Acquisition and Organization of Food Composition Data Using a Neuro-symbolic Approach[1]

Azanzi Jiomekong

Department of Computer Science - University of Yaounde I, Yaounde, Cameroon - fidel.jiomekong@facsciences-uy1.cm

Germain Kangsi

Department of Biochemistry - University of Yaounde I, Yaounde, Cameroon - kansci2000@gmail.com

Jouonzo Josiane

Department of Animal Biology and Physiology - University of Yaounde I, Yaounde, Cameroon - jouonzo1@gmail.com

Jean Petit BIKIM

Department of Computer Science - University of Yaounde I, Yaounde, Cameroon - jean.bikim@facsciences-uy1.cm

Félicitée Nguefack

Faculty of Medicine and Biomedical Sciences - University of Yaounde I, Yaounde, Cameroon - ngeufackfelicite@gmail.com

Isabelle MEKONE

Faculty of Medicine and Biomedical Sciences - University of Yaounde I, Yaounde, Cameroon - isabelle.mekone@fmsb-uy1.cm

Hélène Kamo Selangai

Faculty of Medicine and Biomedical Sciences, University of Garoua, Garoua, Cameroon - nissilena3@gmail.com

Sören Auer

TIB - Leibniz Information Centre for Science and Technology & L3S Research Center, Hannover, Germany - auer@tib.eu

Abstract

To address the problem of malnutrition, in addition to the improvement of agricultural practices, a solution consists of compiling data about the food consumed and its composition into usable databases such as Food Composition Tables (FCTs) or Food Composition Databases (FCDB). However, existing solutions result in isolated datasets with limited interoperability, poor semantic integration, and restricted machine-readability, which hinder large-scale data linking, querying, and reuse. In this work, we propose to leverage large language models (LLMs) and knowledge graphs (KGs) to build a neuro-symbolic knowledge graph (NeSyKG). The latter aims for the acquisition, organization and diffusion of food composition data in compliance with the FAIR (Findable, Accessible, Interoperable and Reusable) principles. The case study consists of the collection, and organization of Cameroonian food composition data using the neuro-symbolic Open Research Knowledge Graph (ORKG). The main results involve: (1) A neuro-symbolic approach for food composition data acquisition and organization; (2) The extraction and organization of Cameroonian food composition data from 15 scientific papers; (3) and a live review describing Cameroonian FCT, freely available from the ORKG platform.

Keywords: Food Consumption, Food Composition Table, Cameroonian Food Composition Table, Neuro-symbolic Knowledge Graph, Linked and Open Food Data

---

[1] The text and materials are free from any copyright violations

1. Introduction

Malnutrition is an inadequate food/nutrient intake which causes undernutrition (wasting, stunting, and micronutrient deficiencies) along with overweight, obesity and diet-related non-communicable Diseases (NCD) within individuals, households and population throughout life [1, 2, 3]. Despite the rich biodiversity of the African continent and the tremendous progress made in food production, Africa is still struggling with the problems of food insecurity, hunger and malnutrition [2, 3]. Africa is also experiencing a rapid increase in overnutrition linked to the development of NCD while the main diet-related problems involve obesity, CVD, diabetes, hypertension [1]. In developing countries like Cameroon, it has been reported that malnutrition remains a major challenge to public health where the childhood malnutrition rate is more than 35% in certain regions [2, 3]. To address the problem of malnutrition, in addition to the improvement of agricultural practices, a solution consists of compiling data about the food consumed and its composition into usable databases such as Food Composition Tables (FCTs) or Food Composition Databases (FCDs) [6, 9, 10]. In the rest of this paper, we'll be using FCT to designate both FCT and FCD. These databases can be used for a variety of purposes such as clinical practice, research, public health/education and nutrition monitoring [6, 9, 10]. To illustrate, Table 1 [12] presents an example of a FCT of some Cameroonian foods. It shows the ingredients composition of these foods and their nutritional composition. This table shows for instance that the consumption of Ekwang, Tenue militaire and Koki (100 g) by children aged between 1-2 years would meet more than 100% of their daily recommended intake for vitamin A. These foods can thus be recommended for children suffering from vitamin A deficiency.

Table 1: Example of the composition of some Cameroonian foods and the dietary reference intake established by the Food and Nutrition Board [12].

Local name	Ekomba	Ekwang	Tenue militaire	Koki	Dietary reference intake
Dish Form	Bundle	Compact paste	Bundle	Bundle	
Ingredient (%)	Dried maize grains (58.2), dried peanut seeds (23.5), crayfish (2.1), salt (0.9), Crude palm oil (1.6), pepper (0.6), water (13.0)	Cocoyam tubers (43.6), Cocoyam leaves (25.6), smoked fish (3.5), crayfish (1.7), water (17.8), pepper (0.4), crude palm oil (5.9), maggi cube (0.1), onion (1.6)	Dried maize grains (41.8), cocoyam leaves (30.7), salt (0.2), maggi cube (0.2), pepper (0.4), crude palm oil (6.9), crayfish (3.4), water (16.5)	Dried cowpea seeds (77.1), crude palm oil (5.6), water (15.4), pepper (1.4), salt (0.4)	
Protein	38.2	10.8	16.4	41.6	13 g/day
Fat	26.6	14.2	17.5	19.5	30 g/day
Dietary fiber	9	1.1	4.1	4.9	19 g/day
Carbohydrates	17.1	11.9	11.1	11.7	130 g/day
Energy	13.4	7.8	8.4	10	1350 cal/day
Calcium	6.1	2.7	4.9	7.8	500 mg/day
Magnesium	38.4	16.2	21.6	30.6	80 mg /day
Sodium	56.8	33.6	37.5	51.6	1000 mg/day
Potassium	3.2	4.2	2.1	6.1	3000 mg/day
Iron	7.2	16.6	21.6	19.4	7 mg/day
Zinc	21.2	8.9	10.6	37.8	3 mg/day
Copper	36.2	55.9	49.9	36.7	0.34 mg/day
Manganese	29.5	23.3	34.4	31.5	1.2 mg/day
Vitamin A	35.9	108.8	119.2	119.1	0.3 mg RAE/day

