to the representation and sharing of complex knowledge (5). They do this by defining terms with human readable definitions alongside machine readable relationships that facilitate the annotation, exchange, analysis, and interpretation of data (6). Establishment of clearly defined ontology classes representing domain-specific terminology is the first step to building common platforms that are of practical value to data curators and to end-users searching for relevant information. An approachable lexical representation of objects or concepts from different perspectives, that also helps reduce ambiguities in terminology for non-specialists, is particularly important for describing datasets in food supply chains (7) (**Supplementary Figure 1**). For instance, nutritional composition may vary depending on factors such as cultivars, cultivation systems, processing variables, food storage and preparation. Moreover, there is a need to distinguish between individual chemical components and the method by which their concentration is determined. In many standard Food Composition Tables and Databases (FCTs/ FCDBs) such information is often conflated or absent (8).

The Open Biomedical and Biological Ontologies Foundry and Library (OBO) is responsible for the establishment and development of a wide range of formal vocabularies in the life-sciences and related domains (9). This includes the ontology for Chemical Entities of Biological Interest (ChEBI) (10), which provides a valuable resource for structured sets of chemical definitions. OBO principles emphasize the value of reusing terms (formally known as classes or properties) between ontologies. The development of the Compositional Dietary Nutrition Ontology (CDNO) (4) was prompted by the need to follow *Findable, Accessible, Interoperable, and Reusable* (FAIR) principles (11) of data sharing. CDNO was initially focused on vocabulary to describe nutritional components in plant-derived materials contributing to human diet, and particularly those that may vary according to crop variety or within genetic resource collections (2, 4). However, we found that the structured reusable definitions of nutritional components were equally applicable to a wide range of food raw materials derived from livestock, fish or any other organic or inorganic source described in the Food Ontology (FoodOn) (12) (**Figure 1B**).

## METHODS

While developing and expanding CDNO, we have followed the OBO principles (13), which emphasize community development of interoperable ontologies. We focused on reuse and import of existing OBO terms, as well as ensuring open discussion within the CDNO GitHub repository (14). In order to generate terms that are subclasses of CDNO ‘*dietary nutritional component*’ [CDNO:0200001], a modified version of the Crop Dietary

