FAO ESS trade processing system

Christian A. Mongeau Ospina
Food and Agriculture Organization of the United Nations

16 June 2020

Abstract

This document describes the steps carried out by the Economic and Social Development Department for the processing of agricultural trade.

Contents

1 Introduction

The trade processing system of the Economic and Social Development Department (ESS) of the Food and Agriculture Organization of the United Nations (FAO) involves two main steps: raw data treatment and data validation.

Raw data treatment includes all operations required to save trade data in FAO-specific datasets with predefined structures, from bilateral raw data. These operations can be roughly summarised as: data harvesting, data content assessment, countries and items codes conversions/aggregation, standardisation of units of measurement, calculation of unit values, imputation of reporter's missing or incomplete quantities, imputation of data for missing countries through "mirroring". At the end of these operations, bilateral and total trade data is saved to FAO's Statistical Working System (SWS)[1] datasets.

Two R[2] routines (called also plugins)[3] deal with this first phase: "Complete Trade Flow CPC" and "Total Trade CPC". The former deals with raw bilateral data, applies the operations mentioned before, and saves data to the `Bilateral Trade Flow (CPC)` SWS dataset; the latter takes as input the data obtained as output from the "Complete Trade Flow CPC" plugin, aggregate all bilateral partner data for each combination of reporter/flow/item, calculates the aggregated unit values, and saves data to the `Bilateral Trade Flow (CPC)` SWS dataset.

[1] The appendix "Statistical Working System" gives background information on what the SWS is and how trade-related information is organised therein. For now, it is sufficient to say that the "Statistical Working System is a corporate platform used in FAO for the processing and storage of statistical datasets, providing the framework needed to use the same methods, standards, classifications and approaches within statistical processes." http://www.fao.org/economic/ess/ess-home/ess-about/statistical-methods/en/

[2] "R is a language and environment for statistical computing and graphics." https://www.r-project.org/

[3] The computational part of the SWS is composed of a series of "plugins", which are R scripts that are run on the R server available in SWS. These plugins take as input data available as SWS datasets/datatables, carry out statistical procedures based on these data and parameters, then save processed data back to SWS datasets/datatables (though some plugins' scope is not saving data, but computing some analysis that is sent back to the user by email). More information of hwo the SWS works is given in the appendix "Statistical Working System".

When data have been processed and saved to the datasets, data validation by ESS analysts takes place. This is done by means of complimentary tools that detect outliers or other type of data that deserves verification, and allows analysts to correct flows in an informed way by means of an interactive validation tool.

The trade module was developed by (in alphabetical order): Marco Garieri, Alexander Matrunich, Christian Mongeau, and Bo Werth. Plugins for data validation have been developed by (in alphabetical order): Carlo Del Bello, Alberto Munisso, Aydan Selek, Sumeda Siriwardena, Cristina Valdivia. The interactive tool for data validation has been developed by Christian Mongeau. Useful feedback has been given by (in alhpabetical order): Katherine Baldwin, Claudia De Vita, Carola Fabi, Dominique Habimana, Salar Tayyib.

Any remaining error/inaccuracy in this document (as of June 2020) can be attributed exclusively to the author. A correction and pull request can be made in the module git repo at:

https://github.com/SWS-Methodology/faoswsTrade/

2 Raw data processing

The sources of bilateral raw data are the United Nations Statistics Division (UNSD) for non-European countries, and Eurostat for European countries. In order for these two data sources to be combined and saved into a single SWS dataset they need to pass trough different processing steps, which will be explained in this section.

2.1 Data harvesting

TODO: HERE A DESCRIPTION OF THE WORKFLOW FOLLOWED BY FRANO WILL BE PRESENT, WITH A DESCRIPTION OF THE SHINY TOOL FRANO SAID IS AVAILABLE FOR DOING THE HARVESTING.

Data of relevance for ESS trade falls under the following "Harmonized System" (HS)[4] chapters[5]:

[4] "The Harmonized Commodity Description and Coding System generally referred to as"Harmonized System" or simply "HS" is a multipurpose international product nomenclature developed by the World Customs Organization (WCO). It comprises about 5,000 commodity groups; each identified by a six digit code, arranged in a legal and logical structure and is supported by well-defined rules to achieve uniform classification. The system is used by more than 200 countries and economies as a basis for their Customs tariffs and for the collection of international trade statistics. Over 98 % of the merchandise in international trade is classified in terms of the HS." http://www.wcoomd.org/en/topics/nomenclature/overview/what-is-the-harmonized-system.aspx

[5] "The HS comprises approximately 5,300 article/product descriptions that appear as headings and subheadings, arranged in 99 chapters, grouped in 21 sections. The six digits can be broken down into three parts. The first two digits (HS-2) identify the chapter the goods are classified in, e.g. 09 = Coffee, Tea, Maté and Spices. The next two digits (HS-4) identify groupings within that chapter, e.g. 09.02 = Tea, whether or not flavoured. The next two digits (HS-6) are even more specific, e.g. 09.02.10 Green tea (not fermented)... Up to the HS-6 digit level, all countries classify products in the same way (a few exceptions exist where some countries apply old versions of the HS)." https://unstats.un.org/unsd/tradekb/Knowledgebase/50018/Harmonized-Commodity-Description-and-Coding-Systems-HS

chapter	description
01	Animals; live
02	Meat and edible meat offal
03	Fish and crustaceans, molluscs and other aquatic invertebrates
04	Dairy produce; birds' eggs; natural honey; edible products of animal origin, not elsewhere specified or included
05	Animal originated products; not elsewhere specified or included
06	Trees and other plants, live; bulbs, roots and the like; cut flowers and ornamental foliage
07	Vegetables and certain roots and tubers; edible
08	Fruit and nuts, edible; peel of citrus fruit or melons
09	Coffee, tea, mate and spices
10	Cereals
11	Products of the milling industry; malt, starches, inulin, wheat gluten
12	Oil seeds and oleaginous fruits; miscellaneous grains, seeds and fruit, industrial or medicinal plants; straw and fo
13	Lac; gums, resins and other vegetable saps and extracts
14	Vegetable plaiting materials; vegetable products not elsewhere specified or included
15	Animal or vegetable fats and oils and their cleavage products; prepared animal fats; animal or vegetable waxes
16	Meat, fish or crustaceans, molluscs or other aquatic invertebrates; preparations thereof
17	Sugars and sugar confectionery
18	Cocoa and cocoa preparations
19	Preparations of cereals, flour, starch or milk; pastrycooks' products
20	Preparations of vegetables, fruit, nuts or other parts of plants
21	Miscellaneous edible preparations
22	Beverages, spirits and vinegar
23	Food industries, residues and wastes thereof; prepared animal fodder
24	Tobacco and manufactured tobacco substitutes
33	Essential oils and resinoids; perfumery, cosmetic or toilet preparations
35	Albuminoidal substances; modified starches; glues; enzymes
38	Chemical products n.e.s.
40	Rubber and articles thereof
41	Raw hides and skins (other than furskins) and leather
43	Furskins and artificial fur; manufactures thereof
50	Silk
51	Wool, fine or coarse animal hair; horsehair yarn and woven fabric
52	Cotton
53	Vegetable textile fibres; paper yarn and woven fabrics of paper yarn

In Table 1, an example of filtered UNSD data is shown (variable names have been modified). It contains information on the reporter, partner, flow (1 = imports; 2 = exports; 3 = re-exports; 4 = re-imports), HS codes (variable length), monetary value (in USD dollars), weight (in kilograms), supplementary quantity, the unit in which the supplementary quantity is given (`qunit`, e.g., 8 = kilograms, 7 = liters), and the chapter.

In Table 2, an example on filtered Eurostat data is reported (also in this case, the original variable names were modified). The main differences are: reporter and partner codes are "geonomenclature" codes (in UNSD they are "M49"); the HS length is maximum 8-digits; monetary values are in thousands of euros; weight is reported in tonnes; the supplementary quantity is always commodity specific (in Tariff line data it can happen that the same HS code is reported in different units).

Besides filtering by chapters, only some HS codes inside the chapters are considered. These codes are given in the `hs6faointerest` datatable, of which Table 3 shows a subset of ten codes.

For both UNSD and Eurostat data, codes that contain non-numeric characters are dropped.[6] An example of this can be seen in the Eurostat table showed before (Table 2), which contains the `10MMM000` code. The only information of the transaction reported with that code is that it corresponds to some cereal (chapter 10 is

[6]Alphanumeric codes (i.e., codes that have non-numeric characters) happen mainly because a country does not want to disclose the full information of a given transaction (or set of transactions).

Table 1: Subset of Tariff line data

year	reporter	partner	flow	hs	value	weight	qty	qunit	chapter
2014	12	699	1	38089119	109821.79	7160.22	7160.22	5	38
2014	600	380	2	08140000000	24456.00	15870.00	15870.00	8	08
2014	398	276	1	400259	532.12	0.20	0.20	8	40
2014	703	616	1	05119910	39.83	6.00	6.00	8	05
2014	251	76	1	20081912	1933.60	148.00	148.00	8	20
2014	702	156	1	382311	117527.41	60000.00	60000.00	8	38
2014	203	682	1	100620	143.00	20.00	NA	1	10
2014	344	840	3	16055900	17245.07	907.00	907.00	8	16
2014	48	682	1	19059093	753403.97	368416.00	368416.00	8	19
2014	616	428	1	401693	362.00	20.00	20.00	8	40
2014	384	466	2	19021900	901389.92	1238270.00	1238270.00	8	19
2014	764	152	2	400942	108.10	NA	NA	1	40
2014	414	380	1	38089990	574.72	2.00	2.00	8	38
2014	158	324	2	19059090006	462.59	103.00	103.00	8	19
2014	690	784	1	18069090	1210.97	28.80	15.00	5	18

Table 2: Subset of Eurostat data

year	reporter	partner	flow	hs	value	weight	qty
201452	003	0346	2	18069060	6.88	2.2	NA
201452	008	0001	2	40119200	26.65	14.8	1225
201452	068	0005	2	10MMM000	42.17	0.0	NA
201452	005	0006	2	52053400	0.11	0.0	NA
201452	006	0647	2	52085100	1361.25	17.7	180134
201452	017	0004	2	09096100	7.45	1.2	NA
201452	001	0011	2	02042300	4.76	0.6	NA
201452	060	0032	2	19049080	17.55	3.1	NA
201452	008	0032	2	02031955	1947.23	438.4	NA
201452	017	0212	2	33051000	282.50	11.4	NA
201452	001	0690	2	03061410	9.05	0.3	NA
201452	008	0001	1	35030080	1002.17	7.4	NA
201452	018	0006	1	52094300	0.33	0.0	93
201452	017	0604	2	22011019	0.09	0.1	70
201452	006	0091	2	38089110	1.71	1.6	NA

Table 3: HS-6 codes of interest

hs6_code	
180520	
020576	
090838	
110243	
150379	
530211	
020892	
210546	
010525	
510425	

"Cereals"). Besides that, it is difficult, or even impossible, to assign it to a specific cereal type (e.g., wheat or barley, which are recorded as `100110` and `100300`, respectively). Thus, given the difficulty in assigning codes with non-numeric characters to detailed items, and given that these cases are relatively marginal, the codes are removed from the raw data and no attempt at assigning them is done.

So far, the discussion has evolved by mentioning only the HS system as the standard coding for items, but the raw data obtained by UNSD/Eurostat does not come (usually) in standard HS codes. Indeed, UNSD data comes as Tariffline, i.e., HS codes that may (and usually do) extend to more than 6 digits, usually 8, 10, 12. These codes are not directly comparable across countries, as each country may have a description for such codes that is relatively different from that of other countries. As for Eurostat data, it comes with "Combined Nomenclature" (CN) codes at 8 digits (CN8), which are EU extensions of HS codes[7].

For Eurostat data, only `stat_regime` equal to 4 is kept[8]:

> In this system *["Statistical regime 4" or "Total trade"]*, the recorded aggregates include all goods entering or leaving the economic territory of a country with the exception of simple transit trade. In particular, all goods received into customs warehouses are recorded as imports, regardless of whether they subsequently go into free circulation in the Member State of receipt. Similarly, outgoing goods from customs warehouses are included in the general trade aggregates, at the time they leave the Member State.

2.2 Raw data: SWS datatables vs. RDS files

As seen above, raw UNSD and Eurostat data are saved into SWS datatables. All plugins, including then the "Complete Trade Flow CPC" and "Total Trade CPC" plugins, read raw data stored directly in the SWS database. This, however, for large amount of data as the trade raw data (which generally contain more than 15 million rows) turns out to be a time-expensive process. For instance, the process of getting the raw UNSD and Eurostat data takes by itself around 40 minutes. In order to overcome this issue, the "Complete Trade Flow CPC" is designed to use RDS files[9] stored in the SWS "R shared drive"[10]. Reading these files takes less than 2 minutes, so the benefits in terms of saving reading (and, thus, overall time employed by the plugin to complete) time is evident.

A more detailed information on how to access the raw data files stored in the R shared drive, and how to updated them, will be given in "Accessing/updating files in the R shared drive" Appendix.

2.3 "Pre-Processing Report" plugin

Data content assessment is carried out by means of different tables (pre-processing tables) compiled from raw data, which give a first idea of the completeness and quality of the data. For instance, if in a given year transactions of a country are 70% less than those reported the previous year, it may indicate that the country file is incomplete[11] so that it can be decided to treat the country as a non-reporter (so that mirroring will be used for imputing all its transactions in that year) until its data is more complete.[12]

[7] https://trade.ec.europa.eu/tradehelp/eu-product-classification-system

[8] See pag. 9 in *DG Trade Statistical Guide*, June 2016, http://trade.ec.europa.eu/doclib/docs/2013/may/tradoc_151348.pdf

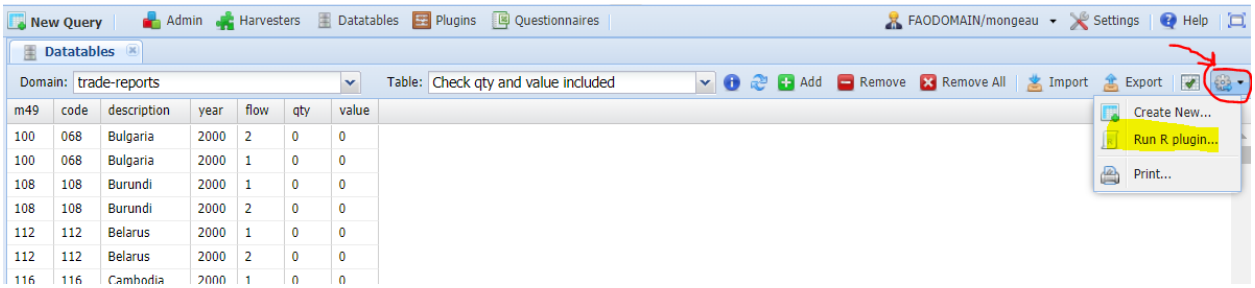
[9] RDS is a file format used to read/write serialized data in R.

[10] The "R shared drive" is a folder in the SWS server where R plugins can write and read data to/from, and can be accessed also by local PCs in the FAO network by means of a Samba service.

[11] Given the country reports at the same number of HS digits. Reports also contain this information.

[12] This is the final decision if no other option in available. What happens when analysis on the raw data content is carried out and some country data seems incomplete is that the data providers are contacted by an FAO officer who will require additional details or other useful information that can help understanding what is the source for such incompleteness. For instance, while it can well be that a country has reported a number of transaction in line with previous years, it may happen that for a large amount of them do not contain any information on quantity. In this case the provider is contacted so to understand whether the country actually reported the data like that, or there was some other issue that made the quantity information unavailable.

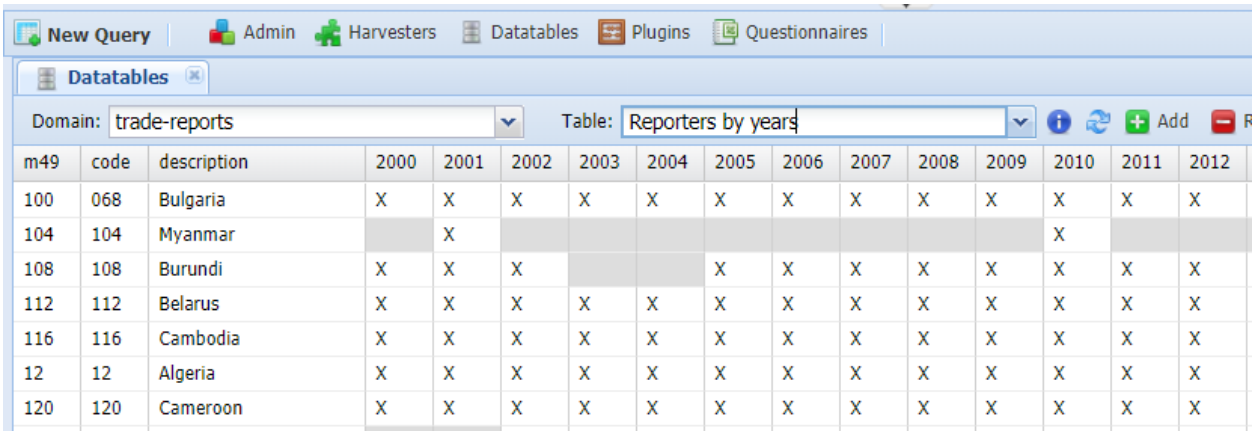
In order to run this plugin, one of the datatables indicated below (found in the "trade-reports" domain) need to be selected, then the user will need to click on the button shown in the Figure below, select "Run plugin...", and select "Pre-Processing Report" in the popup window that will open.



All tables report the "m49" column, indicating the M49 code assigned to the country as it wilkl be saved in SWS, and a "code" column containing the original code assigned by the country either by UNSD (M49-comtrade) or Eurostat (geonom).

2.3.1 "Reporters by years"

This table reporter whether the country is available in the raw data file as a reporter, and it is indicated with an "X" in the specific country/year combination.

