

Matching Food Items with FoodOn Identifiers in a Knowledge Base

Kat Thornton, Kenneth Seals-Nutt, Mika Matsuzaki

15 August 2022

Our Collaboration



Figure 1: Mika Matsuzaki- nutritional epidemiologist,
Kenneth Seals-Nutt- software engineer, **Kat Thornton-** information scientist

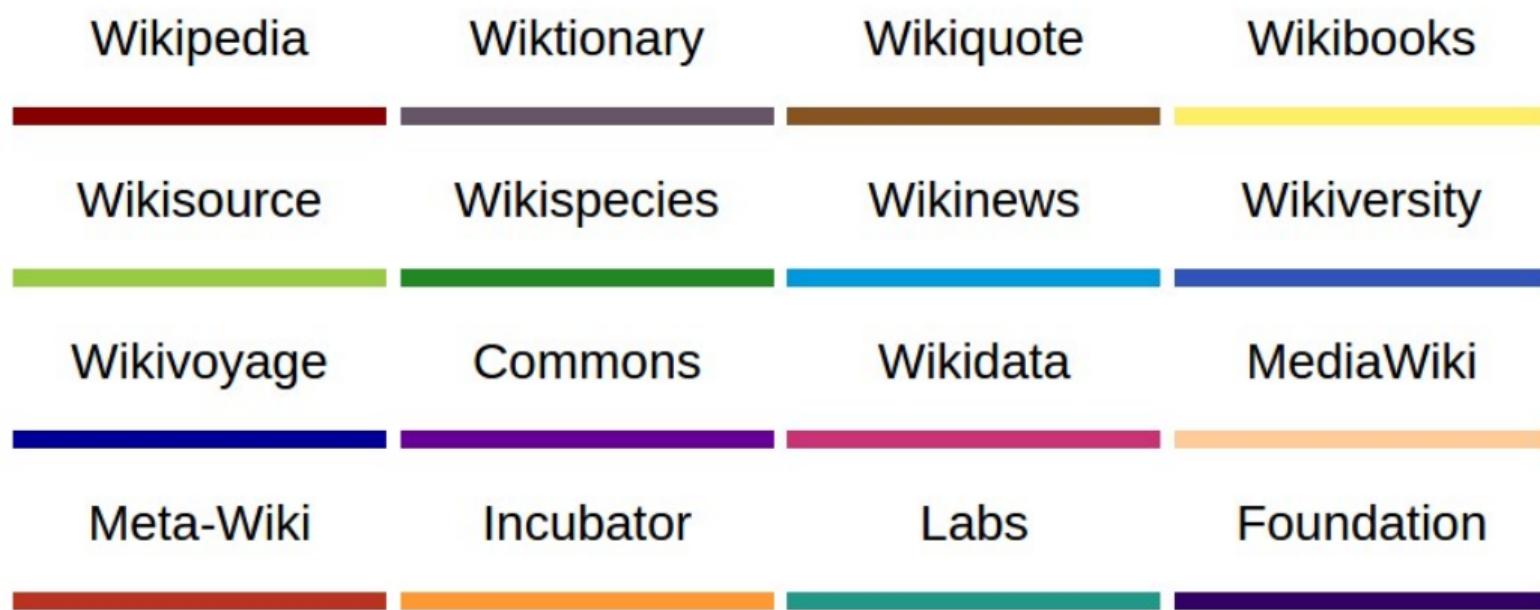
Will the combination of data sourced from global food composition tables and enriched with FoodOn identifiers reduce data disparities for people who record nutrient intake?

