Introducing WikiFCD: Many Food Composition Tables in a Single Knowledge Base

Katherine Thornton[1], Kenneth Seals-Nutt[2] and Mika Matsuzaki[3]

[1] *WikiFCD Collaborative, Olympia, WA, USA*
[2] *WikiFCD Collaborative, New York, New York, USA*
[3] *Johns Hopkins Bloomberg School of Public Health, 615 N Wolfe St, Baltimore, MD 21205, United States*

Abstract

We introduce WikiFCD, a knowledge base of structured food composition data. This knowledge base is designed to accommodate data from different regions of the world, it is multi-lingual, and it supports open participation of any interested editor. We used Wikibase to store data from multiple food composition tables (FCTs). We mapped relevant classes of data to corresponding entities in the Wikidata knowledge base in order to support querying of food composition data alongside data about chemical compounds, metabolites, biological pathways, and data about human genes. We also make use of FoodOn to provide identifiers for food items. This knowledge base contains a growing number of FCTs that provide coverage of a broad range of cuisines and food traditions. Reusing data from this knowledge base can provide greater coverage of foods for nutrient intake tools. This knowledge base will be useful for policy makers, epidemiologists, nutrition researchers, developers of food-related applications, and people interested in food tracking.

Keywords

food composition, knowledge base, Wikidata, Nutri-informatics

1. Introduction

Food is an essential part of our lives, providing energy and nutrients required for health. Suboptimal diet contributes to one in five deaths globally, making dietary improvement one of the highest priorities in global health [1]. Our ability to accurately represent and retrieve information on food items has an unequivocal impact on the quality of prevention and treatment strategies we develop for nutrition-related diseases. Food composition data (FCD) - a central piece connecting foods to health - has a rich history and diverse datasets exist around the world. And yet, the usability as well as the interoperability of these data vary greatly, with a large disparity between high income countries (HIC) and low and middle income countries (LMIC). This disparity has a grave implication for global health as the "triple" burden of malnutrition due to deficiencies or excess in macro and micronutrients are ubiquitous. Accurate and detailed information about the composition of the foods we eat are, more than ever, needed by policy

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

katherine.thornton@yale.edu (K. Thornton); kenneth@seals-nutt.com (K. Seals-Nutt); mmatsuz2@jhu.edu (M. Matsuzaki)

0000-0002-4499-0451 (K. Thornton); 0000-0002-5926-9245 (K. Seals-Nutt); 0000-0002-7020-3757 (M. Matsuzaki)

CEUR Workshop Proceedings (CEUR-WS.org)

makers, researchers, software developers, and consumers as the world faces the epidemic of nutrition-related diseases.

Even in HIC, many food items that are consumed by people every day cannot be considered for dietary analyses despite the existence of nutrient data for these food items [2] . As a result, individuals are often forced to use nutrient data from "similar" food items, which may or may not actually have similar nutrient content. This is especially true for those who consume more ethnic minority foods.

FCD are currently fragmented, unevenly provisioned, and published in formats ill-suited to the web. Nutrient content of fruits, vegetables, staples, meats, and dairy products can also vary for the same item from different areas and times because of changing characteristics such as climate and terroir. However, the current structure for most FCDs are not well-suited for reflecting these changes. Importantly, even though there are also wide regional variations in foods that are commonly consumed, some places lack access to regionally appropriate FCD, up-to-date FCD, or FCD in their own languages, leading to disparities in data availability and accessibility and ultimately, in scientific evidence in health research. Development and maintenance of such databases are difficult if the contributors are limited to small, closed groups of researchers and employees in this field.

However, it is within our power to change this situation and ensure that accurate food composition data are available for the long tail of food items. Our solution to this challenge is to build a knowledge base of structured food composition data using a peer production approach. Not only is this knowledge base designed to accommodate data from different regions of the world, it is multi-lingual, and supports open participation of any interested editor. The knowledge base is also designed so that data is available according to FAIR principles [3]. The free software infrastructure used to power Wikidata - Wikibase - has enabled us to develop a unified resource encompassing many different food composition tables. We are building our knowledge base using Wikibase because it is optimized for both human and algorithmic curation. Opening contribution to anyone interested in this effort will both allow our dataset to grow and involve many more people to participate in the data curation, which, as shown in the examples of Wikipedia and Open Street Map, could lead to a creation of a large, equitable knowledge base.