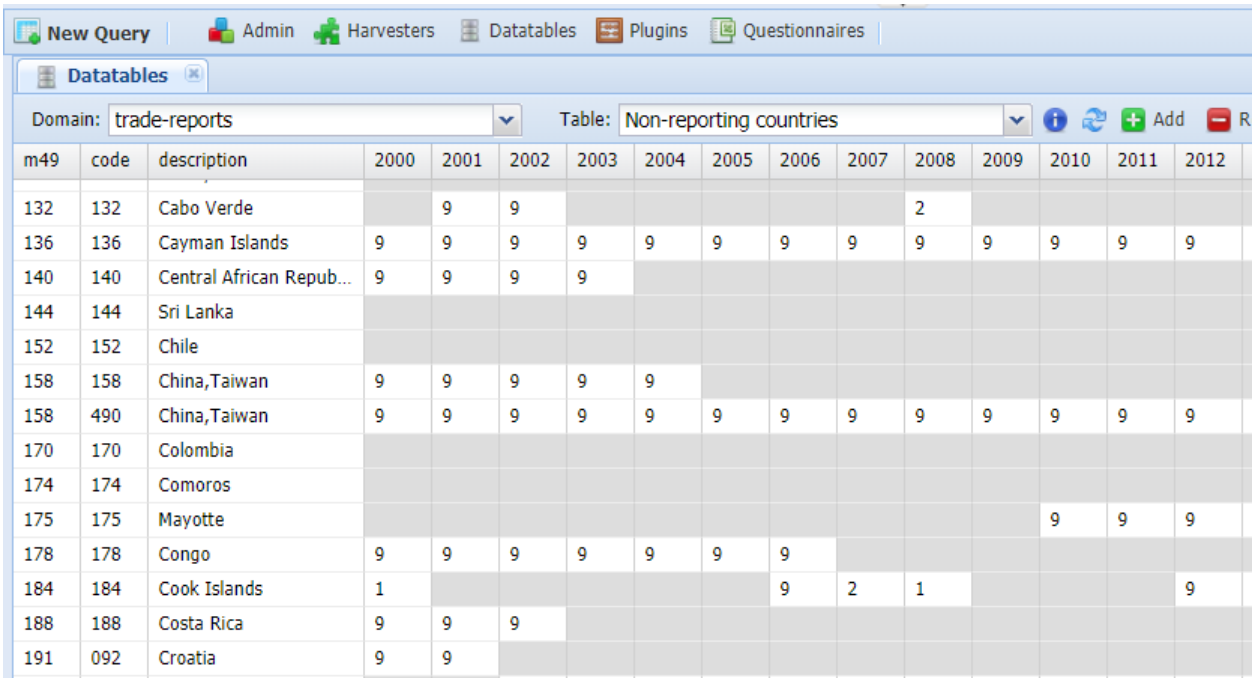


m49	code	description	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
100	068	Bulgaria	X	X	X	X	X	X	X	X	X	X	X	X	X
104	104	Myanmar		X									X		
108	108	Burundi	X	X	X			X	X	X	X	X	X	X	X
112	112	Belarus	X	X	X	X	X	X	X	X	X	X	X	X	X
116	116	Cambodia	X	X	X	X	X	X	X	X	X	X	X	X	X
12	12	Algeria	X	X	X	X	X	X	X	X	X	X	X	X	X
120	120	Cameroon	X	X	X	X	X	X	X	X	X	X	X	X	X

In the Figure above, for instance, Myanmar is among reporter countries in 2001 and 2010, while in the other years it is not a reporter.

2.3.2 "Non-reporting countries"

A country can be non-reporter because it did not report any flow, or because it did not report imports or exports. Scope of this table is showing this information. Countries that did not report imports and exports are marked with a "9", while those that did not report imports or exports are marked with a "1" or "2", respectively.

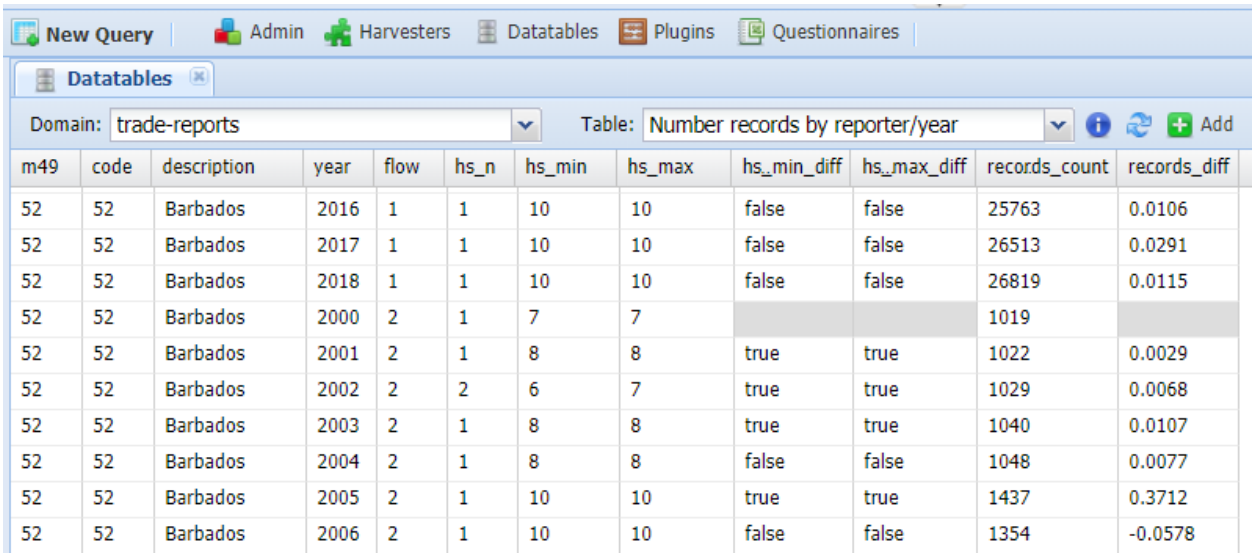


m49	code	description	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012
132	132	Cabo Verde		9	9						2				
136	136	Cayman Islands	9	9	9	9	9	9	9	9	9	9	9	9	9
140	140	Central African Repub...	9	9	9	9									
144	144	Sri Lanka													
152	152	Chile													
158	158	China,Taiwan	9	9	9	9	9								
158	490	China,Taiwan	9	9	9	9	9	9	9	9	9	9	9	9	9
170	170	Colombia													
174	174	Comoros													
175	175	Mayotte											9	9	9
178	178	Congo	9	9	9	9	9	9	9						
184	184	Cook Islands	1						9	2	1				9
188	188	Costa Rica	9	9	9										
191	092	Croatia	9	9											

In the example shown in the Figure, Cabo Verde was a complete non-reporter in 2001 and 2002, while it reported only imports in 2008 (a "2" indicated that exports are missing).

2.3.3 "Number records by reporter/year"

This table contains by country/year/flow, the minimum and maximum number of HS digits reported by the country (`hs_min` and `hs_max`; the number of different HS digits is shown by the `hs_n` variable), whether there has been a difference in this two variables with respect to the previous year (`hs_min_diff` and `hs_max_diff`), the number of records (`records_count`) and the percentage variation of record counts with respect to the previous year (`records_diff`).

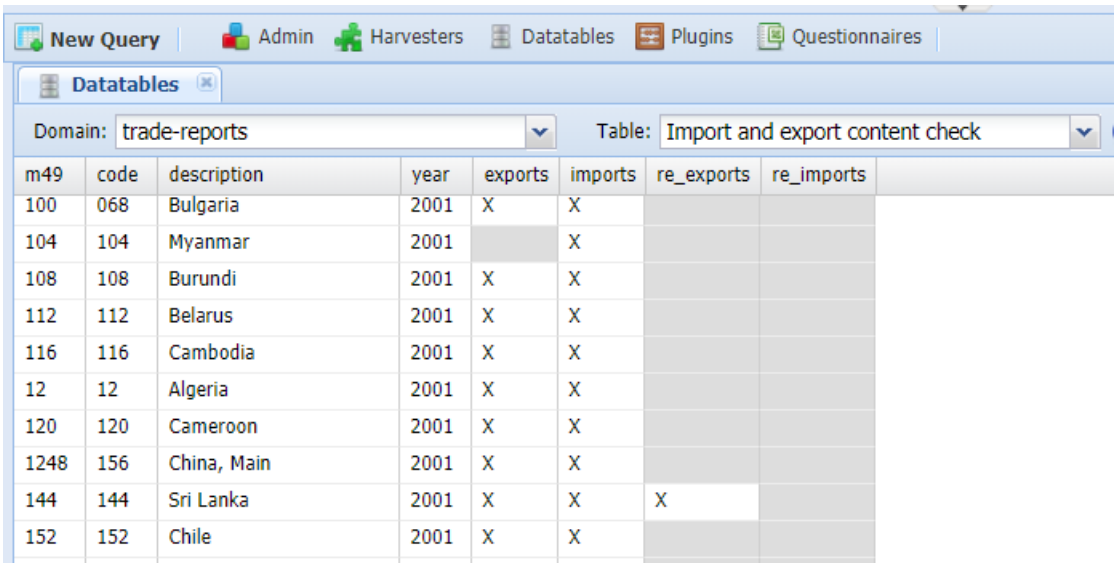


m49	code	description	year	flow	hs_n	hs_min	hs_max	hs_min_diff	hs_max_diff	records_count	records_diff
52	52	Barbados	2016	1	1	10	10	false	false	25763	0.0106
52	52	Barbados	2017	1	1	10	10	false	false	26513	0.0291
52	52	Barbados	2018	1	1	10	10	false	false	26819	0.0115
52	52	Barbados	2000	2	1	7	7			1019	
52	52	Barbados	2001	2	1	8	8	true	true	1022	0.0029
52	52	Barbados	2002	2	2	6	7	true	true	1029	0.0068
52	52	Barbados	2003	2	1	8	8	true	true	1040	0.0107
52	52	Barbados	2004	2	1	8	8	false	false	1048	0.0077
52	52	Barbados	2005	2	1	10	10	true	true	1437	0.3712
52	52	Barbados	2006	2	1	10	10	false	false	1354	-0.0578

In the Figure above, it is possible to see that Barbados reported 1,022 export transactions at the HS 8-digit level in 2001, while it reported 1,029 transactions in 2002 (an increase of 0.685%, as indicated by the `records_diff` variable), which are either at the 6- and 7-digit level.

2.3.4 "Import and export content check"

This table shows by country/year whether the country reported import/exports (and eventuallu re-imports and re-exports) and indicates it with an "X".



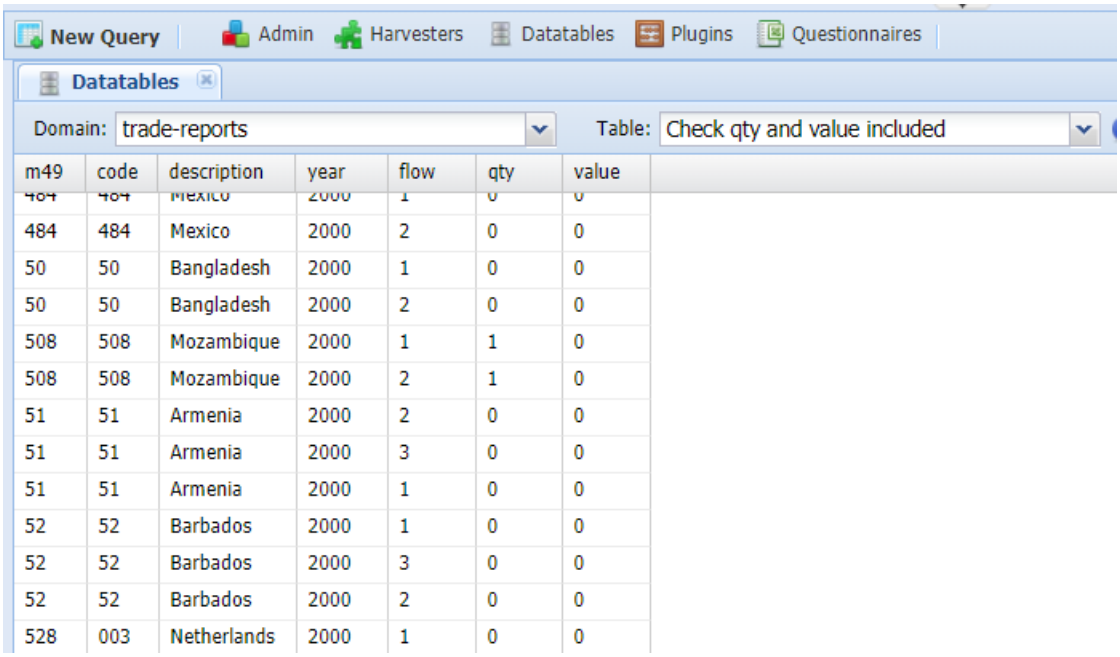
m49	code	description	year	exports	imports	re_exports	re_imports
100	068	Bulgaria	2001	X	X		
104	104	Myanmar	2001		X		
108	108	Burundi	2001	X	X		
112	112	Belarus	2001	X	X		
116	116	Cambodia	2001	X	X		
12	12	Algeria	2001	X	X		
120	120	Cameroon	2001	X	X		
1248	156	China, Main	2001	X	X		
144	144	Sri Lanka	2001	X	X	X	
152	152	Chile	2001	X	X		

Except Myanmar, in the previous Figure all countries reported imports and exports (Myanmar reported only imports); Sri Lanka also reported re-exports.

2.3.5 "Check qty and value included"

This reports shows whether information is missing for quantity or value by country/flow. If the information is missing, this is indicated with a "1".

**NOTE: THIS NEED TO BE CHANGED, SO TO DO IT AS IN OTHER TABLES WHERE BINARY INFORMATION IS GIVEN WITH X/BLANK INSTEAD OF 1/0.

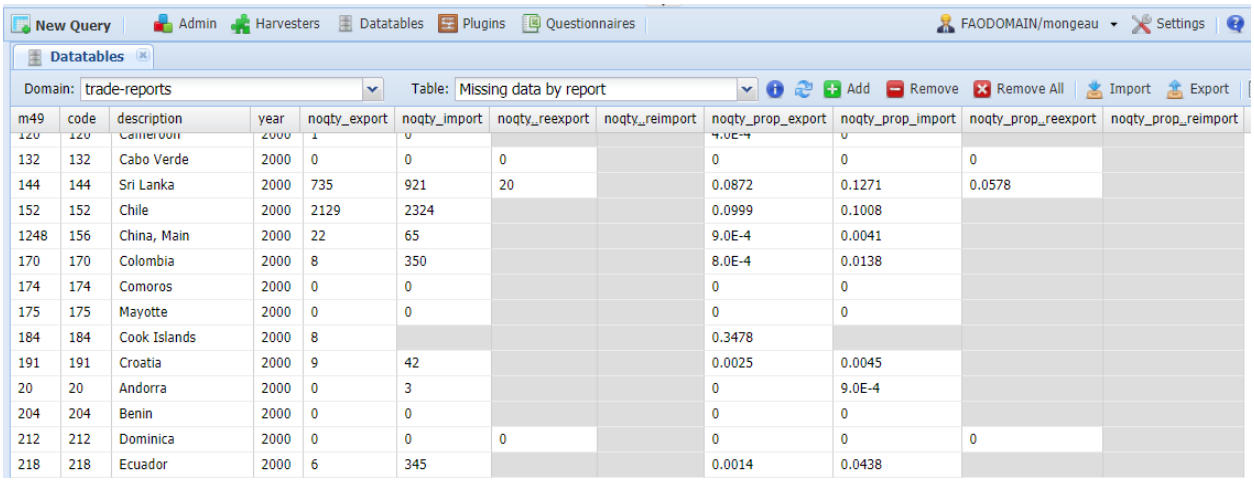


m49	code	description	year	flow	qty	value
484	484	Mexico	2000	1	0	0
484	484	Mexico	2000	2	0	0
50	50	Bangladesh	2000	1	0	0
50	50	Bangladesh	2000	2	0	0
508	508	Mozambique	2000	1	1	0
508	508	Mozambique	2000	2	1	0
51	51	Armenia	2000	2	0	0
51	51	Armenia	2000	3	0	0
51	51	Armenia	2000	1	0	0
52	52	Barbados	2000	1	0	0
52	52	Barbados	2000	3	0	0
52	52	Barbados	2000	2	0	0
528	003	Netherlands	2000	1	0	0

2.3.6 "Missing data by report"

This table contains information of missing or zero quantities by reporter and year. This information is encoded as $noqty_F$ and $noqty_prop_F$, where F is the flow (import, export, re-import, re-export). The former indicates how many observations are either missing or zero, and the latter is the proportion of the former to the total number of observations.

In the example below, Sri Lanka had 735 quantity observations which were either missing or zero, and the proportion of these to all of Sri Lanka's exports was 8.72%.



m49	code	description	year	noqty_export	noqty_import	noqty_reexport	noqty_reimport	noqty_prop_export	noqty_prop_import	noqty_prop_reexport	noqty_prop_reimport
120	120	Cameroon	2000	1	0			4.0E-4	0		
132	132	Cabo Verde	2000	0	0	0		0	0	0	
144	144	Sri Lanka	2000	735	921	20		0.0872	0.1271	0.0578	
152	152	Chile	2000	2129	2324			0.0999	0.1008		
1248	156	China, Main	2000	22	65			9.0E-4	0.0041		
170	170	Colombia	2000	8	350			8.0E-4	0.0138		
174	174	Comoros	2000	0	0			0	0		
175	175	Mayotte	2000	0	0			0	0		
184	184	Cook Islands	2000	8				0.3478			
191	191	Croatia	2000	9	42			0.0025	0.0045		
20	20	Andorra	2000	0	3			0	9.0E-4		
204	204	Benin	2000	0	0			0	0		
212	212	Dominica	2000	0	0	0		0	0	0	
218	218	Ecuador	2000	6	345			0.0014	0.0438		

3 Bilateral trade processing

After raw data have been harvested from UNSD/Eurostat, and after a decision on data completeness and quality has been achieved, bilateral raw data can be processed with the "Complete Trade Flow CPC" plugin. When this completes, its output is saved to the "Bilateral Trade Flow (CPC)" dataset, were bilateral transactions are saved by year as unique combinations of reporter and partner M49 codes, CPC item codes, SWS element codes that represent the flow and the type of variable (e.g., 5622 indicates imports in 1,000 US dollar, while 5910 indicates exports in metric tons). Once bilateral transactions are stored on SWS, these are aggregated by summing all partners for each reporter/item/element/year combinations by the "Total Trade CPC" plugin and saved into the "Total Trade (CPC)" SWS dataset.

This Section contains a description of the operations carried out by the "Complete Trade Flow CPC" and "Total Trade CPC" plugins, devoting more space to the former as it is the main plugin (the latter being basically an aggregation plugin).