At the Africa level, many strategies such as African Regional Nutrition Strategy (2015-2025), the African Union Malabo Declaration (2014) and the African Union Agenda 2063 outline the specific role of the African Union (AU) and the AU Commission (AUC) in the elimination of hunger and malnutrition. On the other hand, African Network of Food Data systems (AFROFOODS) aims to coordinate at the regional level the activities on food composition. Thus, several Food Composition Databases (FCDB) have been developed [7, 8]. However, these solutions suffer many problems, the main problems are: (1) Static databases sometimes in print form or in Excel format - most FCT (particularly, those published in scientific literature) are compiled using their own procedures, classification and indexing systems. Consequently, the data from these datasets are not only scattered, but also not harmonized; (2) Whilst many countries have a national or regional FCT, most of them contain incomplete data - the comparison of several FCT (http://www.orkg.link2comparisontable) showed for instance that, FCT should be always updated because eating habits change over time; (3) FCTs are not easily accessible to the general public. For the particular case of Cameroon, the most recent versions of the Cameroonian Food Composition Tables (FCT) were published in 1957 [4] and 1966 [5]. These resources are outdated given the current evolution of the country in terms of food habits.

In this paper, we propose a novel approach for the acquisition and organization of food composition data. This approach consists of leveraging large language models (LLMs) and knowledge graphs (KGs) [9] to build a neuro-symbolic knowledge graph (NeSyKG) for the acquisition, organization and diffusion of food composition data in compliance with the FAIR (Findable, Accessible, Interoperable and Reusable) principles. The case study consisted of the extraction and organization of Cameroonian food composition data using the the neuro-symbolic Open Research Knowledge Graph (ORKG) [13, 14, 15]. In the rest of this paper, Section 2 presents the acquisition and organization of food composition data, Section 3 presents the neuro-symbolic approach proposed in this work for the acquisition and organization of food composition data. Section 4 presents the Cameroonian case study and the conclusion and future work is presented in Section 5.

2. Acquisition and Organization of Food Composition Data

Food Composition Data Acquisition. There are two ways to build food composition tables [7, 8, 9, 10, 11, 12]. The first consists of chemical analysis of food. This method is expensive and needs appropriate materials and skills. The second method consists of extracting food composition knowledge from existing resources such as databases and scholarly data. Currently, the acquisition of food composition data for the construction of FCTs is primarily performed manually by domain experts. In many cases, data are also borrowed or adapted from existing FCTs [7, 8, 9, 10]. While acquisition offers the advantage of expert validation and high reliability, it presents several important limitations, including high time consumption, limited scalability, susceptibility to human error, and challenges in keeping datasets up to date as new scientific evidence becomes available. For instance, the development of an updated version of the West African FCT took around four years and involved numerous researchers, nutritionists, and institutions [7]. In a previous work, we proposed leveraging machine learning (ML) to automatically extract food composition data from scholarly communication [11]. During this work, we found that automated approaches can accelerate data extraction, support large-scale integration of existing FCTs, and facilitate continuous updating. However, automation must be designed carefully to preserve data quality.

Food Composition Data Organization. Once extracted, food composition data organization aims to put the food information in a structured way and ideally, machine readable. Food information can be organized using tabular data or using AI techniques. AI techniques for organizing food information consist of symbolic organization, connectionist organization and neuro-symbolic organization [6].

- **Tabular organization**: In many cases, food composition data is organized into a tabular form. This consists of storing food, food ingredients and food composition using either spreadsheet or relational databases [6, 7, 8, 9, 10, 11, 12]. Thereafter, SQL (or equivalent query language) is used to query the database for several purposes such as food ingredients, the composition of a particular food, etc. To help countries develop their FCT, FAO provides FAO/INFOODS compilation tools in MS Excel format for storing, documenting and managing FCT electronically [12]. This standard has been used to build several food composition tables. But in many cases, countries developed their own FCT and stored it in relational databases. Tabular organization results in isolated datasets (published on the INFOOD website or country/region website (https://www.fao.org/infoods/infoods/tables-and-databases/en/)) with limited interoperability, poor semantic integration, and restricted machine-readability, which hinder large-scale data linking, querying, and reuse.

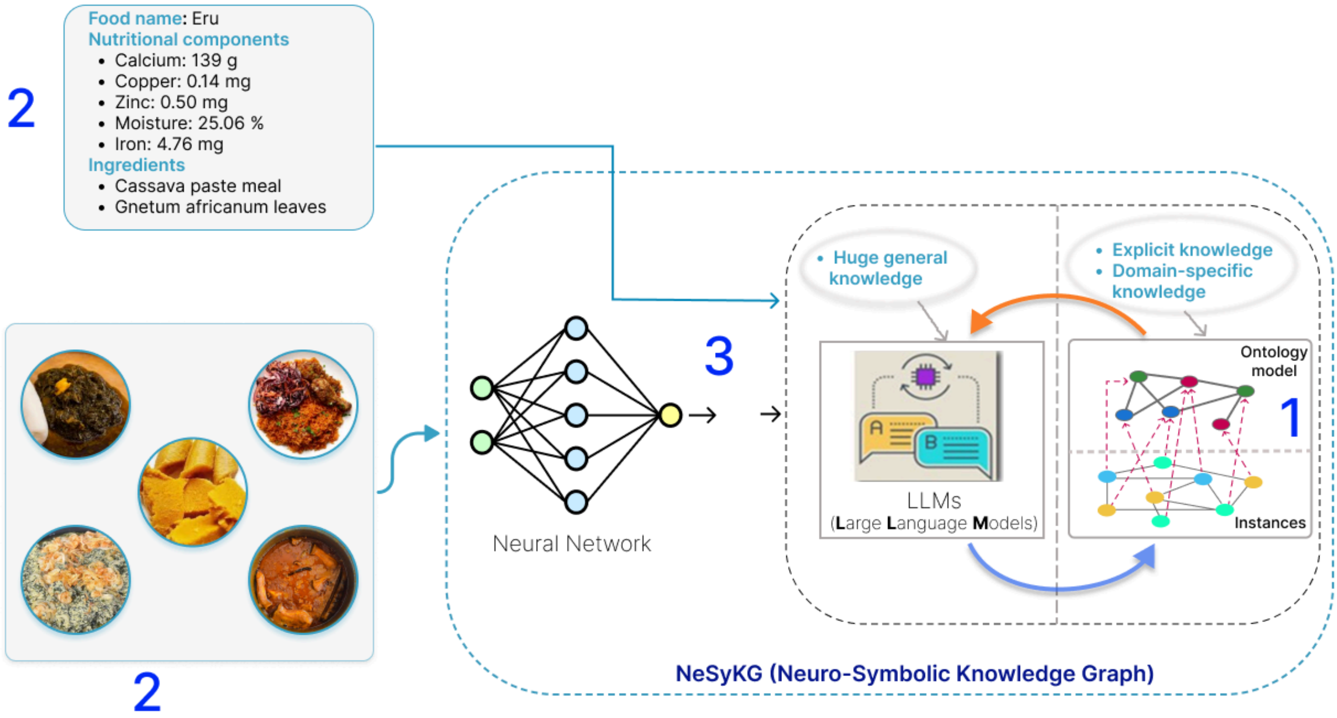
- **Symbolic organization** [6, 14]: the symbolic organization uses symbols to represent background food knowledge. Due to the large-scale nature of these data, they should be organized so as to be in a machine and human readable form and facilitate their exploration and exploitation. To this end, food information is linked together forming either food classification systems, food ontologies, food knowledge graphs or food linked data [17, 18]. An ultimate example is the AGROVOC Linked Open Dataset about agriculture [18].
- **Connectionist organization** [6, 14]: connectionist organization consists of learning associations from food data and storing these information in the form of connections between neurons. To this end, it requires a lot of data. Using these data, it adjusts the strength of the connections (weights) between its nodes (or neurons). These models are used to provide food composition data to users given a prompt. However, the main problem in Africa is the lack of food datasets for developing connectionist models.
- **Neuro-symbolic organization** [6, 14]: neuro-symbolic organization proposes the combination of connectionist and symbolic organization and simultaneously leverages their advantages.