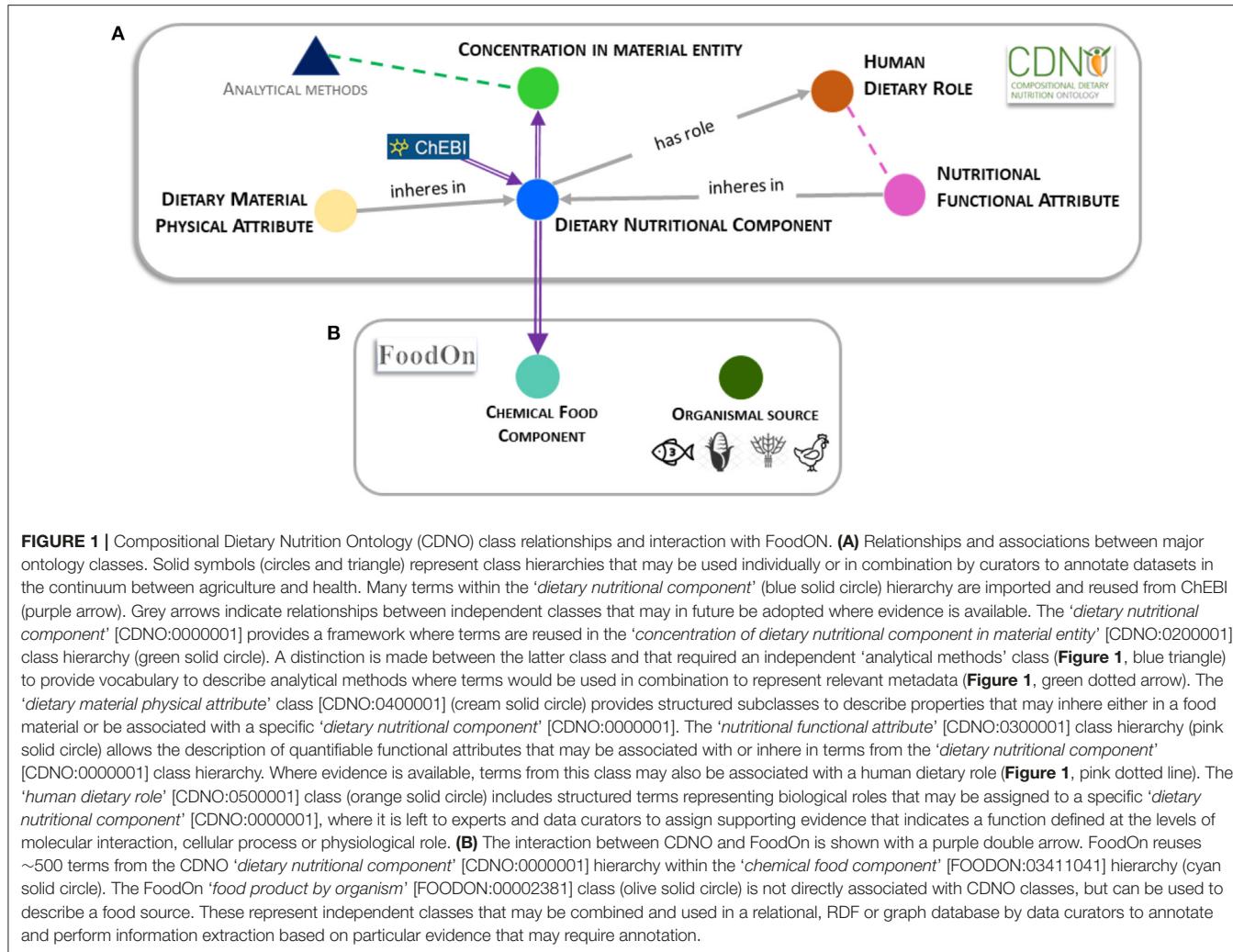
Nutrition Data Framework (CDN-DF) v.1.0 from Halimi et al. (15) was used, with definition and organization of additional terms arising from discussions with plant chemist domain specialists and curators from the International Network of Food Data Systems (INFOODS) collated by the Food and Agriculture Organization (FAO) (16), USDA FoodData Central (17), and the European Food Information Resource (EuroFIR) (18) food composition databases and repositories. The CDN-DF v.2.0 was used as an input for a Python script that parsed the CDN-DF\_v.2.0.xlsx into the nutritional\_components\_framework.csv and sugar\_derivatives.csv files, which were converted into input files for ROBOT templates. These templates were used to generate a revised organization of classes/terms compiled into the reference CDNO in Web Ontology Language (OWL) (19) file. Dietary nutritional components not present in the ChEBI were proposed and accepted as new entities using the ChEBI submission tool and imported into CDNO. The remaining terms that did not fit within the ChEBI scope were formally defined in CDNO, supported by reference to peer reviewed literature and authoritative online resources. These terms were described by following existing ontology definition guidelines for development of genus-differentia definitions (20). The class ‘*concentration of dietary nutritional component in material entity*’ [CDNO:0200001], as well as its subclasses were created using a Dead Simple OWL Design Pattern (DOS-DP) (21) modified from The Environment Ontology (ENVO) (22, 23). The DOS-DP combined terms from the Phenotype and Trait Ontology (PATO) (24), CDNO and the Basic Formal Ontology (BFO) (25) with OWL equivalence axioms. The remaining major classes were proposed and discussed via the CDNO GitHub issue forum (14) and in online workshops and seminars.

The CDNO ontology and accompanying code was initially created using the Ontology-Development-Kit (ODK) (26), and later versions of CDNO were developed using the templates module from the ROBOT software (27). The reference CDNO OWL file and the source code are available from Github CDNO repository (14). Additional database tables were added to the core CropStoreDB MySQL schema (28) to manage different nutritional data sources, along with an ‘ontology register’ lookup table to CDNO, FoodOn, ChEBI, NCBI taxon (29) and Plant Ontology (PO) (30) terms.

## RESULTS AND DISCUSSION

CDNO is registered as part of the OBO Foundry with terms and definitions searchable via Ontobee (31) or the Ontology Lookup Service (OLS) (32). Version 2.2 of the CDNO comprises five top level classes ‘*dietary nutritional component*’ [CDNO:0000001], ‘*concentration of dietary nutritional component in material entity*’ [CDNO:0200001], ‘*nutritional functional attribute*’ [CDNO:0300001], ‘*dietary material physical attribute*’ [CDNO:0400001], and ‘*human dietary role*’ [CDNO:0500001] (**Figure 1A**).

While dietary nutrients within food substrates are often present as complex and dynamic physical and chemical structures or mixtures, food labelling and FCTs/ FCDBs typically