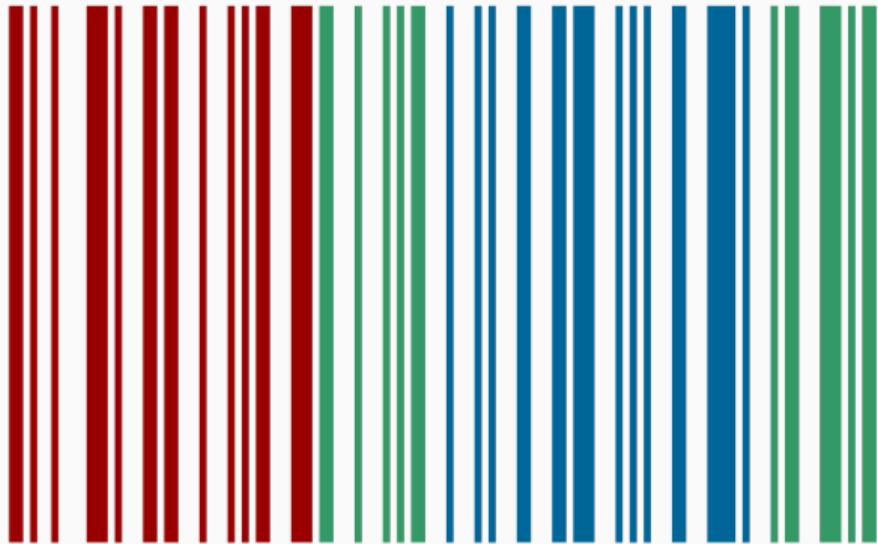
Wikipedias



WIKIPEDIA
The Free Encyclopedia

Wikimedia Foundation Projects





WIKIDATA

Wikibase

Wikibase is the software used by Wikidata that lets us store, manage, and access structured data using peer production.



Goal of the project

To create a free, open, dynamic, and collaborative global food composition database using a peer production model.

We have been creating our data models using the following examples.

[Visit WikiFCD!](#)

- USDA Food Data Central
- SMILING Vietnam
- SMILING Indonesia
- SMILING Laos
- SMILING Thailand
- FCT Malawi
- FCT Nigeria
- FCT Uganda

Food Composition Table Examples

Table 1. Proximate composition, content of cholesterol and phytosterols and the proportion of fatty acidated and polyunsaturated fatty acids of food samples from Ghana

Food table no.	Food	Lab code	Proximate composition, weight/100 g edible portion						
			Water g	Protein g	Fat g	Total carbohydrate g	Cholesterol mg	Phytosterols mg	
<i>Cereal products</i>									
10	akassa, koko	554	93.8	0.5	0.4	5.3	—	—	—
11	akpler	586	66.6	1.4	0.8	31.2	—	—	—
12	bafia	555	72.3	2.1	1.3	5.4	—	—	—
19	kenkey, Ga	556	64.2	3.6	1.7	30.5	—	—	—
20	kenkey, Fante	557	70.5	3.0	1.4	25.1	—	—	—
29	rice, raw	558	8.4	8.1	1.0	82.5	—	—	—
31	rice, cooked	559	64.7	2.9	0.4	32.0	—	—	—
35	wheat bread	560	24.3	11.7	3.7	60.5	—	—	—
37	yeast	573	35.4	4.4	4.4	54.4	—	—	—
350	biscuits	582	35.4	8.8	6.7	49.1	—	—	—
429	ricewater	609	90.6	1.0	0.3	8.1	—	—	—
434	wheat	610	62.4	4.6	1.8	31.2	—	—	—
<i>Sugary root and fruit</i>									
39	cassava, raw	561	57.2	0.7	0.5	41.6	—	—	—
40	cassava, boiled	587	59.6	0.7	0.5	39.2	—	—	—
44	cassava balls, fried	588	19.5	0.7	6.8	73.0	—	—	—
46	gari	562	7.7	0.9	1.1	50.6	—	—	—
47	gari	563	7.1	0.8	0.5	57.0	—	—	—
49	yake yake	589	53.2	0.4	0.5	45.9	—	—	—
53	cocoyam, boiled	564	59.2	3.0	0.6	37.2	—	—	—