Although several works have been done on representing food information using symbolic and connectionist AI, neuro-symbolic AI is still rarely used [6]. However, this model should be considered for several reasons: **(1) Multimodal nature of food information:** Food information involves food images and text. Therefore, one may consider a scenario in which a connectionist model for food recognition is combined with the symbolic model (e.g., a food composition knowledge graph) for nutrient generation; **(2) Explainability:** Because the processing of connectionist models is often opaque, infusing symbolic knowledge into these models may help to understand the predictions obtained. For instance, one may combine a Large Language Model (LLM) for food Question Answering with a food KG in order to understand the prediction of the LLM; **(3) Enable multi-step inference:** Connectionist models can be trained on symbolic models such as KG to enable multi-step inference. An example of a three-step inference: from a food name/image, infer the list of ingredients composition; from the food ingredient, infer the nutrient; from the nutrient, infer if the food can be recommended to a person in diet.

3. Methodology

In this work, we propose a new approach for the acquisition and organization of food composition data. The overall workflow of the proposed approach is illustrated in Fig. 1. Unlike conventional approaches proposed and used by INFOODS that primarily rely on spreadsheets, relational databases, or static PDF documents, our method explicitly considers the multimodal nature of food information. Actually, food composition data are not limited to textual nutrient description; they also include visual representations of food items. However, existing FCTs typically provide textual and numerical data only, and when images are included, they are embedded as static figures in PDF documents. These images are usually limited to one illustration per food item and are not structured or reusable for computational purposes such as training image recognition models.

To overcome these limitations, we propose in this work a neuro-symbolic approach for the collection and the organization of food composition data that integrates textual, numerical, and visual information. It allows us to link structured food composition data to corresponding image data within a unified neuro-symbolic knowledge graph. For clarity, the steps shown in Fig. 1 are numbered and described below.



Food name: Eru
Nutritional components
- Calcium: 139 g
- Copper: 0.14 mg
- Zinc: 0.50 mg
- Moisture: 25.06 %
- Iron: 4.76 mg

Ingredients
- Cassava paste meal
- Gnetum africanum leaves

Figure 1: Neuro-symbolic framework for food composition data acquisition and organization

Step 1 (1): Ontology Design - Ontology design step is the fundamental step of our methodology

An ontology is a formal, explicit specification of a shared conceptualization of a domain [16]. It defines entities (classes), their properties (attributes), relationships between entities, and constraints governing their interactions. In this work, we designed a food composition ontology with the guidance of a professor in food science and nutrition (see Fig. 2) that defines *food entities* (e.g., raw food, composite dishes), *food ingredients* (e.g., banana, pepper), *food composition* (e.g., carbohydrate, fat), *data source* (e.g., scientific publication, database, chemical analysis of food) and *report entities* (e.g., Cameroon FCT, Chat FCT). In addition to textual and numerical properties, the ontology includes object properties that link food entities to corresponding image URLs. This enables the development of a multi-modal dataset in which structured text data and image data coexist.

Step 2 (2): Data Collection

Data collection involves gathering food composition information (either from chemical analysis or from research reports) in textual and numerical form, along with corresponding image data. The objective is to construct a rich, multi-source dataset that reflects both the compositional and visual dimensions of food items. Textual data includes food items, dishes, ingredient list, food composition data, food description, and metadata such as country of origin. Image data includes multiple images per food item when available (see Fig. 1). Unlike traditional FCT compilation, which focuses exclusively on tabular nutrient value, our collection ensures that each food entity is associated with structured semantic information and one or more representative images.

