Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at *Nutrients* **2019**, *11*, 1300; doi:10.3390/nu11061300

Article

An ontology to standardize nutritional epidemiologic research output: from paper-based standards to linked content

Chen Yang[1], Henry Ambayo[1], Bernard De Baets[2], Patrick Kolsteren[1], Nattapon Thanintorn[3], Dana Hawwash[1], Jildau Bouwman[4], Antoon Bronselaer[5], Filip Pattyn[6¶], Carl Lachat[1¶*]

[1] Department of Food Technology, Safety and Health, Ghent University, Ghent, Belgium; Chen.Yang@UGent.be; henryambayo@gmail.com; Patrick.Kolsteren@UGent.be; Dana.Hawwash@UGent.be; Carl.Lachat@UGent.be;
[2] KERMIT, Department of Data Analysis and Mathematical Modelling, Ghent University, Ghent, Belgium; Bernard.DeBaets@UGent.be;
[3] Department of Pathology and Anatomical Sciences, University of Missouri, USA; nthanintorn@hotmail.com;
[4] Netherlands Organization for Applied Scientific Research, Zeist, Netherlands; jildau.bouwman@tno.nl;
[5] Department of Telecommunications and information processing, Ghent University, Ghent, Belgium; Antoon.Bronselaer@UGent.be;
[6] ONTOFORCE, Ghent, Belgium; filip.pattyn@ontoforce.com;

* Correspondence: Carl.Lachat@UGent.be; Tel.: +32-(0)9 264 93 77

Received: date; Accepted: date; Published: date

Abstract: 1) Background: The use of linked data in Semantic Web are promising approaches to add value to nutrition research. An ontology, which defines the logical relationships between well-defined taxonomic terms, enables linking and harmonizing research output. To enable the description of domain-specific output in nutritional epidemiology, we propose the Ontology for Nutritional Epidemiology (ONE) according to authoritative guidance for nutritional epidemiologic research; 2) Methods: First, a scoping review was conducted to identify existing ontology terms for reuse in ONE. Second, existing data standards and manuscript reporting guidelines for nutritional epidemiology were converted into ontology, and the terms used in the standards were summarized and listed separately in a taxonomic hierarchy. Third, the ontologies of the nutritional epidemiologic standards, reporting guidelines and the core concepts were gathered in ONE. Three case studies were illustrated for its potential applications. (i) annotation of existing manuscripts and data, (ii) ontology-based inference, and (iii) estimation of reporting completeness in a sample of nine manuscripts; 3) Results: Ontologies for "food and nutrition" (n=33), "disease and special population" (n=86), "data description" (n=21), "research description" (n=32) and "supplementary (meta) data description" (n=44) were reviewed and listed. ONE consists of 339 classes (79 new classes to describe nutrition data and 24 new classes to describe the content of nutrition manuscripts). The case studies demonstrated the application of ONE. 4) Conclusion: ONE is a resource to automate data integration, searching and browsing, and can be used to assess reporting completeness in nutritional epidemiology.

Keywords: ontology, nutritional epidemiology, minimal data information, data quality descriptors, study reporting guidelines, Semantic Web

1. Introduction

Nutritional epidemiology provides evidence regarding the effects of human diets on health [1]. Unfortunately, most evidence is produced by short-term randomized trials or observational studies with small effect sizes [2]. Large-scale studies are time-consuming and demand substantial involvement of participants. Integrated analysis of shared data could increase the power of analysis

© 2019 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

and add considerable value to research [3]. However, due to the various descriptions of data and research output at large, retrieval and use of shared nutritional epidemiologic data is challenging. Reporting guidelines describe essential information for research output and are useful to standardize the description of nutritional epidemiologic output [4].

An ontology framework developed based on such guideline enables a standardized method of data descriptions in Semantic Web [5, 6]. An ontology consists of terms and relations of terms to structure the description of shared data in Semantic Web (9). While a terminology defines the terms, an ontology defines the relations between the well-defined terms to structure the description of shared data. Ontology terms and relations are human-readable, but their electronic identifiers also enable computer processing such as inferencing and machine learning [7, 8]. An introduction to ontology with simple examples is given by Noy and McGuinness [9].

Ontologies can contribute to make research output (e.g. data, manuscript and study protocols) Findable, Accessible, Interoperable and Re-usable (FAIR) [10]. FAIR research output is now made mandatory by research funders such as the European Commission for the establishment of a European Open Science Cloud [11].

The development of a virtual research infrastructure to share research output with researchers, consumers, the public and the private sector is a promising prospect for nutrition science [12]. Despite calls since 2007 [13], progress towards an ontology for nutritional epidemiology is poor. FoodOn was developed as a taxonomy for food classification and description [14], with subsequent identifiers in Langual and FoodEx2 [15, 16]. Although generic ontologies such as the Ontology for Nutritional Studies [17] and Bionutrition Ontology [18] are available, none of these include structure for describing nutritional epidemiologic output.

We present the Ontology for Nutritional Epidemiology (ONE), as well as case studies to illustrate its potential applications. The purpose of developing ONE is not to introduce a novel controlled vocabulary or terminology, but to define the relations between (often existing) terms for describing nutritional epidemiologic research. ONE hence identifies relevant existing ontology terms and introduces a minimum of new terms.

ONE has three components: 1) “descriptors for nutritional epidemiologic data”: meta-data descriptions for nutritional epidemiologic data; 2) “STROBE-nut (STrengthening the Reporting of OBservational studies in Epidemiology - an extension for nutritional epidemiology) items”: quality descriptors for reporting nutritional epidemiologic studies and 3) “Nutritional epidemiologic terms”: core nutritional epidemiologic concepts. ONE has been registered on Bioportal (https://bioportal.bioontology.org/ontologies/ONE).

The present study was conducted in the context of the European Nutritional Phenotype Assessment and Data Sharing Initiative, a collaborative effort of 16 multidisciplinary consortia from 50 research centers in 9 countries, aiming to promote data sharing in nutrition.

2. Materials and Methods

A scoping review of existing ontology terms provided a basis for the development of ONE. Next, ONE was developed by converting paper-based nutritional epidemiologic standards [4, 19, 20] into an ontology representation, including a separate taxonomic hierarchy of specific nutritional epidemiologic terms. Finally, ONE was applied in three case studies to illustrate its potential applications.

2.1 Review and summary of existing ontologies for use in nutritional epidemiology

Nutritional epidemiology is an interdisciplinary science that builds on other disciplines such as nutrition, food science, medicine and epidemiology. Instead of developing a new stand-alone ontology, we therefore first considered existing ontologies in other disciplines, and then identified missing elements for nutritional epidemiology [13]. On April 13th 2018, all ontologies in the three main medical ontology libraries [21, 22]: OBO Foundry (http://www.obofoundry.org/) [23], BioPortal

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

(https://bioportal.bioontology.org/) [24] and Ontology Lookup Service (https://www.ebi.ac.uk/ols/index) [25] were reviewed by CY and HA independently. Ontologies were included if their scope met part of the controlled vocabulary requirement of nutritional epidemiology as shown in Supplemental Table 1.

A pre-established data extraction spreadsheet was used to list all ontologies for review. Three review rounds were conducted. During the first review round, the full names of all the ontologies were assessed. During the second review round, the short descriptions of the ontologies on their BioPortal homepage were reviewed. If the information from the descriptions was insufficient or reviewer disagreement appeared, ontologies were included for the next review round; Finally, during the third review round, the included terms and taxonomies of the ontologies were reviewed. Disagreements were resolved through discussion until a consensus was reached. In case some ontologies were inaccessible, information on these ontologies was reviewed in relevant publications or web pages.

The FAIR principles provide essential guidance to search for and integrate data at individual and meta-level. The required types of controlled vocabulary to achieve FAIR principles in nutritional epidemiology were summarized (Table A1), and the selected ontologies were classified accordingly. A quality assessment of the selected ontologies was conducted using the modules by Burton-Jones, Storey [26]. Minor changes were made to present the quality of multiple medical ontologies. On May 16th 2018, statistics were collected through BioPortal (https://bioportal.bioontology.org/), Agroportal (http://agroportal.lirmm.fr/) and Ontobee (http://www.ontobee.org/).

2.2 Development of ONE

The ontology is represented in the Resource Description Framework (RDF) format [27] and edited using the default text editor of Microsoft Windows 7. A quality assessment of existing ontologies in step 1 was conducted as proposed by Burton-Jones, Storey [26]. The relevance, authority and history module were not assessed however, as they require data collection after publishing the ontology.

2.2.1 Existing data standards in nutritional epidemiology

The terms of two existing standards for nutrition research (i.e. minimal meta-data descriptors [19] and data quality descriptors [20]) were represented in ONE. The ontology terms were grouped as "descriptors for nutritional epidemiologic data".

In case certain terms were found in more than one ontology, the term with the definition that best described the intended term was selected by a domain expert. When no exact terms were found in the selected ontologies, a synonym term was obtained from a domain expert if the definition was suitable.

However, if the exact term or the synonym could not be retrieved from existing ontologies, a new electronic identifier was attributed: 1) for terms only used in nutritional epidemiologic research, the identifier "one:nexxxxx" (xxxxx=5 digits) was assigned, where "one" represents "ontology for nutritional epidemiology", "ne" represents "only used in nutritional epidemiology" (e.g. identifier for "dietary assessment administration": one:ne00057); 2) for other terms that can be also used in other subjects, identifier "one:Txxxxx" (xxxxxx=5 digits) was assigned, where "one" represents "ontology for nutritional epidemiology" and "T" represents "temporary" (e.g. identifier for "food composition table": one:T00027). Terms indicated with "T" should hence be developed in their corresponding domain ontology. The list of temporary terms will be reviewed on a regular basis and updated where needed.

2.2.2 Reporting guidelines in nutritional epidemiology

"The STrengthening the Reporting of OBservational studies in Epidemiology (STROBE)- reporting guidelines for nutritional epidemiology" [4] were used as the basis to develop the ontology for

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

manuscript reporting in nutritional epidemiology. The collection of ontology terms is allocated under the term "STROBE-nut items" in ONE. For the STROBE-nut reporting items (e.g. title, abstract, etc.), electronic identifier "one:reportxxxxx" (xxxxx=5 digits) was given, where "one" represents "ontology for nutritional epidemiology", "report" represents "reporting items" (e.g. identifier for "title": one:report00001); 2) for the STROBE-nut recommendations, identifier "one:report/nut-x" (x = digits) was assigned, where "one" represents "ontology for nutritional epidemiology", "report/nut-x" represents "the STROBE-nut recommendations for reporting on items" (e.g. identifier for "STROBE-nut recommendation 1": one:report/nut-1).

2.2.3 Nutritional epidemiologic terms

The term "nutritional epidemiologic terms" (electronic identifier: one:terms) was used to group the specific nutritional epidemiologic terms summarized from the standard descriptions during the previous steps. The taxonomy presents terms for describing the core concepts, study design and data measurement characteristics of nutritional epidemiology. However, those terms do not cover generic information to report study findings, such as study name, study duration, study area, etc. Terms used for generic study information, however, are considered minimal data requirements and quality descriptors, and were hence mainly retrieved from other existing ontologies.

2.3 Applications of ONE

ONE was applied in three case studies to illustrate its potential applications: (i) study annotation and term query, (ii) ontology-based inference, and (iii) estimation of reporting completeness in a sample of nine manuscripts.

First, an existing manuscript [28] and one of its corresponding datasets were annotated manually using ONE terms (Syntax available on Bioportal). Terms from other ontologies were also used to annotate nutrition information that was not related to nutrition (e.g. geography, season, etc.).

Second, a potential ontology-based inference was illustrated. Three terms used for annotating the existing manuscript were selected for the case [28]. By showing partial taxonomic hierarchies of the three terms, we explained how to infer unknown information from available information. Inference on the basis of the taxonomy of terms can significantly improve the quality of data search and integration.

Third, an assessment of reporting completeness was conducted using the ontology, similar to the ontology-based meta-analyses by Kupershmidt, Su [29] and Ramaprasad and Syn [30]. A convenient sample of nine published manuscripts [31-39] was manually annotated using STROBE-nut terms of ONE for this purpose. By querying the electronic identifiers of STROBE-nut terms, the reporting frequencies of STROBE-nut terms were obtained. The hierarchies of STROBE-nut terms and one annotated manuscript were compared to illustrate where STROBE-nut terms were reported in the manuscript.

