EuroFIR
European Food Information Resource

MAPPING DATA DOCUMENTATION BETWEEN FOOD COMPOSITION DATABASES

Jayne D. Ireland & Anders Møller
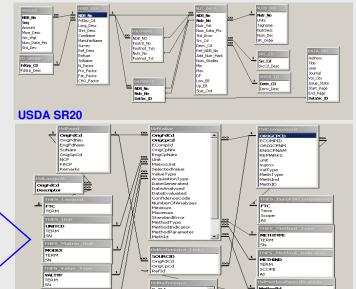*Danish Food Information (DFI), Roskilde, Denmark (ji@danfood.info)*

Objective

Nutrient databases include information referring to the description and identification of foods and their content of components. However, different databases may document their data in different ways, rendering comparisons difficult. The objective of this paper is to propose a mapping of value types and method types between the USDA nutrient database and the network of European databases.
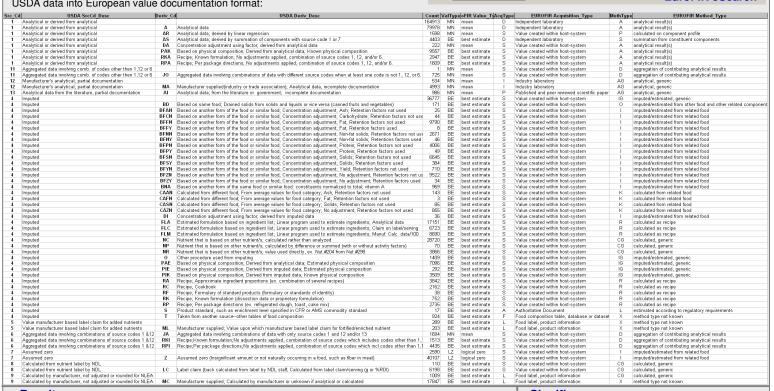
Methods and Materials

USDA's National Nutrient Database for Standard Reference is the most cited food composition database worldwide because of its quality and online accessibility (http://www.ars.usda.gov/ba/bhnrc/ndl). The USDA nutrient database characterizes nutrient values with *Source Code* indicating data type (e.g. analytical, calculated, assumed zero) and *Data Derivation Code* with more information on how the nutrient values were calculated or imputed (e.g. based on another form of the food or a similar food).

On the other hand, the **European Food Information Resource Network** (EuroFIR) has developed recommendations for compiling and disseminating food composition data that describes nutrient data with 3 slightly different concepts: *Value Type* (e.g. Mean, Best estimate, Logical zero), *Acquisition Type* (e.g. Industry laboratory, Peer reviewed scientific paper) and *Method Type* (e.g. Analytical result, Summation from constituent components, Imputed from related food). These thesauri are available on the EuroFIR Technical website (http://www.eurofir.org/eurofir/EuroFIRThesauri.asp).

A **correspondence table** was thus created between observed combinations of *Source Code* and *Data Derivation Code* in the USDA nutrient database, with appropriate *Value Type, Acquisition Type* and *Method Type* codes from the EuroFIR thesauri, to extract USDA data into European value documentation format:

USDA SR20

EuroFIR eSearch

Src_Cd	USDA SrcCd_Desc	Deriv_Cd	USDA Deriv_Desc	Count	ValType	EuroFIR Value_Ty	AcqType	EUROFIR Acquisition_Type	MethType	EUROFIR Method_Type
1	Analytical or derived from analytical			164913	MN	mean	D	Independent laboratory	A	analytical result(s)
1	Analytical or derived from analytical	A	Analytical data	79978	MN	mean	D	Independent laboratory	A	analytical result(s)
1	Analytical or derived from analytical	AR	Analytical data; derived by linear regression	1598	MN	mean	S	Value created within host-system	P	calculated on component profile
1	Analytical or derived from analytical	AS	Analytical data; derived by summation of components with source code 1 or 7	4403	BE	best estimate	D	Independent laboratory	S	summation from constituent components
1	Analytical or derived from analytical	DA	Concentration adjustment using factor; derived from analytical data	222	MN	mean	S	Value created within host-system	A	analytical result(s)
1	Analytical or derived from analytical	PAK	Based on physical composition; Derived from analytical data; Known physical composition	9557	BE	best estimate	S	Value created within host-system	A	analytical result(s)
1	Analytical or derived from analytical	RKA	Recipe; Known formulation; No adjustments applied, combination of source codes 1, 12, and/or 6.	2947	BE	best estimate	S	Value created within host-system	A	analytical result(s)
1	Analytical or derived from analytical	RPA	Recipe; Per package directions; No adjustments applied, combination of source codes 1, 12, and/or 6.	1609	BE	best estimate	S	Value created within host-system	A	analytical result(s)
11	Aggregated data involving comb. of codes other then 1,12 or 6			1	MN	mean	S	Value created within host-system	D	aggregation of contributing analytical results
11	Aggregated data involving comb. of codes other then 1,12 or 6	JO	Aggregated data involving combinations of data with different source codes when at least one code is not 1, 12, or 6.	725	MN	mean	S	Value created within host-system	D	aggregation of contributing analytical results
12	Manufacturer's analytical; partial documentation			534	MN	mean	I	Industry laboratory	AG	analytical, generic
12	Manufacturer's analytical; partial documentation	MA	Manufacturer supplied(industry or trade association); Analytical data, incomplete documentation	4993	MN	mean	I	Industry laboratory	AG	analytical, generic
13	Analytical data from the literature, partial documentation	AI	Analytical data; from the literature or government; incomplete documentation	986	MN	mean	P	Published and peer reviewed scientific paper	AG	analytical, generic
4	Imputed			36777	BE	best estimate	S	Value created within host-system	IG	imputed/estimated, generic
4	Imputed	BD	Based on same food; Drained solids from solids and liquids or vice versa (canned fruits and vegetables)	171	BE	best estimate	S	Value created within host-system	O	imputed/estimated from other food and other related component
4	Imputed	BFAN	Based on another form of the food or similar food; Concentration adjustment; Ash; Retention factors not used	25	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFCN	Based on another form of the food or similar food; Concentration adjustment; Carbohydrate; Retention factors not use	44	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFFN	Based on another form of the food or similar food; Concentration adjustment; Fat; Retention factors not used	9790	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFFY	Based on another form of the food or similar food; Concentration adjustment; Fat; Retention factors used	8	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFNN	Based on another form of the food or similar food; Concentration adjustment; Non-fat solids; Retention factors not use	2671	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFNY	Based on another form of the food or similar food; Concentration adjustment; Non-fat solids; Retentions factors used	46	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFPN	Based on another form of the food or similar food; Concentration adjustment; Protein; Retention factors not used	4006	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFPY	Based on another form of the food or similar food; Concentration adjustment; Protein; Retention factors used	49	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFSN	Based on another form of the food or similar food; Concentration adjustment; Solids; Retention factors not used	6645	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFSY	Based on another form of the food or similar food; Concentration adjustment; Solids; Retention factors used	384	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFYN	Based on another form of the food or similar food; Concentration adjustment; Yield; Retention factors not used	710	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFZN	Based on another form of the food or similar food; Concentration adjustment; No adjustment; Retention factors not us	9522	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BFZY	Based on another form of the food or similar food; Concentration adjustment; No adjustment; Retention factors used	34	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	BNA	Based on another form of the same food or similar food: constituents normalized to total; vitamin A	969	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	CAAN	Calculated from different food; From average values for food category; Ash; Retention factors not used	143	BE	best estimate	S	Value created within host-system	K	calculated from related food
4	Imputed	CAFN	Calculated from different food; From average values for food category; Fat; Retention factors not used	3	BE	best estimate	S	Value created within host-system	K	calculated from related food
4	Imputed	CASN	Calculated from different food; From average values for food category; Solids; Retention factors not used	85	BE	best estimate	S	Value created within host-system	K	calculated from related food
4	Imputed	CAZN	Calculated from different food; From average values for food category; No adjustment; Retention factors not used	855	BE	best estimate	S	Value created within host-system	K	calculated from related food
4	Imputed	DI	Concentration adjustment using factor; derived from imputed data	36	BE	best estimate	S	Value created within host-system	I	imputed/estimated from related food
