



dataFIELD v. 1.0 Documentation

dataFIELD (database of Food Impacts on the Environment for Linking to Diets) aggregates data on the greenhouse gas emissions (GHGE) and cumulative energy demand (CED) associated with production of specific foods and assigns those impact factors to food commodities to facilitate linkages with self-selected individual diets in the US National Health and Nutrition Examination Survey (NHANES). Data were gathered from the life cycle assessment literature as described in Heller et al, 2018 (<https://doi.org/10.1088/1748-9326/aab0ac>), compiled by food type, and assigned to the food commodities in the Food Commodity Intake Database (FCID) (<http://fcid.foodrisk.org/>). Proxy assignments were made when direct food type matches were unavailable, and adjustments were required to account for differences in mass basis (for example, excluding inedible portions).

Additional datasets and documentation for linking dataFIELD to NHANES will be available at: <https://sph.tulane.edu/gchb/diet-environmental-impacts>

This document describes the dataFIELD workbook in detail.

The dataFIELD workbook contains seven sheets:

1. Description – a high level description of the workbook
2. FCID Linkages – (detailed below) contains final (utilized) impact factors
3. Literature entries – (detailed below) raw data entries logged in literature review, including metadata, references
4. Pivot Table_GHG – GHGE data from "literature entries" sheet aggregated by specific food type and boundary condition (farm gate vs. processor gate)
5. Pivot Table_CED – CED data from "literature entries" sheet aggregated by specific food type and boundary condition (farm gate vs. processor gate)
6. Proxy def'n's – contains proxy definitions used to fill FCID commodities without direct matches.
7. MassConv – compilation of mass conversions used to account for inedible portions, etc.

FCID Linkages sheet

This sheet is the heart of dataFIELD and is where all the linkages, assignments and calculations occur. Descriptions of each column and how the spreadsheet functions are provided below.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1						UTILIZED IMPACT FACTORS				farm gate (=1)	PROXY DEFINITION		ENERGY USE MJ/kg		
2					number of	CED	GHG					PROXY?			
3	FCID_Code	FCID_Desc	Weight_Basis_as defined by FCID	Comments (from FCID)	N_Cases	MJ / kg	CO ₂ eq / kg	weight Conversion Factor	conversion factor source	or processor gate (=2)			# of datapoint	N-average	Std. Dev
4	101050000	Beet, garden, roots	weight of roots; juice		659	1.221	0.194	1.11	AgHB	1		DB value	1	1.100	0.000
5	101052000	Beet, sugar	dry weight of sugar (Sucrose is a disaccharide)		49979	4.574	0.391	1	FICRCD	2		DB value	4	4.574	0.637
6	101053000	Beet, sugar, molasses	weight of molasses		7656	3.476	0.288	1	FICRCD	2		DB value	4	3.476	0.378
7	101067000	Burdock	weight of roots		2	1.891	0.284	1.33	NNDSR28	1	roots	PROXY	14	1.422	0.574

- Column A:
ID code assigned by FCID.
- Column B:
FCID commodity description. Note that the specific text in this column is used for lookup functions elsewhere in the sheet
- Column C:
Weight basis as defined by FCID. Impact factors are adjusted to align with this weight basis definition in order to allow compatibility with FCID.
- Column D (hidden):
Additional comments from FCID documentation on inclusions within FCID codes.
- Column E (hidden):
Offers an indication of how prevalent a given commodity is within the FCID recipe files. Higher numbers indicates more occurrences.
- Columns F&G:
Contain the “final” impact factors utilized in our analyses. Note that these values are on a weight basis as defined by the FCID recipe files (Column C). If linking to other databases, adjustments may be necessary to properly match weight bases (see below)
- Column H:
Contains the “weight conversion factor” used to adjust from the weight basis as defined in LCA studies and the weight basis utilized by FCID. This conversion primarily pertains to inedible portions of fruits and vegetables (peels, rinds, cores, etc). For example, the weight basis for “orange” as defined by FCID is “weight of pulp; excluding seeds and peel” whereas LCA results are on the basis of the whole orange (including peel). A conversion factor is used to account for these differences. Weight conversion factors are also used to proxy the impacts of juices and dried fruits by adjusting the LCA data from the base fruit/vegetable by the removal of the weight of pulp or water.
- Column I:

Indicates the source of the weight conversion factor; sources are aggregated in the sheet “MassConv”.

- Column J:

Indicates the boundary condition applied for a given commodity. In most cases, boundary conditions account for ‘cradle-to-farm gate’ impacts, i.e., only the GHGE and CED associated with farm-level production (and upstream resources) are included. For commodities requiring processing (flours, refined sugars, vegetable oils, etc), the boundary condition has been chosen as ‘cradle-to-processor gate’.

- Column K:

When a direct match is not available for a given FCID commodity and a proxy is used to estimate impacts, Column K defines the proxy used. These proxy definitions refer to entries or groups of entries defined in the sheet ‘Proxy def’ns’. Often, averages across food groupings (e.g. berries, brassicas, brassica greens, citrus, fresh herbs, grains, other greens, nuts, roots, dried spices, other tree fruit, tropical fruit) are used as a proxy for a specific food item in that same grouping that was lacking data. Failing this approach, other proxies of foods with similar form were then assigned. These assignments were based on similarities of specific crops in their botany and, most importantly, production methods, as determined by the expertise of our research team. Note that in some cases, use of the “proxy” mechanism is largely a means of automating data lookup processes for similar items. For example, separate FCID codes exist for “Apple, fruit with peel” and “Apple, peeled fruit”; these use the same underlying data (albeit with slightly different weight conversion factors) but a “proxy” is used for the latter to allow for automated lookup of inexact commodity names.