3. Results

3.1 Review and summary of existing ontology vocabulary for use in nutritional epidemiology

In total, 1053 ontologies were retrieved and 216 ontologies were selected and classified according to their scope (Figure 1). As shown in Supplemental Table 2, 140 ontologies were selected for data annotation (29 ontologies for food/dietary agricultural products, 4 ontologies for nutrients/chemical compounds, 86 ontologies for disease & special population and 21 ontologies for data management) and 32 for metadata annotation (32 ontologies for research terminology and no ontology for metadata representation). There are also 44 ontologies that can be used for supplementary (meta) data (e.g. ethical issues, demographics, fundamental ontology knowledge frameworks, etc.). Among the ontologies found, no ontology was developed as a frame (e.g. guidance and guidelines) to present nutritional epidemiologic information (i.e. meta-data representation) [26].

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

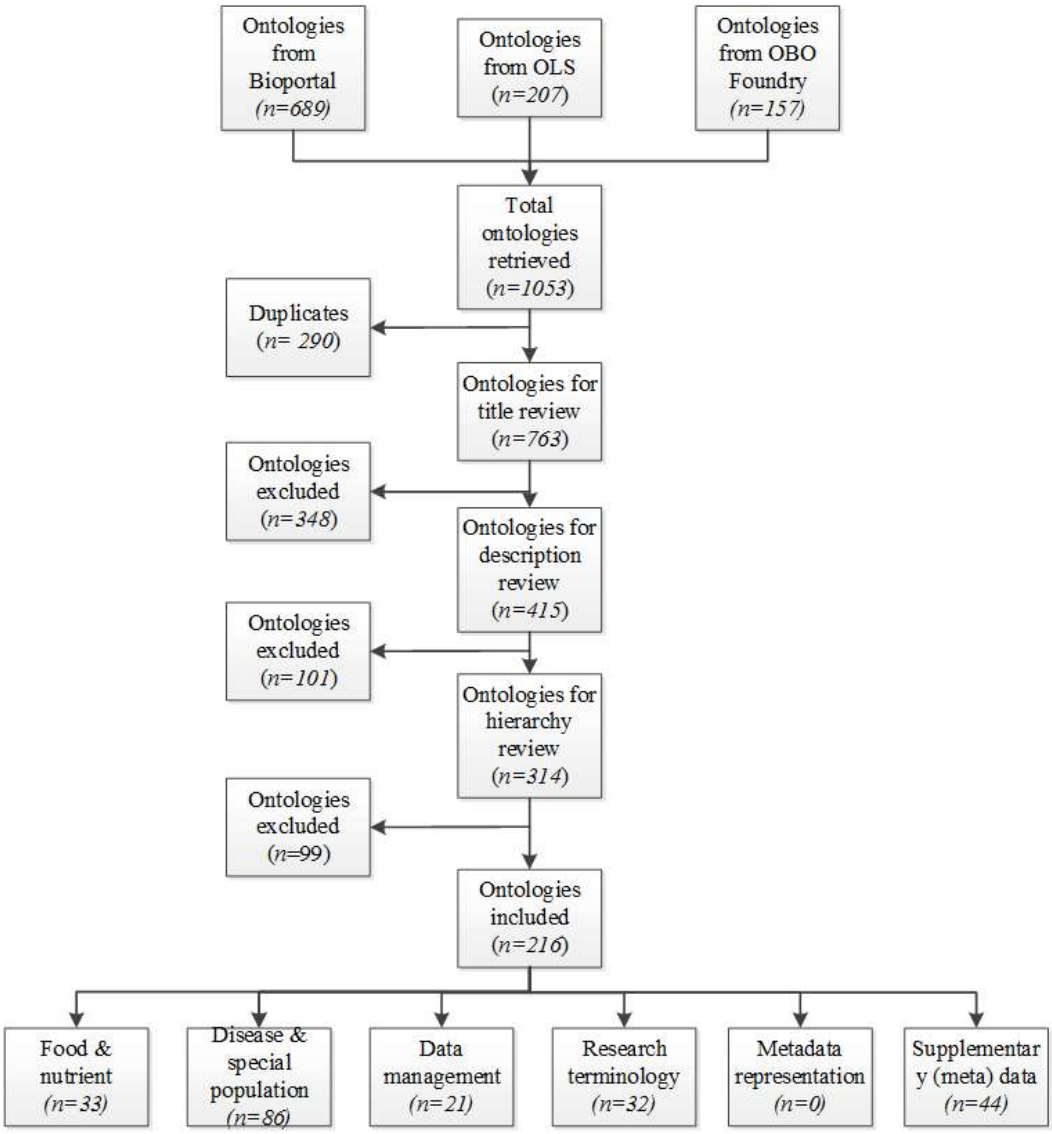


Figure 1. Review and selection process of ontologies for nutritional epidemiology

The quality assessment (Figure A1-a) shows that 14% of the selected ontologies have less than 100 terms. Most of the selected ontologies (64%) have 101-10000 terms, while 15% of the selected ontologies have more than 10000 terms. The richness module (Figure A1-b) shows that 16% of the selected ontologies have no properties, 24% of the selected ontologies have 1-10 properties, 54% of the selected ontologies have more than 10 properties, including 13% of the selected ontologies with over 100 properties. Figures A1-c and 1-d indicate that 23% of the terms have no definitions, and 93% of the selected ontologies are not peer-reviewed. The lawfulness module (Figure A-e), authority module (Figure A1-f) and history module (Figure A1-g) represent the practicality of the selected ontologies. Only 2% of the selected ontologies are inaccessible due to error ontology files (Figure A1-e). Only 6% of the selected ontologies are not mapped, while 22% of the selected ontologies are made of more than 300 mapped ontologies (Figure A1-f). Less than half (42%) of the selected ontologies were visited less than ten times per month (Figure A1-g).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

3.2 Development of ONE

The structure of ONE is shown in Figure 2 and a quality description is included in Table A3. ONE consists of 339 classes. It reused classes from 22 existing ontologies, the main referred medical ontologies are NCIT (National Cancer Institute Thesaurus, 43 classes) and MeSH (Medical Subject Headings, 33 classes). ONE proposes 79 new classes to describe nutrition data and 24 new classes to describe the content of nutrition manuscripts.

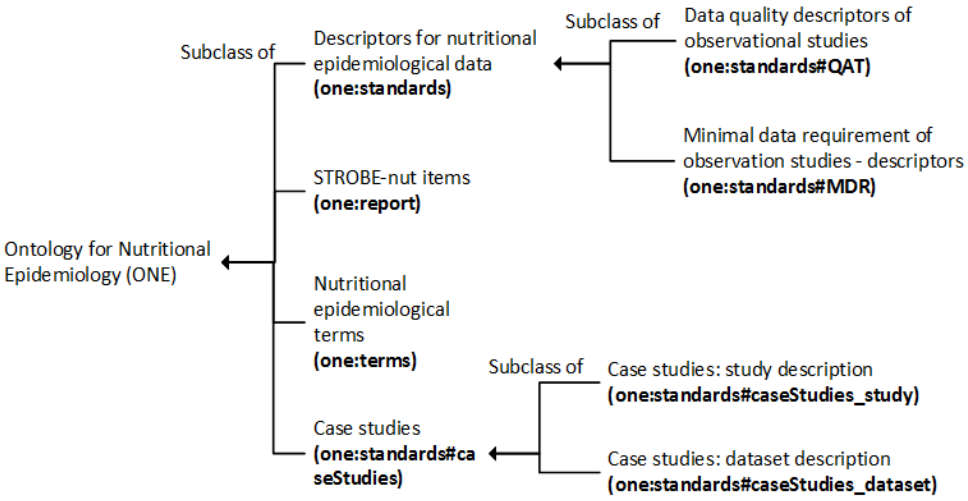
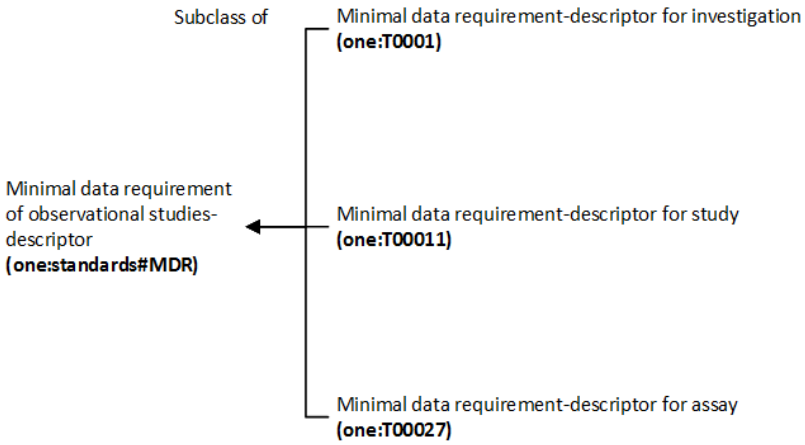


Figure 2. The overall structure of the ontology for nutritional epidemiology (ONE)

The electronic identifiers of terms are written after the corresponding terms. The electronic identifiers (e.g. NCIT:C94729) consist of two parts: 1) an ontology acronym (e.g. "NCIT" is the acronym of "ontology for National Cancer Institute Thesaurus") and 2) a code of the term in the ontology (e.g. C94729 is code of "season" in NCIT ontology).

3.2.1 Existing data standards in nutritional epidemiology

The main taxonomies of minimal data requirements and data quality descriptors are shown in Figures 3 and 4, respectively The collection of ontology terms is reported in Tables A8 and A9, respectively. Recommendations for generic terms that could not be found in existing ontologies of other subjects are indicated as footnotes of Tables A8 and A9.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Figure 3. The ontology taxonomy of minimal data requirements

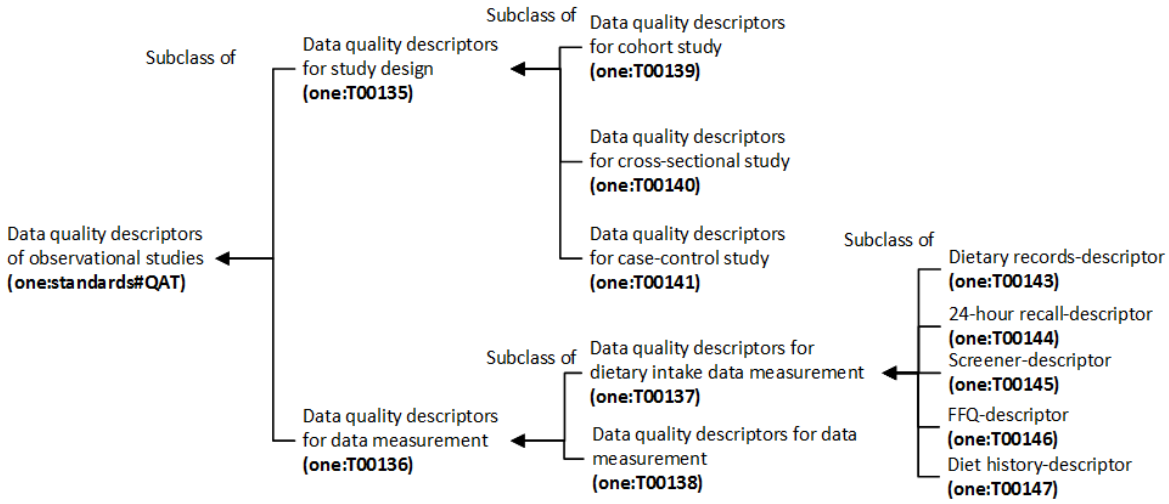


Figure 4. The ontology taxonomy of data quality descriptors of observational studies

3.2.2 STROBE-nut reporting guidelines in nutritional epidemiology

For the collection of ontology terms for STROBE-nut reporting guidelines, the STROBE reporting items (e.g. title, abstract, etc.) were used as a taxonomic hierarchy of terms. The specific STROBE-nut recommendations were arranged under their corresponding STROBE reporting items (Figure 5).