In addition to the sheer number of potential contributors to this project - Wikidata currently has over 250,000 active users - there are other advantages of this peer production and Wikibase based approach to traditional methods of FCD development. First, this Wikibase instance substantially improves the usability of FCDs from different sources for diverse users - from WikiProjects and Wikipedia editors and viewers to academic researchers to public health workers. Researchers have found that the absence of culturally-diverse foods in apps such as MyFitnessPal is a barrier to using them in research [4, 5, 6, 7].

Building a structured dataset is also a key step in identifying most appropriate data to borrow in resource-poor settings where up-to-date, detailed, and regionally appropriate FCD are not readily available. This new database will also open up ways to explore new research questions to explore more nuanced nutrition data (e.g. changes in nutrient content of the same product, depending on the climate conditions of the year), which can potentially make substantial advances in nutrition and health research.

We created an instance of Wikibase for this project and designed our own data models which are flexible enough to allow us to incorporate data from heterogeneous data sources. Connecting this knowledge base with Wikidata allows us to combine this data with cross-domain data related to micronutrients, chemical compounds, biological pathways, human genes, and disease information using databases like the Human Metabolome Database and Wikipathways. Connecting the data in this way allows us to ask questions about how food choices may impact health in a broad range of ways.

2. Development of WikiFCD

2.1. Wikidata

Wikidata went live in late 2012 [8]. The infrastructure of Wikidata is collaboratively built via commons-based peer production [9, 10, 11]. Commons-based peer production is the name given to open collaboration systems where users are creating content under the agreement that all content will remain in the public domain. This means that content created by the community can be freely reused by others. The peer production aspect refers to how users coordinate work themselves, rather than some members of the community organizing the work tasks of other members. Wikidata is edited by volunteers from all over the world in more than 350 languages [12].

In addition to a free software infrastructure, the Wikidata community also publishes all content in the knowledge base under a Creative Commons Zero License. The Wikidata community makes dumps of previous versions of the content of the knowledge base available. The infrastructure of the Wikidata knowledge base is maintained by an international community of people. For cultural heritage institutions who find the structured data in Wikidata useful for work flows, this mans that there will be much less staff time necessary to design, build and maintain infrastructure for this data. The data in WikiFCD complements the work of several active communities curating data in Wikidata: the GeneWiki initiative [13, 14], the WikiPathways community [15] The LOTUS initiative [16] and the Scholia project [17].

Wikidata is a multilingual knowledge base, leveraging the concept mappings created through years of conceptual alignment among the different language versions of Wikipedia [18]. This means that more users will have access to data in their language, an important step in reducing the dominance of the English language which disadvantages other linguistic communities.

2.2. Reusing Wikibase

We chose to create this knowledge graph of structured data published in a publicly-available instance of the Wikibase platform, called WikiFCD[1]. Wikibase is a set of extensions to the MediaWiki software platform and is developed by the Wikimedia Foundation as free software. Wikibase is a novel infrastructural platform for data management suitable for data from many domains. This is the first application built on Wikibase tailored to the needs of the epidemiological community. The output of this project will be a knowledge graph of structured data in the form of a Wikibase instance populated with data from heterogeneous food composition tables.

[1] https://wikifcd.wiki.opencura.com/wiki/Main_Page

We reused three subsets from Wikidata to create some of the basic structure of our knowledge base. Identifying and reusing subsets of Wikidata is still an emerging practice [19]. We wrote SPARQL queries to identify all taxa with an identifier in the Germplasm Resources Information Network (GRIN) in Wikidata. We then wrote a bot to populate these items to WikiFCD with mappings back to the Wikidata item. The purpose of having these taxa in WikiFCD was to be able to create statements about food items that are derived from a taxon. We did the same process with the set of human languages and the set of countries/states. We did this so that we could use language and country items in our statements about individual food composition tables and to provide linguistic information about the common names of food items.

We have systematically mapped data in WikiFCD to corresponding items and properties in Wikidata itself. These mappings allow us to ask questions of both data sets and to make use of the mappings between Wikidata and thousands of external data sources. These mappings increase the breadth and complexity of data combinations we can create, using Wikidata as the hub of connection. Multiple data visualization options are available via the Query Service of our Wikibase instance. The Query Service is a SPARQL endpoint which supports querying the data in the knowledge graph via the SPARQL query language. Graphs, charts, network diagrams, and maps are some of the visualizations we will be able to offer end-users of this knowledge base [20].