3.1 "Complete Trade Flow CPC" plugin (bilateral transactions)

The "Complete Trade Flow CPC" plugin (SWS id: 5636) takes as input the raw UNSD and Eurostat data and saves bilateral transactions for reporters and partners expressed in M49 codes, items expressed in CPC codes, and elements in SWS-specific element codes, in the "Bilateral Trade Flow (CPC)" SWS dataset. In between, the plugin carries out several operations. These are explained below.

3.1.1 Aggregation of UNSD Tariff line shipments

The tariffline data from UNSD contains multiple rows with identical `reporter/partner/commodity/flow/year/qunit` combinations. See, e.g., the data reported in Table 4, where for the same year/reporter/partner/flow/qunit there are 15 recorded transactions involving the the item with code `22071090`.

Table 4: Example of multiple transactions by 'reporter' / 'partner' / 'flow' / 'year' / 'hs'

year	reporter	partner	flow	hs	value	weight	qty	qunit	chapter
2014	508	710	1	22071090	99570	NA	126000	7	22
2014	508	710	1	22071090	126530	NA	168000	7	22
2014	508	710	1	22071090	87950	NA	77141	7	22
2014	508	710	1	22071090	194740	NA	190719	7	22
2014	508	710	1	22071090	69580	116332	116261	7	22
2014	508	710	1	22071090	1050	NA	2871	7	22
2014	508	710	1	22071090	109770	NA	126000	7	22
2014	508	710	1	22071090	30050	NA	40000	7	22
2014	508	710	1	22071090	147840	NA	210000	7	22
2014	508	710	1	22071090	252100	NA	230538	7	22
2014	508	710	1	22071090	300	500	23	7	22
2014	508	710	1	22071090	28690	NA	40000	7	22
2014	508	710	1	22071090	36360	37847	42020	7	22
2014	508	710	1	22071090	2240	3197	2500	7	22
2014	508	710	1	22071090	75540	101700	126000	7	22

Table 5: Results of the aggregation of multiple transactions

year	reporter	partner	flow	hs	qunit	value	weight	qty	nrows
2014	508	710	1	22071090	7	184020	259576	286804	5
2014	508	710	1	22071090	7	1078290	NA	1211269	10

Multiple transactions can be aggregated, taking into account that transactions with missing weight and/or quantity can be aggregated with other transactions that have the same variables missing.

The results of the aggregation of the example reported in Table 4 are shown in the Table 5. In this case, even if there is a unique combination of `reporter/partner/flow/year/hs`, the rows are two: indeed, one reports the aggregated transactions where the `weight` variable was available, and the other reports the aggregated cases where it was missing.

3.1.2 Mapping UNSD Tariff line and Eurostat data

At this stage a standardization/mapping step is performed. The details are divided between UNSD Tariff line and Eurostat due to the nature of the differences among the two datasets.

3.1.2.1 UNSD Tariff line

1. UNSD Tariff line data reports area code with Tariff line M49 standard codes (which are different from official M49). The area code is converted in FAO country code using a specific conversion table provided by Team ENV. The table is contained in the `faoswsTrade` package as `faoswsTrade:m49faomap`. Table 6 contains a subset of it.

2. Countries that are not supposed to exist in the year for which the module runs are removed from the data (e.g., Serbia did not exist as a single official country before 2006). The information is retrieved from the `geographicAreaM49Reporter` SWS codelist. Table 7 contains some examples.

3. European countries (as reporters) are removed, as Eurostat, which contains this information, data will be used.

Table 6: Subset of the M49 to FAO codes mapping table

m49	fao
72	20
278	79
270	75
530	151
352	99
50	16
230	62
434	124
175	270
839	252

Table 7: Subset of the tables with countries and their 'startDate' and 'endDate'

code	description	selectionOnly	type	startDate	endDate
272	Serbia, Republic of	FALSE	country	2006-01-01	2999-12-31
183	Romania	FALSE	country	1900-01-01	2999-12-31
178	Eritrea	FALSE	country	1993-01-01	2999-12-31
40	Chile	FALSE	country	1900-01-01	2999-12-31
194	Saudi Arabia	FALSE	country	1900-01-01	2999-12-31
190	Saint Pierre & Miquelon	FALSE	country	1900-01-01	2999-12-31
252	Unspecified	FALSE	country	1900-01-01	2999-12-31
195	Senegal	FALSE	country	1900-01-01	2999-12-31
208	Tajikistan	FALSE	country	1992-01-01	2999-12-31
212	Syrian Arab Republic	FALSE	country	1900-01-01	2999-12-31
64	Faeroe Islands	FALSE	country	1900-01-01	2999-12-31
31	Bouvet Island	FALSE	country	1900-01-01	2999-12-31
211	Switzerland	FALSE	country	1900-01-01	2999-12-31
167	Czechia	FALSE	country	1993-01-01	2999-12-31
157	Nicaragua	FALSE	country	1900-01-01	2999-12-31

Table 8: Example of unmapped countries

m49	fao
270	75
280	79
716	181
634	179
471	252
51	1

4. Area codes that do not mapp to any FAO country code are removed and will be mirrored in a later step. All countries mapping to code 252 (which corresponds to undefined areas) are mapped to the 896 M49 code ("Other nei"). Table 8 show some examples.

5. The flow codes of re-Import (code 4) are recoded into Import (code 1) and codes of re-Export (code 3) to Export (code 2). This procedure is applied following UNSD standards:

   > Exports of a country can be distinguished as exports of domestic goods and exports of foreign goods. The second class is generally referred to as re-exports. The exports shown in our database contain both the exports of domestic and foreign goods. Re-exports are exports of foreign goods in the same state as previously imported; they are to be included in the country exports. It is recommended that they be recorded separately for analytical purposes. This may require the use of supplementary sources of information in order to determine the origin of re-exports, i.e., to determine that the goods in question are indeed re-exports rather than the export of goods that have acquired domestic origin through processing. Re-imports are goods imported in the same state as previously exported. They are included in the country imports. It is recommended that they be recorded separately for analytical purposes. This may require the use of supplementary sources of information in order to determine the origin of re-imports, i.e., to determine that the goods in question are indeed re-imports rather than the import of goods that have acquired foreign origin through processing. There are several reasons why an exported good might return to the country of origin. The exported good might be defective, the importer might have defaulted on payments or cancelled the order, the authorities might have imposed an import barrier, or demand or prices in the country of origin might have made it worthwhile to bring the good back.

   See: http://unstats.un.org/unsd/tradekb/Knowledgebase/Reexports-and-Reimports

6. HS codes are converted in FCL (*FAO Commodity List*) codes. More information on the HS-FCL-CPC mapping will be given below.

7. Information of the FCL units is added, i.e, to each FCL code its final unit of measurement is assigned.

8. Data conversion of units of measurements are applied to meet FAO standards, where all weights are reported in tonnes, animals in heads or 1000 heads and, for some commodities, just the value is provided. For example, if the originally-reported quantity is "units" and the FAO unit is "1000 heads", the quantity is divided by 1000. Some of these conversions, as the example just made, are simple mathematical conversions that involve multiplying by a fixed factor known *a priori*. Other conversions are less trivial and need to be obtained with information derived from within the data. This is handled in the next step

9. Non-livestock commodity specific conversions are added. If a quantity is originally expressed in "units", and weight is not available, a conversion factor for converting units into kilograms is required. In order to obtain these conversion factors, all transactions with non-standard quantities that have also the weight available are used to obtain "empirical conversion factors": weight are divided by quantity and the median by measurement of unit of this ratio is obtained:

Table 9: Subset of FCL units

fcl	fclunit
1096	heads
1068	1000 heads
521	mt
1083	1000 heads
1150	heads
44	mt
840	mt
1098	mt
1181	number
1126	heads

$$qw_u = \text{median}(\text{weight}/\text{quantity}_u)$$

where u is a given non-standard unit of measurement. This factor (qw_u) is the applied to all correspondent transactions that have no weight but are reported in unit u:

$$\text{weight} = \text{quantity}_u \times qw_u$$

To make this clear an example could be useful: suppose that a country does not report the weight of eggs, but reports units, in this case, we compute the median of the weight/quantity/1000 ratio for all countries where both weight and quantity are reported and then apply this median in order to have an idea of how many tonnes the reported quantities of that country weighs.[13] This procedure is not applied to livestock: in this case, a country/item specific datatable exists where weights of livestock are present (next step).

10. For livestock, the conversion from weight to final units (head/1,000 heads) where the reported unit is either missing or different from the final unit, is made by dividing the reported (or imputed as per previous steps) the weight by an average livestock weight by country and species. The table used has being compiled by both AGA division and by team CLFS, and is available in SWS as `Livestock weights` in the `trade-reference-files` domain. By calling these averages as aw, the final unit in heads/1,000 heads will be obtained as:

$$\text{quantity} = \frac{\text{weight}}{aw}$$

3.1.2.2 Eurostat

1. Eurostat classifies areas in their "geonomenclature" coding system. These codes are converted in FAO country codes using a specific conversion table, stored in the SWS in the `geonom2fao` datatable of the `trade-reference-files` domain. Table 10 contains the geonomenclature-FAO code correspondence table for some countries.

   Area codes not mapping to any FAO country code are reported and the records for these area codes are removed. All countries mapping to code 252 (which corresponds to undefined areas) are mapped to the 896 M49 code ("Other nei").

[13] Probably a weak point of this procedure is that not in all countries eggs weigh the same, thus a more realistic approach would be to compute regional medians (e.g., for Thailand use the Asian median of the weight/quantity ratio). The first-best approach, in any case, would be to have speficic tables derived from external sources. This is currently under investigation.

Table 10: Subset of geonom (Eurostat) to FAO codes mapping

code	faostat	active	name
1	68	68	France
2	15	15	Belg.-Luxbg
3	150	150	Netherlands
4	79	79	Fr Germany
5	106	106	Italy
6	229	229	Utd. Kingdom
7	104	104	Ireland
8	54	54	Denmark
9	84	84	Greece
10	174	174	Portugal

Table 11: EUR/USD exchange rates

eusd_year	eusd_exchangerate
2006	1.25565
2007	1.37064
2008	1.47171
2009	1.3928
2010	1.32689
2011	1.39141
2012	1.28557
2013	1.32816
2014	1.32884
2015	1.109625

2. CN8 codes are converted in FCL (*FAO Commodity List*) codes. This step is performed using the same HS to FCL mapping table as for UNSD Tariff line. More details will be given below.

3. Information of the FCL units is added. This step is straightforward since for Eurostat the units are for the vast majority the same as FAO units.

4. Some commodity specific conversions are needed as Eurostat reports the figures in a different unit with respect to FAO. With respect to UNSD data, this is only needed for few commodities, namely (codes are FLC): 1057 ("Chickens"), 1068 ("Ducks"), 1072 ("Geese"), 1079 ("Turkeys"), 1083 ("Pigeons Other Birds"), 1140 ("Rabbits"), 1181 ("Beehives").

5. Values are converted from EUR to USD using a table, stored in the SWS, with the official EUR/USD exchange rate for each year provided by the CLFS Team. The table is stored in the `EUR/USD conversion` SWS datatable of the `trade-reference-files` domain. Table 11 contains the exchange rates for some years.

3.1.2.3 HS-FCL-CPC code mapping

ESS has built a dataset for the conversion of country-specific HS codes to the FCL. Given that FCL-CPC conversions are, with rare exceptions, 1-to-1, once the HS codes are mapped to FCL they get converted to CPC, given that it is the classification system used in the trade datasets of the SWS.

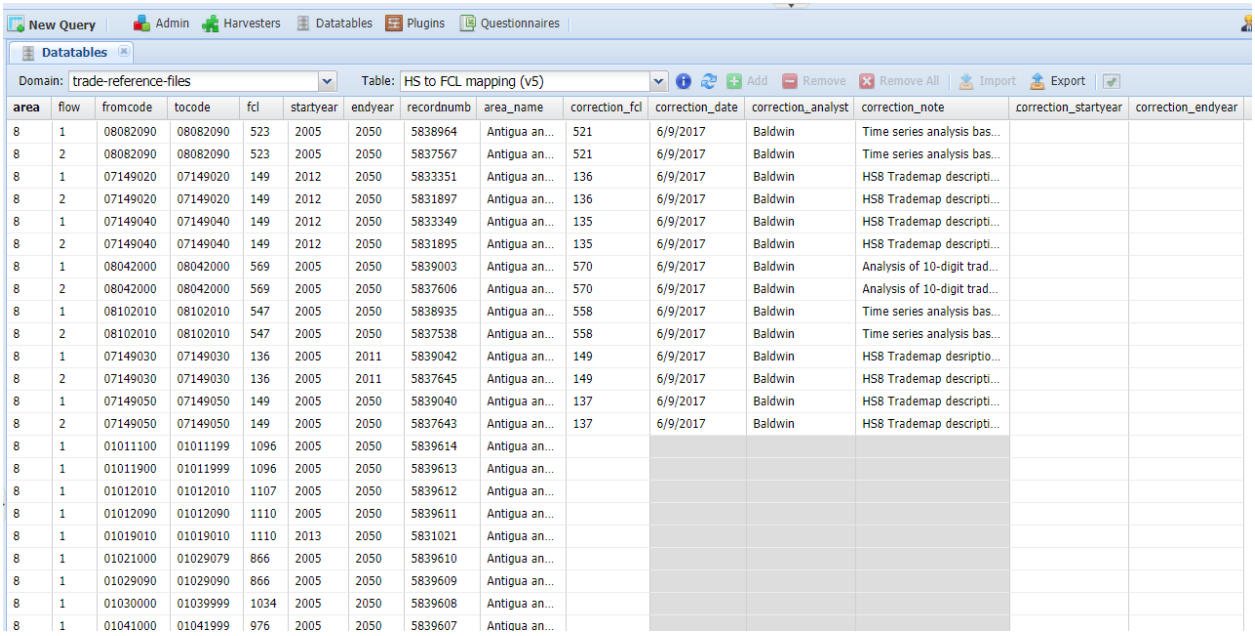
An example of the HS-FCL is given in Table 12. In the table, `area` (i.e., reporters) codes are given in FAO area list codes. There are two set of ranges: one is for the HS codes (`fromcode` and `tocode`) the other one is for dates (`startyear` and `endyear`). To give a concrete example, in the first row of the table, the HS codes

Table 12: Subset of the HS to FCL mapping table

area	flow	fromcode	tocode	fcl	startyear	endyear
109	2	5301210000	5301219999	773	2007	2050
81	2	15159060	15159060	340	2000	2000
58	1	2009790000	2009790000	519	2003	2012
66	2	012355	012355	1069	2000	2001
134	2	0809401000	0809401000	536	2000	2003
17	2	04041004	04041004	900	2007	2008
165	1	11062039	11062039	150	2003	2050
119	2	220410	220410	564	2002	2003
131	1	071350	071350	181	2008	2050
181	2	19019091	19019091	115	2000	2050

between `5301210000` and `5301219999` for any year between 2007 and 2050 for exports (flow = 2) of reporter 109 (Jamaica) will be converted to FCL 773 ("Flax Fibre and Tow").[14]

Said dataset was built starting from the MDB files used in the previous system (known as *Shark/Jellyfish*) and is called "historical dataset" given that it was built during the years ESS received data directly by the countries (contrary to the current system, where data comes from both UNSD and Eurostat). It is available in the "HS to FCL mapping (v5)" datatable of the "trade-reference-files" domain. An example of the actual table is shown below.



area	flow	fromcode	tocode	fcl	startyear	endyear	recordnumb	area_name	correction_fcl	correction_date	correction_analyst	correction_note	correction_startyear	correction_endyear
8	1	08082090	08082090	523	2005	2050	5838964	Antigua an...	521	6/9/2017	Baldwin	Time series analysis bas...		
8	2	08082090	08082090	523	2005	2050	5837567	Antigua an...	521	6/9/2017	Baldwin	Time series analysis bas...		
8	1	07149020	07149020	149	2012	2050	5833351	Antigua an...	136	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	2	07149020	07149020	149	2012	2050	5831897	Antigua an...	136	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	07149040	07149040	149	2012	2050	5833349	Antigua an...	135	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	2	07149040	07149040	149	2012	2050	5831895	Antigua an...	135	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	08042000	08042000	569	2005	2050	5839003	Antigua an...	570	6/9/2017	Baldwin	Analysis of 10-digit trad...		
8	2	08042000	08042000	569	2005	2050	5837606	Antigua an...	570	6/9/2017	Baldwin	Analysis of 10-digit trad...		
8	1	08102010	08102010	547	2005	2050	5838935	Antigua an...	558	6/9/2017	Baldwin	Time series analysis bas...		
8	2	08102010	08102010	547	2005	2050	5837538	Antigua an...	558	6/9/2017	Baldwin	Time series analysis bas...		
8	1	07149030	07149030	136	2005	2011	5839042	Antigua an...	149	6/9/2017	Baldwin	HS8 Trademap desriptio...		
8	2	07149030	07149030	136	2005	2011	5837645	Antigua an...	149	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	07149050	07149050	149	2005	2050	5839040	Antigua an...	137	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	2	07149050	07149050	149	2005	2050	5837643	Antigua an...	137	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	01011100	01011199	1096	2005	2050	5839614	Antigua an...						
8	1	01011900	01011999	1096	2005	2050	5839613	Antigua an...						
8	1	01012010	01012010	1107	2005	2050	5839612	Antigua an...						
8	1	01012090	01012090	1110	2005	2050	5839611	Antigua an...						
8	1	01019010	01019010	1110	2013	2050	5831021	Antigua an...						
8	1	01021000	01029079	866	2005	2050	5839610	Antigua an...						
8	1	01029090	01029090	866	2005	2050	5839609	Antigua an...						
8	1	01030000	01039999	1034	2005	2050	5839608	Antigua an...						
8	1	01041000	01041999	976	2005	2050	5839607	Antigua an...						

The difference between the previous table and the one reported as example in Table 12 is that it contains some `correction_X` variables, where `X` is one of `fcl`, `startyear`, and `endyear`, and are used to correct previous mapping entries that were not correct. It also contains an identifier of the row (`recordnumb`) and some informative variables, as the country name (`area_name`), and some notes inserted after a correction was done (`correction_date`, `correction_analyst`, `correction_note`).