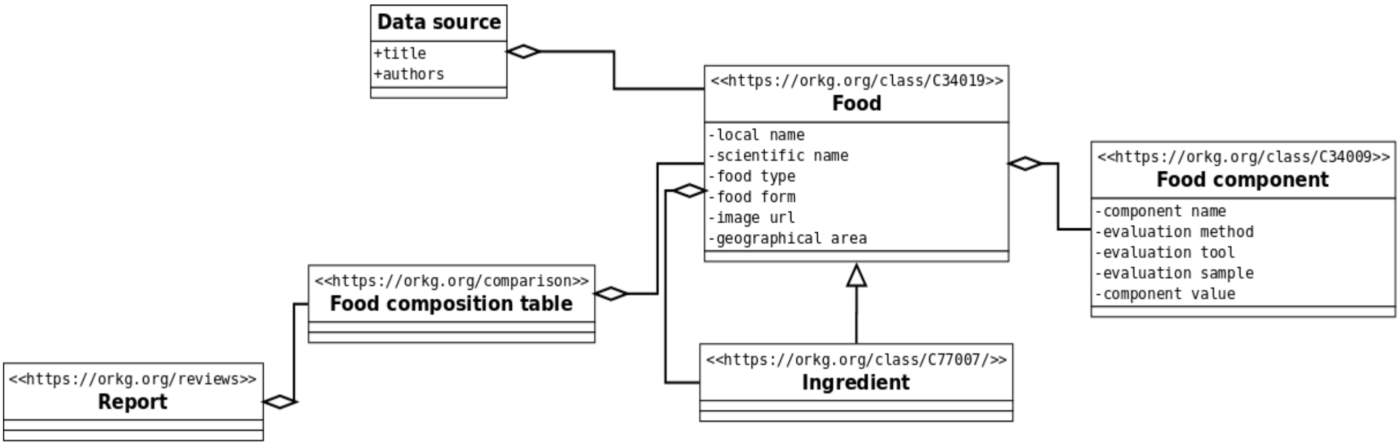


Figure 2: Food Composition Ontology

Step 3 (3): Data Organization in a neuro-symbolic knowledge graph

Once collected, the data are organized within a neuro-symbolic knowledge graph. This graph consists of a symbolic component (the Knowledge Graph) and neural components (a LLM and a deep learning model). The KG is used to organize textual and numerical data by instantiating ontology classes and relationships. Each food item (e.g., koki in Fig. 1) becomes a node in the graph, linked to its ingredients, components, etc. During the process of ontology population, a LLM operates as an intelligent assistant within a human-in-the-loop validation. Unlike traditional approaches presented in Section 2, where domain experts manually interpret and enter data in the FCT, in this approach, the LLM is used for suggesting appropriate classes and properties for newly entered data, and experts perform validation and refinement when necessary. In our experience, this significantly reduces manual workload while maintaining data quality and reliability. Another component of the neuro-symbolic KG is a deep learning model trained on the collected food images for food image recognition. Once food entities are structured within the KG, FCTs are manually constructed from the graph. Unlike static spreadsheets, these tables are machine-readable, dynamically updatable, and linked to image data. The FCTs is used to write structured reports and review documents (like the ones on the INFOOD website) describing food composition data, including references to data sources and visual representations.

4. Case study

This section describes how Cameroonian foods, their ingredients, and their nutritional components are extracted and organized to construct a Cameroonian food composition knowledge graph using the ORKG [13-15]. This work is conducted by a multidisciplinary team composed of artificial intelligence experts, nutritionists, and medical doctors. Nutritionists are responsible for ensuring data quality and medical doctors contribute domain knowledge related to public health and clinical nutrition, with the objective of identifying which foods are appropriate for different forms of malnutrition based on their nutritional composition. AI experts contribute to the acquisition and organization of food data and the training of data curators. Before delving into the case study, the ORKG is presented.

4.1. The Open Research Knowledge Graph

The ORKG is a neuro-symbolic knowledge graph designed to semantically structure and interconnect scholarly information, blending the strengths of symbolic AI with neural network approaches [13, 14, 15]. It is built according to the principles of Open Science, Open Data, and Open Source. Therefore, all the data ingested in ORKG are freely available on the Web and via API, dump, and integrations, can be identified by a Uniform Resource Identifier (URI), and can be accessed via HTTP or using a software library.

In this work, the following features of the ORKG are used:

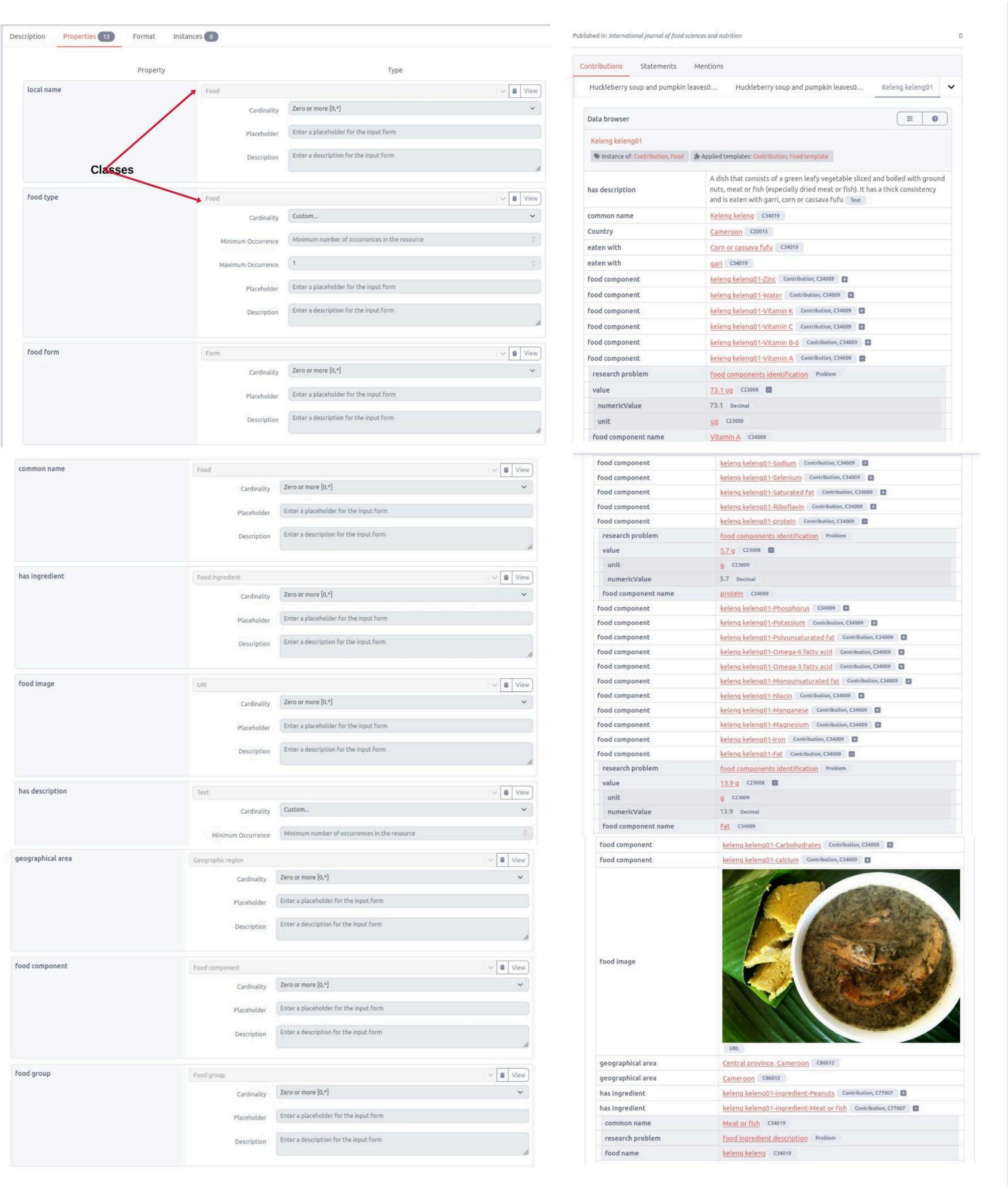
- **ORKG lists:** Allow us to organize resources of a specific research domain. In this work, the lists are used to group research papers related to the Cameronian food.
- **Resources metadata:** This feature aims to extract and organize metadata (e.g., title, authors, journal, book title, etc.,) of scholarly resources.
- **Semantic description of resources:** The semantic description of a resource consists of the annotation of this resource with key insights extracted from them and organizing these elements into research contributions. This allows us to put a scientific paper in a machine-readable form following the RDF paradigm. Contributions can be compared between them or by other contributions from other resources in the ORKG. During the process of the semantification of a resource, ChatGPT is used for suggestions to users. The user can choose to use the ChatGPT suggestions directly, refine them before use, or decide not to use them.
- **Templates:** These are used during the description of contributions to facilitate data entry and comparability. To build templates, classes, and properties allowing the description of contributions of the domain are identified.
- **Comparison Tables:** The structured content descriptions of resources into research contributions is done in such a way that the contribution becomes comparable with other contributions. For instance, in our case, this feature is used to build food composition tables in which different foods are presented as contributions and properties consisting of ingredients and food components. Once built, food composition tables can be published with a DOI, exported in various formats such as RDF, LaTeX, PDF, CSV, and integrated into a living dynamic FCT such as WAFCT. It should be noted that unlike existing FCT, these tables can be updated, by correcting errors or including additional data (e.g., additional food, ingredients, components).
- **Smart Reviews:** The ORKG provides a "What You See Is What You Get (WYSIWIG)" editor for helping users summarize the work in the system. In this work, this feature is used for organizing the whole FCT into a document such as the WAFCT. As the comparison tables, Smart Reviews can be edited, corrected, improved and re-published.