To collect a list of food composition tables (FCTs) representative of international communities, we consulted the resources described by the United Nations Food and Agriculture Organization (FAO)[2]. We worked from the FAO's list of food composition tables to identify existing FCTs that we could add to our Wikibase. We then found copies of these FCTs where possible. We then extracted the data from these tables. The FCTs were originally published as CSV or as tabular data encoded in a PDF.

We populated WikiFCD with data from the USDA's Food Data Central database. Food Data Central has a set of APIs that can be used to access data. We wrote a client to collect data from Food Data Central and then wrote a bot to populate WikiFCD with the data.

We created a database model that can represent heterogeneous food composition tables. We used this model to map multiple food composition tables so that we could then import them into a Wikibase instance. We also support the addition of data sourced from the literature that covers a single food. This is an advantage of our data model as well as our contribution model. While other multi-country food composition data bases (FCDBs) combine national level FCTs [21], we include foods that are not yet found in any country's official FCT. We aim for broad representation of food ways, striving to include food composition data for wild, foraged foods, and less-commonly-eaten plant foods.

Our alignment of food composition table data with Wikidata allows us to leverage the sum of knowledge in the projects of the Wikimedia foundation. Because Wikimedia Commons, the media repository of Wikimedia projects, has also been aligned with Wikidata, we will be able to easily reuse images of food items, molecular structure models, and food dishes alongside our projects. This query from our SPARQL endpoint[3] lists all of the food items in our project Wikibase that have an associated image in Wikimedia Commons.

[2] http://www.fao.org/infoods/infoods/tables-and-databases/en/

[3] https://tinyurl.com/y99qtk7p

We used the wbstack platform to create an instance of Wikibase for testing[4]. The wbstack service provides a hosted version of Wikibase that users can load with their own data. Wikibase is the software used to support Wikidata itself. In order to populate our system with data we used a tool called WikidataIntegrator (WDI). WDI is a python library for interacting with data from Wikidata [22]. WDI was created by the Su Lab of Scripps Research Institute and shared under an open-source software license via GitHub[5]. Using WDI as a framework, we wrote bots to transfer data from FCTs to our Wikibase.

The largest class of items in this system is that of `food item`. There are currently about 400,000 food items in the system. We have more than 300 properties in the system which we use to describe the items. Examples of properties are Dietary Fiber (P11), and Fatty acids, total saturated (P86).

In order to group food items we assign identifiers from FoodOn. FoodOn is an ontology that describes foods and the organisms from which they are derived [23]. By making use of the FoodOn ontology we can bring together food items across diverse FCT sources. FoodOn reuses the Composition Dietary Nutrition Ontology [24]. We plan to map our nutrient properties to the relevant components of CDNO.

After importing data from Food Data Central of the United States, we next imported data from the Malawian Food Composition Table 2019 [25]. We used WikidataIntegrator to write a bot to read data from a CSV table version of the FCT and write it to WikiFCD according to our data model. We took several steps to prepare the data before ingest. We split out values from the column "Food Item" and created three additional columns. We left the English language name of the food item in the "Food Item" column. We created a new column "Taxon" to accommodate the binomial names in italics that were previously in the "Food Item" column after the English-language label. It was very helpful to see that the binomial names were included for so many food items in this FCT. This information allows us to disambiguate food items. We then created a column "Taxon ID" for the Qid of the taxon in our knowledge base. We created another new column "Common name" for the name in parentheses for local names of these food items. The reason we created separate columns was to prepare the file for use by our bot. For each of these new columns the data is mapped to a separate property in our data model and our bot will write different statements to WikiFCD using the appropriate properties. We also had to remove the quotation marks, square brackets, and parentheses around some of the nutrient values reported. We could not accommodate those characters in our knowledge base, so we removed them before the bot run.

Some creators of FCTs include references for the publications where they sourced their data. These references are very useful for understanding how the FCT was compiled. These references can be difficult for us to incorporate into our data model because Wikibase was designed to accommodate references for each statement [26]. If we are unable to determine which values were sourced from which publication, the we do not have clarity about which statement(s) on which to put the reference. The Malawian 2019 FCT clearly indicated for each row of data which reference was used to source data. We wrote additional bots for each FCT we ingested into the WikiFCD system.