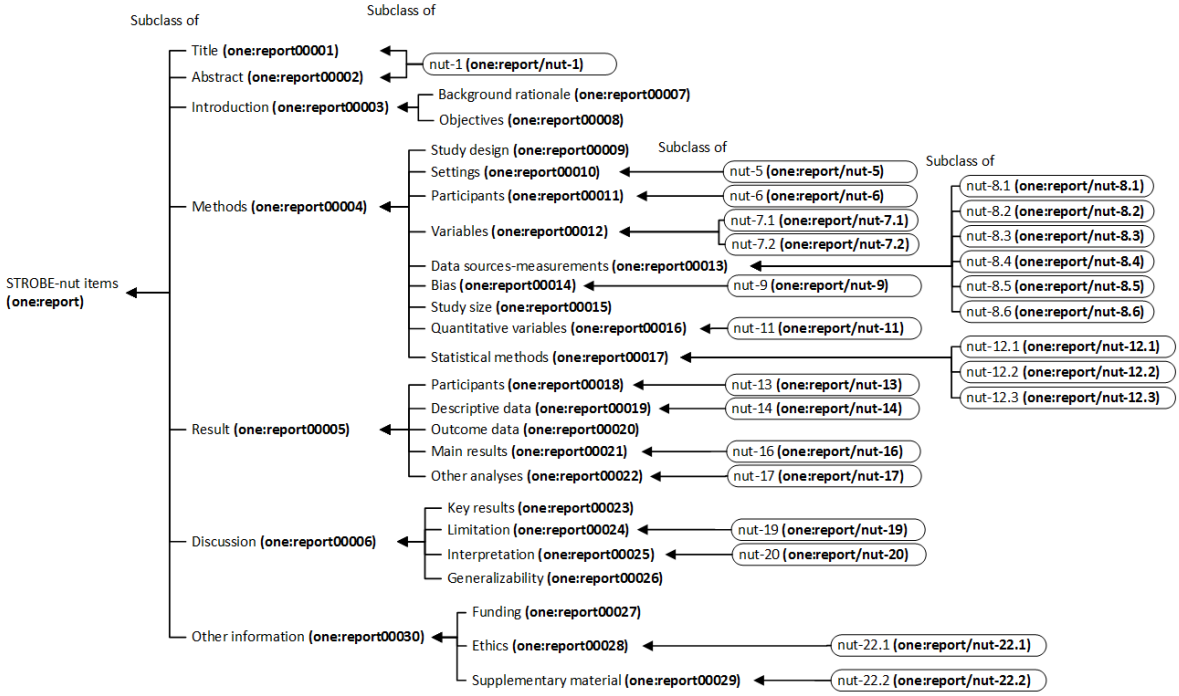


Figure 5. The ontology taxonomy of STROBE-nut items

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

3.2.3 Nutritional epidemiologic terms

ONE terms and their taxonomic hierarchy are shown in Table 1. The taxonomical hierarchy also includes relevant terms of other ontologies. The present ontology has concepts related to dietary assessment method, dietary assessment questionnaire, dietary data validation, dietary data processing and dietary data quality descriptions.

Table 1. Hierarchical structure of nutritional epidemiologic terms

1st hierarchical level	2nd hierarchical level	3rd hierarchical level
Dietary assessment method (one:ne00001)	Dietary records (one:ne00002)	Dietary record: short term (one:00042) Dietary record :Long term weighted (>7 days) (one:ne00043) Dietary records: PDA-technologies (one:ne00007) Dietary records: Mobile phone-based technologies (one:ne00008) Dietary records: Camera-recorder–based technologies (one:ne00009) Dietary records: Tape-recorder–based technologies (one:ne00010)
	24-Hour Recall (one:ne00003)	24-Hour Recall: Interactive computer-based technologies (one: 00011) 24-Hour Recall: Interactive web-based technologies (one: 00012)
	Screener (one:ne00004)	Screener: Interactive computer-based technologies (one:ne00013) Screener: Interactive web-based technologies (one:ne00014) Screener: qualitative (only frequency) (one:ne00015) Screener: semi-quantitative (one:ne00016) Screener: quantitative (one:ne00017)
	Food Frequency Questionnaire (one:ne00005)	FFQ: Interactive computer-based technologies (one:ne00018) FFQ: Interactive web-based technologies (one:ne00019) FFQ: qualitative (only frequency) (one:ne00020) FFQ: semi-quantitative (one:ne00021) FFQ: quantitative (one:ne00022)
	Diet History (one:ne00006)	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

1st hierarchical level	2nd hierarchical level	3rd hierarchical level
Dietary intake data (one:ne00023)	Unadjusted data (preferred option) (one:ne00024) Adjusted data for total energy intake using density method (one:ne00025) Adjusted data for total energy intake using residual method (one:ne00026) Estimates of usual intake from short-term measurements (one:ne00027)	
(External upper level: Administration (NCIT:C25409)) Dietary assessment administration (one:ne00028)	Proxy-administered (one:ne00029) Self-administered not verified by interviewer (one:ne00030) Self-administered and checked by interviewer (one:ne00031) Interview-administered (one:ne00032) Interview-administered using AMPM (one:ne00033)	
(External upper level: Questionnaire (NCIT_C17048)) Dietary assessment questionnaire (one:ne00034)	Self-developed questionnaires (one:ne00035) Use of standardized questionnaire (one:ne00036) Adopted other Questionnaires (one:ne00037)	
(External upper level: Content validity (NCIT_C78690)) Content validity of dietary assessment questionnaire (one:ne00038)	Verified content validity in another population (one:ne00039) Verified content validity in a comparable population in terms of both age and dietary habits (one:ne00040)	
Reference of dietary assessment questionnaire validation (one:ne00041)	Dietary assessment methods (one:ne00001)	
	Objective methods (one:ne00044)	Biomarker of dietary intake (one:ne00045)
Validated information (OBI_0302838)	Properties of dietary assessment questionnaire (one:ne00047)	Inter-rater reliability (NCIT_C78688)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

1st hierarchical level	2nd hierarchical level	3rd hierarchical level
validated information of dietary assessment questionnaire (one:ne00046)	Frequency options to identify between-person variations (one:ne00048)	
	Food items lead to underestimated target nutrients intake (one:ne00049)	
Validation type for dietary assessment questionnaire (one:ne00050)	Concurrent validity (OBCS_0000160) precision (NCIT_C48045)	
Quantification of portion sizes (one:ne00051)	Not quantified (one:ne00052) Standard portion sizes without aids (one:ne00053) Standard portion sizes with aids (one:ne00054) Portion sizes are assessed digitally but not verified by trained staff (one:ne00055) Portion sizes are assessed digitally and verified by trained staff (or packaging) (one:ne00056)	
Portion size of dietary intake data (one:ne00057)	directly expressed portion size (one:ne00058) converted portion size (one:ne00059) unconverted portion size (one:ne00060)	
Matched consumed food to referred food composition data (one:ne00060)	exact matching (one:ne00061) matched to means of min. 3 food items (one:ne00062) matched to same food items with similar moisture content (one:ne00063) matched to a different food (one:ne00064) Percentage in xsd:decimal	
Representativeness of the week/weekend days (one:ne00065)	Weekend (NCIT_C137684) Weekday (NCIT_C86936)	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

1st hierarchical level	2nd hierarchical level	3rd hierarchical level
Number of recall/measurement days per individual (one:ne00066)	xsd:integer	
Selection of recall/measurement days (one:ne00067)	Convenience selection (one:ne00068) Consecutive days (one:ne00069) Non-consecutive, non-random days (one:ne00070) Randomly over the week (one:ne00071)	
The time of diet records (one:ne00072)	Not during eating occasions nor immediately after (one:ne00073) Immediately after eating occasion (one:ne00074) During eating occasion (one:ne00075)	
Food quantification method (one:ne00076)	Food quantification method tailored to the characteristics of the population (one:ne00077) Food quantification method not specifically tailored to the characteristics of the population (one:ne00078)	

3.3 *Application* of ONE

Case study 1: study annotation and term query

The annotations for a manuscript [28] and its dataset collected in Cameroon [40] are presented in Tables A4 and A5, respectively. By using ONE terms (e.g. "Study name", "Study objective", "Study population", etc.) to build the relationships between the manuscript/dataset and its meta-data, the manuscript/dataset is annotated according to the data standards and STROBE-nut reporting guidelines included in ONE. Applying ONE avoids confusion when annotating the manuscript and dataset since all term definitions are available. This facilitates the correct understanding by annotators and users of annotated manuscripts and datasets.

Case study 2: ontology-based inference

Using the annotation in case study 1, the potential for ontology-based inferencing is presented in Table A6. Using "Country", "Study" and "Method" as relations between the manuscript and its meta-data, the manuscript is annotated as "a cross-sectional study collecting data in Cameroon by 24-hour recall method". The annotation is inferred to a more generic annotation through the taxonomies of terms in the United States National Library of Medicine Medical Subject Headings (MeSH) and ONE ontology. The upper level terms of "MeSH:D002163", "MeSH:D03430" and "one:ne00003" are "MeSH:D000350", "MeSH:D016021" and "one:ne00001" (second column), respectively. According to the labels of the three upper level terms, the inferred information (third column) is obtained: "this is an epidemiologic study collecting data in Central Africa by dietary assessment method". The ontology inference now enables integration and a wider search of data. For

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

example, when searching information annotated for "Central Africa", the present data from "Cameroon" will be identifiable.

Case study 3: estimation of reporting completeness in a sample of nine manuscripts

The STROBE-nut annotation of nine manuscripts is put under ONE class "Case studies: study description" [41]. Table A7-a counts the number of STROBE-nut items described in each manuscript, while Table A7-b reports the frequency of each STROBE-nut item reported in the nine manuscripts. Additional details on the hierarchy of annotation is available in Table A7-c. For instance, Mills, Brown [31] indicates three STROBE-nut items (i.e. Nut-13, Nut-14 and Nut-16) that were reported in the "methods section", instead of the "results section" of manuscripts as recommended by STROBE-nut.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

4. Discussion

We reviewed existing ontologies to identify terms for annotating nutritional epidemiologic research output. Ontology terms were collected to describe the minimal information needed to annotate and link research outputs such as data, manuscript and study protocols to facilitate study identification, retrieval, integration and re-use.

To date, an ontology for study level description in nutrition epidemiology was missing. The present work adds value to the Cochrane PICO (i.e. Patient, Population or Problem; Intervention, Comparison, Outcome) ontology [5], which is being developed to formulate research questions, search and characterize clinical studies as well as meta-analyses. ONE complements the work of GODAN [42], LanguaL [16] and FoodEx2 [14] initiatives, which have focused on terms for food items and their properties.

To our knowledge, it is the first time that an ontology is developed based on manuscript reporting guidelines such as STROBE-nut [43]. Reporting guidelines are widely applied and endorsed by journals as tools to improve completeness of reporting in biomedical research, to enable easier searching, filtering and navigation of research findings for further policy, practice or research [44, 45]. However, reporting guidelines have remained a paper-based initiative for manuscripts. The conversion into a machine-readable representation could expand the use of reporting guidelines to searching and inferring. Converting other research reporting guidelines such as CONSORT [46], PRISMA [47] into ontologies, would significantly improve the scope of their application. For instance, assessment of reporting completeness has remained a manual and ad-hoc exercise and has only been attempted in a handful of cases [48-50]. The application of ontologies could potentially be used for automatic monitoring of reporting completeness of manuscripts. It would enable identification of frequently and rarely reported STROBE-nut items and where they are applied in the manuscripts, and as such provide insights to update the standards [51]. Other potential applications of ontologies for research output are monitoring of trends in research and identification of neglected areas, as shown in the use of the Gene Ontology for genetic research [52]. Similar applications are useful for recommendations for minimal data requirements and data quality descriptions.

To update ONE, automated processes will be required [13]. “Ontology learning”, a process where machines are thought by humans how to build ontologies from text, provide useful prospects in this regard [53]. Ontology learning from text has been demonstrated earlier [54]. For instance, Arguello Casteleiro, Demetriou [55] have applied deep learning to extract a cardiovascular disease ontology from biomedical literature. However, considerable technical challenges remain and sustained effort by nutritionists and machine learning expertise will be required.

Development of user-friendly applications of ontology-based annotation will be required to apply ONE and minimize the burden of ongoing work by researchers. To date, most researchers in nutritional epidemiology are unfamiliar with using ontologies. Further ontology development in nutritional epidemiology will require the contribution of researchers working in multiple research areas. Additional training and capacity building efforts are needed to ensure uptake and ownership by the nutrition research community. Ad-hoc training sessions have been organized previously [56], but will require further development and integration in academic curricula.

The strength of the current work is the use of existing standards and recommendations that are developed for nutrition research [44, 57]. Those standards are developed by and used in the nutrition research community and ensure validity of ONE in the wider research community. Existing ontologies were reviewed as a preparation for converting the existing standards into an ontology. The review is a useful resource for researchers and ontology developers in nutritional epidemiology. However, some of the reviewed ontologies, did not contain terms that were essential for ONE and consequent ontology-based inferring.

The reviewed existing ontologies, including ONE, are not yet able to annotate all aspects in nutritional epidemiology. For example, an ontology to connect dietary intake data to food nutrition

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at *Nutrients* **2019**, *11*, 1300; doi:10.3390/nu11061300

composition data based on international/local food composition tables is still missing. To enable ontology applications in nutritional epidemiology, contributions are still required from researchers working on multiple research areas.

Five reviewed ontologies (i.e. Ontology of Physical Exercises (OPE) [58]; Randomized Controlled Trials Ontology (RCTONT) [59]; Non-Randomized Controlled Trials Ontology (NONRCTO) [60], Immune Disorder Ontology (IMMDIS) [61] and Neglected Tropical Disease Ontology (NTDO) [62]) however, contained errors in the formats and could not be assessed. Identifying these data gaps is hopefully an incentive to address the missing elements.