4.2. Data Collection

Data collection consists of the collection of food composition data and food images. Chemical analysis of foods is the most direct and reliable method for acquiring nutrient composition data [9, 12]. However, in this work, we could not afford required costs and skilled personnel needed. Therefore, for the proof of concept, we decided to adopt an indirect method based on pre-existing data from scholarly publications. It should be noted that this work can be easily extended to incorporate data from chemical analysis. The sources of food composition data were scholarly publications describing the chemical analysis of Cameroonian foods [19-32]. More than 15 papers were collected and organized using the ORKG list (https://orkg.org/lists/R1355426). Images corresponding to Cameroonian foods are collected from the Internet and mobile phone.

4.3. Ontology Implementation and Population

The food composition ontology was implemented using the ORKG and is publicly available. To this end, main classes such as *Food, Food Components and Food Ingredient* were implemented as classes in the ORKG. Each class is linked to the corresponding template. As presented in Section 4.1, ORKG templates provide a uniform structure for populating ontology classes. Each instance of a class is entered through its corresponding template, ensuring consistency and standardization. In addition, templates enable the creation of semantic relationships between ontology classes. For instance, for the class Food, a dedicated food template has been designed (https://orkg.org/templates/R677398/properties). For the statement *"Food has Food Composition"*, the property *food component* is defined, enabling the association between a food entity and its corresponding food components. It should be noted that *Food Component* class also has its own template (https://orkg.org/templates/R684336/properties), containing multiple properties specified. Fig. 3 illustrates an example in which a food template is semantically connected to both the food component and food ingredients (on the left). This figure also presents an example of a food instance within the ORKG (on the right).



Property | Type

local name: Food
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

food type: Food
Cardinality: Custom...
Minimum Occurrence: Minimum number of occurrences in the resource
Maximum Occurrence: 1
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

food form: Form
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

common name: Food
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

has ingredient: Food ingredient
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

food image: URI
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

has description: Text
Cardinality: Custom...
Minimum Occurrence: Minimum number of occurrences in the resource

geographical area: Geographic region
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

food component: Food component
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

food group: Food group
Cardinality: Zero or more [0,*]
Placeholder: Enter a placeholder for the input form
Description: Enter a description for the input form