**FIGURE 1 |** Compositional Dietary Nutrition Ontology (CDNO) class relationships and interaction with FoodON. **(A)** Relationships and associations between major ontology classes. Solid symbols (circles and triangle) represent class hierarchies that may be used individually or in combination by curators to annotate datasets in the continuum between agriculture and health. Many terms within the '*dietary nutritional component*' (blue solid circle) hierarchy are imported and reused from ChEBI (purple arrow). Grey arrows indicate relationships between independent classes that may in future be adopted where evidence is available. The '*dietary nutritional component*' [CDNO:0000001] provides a framework where terms are reused in the '*concentration of dietary nutritional component in material entity*' [CDNO:0200001] class hierarchy (green solid circle). A distinction is made between the latter class and that required an independent '*analytical methods*' class (**Figure 1**, blue triangle) to provide vocabulary to describe analytical methods where terms would be used in combination to represent relevant metadata (**Figure 1**, green dotted arrow). The '*dietary material physical attribute*' class [CDNO:0400001] (cream solid circle) provides structured subclasses to describe properties that may inhere either in a food material or be associated with a specific '*dietary nutritional component*' [CDNO:0000001]. The '*nutritional functional attribute*' [CDNO:0300001] class hierarchy (pink solid circle) allows the description of quantifiable functional attributes that may be associated with or inhere in terms from the '*dietary nutritional component*' [CDNO:0000001] class hierarchy. Where evidence is available, terms from this class may also be associated with a human dietary role (**Figure 1**, pink dotted line). The '*human dietary role*' [CDNO:0500001] class (orange solid circle) includes structured terms representing biological roles that may be assigned to a specific '*dietary nutritional component*' [CDNO:0000001], where it is left to experts and data curators to assign supporting evidence that indicates a function defined at the levels of molecular interaction, cellular process or physiological role. **(B)** The interaction between CDNO and FoodOn is shown with a purple double arrow. FoodOn reuses ~500 terms from the CDNO '*dietary nutritional component*' [CDNO:0000001] hierarchy within the '*chemical food component*' [FOODON:03411041] hierarchy (cyan solid circle). The FoodOn '*food product by organism*' [FOODON:00002381] class (olive solid circle) is not directly associated with CDNO classes, but can be used to describe a food source. These represent independent classes that may be combined and used in a relational, RDF or graph database by data curators to annotate and perform information extraction based on particular evidence that may require annotation.

represent proximate and individual chemical components, alongside properties such as energy. Within CDNO, the primary '*dietary nutritional component*' [CDNO:0000001] class is formally defined as: "A material entity taken in by an organism that contributes to the survival, growth, development, or other biological function of itself, its biota, or its holobionts." This class is structured with 10 subclasses corresponding to the major commonly used proximate classifications of chemical food composition, such as proteins, carbohydrates, and vitamins. The component terms are organized in an acyclic hierarchical structure with up to five additional levels. At present (v3.1) this class includes over 685 chemical terms, including 29 unique CDNO nutritional chemical components and their definitions. Within FoodOn, the '*dietary nutritional component*' [CDNO:0000001] class structure has been imported as a subclass of '*chemical food component*' [FOODON:03411041] class (**Figure 1B**). A similar hierarchical classification of dietary nutritional components that lacked ontological relationships and definitions had previously been proposed by the EuroFIR project (18, 33). The current versions of EuroFIR thesauri are available online

(34). This was shared following exchange of the original CDNO framework.

Ensuring interoperability of terms defined within CDNO, along with their labels and synonyms, requires ongoing consultation with a range of specialists from different domains. The '*dietary nutritional component*' [CDNO:0200001] class imports many terms defined within ChEBI, where relationships are primarily determined by formal chemical classifications. However, we were keen to establish a hierarchy that focuses on and accommodates terms organized according to sub-categories recognized by nutritionists and different domain experts such as food scientists and chemists. We generated and defined subclasses as required, and included synonyms used in different English-speaking countries. As an example, the term '*available carbohydrate*' [CDNO:0000003] has the synonym "digestible carbohydrate" according to Englyst et al. (35), but should not be confused with the term "total carbohydrate" used in some food tables. The latter term is used in the USDA FCT to refer to a specific method used for carbohydrate determination, calculated by subtraction of the sum of the crude protein, total fat, moisture, and ash from the total weight of the food (36).

In order to accommodate such conceptual discrepancies and reduce ambiguity, the term '*concentration of carbohydrate in material entity*' [CDNO:0200005] can be used to refer to total carbohydrate, without making any assumption as to a specific type of carbohydrates.

## Context and Use of Major Classes

A major intended use of the '*dietary nutritional component*' [CDNO:0000001] class is to harmonize the annotation and exchange of dietary composition datasets from a diverse range of sources that quantify concentration of chemical nutritional components (37) (Figure 1; Supplementary Figure 2). These may include data generated by analytical laboratories for production, reference and research (Figure 1A), as well as derived from existing FCTs/ FCDBs or food labelling. Such data may also be used when evaluating evidence in relation to dietary role. The '*concentration of dietary nutritional component in material entity*' [CDNO:0200001] class is formally defined as: "The concentration of dietary nutritional component when measured in some material entity". In addition, the '*dietary material physical attribute*' [CDNO:0400001] class is defined as: "A physical property that inheres in a food material or one or more dietary nutritional components." This enables a formal distinction to be made between chemical components and physical properties (or qualities) such as "potential energy" that may inhere in a food material. At present, in most FCTs/ FCDBs the tag for "energy" appears equivalent to or alongside chemical components such as sugars (Figure 1A).

