

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

Wikipedia

Wiktionary

Wikiquote

Wikibooks

Wikisource

Wikispecies

Wikinews

Wikiversity

Wikivoyage

Commons

Wikidata

MediaWiki

Meta-Wiki

Incubator

Labs

Foundation



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

Example FCT

Table 4. Carotenoids

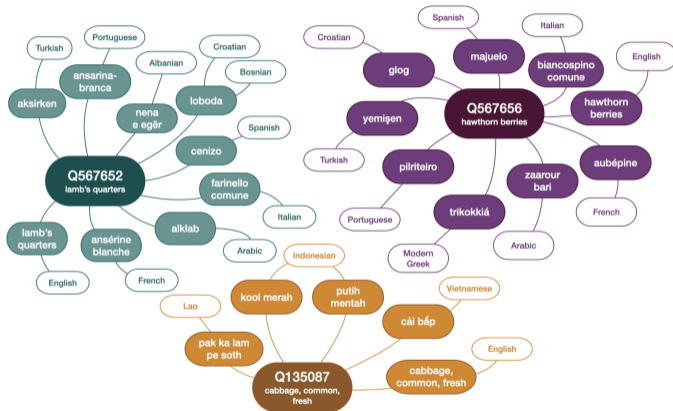
Food code	Food Name	No. of Regions	Lutein	Zeaxanthin	Lycopene	β -Cryptoxanthin	γ -Carotene	α -Carotene	β -Carotene	Total Carotenoids
			μg							
			LUTN	ZEAX	LYCPN	CRYPXB	CARTG	CARTA	CARTB	CARTOID
E020	Fig (<i>Ficus carica</i>)	6	19.0±2.27	2.01±0.64					2.40±0.59	454±51.3
E021	Gooseberry (<i>Embelica 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	90.3±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±1.96	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

Matching Food Items to FoodOn Identifiers

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

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