5. Conclusions

To conclude, this study introduced a comprehensive ontology for reporting nutritional epidemiologic studies and data. Application of ONE will enable monitoring of reporting completeness in the biomedical literature. Ultimately, the generated ontologies should be integrated with other linked data and applied in data collection tools, text editors, journal submission systems or data repositories for convenient and scalable search, quality checking, etc.

Supplementary Materials: N/A

Author Contributions: Conceptualization, Bernard De Baets and Patrick Kolsteren; Investigation, Chen Yang and Henry Ambayo; Methodology, Chen Yang and Carl Lachat; Software, Chen Yang, Nattapon Thanintorn and Filip Pattyn; Supervision, Filip Pattyn and Carl Lachat; Writing – original draft, Chen Yang and Carl Lachat; Writing – review & editing, Chen Yang, Henry Ambayo, Bernard De Baets, Patrick Kolsteren, Nattapon Thanintorn, Dana Hawwash, Jildau Bouwman, Antoon Bronselaer, Filip Pattyn and Carl Lachat.

Funding: the European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI) is part of the Joint Programming Initiative (JPI) "A healthy diet for a healthy life" and is funded by national funding agencies in 9 European countries. The work in this manuscript is funded by the FWO Research Foundation - Flanders, grant number G0D4815N. CY is funded by a scholarship from the Chinese Scholarship Council. CL received funding from Bioversity Int for the work on standards for dietary assessment.

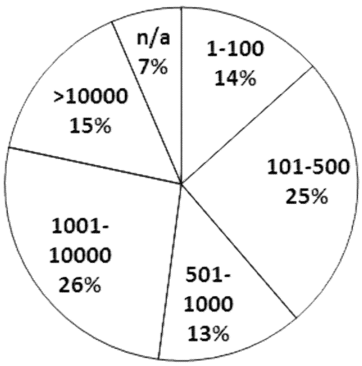
Acknowledgments: We would like to thank Giles Hanley-Cook for the language review.

Conflicts of Interest: The authors declare no conflict of interest.

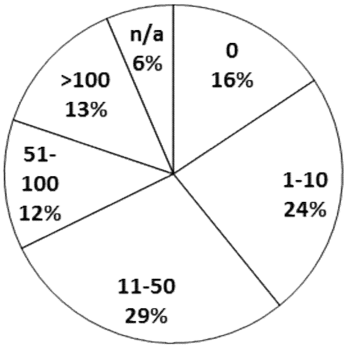
Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

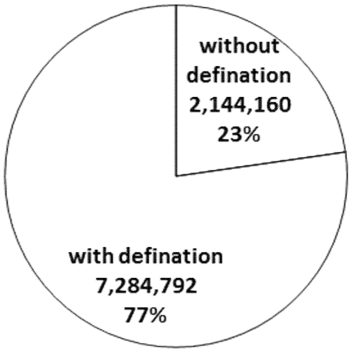
Appendix A



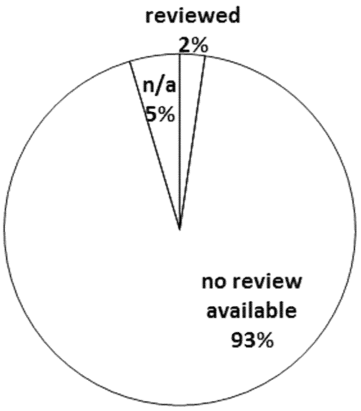
(a) Comprehensibility module: values are the sum of terms per ontology



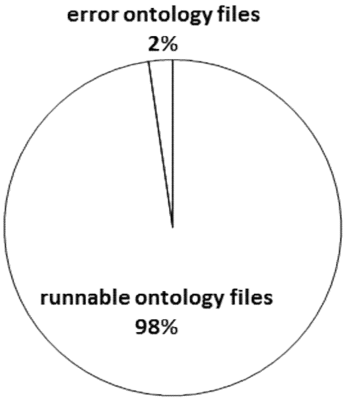
(b) Richness module: values are the sum of properties per ontology



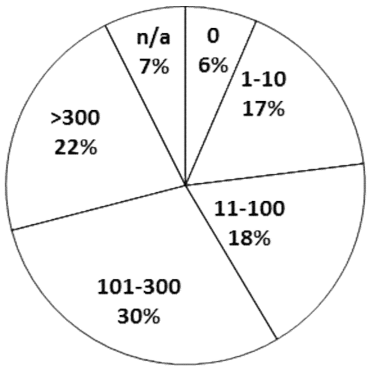
(c) Interpretability & clarity module: values are the number of terms with/without definition



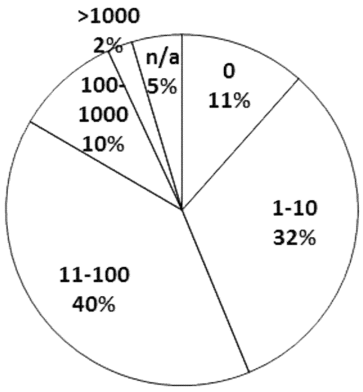
(d) Consistency & accuracy module: reviewed ontologies



(e) Lawfulness module: values are the percentage of runnable ontology files



(f) Authority module: values are the number of mapped ontologies of each ontology



(g) History module: values are everage visits/month of each ontology during Feb,Mar and Apr, 2018

Figure A1. Quality characteristics of selected ontologies for nutritional epidemiology

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Table A1. Scope of nutritional epidemiology's controlled vocabulary for achieving FAIR principle

FAIR principle	Requires controlled vocabulary on		Applications
	Data-level	Metadata-level	
Findable Reusable	Food, nutrients, disease & special population, supplementary data, data management,	Research terminology, metadata representation, supplementary metadata	Data search
			Data integration

Table A2. Classification of selected ontologies according to the scope of ONE (complete list)