We make an important distinction between the class '*concentration of dietary nutritional component in material entity*' [CDNO:0200001] and terms used to describe the analytical method, by which a specific concentration is established. Diverse methods and units of measurement are associated with quantitative data in research literature, for supply chain quality assurance and control, or to inform Food Composition Tables (FCT) and labelling (Figure 1B). This requires appropriate vocabulary (potentially in an independent '*analytical methods*' class) to describe the distinct steps in the process to quantify concentration (Figure 1A), including methodologies and protocols used for sampling, extraction, and analysis that may re-use terms from existing ontologies such as the Chemical Method Ontology (CHMO) (38) and the Ontology for Biomedical Investigations (OBI) (39). Many (FCTs/ FCDBs) tags conflate methods used and nutritional components measured, such as INFOODs tag names [GLYCERA] defined as "glycerides, total; determined by analysis," [LACSM] as "lactose; expressed in monosaccharide equivalents", or [CHOT] "carbohydrate, total; calculated by summation." In future such tags could be annotated with a combination of *nutritional component* and *analytical method* term IDs.

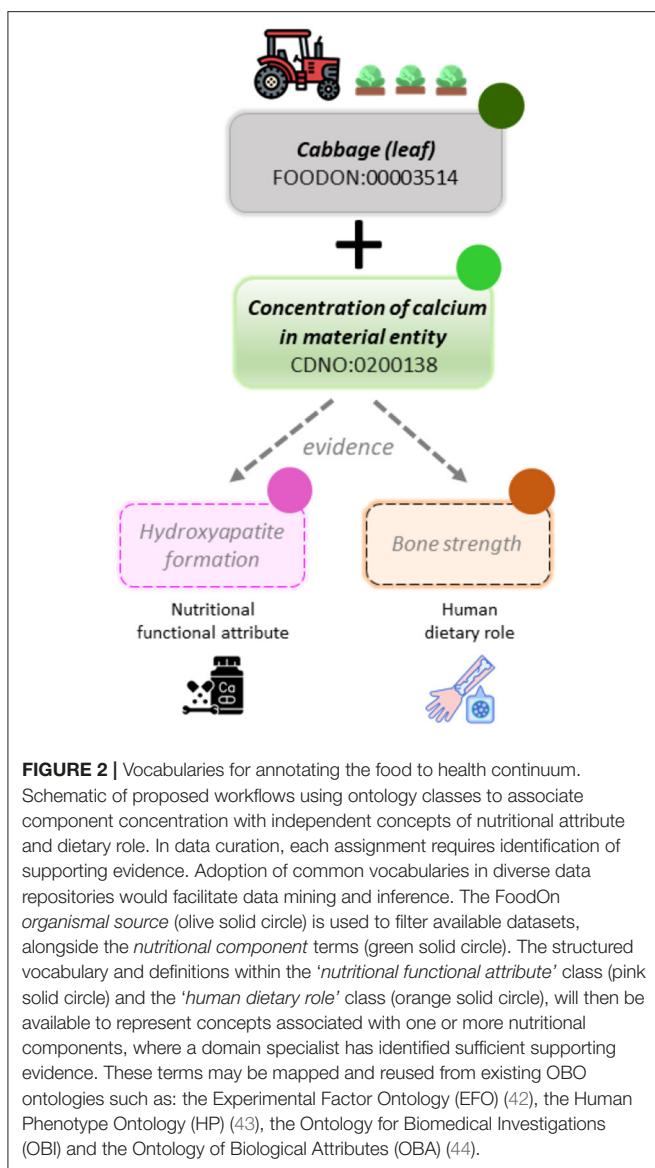
The CDNO team has recently received positive feedback for the classification and description of additional terms to address the increasing clinical, consumer and market interest in the relationship between the nutritional composition of food and ingredients, its provenance and its ability to affect personal and public health outcomes. There is a growing but dispersed evidence base associated with "functional foods"

and "nutraceuticals." We therefore sought to provide a clear framework that distinguishes between chemical components and physical properties of material entities, the functional attributes they may possess, and any proposed associated human dietary role (Figures 1 and 2B). The CDNO '*nutritional functional attribute*' [CDNO:0300001] class hierarchy (Figure 1A) is therefore defined as: "A functional attribute that inheres in one or more dietary nutritional components (or food material) and may contribute to a dietary role." This provides a structured vocabulary that allows description of quantifiable knowledge (Figure 1A), with terms such as "antioxidant status" or "glycemic index."