Published in: *International journal of food sciences and nutrition*

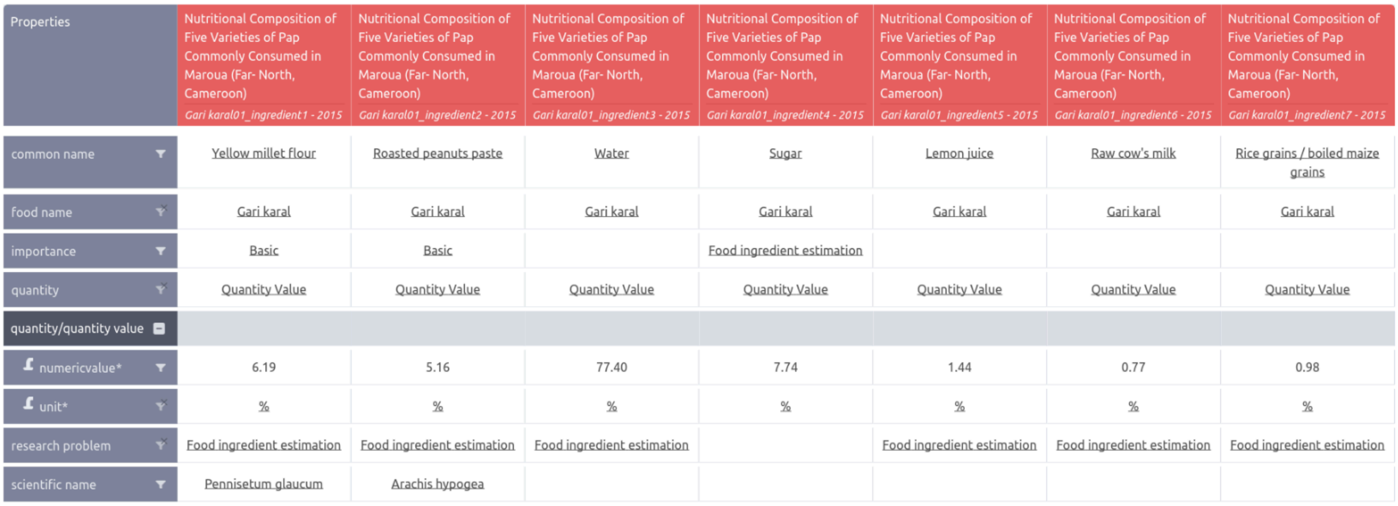
Keleng keleng01

Instance of: Contribution, Food | Applied templates: Contribution, Food template

has description	A dish that consists of a green leafy vegetable sliced and boiled with ground nuts, meat or fish (especially dried meat or fish). It has a thick consistency and is eaten with garri, corn or cassava fufu Text
common name	Keleng keleng C34019
Country	Cameroon C20015
eaten with	Corn or cassava fufu C34019
eaten with	gari C34019
food component	keleng keleng01-Zinc Contribution, C34009
food component	keleng keleng01-Water Contribution, C34009
food component	keleng keleng01-Vitamin K Contribution, C34009
food component	keleng keleng01-Vitamin C Contribution, C34009
food component	keleng keleng01-Vitamin B-6 Contribution, C34009
food component	keleng keleng01-Vitamin A Contribution, C34009
research problem	food components identification Problem
value	73.1 ug C23008
numericValue	73.1 Decimal
unit	ug C23009
food component name	Vitamin A C34009
food component	keleng keleng01-Sodium Contribution, C34009
food component	keleng keleng01-Selenium Contribution, C34009
food component	keleng keleng01-Saturated fat Contribution, C34009
food component	keleng keleng01-Riboflavin Contribution, C34009
food component	keleng keleng01-protein Contribution, C34009
research problem	food components identification Problem
value	5.7 g C23008
unit	g C23009
numericValue	5.7 Decimal
food component name	protein C34009
food component	keleng keleng01-Phosphorus C34009
food component	keleng keleng01-Potassium Contribution, C34009
food component	keleng keleng01-Polyunsaturated fat Contribution, C34009
food component	keleng keleng01-Omega-6 fatty acid Contribution, C34009
food component	keleng keleng01-Omega-3 fatty acid Contribution, C34009
food component	keleng keleng01-Monounsaturated fat Contribution, C34009
food component	keleng keleng01-Niacin Contribution, C34009
food component	keleng keleng01-Manganese Contribution, C34009
food component	keleng keleng01-Magnesium Contribution, C34009
food component	keleng keleng01-Iron Contribution, C34009
food component	keleng keleng01-Fat Contribution, C34009
research problem	food components identification Problem
value	13.9 g C23008
unit	g C23009
numericValue	13.9 Decimal
food component name	Fat C34009
food component	keleng keleng01-Carbohydrates Contribution, C34009
food component	keleng keleng01-calcium Contribution, C34009
food image	URL
geographical area	Central province, Cameroon C86012
geographical area	Cameroon C86012
has ingredient	keleng keleng01-ingredient-Peanuts Contribution, C77007
has ingredient	keleng keleng01-ingredient-Meat or fish Contribution, C77007
common name	Meat or fish C34019
research problem	food ingredient description Problem
food name	keleng keleng C34019

Figure 3: An example of a food template (on the left) and food instance (on the right)

Once the ontology was implemented, its population consisted of data entry and validation with ChatGPT assistance to suggest ontology classes, properties, and relationships between foods, nutrients. Food names were chosen to describe the food items as appropriately as possible. Most of the composite dishes are given by their Cameroonian names. For each food item, multiple images are collected when available to improve variability for potential image-based food recognition tasks and each image is manually linked to its corresponding food entry. Foods and food components are used to build FCT. An excerpt of several foods are presented by Fig. 4, 5, and 6. These figures are provided as additional materials.

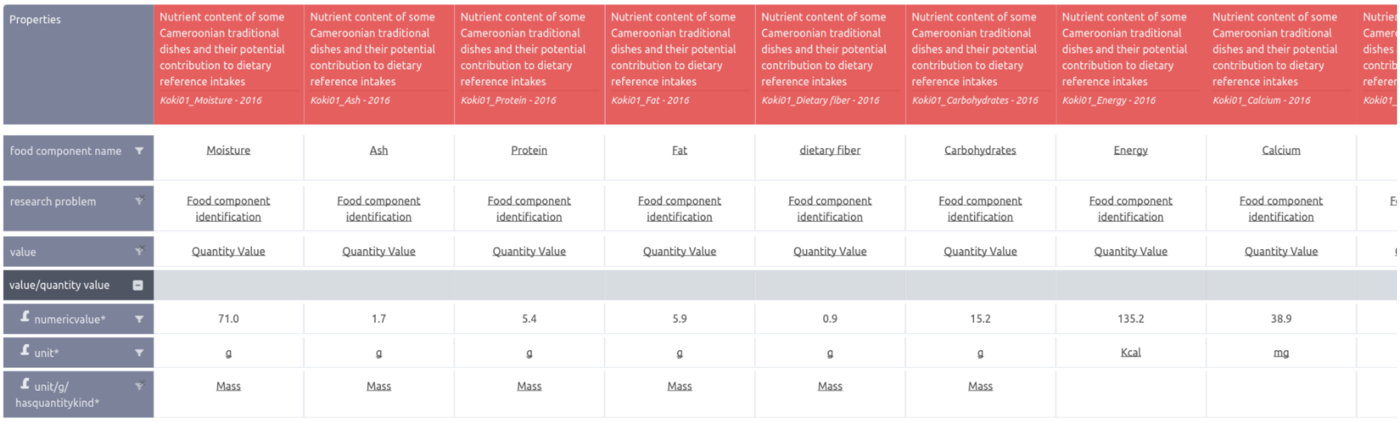


Properties	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient1 - 2015*	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient2 - 2015*	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient3 - 2015*	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient4 - 2015*	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient5 - 2015*	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient6 - 2015*	Nutritional Composition of Five Varieties of Pap Commonly Consumed in Maroua (Far- North, Cameroon) *Gari karal01_ingredient7 - 2015*
common name	Yellow millet flour	Roasted peanuts paste	Water	Sugar	Lemon juice	Raw cow's milk	Rice grains / boiled maize grains
food name	Gari karal	Gari karal	Gari karal	Gari karal	Gari karal	Gari karal	Gari karal
importance	Basic	Basic		Food ingredient estimation			
quantity	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value
quantity/quantity value							
numericvalue*	6.19	5.16	77.40	7.74	1.44	0.77	0.98
unit*	%	%	%	%	%	%	%
research problem	Food ingredient estimation	Food ingredient estimation	Food ingredient estimation		Food ingredient estimation	Food ingredient estimation	Food ingredient estimation
scientific name	Pennisetum glaucum	Arachis hypogea					