Food & nutrient (n=33)
Food/Dietary agricultural product (n=29):
Barley Trait Dictionary ontology (https://www.ebi.ac.uk/ols/ontologies/co_323);
Brassica ontology (https://www.ebi.ac.uk/ols/ontologies/co_348);
Cassava ontology (https://www.ebi.ac.uk/ols/ontologies/co_334);
Castor bean ontology (https://www.ebi.ac.uk/ols/ontologies/co_347);
Chickpea ontology (https://www.ebi.ac.uk/ols/ontologies/co_338);
Common bean ontology (https://www.ebi.ac.uk/ols/ontologies/co_335);
Cowpea ontology (https://www.ebi.ac.uk/ols/ontologies/co_340);
Fish Ontology (FISHO) (https://bioportal.bioontology.org/ontologies/FISHO);
Groundnut ontology (https://www.ebi.ac.uk/ols/ontologies/co_337);
Lentil ontology (https://www.ebi.ac.uk/ols/ontologies/co_339);
Livestock Product Trait Ontology (LPT) (https://bioportal.bioontology.org/ontologies/LPT);
Maize ontology (https://www.ebi.ac.uk/ols/ontologies/co_322);
Mungbean ontology (https://www.ebi.ac.uk/ols/ontologies/co_346);
Natural Products Ontology (NATPRO) (https://bioportal.bioontology.org/ontologies/NATPRO);
Oat ontology (https://www.ebi.ac.uk/ols/ontologies/co_350);
Pearl millet ontology (https://www.ebi.ac.uk/ols/ontologies/co_327);
Pigeonpea ontology (https://www.ebi.ac.uk/ols/ontologies/co_341);
Potato ontology (https://www.ebi.ac.uk/ols/ontologies/co_330);
Rice ontology (https://www.ebi.ac.uk/ols/ontologies/co_320);
Sorghum ontology (https://www.ebi.ac.uk/ols/ontologies/co_324);
Soy Ontology (SOY) (https://bioportal.bioontology.org/ontologies/SOY);
Soybean ontology (https://www.ebi.ac.uk/ols/ontologies/co_336);
Sugar Kelp trait ontology (https://www.ebi.ac.uk/ols/ontologies/co_360);
Sweet Potato ontology (https://www.ebi.ac.uk/ols/ontologies/co_331);
Vitis ontology (https://www.ebi.ac.uk/ols/ontologies/co_356);
Wheat ontology (https://www.ebi.ac.uk/ols/ontologies/co_321);
Yam ontology (https://www.ebi.ac.uk/ols/ontologies/co_343);
FOODON (https://bioportal.bioontology.org/ontologies/FOODON);
OntoFood (OF) (https://bioportal.bioontology.org/ontologies/OF);
Nutrients/chemical compounds (n=4):
Amino Acid Ontology (AMINO-ACID) (https://bioportal.bioontology.org/ontologies/AMINO-ACID);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Lipid Ontology (LIPRO) (https://bioportal.bioontology.org/ontologies/LIPRO); Protein Ontology (PR) (https://bioportal.bioontology.org/ontologies/PR); Chemical Entities of Biological Interest (ChEBI) (https://bioportal.bioontology.org/ontologies/CHEBI).
Disease & Special population (n=86)
Computer Assisted Brain Injury Rehabilitation Ontology (CABRO) (https://bioportal.bioontology.org/ontologies/CABRO); Computer-Based Patient Record Ontology (CPRO) (https://bioportal.bioontology.org/ontologies/CPRO); Allergy Detector II (ALLERGYDETECTOR) (https://bioportal.bioontology.org/ontologies/ALLERGYDETECTOR); Alzheimer's disease ontology (ADO) (https://bioportal.bioontology.org/ontologies/ADO); Asthma Ontology (AO) (https://bioportal.bioontology.org/ontologies/AO); Autism DSM-ADI-R ontology (ADAR) (https://bioportal.bioontology.org/ontologies/ADAR); Bilingual Ontology of Alzheimer's Disease and Related Diseases (ONTOAD) (https://bioportal.bioontology.org/ontologies/ONTOAD); BioMedBridges Diabetes Ontology (DIAB) (https://bioportal.bioontology.org/ontologies/DIAB); Bleeding History Phenotype Ontology (BHO) (https://bioportal.bioontology.org/ontologies/BHO); Breast Cancer Grading Ontology (BCGO) (https://bioportal.bioontology.org/ontologies/BCGO); Cancer Research and Management ACGT Master Ontology (ACGT-MO) (https://bioportal.bioontology.org/ontologies/ACGT-MO); Cardiovascular Disease Ontology (www.obofoundry.org/ontology/cvdo.html); Chronic Kidney Disease Ontology (CKDO) (https://bioportal.bioontology.org/ontologies/CKDO); Cigarette Smoke Exposure Ontology (CSEO) (https://bioportal.bioontology.org/ontologies/CSEO); Congenital Heart Defects Ontology (CHD) (https://bioportal.bioontology.org/ontologies/CHD); COPD Ontology (COPDO) (https://bioportal.bioontology.org/ontologies/COPDO); Dengue Fever Ontology (IDODEN) (https://bioportal.bioontology.org/ontologies/IDODEN); Dermatology Lexicon (DERMLEX) (https://bioportal.bioontology.org/ontologies/DERMLEX); Diabetes Mellitus Diagnosis Ontology (DDO) (https://bioportal.bioontology.org/ontologies/DDO); Diabetes Mellitus Treatment Ontology (DMTO) (https://bioportal.bioontology.org/ontologies/DMTO); Diagnosis Ontology of Clinical Care Classification (DOCCC) (https://bioportal.bioontology.org/ontologies/DOCCC); Diagnostic Ontology (DIAGONT) (https://bioportal.bioontology.org/ontologies/DIAGONT); Disease core ontology applied to Rare Diseases (HRDO) (https://bioportal.bioontology.org/ontologies/HRDO); Disorders cluster (APADISORDERS) (https://bioportal.bioontology.org/ontologies/APADISORDERS); Dispedia Core Ontology (DCO) (https://bioportal.bioontology.org/ontologies/DCO); Eligibility Feature Hierarchy (ELIG) (https://bioportal.bioontology.org/ontologies/ELIG); EmpowerBP (EBP) (https://bioportal.bioontology.org/ontologies/EBP); Environment Ontology (ENVO) (https://bioportal.bioontology.org/ontologies/ENVO); Epilepsy and Seizure Ontology (EPSO) (https://bioportal.bioontology.org/ontologies/EPSO); Family Health History Ontology (FHHO) (https://bioportal.bioontology.org/ontologies/FHHO); Fanconi Anemia Ontology (IFAR) (https://bioportal.bioontology.org/ontologies/IFAR); Glioblastoma (GBM) (https://bioportal.bioontology.org/ontologies/GBM);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Health Level Seven Reference Implementation Model, Version 3 (HL7) (https://bioportal.bioontology.org/ontologies/HL7);
Heart Failure Ontology (HFO) (https://bioportal.bioontology.org/ontologies/HFO);
HIV ontology (HIV) (https://bioportal.bioontology.org/ontologies/HIV);
Holistic Ontology of Rare Diseases (HORD) (https://bioportal.bioontology.org/ontologies/HORD);
Human Dermatological Disease Ontology (DERMO) (https://bioportal.bioontology.org/ontologies/DERMO);
Infectious Disease Ontology (IDO) (https://bioportal.bioontology.org/ontologies/IDO);
Influenza Ontology (FLU) (https://bioportal.bioontology.org/ontologies/FLU);
International Classification of Wellness (ICW) (https://bioportal.bioontology.org/ontologies/ICW);
Malaria Ontology (https://bioportal.bioontology.org/ontologies/IDOMAL);
Mental Functioning Ontology (MF) (https://bioportal.bioontology.org/ontologies/MF);
MFO Mental Disease Ontology (MFOMD) (https://bioportal.bioontology.org/ontologies/MFOMD);
Monarch Disease Ontology (MONDO) (https://bioportal.bioontology.org/ontologies/MONDO);
Multiple sclerosis ontology (MSO) (https://bioportal.bioontology.org/ontologies/MSO);
National Institutes of Health Stroke Scale Ontology (NIHSS) (https://bioportal.bioontology.org/ontologies/NIHSS);
NCCN EHR Oncology Categories (NCCNEHR) (https://bioportal.bioontology.org/ontologies/NCCNEHR);
Neomark Oral Cancer Ontology, version 3 (NEOMARK3) (https://bioportal.bioontology.org/ontologies/NEOMARK3);
Neomark Oral Cancer Ontology, version 4 (NEOMARK4) (https://bioportal.bioontology.org/ontologies/NEOMARK4);
Obstetric and Neonatal Ontology (ONTONEO) (https://bioportal.bioontology.org/ontologies/ONTONEO);
Ontological Knowledge Base Model for Cystic Fibrosis (ONTOKBCF) (https://bioportal.bioontology.org/ontologies/ONTOKBCF);
Ontology for BIoBanking (OBIB) (https://bioportal.bioontology.org/ontologies/OBIB);
Ontology of amyotrophic lateral sclerosis, social module (ONTOPARON_SOCIAL) (https://purl.bioontology.org/ontology/ONTOPARON_SOCIAL);
Ontology of Craniofacial Development and Malformation (OCDM) (https://bioportal.bioontology.org/ontologies/OCDM);
Ontology of Glucose Metabolism Disorder (OGMD) (https://www.bioportal.bioontology.org/ontologies/OGMD);
Ontology of Pneumology (ONTOPNEUMO) (https://bioportal.bioontology.org/ontologies/ONTOPNEUMO);
Orphanet Rare Disease Ontology (ORDO) (https://bioportal.bioontology.org/ontologies/ORDO);
Parkinson's Disease Ontology (PDON) (https://bioportal.bioontology.org/ontologies/PDON);
Pathogenic Disease Ontology (PDO) (https://bioportal.bioontology.org/ontologies/PDO);
Pre-eclampsia Ontology (PE-O) (https://bioportal.bioontology.org/ontologies/PE-O);
Pulmonary Embolism Ontology (PE) (https://bioportal.bioontology.org/ontologies/PE);
RegenBase ontology (RB) (https://bioportal.bioontology.org/ontologies/RB);
Sickle Cell Disease Ontology (SCDO) (https://bioportal.bioontology.org/ontologies/SCDO);
Spinal Cord Injury Ontology (SCIO) (https://bioportal.bioontology.org/ontologies/SCIO);
The Oral Health and Disease Ontology (OHD) (https://bioportal.bioontology.org/ontologies/OHD);
Anthology of Biosurveillance Diseases (ABD) (https://bioportal.bioontology.org/ontologies/ABD);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Children's Health Exposure Analysis Resource (CHEAR) (https://bioportal.bioontology.org/ontologies/CHEAR);
Codificacion De Enfermedades Pediatricas (En Edición) (CEI_10) (https://bioportal.bioontology.org/ontologies/CEI_10);
Human Disease Ontology (DOID) (https://bioportal.bioontology.org/ontologies/DOID/);
International Classification of Diseases, Version 10 - Clinical Modification (ICD10CM) (https://bioportal.bioontology.org/ontologies/ICD10CM);
International Classification of Diseases, Version 10 - Procedure Coding System (ICD10PCS) (https://bioportal.bioontology.org/ontologies/ICD10PCS);
International Classification of Diseases, Version 10 (ICD10) (https://bioportal.bioontology.org/ontologies/ICD10);
International Classification of Diseases, Version 9 - Clinical Modification (ICD9CM) (https://bioportal.bioontology.org/ontologies/ICD9CM);
International Classification of External Causes of Injuries (ICECI) (https://bioportal.bioontology.org/ontologies/ICECI);
International Classification of Primary Care - 2 PLUS (ICPC2P) (https://bioportal.bioontology.org/ontologies/ICPC2P);
International Classification of Primary Care (ICPC) (https://bioportal.bioontology.org/ontologies/ICPC);
International Classification of Wellness (ICW) (https://bioportal.bioontology.org/ontologies/ICW);
Online Mendelian Inheritance in Man (OMIM) (https://bioportal.bioontology.org/ontologies/OMIM);
Regional Healthcare System Interoperability and Information Exchange Measurement Ontology (HEIO) (https://bioportal.bioontology.org/ontologies/HEIO);
STO (CVAO) (https://bioportal.bioontology.org/ontologies/CVAO);
Student Health Record Ontology (SHR) (https://bioportal.bioontology.org/ontologies/SHR);
Symptom Ontology (SYMP) (https://bioportal.bioontology.org/ontologies/SYMP);
Taxonomy for Rehabilitation of Knee Conditions (TRAK) (https://bioportal.bioontology.org/ontologies/TRAK);
Upper-Level Cancer Ontology (CANONT) (https://bioportal.bioontology.org/ontologies/CANONT);
Immune Disorder Ontology (IMMDIS) (inaccessible) (https://bioportal.bioontology.org/ontologies/IMMDIS);
Neglected Tropical Disease Ontology (NTDO) (inaccessible). (https://bioportal.bioontology.org/ontologies/NTDO)
Data management (n=21)
Bioinformatics operations, data types, formats, identifiers and topics (EDAM) (https://bioportal.bioontology.org/ontologies/EDAM);
Comparative Data Analysis Ontology (CDAO) (https://bioportal.bioontology.org/ontologies/CDAO);
Computer Retrieval of Information on Scientific Projects Thesaurus (CRISP) (https://bioportal.bioontology.org/ontologies/CRISP);
Mathematical modeling ontology (MAMO) (https://bioportal.bioontology.org/ontologies/MAMO);
Ontology of Core Data Mining Entities (ONTODM-CORE) (https://bioportal.bioontology.org/ontologies/ONTODM-CORE);
Ontology of Data Mining Investigations (ONTODM-KDD) (https://bioportal.bioontology.org/ontologies/ONTODM-KDD);
Confidence Information Ontology (CIO) (https://bioportal.bioontology.org/ontologies/CIO);
Data Collection Ontology (GDCO) (https://bioportal.bioontology.org/ontologies/GDCO);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

SMASH (Semantic Mining of Activity, Social, and Health data) Ontology (SMASH) (https://bioportal.bioontology.org/ontologies/SMASH); The Data Use Ontology (DUO) (https://bioportal.bioontology.org/ontologies/DUO); The Statistical Methods Ontology (STATO) (https://bioportal.bioontology.org/ontologies/STATO); APA Statistical Cluster (APASTATISTICAL) (https://bioportal.bioontology.org/ontologies/APASTATISTICAL); Biomedical Informatics Research Network Project Lexicon (BIRNLEX) (https://bioportal.bioontology.org/ontologies/BIRNLEX); Data Catalog Vocabulary (DCAT) (https://bioportal.bioontology.org/ontologies/DCAT); Image and Data Quality Assessment Ontology (IDQA) (https://bioportal.bioontology.org/projects/IDQA); Ontology of Biological and Clinical Statistics (OBCS) (https://bioportal.bioontology.org/ontologies/OBCS); Probability Distribution Ontology (PROBONTO) (https://www.ebi.ac.uk/ols/ontologies/probonto); Quantities, Units, Dimensions, and Types Ontology (QUDT) (https://bioportal.bioontology.org/ontologies/QUDT); RDL (RDL) (https://bioportal.bioontology.org/ontologies/RDL); schema.org (SCHEMA) (https://bioportal.bioontology.org/ontologies/SCHEMA); Semantic DICOM Ontology (SEDI) (https://bioportal.bioontology.org/ontologies/SEDI)
Research terminology (n=32)
Bionutrition Ontology (BNO) (https://bioportal.bioontology.org/ontologies/BNO); Clinical Measurement Ontology (CMO) (https://bioportal.bioontology.org/ontologies/CMO); Clinical Signs and Symptoms Ontology (CSSO) (https://bioportal.bioontology.org/ontologies/CSSO); Clinical Study Ontology (CSO) (https://bioportal.bioontology.org/ontologies/CSO); Clinical Trials Ontology (CTO) (https://bioportal.bioontology.org/ontologies/CTO); EDDA Study Designs Taxonomy (EDDA) (https://bioportal.bioontology.org/ontologies/EDDA); Epidemiology Ontology (https://www.ebi.ac.uk/ols/ontologies/epo); Mass spectrometry ontology (https://bioportal.bioontology.org/ontologies/MS); Non-Pharmacological Interventions (NPIs/NPI) (https://bioportal.bioontology.org/ontologies/NPI); Ontology for Nutritional Studies (ONS) (https://bioportal.bioontology.org/ontologies/ONS); Ontology of Clinical Research (OCRE) (https://bioportal.bioontology.org/ontologies/OCRE); SMART Protocols (SP) (https://bioportal.bioontology.org/ontologies/SP); Biomedical Research Integrated Domain Group Model (BRIDG) (https://bioportal.bioontology.org/ontologies/BRIDG); Biomedical Resource Ontology (BRO) (https://bioportal.bioontology.org/ontologies/BRO); Biomedical Topics (BMT) (https://bioportal.bioontology.org/ontologies/BMT); Current Procedural Terminology (CPT) (https://bioportal.bioontology.org/ontologies/CPT); eagle-i resource ontology (ERO) (https://bioportal.bioontology.org/ontologies/ERO); Experimental Conditions Ontology (XCO) (https://bioportal.bioontology.org/ontologies/XCO); Experimental Factor Ontology (EFO) (https://bioportal.bioontology.org/ontologies/EFO); Medical Subject Headings (MESH) (https://bioportal.bioontology.org/ontologies/MESH); MedlinePlus Health Topics (MEDLINEPLUS) (https://bioportal.bioontology.org/ontologies/MEDLINEPLUS); National Cancer Institute Thesaurus (NCIT) (https://bioportal.bioontology.org/ontologies/NCIT); Ontology for Biomedical Investigation (https://bioportal.bioontology.org/ontologies/OBI); Ontology for General Medical Science (OGMS) (https://bioportal.bioontology.org/ontologies/OGMS);