- Columns L-O: Data lookup for cumulative energy use.

- Column L: determines whether a direct match to the FCID commodity name exists in the database; if yes, “DB value” is returned, if no, “PROXY” is returned.
- If “DB value”, columns M-O then look up statistics from the “Pivot Table_CED” sheet (number of datapoints included in average, average CED value, and standard deviation).
- If column L returns “PROXY”, columns M-O construct values for number of datapoints, average value, and standard deviation, based on the proxy defined in Column K. These constructions are via DATABASE functions (DCOUNT, DAVERAGE, DSTDEV) and directly access the “Literature entries” sheet.
- Note that all formulas respond to the boundary condition choice in Column J.

P	Q	R	S	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG
GHG Emissions (kg CO2eq/kg)				VARIANCE CALCULATIONS											
PROXY?				Energy Use (MJ/kg)					GHGE						
	# of datapoint	N-average	Std. Dev	basis	n	proxy group mean	SD (group or indiv food)	scaled SD (group SD*food)	basis	n	proxy group mean	SD (group or indiv food)	scaled SD (group SD*food)		
DB value	2	0.175	0.092	roots	14	1.4220714	0.574420439	0.444325419	roots	25	0.2134	0.179920353	0.147544807		
DB value	4	0.391	0.076	sweeteners	14	8.8453456	17.46832841	9.033479887	sweeteners	13	1.0611859	1.805599706	0.664914384		
DB value	4	0.288	0.047	sweeteners	14	8.8453456	17.46832841	6.864320758	sweeteners	13	1.0611859	1.805599706	0.489336436		
PROXY	25	0.213	0.180	roots	14	1.4220714	0.574420439	0.574420439	roots	25	0.2134	0.179920353	0.179920353		

- Columns P-S:

Data lookup for GHGE. Lookup structure is parallel to that described for cumulative energy demand (columns L-O).

- Column T&U (hidden)

Lack of literature data for life cycle CED and GHGE for certain processed foods contained in the FCID listing required a combination of proxy data and other sources to develop valid estimates. Estimates of the processing energy (and subsequently, GHGE) required to convert fruits and vegetables into juiced form, as well as vinegar and maple syrup production, are included in these columns. These components are added to the final values in columns F&G after mass conversion factors are applied to the base fruit or vegetable. See the *ERL* journal article [Supporting Information](#) section, “Additional food processing estimates” for more details.

- Column V (intentionally blank)
- Columns W-AF:

Compile variability estimates for each impact factor and commodity. The method here assures that at least 5 datapoints are used in calculating a standard deviation. If the original assigned impact factor was based on fewer than 5 datapoints, a broader proxy group is assigned and the standard deviation from that group is used as the variance basis. See the *ERL* journal article [Supporting Information](#) for details on variance calculations.

Literature Entries sheet

This sheet contains the raw data entries gathered from literature studies. It can be used to track references, investigate differences within a given food type on the basis of production methods or country, identify studies that consider life cycle stages beyond farm gate, and explore distribution of impacts across life cycle stages. Brief descriptions of the sheet contents follow.

	A	C	D	E	F	G	H	I	J	L	M	AD	AE	C
		FOOD TYPE	SUB TYPE	FOOD FORM	PRODUCT OR PRODUCTION SPECIFICS	COMPARES PRODUCTION STRATEGIES / METHODS?	COUNTRY OF ORIGIN	Publication Year	Publication Authors	Publication Source Type	Bibliographic Reference	Reported Functional Unit	Adjustment of FU for KG Edible Food	GA
1														
2	1	Dairy	Fluid Milk, Cow	Fresh or Not Specified	Liquid Milk	No	USA	2013	Greg Thomas	Peer Reviewed	Thoma, G., Pop	1 kg of milk con	No	
3	2	Meat	Sheep	Fresh or Not Specified	lowland farms	Yes	England & W	2014	A.K. Jones, D.L.	Peer Reviewed	Jones, A. K., Jon	1 kg LW	dressing yield=50.8%	
4	3	Meat	Sheep	Fresh or Not Specified	highland farms	Yes	England & W	2014	A.K. Jones, D.L.	Peer Reviewed	Jones, A. K., Jon	1 kg LW	dressing yield=50.8%	
5	4	Meat	Sheep	Fresh or Not Specified	hill farms	Yes	England & W	2014	A.K. Jones; D.L.	Peer Reviewed	Jones, A. K., Jon	1 kg LW	dressing yield=50.8%	

- Column A: simple indexing number
- Column C: Food group
- Column D: specific food commodity. Text should correspond with FCID commodity descriptions in ‘FCID Linkages’ sheet
- Column E: information on food form: fresh, canned, frozen, etc
- Column F: provides other relevant information on the product or production method
- Column G: indicates whether production methods are compared within the study
- Column H: country of production modeled in study
- Columns I-M: bibliographic information for the data source.
- Columns N-AC (hidden): these columns contain yes/no responses as to whether specific life cycle stages are included/considered in the source study, along with additional scope/boundary notes.
- Column AD indicates the functional unit reported in the study
- Column AE indicates adjustments applied to normalize to a standard “per kg” functional unit (per kg boneless, edible in the case of animal flesh products).
- Columns AF-AI: results accessed in other parts of dataFIELD are aggregated here.
- Columns AJ-AV: specific CED data, including contributions from specific life cycle stages, if specified in study, farm gate and total LC values, method indications and additional notes
- Columns AX-BJ: specific GHGE data, including contributions from specific life cycle stages, if specified in study, farm gate and total LC values, method indications and additional notes