During trade processing, some HS-codes could not have been mapped because a specific correspondence has not been found in the "HS to FCL mapping (v5)" datatable. For these cases, an additional table was created

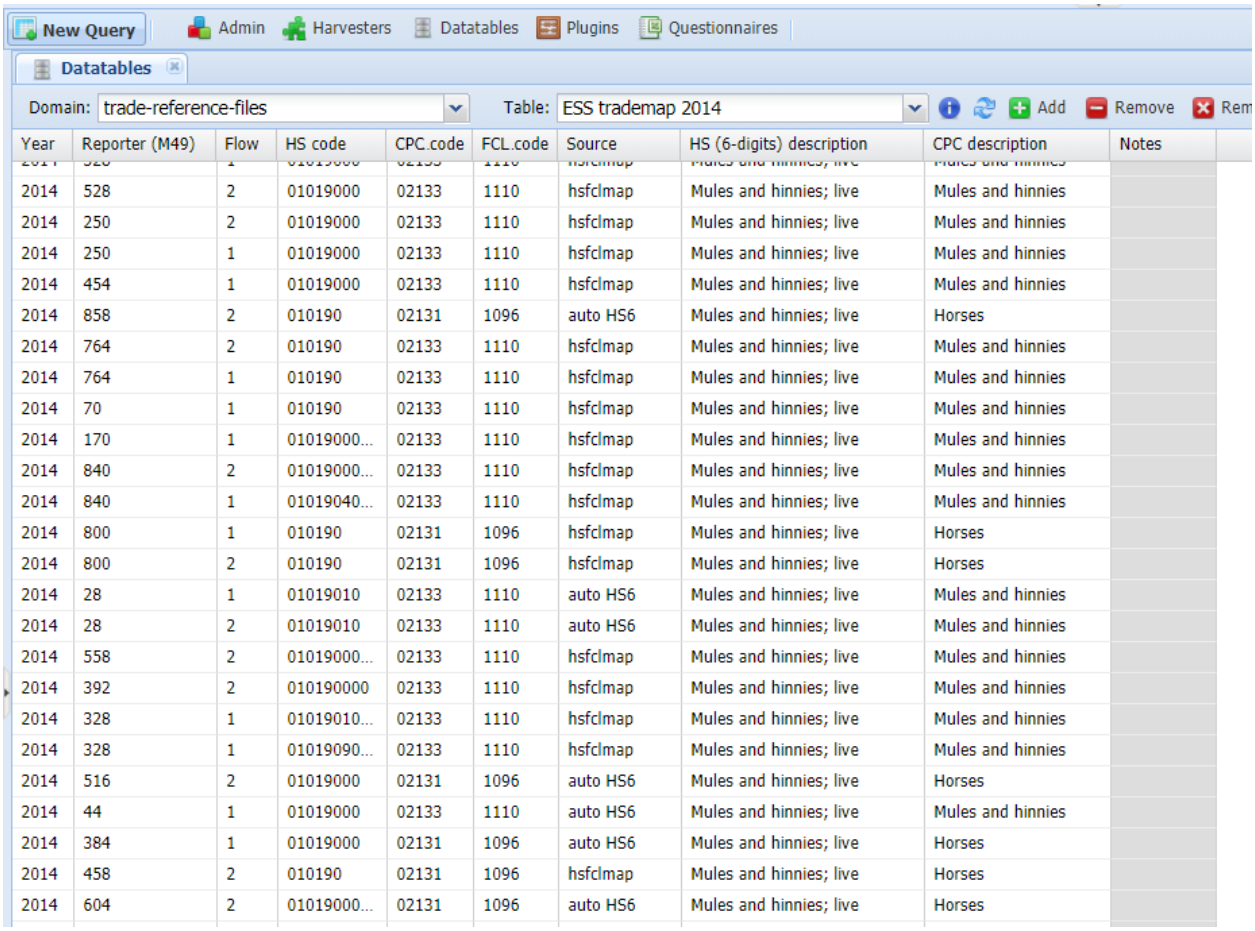
[14] The ranges do not necessarily need to contain multiple elements. For instance, many entries in the table show that `fromcode` and `tocode`, and/or `startyear` and `endyear` are the same.

Table 13: Subset of additions to the HS to FCL mapping table

year	reporter_fao	flow	hs	fcl	details
2002	33	1	3301230000	753	GTIS TL description
2014	219	1	7099100	0	NA
2015	107	1	80261	234	Generic HS2012 to FCL unique six-digit match
2015	146	2	80830901	0	NA
2014	299	1	110510	0	NA
2015	230	1	80261	0	NA
2014	68	2	3081100	0	FISHERY CODE: DROP
2014	299	1	190240	0	NA
2015	299	1	100199	0	NA
2014	7	1	2075200	1073	Standard_HS12
2014	230	2	20752	0	NA
2014	299	1	220600	517	Generic HS2012 to FCL (could also be mapped to 26,39)
2014	153	1	10085000	0	NA
2014	180	1	9011200	0	NA
2013	169	1	3083000000	0	FISHERY CODE: DROP

so to extend the main correspondence table for missing connections. This table is available as "HS to FCL mapping (v4)" on the "trade-reference-files" SWS domain. An example of this table is given in Table 13.

The previous two tables ("HS to FCL mapping (v4)" and "HS to FCL mapping (v5)"; the two combined will be called simply "HS to FCL mapping") are used as the basis of the current mapping mechanism, which is now done once new data arrives and is saved into year-specific "trademaps", available on SWS as `ESS trademap YEAR` datatables of the "trade-reference-files" domain, where `YEAR` is a specific year. Some entries of the 2014 `ESS trademap 2014` datatable are reported below.



Year	Reporter (M49)	Flow	HS code	CPC.code	FCL.code	Source	HS (6-digits) description	CPC description	Notes
2014	528	2	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	250	2	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	250	1	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	454	1	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	858	2	010190	02131	1096	auto HS6	Mules and hinnies; live	Horses	
2014	764	2	010190	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	764	1	010190	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	70	1	010190	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	170	1	01019000...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	840	2	01019000...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	840	1	01019040...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	800	1	010190	02131	1096	hsfclmap	Mules and hinnies; live	Horses	
2014	800	2	010190	02131	1096	hsfclmap	Mules and hinnies; live	Horses	
2014	28	1	01019010	02133	1110	auto HS6	Mules and hinnies; live	Mules and hinnies	
2014	28	2	01019010	02133	1110	auto HS6	Mules and hinnies; live	Mules and hinnies	
2014	558	2	01019000...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	392	2	010190000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	328	1	01019010...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	328	1	01019090...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	516	2	01019000	02131	1096	auto HS6	Mules and hinnies; live	Horses	
2014	44	1	01019000	02133	1110	auto HS6	Mules and hinnies; live	Mules and hinnies	
2014	384	1	01019000	02131	1096	auto HS6	Mules and hinnies; live	Horses	
2014	458	2	010190	02131	1096	hsfclmap	Mules and hinnies; live	Horses	
2014	604	2	01019000...	02131	1096	auto HS6	Mules and hinnies; live	Horses	

The table contains a `Source` variable that indicates how the correspondence was obtained. There are four cases:

- `hsfclmap`: this indicates that the mapping procedure found the specific HS code in the "HS to FCL mapping" table; for instance, by considering the examples reported in Table 12, supposing that Jamaica reports an export with HS 5301211020 in 2018, then it will be mapped to FCL 773, given that it falls in the ranges set by `fromcode-tocode` and `startyear-endyear` (first row of the table).
- `standard`: in cases where no direct connection can be established as `hsfclmap`, the mapping procedure extract the HS at 6-digits (HS6) from the specific HS code reported by the country and checks whether there is a 1-to-1 connection of that HS6 to an FCL in the "HS2012-6 digits Standard" SWS datatable of the "trade-reference-files" domain; if there is one, this get assigned to the specific HS code.
- `auto HS6`: when the HS code could not be mapped with neither the `hsfclmap` nor the `standard` method, then an empirical approach is used, i.e., all HS codes reported are cut at the 6-digits level, then the most common correspondence HS6-FCL is assigned.
- `manual`: It can happen that none of the previous methods were able to assign an FCL code to some HS codes, and these remain "unmapped"; an email containing these codes is sent to officers (subject is "Trade plugin: unmapped codes, year YEAR") and they will take care of map the codes themselves, or sent these to the country analysts.
- `previous`: this indicates that the correspondence has been copied from a previous trademap.

Each new trade validation round (involving a completely new year), will copy the previous year trademap and will assign `previous` to the `Source` variable. For instance, when year 2019 will be compiled (granted it is the first time it is), the map `ESS trademap 2018` will be copied, all `Souce` codes will be made `previous`, and will be finally saved as `ESS trademap 2019`. It is usually the case that countries have new codes, so these

will be initially "unmapped". For those codes for which a conversion with any of the `hsfclmap`, `standard`, or `auto HS6` is possible, the plugin will save the corresponding correspondencies, otherwise it will keep them unmapped. For this the `manual` intervention is required.

3.1.3 Unified official trade flows dataset

UNSD Tariff line and Eurostat datasets are ready to be merged together. Thus, the resulting table has all the countries worldwide.

3.1.4 Standardization, editing and outlier detection

3.1.4.1 Unit values computation

For each record having both quantity and value, the unit value (uv) is computed as follows:

$$uv = \frac{value}{quantity}$$

3.1.4.2 Missing quantities imputation

For records where the commodity has to be reported in quantity and the quantity is missing and the value is present, the corresponding quantity is imputed dividing the corresponding value by a median unit value:

$$quantity = \frac{value}{uv_{median}}$$

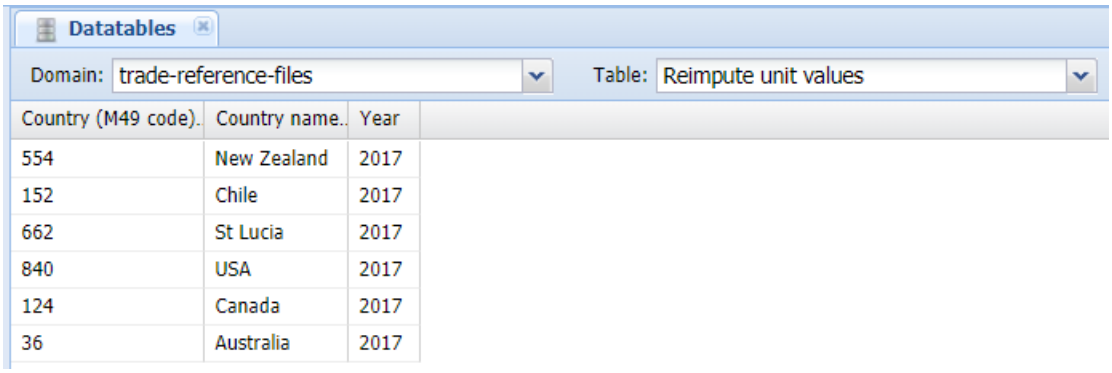
The median unit value (uv_{median}) is obtained in a specific-to-generic fashion (in all cases, the unit values are calculated separately for imports and exports). A first attempt is done by calculating unit values at the most specific HS level (i.e., the one at which the quantity is expressed). If the number of partners for which this unit value can be calculated is greater than a certain threshold (currently 10) the median unit value across partner is calculated and used for imputation. If the first attempt fails (i.e., it is not possible to calculate a unit value at the most specific HS level), then the same approach is used by taking into account more generic HS levels, in particular at eight and six digits, and the most generic level for which a sufficient number of partners (at least 10) is available is used for calculating the median unit value. Usually a suitable median can be calculated at the 8-digit level or, at least, at the 6-digit level. However, if the previous strategies fail (which implies that there is not a sufficient number of partners in order to calculate the median) two attempts at calculating a non reporter-specific median unit values (i.e., median unit values valid for all reporters) are sequentially undertaken: by HS and by FCL. In most cases it should be possible to calculate the median unit value by HS, thus that the FCL level is used as the strategy of last resort. Actually, for completeness sake, the very last *fallback* is the median unit value by flow. This is the most generic unit value that can be used for imputation and is calculated just for precaution, as it is very unlikely that an appropriate more specific median unit value can not be calculated.

In short, the first one of the following median unit values that can be calculated is used for imputation (import and export unit values are always calculated separately):

1. most specific HS code, across partners;
2. 8-digit HS level, across partners;
3. 6-digit HS level, across partners;
4. most specific HS code, across reporters/partners;
5. FCL code of the most specific HS code, across reporters/partners;
6. by flow (without taking into account any commodity code; this is very unlikely to be applied).

For how flags are assigned, please refer to Section "Flag management".

3.1.4.3 Re-imputation of flows Given the cross-section nature of the "Complete Trade Flow CPC" plugin, there can be cases where imputations of missing quantities are not of the same order of magnitude with respect to the time series of the country. For instace, if a country reports quantity data for 19 out of 20 partners, but the reported data was biased upwards by a factor of, say, 1,000, then when an imputation is done for the missing quantity partner, it will be also be biased by a factor of 1,000. In order to take this into account, a "re-imputation" is carried out for specific countries where imputations of low quality have been found during the validation process. The countries for which the re-imputation is carried out are listed in the "Reimpute unit values" SWS datatable of the "trade-reference-files" domain. An extract of this table is shown below.



Country (M49 code).	Country name..	Year
554	New Zealand	2017
152	Chile	2017
662	St Lucia	2017
840	USA	2017
124	Canada	2017
36	Australia	2017

The re-imputation process consists in temporary aggregating the bilateral flows so to obtain what the totals would be, then compute the total unit value and compare it with the previous years average. If the current unit value is greater than 150% of its average, then the current unit value is re-calculated by taking its previous value and applying a variation equal to the variation of the median unit value for all reporters (of the specific item and flow considered), and the quantity is re-calculated with the new unit value. More specifically, if

$$UV/UV_{prev} > 1.5$$

where UV is the current unit value and UV_{prev} is the average of the unit value of the previous three years, then the plugin calculates

$$UV_{new} = UV_{prev} \times (1 + UVT_{var})$$

where UVT_{var} is the obtained as

$$UVT_{var} = \frac{\text{median}(UVT_t)}{\text{median}(UVT_{t-1})}$$

which represents the variation rate of the total unit value for all reporters for a specific item/flow. It is constrained to be included in the -50% and +50%, so to avoid using an outlying variation.

3.1.4.4 Outlier detection and imputation

In the current version of the module, **no automatic outlier imputation is carried out**. The reason is that by comparing the results of the module by correcting outliers and previous FAOSTAT data, the two different datasets presented remarkable differences. Indeed, it was found that the *uncorrected* data was on average more similar to previous FAOSTAT data. For this reason, **automatic** correction is not currently being used, relying on a semi-automatic (guided) correction workflow that is done through an external validation tool. The steps of the guided validation are the following:

Table 14: Protected flags

flagObservationStatus	flagMethod
E	c
E	f
E	h
I	c
M	-
T	-
T	h
T	p
BLANK	-
BLANK	p
BLANK	q

1. a validation plugin for total trade is used that indicates which `reporter / commodity / flow / year` combination is likely to be an outlier. It computes various outlier detection routines on the data and assigns scores based on how many times a transaction has been found to be an outlier. This information is displayed on SWS by increasing levels of colouring that go from pale yellow to red, the last one being the one that indicates that a particular transaction has been found to be an outlier by all methods.
2. analysts select the series with outliers (by going from the most to the least severe cases) and use an interactive validation tool that allows to dig into the composition of total trade flows as it uses bilateral data. The tool displays the outliers, allows to use different methods for correcting them, and stores the correction that the analyst deemed required in a `corrections` dataset that is integrated into subsequent runs of the "Complete Trade Flow CPC" trade module. In order to have more information on this topic, please see the validation tool documentation.

When no statistical-based imputation seems appropriate, analysts can "force" some values (e.g., obtained by consulting external sources) by overwriting the values saved on SWS and using "protected" flags. When the modules are run, these protected figures will not be overwritten by any figure generated by the module. The list of protected flags is shown in the following table ("BLANK" stands for an empty flag).

The previous set of flags is a subset of the protected flags in `faoswsFlag::flagValidTable`. The difference is that as trade is concerned, the **(BLANK, c)**, **(BLANK, h)**, and **(T, c)** are flags actually given by the module, thus they should not be considered protected. See the "Flags management" Section to have more details on flags.

3.1.5 Mirroring

The module generates a list of non-reporting countries, i.e., those present as partners but missing as reporters. For these countries the mirroring routine is applied: the corresponding trade of the non-reporting countries are extracted from the partners, inverting the flows. The quantities are the same while the values are corrected by a factor of 12% due to the CIF/FOB (*Cost, Insurance and Freight / Free on Board*) conversion (i.e., original imports are divided by 1.12, while original exports are multiplied by 1.12).[15]

There is also another condition for which mirroring is applied: when a flow is completely missing for a country when it is a reporter. For instance, if in a given year Tanzania did not report any flow as export, but did so for imports, the mirroring procedure will be used for exports. While cases like this are relatively less frequent than those for which a country is a complete non-reporter, they do happen.

[15]CIF "rely on importers' declarations, and include all trade costs (except tariffs and domestic taxes after the border)", while FOB rely "on exporters' declarations, and does not include trade costs." http://www.cepii.fr/PDF_PUB/wp/2011/wp2011-10.pdf

The following table shows an example of the mirroring procedure. Country A is a non-reporter in 2018, so country B's exports will be used to impute country A's imports. First, the quantities will be copied as reported by country B (890); values from country B will be increased by the CIF/FOB margin of 12% (1,400 $\times$ 1.12 = 1,568).

	Country A			Country B	
	imports			exports	
year	value	quantity		value	quantity
2016	1450	800		1250	800
2017	1550	950		1390	925
2018	1568	890	←	1400	890

Notice in the example, that in the years that both countries are reporter:

- Not necessarily quantities are the same (e.g., in 2017 they are different): quantities may not be the same because of several reasons: countries use Tariff line descriptions that can bring some difference on how items are recorded as they are aggregated into the CPC; timing differences in how transactions are recorded (e.g., a transaction can be exported in December, but arrives at destinatio in January); transaction may have been reported with non-standard units by at least one reporter; net weight was used by one reporter, while gross weight was used by the other; etc.
- Not necessarily the CIF/FOB margin is 12% (e.g., in 2016 it is 16%): our choice of a fixed CIF/FOB markup is simply for convenience, as these margins are neither time- nor distant-constant, i.e., they do evolve on time and they are different depending on geographic proximity of reporters/partners. making use of different CIF/FOB margins will be explored in the future.