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Read Clinical Terminology Version 2 (RCTV2) (https://bioportal.bioontology.org/ontologies/RCTV2); Robert Hoehndorf Version of MeSH (RH-MESH) (https://bioportal.bioontology.org/ontologies/RH-MESH); SNOMED CT (SNOMEDCT) (https://bioportal.bioontology.org/ontologies/SNOMEDCT); Read Codes, Clinical Terms Version 3 (CTV3) (RCD) (https://bioportal.bioontology.org/ontologies/RCD); CARRE Risk Factor ontology (CARRE) (https://bioportal.bioontology.org/ontologies/CARRE); Ontology of Physical Exercises (OPE) (https://bioportal.bioontology.org/ontologies/OPE) (inaccessible); Randomized Controlled Trials Ontology (RCTONT) (https://bioportal.bioontology.org/ontologies/RCTONT) (inaccessible); Non-Randomized Controlled Trials Ontology (NONRCTO) (https://bioportal.bioontology.org/ontologies/NONRCTO) (inaccessible).
Metadata representation (n=0)
Supplementary (meta)data (n=44)
VIVO Ontology for Researcher Discovery (VIVO) (https://bioportal.bioontology.org/ontologies/VIVO); Human Ancestry Ontology (HANCESTRO) (https://bioportal.bioontology.org/ontologies/HANCESTRO); APA Occupational and Employment cluster (APAOCUEMPLOY) (https://bioportal.bioontology.org/ontologies/APAOCUEMPLOY); EDDA Publication Types Taxonomy (EDDA_PT) (https://bioportal.bioontology.org/ontologies/EDDA_PT); Ethnicity Ontoloy (EO) (https://bioportal.bioontology.org/ontologies/EO); Geographical Entity Ontology (GEO) (https://bioportal.bioontology.org/ontologies/GEO); Informed Consent Ontology (ICO) (https://bioportal.bioontology.org/ontologies/ICO); Ontology of Geographical Region (OGR) (https://bioportal.bioontology.org/ontologies/OGR); Provenance Ontology (PROVO) (https://bioportal.bioontology.org/ontologies/PROVO); Scientific Evidence and Provenance Information Ontology (SEPIO) (www.obofoundry.org/ontology/sepio.html); Time Event Ontology (TEO) (https://bioportal.bioontology.org/projects/TEO); BioPortal Metadata Ontology (BP-METADATA) (https://bioportal.bioontology.org/ontologies/BP-METADATA); Evidence and Conclusion Ontology (ECO) (https://bioportal.bioontology.org/ontologies/ECO); Gazetteer (https://bioportal.bioontology.org/ontologies/GAZ); OBO Relations Ontology (https://bioportal.bioontology.org/ontologies/OBOREL); Ontology Metadata Vocabulary (OMV) (https://bioportal.bioontology.org/ontologies/OMV); Ontology of Medically Related Social Entities (OMRSE) (https://bioportal.bioontology.org/ontologies/OMRSE); Provenance, Authoring and Versioning (PAV) (https://bioportal.bioontology.org/ontologies/PAV); PLOS Thesaurus (PLOSTHES) (https://bioportal.bioontology.org/ontologies/PLOSTHES); Population and Community Ontology (PCO) (https://bioportal.bioontology.org/ontologies/PCO); Role Ontology (ROLEO) (https://bioportal.bioontology.org/ontologies/ROLEO); Basic Formal Ontology (BFO) (https://bioportal.bioontology.org/ontologies/BFO); BIBFRAME 2.0 (BIBFRAME) (https://bioportal.bioontology.org/ontologies/BIBFRAME); CEDAR Value Sets (CEDARVS) (https://bioportal.bioontology.org/ontologies/CEDARVS); Contributor Role Ontology (ROLEO) (https://bioportal.bioontology.org/ontologies/ROLEO); DC Terms (DCT) (https://bioportal.bioontology.org/ontologies/DCT); DCMI Metadata Terms: properties in /terms/ namespace (DCTERMS)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

(https://bioportal.bioontology.org/ontologies/dcterms);
DCMI Terms (DCMI) (https://bioportal.bioontology.org/ontologies/DCMI);
DCMI Type Vocabulary (DCMITYPE) (https://bioportal.bioontology.org/ontologies/DCMITYPE);
Dublin Core (DC) (https://bioportal.bioontology.org/ontologies/DC);
Dublin Core Collection Description Frequency Vocabulary (DCCDFV)
(https://bioportal.bioontology.org/ontologies/DCCDFV);
General Formal Ontology (GFO) (https://bioportal.bioontology.org/ontologies/GFO);
General Formal Ontology for Biology (GFO-BIO)
(https://bioportal.bioontology.org/ontologies/GFO-BIO);
Information Artifact Ontology (IAO) (https://bioportal.bioontology.org/ontologies/IAO);
ISO 639-2: Codes for the Representation of Names of Languages (ISO639-2)
(https://bioportal.bioontology.org/ontologies/ISO639-2);
NIH NLM Value Sets (NLMVS) (https://bioportal.bioontology.org/ontologies/NLMVS);
Ontology of Datatypes (ONTODT) (https://bioportal.bioontology.org/ontologies/ONTODT);
OWL-Time (TIME) (https://bioportal.bioontology.org/ontologies/TIME);
Semantic Types Ontology (STY) (https://bioportal.bioontology.org/ontologies/STY);
Semantic science Integrated Ontology (SIO) (https://bioportal.bioontology.org/ontologies/SIO);
Terminological and Ontological Knowledge Resources Ontology (TOK)
(https://bioportal.bioontology.org/ontologies/TOK);
vCard Ontology - for describing People and Organizations (VCARD)
(https://bioportal.bioontology.org/ontologies/VCARD);
VIVO-Integrated Semantic Framework (VIVO-ISF)
(https://bioportal.bioontology.org/ontologies/VIVO-ISF);
Bro_Name (BRO_ACRONYM) (http://bioportal.bioontology.org/ontologies/BRO_ACRONYM)

Table 3A. Metrics for quality assessment of ONE

Metrics suite	Attributes	Description	Assessment for ONE
Syntactic quality	Lawfulness	Correctness of syntax	no error detected
	Richness	Breadth of syntax used	1 defined property, but all ONE classes can be converted to properties.
Semantic quality	Interpretability	Meaningfulness of terms	terms come from well-defined guidelines.
	Consistency	Consistency of meaning of terms	no term is used in more than 1 way in the ontology.
	Clarity	Average number of word senses	Close to 1, because they are all academic terms

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Pragmatic quality	Comprehensiveness	Number of classes and properties	339 classes and 1 properties.
	Accuracy	Accuracy of information	Checked manually, no error detected.
	Relevance	Relevance of information for a task	n/a, assess in the future
Social quality	Authority	Extent to which other ontologies rely on it	n/a, assess in the future
	History	Number of times ontology has been used	n/a, assess in the future

Table 4A. Case study-Dietary species richness as a measure of food biodiversity and nutritional quality of diet (Lachat et al. 2018), study description

Preferred Name	Lachat C et al. 2018 PNAS
ID	http://one.ugent.be/standards#lachatc2018pnas
Study Name	Dietary species richness as a measure of food biodiversity and nutritional quality of diet
Study objective	Assess the intricate relationship between food biodiversity and diet quality
Study population	General population
Study terminated	06/06/2017
Study description	We applied biodiversity indicators to dietary intake data from and assessed associations with diet quality of women and young children.
age.max	43
age.min	0.5
Data analysis permission	accessible raw data
Data sharing policy	Publicly accessible
Metadata	Publicly accessible
Aggregate data sharing policy	Publicly accessible
Contact information	Carl.Lachat@UGent.be
Contact person	Lachat C (orcid)
Country	Sri Lanka Cameroon Congo Benin Vietnam Kenya Ecuador
DOI	http://doi.org/10.1073/pnas.1709194115
Epidemiologic Studies	Cross-sectional studies
Funding Organization	http://www.fwo.be/en
label	Lachat C et al. 2018 PNAS
Population Characteristics	Women

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Rural population
	Child
prefixIRI	lachatc2018pnas
prefLabel	Lachat C et al. 2018 PNAS
Principal Investigator	Lachat C (orcid)
Publications	http://www.pnas.org/content/115/1/127
Recruitment period	Benin:01/10/2013-31/12/2013,01/05/2014-31/07/2014; Cameroon:01/07/2013-31/08/2013; Congo:01/07/2009-30/09/2009; Ecuador:01/03/2011-31/03/2011; Kenya:01/09/2014-30/09/2014; 01/04/2015-30/04/2015; Sir Lanka: 01/07/2013-30/09/2013; Vietnam: 01/08/2014-31/12/2014
Sampling method	Convenience sampling
strobe-nut	nut-22.1
	nut-8.1
	nut-20
	nut-8.3
	nut-11
	nut-22.2
	nut-12.3
	nut-8.5
	nut-5
	nut-1
	nut-8.2
	nut-7.1
	nut-12.1
	nut-19
Total number of females recruited	2188
Total number of participants recruited	6226
subClassOf	Case studies: study description

Lachat C, Raneri JE, Smith KW, Kolsteren P, Van Damme P, Verzelen K, Penafiel D, Vanhove W, Kennedy G, Hunter D, et al. Dietary species richness as a measure of food biodiversity and nutritional quality of diets. Proc Natl Acad Sci U S A 2018;115(1):127-32. doi: 10.1073/pnas.1709194115.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Table 5A. Case study-Dietary species richness as a measure of food biodiversity and nutritional quality of diet (Lachat et al. 2018), Cameroon dataset description

Preferred Name	Cameroon dataset-Lachat C et al. 2018 PNAS
ID	http://one.ugent.be/standards#lachatc2018pnasCameroon
Country	Cameroon
Dietary assessment administration	Interview-administered
Dietary assessment method	24-Hour recall
Dietary assessment questionnaire	Self-developed questionnaires
Dietary intake data	Unadjusted data
Food composition table	West Africa Food Composition Table (2012), FAO
Food quantification method	Food quantification method not specifically tailored to the characteristics of the population
Health outcomes	01/07/2013-31/08/2013
label	Cameroon dataset-Lachat C et al. 2018 PNAS
Matched consumed food to referred food composition data	Exact matching Matched to a different food
Number of recall/measurement days per individual	2
Portion size of dietary intake data	Converted portion size Directly expressed portion size
prefixIRI	lachatc2018pnasCameroon
prefLabel	Cameroon dataset-Lachat C et al. 2018 PNAS
Quantification of portion sizes	Portion sizes are assessed digitally and verified by trained staff (or packaging)
Random selection	Convenience sampling
Sample representativeness	Non-representative sample
Sampling	01/07/2013-31/08/2013
Seasons	Rainy season
Selection of recall/measurement days	Non-consecutive, non-random days
The time of diet records	Not during eating occasions nor immediately after
subClassOf	Case studies: dataset description

Lachat C, Raneri JE, Smith KW, Kolsteren P, Van Damme P, Verzelen K, Penafiel D, Vanhove W, Kennedy G, Hunter D, et al. Dietary species richness as a measure of food biodiversity and nutritional quality of diets. Proc Natl Acad Sci U S A 2018;115(1):127-32. doi: 10.1073/pnas.1709194115.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Table 6A. Case study: ontology-based inferences

Annotations of Carl et al. 2018	Upper level terms according to their taxonomic hierarchies	Inferred information
Country: Cameroon (MeSH:D002163)	Africa, Central (MeSH:D000350) Cameroon (MeSH:D002163)	The study was conducted in central Africa.
Study: cross-sectional study (MeSH:D03430)	Epidemiologic studies (MeSH:D016021) cross-sectional study (MeSH:D03430)	This study is an epidemiologic study.
Method: 24-hour recall (one:ne00003)	Dietary assessment method (one:ne00001) 24-hour recall (one:ne00003)	The study used a dietary assessment method.

Table 7A. Case study: estimation of reporting completeness in a sample of nine manuscripts (reporting quality reference by using STROBE-nut terms)

(a) mapped STROBE-nut terms per manuscript

Publications	Number of STROBE-nut items (Mapped/total)
Mills et al. 2017	21/24
Abris et al. 2018	17/24
Chatelan et al. 2017	18/24
Lam et al. 2017	16/24
Llanaj et al. 2018	15/24
Arsenault et al. 2014	15/24
De Cock et al. 2016	15/24
Mills et al. 2018	14/24
Workicho et al. 2016	9/24

(b) mapping rate of each STROBE-nut term

Mapping rate (%)	Number of items	STROBE-nut items
100% mapping rate	3	1; 8.1; 19
high mapping rate (100%-75%)	9	5; 6; 7.1; 7.2; 8.5; 11; 14; 20; 22.1
medium mapping rate (75%-50%)	5	8.2; 8.6; 12.1; 12.2; 13
low mapping rate (50%-25%)	3	8.3; 9; 22.2
extreme low mapping rate (<25%)	4	8.4; 12.3; 16; 17

(c) hierarchy mapping

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at *Nutrients* **2019**, *11*, 1300; doi:10.3390/nu11061300

STROBE-nut reporting guideline	**Mills et al. 2017**
… ❖ Methods ➢ … ❖ Result ➢ Nut-13 ➢ Nut-14 ➢ Nut-16 ❖ Discussion ➢ … …	… ❖ Methods ➢ Nut-13 ➢ Nut-14 ➢ Nut-16 ❖ Result ➢ … ❖ Discussion ➢ … …

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Table 8A. Ontology view of minimal data requirement of observational studies