[4] https://www.wbstack.com/

[5] https://github.com/SuLab/WikidataIntegrator

2.3. Populating WikiFCD with Data

Data in WikiFCD is FAIR data. FAIR is a set of data principles [3]. By creating data that aligns with the FAIR data principles, we ensure that this metadata is easy to find and easy to reuse. Redundant, fragmented descriptions in siloed repositories are frustratingly incomplete. Many governmental bodies and international consortia have endorsed the FAIR data principles as a key aspect of their open science or open data initiatives [27]. FAIR is an acronym for findable, accessible, interoperable and reusable. Food composition data in WikiFCD are **findable** in that WikiFCD is available on the web and is openly accessible. The Qids assigned to WikiFCD items are their unique, persistent identifiers.

These metadata are **accessible** because the entity data associated with their unique ids (all statements and references asserted about an item) are dereferencable via the HTTP protocol. They are **interoperable** in that they link to many other databases and systems through the Wikidata mappings which connect to external ids.

These metadata are **reusable** due to the use of the CCO license for the content of WikiFCD. Anyone can reuse WikiFCD data for any purpose. Publishing data in the WikiFCD knowledge base fulfills the most complete degree of FAIRness, level F, "FAIR data, Open Access, Functionally Linked", as described in [27].

We have so far curated data from the United States Department of Agriculture's FoodData-Central database, SMILING Indonesia, SMILING Vietnam, SMILING Thailand, SMILING Laos, and Malawi. Our initial goal is to curate data from low and middle income countries in WikiFCD with the aim to reduce the aforementioned data disparities in nutrition.

2.4. Mapping food items to FoodOn

FoodOn is an ontology for foods [23]. FoodOn reuses many food categories from LanguaL and is developed according to the ontology principles of the OBO Foundary. We decided to reuse FoodOn identifiers on our food items in WikiFCD in order to create a bridge between our food composition data and the FoodOn ontology. We have mapped some of our food items to their FoodOn identifiers manually as a test set. In the future we will be able to match some food items in a semi-automated manner if we have data about the taxon from which the food item is derived. Some food composition tables provide this information. If this information is not provided in the FCT we will then map them manually.

3. Use Case

Even though we have only a small fraction of existing FCDs in the world, the benefit of the creation of this Wikibase instance is apparent. We are able to query for values across all FCTs in WikiFCD. For example we can query for a ranked list of foods that have the most to least Docosahexaenoic acid (DHA) per 100 grams[6].

We have also tested several federated queries that allow data from additional SPARQL endpoints to be included. For the subset of items that we have already mapped to FoodOn, we were interested to know what metabolites are produced when humans consume these foods. We

[6] https://tinyurl.com/y56qvvr6

wrote a federated query between WikiFCD and Wikidata to ask about the food items, FoodOn ids, and taxa from which these foods are derived (facts stored in WikiFCD) with data about metabolites available from Wikidata[7].

We explored the reuse of information about biological pathways from Wikidata as well as the supporting scientific literature from which the information was sourced by writing a federated query between WikiFCD and Wikidata[8]. The query asks for chemical compounds that are part of a biological pathway in homo sapiens and the scientific articles that provide evidence.

We can use Wikidata as a hub of identifiers that provide cross-references to additional databases [28]. This means that once we have the Wikidata Qid for a resource, we find many other identifiers for that resource from a broad range of other databases and information systems. For example many chemical compounds have an external identifier for the Human Metabolome Database (HMDB). We wrote a federated query for taxa listed in WikiFCD in which certain chemical compounds are found along with the HMDB identifiers for those compounds. This query allows us to connect food items that are derived from specific plants with a profile of metabolites that are relevant for human health. The microbiome is recognized as playing a role in health inequities [29]. Being able to combine these data is an important step in preparing additional research.

Items in Wikidata are connected to external databases or collections through the use of properties that have the data type “external id”. More than half of all properties in Wikidata are external id properties. Connecting Wikidata items to other resources in this way is a powerful feature allowing us to fulfill the promise of linked open data [30]. By following external id links, users can discover more information about the item of interest. We prioritize connecting to multiple external projects in our curation activities. As more external identifiers are published to the Wikidata knowledge base it grows in prominence as a cross-switch for identifiers and vocabularies [31]. Wikidata is becoming a hub of persistent identifiers [28]. As users contribute additional data to Wikidata it will become even more valuable.

4. Discussion

4.1. Lessons Learned

Through this development of the pilot project, we have learned several valuable lessons in creating a global FCD. Chan et al. detail the importance of standardizing nutrition data [32]. Our experience importing FCTs into WikiFCD have illustrated how the lack of a standardized template for food composition tables impedes data interoperability. We encourage future creators of FCTs to use INFOODS tag names [33, 34]. Currently we need to develop a unique bot for each FCT. In the future, if a standardized FCT were adopted, we could accomplish the same work with a single bot built to understand the structure of the standard. This would reduce the time researchers need to spend reusing data from different FCTs.