3.1.5.1 Forced mirroring Additionally to countries that are completely missing, or those that appear but that only report one of the two flows (either exports or imports), some countries that do appear with both imports and exports are mirrored. This is called "Forced mirroring". This happens for countries that have been conmsidered to have incomplete or low quality data (for any reason). This is *not* an automatic process and is decided after having analysed the dataset through the pre-processing reports. When a country is considered to have incomplete data, so that the mirroring process would improve the coverage, it is placed in the "Force mirroring" datatable of the "trade-reference-files" domain.

3.1.5.2 Discarding incompatible mirror data ("TP criterion")

The mirroring procedure is an approximation of the actual flows of a country, as they are inferred from the declarations made by reporting countries. It may happen that these are structurally similar, so that the mirror gives a very good approximation, but it can happen that the mirroring mechanism results in data that is not in line with the time series of the mirrored country. For instance, if a main partner for a given commodity is also a non-reporter for a given year, then the resulting total flow will be much lower than it should be. This is why the mirrored flows are checked for completeness, and if they are considered to be too low, then the mirrored flows get removed so that they are checked, and eventually supplemented with external data or manual imputation by analysts during the validation stage.

The criteria for which a mirrored flow is incomplete is the following:

- $value/value_{mean} < 0.6$
- $quantity/quantity_{mean} < 0.6$
- $quantity_{mean} > 1,000$ tonnes

where $value_{mean}$ and $quantity_{mean}$ are the 5-year average of the lagged value of *value* and *quantity*, respectively, which are the aggregated monetary value and quantity of a reporter's given item/flow.

The following table contains an example of a country that is non-reporter in 2014 and 2018. The upper side of the table contains bilateral flows with its threee partners, as the partners declared; the lower part contains the aggregation of the bilateral flows declared by the country, indicated as "total", and the sum of the flows declared by the other partners (i.e., those in the upper part of the table), indicated as "mirrored". Given that 2018 is missing (orange highlighted part), the sum of flows declared by partners (in the yellow highlighted cell) are considered to fill the gap. However, these represent only around 40% of what is usually declared by the country (making the comparison with its 5-year average). Thus, according to the criteria shown above, such aggregation is not used to impute the missing flows.[16]

		2013	2014	2015	2016	2017	2018	*mean*	*ratio*
Partner A									
	value	150	165	182	200	220	242		
	quantity	100	110	121	133	146	161		
Partner B									
	value	300	330	363	399	439	483		
	quantity	210	231	254	280	307	338		
Partner C									
	value	1,000	1,100	1,210	1,331	1,464			
	quantity	800	880	968	1,065	1,171			
Total									
	value	1706	1500	1903	2017	2200			
	quantity	1220	1101	1330	1421	1548			
Mirrored									
	value	1,450	1,595	1,755	1,930	2,123	725	*1,770*	*0.41*
	quantity	1,110	1,221	1,343	1,477	1,625	499	*1,355*	*0.37*

When the plugin removes mirrored flows because they did not meet the 60% threshold, then an email is sent to officers informing them that the removal took place. The subject of the email is (e.g., for 2018) "Trade plugin: Excluded Tp, year 2018".

3.2 Flags management

The module assigns two types of flags ("Observation Status" and "Method") once some conditions are met.

The first flags that all data are given are a "BLANK" Observation Status flag and an "h" Method flag. They indicate that data are official and were harvested, respectively. After these, flags are assigned depending on the type of operation is done on the data . The different kind of flags, and the conditions that should be met in order to assign them, are reported in Table 15.

An observation can have multiple Observation Status flags and Method flags associated with it. The final flag is the "weakest" flag: the `flagWeightTable` table of the `faoswsFlag` R package contains the weights

[16]The example shows the reason why the aggregated mirrored transaction did not meet the threshold: in this case, an important partner (partner c) was missing (probably because it was also a non-reporter).

Table 15: Trade flags

ObservationStatus	Method	When
BLANK	h	Official data, harvested (No unit or currency conversion made) representing only one flow.
BLANK	i	Official data, identity (with unit conversion or currency conversion made) e.g: kg/ton; eur/$; $/ton (so can be only one or multiple flows)
BLANK	s	Official data, sum (where shipments have been added/aggregated; and no conversions have been done)
BLANK	p	Official data, publication (collected from official country publication, national website, UNCOMTRADE, TRADEMAP)
I	e	Imputation, module (if the quantities have been imputed based on median unit values)
I	i	Imputation, identity (for unit value if quantities or values are imputed)
I	s	Imputation, sum (if one of the aggregated items into one FCL/CPC was imputed). i.e. if one or more flows representing >10% of total have been modified.
T	i	Mirror data, identity (only for $ values due to +/-12% CIF/FOB conversion)
T	c	Mirror data, of only one flow (includes "blank/i" flows)
T	s	Mirror data, sum (for quantities/values if they are aggregated)
T	p	Mirror data from TRADEMAP, USDA, publications (for quantities/values if they are collected from publications and online databases)
E	f	Manual estimation
X	p	Estimation/imputation from UNCOMTRADE or TRADEMAP

that should be assigned to all flags, and the one with the lowest value prevails over the other flags. For instance, if one shipment if official ("BLANK" Observation Status flag) needs to be aggregated with another shipment that was originally missing the quantity and was thus imputed ("I" observation status flag), the final flag will be "I" as it is the weakest between the two flags involved.

There is an exception in flag aggregation, for which the standard "weakest flag is assigned" rule is not applied: when aggregation is done and some imputed observations are addends, the final flag will be "I" if and only if the imputed data accounts for more than 10% of the final aggregation.

An example of flag aggregation is reported below. It shows two items for the same reporter, both having 4 shipments with the fourth having missing quantity in both cases. After the shipment with missing data is imputed (it gets the "I" Observation Status flag), the total bilateral transaction is computed and the final Observation Status flag will be "BLANK" for Item A because the imputed shipment accounts for less than 10% of the total transacttion (specifically, 7.8%), while the flag for Item B is "I" because the imputation is more than 10% of the total (17.5%).

	Item A		**Item B**	
	quantity	flag	quantity	flag
Shipment 1	200	BLANK	950	BLANK
Shipment 2	1,250	BLANK	180	BLANK
Shipment 3	480	BLANK	50	BLANK
Shipment 4	150	I	250	I
% imputed	7.2%		17.5%	
Total bilateral transaction	**2,080**	**BLANK**	**1,430**	**I**

Flags are a FAO standard endorsed by the IDWG-TTF on Statistics. See the "Observation Status Code, Flags" document of the "Statistical Standard Series" to have a complete description. The document is available at:

http://intranet.fao.org/fileadmin/user_upload/scp/Standards_for_quality_compliance/SSS_Observation_Status_Codes___Flags___endorsed___December_2016_.pdf

3.3 Conversion to FAO SWS standards

At this point data is almost ready to be saved in the SWS. Additional mapping and aggregation are necessary in order to respect the SWS standards:

- Conversion of FCL into CPC codes. This conversion is based on the table of conversion 2.1 expanded. If some FCL codes are not mapped to CPC, the corresponding records are filtered out. Since the mapping between FCL and CPC is one-to-one there is no aggregation at this point. The routine just adds the corresponding CPC code.
- Conversion from FAO country code to M49.
- Each row of the final output must be either quantity- or value-specific, while so far the module keeps this information in one row. The information is therefore split in two separate rows.

The first submodule saves the final output in the `completed_tf_cpc_m49` SWS dataset, within the `trade` domain.

3.4 Use of validation corrections

While details of validation will be given in the next section, here it is necessary to say that if corrections for the year for which the plugin is run exist, they will replace the data generated by the plugin. For instance, if the plugin generated 2 million tonnes for some item in some country, but in the validation process it was found that this figure suffers from an "order of magnitude" problem because it was saved as tonnes while it should have actually be kilograms so that it got replaced by 2 thousand tonnes, 2 millions will be replaced by 2 thousand in the data that is processed. When values are replaced by validated figures, they get the **(I, e)** flags combination. Moreover, if a mirror flow exists it will be changed accordingly (in this case, the flags combination will be set to **(T, e)**).

The "corrections" mechanism was designed so that if the original figure that will be replaced is different from what the analyst corrected, the correction will be dropped. This can happen for different reasons, among which: raw data changed and pre-existing errors were corrected; the mapping table of a commodity was modified; etc.

Metadata for corrected figures is generated and will be saved on SWS.

More details on corrections generated during validation will be given in the "Data Validation" section.

3.5 Remove non-existent transactions

The plugin checks whether there are combinations of `reporter` / `partner` / `item` / `element` stored on SWS that are not generated by the plugin that are unprotected. If there are, said combination(s) will be removed from SWS. Indeed, the plugin should generate all possible combinations of those dimensions, and combinations that are not generated should not exist, except if they have been inserted manually (and thus have a protected flag). These "non-existing" combinations can be present on SWS because they were generated in the past but they should not have been. Also in this case, possible reasons can be multiple: the plugin had a bug that got corrected; the mapping table was modified; etc. Given that the SWS R-API does not have an option to check whether some combinations are not going to be overwritten, this needs to be done code-side by performing a set difference on the combinations available in the plugin and those generated by the plugin. Values and flags of the resulting combinations will be set to NA as there is actually no way to remove the observation. In any case, the result is substantially the same.

3.6 "Total Trade CPC" plugin (totals by reporter)

The "Total Trade CPC" plugin (SWS id: 5662) uses as input the output of the previous submodule. It aggregates total trade flows by reporting country for partners countries to a single total trade for each unique CPC commodity code. As the previous one, this submodule works by year.

The module saves the output into the dataset `Total Trade CPC`, within the `trade` domain.

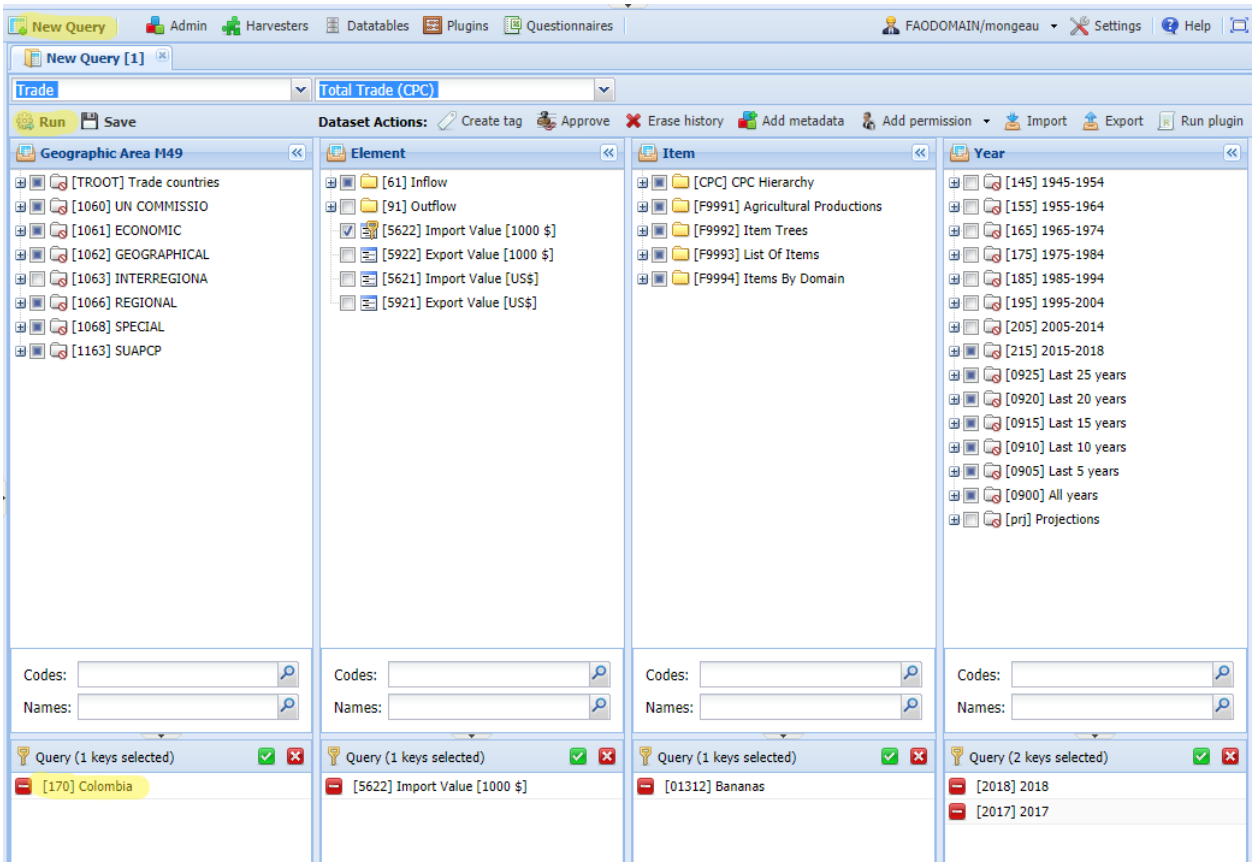
As for the "Complete Trade Flow CPC", non-existent transactions are removed. In this case, the combinations to compare are composed by `reporter` / `item` / `element`.

Flags follow the same aggregation rules as shown for the "Complete Trade Flow CPC" plugin, including the exception for which imputed flows that do not account for more that 10% of the total flow does not make the total flow to be considered as imputed.

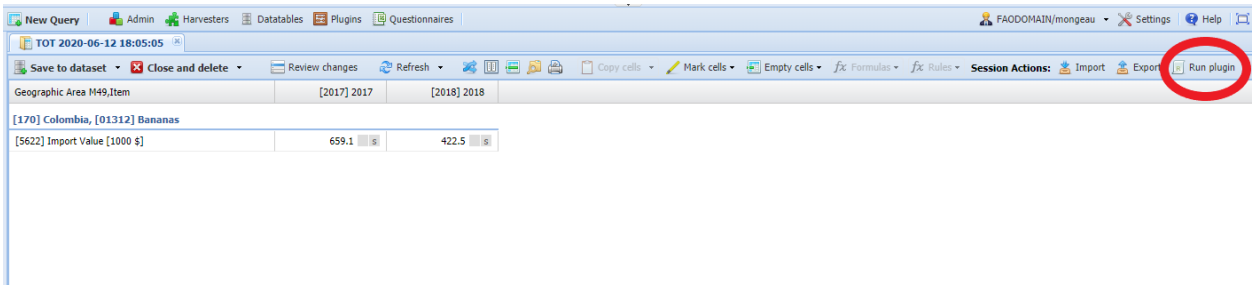
4 Data validation

Trade data validation is made with some helper plugins that allow to spot cases that need to be checked, and through an interactive tool for data validation. These plugins and the tool will be described in this section.

The plugins usually require that the user opens a session in the "Total Trade (CPC)" dataset, where the country that he/she is validating is selected. In order to open a session the user needs to click on "New query", then select the "Trade" domain, and the "Total Trade (CPC)" dataset. After that, the country for which validation is being carried out needs to be selected in the "Geographic Area M49" dimension (remaining dime4nsions can be set arbitrarily), then ther "Run" button needs to ne selected.



When the session is created the user can run a plugin by clicking on "Run plugin" in the upper right corner of the SWS session.



Running the plugins will allow the analyst understand which flows at the total level require attention. As the analyst collects a list of "suspicious" data at the total level, he/she can proceed the investigation and eventually correction at the bilateral level with the interactive validation tool (Shiny tool).

4.1 List of validation plugins

4.1.1 `Tradeoutliers_most_recent_version` plugin

This plugin calculates the 5-year average value of validated years for each item/flow/quantity (`meanOld`) and calculates the ratio of the current year to the average:

$$ratio = quantity/meanOld$$

If this ratio is lower than 0.25 or higher than 4, then the correpondent flow/item combination will be marked as outlier:

$$\text{if } ratio < 0.25 \text{ OR } ratio > 4 \implies outlier$$

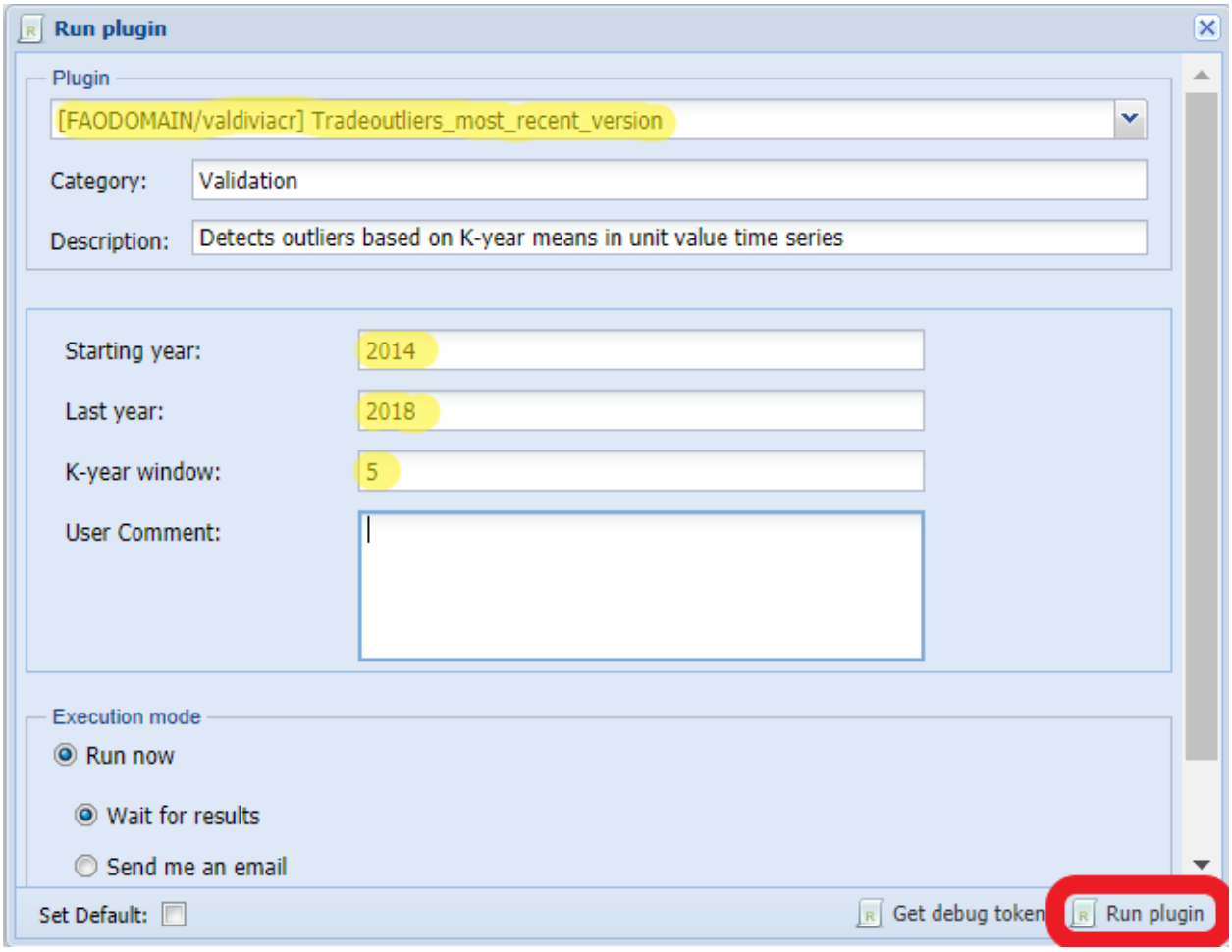
Also, a check on unit values is carried out and it check whether the unit value decreases by less than 50% or increases more then 100%:

$$\text{if } \widehat{UV} < -0.5 \text{ OR } \widehat{UV} > 1 \implies outlier$$

where $\widehat{UV} = UV_t/U_{t-1} - 1$ is the growth rate of UV (Unit Value).