FOOD_NAME_ENGLISH	FOOD_NAME_LOCAL	SCIENTIFIC_NAME	FOOD_GROUP	FOOD_SUB_GROUP	COUNTRY	UN
2	Rice noodle, fermented	Ricea ppol	Grains & grain products	Refined grains and products, unrefined grains and products, wholegrain and products, unrefined/unrefined	Lao	001
3	Wheat noodle (instant), instant	Thas sam lat heob	Grains & grain products	Unrefined grains and products, wholegrain and products, unrefined/unrefined	Lao	001
4	Straka rice (cooked), steamed	Khao nrao roun	Grase saline	Grains & grain products	Lao	001
5	Rice steamed, white	Khao chae foun	Grase saline	Grains & grain products	Lao	001
6	Rice noodle (rice starch)	Sen khao ppol	Grains & grain products	Unrefined grains and products, unrefined/unrefined	Lao	001
7	Rice, white, steamed	Khao leav	Grase glutinous	Grains & grain products	Lao	001
8	Rice, white, fried	Khao vun khao	Grase glutinous	Grains & grain products	Lao	001
9	Rice, white	Khao chev	Grase saline	Grains & grain products	Lao	001
10	Steamed Strikay rice (white), plain	Khao nrao roun ping	Grase saline	Refined grains and products, unrefined grains and products, enriched/unrefined	Lao	001
11	Baby food, cereals	Machok daek ta Cereali	Grase saline	Grains & grain products	Lao	001
12	Peanut boiled	Thong daem leav	Legumes, nuts & seeds	Cooked beans, lentils, peas	Lao	001
13	Coconuts, seeds, black, dried, boiled	Thoum dem krom	Vinegar unpasteurized	Legumes, nuts & seeds	Lao	001
14	Sesame seeds, white, roasted	Hauda npa khao	Sesame/milieu	Cooked beans, lentils, peas	Lao	001
15	Banana, flowers, fresh	Hauda pre	Musa bellidifolia	Nuts, seeds, and unpeeled	Lao	001
16	Cabbage, common, fresh	Pat ka leu po soth	Breseece breseece	Vegetables	Lao	001
17	Chinese cabbage blanched	Pat kaed khao loek	Breseece polystachys	Vegetables	Lao	001
18	Moringa pphy/leaves/cabbage, fresh	Pat leung soth	Breseece polystachys	Vegetables	Lao	001
19	Fennel common leaves, fresh	Pat kaed aer	Iurosia aquatica	Vegetables	Lao	001
20	Mustard green, stem and leaves	Pat kaed khao soth	Breseece juncea	Condiment vegetables	Lao	001

USDA U.S. DEPARTMENT OF AGRICULTURE
Agricultural Research Service

ARS HOME ABOUT ARS CONTACT ARS

HOME DATA TYPE DOCUMENTATION DOWNLOAD DATA API GUIDE HELP FAQ ABOUT US

CONTACT FOODDATA-CENTRAL



FoodData Central

Search FoodData Central:

Download Data Get an API Key API Guide

```

{
  "wid": 12345,
  "widFoodCategory": {
    "widFoodCategoryCode": "4602",
    "widFoodCategoryDescription": "Tomatoes"
  },
  "description": "Tomatoes, fresh, cooked",
  "foodAttributes": [
    {
      "id": 97839,
      "value": "from restaurant",
      "sequenceNumber": 2,
      "foodAttributeType": {
        "id": 1001,
        "name": "Additional Description",
        "description": "Additional descriptions for the food."
      }
    },
    {
      "id": 97838,
      "value": "with or without fat",
      "sequenceNumber": 1,
      "foodAttributeType": {
        "id": 1001,
        "name": "Additional Description",
        "description": "Additional descriptions for the food."
      }
    }
  ]
}