In order to represent the distinct concepts relating to potential role in the context of health and wellbeing, we then defined the '*human dietary role*' [CDNO:0500001] class as: "A biological role that may be assigned to a dietary nutritional component based on evidence, supported at the levels of molecular interaction, cellular process or physiological role." The value of establishing these distinct classes is demonstrated by the ambiguity associated with use of the word "vitamin," which may refer either to a role [CHEBI:33229] and/or to a chemical entity [CDNO:0000014], depending on context. It is recognized that any conjecture made in relation to role [BFO:0000023] is dependent on an evidence base (Figure 2B), and so the terms defined within the '*human dietary role*' class hierarchy are made available primarily for data curators and specialists to associate or annotate with evidence-based datasets. Moreover, a role may be dependent on many variables, including but not limited to concentration (dose), physical form (bioavailability), demographic, genetic, developmental stage and/or health status of the human subject, as well as intake of other dietary components. Such variables may be defined within other OBO ontologies such as the Ontology for Nutritional Studies (ONS) (40), which provides a framework for evidence-based studies structured according to various parameters or the Environmental Conditions, Treatments, and Exposures Ontology (ECTO) which supports modelling of exposure processes such as dietary exposures (41). We anticipate that further development of the CDNO classes described above may benefit from reuse of terms from additional OBO ontologies such as the Experimental Factor Ontology (EFO) (42), the Human Phenotype Ontology (HP) (43), OBI, and the Ontology of Biological Attributes (OBA) (44). However, substantial work between domain experts and OBO ontology curators will be required to resolve any discrepancies and allocate appropriate terms to the '*nutritional functional attribute*' or '*human dietary role*' classes (Figure 2).

## CDNO for Data Curation and Retrieval

The CDNO is a live open-source project that encourages regular discussion and enhancements to be proposed by stakeholders, in the food supply and nutritional domain. Current updates by the CDNO developers are made in consultation with the OBO community. The value of a common vocabulary is demonstrated by the ease with which specific terms may be associated with distinct data sources. Since CDNO is expected to facilitate the compilation and analysis of a diverse range of datasets, we present a use case interface that demonstrates how variation in



the concentration of nutritional components may be compared between datasets from different sources and levels of abstraction (**Supplementary Figure 2**).

Analytical samples may be derived from any stage in the characterization of biodiversity, plant breeding, cultivar improvement and deployment, through the food production, processing and supply chains, as well as from food storage, preparation, consumption and digestion (**Supplementary Figure 1; Figure 1A**). In many cases, the '*concentration of dietary nutritional component in material entity*' [CDNO:0200001] terms would be used to annotate data in a curation pipeline in conjunction with FoodOn terms that define organismal source (e.g., from crops, livestock, fisheries) and organismal part (e.g., grain, liver, fin) (**Figure 1B**).

In order to demonstrate a use-case we curated data that used a series of metadata terms to describe specific sets of nutritional component concentration data. This included the FoodOn '*Food*

*product by organism*' [FOODON:00002381] class (**Figure 1**), and the National Center for Biotechnology Information (NCBI) organismal classification ontology (NCBITaxon) entity (**Supplementary Figure 2**). We then developed an online data retrieval interface that allows selection of specific CDNO terms to filter and access multiple sources of nutritional data derived from a crop biodiversity database (45), a national food composition database (46) and a geo-spatial dietary nutrition study (47) (**Supplementary Figure 2**).

One valuable outcome of generating an ontology is the opportunity to include structured annotation within the reference OWL file. This may include authoritative citations, as well as formal cross-references to published databases and other reference data sources. We therefore incorporated provisional cross-references between specific terms and widely used FCT/FCDB tagnames/code numbers. For example, the term '*concentration of proline in material entity*' [CDNO:0200062] was associated with the tagname "PRO" from INFOODs and the code "1226" from USDA. Ongoing maintenance of this feature will facilitate the harmonization of vocabularies used in different FCDBs.

## CONCLUSION

CDNO is a new, open source, community developed, web-accessible vocabulary providing a formal representation of commonly used dietary and nutritional terminology. Ongoing development of this ontology (48) will contribute toward data sharing and interoperability, particularly for initiatives where a wide range of foodstuffs are analyzed to diversify diet and agricultural production (49). We anticipate that extending the range and harmonization of terminologies in food systems will facilitate the sourcing and management of nutritional resources, and stimulate development of information-led markets (50, 51). In particular, there is scope for large-scale data integration that enables downstream meta-analyses to complement advances in human, crop and livestock genomics and high throughput analytical chemistry.

## DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://github.com/Southern-Cross-Plant-Science/cdno>.

## AUTHOR CONTRIBUTIONS

LA-H, KB, RW, DD, RM, and GK were members of the working group that conceptualized and design of the CDNO. LA-H, KB, RW, DD, RM, ML, MW, LC, AMa, AMø, JI, LS, XZ, BB-F, PM, SF, AS, LL, and GK provided significant analysis and interpretation of research data. LA-H and GK wrote the first draft of the manuscript. KB and LA-H created the codebase for the CDNO. All authors provided critical revisions on manuscript drafts and approved the final manuscript.

## FUNDING

LA-H performed this work as part of a Scholarship from Southern Cross University and Crops For the Future, with additional and subsequent post-doc funding from WISEcode LLC.