In order to avoid checking small quantities, which proxies relatively unimportant items for the country (at least trade is concerned), quantities below given country-specific thresholds are not checked for outliers. The list of these thresholds is available in the "Trade outlier country thresholds" of the "trade-reference-files" domain.[17]

[17] When a country-specific threshold is not available, 1,000 is used.



After successful completion, the plugin sends an email with subject "outlierList" with an Excel file like the following:

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	geographi	geographi	measured	measured	measured	measuredElement	timePoint	Value	flagObserv	flagMetho	meanOld	growth_rate	ratio
2	170	Colombia	'0114	Sorghum	5630	Import Unit Value	2017	3933.517		i	230.5125	20.2346902	17.06422
3	170	Colombia	'0114	Sorghum	5630	Import Unit Value	2018	186.5422		i	230.5125	-0.95257622	0.80925
4	170	Colombia	'01252	Green garl	5630	Import Unit Value	2018	674.0979		i	941.0996	-0.52052805	0.716288
5	170	Colombia	'01322	Lemons ar	5630	Import Unit Value	2015	1200.95		i	262.8872	0.83944153	4.568309
6	170	Colombia	'01323	Oranges	5930	Export Unit Value	2018	433.9163		i	425.9218	1.28164295	1.01877
7	170	Colombia	'01324	Tangerines	5630	Import Unit Value	2015	750.8247		i	287.0833	1.50493776	2.615355
8	170	Colombia	'01324	Tangerines	5630	Import Unit Value	2017	395.7347		i	287.0833	-0.62406795	1.378467
9	170	Colombia	'01802	Sugar cane	5930	Export Unit Value	2015	3043.412		i	53.7385	26.5088913	56.63374
10	170	Colombia	'01802	Sugar cane	5930	Export Unit Value	2017	17000		i	53.7385	NA	316.3468
11	170	Colombia	'01802	Sugar cane	5930	Export Unit Value	2018	1963.256		i	53.7385	-0.88451438	36.5335
12	170	Colombia	'21932.01	Degras	5630	Import Unit Value	2018	479.0001		i	127.6722	1.24541956	3.751796
13	170	Colombia	'24320	Malt, whe	5930	Export Unit Value	2017	392.5733		i	509.5064	-0.6419074	0.770497
14	170	Colombia	'39120.14	Bran of Pu	5630	Import Unit Value	2017	342.6589		i	97.38873	1.73756389	3.518466
15													

The email body indicates which quantity threshold has been used to check for outliers.

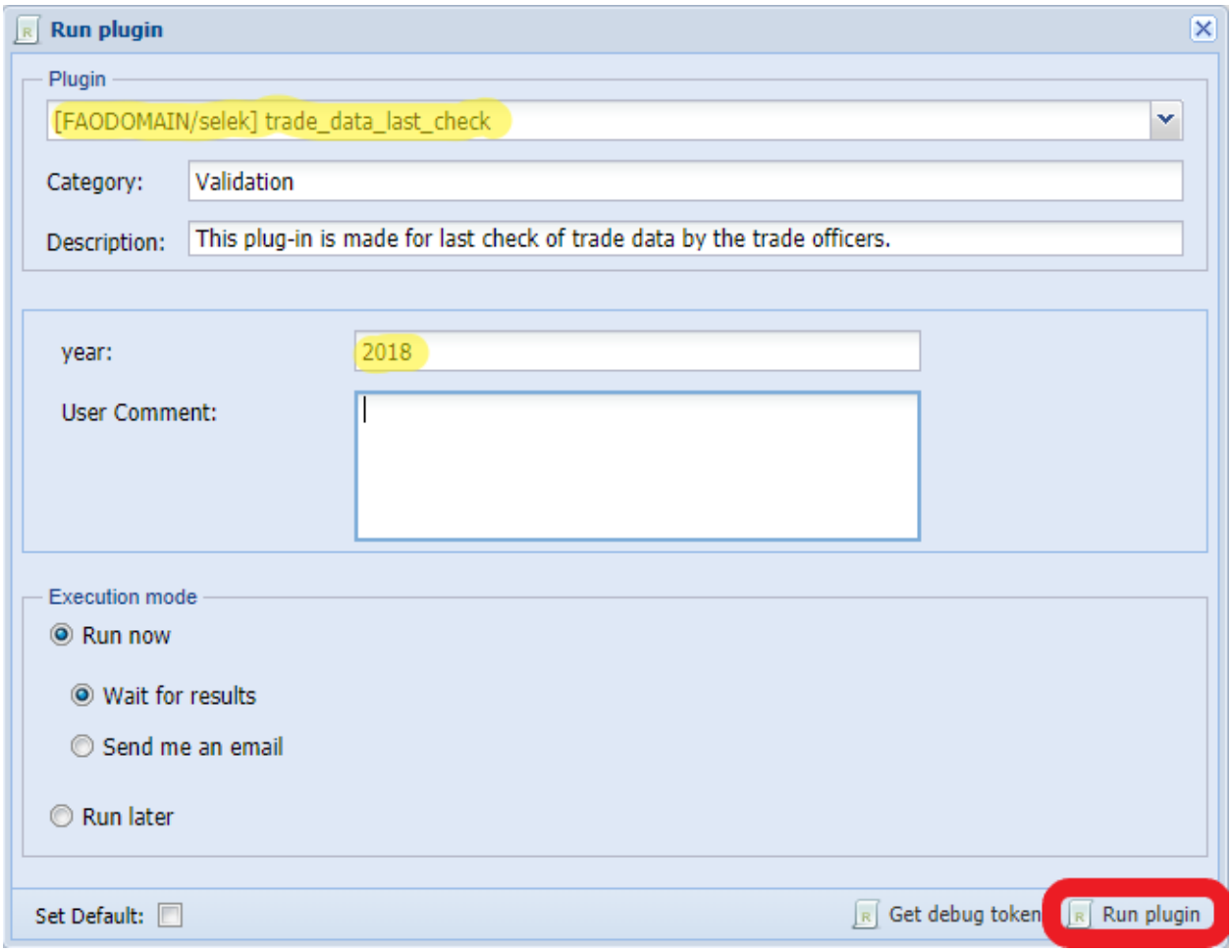
4.1.2 `trade_data_last_check` plugin

This plugin calculate the ratio of the current year quantity and diveds it by the 5-year moving average (`movav`) of the same flow/item:

quantity/movav

It sends an Excel file that indicates:

- yellow cells: $quantity/movav < 0.5$, i.e., values less than the 5-year average of at least 50%
- red cells: missing values
- boldface: official data

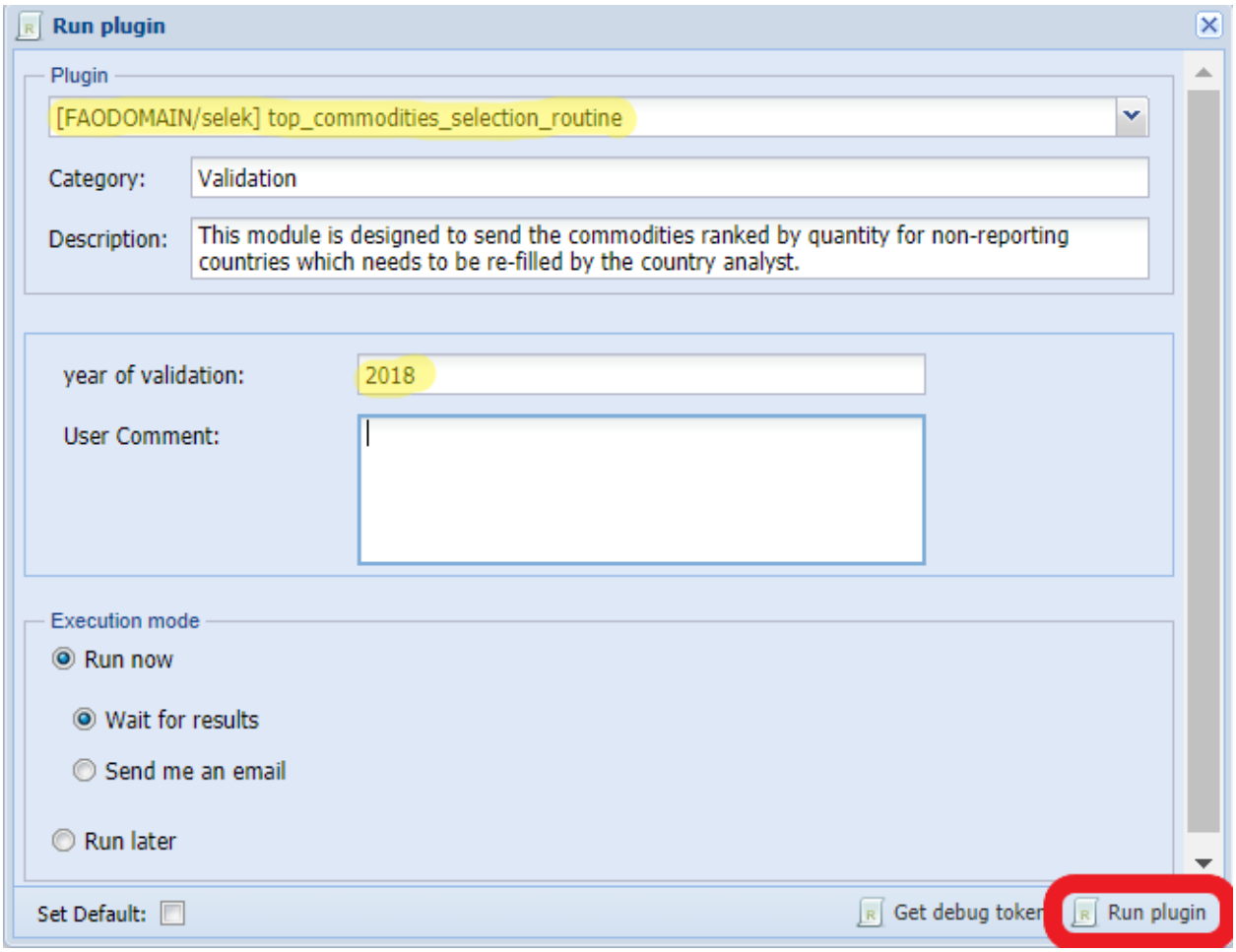


When the plugin completes, it sends an email with subject "trade_data_last_check" with an Excel file like the following:

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	geographi	measure	measured	geographic	measuredItemCPC	measuredElemer	2013	2014	2015	2016	2017	2018	5_year_average	
2	170	5610	'0112	Colombia	Maize (corn)	Import Quantity	3,635,280	**3,961,571**	**4,717,637**	**4,586,084**	**4,930,065**	**5,409,552**	4,366,128	
3	170	5610	'0111	Colombia	Wheat	Import Quantity	1,413,789	**1,824,499**	**1,704,727**	**2,095,562**	**1,890,615**	**1,722,001**	1,785,838	
4	170	5610	'21910.03	Colombia	Cake of Soya bean	Import Quantity	1,048,949	**1,016,341**	**1,160,890**	**1,235,220**	**1,322,651**	**1,395,094**	1,156,810	
5	170	5610	'0141	Colombia	Soya beans	Import Quantity	360,624	**448,170**	**580,264**	**530,405**	**538,116**	**650,474**	491,516	
6	170	5610	'24490	Colombia	Other non-alcoholi	Import Quantity	26,007	**30,865**	**329,259**	**671,772**	**684,639**	**664,926**	348,508	
7	170	5610	'2161	Colombia	Soya bean oil	Import Quantity	223,694	**298,781**	**310,672**	**365,747**	**353,639**	**344,631**	310,506	
8	170	5610	'0115	Colombia	Barley	Import Quantity	254,577	**253,469**	**253,811**	**297,069**	**293,217**	**321,910**	270,429	
9	170	5610	'23520	Colombia	Refined sugar	Import Quantity	287,111	**109,371**	**61,695**	**232,361**	**196,583**	**105,782**	177,424	
10	170	5610	'2165	Colombia	Palm oil	Import Quantity	117,534	**121,728**	**125,495**	**227,409**	**195,964**	**310,921**	157,626	
11	170	5610	'23161.02	Colombia	Rice, Milled	Import Quantity	146,117	**81,051**	**227,692**	**243,548**	**68,647**	**57,939**	153,411	
12	170	5610	'39160	Colombia	Brewing or distillin	Import Quantity	92,554	**128,588**	**131,241**	**189,092**	**216,177**	**218,406**	151,530	
13	170	5610	'0114	Colombia	Sorghum	Import Quantity	496,872	**101,404**	**23,135**	**43,078**	**125**	**54,908**	132,923	
14	170	5610	'39130.04	Colombia	Gluten Feed and M	Import Quantity	109,101	**105,511**	**131,005**	**153,352**	**128,448**	**128,923**	125,484	
15	170	5610	'01341	Colombia	Apples	Import Quantity	100,278	**113,383**	**100,026**	**97,500**	**101,493**	**102,233**	102,536	
16	170	5610	'24110	Colombia	Undenatured ethyl	Import Quantity	111,487	**78,864**	**87,053**	**87,084**	**130,852**	**210,467**	99,068	

4.1.3 `top_commodities_selection_routine` plugin

This plugin sorts in decreasing order commodities by quantity for non-reporting countries, showing which items have been deleted by the "TP criterion" (see the "Discarding incompatible mirror data ("TP criterion")" Section), or are relatively low when compared to their 5-year average.



When the plugin completes, it sends an email with subject "main_commodities" with an Excel that indicates:

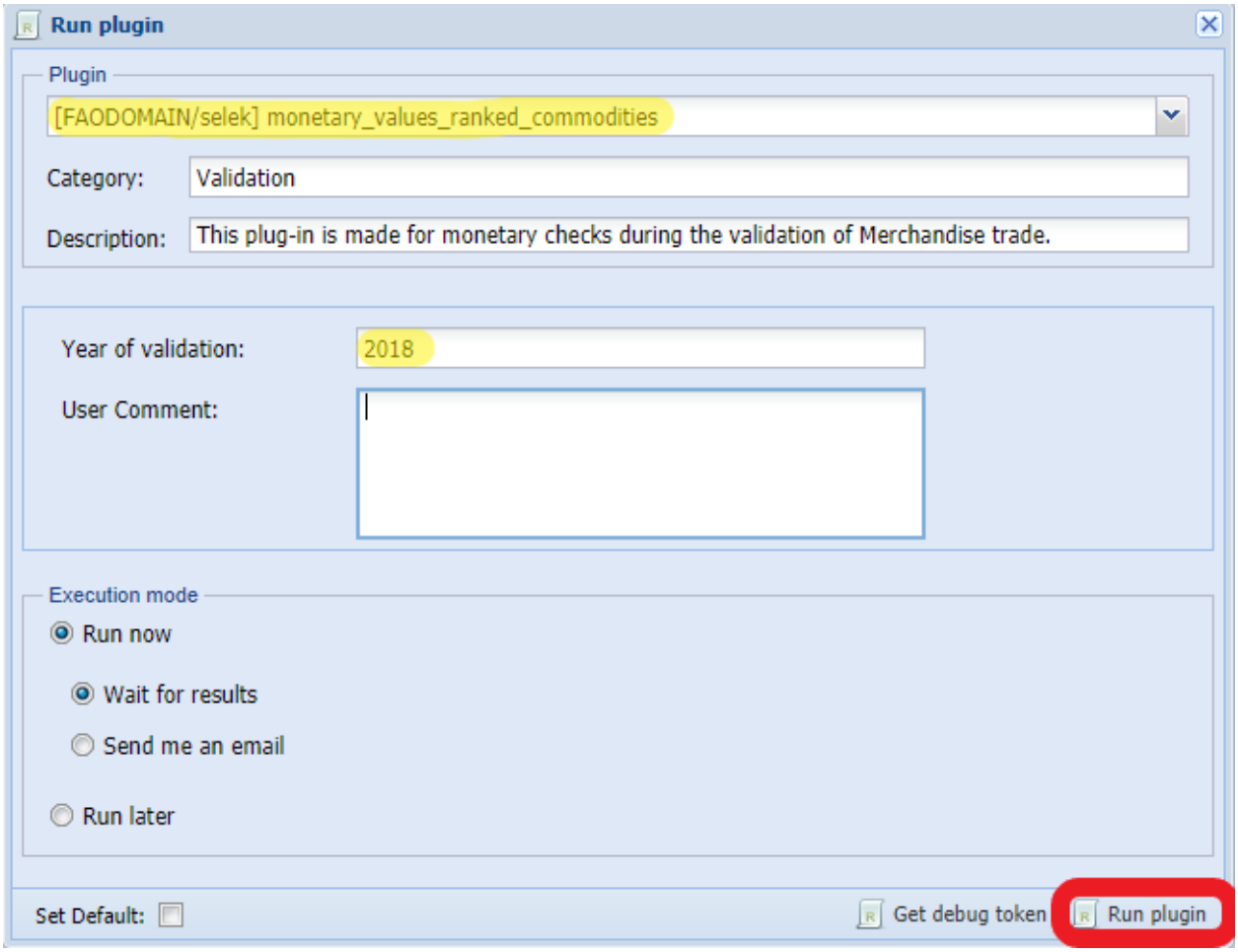
- yellow cells: values less than 50% of their 5-year average