Figure 4: An example of a comparison table, describing and comparing ingredient composition of Gari karal

Properties	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_calcium*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Magnesium*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Phosporus*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Potassium*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Zinc*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Copper*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Iron*	Mineral content in some Cameroonian household foods eaten in Douala *Macabo Ndolè01_Manganese*
food component name	calcium	magnesium	phosporus	potassium	zinc	copper	Iron	manganese
research problem	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification
sample	100g fresh weight	100g fresh weight	100g fresh weight	100g fresh weight	100g fresh weight	100g fresh weight	100g fresh weight	100g fresh weight
value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value
value/quantity value								
numericvalue*	309.14	263.83	742.88	3243.90	20.57	0.87	15.15	2.86
unit*	mg	mg	mg	mg	mg	mg	mg	mg

Figure 5: "Macabo Ndolè" nutritional composition (https://doi.org/10.48366/R1586215)



Properties	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Moisture - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Ash - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Protein - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Fat - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Dietary fiber - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Carbohydrates - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Energy - 2016*	Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes *Koki01_Calcium - 2016*	Nutrien Camer dishes contrib referer *Koki01_*
food component name	Moisture	Ash	Protein	Fat	dietary fiber	Carbohydrates	Energy	Calcium	
research problem	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	Food component identification	F
value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	Quantity Value	
value/quantity value									
numericvalue*	71.0	1.7	5.4	5.9	0.9	15.2	135.2	38.9	
unit*	g	g	g	g	g	g	Kcal	mg	
unit/g/ hasquantitykind*	Mass	Mass	Mass	Mass	Mass	Mass			

Figure 6: “Koki” nutritional composition (https://doi.org/10.48366/R1586241)

4.5. Use case Summary

The Cameroonian FCT presented in this case study provides, for the first time since 1960, information on relevant food components of the main food items consumed in Cameroon. This updated version describes 253 foods, 85 food ingredients and 71 food components. Compared to the 1966, 1957 Cameroonian FCT, this updated version is a significant improvement in terms of completeness. Despite its limitations, it is a useful tool for assessing Cameroonian nutrient intakes, accessible to the public, and it should be continuously updated. Finally, we hope that this work will be considered as a first step for producing future releases with more chemical analysis (many do not have many chemical analysis), a high number of food items, and high-quality composition data. Indeed, more research and more collaborative work are necessary to achieve this objective.

The main results of this case study involves:

- A list of papers used for the collection and organization of Cameroonian FCT (https://orkg.org/lists/R1355426)
- The description of more than 253 Cameroonian foods 85 ingredients and 71 components
- The creation of templates used for the description of food (https://orkg.org/templates/R677398/), food ingredients (https://orkg.org/templates/R677408), and food component (https://orkg.org/templates/R684336/)
- A life review describing Cameroonian FCT (https://orkg.org/reviews/R700894)

5. Conclusion and future work

This paper presents an approach for the acquisition and organization of food composition data based on a neuro-symbolic framework. Unlike traditional FCT development processes, which primarily rely on manual curation, spreadsheets, static PDF documents, or isolated databases, the proposed approach explicitly considers the multimodal nature of food information by integrating textual, numerical, and image data. Through the Cameroonian case study, we demonstrated the feasibility and relevance of the proposed approach in a context where food data are fragmented, heterogeneous, and often difficult to access.

Future work will extend the proposed approach by integrating additional Cameroonian food, additional regional datasets, (West African FCT, Kenyan FCT, Tunisia FCT, etc.), agricultural data, and evaluating the system in real-world agriculture, nutrition and public health settings. We are currently using the structured data in the ORKG to develop an intelligent food assistant capable of estimating nutritional values from food images. A prototype is already available online (https://tsotsa.org/tsotsa-assistant). Such a system could support dietary decision-making in context like Cameroon, where malnutrition and diet-related health challenges remain prevalent. Finally, we aim to design a semi-automated pipeline for generating FCT with human-in-the-loop validation, which would be particularly beneficial for many developing countries lacking national FCT resources.

References

[1] Owino VO. Challenges and opportunities to tackle the rising prevalence of diet-related non-communicable diseases in Africa. Proceedings of the Nutrition Society. 2019 Nov;78(4):506-12.

[2] Bita AI, Chukuwchindun BA, Biwole TE, Omgba AN, Eyong JB, Sanou MS. Qualitative Analysis of the Determinants of Malnutrition among Children Under-five in Households in the Mayo-Danay Division, Cameroon. Budapest International Research in Exact Sciences (BirEx) Journal. 2020;3:755-66.

[3] Jiomekong A, Nguefac F. First Paediatric Nutrition Life Cycle Congress 2026 (2026) [Internet]. Available from https://sites.google.com/view/sonup-cv-congres2026/2026-proceedings.

[4] Bergeret B, Masseyeff R. Table provisoire de composition des aliments du Sud-Cameroun. Ann. Nutrition et Aliment. 1957;11:47.

[5] Joseph P, Simone B. Les Aliments D’Origine Végétale Au Cameroun. Le Cameroun Agricole, Pastoral et Forestier. 1967;110:49-66.

[6] Jiomekong A, Oelen A, Auer S, Anna‐Lena L, Lars V. Food information engineering. AI Magazine. 2024 Sep;45(3):338-53.