## ACKNOWLEDGMENTS

The authors thank Edward Joy for valuable discussion about the geospatial nutritional data. The authors also thank the Open Biomedical Ontologies personnel who participated in the development, release, and constant improvement of this work. The authors also thank Naomi Fukagawa from USDA for providing feedback on the initial CDN-DF framework.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fnut.2022.928837/full#supplementary-material>

**Supplementary Figure 1** | Representation of nutritional information in the continuum between agriculture and health. Knowledge is represented in the context of different domains that may contribute to variation in nutritional

outcomes. The symbols under each domain correspond to each of the major CDNO and FoodOn classes represented in **Figure 1**.

**Supplementary Figure 2** | Use-case of CDNO terms used to mine and compare data from diverse sources. Online graphical interface developed to demonstrate how the 'concentration of dietary nutritional component in material entity' [CDNO:0200001] classes are associated with quantitative data derived from (i) genetic resources, (ii) food composition tables and (iii) a geo-spatial study. Data associated with distinct data sources (see below) have been curated within a relational schema to enable semantic search and filtering based on annotation of key records, where ontology terms are managed within an 'ontology register' table. The user may navigate the hierarchical class tree (here, CDNO v3.0) and select (green solid circle) a nutritional component (e.g. CDNO:0200138 - 'calcium concentration'). Food or crop groups may be selected (olive solid circle) and then refined to e.g. specific crops or organismal parts used for food. Data from the filtered datasets may be represented by a combination of box-whisker plots, frequency distributions, or single values. Calcium concentration data were sourced to represent variation within (i) a crop biodiversity collection representing the vegetable *Brassica oleracea* genepool (52), the Brassica Information Portal (45) (ii) vegetables and vegetable products from a food composition database - the Canadian Nutrient File (53) - cited in (46) and (iii) a geo-spatial study where crop-based edible food items were sampled from multiple locations in Malawi, with data presented for variation in the underutilized crop *Moringa oleifera* (47). Recorded values have been adjusted to enable direct comparison on a consistent y-axis, as the original units varied according to each study (i) %, (ii) mg/100g, (iii) mg per kg. For (ii), as for many food composition databases, only single values per nutrient component per food are available (15), although the number of original samples from which the mean value is derived is stated.

## REFERENCES

1. Fanzo J, Bellows AL, Spiker ML, Thorne-Lyman AL, Bloem MW. The importance of food systems and the environment for nutrition. *Am J Clin Nutr.* (2021) 113:7–16. doi: 10.1093/ajcn/nqaa313
2. Azman Halimi R, Barkla BJ, Mayes S, King GJ. The potential of the underutilized pulse bambara groundnut (*Vigna subterranea* (L) Verdc) for nutritional food security. *J Food Compos Anal.* (2019) 77:47–59. doi: 10.1016/j.jfca.2018.12.008
3. Chan L, Vasilevsky N, Thessen A, McMurry J, Haendel M. The landscape of nutri-informatics: a review of current resources and challenges for integrative nutrition research. *Database.* (2021) 2021:1–20. doi: 10.1093/database/baab003
4. Andrés-Hernández L, Baten A, Halimi RA, Walls R, King GJ. Knowledge representation and data sharing to unlock crop variation for nutritional food security. *Crop Sci.* (2020) 60:1–14. doi: 10.1002/csc2.20092
5. Munir K, Sheraz Anjum M. The use of ontologies for effective knowledge modelling and information retrieval. *Appl Comput Informatics.* (2018) 14:116–26. doi: 10.1016/j.aci.2017.07.003
6. Gan M, Dou X, Jiang R. From ontology to semantic similarity: calculation of ontology-based semantic similarity. *Sci World J.* (2013) 2013:793091. doi: 10.1155/2013/793091
7. Davis KE, Downs S, Gephart JA. Towards food supply chain resilience to environmental shocks. *Nat Food.* (2021) 2:54–65. doi: 10.1038/s43016-020-00196-3
8. Greenfield H, Southgate DAT. *Food Composition Data: Production, Management, and Use.* 2nd ed. Food and Agriculture Organization of the United Nations (2003).
9. Jackson R, Matentzoglu N, Overton JA, Vita R, Balhoff JP, Buttigieg PL, et al. OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. *Database.* (2021) 2021:1–9. doi: 10.1093/database/baab069
10. Degtyarenko K, De matos P, Ennis M, Hastings J, Zbinden M, Mcnaught A, et al. ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res.* (2008) 36:344–50. doi: 10.1093/nar/gkm791
11. Wilkinson MD. Comment : the FAIR guiding principles for scientific data management and stewardship. *Nature.* (2016) 3:1–9.
12. Dooley DM, Grif EJ, Gosal GS, Buttigieg PL, Hoehndorf R, Schriml LM, et al. FoodOn : a harmonized food ontology to increase global food traceability, quality control and data integration. (2018) 2:23. doi: 10.1038/s41538-018-0032-6
13. OBO Foundry. *Principles: Overview.* (2022). Available online at: <http://www.obofoundry.org/principles/fp-000-summary.html> (accessed January 20, 2022).
14. Andrés-Hernández L. *Github CDNO Repository.* (2021). Available online at: <https://github.com/Southern-Cross-Plant-Science/cdno> (accessed January 20, 2022).
15. Azman Halimi R, Barkla B, Andrés-Hernández L, Mayes S, King GJ. Bridging the food security gap : an information-led approach to connect dietary nutrition, food composition. *J Sci food Agric.* (2019) 100:1495–504. doi: 10.1002/jsfa.10157
16. Charrondiere UR, Burlingame B. Report on the FAO/INFOODS Compilation Tool: A simple system to manage food composition data. (2011) 24:711–5. doi: 10.1016/j.jfca.2010.09.006
17. Haytowitz DB, Pehrsson PR. USDA's National Food and Nutrient Analysis Program (NFNAP) produces high-quality data for USDA food composition databases: two decades of collaboration. *Food Chem.* (2018) 238:134–8. doi: 10.1016/j.foodchem.2016.11.082
18. Møller A, Unwin ID, Becker W, Ireland J. EuroFIR's food databank systems for nutrients and bioactives. *Trends Food Sci Technol.* (2007) 18:428–33. doi: 10.1016/j.tifs.2007.02.003
19. Bechhofer S. OWL: Web Ontology LanguageIn: LIU L, Özsü MT, editors *Encyclopedia of Database Systems.* Boston, MA: Springer US (2009). doi: 10.1007/978-0-387-39940-9\_1073
20. Seppälä S, Ruttenberg A, Smith B. Guidelines for writing definitions in ontologies. *Cienc da Inf.* (2017) 46:73–88. doi: 10.18225/ci.inf.v46i1.4015
21. Osumi-Sutherland D, Courtot M, Balhoff JP, Mungall C. Dead simple OWL design patterns. *J Biomed Semantics.* (2017) 8:1–7. doi: 10.1186/s13326-017-0126-0
22. Buttigieg P, Morrison N, Smith B, Mungall CJ, Lewis SE. The environment ontology: contextualising biological and biomedical entities. *J Biomed Semantics.* (2013) 4:43. doi: 10.1186/2041-1480-4-43