- red cells: missing or deleted values considered to be incompatible with their time series ("bad TP")
- boldface: official data

geographi	geographi	measured	measured	measured	measured	2013	2014	2015	2016	2017	2018	5_year_average
626	Timor-Lest	'23161.02	Rice, Mille	5610	Import Qu	12,900	44,111	19,602	19,751	**93,712**	88,757	38,015
626	Timor-Lest	'17400	Ice and sn	5610	Import Qu	27,898	28,211	33,371	43,371	**42,024**	42,313	34,975
626	Timor-Lest	'23520	Refined su	5610	Import Qu	11,693	69,730	31,050	12,305	**971**	31,803	28,361
626	Timor-Lest	'24490	Other non	5610	Import Qu	18,901	15,487	16,644	26,194	**19,239**	18,485	19,293
626	Timor-Lest	'23110	Wheat and	5610	Import Qu	8,700	15,438	16,328	15,637	**16,953**	17,300	14,611
626	Timor-Lest	'2165	Palm oil	5610	Import Qu	8,991	10,381	11,663	10,156	**8,179**	11,883	9,874
626	Timor-Lest	'2351f	Raw cane	5610	Import Qu	4	19,652	6,600	6	**10,447**	0	9,176
626	Timor-Lest	'F1232	food prepa	5610	Import Qu	9,602	11,418	8,780	10,600	**4,466**	11,721	8,973
626	Timor-Lest	'21121	Meat of ch	5610	Import Qu	5,614	4,748	6,095	7,645	**11,951**	9,254	7,210
626	Timor-Lest	'24310.01	Beer of Ba	5610	Import Qu	4,906	6,054	7,813	7,115	**6,893**	1,025	6,556
626	Timor-Lest	'F0022	pastry	5610	Import Qu	2,839	3,764	3,397	3,985	**4,221**	4,298	3,641
626	Timor-Lest	'01919.96	Other fora	5610	Import Qu	3,362	2,711	2	0		65	3,036
626	Timor-Lest	'01290.90	Other vege	5610	Import Qu	5,000	23	11	3	**21**	21	2,511
626	Timor-Lest	'22230.03	Buttermilk	5610	Import Qu	2,659	3,100	2,491	1,791	**1,759**	1,804	2,360
626	Timor-Lest	'39170.01	Food Wast	5610	Import Qu	7	847	3,854	4,108	**2,156**	4,299	2,195
626	Timor-Lest	'23710	Uncooked	5610	Import Qu	0	691	1,112	423	**6,484**	202	1,742
626	Timor-Lest	'23161.03	Rice, Broke	5610	Import Qu	9		1,500	4,974	**467**	1,300	1,738
626	Timor-Lest	'01510	Potatoes	5610	Import Qu	1,133	1,927	2,539	1,553	**580**	2,134	1,546
626	Timor-Lest	'23670.01	Sugar Con	5610	Import Qu	1,842	1,665	1,372	1,335	**1,203**	1,060	1,483
626	Timor-Lest	'0231	Hen eggs i	5610	Import Qu	938	1,079	1,494	1,618	**594**	1,177	1,144
626	Timor-Lest	'23912.02	Coffee Ext	5610	Import Qu	1,846	1,576	1,008	794	**110**	702	1,135
626	Timor-Lest	'21184.02	Sausages a	5610	Import Qu	1,277	1,167	1,231	959	**504**	1,544	1,028
626	Timor-Lest	'2166	Coconut o	5610	Import Qu	1			0	**2,975**		992
626	Timor-Lest	'25020.01	Cigarettes	5610	Import Qu	746	872	1,061	1,306	**964**	1,426	990
626	Timor-Lest	'21439.9	Juice of fru	5610	Import Qu	328	856	973	783	**1,733**	1,146	934

4.1.4 `monetary_values_ranked_commodities` plugin

This plugin is used for the validation of total agricultural/merchandise trade data. It calculates and ranks 5-year averages of the monetary values of all items and facilitate the comparison of "Total Merchandise Trade" (CPC F1881) and "Agricult.Products,Total" (CPC F1882) by adding the "SHARE of Agriculture to Merchandise" ($F1881/F1882 \times 100$).



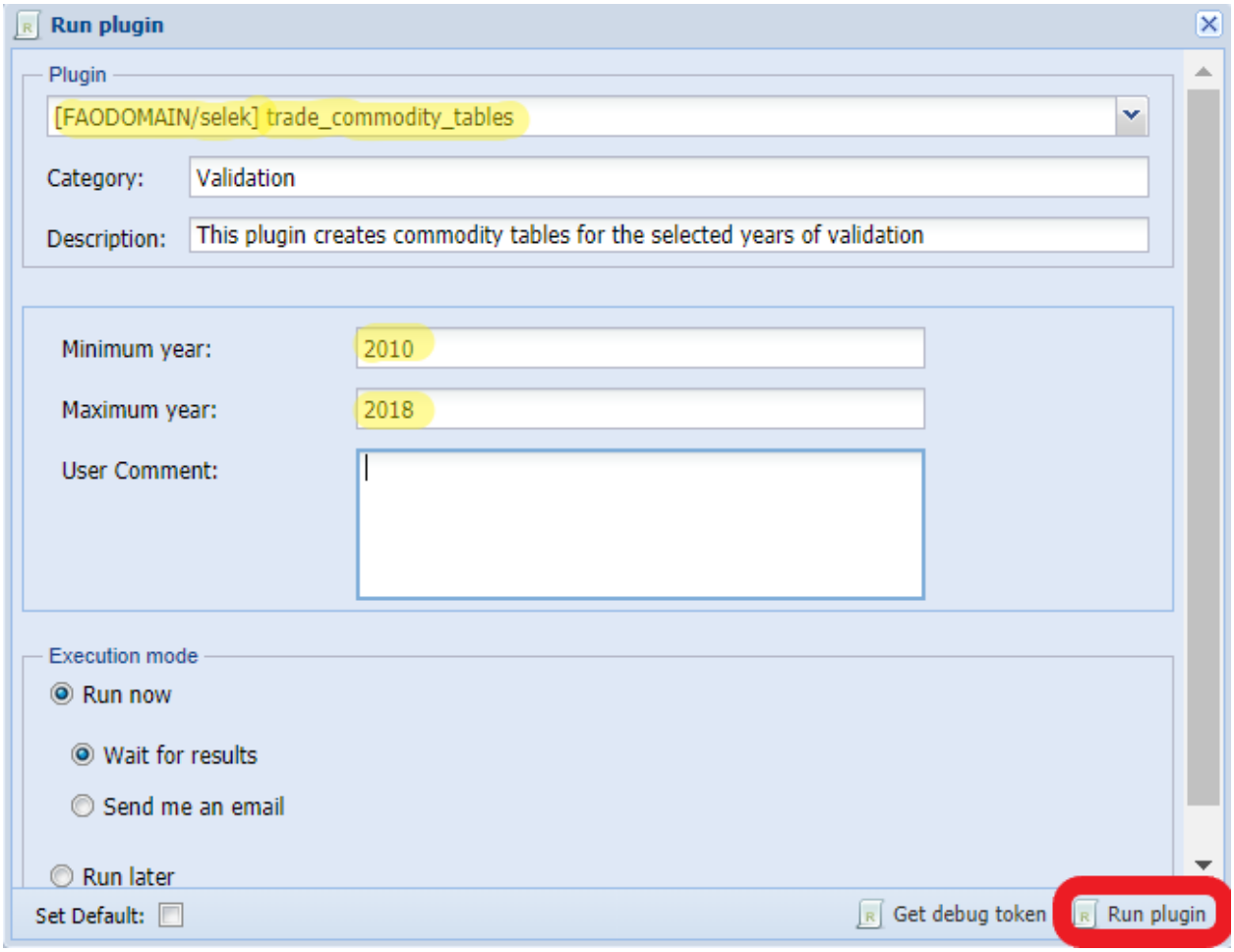
When the plugin completes successfully, it sends an email with subject "Monetary values ranked commodities" with an Excel that indicates:

- yellow cells: values less than 50% of their 5-year average
- red cells: *SHARE of Agriculture to Merchandise* is greater than 50%
- pink cells: missing value
- boldface: official data

geogr	geographicAre	measured	measuredItemCPC_description	measu	measuredElementTrad	2013	2014	2015	2016	2017	2018	5_year_average
624	Guinea-Bissau	'F1881	Total Merchandise Trade	5622	Import Value [1000 $]	240,000	214,000	207,000	230,000	280,000	330,383	234,200
624	Guinea-Bissau	'F1882	Agricult.Products,Total	5622	Import Value [1000 $]	144,706	139,869	117,325	136,983	157,213	150,116	139,219
624		'NA	SHARE of Agriculture to Merch	5622		60	65	57	60	56	45	0
624	Guinea-Bissau	'23161.03	Rice, Broken	5622	Import Value [1000 $]	18,184	19,259	19,963	21,193	16,526	19,760	19,025
624	Guinea-Bissau	'F1232	food preparations n.e.	5622	Import Value [1000 $]	12,950	18,539	15,121	14,228	13,357	12,631	14,839
624	Guinea-Bissau	'23161.02	Rice, Milled	5622	Import Value [1000 $]	11,080	6,676	5,667	6,387	37,937	26,428	13,549
624	Guinea-Bissau	'23999.01	Malt Extract	5622	Import Value [1000 $]	13,848	9,523	11,064	8,733	4,971	11,743	9,628
624	Guinea-Bissau	'2165	Palm oil	5622	Import Value [1000 $]	10,035	7,403	10,973	4,403	7,344	4,223	8,032
624	Guinea-Bissau	'24310.01	Beer of Barley, malted	5622	Import Value [1000 $]	7,364	10,082	6,228	6,536	7,226	7,026	7,487
624	Guinea-Bissau	'24212.02	Wine	5622	Import Value [1000 $]	6,408	7,459	5,515	6,213	6,527	7,503	6,424
624	Guinea-Bissau	'24490	Other non-alcoholic caloric bev	5622	Import Value [1000 $]	4,790	5,961	3,659	8,721	7,434	6,946	6,113
624	Guinea-Bissau	'23110	Wheat and meslin flour	5622	Import Value [1000 $]	6,281	3,309	2,880	3,934	9,120	9,528	5,105
624	Guinea-Bissau	'22211	Whole milk powder	5622	Import Value [1000 $]	13,849	2,725	1,525	1,876	1,317	1,302	4,258
624	Guinea-Bissau	'25020.01	Cigarettes	5622	Import Value [1000 $]	3,454	5,079	1,837	2,951	2,645	2,527	3,193
624	Guinea-Bissau	'21121	Meat of chickens, fresh or chill	5622	Import Value [1000 $]	2,023	2,714	3,076	3,073	4,735	4,228	3,124
624	Guinea-Bissau	'23520	Refined sugar	5622	Import Value [1000 $]	4,200	2,128	1,087	4,815	1,660	2,542	2,778
624	Guinea-Bissau	'23710	Uncooked pasta, not stuffed or	5622	Import Value [1000 $]	1,027	1,203	1,055	7,602	2,145	3,255	2,607
624	Guinea-Bissau	'23999.02	Food Preparations of Flour, Me	5622	Import Value [1000 $]	2,522	2,143	2,733	2,564	2,914	1,855	2,575

4.1.5 `trade_commodity_tables` plugin

This plugin is designed to validate imports and exports nat the global level. It runs only for selected items, available in the `commodity_list_for_total_trade_tables` datatable.



The plugin sends two emails as output. The first one is sent to the analyst and has subject "World" and contains an Excel file with global trade for all the items: red cells indicate that there is no balance between imports and exports (+/- 5%).

Country G	Commodity name	Comm	Trade Dimension	2010	2011	2012	2013	2014	2015	2016
World	Rice	0113	Import_Quantity (t)	2,669,203	2,483,866	2,644,177	2,693,441	2,484,405	2,754,559	2,840,496
World	Rice	0113	Export_Quantity (t)	2,637,439	2,023,015	2,518,261	2,513,456	2,505,251	2,378,280	2,825,658
World	Rice	0113	Import - Export	31,764	460,851	125,916	179,985	-20,846	376,279	14,838
World	Rice	0113	[(Import/Export) - 1] in %	1	23	5	7	-1	16	1
World	Rice	0113	[Import_growth] in %		-7	6	2	-8	11	3
World	Rice	0113	[Export_growth] in %		-23	24	0	0	-5	19
World	Rice	0113	Status	1	0	1	0	1	0	1
World	Rice	0113	Import Value [1000 $]	1,130,940	1,083,166	1,159,110	1,253,957	1,302,219	1,127,926	1,081,567
World	Rice	0113	Export Value [1000 $]	1,030,289	797,217	1,020,067	1,074,274	1,062,074	803,511	912,181
World	Sorghum	0114	Import_Quantity (t)	6,782,819	6,805,501	6,100,299	6,915,139	8,410,842	13,464,345	9,387,032
World	Sorghum	0114	Export_Quantity (t)	6,318,313	6,274,657	6,516,222	6,403,415	9,412,567	13,233,327	8,706,781
World	Sorghum	0114	Import - Export	464,506	530,844	-415,923	511,724	-1,001,725	231,018	680,251
World	Sorghum	0114	[(Import/Export) - 1] in %	7	8	-6	8	-11	2	8
World	Sorghum	0114	[Import_growth] in %		0	-10	13	22	60	-30
World	Sorghum	0114	[Export_growth] in %		-1	4	-2	47	41	-34
World	Sorghum	0114	Status	0	0	0	0	0	1	0
World	Sorghum	0114	Import Value [1000 $]	1,531,661	2,051,660	1,759,497	2,158,294	2,465,444	3,759,855	2,154,056
World	Sorghum	0114	Export Value [1000 $]	1,162,060	1,717,325	1,596,977	1,603,257	2,231,901	3,117,011	1,787,115

The second email is sent to the officer and contains the single commodity tables for each item. These Excel files have three sheets, which are "World_summary", "Country_details", and "Regions", with details of global, country and region, respectively.

Country Group	Country	Commodity	Commodity	Trade Dimension	2010	Status	Metho	2011	Status	Metho	2012	Status	Metho
Asia	Afghanistan	Barley	0115	Import_Quantity (t)	3,109		-	33,514	T	p	3,607	T	p
Asia	Afghanistan	Barley	0115	Import Value [1000 $]	371		-	7,473	T	p	777	T	p
Asia	Afghanistan	Barley	0115	Import UV [$/t]	119	I	i	223	I	i	215	I	i
Asia	Afghanistan	Barley	0115	Export_Quantity (t)	0			0			0		
Asia	Afghanistan	Barley	0115	Export Value [1000 $]	0	M	u	0	M	u	0	M	u
Asia	Afghanistan	Barley	0115	Export UV [$/t]	0			0			0		
Europe	Albania	Barley	0115	Import_Quantity (t)	2,141		-	2,181		-	1,282		-
Europe	Albania	Barley	0115	Import Value [1000 $]	453		-	599		-	451		-
Europe	Albania	Barley	0115	Import UV [$/t]	212	I	i	275	I	i	352	I	i
Europe	Albania	Barley	0115	Export_Quantity (t)	0	E	t	0	E	t	0	E	t
Europe	Albania	Barley	0115	Export Value [1000 $]	0	E	t	0	E	t	0	E	t
Europe	Albania	Barley	0115	Export UV [$/t]	0	M	u	0	M	u	0	M	u
Africa	Algeria	Barley	0115	Import_Quantity (t)	160,000	E	f	385,153		-	401,781		-
Africa	Algeria	Barley	0115	Import Value [1000 $]	41,000	E	f	115,620		-	113,895		-
Africa	Algeria	Barley	0115	Import UV [$/t]	256	I	i	300	I	i	283	I	i

4.2 Interactive tool for data validation (Shiny tool)

The interactive tool for trade data validation allows to query the total and bilateral trade datasets, and eventually apply corrections to the bilateral flows. Any modification to the bilateral flows is immediately visible in the tool. Once the validation is completed, modified data can be sent back to SWS, where its can be saved to the dataset.