We recommend that teams creating FCTs in the future consider providing mappings for food items to their corresponding FoodOn identifiers. This step will increase precision by providing

[7] https://tinyurl.com/yz5seocf

[8] https://tinyurl.com/ybtgwgby

unambiguous indications of the taxonomic source of the food and which part (eg. plant leaves vs. plant roots). Currently this information is indicated in the label of the food item in many FCTs, but the languages for describing organisms varies. Reuse of FoodOn identifiers will also reduce confusion related to naming differences for foods at the regional and national levels.

In WikiFCD we established mappings from certain items to their corresponding items in Wikidata. Queries on the WikiFCD SPARQL query endpoint can be written to include data from Wikidata because of the fact that the endpoint supports federated querying. This allows users to ask questions of our data that go far beyond what our dataset can answer. The ability to connect a global FCD to a general-purpose knowledge base increases the utility of the FCD.

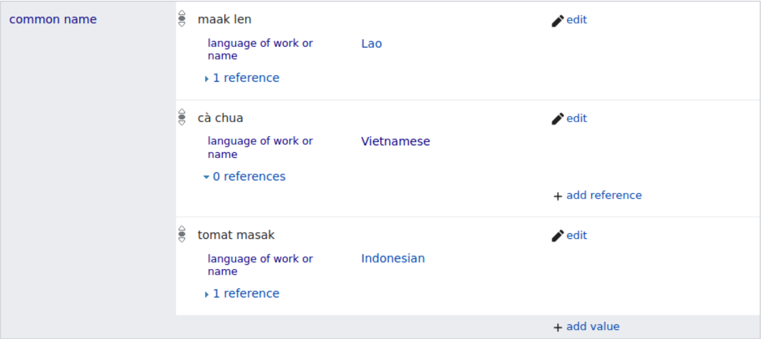
Maintaining a set of mappings to Wikidata also allows us to be strategic in our curation. As researchers estimate that there are 200,000 to 1,000,000 different metabolites synthesized by plants [35], we determined that it is beyond the scope of our knowledge base to store these metabolites in WikiFCD. Instead connect WikiFCD items for taxon names to the corresponding items in Wikidata. These mappings then allow us to make federated SPARQL queries to ask questions about plant metabolites such as "What metabolites are found in foods that are natural products of *Vaccinium deliciosum*, and with what do they physically interact[9].

Connecting Wikidata items to resources in external databases or systems allows for software agents to discover related content automatically. This allows us to benefit from complementary work and provides infrastructure for connecting information that was previously fragmented across multiple systems. In the domain of food, Wikidata has external identifiers for several large food databases. For example, `Property P4729` "INRAN Italian Food ID" is used to link food items with the Italian national nutrient database. Through the use of `Property P4729` the pages dedicated to these resources can be connected to their corresponding items in Wikidata. By making use of this property, we can use the INRAN identifier to find additional information about the food item in the INRAN database. In this way, Wikidata serves as a hub of identifiers that connect to external resources.

4.2. Building the WikiFCD Community

Using Wikibase as infrastructure has allowed the Wikidata community to engage in peer-production and collaborative ontology engineering [11]. We identified peer-production and collaborative ontology engineering as vital components to include in the vision of WikiFCD. Wikibase offers a novel method to change the current state of FCDs and bring a peer-produced knowledge base to the field of nutrition research. The current project explores whether a Wikibase-based FCD can be an effective method of developing a more equitable and comprehensive knowledge base in nutrition. WikiFCD is distinct from previous attempts in compiling a global FCD in that it allows the community members, or "peers", to become involved in the database development directly. The project aims to empower users from low resource settings to fully utilize available nutrient data to answer their own questions, identify knowledge gaps, and engage in improving the database. In successful peer production communities like Wikipedia, projects have garnered efforts from hundreds of thousands of volunteers. The involvement of a large number of "peers" in the production has a potential to successfully building a global FCD.