[7] Vincent A, Grande F, Compaoré E, Amponsah Annor G, Addy PA, Aburime LC, Ahmed D, Bih Loh AM, Dahdouh Cabia S, Deflache N, Dembélé FM. FAO/INFOODS Food Composition Table for Western Africa (2019)/Table de composition des aliments FAO/INFOODS pour l’Afrique de l’Ouest (2019).

[8] Mwai J, Kimani A, Mbelenga E, Charrondiere UR, Grande F, Rittenschober D, Vincent A, Murugu D. Kenya food composition tables 2018. FAO: Rome, Italy. 2018.

[9] Greenfield H, Southgate DA. Food composition data: production, management, and use. Food & Agriculture Org.; 2003.

[10] Pennington JA. Applications of food composition data: Data sources and considerations for use. Journal of Food Composition and Analysis. 2008 Feb 1;21:S3-12.

[11] Jiomekong A, Folefac M, Tapamo H. Food composition knowledge extraction from scientific literature. InInternational Conference on Artificial Intelligence: Towards Sustainable Intelligence 2023 Sep 4 (pp. 89-103). Cham: Springer Nature Switzerland.

[12] Charrondiere UR, Burlingame B. Report on the FAO/INFOODS Compilation Tool: A simple system to manage food composition data. Journal of Food Composition and Analysis. 2011 Jun 1;24(4-5):711-5.

[13] Jiomekong, A., Auer, S., & Oelen, A. (2024, May). Linked Open Literature Review using the Neuro-symbolic Open Research Knowledge Graph. In *Companion Proceedings of the ACM on Web Conference 2024* (pp. 1015-1018).

[14] Auer S, D’Souza J, Farfar KE, Jaradeh MY, Jiomekong A, Karras O, Oelen A, Snyder L, Stocker M, Vogt L. Open research knowledge graph: a large-scale neuro-symbolic knowledge organization system. InHandbook on Neurosymbolic AI and Knowledge Graphs 2025 (pp. 385-420). IOS Press.

[15] Jiomekong A, Tiwari S. An approach based on open research knowledge graph for knowledge acquisition from scientific papers. The Electronic Library. 2024 Jun 27;42(3):413-42.

[16] Guarino N, Oberle D, Staab S. What is an ontology?. InHandbook on ontologies 2009 May 22 (pp. 1-17). Berlin, Heidelberg: Springer Berlin Heidelberg.

[17] Kolchin M, Chistyakov A, Lapaev M, Khaydarova R. FOODpedia: Russian food products as a linked data dataset. InEuropean Semantic Web Conference 2015 May 31 (pp. 87-90). Cham: Springer International Publishing.

[18] Caracciolo C, Stellato A, Morshed A, Johannsen G, Rajbhandari S, Jaques Y, Keizer J. The AGROVOC linked dataset. Semantic Web. 2013 Aug;4(3):341-8.

[19] Ponka R, Fokou E, Beaucher E, Piot M, Gaucheron F. Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes. Food Science & Nutrition. 2016 Sep;4(5):696-705.

[20] Sop MM, Gouado I, Mofor CT, Smriga M, Fotso M, Ekoe T. Mineral content in some Cameroonian household foods eaten in Douala. African Journal of Biotechnology. 2008;7(17).

[21] Gimou MM, Charrondière UR, Leblanc JC, Pouillot R, Noël L, Guérin T. Concentration data for 25 elements in foodstuffs in Yaounde: the Cameroonian Total Diet Study. Journal of food composition and analysis. 2014 May 1;34(1):39-55.

[22] Ponka R, Fokou E, Leke R, Fotso M, Souopgui J, Bih MA, Tchouanguep FM. Methods of preparation and nutritional evaluation of dishes consumed in a malaria endemic zone in Cameroon (Ngali II). African Journal of Biotechnology. 2005 Aug 12;4(3):273-8.

[23] Sharma S, Claude Mbanya J, Cruickshank K, Cade J, Tanya AK, Cao X, Hurbos M, Wong MR. Nutritional composition of commonly consumed composite dishes from the Central Province of Cameroon. International journal of food sciences and nutrition. 2007 Jan 1;58(6):475-85.

[24] Elie F, Roger P, Honoris TD, Brice DK, Blonde TL, Bih AM, Martin F. Methods of preparation and nutritive value of some dishes consumed in the West Region of Cameroon. Pakistan Journal of Nutrition. 2009 Jul 15;8(8):1190-5.

[25] Kansci G, Fokou E, Piot M, Leonil J, Ponka R, Gaucheron F, Beaucher E. Recipes proximate and mineral compositions of some traditional sauces consumed in the Far North Region of Cameroon.

[26] Ponka R, Bouba AA, Fokou E, Beaucher E, Piot M, Leonil J, Gaucheron F. Nutritional composition of five varieties of pap commonly consumed in Maroua (Far-North, Cameroon). Polish Journal of Food and Nutrition Sciences. 2015;65(3).

[27] Feumba R. Recipes and Proximate Composition of Some Traditional African Nightshade Dishes Consumed in the North West Region of Cameroon. Journal of food sciences. 2023 Feb 26.

[28] Ponka R, Fokou E, Kansci G, Beaucher E, Piot M, Leonil J, Gaucheron F. Amino acids, major carotenoids and vitamin A activity of some traditional sauces consumed in the Far North Region of Cameroon. Journal of Food Composition and Analysis. 2015 Nov 1;43:88-95.

[29] Ponka R, Fokou E, Beaucher E, Piot M, Gaucheron F. Nutrient content of some Cameroonian traditional dishes and their potential contribution to dietary reference intakes. Food Science & Nutrition. 2016 Sep;4(5):696-705.

[30] Fernande NB, Raissa N, Sandrine BN, Diana MN, Martine NE, Nadege NM. Macronutrient values of local meals of some Cameroonian Traditional Communities Living in Yaounde. Journal of Food and Nutrition Sciences. 2021;9(2):57-63.

[31] Ejoh RA, Nkonga DV, Inocent G, Moses MC. Nutritional components of some non-conventional leafy vegetables consumed in Cameroon. Pakistan Journal of Nutrition. 2007 Oct 15;6(6):712-7.

[32] Ngaha DW, Fombang EN, Ejoh RA. Dietary intake of vitamin A and macronutrients among pregnant women in Ngaoundere Town, Adamawa Region, Cameroon. Food and Nutrition Sciences. 2014 Nov 20;5(21):2071-80.