	Descriptors	Options
	[b, c] ISA framework-Investigation (one:T00001)	
1	Study name (NCIT_C686631)	Acronym (NCIT_C93495)
2	Country (ancestro_0003)	(ancestro ontology)
3	Study aim (NCIT_C94090)	
4	Principal Investigator (NCIT_C19924)	
5	Contact information (NCIT_C60776); contact person (NCIT_C25461)	
6	Funding Organization (VIVO_core#FundingOrganization)	
7	Upload (NCIT_C48914) URL (HL7_C1710546)	Study reference link page description (NCIT_C94131) [b] Study registration number (one:T00002) IRB-IEC Approval (CARELEX_ IRB-IEC_Approval) Informed consent (MeSH_D007258) Study protocol (NCIT_C70817) Questionnaires (NCIT_C17048) Standard Operating Procedures (SIO_000964) Publications (MeSH:D011642): Type (MeSH:D011642 subclasses), DOI (EDAM_data_1188), URL (HL7_C1710546) Other
8	Study terminated (NCIT_C70757)	DD/MM/YYYY (xsd:datetime)
9	[b, d] Data sharing policy (one:T00003)	
10	[b, d] Aggregate Data sharing policy (one:T00004)	[b, d] Publicly accessible (one:T00005) [b, d] Accessible upon request (one:T00006) [b, d] Not publicly accessible (one:T00007)
11	Metadata (MeSH: D000071253)	
12	[b, d] Data analysis permission (one:T00008)	[b, d] accessible raw data (one:T00009) [b ,d] federated analysis (one:T00010)
	[b, c] ISA framework-Study(one:T00011)	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Descriptors	Options
1	Epidemiologic Studies (MeSH_D016021)	Cohort (MeSH_D015331) Cross-sectional (MeSH_D003430) Case-control (MeSH_D016022) Seroepidemiologic study (MeSH_D016036) Other (subclasses of MeSH_D016021)
2	Study description (NCIT_C142704)	
3	Study population (NCIT_C70833)	General population (NCIT_C18241)
4	Population characteristics (MeSH_D011154)	MeSH_D011154 subclasses
5	b, e population representativeness (one:T00012)	[b, e] National level (one:T00013) [b, e] Subnational level (one:T00014) [b, e] Community level (one:T00015)
6	Type of sampling (NCIT_C71492)	Equal probability sampling method (NCIT_C71517) - [b, g] Simple Random Sampling (one:T00016), - [b, g] Stratified Random Sampling (one:T00017) - [b, g] Multi-Stage Sampling (one:T00018) Non-probability sampling (NCIT_C127781) - [b, g] Voluntary response sampling (one:T00019) - [b, g] Judgement sampling (one:T00020) - [b, g] Convenience sampling (one:T00021)
7	Control group (MeSH_D035061, NCIT_C28143)	
8	Type of controls (NCIT_C49647)	
9	Recruitment period (NCIT_C142664)	DD/MM/YYYY (xsd:datetime)
10	Follow-ups (NCIT_C16033)	time (xsd:datetime) actions (CTV3_X79tx)
11	Total number of participants recruited (MeSH_D011153)	[b, f] total number of males (one:T00022) [b, f] total number of females (one:T00023)
12	[b] Participants age range (one:T00024)	[b, i] age.min (one:T00025) [b, i] age.max (one:T00026)
	[b, c] ISA framework-Assay (one:T00027)	

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Descriptors	Options
1	[a] Dietary assessment method (one:ne00001)	[a] Dietary records (one:ne00002) - [a] Dietary records: PDA-technologies (one:ne00007) - [a] Dietary records: Mobile phone-based technologies (one:ne00008) - [a] Dietary records: Camera-recorder–based technologies (one:ne00009) - [a] Dietary records: Tape-recorder–based technologies (one:ne00010) a 24-Hour Recall (one:ne00003) - [a] 24-Hour Recall: Interactive computer-based technologies (one: 00011) - [a] 24-Hour Recall: Interactive web-based technologies (one: 00012) [a] Screener (one:ne00004) - [a] Screener: Interactive computer-based technologies (one:ne00013) - [a] Screener: Interactive web-based technologies (one:ne00014) - [a] Screener: qualitative (only frequency) (one:ne00015) - [a] Screener: semi-quantitative (one:ne00016) - [a] Screener: quantitative (one:ne00017) [a] Food Frequency Questionnaire (one:ne00005) - [a] FFQ: Interactive computer-based technologies (one:ne00018) - [a] FFQ: Interactive web-based technologies (one:ne00019) - [a] FFQ: qualitative (only frequency) (one:ne00020) - [a] FFQ: semi-quantitative (one:ne00021) - [a] FFQ: quantitative (one:ne00022) [a] Diet History (one:ne00006) [a] Other: please specify
2	[b, j] Food composition Table (one: T00027)	
3	Food product type (FoodOn_03400361)	Food, Drinks, Dietary supplements (classes of FoodOn)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Descriptors	Options
4	[a] Dietary intake data (one:ne00023)	[a] Unadjusted data (preferred option) (one:ne00024) [a] Adjusted data for total energy intake using density method (one:ne00025) [a] Adjusted data for total energy intake using residual method (one:ne00026) [a] Estimates of usual intake from short-term measurements (one:ne00027) Other: describe
5	Physical activity measurement (NCIT_C120914)	[b, h] Objective measurement (one:T00028) [b, h] Subjective measurement (one:T00029)
6	Tobacco use (MeSH_D064424)	
7	Alcohol consumption (NCIT_C16273)	
8	Anthropometry (MeSH_D000886)	Weight (MeSH_DD001835) Height (MeSH_D001827) Waist circumference (MeSH_D055105) BMI status (MeSH_D015992) Body fat distribution (MeSH_D050218)
9	Socio-demographic factor (ONTOAD_AD000403)	
10	Health outcomes (HL7_C1550208)	xsd:datetime
11-12	Genitourinary samples (CTV_X7ADQ)	Blood sample (CTV3_X7ADI) Serum sample (CTV3_X7AE4) Plasma sample (CTV3_X7AEI) Urine sample (CTV3_X7ABI) Saliva sample (CTV3_4128) Faeces sample (CTV3_x7AAR) Other: please specify (subclasses of CTV3_X7ADQ)
13	Fasting (CTV3_X78x9)	
14	sampling (NCIT_C25662)	xsd:datetime
15	Omics (EDAM_topic3391)	Biomarkers (EDAM_topic3360) Metabolomics (EDAM_topic3172) Proteomics (EDAM_topic0121) Genomics (EDAM_topic0622) Transcriptomics (EDAM_topic3308)
16	Metabolite profiling (OBI_0000366)	
17	mass spectrometry (MeSH_D013058) chromatography (MeSH_D002845)	

[a] undefined nutritional epidemiologic term; [b] other undefined terms;

Recommendation: put undeveloped term (s) in selected ontology: [c]: GODAN framework; NCIT: subclasses of Body Weight Measurement (NCIT_C92648). ISA framework; [d] FAIR guiding principle, under "To be accessible" and "To be reusable"; [e] MeSH term, subclasses of "population characteristics MeSH_D011154", [f] MeSH term,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at *Nutrients* **2019**, *11*, 1300; doi:10.3390/nu11061300

subclasses of MeSH_D011153; [g] NCIT: subclasses of NCIT_C71517/NCIT_C127781; [h] NCIT: subclasses of NCIT_120914; [i] XML schema (XSD); [j] GODAN project;

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

Table 9A. Ontology view of data quality descriptors of observational studies

	Descriptors	Options
	Study design (NCIT_C15320) Cohort (MeSH_D015331) Cross-sectional (MeSH_D003430) Case-control (MeSH_D016022)	
1	Response rate (EO:0000139)	Response rate (EO:0000139) [b] Cooperation rate (one:T00101)
2	Covariates (NCIT_C142645) Cofounding factors (MESH/D015986)	
3	[b] Method for confirming diagnosis (one:T00102)	owl:class (i.e. method) [b] non-validated diagnosis (one:T00103)
4	missing data (NCIT_C142610) - [b] missing data-exposure (one:T00104) - [b] missing data-outcome (one:T00105)	xsd:decimal
5	missing data (NCIT_C142610)	[b] Missing (completely) at random (one:T00106) [b] Missing not at random (one:T00107)
6	Random selection (OBCS_0000063)	
7	** sample representativeness (one:T00108)	[b] Representative sample (one:T00109) [b] Non-representative sample (one:T00110)
8	Incidence (NCIT_C61299)	[b] Incident cases (one:T00111)
9	Control groups (NCIT_C28143)	[b] Control group from same population as cases (one:T00112) [b] Controls group from similar population as cases (one:T00113) [b] Controls group from another population (one:T00114)
10	Lost to follow-up (MESH/D059012, (NCIT_C48227)	xsd:decimal
	[a] Dietary assessment method (one:ne00001): [a] Dietary records (one:ne00002), [a] 24-Hour Recall (one:ne00003), [a] Screener (one:ne00004), [a] Food Frequency Questionnaire (one:ne00005), [a] Diet History (one:ne00006)	
1	Administration (NCIT:C25409) - [a] Dietary assessment administration (one:ne00028)	[a] Dietary assessment administration (one:ne00028) - [a] Proxy-administered (one:ne00029) - [a] Self-administered not verified by interviewer (one:ne00030) - [a] Self-administered and checked by interviewer (one:ne00031) - [a] Interview-administered (one:ne00032) - [a] Interview-administered using AMPM (one:ne00033)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Descriptors	Options
2	Questionnaire (NCIT_C64253) - [a] Dietary assessment questionnaire (one:ne00034)	[a] Dietary assessment questionnaire (one:ne00034) - [a] Self-developed questionnaires (one:ne00035) - [a] Use of standardized questionnaire (one:ne00036) - [a] Adopted other Questionnaires (one:ne00037)
3	Content validity (NCIT_C78690) - [a] Content validity of dietary assessment questionnaire (one:ne00038)	[a] Content validity of dietary assessment questionnaire (one:ne00038) - [a] verified content validity in another population (one:ne00039) - [a] verified content validity in a comparable population in terms of both age and dietary habits (one:ne00040)
4	[a] Reference of dietary assessment questionnaire validation (one:ne00041)	[a] Reference of the dietary assessment questionnaire validation (one:ne00041) - [a] dietary assessment methods (one:ne00001) - [a] Food Frequency Questionnaire (one:ne00005) - [a] 24-Hour Recall (one:ne00003) - [a] Dietary records (one:ne00002) - [a] short term dietary record (one:ne00042) - [a] long term weighted dietary record (>7 days) (one:ne00043) - [a] objective methods (one:ne00044) - [a] biomarker of dietary intake (one:ne00045)
5	Validated information (OBI_0302838) - [a] validated information of dietary assessment questionnaire (one:ne00046)	[a] Properties of dietary assessment questionnaire (one:ne00047) - [a] inter-rater reliability (NCIT_C78688) [a] Frequency options to identify between-person variations (one:ne00048) [a] Food items lead to underestimated target nutrients intake (one:ne00049)
6	[a] Validation type for dietary assessment questionnaire (one:ne00050)	Concurrent validity (OBCS_0000160) precision (NCIT_C48045)
7	Season (NCIT_C94729)	Season (NCIT_C94729) - [b] All seasons (one:T00115) - Summer (NCIT_C94732) - Winter (NCIT_C94730) - Spring (NCIT_C94731) - Autumn (NCIT_C94733)
8	[a] Quantification of portion sizes (one:ne00051)	[a] Quantification of portion sizes (one:ne00051) - [a] Not quantified (one:ne00052) - [a] Standard portion sizes without aids (one:ne00053) - [a] Standard portion sizes with aids (one:ne00054) skos:definition such as pictures, models, standard household measure, utensils, etc. - [a] Portion sizes are assessed digitally but not verified by trained staff (one:ne00055) - [a] Portion sizes are assessed digitally and verified by trained staff (or packaging) (one:ne00056)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Descriptors	Options
9	[a] Portion size of dietary intake data (one:ne00057)	[a] Portion size of dietary intake data (one:ne00057) - [a] directly expressed portion size (one:ne00058) - [a] converted portion size (one:ne00059) - [a] unconverted portion size (one:ne00060)
10	[b, c] Food composition Table (one: T00027) - [b] Geographically-specific food composition data (one:T00116)	
11	[a] Matched consumed food to referred food composition data (one:ne00060)	[a] Matched consumed food to referred food composition data (one:ne00060) - [a] exact matching (one:ne00061) - [a] matched to means of min. 3 food items (one:ne00062) - [a] matched to same food items with similar moisture content (one:ne00063) - [a] matched to a different food (one:ne00064) Percentage in xsd:decimal
12	[a] Representativeness of the week/weekend days (one:ne00065)	Weekend (NCIT_C137684) Weekday (NCIT_C86936)
13	[a] Number of recall/measurement days per individual (one:ne00066)	xsd:integer
14	[a] Selection of recall/measurement days (one:ne00067)	[a] Selection of recall/measurement days (one:ne00067) - [a] Convenience selection (one:ne00068) - [a] Consecutive days (one:ne00069) - [a] Non-consecutive, non-random days (one:ne00070) - [a] Randomly over the week (one:ne00071)
15	[a] The time of diet records (one:ne00072)	[a] The time of diet records (one:ne00072) - [a] Not during eating occasions nor immediately after (one:ne00073) - [a] Immediately after eating occasion (one:ne00074) - [a] During eating occasion (one:ne00075)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