[9] https://tinyurl.com/y7qplyjh



common name

maak len — edit
language of work or name: Lao
1 reference

cà chua — edit
language of work or name: Vietnamese
0 references
+ add reference

tomat masak — edit
language of work or name: Indonesian
1 reference

+ add value

Figure 1: Common names listed for Tomato in WikiFCD

Developing a successful online community can be challenging. Given strong interests and support we have received from the communities of nutrition researchers, we believe that we will be able to attract participants of the WikiFCD community. Additionally, our team includes experienced Wikimedians and academic researchers in the field of online communities and peer production who are equipped with extensive experiences in peer production communities and in-depth knowledge on theories of online community development. Furthermore, we will also be developing a mobile meal-planning application with options to contribute missing data to WikiFCD, which helps users to connect personal needs to community needs and lowers hurdles in contributing to this knowledge base.

One of the reasons we chose Wikibase was the support for multiple concurrent editors. This means that many different people can contribute data to WikiFCD at the same time. If a community has recently gathered data for their own FCT, they are welcome to add their data to WikiFCD. Use the search box to find a food item. If you'd like to add data for tomato, then start editing the item for `Tomato, fresh`[10]. If you'd like to add a name for this food item in a language that is not yet listed under `common name` then click on `add value` as seen in Figure 1. After entering text, then click on `add qualifier` select `language of work or name` and then enter the language of the word you just contributed. Then provide a reference for your statement. You can reference a webpage or a published source. If you want to reference a published source, simply create a new item for that source in WikiFCD using the `create new item` link available from the sidebar menu.

If there is an image of your food item available within Wikimedia Commons, then it is also possible to connect that image to the WikiFCD food item. If you would like to connect a food item to an image from Commons, you can use P59 “image” as the connecting property[11].

4.3. WikiFCD as a Model Database

Finally, WikiFCD will also serve as an example of setting up a peer-produced knowledge base, helping others who are interested in creating one for their own needs (e.g. local organic farming

[10] https://wikifcd.wiki.opencura.com/wiki/Item:Q135084

[11] https://wikifcd.wiki.opencura.com/wiki/Property:P59

communities) while retaining an ability to make federated queries to other Wikibase-based databases like Wikidata. This knowledge base will be useful for policy makers, epidemiologists, nutrition researchers, developers of food-related applications, and people interested in food tracking. This knowledge base will provide a low-cost data publishing option for areas of the world with limited budgetary resources for data promulgation. From a technology perspective, many national food composition tables are currently publishing one-star or two-star data. We will provide the enabling technology for any organization to publish five-star linked open data that meets the FAIR data guidelines at no cost. We hope that this project becomes the first step to creating a federated community of food and nutrition knowledge producers.

5. Conclusion

Providing infrastructure for researchers and policy makers who need accurate food composition data requires a team of technologists working in close collaboration with domain experts. Populating the resource with data is work that can be shared by anyone interested in food data. We have created a resource that emphasizes ease of data reuse as well as ease of data addition.

If successful, WikiFCD can lead to reduction in data disparities and also enable users to pursue research questions and projects that are currently difficult to explore. WikiFCD will also be able to identify knowledge gaps in FCDs (e.g. missing nutrient information for regional foods). Our system also has the advantage of making federated queries to other Wikibase databases, which will substantially expand the scope of research questions that can be explored. Furthermore, if subsets of the data are appropriate for other Wikibase instances like Wikidata, we will be able to provide machine-actionable ShEx schemas that will help us prepare data for other systems. In this way the data will be readily-available for incorporation into other Wikibase instances if desired.

References

[1] R. Fallaize, A. L. Macready, L. Butler, J. Ellis, J. Lovegrove, An insight into the public acceptance of nutrigenomic-based personalised nutrition, Nutrition research reviews 26 (2013) 39–48.

[2] M. C. Ocké, S. Westenbrink, C. T. van Rossum, E. H. Temme, W. van der Vossen-Wijmenga, J. Verkaik-Kloosterman, The essential role of food composition databases for public health nutrition – experiences from the netherlands, Journal of Food Composition and Analysis 101 (2021) 103967. URL: https://www.sciencedirect.com/science/article/pii/S0889157521001678. doi:https://doi.org/10.1016/j.jfca.2021.103967.

[3] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The fair guiding principles for scientific data management and stewardship, Scientific data 3 (2016) 160018.

[4] M. Egan, A. Fragodt, M. Raats, C. Hodgkins, M. Lumbers, The importance of harmonizing food composition data across europe, European journal of clinical nutrition 61 (2007) 813–821.

[5] S. Shimbo, A. Hayase, M. Murakami, I. Hatai, K. Higashikawa, C.-S. Moon, Z.-W. Zhang, T. Watanabe, H. Iguchi, M. Ikeda, Use of a food composition database to estimate daily dietary intake of nutrient or trace elements in japan, with reference to its limitation, Food Additives & Contaminants 13 (1996) 775–786.