23. 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 interoperability. *J Biomed Semantics*. (2016) 7:1–12. doi: 10.1186/s13326-016-0097-6

24. Gkoutos G V, Green ECJ, Mallon A, Hancock JM, Davidson D. Using ontologies to describe mouse phenotypes. *Genome Biol*. (2005) 6:R8. doi: 10.1186/gb-2004-6-1-r8

25. Smith B. On classifying material entities in Basic Formal Ontology. In: *Proceedings of the Third Interdisciplinary Ontology Meeting*. Tokyo: Keio University Press (2012) 1–13.

26. OBO Foundry Revision b5bcde0f. *Creating a new Ontology using the ODK*. (2012). Available online at: <https://ontology-development-kit.readthedocs.io/en/latest/NewRepo.html> (accessed June 12, 2020).

27. Jackson RC, Balhoff JP, Douglass E, Harris NL, Mungall CJ, Overton JA, et al. ROBOT: a tool for automating ontology workflows. *BMC Bioinformatics*. (2019) 20:1–10. doi: 10.1186/s12859-019-3002-3

28. Leibovici DG, Anand S, Santos R, Mayes S, Ray R, Al-Azri M, et al. Geospatial binding for transdisciplinary research in crop science: the GRASPgfs initiative. *Open Geospatial Data, Softw Stand*. (2017) 2:20. doi: 10.1186/s40965-017-0034-3

29. Schulz S, Stenzhorn H, Boeker M. The ontology of biological taxa. *Bioinformatics*. (2008) 24:313–21. doi: 10.1093/bioinformatics/btn158

30. Jaiswal P, Ware D, Ni J, Chang K, Zhao W, Schmidt S, et al. Gramene: development and integration of trait and gene ontologies for rice. *Comp Funct Genomics*. (2002) 3:132–6. doi: 10.1002/cfg.156

31. Ontobee CDNO. (2022). Available online at: <https://www.ontobee.org/ontology/CDNO> (accessed January 20, 2022).

32. The Ontology Lookup Service. (2022). Available online at: <https://www.ebi.ac.uk/ols/ontologies/cdno> (accessed March 20, 2022).

33. Møller A, Unwin ID, Ireland J, Roe MA, Becker W, Colombani P. *The EuroFIR Thesauri 2008*. EuroFIR Technical Report, Denmark: Danish Food Information (2008).

34. EuroFIR thesauri. (2020). Available online at: <https://www.eurofir.org/our-resources/eurofir-thesauri/> (accessed January 20, 2022).