	Descriptors	Options
16	[a] Food quantification method (one:ne00076)	[a] Food quantification method (one:ne00076) - [a] Food quantification method tailored to the characteristics of the population (one:ne00077) - [a] Food quantification method not specifically tailored to the characteristics of the population (one:ne00078)
	Anthropometry (MeSH:D000886)	
1	[b] Training of assessor (one:T00117)	[b] Training of assessors (one:T00117) - [b] without assessors (one:T00118) = Self-report (NCIT_C74528) - [b] trained assessors (one:T00119) - [b] trained assessors using Standard Operating Procedures (one:T00120) - [b] trained assessors not using Standard Operating Procedures (one:T00121) - [b] untrained assessors using Standard Operating Procedures (one:T00122)
2	Body Weight Measurement (NCIT_C92648)	Body Weight Measurement (NCIT_C92648) - Self-Report (NCIT_C74528) - Proxy Data Origin (NCIT_C142651) - [b, d] Measured with no clothing instructions by an assessor (one:T00123) - [b, d] Measured naked or with only light clothing by an assessor (one:T00124)
3	[b] Height measurement (one:T00125)	[b] Height measurement (one:T00125) - Self-Report (NCIT_C74528) - Proxy Data Origin (NCIT_C142651) - [b] Measured with shoes (one:T00126) - [b] Measured barefoot (one:T00127)
4	[b] Waist circumference measurement (one:T00128)	[b] Waist circumference measurement (one:T00128) - Self-Report (NCIT_C74528) - Proxy Data Origin (NCIT_C142651) - [b] Measured with no clothing instructions (one:T00129) - [b] Measured naked or with only light clothing (one:T00130)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at *Nutrients* **2019**, *11*, 1300; doi:10.3390/nu11061300

	Descriptors	Options
5	Measurement of body mass index (SNOMEDCT_698094009)	Measurement of body mass index (SNOMEDCT_698094009) - Self-Report (NCIT_C74528) - [b] Assessed using pictograms or silhouettes (one:T00131) - Objective Measurement (NCIT_C142618): xsd:definition weight & height, body scanner, etc.
6	[b] Adiposity measurement (one:T00132)	bioelectrical impedance analysis (NCIT_C43545) Dual X-ray Absorptiometry (NCIT_C48789) Waist-to-hip ratio (NCIT_C17651) Skin fold (CMO_0000246)

[a] undefined nutritional epidemiologic term; [b] other undefined terms;

Recommendation: put undeveloped term (s) in selected ontology: [c] GODAN framework; NCIT: subclasses of Body Weight Measurement (NCIT_C92648).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

References

[1] Willett W. Nutritional epidemiology. 2nd Edition ed. New York: Oxford University Press; 1998.

[2] Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. N Engl J Med. 2000;342:1887-92.

[3] Chan AW, Song F, Vickers A, Jefferson T, Dickersin K, Gotzsche PC, et al. Increasing value and reducing waste: addressing inaccessible research. Lancet. 2014;383:257-66.

[4] Lachat C, Hawwash D, Ocké MC, Berg C, Forsum E, Hörnell A, et al. STrengthening the Reporting of OBservational studies in Epidemiology – Nutritional Epidemiology (STROBE-nut): an extension of the STROBE statement. Plos Medicine. 2016.

[5] Cochrance Linked data. Cochrane PICO Ontology.

[6] Institute for Health Metrics and Evaluation, Global Burden of Disease Study 2015 (GBD 2015) Data Input Sources Tool. Global Burden of Disease Study 2015 (GBD 2015) Data Input Sources Tool. 2015.

[7] Rhee SY, Wood V, Dolinski K, Draghici S. Use and misuse of the gene ontology annotations. Nat Rev Genet. 2008;9:509-15.

[8] Fang WD, Zhang L, Wang YX, Dong SB. Toward a semantic search engine based on ontologies. Proceedings of 2005 International Conference on Machine Learning and Cybernetics, Vols 1-9. 2005:1913-8.

[9] Noy NF, McGuinness DL. Ontology development 101: A guide to creating your first ontology. 2001.

[10] Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data. 2016;3:160018.

[11] Boulos M, Yassine A, Shirmohammadi S, Namahoot C, Brückner M. Towards an "Internet of Food": Food Ontologies for the Internet of Things. Future Internet. 2015;7:372-92.

[12] Snoek HM, Eijssen LMT, Geurts M, Vors C, Brown KA, Bogaardt M-J, et al. Advancing food, nutrition, and health research in Europe by connecting and building research infrastructures in a DISH-RI: Results of the EuroDISH project. Trends Food Sci Tech. 2018;73:58-66.

[13] Lemay DG, Zivkovic AM, German JB. Building the bridges to bioinformatics in nutrition research. Am J Clin Nutr. 2007;86:1261-9.

[14] European Food Safety Authority. The food classification and description system FoodEx2

[15] Dooley MD, Griffiths JE, Gosal SG, Buttigieg LP, Hoehndorf R, Lange CM, et al. FoodOn: a harmonized food ontology to increase global food traceability, quality control and data integration. npj science of food. 2018;2.

[16] Danish Food Informatics. LanguaL - the International Framework for Food Description. 2015.

[17] 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;13:12.

[18] Coffran C, Meehan D, Ronning A. Bionutrition Ontology. 2013.

[19] Pinart M, Nimptsch K, Bouwman J, Dragsted LO, Yang C, De Cock N, et al. Joint Data Analysis in Nutritional Epidemiology: Identification of Observational Studies and Minimal Requirements. J Nutr. 2018;148:285-97.

[20] Yang C, Pinart M, Kolsteren P, Van Camp J, De Cock N, Nimptsch K, et al. Perspective: Essential Study Quality Descriptors for Data from Nutritional Epidemiologic Research. Adv Nutr. 2017;8:639-51.

[21] d'Aquin M, Noy NF. Where to Publish and Find Ontologies? A Survey of Ontology Libraries. Web Semant. 2012;11:96-111.

[22] World Wide Web Consortium. Ontology repositories. World Wide Web Consortium; 2012.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

[23] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, et al. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol. 2007;25:1251-5.

[24] Noy NF, Shah NH, Whetzel PL, Dai B, Dorf M, Griffith N, et al. BioPortal: ontologies and integrated data resources at the click of a mouse. Nucleic Acids Res. 2009;37:W170-3.

[25] Cote RG, Jones P, Apweiler R, Hermjakob H. The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. BMC Bioinformatics. 2006;7:97.

[26] Burton-Jones A, Storey VC, Sugumaran V, Ahluwalia P. A semiotic metrics suite for assessing the quality of ontologies. Data & Knowledge Engineering. 2005;55:84-102.

[27] RDF Working Group. Resource Description Framework (RDF). 2014.

[28] Lachat C, Raneri JE, Smith KW, Kolsteren P, Van Damme P, Verzelen K, et al. Dietary species richness as a measure of food biodiversity and nutritional quality of diets. Proc Natl Acad Sci U S A. 2018;115:127-32.

[29] Kupershmidt I, Su QJ, Grewal A, Sundaresh S, Halperin I, Flynn J, et al. Ontology-based meta-analysis of global collections of high-throughput public data. PLoS One. 2010;5.

[30] Ramaprasad A, Syn T. Ontological Meta-Analysis and Synthesis Communications of the Association for Information Systems. 2015;37.

[31] Mills S, Brown H, Wrieden W, White M, Adams J. Frequency of eating home cooked meals and potential benefits for diet and health: cross-sectional analysis of a population-based cohort study. Int J Behav Nutr Phys Act. 2017;14:109.

[32] Mills S, Adams J, Wrieden W, White M, Brown H. Sociodemographic characteristics and frequency of consuming home-cooked meals and meals from out-of-home sources: cross-sectional analysis of a population-based cohort study. Public Health Nutr. 2018;21:2255-66.

[33] Abris GP, Kim NH, Provido SMP, Hong S, Yu SH, Lee CB, et al. Dietary diversity and nutritional adequacy among married Filipino immigrant women: The Filipino Women's Diet and Health Study (FiLWHEL). BMC Public Health. 2018;18:359.

[34] Chatelan A, Beer-Borst S, Randriamiharisoa A, Pasquier J, Blanco JM, Siegenthaler S, et al. Major Differences in Diet across Three Linguistic Regions of Switzerland: Results from the First National Nutrition Survey menuCH. Nutrients. 2017;9.

[35] Lam MCL, Adams J. Association between home food preparation skills and behaviour, and consumption of ultra-processed foods: Cross-sectional analysis of the UK National Diet and nutrition survey (2008-2009). Int J Behav Nutr Phys Act. 2017;14:68.

[36] Llanaj E, Ádány R, Lachat C, D'Haese M. Examining food intake and eating out of home patterns among university students. 2018.

[37] Arsenault JE, Nikiema L, Allemand P, Ayassou KA, Lanou H, Moursi M, et al. Seasonal differences in food and nutrient intakes among young children and their mothers in rural Burkina Faso. J Nutr Sci. 2014;3:e55.

[38] De Cock N, Van Lippevelde W, Vervoort L, Vangeel J, Maes L, Eggermont S, et al. Sensitivity to reward is associated with snack and sugar-sweetened beverage consumption in adolescents. Eur J Nutr. 2016;55:1623-32.

[39] Workicho A, Belachew T, Feyissa GT, Wondafrash B, Lachat C, Verstraeten R, et al. Household dietary diversity and Animal Source Food consumption in Ethiopia: evidence from the 2011 Welfare Monitoring Survey. BMC Public Health. 2016;16:1192.

[40] Bechem M, Huybregts L, van Damme P. Biodiversity and complementary feeding practices of children in the north west region of Cameroon. Harvard dataverse.2014.

[41] Yang C. Case studies: study description Ontology for Nutritional Epidemiology. BioPortal; 2018.

[42] GODAN. Global Open Data for Agriculture and Nutrition. 2017.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 15 May 2019

Peer-reviewed version available at Nutrients 2019, 11, 1300; doi:10.3390/nu11061300

[43] Glasziou P, Altman DG, Bossuyt P, Boutron I, Clarke M, Julious S, et al. Reducing waste from incomplete or unusable reports of biomedical research. Lancet. 2014;383:267-76.

[44] von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE)statement: guidelines for reporting observational studies. PLoS Med. 2007;4:e296.

[45] the Centre for Statistics in Medicine (CSM). Equator: Enhancing the QUAlity and Transparency Of health Research. the UK EQUATOR Centre.

[46] Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. BMJ. 2010;340:c869.

[47] Moher D, Liberati A, Tetzlaff J, Altman DG, Group P. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. PLoS Med. 2009;6:e1000097.

[48] Moher D, Jones A, Lepage L, for the CG. Use of the CONSORT Statement and Quality of Reports of Randomized Trials. Jama. 2001;285:1992.

[49] Panic N, Leoncini E, de Belvis G, Ricciardi W, Boccia S. Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. PLoS One. 2013;8:e83138.

[50] Sorensen AA, Wojahn RD, Manske MC, Calfee RP. Using the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement to assess reporting of observational trials in hand surgery. J Hand Surg Am. 2013;38:1584-9 e2.

[51] Wilkinson MD, Verborgh R, Bonino da Silva Santos LO, Clark T, Swertz MA, Kelpin FDL, et al. Interoperability and FAIRness through a novel combination of Web technologies. PeerJ Computer Science. 2017;3:e110.

[52] Stoeger T, Gerlach M, Morimoto RI, Nunes Amaral LA. Large-scale investigation of the reasons why potentially important genes are ignored. PLoS Biol. 2018;16:e2006643.

[53] Yang H, Callan J. Human-Guided Ontology Learning. Proceeding of HCIR. 2008:26-9.

[54] Gomez-Perez A, Manzano-Macho D. An overview of methods and tools for ontology learning from texts. Knowl Eng Rev. 2004;19:187-212.

[55] Arguello Casteleiro M, Demetriou G, Read W, Fernandez Prieto MJ, Maroto N, Maseda Fernandez D, et al. Deep learning meets ontologies: experiments to anchor the cardiovascular disease ontology in the biomedical literature. J Biomed Semantics. 2018;9:13.

[56] ENPADASI consortium. Project deliverable report (D6.3, D6.5 workshops). 2017.

[57] ENPADASI consortium. European Nutritional Phenotype Assessment and Data Sharing Initiative (ENPADASI).

[58] Foust JC. The Ontology of Physical Exercises (OPE) BioPortal2013.

[59] Zaveri A. Randomized Controlled Trials Ontology. BioPortal2011.

[60] Zaveri A. Non-Randomized Controlled Trials Ontology. BioPortal2012.

[61] Scheuermann R. Immune Disorder Ontology. BioPortal2012.

[62] Santana F. Neglected Tropical Disease Ontology. BioPortal2012.