[6] A. Durazzo, E. Camilli, S. Marconi, S. Lisciani, P. Gabrielli, L. Gambelli, A. Aguzzi, M. Lucarini, J. Kiefer, L. Marletta, Nutritional composition and dietary intake of composite dishes traditionally consumed in italy, Journal of Food Composition and Analysis 77 (2019) 115–124.

[7] A. Trichopoulou, S. Soukara, E. Vasilopoulou, Traditional foods: a science and society perspective, Trends in Food Science & Technology 18 (2007) 420–427.

[8] D. Vrandečić, Wikidata: A new platform for collaborative data collection, in: Proceedings of the 21st International Conference Companion on World Wide Web, ACM, 2012, pp. 1063–1064.

[9] Y. Benkler, Coase's penguin, or, linux and the nature of the firm, Yale Law Journal (2002) 369–446.

[10] Y. Benkler, A. Shaw, B. M. Hill, Peer production: a modality of collective intelligence, Collective Intelligence (2013).

[11] C. Müller-Birn, B. Karran, J. Lehmann, M. Luczak-Rösch, Peer-production system or collaborative ontology engineering effort: What is wikidata?, in: Proceedings of the 11th International Symposium on Open Collaboration, ACM, 2015, p. 20.

[12] F. Erxleben, M. Günther, M. Krötzsch, J. Mendez, D. Vrandečić, Introducing wikidata to the linked data web, in: The Semantic Web–ISWC 2014, Springer, 2014, pp. 50–65.

[13] A. Waagmeester, G. Stupp, S. Burgstaller-Muehlbacher, B. M. Good, M. Griffith, O. L. Griffith, K. Hanspers, H. Hermjakob, T. S. Hudson, K. Hybiske, S. M. Keating, M. Manske, M. Mayers, D. Mietchen, E. Mitraka, A. R. Pico, T. Putman, A. Timothy, N. Queralt-Rosinach, L. M. Schriml, T. Shafee, D. Slenter, R. Stephan, K. Thornton, G. Tsueng, R. Tu, S. Ul-Hasan, E. Willighagen, C. Wu, A. I. Su, Wikidata as a knowledge graph for the life sciences, Elife 9 (2020) e52614. URL: https://doi.org/10.7554/ELIFE.52614.

[14] E. Mitraka, A. Waagmeester, S. Burgstaller-Muehlbacher, L. M. Schriml, A. I. Su, B. M. Good, Wikidata: A platform for data integration and dissemination for the life sciences and beyond, bioRxiv (2015) 031971.

[15] M. Martens, A. Ammar, A. Riutta, A. Waagmeester, D. N. Slenter, K. Hanspers, R. A. Miller, D. Digles, E. N. Lopes, F. Ehrhart, et al., Wikipathways: connecting communities, Nucleic Acids Research 49 (2021) D613–D621.

[16] A. Rutz, M. Sorokina, J. Galgonek, D. Mietchen, E. Willighagen, J. Graham, R. Stephan, R. Page, J. Vondrášek, C. Steinbeck, et al., Open natural products research: Curation and dissemination of biological occurrences of chemical structures through wikidata, bioArxiv (2021). URL: https://doi.org/10.1101/2021.02.28.433265.

[17] F. Å. Nielsen, D. Mietchen, E. Willighagen, Scholia, scientometrics and wikidata, in: European Semantic Web Conference, Springer, 2017, pp. 237–259.

[18] S. Burgstaller-Muehlbacher, A. Waagmeester, E. Mitraka, J. Turner, T. Putman, J. Leong, C. Naik, P. Pavlidis, L. Schriml, B. M. Good, et al., Wikidata as a semantic framework for the gene wiki initiative, Database 2016 (2016) baw015.

[19] J. E. Labra-Gayo, A. Ammar, D. Brickley, D. F. Álvarez, A. G. Hevia, A. J. Gray, E. Prud'hom-

meaux, D. Slater, H. Solbrig, S. A. H. Beghaeiraveri, et al., Knowledge graphs and wikidata subsetting, BioHackathon Europe 2020 (2021). URL: https://biohackrxiv.org/wu9et/.

[20] S. Malyshev, M. Krötzsch, L. González, J. Gonsior, A. Bielefeldt, Getting the most out of wikidata: semantic technology usage in wikipedia's knowledge graph, in: International Semantic Web Conference, Springer, 2018, pp. 376–394.