```

Example FCT

Food code	Food Name	No. of Regions	Lutein	Zeaxanthin	Lycopene	β -Cryptoxanthin	γ -Carotene	α -Carotene	β -Carotene	Total Carotenoids
			μg							
			LUTN	ZEA	LYCPN	CRYPXB	CARTG	CARTA	CARTB	CARTOID
E020	Fig (<i>Ficus carica</i>)	6	19.0±2.27	2.01±0.64					2.40±0.59	454±51.3
E021	Gosberry (<i>Emblica officinalis</i>)	5	38.7±9.78	2.86±1.61					1.58±0.09	62.01±6.45
E022	Grapes, seeded, round, black (<i>Vitis vinifera</i>)	4	27.8±8.85	3.59±2.20					29.36±8.37	213±29.4
E023	Grapes, seeded, round, green (<i>Vitis vinifera</i>)	5	31.3±8.77	4.06±1.60					30.77±11.10	208±28.8
E024	Grapes, seeded, round, red (<i>Vitis vinifera</i>)	5	17.2±3.04	7.58±2.28					19.94±1.29	314±99.9
E025	Grapes, seedless, oval, black (<i>Vitis vinifera</i>)	5	19.4±3.07	8.14±1.73					19.73±1.76	321±50.6
E026	Grapes, seedless, round, green (<i>Vitis vinifera</i>)	5	25.8±7.90	6.82±3.77					25.46±6.66	216±31.3
E027	Grapes, seedless, round, black (<i>Vitis vinifera</i>)	5	16.1±2.90	7.95±1.73					20.58±2.36	232±27.9
E028	Guava, white flesh (<i>Psidium guajava</i>)	5	19.7±4.59	5.92±3.06					298±32.8	996±108
E029	Guava, pink flesh (<i>Psidium guajava</i>)	5	90.0±11.5	8.94±2.21	2843±340				267±88.2	4078±477
E030	Jack fruit, ripe (<i>Artocarpus heterophyllus</i>)	5	19.3±6.91	2.42±1.27					23.53±1.24	59.61±4.90
E031	Jambu fruit, ripe (<i>Syzygium cumini</i>)	2	18.49	5.68					1.55	66.50
E032	Karonda fruit (<i>Carissa carandas</i>)	1	6.12	1.14					15.64	55.89
E033	Lemon, juice (<i>Citrus limon</i>)	6	9.68±2.22	1.72±0.50					2.62±0.34	85.99±3.36
E034	Lime, sweet,pulp (<i>Citrus limetta</i>)	6	13.2±3.54	1.89±0.44					2.54±0.30	85.13±10.71
E035	Litchi (<i>Nephelium litchi</i>)	4	27.33±9.81	1.61±0.62					1.47±0.21	129±15.1
E036	Mango, ripe, banganapalli (<i>Mangifera indica</i>)	6	3.19±1.51	1.62±0.22	2.72±1.31				1168±141	1424±308
E037	Mango, ripe, gulabkhas (<i>Mangifera indica</i>)	2	4.18	1.64	4.58				666	1350
E038	Mango, ripe, himsagar (<i>Mangifera indica</i>)	3	6.26±2.09	1.19±0.14	5.60±3.84				1181±136	1423±201
E039	Mango, ripe, kesar (<i>Mangifera indica</i>)	4	3.71±2.21	2.33±0.17	6.96±3.31				1264±172	1438±314
E040	Mango, ripe, neelam (<i>Mangifera indica</i>)	2	2.95	1.67	2.51				1291	1376

102

Matching Food Items to FoodOn Identifiers

- For each new FCT we add we need to map hundreds of food items to FoodOn identifiers

LexMapr

Sample_Id	Sample_Desc	Processed_Sample	Processed_Sample (With Scientific Name)	Matched_Components	Match_Status(Macro Level)
1	Acha, black, whole grain, Raw	acha black whole grain raw	acha black whole grain raw	['food (raw):FOODON_03311126', 'whole grain:FOODON_00003950']	Component Match
2	Acha, white, whole grain, Raw	acha white whole grain raw	acha white whole grain raw	['food (raw):FOODON_03311126', 'white:PATO_0000323', 'whole grain:FOODON_00003950']	Component Match
3	Acha, white, whole grain, Boiled	acha white whole grain boiled	acha white whole grain boiled	['food (boiled):FOODON_00002688', 'white:PATO_0000323', 'whole grain:FOODON_00003950']	Component Match
4	Maize, white, whole kernel, dried, raw	maize white whole kernel dried raw	maize {zea mays} white whole kernel dried raw	['food (dried):FOODON_03307539', 'food (raw):FOODON_03311126', 'maize kernel:FOODON_00003427', 'seed, skin present, germ present:FOODON_03420133', 'white:PATO_0000323']	Component Match

Multilingual Data in WikiFCD



Demo



Thank you!

katherine.thornton@yale.edu