4	Imputed	FLA	Estimated formulation based on ingredient list; Linear program used to estimate ingredients; Analytical data	17151	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	FLC	Estimated formulation based on ingredient list; Linear program used to estimate ingredients; Claim on label/serving	6723	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	FLM	Estimated formulation based on ingredient list; Linear program used to estimate ingredients; Manuf. Calc. data/100	8690	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	NC	Nutrient that is based on other nutrient/s, calculated rather than analyzed	28720	BE	best estimate	S	Value created within host-system	CG	calculated, generic
4	Imputed	NP	Nutrient that is based on other nutrient/s; calculated by difference or summed (with or without activity factors)	70	BE	best estimate	S	Value created within host-system	CG	calculated, generic
4	Imputed	NR	Nutrient that is based on other nutrient/s; value used directly, ex. Nut.#204 from Nut.#298	3866	BE	best estimate	S	Value created within host-system	CG	calculated, generic
4	Imputed	O	Other procedure used from imputing	1409	BE	best estimate	S	Value created within host-system	IG	imputed/estimated, generic
4	Imputed	PAE	Based on physical composition; Derived from analytical data; Estimated physical composition	7096	BE	best estimate	S	Value created within host-system	IG	imputed/estimated, generic
4	Imputed	PIE	Based on physical composition; Derived from imputed data; Estimated physical composition	292	BE	best estimate	S	Value created within host-system	IG	imputed/estimated, generic
4	Imputed	PIK	Based on physical composition; Derived from imputed data; Known physical composition	3509	BE	best estimate	S	Value created within host-system	IG	imputed/estimated, generic
4	Imputed	RA	Recipe; Approximate ingredient proportions (ex. combination of several recipes)	3842	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	RC	Recipe; Cookbook	2162	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	RF	Recipe; Formulary of standard products (formulary or standards of identity)	38	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	RK	Recipe; Known formulation (dissection data or proprietary formulation)	752	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	RP	Recipe; Per package directions (ex. refrigerated dough, toast, cake mix)	2735	BE	best estimate	S	Value created within host-system	R	calculated as recipe
4	Imputed	S	Product standard, such as enrichment level specified in CFR or AMS commodity standard	17	BE	best estimate	A	Authoritative Document	L	estimated according to regulatory requirements
4	Imputed	T	Taken from another source--other tables of food composition	924	BE	best estimate	F	Food composition table, database or dataset	X	method type not known
5	Value manufacturer based label claim for added nutrients			289	BE	best estimate	L	Food label, product information	X	method type not known
5	Value manufacturer based label claim for added nutrients	ML	Manufacturer supplied; Value upon which manufacturer based label claim for fortified/enriched nutrient	203	BE	best estimate	L	Food label, product information	X	method type not known
6	Aggregated data involving combinations of source codes 1 &12	JA	Aggregated data involving combinations of data with only source codes 1 and 12 and/or 13	1694	MN	mean	S	Value created within host-system	D	aggregation of contributing analytical results
6	Aggregated data involving combinations of source codes 1 &12	RKI	Recipe;Known formulation;No adjustments applied, combination of source codes which includes codes other than 1,	1513	BE	best estimate	S	Value created within host-system	D	aggregation of contributing analytical results
6	Aggregated data involving combinations of source codes 1 &12	RPI	Recipe;Per package directions;No adjustments applied, combination of source codes which incl codes other than 1,1	4435	BE	best estimate	S	Value created within host-system	D	aggregation of contributing analytical results
7	Assumed zero			2590	LZ	logical zero	S	Value created within host-system	I	imputed/estimated from related food
7	Assumed zero	Z	Assumed zero (Insignificant amount or not naturally occurring in a food, such as fiber in meat)	40107	LZ	logical zero	S	Value created within host-system	I	imputed/estimated from related food
8	Calculated from nutrient label by NDL			110	BE	best estimate	S	Value created within host-system	CG	calculated, generic
8	Calculated from nutrient label by NDL	LC	Label claim (back calculated from label by NDL staff; Calculated from label claim/serving (g or %RDI)	6198	BE	best estimate	S	Value created within host-system	CG	calculated, generic
9	Calculated by manufacturer, not adjusted or rounded for NLEA			1009	BE	best estimate	L	Food label, product information	CG	calculated, generic
9	Calculated by manufacturer, not adjusted or rounded for NLEA	MC	Manufacturer supplied; Calculated by manufacturer or unknown if analytical or calculated	17847	BE	best estimate	L	Food label, product information	X	method type not known

Results

Translation is not automatic, due to overlapping classifications and lack of scope notes for some codes, but allows data to be used in both systems. *ValueType, Acquisition Type* and *Method Type* codes were given to USDA values and then updated according to additional information given in the database.

- EuroFIR *Method Indicators* for energy were further updated according to the accompanying USDA document describing the files, fields and relationships (http://www.nal.usda.gov/fnic/foodcomp/Data/SR21/sr21_doc.pdf).

- Information on analytical methods given in the USDA document was recorded in the EuroFIR *Method Specification* table and linked to analytical values in the *Value* table.
- USDA *Footnotes* were added to Remarks fields in the EuroFIR Food and/or Value tables.
- USDA data sources (DATA_SRC table) were also completed with *Acquisition Types, Reference types* and Citation links (CiteXplore, http://www.ebi.ac.uk/citexplore/).

Significance

Standardized data documentation allows foods to be linked and compared across borders.

It is hoped that the correspondence table of value types between the USDA and European database systems will allow comparison of data with little loss of information and will also open international discussions for clarification and standardization of terms.