[21] P. M. Finglas, R. Berry, S. Astley, Assessing and improving the quality of food composition databases for nutrition and health applications in europe: the contribution of eurofir, Advances in Nutrition 5 (2014) 608S–614S.

[22] W. Andra, G. Stupp, B.-M. Sebastian, B. M. Good, G. Malachi, O. L. Griffith, H. Kristina, H. Henning, T. S. Hudson, H. Kevin, et al., Wikidata as a knowledge graph for the life sciences, eLife 9 (2020).

[23] D. M. Dooley, E. J. Griffiths, G. S. Gosal, P. L. Buttigieg, R. Hoehndorf, M. C. Lange, L. M. Schriml, F. S. Brinkman, W. W. Hsiao, Foodon: a harmonized food ontology to increase global food traceability, quality control and data integration, npj Science of Food 2 (2018) 1–10.

[24] L. Andrés-Hernández, A. Baten, R. Azman Halimi, R. Walls, G. J. King, Knowledge representation and data sharing to unlock crop variation for nutritional food security, Crop Science 60 (2020) 516–529.

[25] S. N. D. A. van Graan, S. K. D. W. A. Masters, K. S. D. F. P. Phiri, A. M. Mwangwela, The malawi food composition database (mafoods) (2020).

[26] D. Vrandečić, M. Krötzsch, Wikidata: a free collaborative knowledgebase, Communications of the ACM 57 (2014) 78–85. URL: https://web.archive.org/web/20190311200511/http://cacm.acm.org/magazines/2014/10/178785-wikidata/fulltext. doi:10.1145/2629489.

[27] B. Mons, C. Neylon, J. Velterop, M. Dumontier, L. O. B. da Silva Santos, M. D. Wilkinson, Cloudy, increasingly fair; revisiting the fair data guiding principles for the european open science cloud, Information Services & Use (2017) 1–8.

[28] J. Neubert, Wikidata as a linking hub for knowledge organization systems? integrating an authority mapping into wikidata and learning lessons for KOS mappings, in: Proceedings of the 17th European Networked Knowledge Organization Systems Workshop co-located with the 21st International Conference on Theory and Practice of Digital Libraries 2017 (TPDL 2017), Thessaloniki, Greece, September 21st, 2017., 2017, pp. 14–25. URL: http://ceur-ws.org/Vol-1937/paper2.pdf.

[29] K. R. Amato, M.-C. Arrieta, M. B. Azad, M. T. Bailey, J. L. Broussard, C. E. Bruggeling, E. C. Claud, E. K. Costello, E. R. Davenport, B. E. Dutilh, H. A. Swain Ewald, P. Ewald, E. C. Hanlon, W. Julion, A. Keshavarzian, C. F. Maurice, G. E. Miller, G. A. Preidis, L. Segurel, B. Singer, S. Subramanian, L. Zhao, C. W. Kuzawa, The human gut microbiome and health inequities, Proceedings of the National Academy of Sciences 118 (2021). URL: https://www.pnas.org/content/118/25/e2017947118. doi:10.1073/pnas.2017947118. arXiv:https://www.pnas.org/content/118/25/e2017947118.full.pdf.

[30] E. Hyvönen, Publishing and using cultural heritage linked data on the semantic web, Synthesis Lectures on the Semantic Web: Theory and Technology 2 (2012) 1–159.

[31] M. L. Zeng, J. Qin, Metadata, American Library Association, 2016.

[32] L. Chan, N. Vasilevsky, A. Thessen, J. McMurry, M. Haendel, The landscape of nutri-informatics: a review of current resources and challenges for integrative nutri-

tion research, Database 2021 (2021). URL: https://doi.org/10.1093/database/baab003. doi:10.1093/database/baab003. arXiv:https://academic.oup.com/database/article-pdf/doi/10.1093/database/baab003/36110502/baab003.pdf, baab003.

[33] P. Puwastien, Issues in the development and use of food composition databases, Public health nutrition 5 (2002) 991–999.

[34] U. Charrondiere, B. Burlingame, Identifying food components: Infoods tagnames and other component identification systems, Journal of Food Composition and Analysis 20 (2007) 713–716.

[35] A. Durazzo, L. D'Addezio, E. Camilli, R. Piccinelli, A. Turrini, L. Marletta, S. Marconi, M. Lucarini, S. Lisciani, P. Gabrielli, et al., From plant compounds to botanicals and back: A current snapshot, Molecules 23 (2018) 1844.