35. Englyst KN, Liu S, Englyst HN. Nutritional characterization and measurement of dietary carbohydrates. *Eur J Clin Nutr*. (2007) 61:S19–39. doi: 10.1038/sj.ejcn.1602937

36. Cummings JH, Stephen AM. Carbohydrate terminology and classification. *Eur J Clin Nutr*. (2007) 61:S5–S18. doi: 10.1038/sj.ejcn.1602936

37. Finglas PM, Berry R, Astley S. Assessing and improving the quality of food composition databases for nutrition and health applications in Europe: the contribution of EuroFIR. *Adv Nutr*. (2014) 5:608S–614S. doi: 10.3945/an.113.005470

38. Batchelor C. *The Chemical Methods Ontology*. (2019). Available online at: <https://github.com/rsc-ontologies/rsc-cmo> (accessed February 13, 2021).

39. Bandrowski A, Brinkman R, Brochhausen M, Brush MH, Bug B, Chibucos MC, et al. The ontology for biomedical investigations. *PLoS ONE*. (2016) 11:1–19. doi: 10.1371/journal.pone.0154556

40. Vitali F, Lombardo R, Rivero D, Mattivi F, Franceschi P, Bordoni A, et al. ONS: an ontology for a standardized description of interventions and observational studies in nutrition. *Genes Nutr*. (2018) 13:1–9. doi: 10.1186/s12263-018-0601-y

41. Chan LE, Thessen AE, Duncan WD, Matentzoglu N, Schmitt C, Grondin C, et al. The Environmental Conditions, Treatments, and Exposures Ontology (ECTO): connecting toxicology and exposure to human health and beyond. (2022):1–25. doi: 10.5281/zenodo.6360645%0A

42. Malone J, Adamusiaik T, Holloway E, Parkinson H. Developing an application ontology for annotation of experimental variables – Experimental Factor Ontology. *Nat Preced*. (2009) 10:1–1. doi: 10.1038/npre.2009.3806.1

43. Robinson PN, Köhler S, Bauer S, Seelow D, Horn D, Mundlos S. The human phenotype ontology: a tool for annotating and analyzing human hereditary disease. *Am J Hum Genet*. (2008) 83:610–5. doi: 10.1016/j.ajhg.2008.09.017

44. *TermGenie for the ontology of biological attributes (OBA)*. Available online at: <http://oba.termgenie.org> (accessed March 9, 2022).

45. Eckes AH, Gubala T, Nowakowski P, Szymczyszyn T, Wells R, Irwin JA, et al. Introducing the Brassica information portal: towards integrating genotypic and phenotypic Brassica crop data. *F1000Research*. (2017) 6:465. doi: 10.12688/f1000research.11301.1

46. Deeks J, Verreault MF, Cheung W. Canadian Nutrient File (CNF): update on Canadian food composition activities. *J Food Compos Anal*. (2017) 64:43–7. doi: 10.1016/j.jfca.2017.04.009

47. Joy EJM, Broadley MR, Young SD, Black CR, Chilimba ADC, Ander EL, et al. Soil type influences crop mineral composition in Malawi. *Sci Total Environ*. (2015) 505:587–95. doi: 10.1016/j.scitotenv.2014.10.038

48. King GJ. The Compositional Dietary Nutrition Ontology Webpage. (2022). Available online at: [www.cdno.info](http://www.cdno.info) (accessed April 8, 2022).

49. Cifelli CJ. Looking beyond traditional nutrients: the role of bioactives and the food matrix on health. *Nutr Rev*. (2021) 79:1–3. doi: 10.1093/nutrit/nuab100

50. Mazac R, Tuomisto HL. The post-anthropocene diet: navigating future diets for sustainable food systems. *Sustain*. (2020) 12:1–15. doi: 10.3390/su12062355

51. Zeb A, Soininen J-P, Sozer N. Data harmonisation as a key to enable digitalisation of the food sector: a review. *Food Bioprod Process*. (2021) 12:360–70. doi: 10.1016/j.fbp.2021.02.005

52. Broadley MR, Hammond JP, King GJ, Astley D, Bowen HC, Meacham MC, et al. Shoot calcium and magnesium concentrations differ between subtaxa, are highly heritable, and associate with potentially pleiotropic loci in Brassica oleracea. *Plant Physiol*. (2008) 146:1707–20. doi: 10.1104/pp.107.114645

53. The Canadian Nutrient File (2021). Available online at: <https://www.canada.ca/en/health-canada/services/food-nutrition/healthy-eating/nutrient-data/canadian-nutrient-file-2015-download-files.html> (accessed January 20, 2022).

**Conflict of Interest:** PM and AS are employed by WISECode LLC. This study received funding from WISECode LLC. The funder had the following involvement with the study: provision of domain-specific know-how in establishing scope of the ontology and classes, along with key insights into practical end use in data management. SF was employed by Nutrient Institute LLC.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Andrés-Hernández, Blumberg, Walls, Dooley, Mauleon, Lange, Weber, Chan, Malik, Møller, Ireland, Segovia, Zhang, Burton-Freeman, Magelli, Schriever, Forester, Liu and King. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.