It can be used at the following URL:

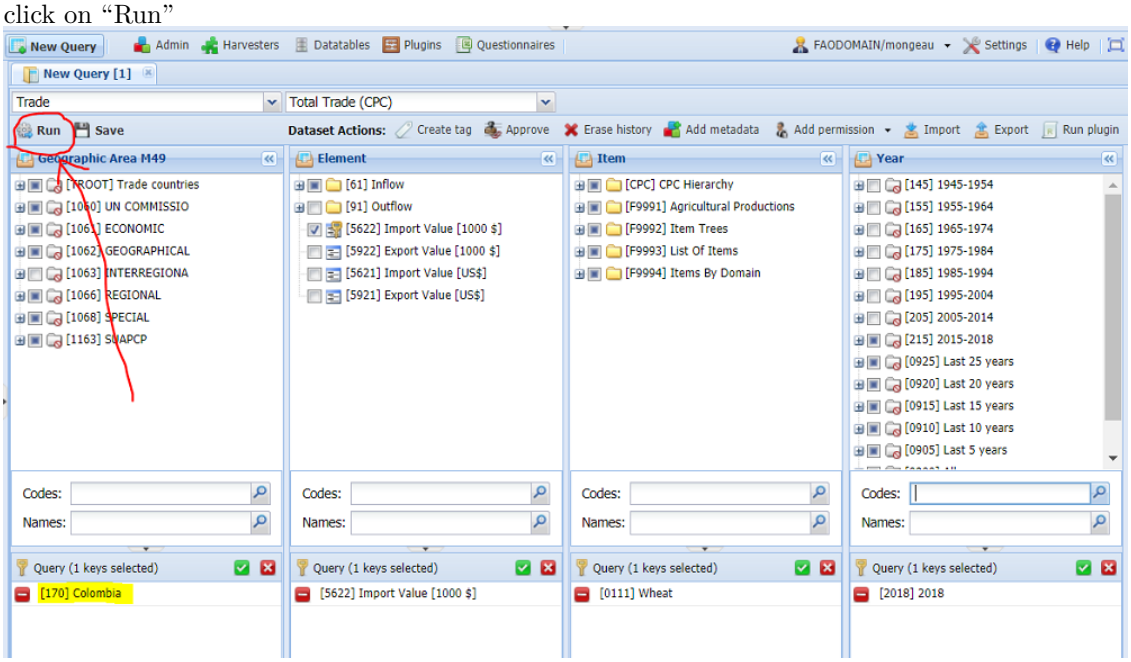
http://hqlprsws1.hq.un.fao.org:3838/TRADEvalidation_test/

4.2.1 Authentication

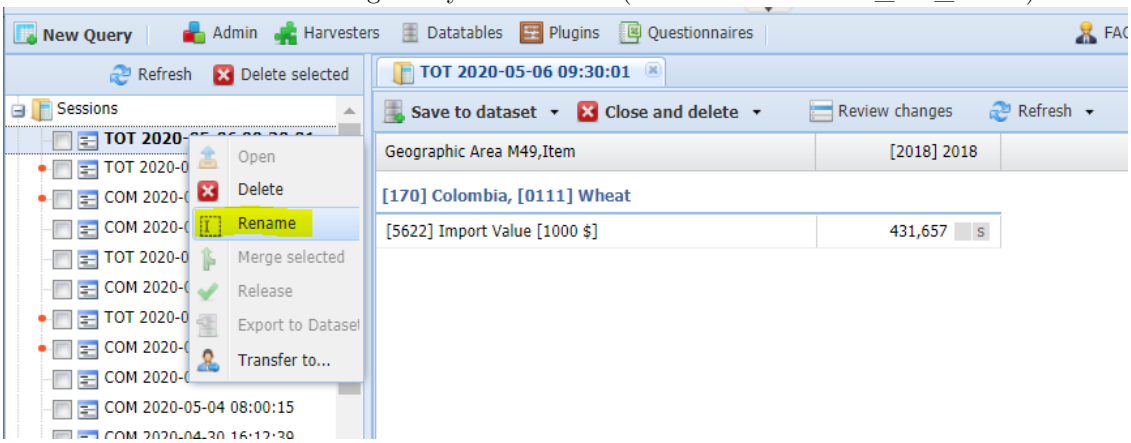
To use the tool, you will need to authenticate to the SWS. The steps for doing so are (all mentioned datasets are in the "Trade" domain):

1. create a session in the "Total Trade (CPC)" dataset
    1. in the country field, insert the country name/code of the reporter that you will validate

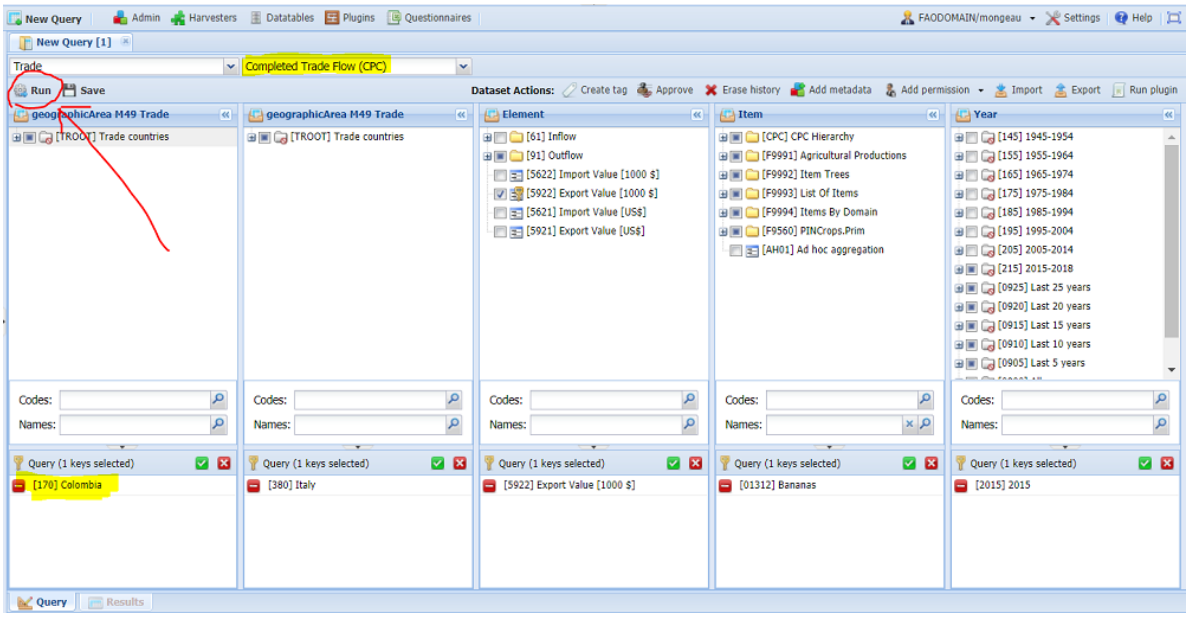
2. in the remaining fields you can put whatever you like
3. click on “Run”



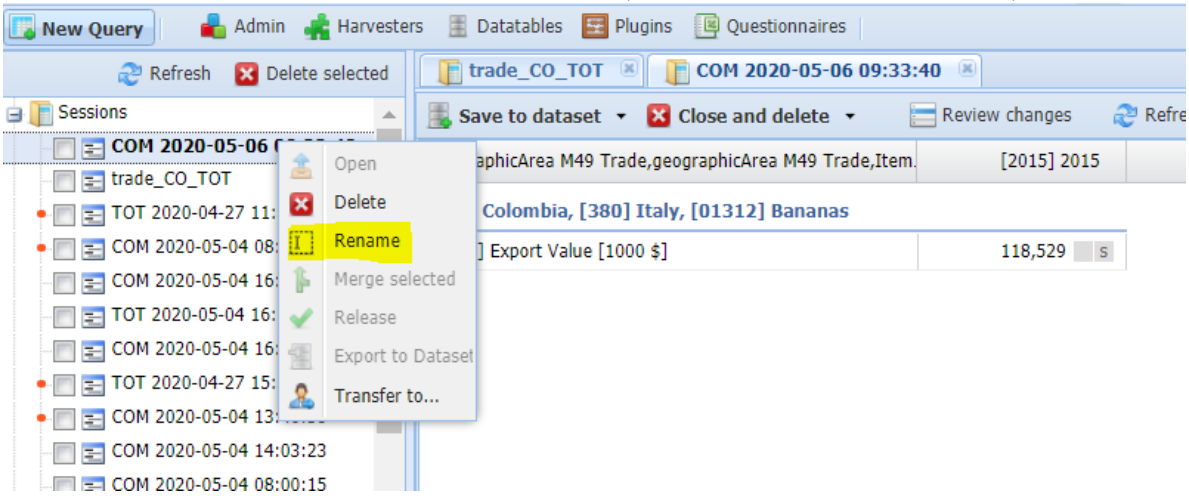
4. rename the session to something that you remember (here it will be “trade_CO_TOT”)



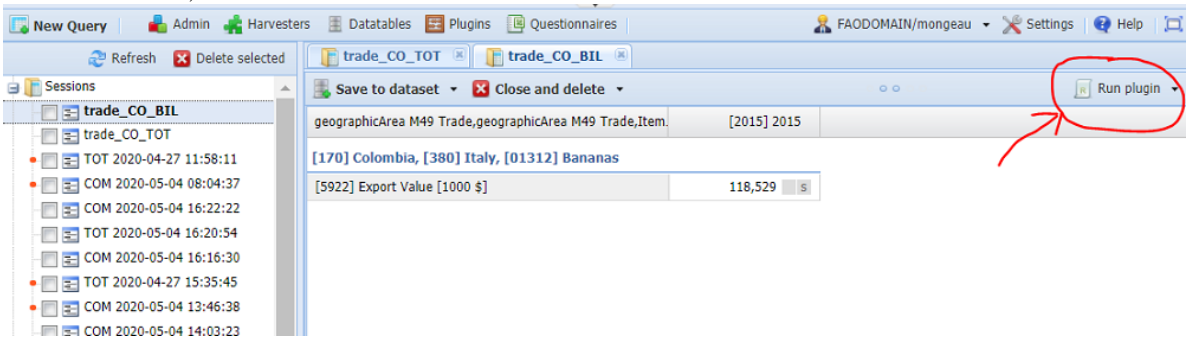
2. create a session in the “Completed Trade Flow (CPC)” dataset
3. in the first country field, insert the country name/code of the reporter that you will validate
4. select the other dimensions
5. click on “Run”



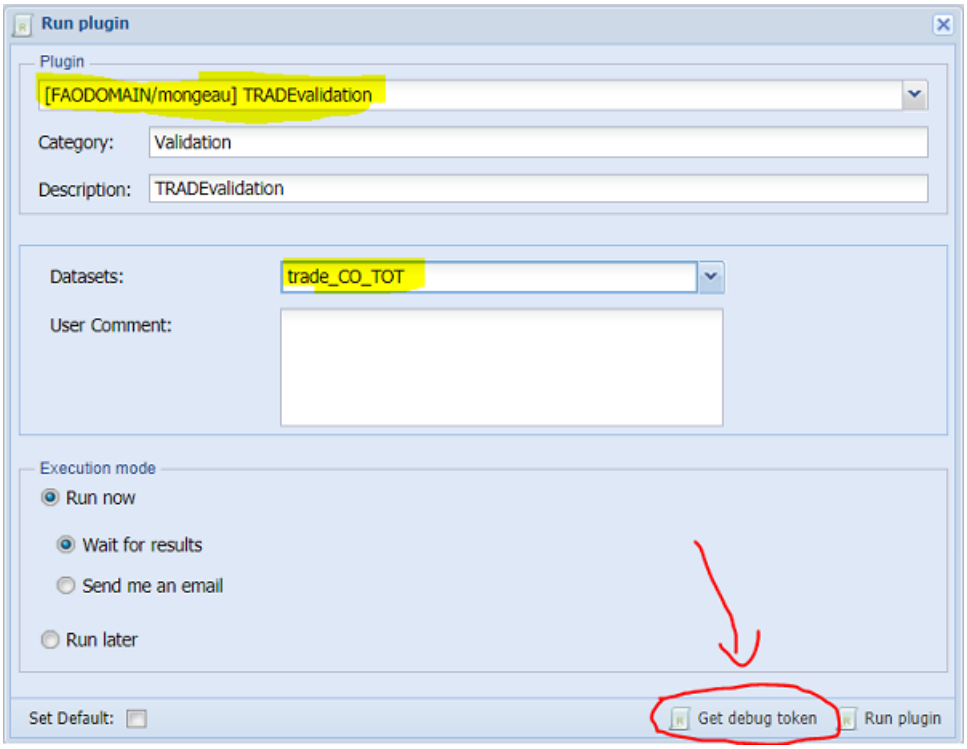
6. rename the session to something that you remember (here it will be "trade_CO_BIL")



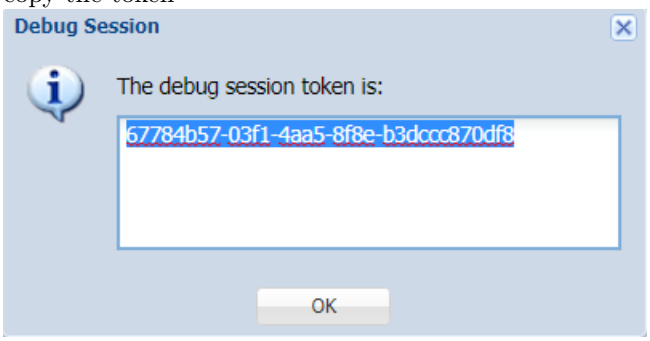
7. click on "Run plugin" in the session on the complete trade dataset (in the example above, in trade_CO_BIL)



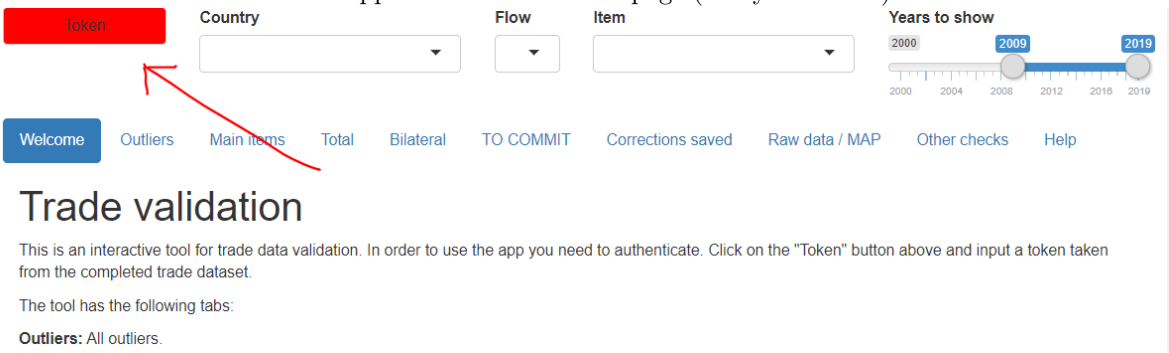
8. select "[FAODOMAIN/mongeau] TRADEvalidation"
9. select the session you created in the "Total Trade (CPC)" dataset
10. click on "Get debug token"



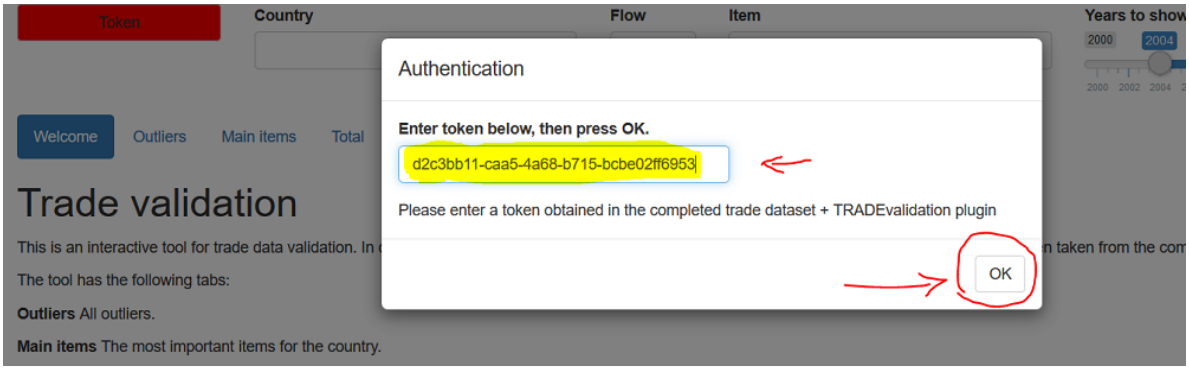
11. copy the token



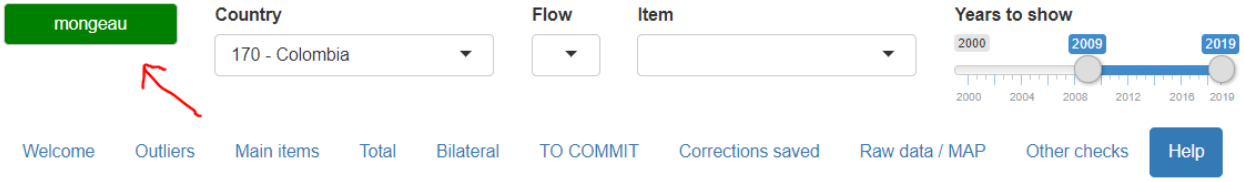
12. open http://hqlprsws1.hq.un.fao.org:3838/TRADEvalidation_test/ (well, this.)
13. click on the red button on the upper left corner on the page (it says “Token”)



14. paste the token you copied before and click “OK”



After you do these steps, if everything goes fine, you will see your username in the button in the upper left corner (and it should become green).



You are now authenticated, and can start/continue your validation.

Please note that the country you selected in your session (i.e., the country that you will validate), will appear in the "Country" menu. At this point you can start your validation by

1. selecting the "Outliers" tab
2. selecting the "Main items" tab

If you want to have a look at a specific flow/item you can choose this combination by filling the "Flow" and "Item" menus in the upper side of the page. Also, you can modify the start and end years in the "Years to show" sliders, but once you do so avoid changing them in middle of a validation of a flow/item combination (to see why, read the final notes at the bottom of the page).

Below are described the contents of the tool's tabs.

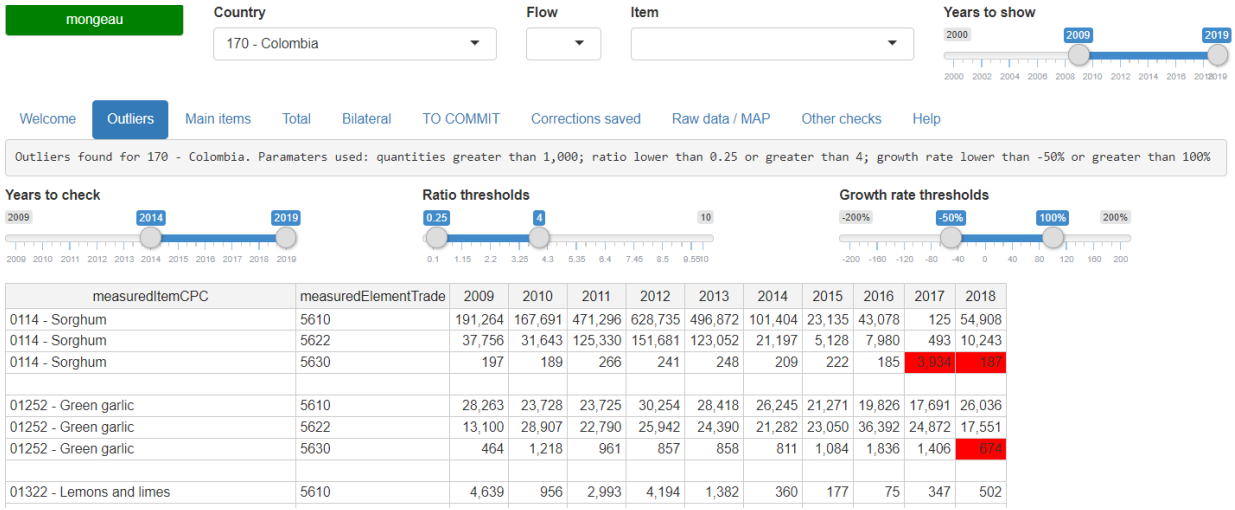
4.2.2 Tool tabs

4.2.2.1 Welcome

This is the first tab you see when you arrive here. It contains a small reminder on what you find and a "recommended workflow". If you don't feel to read all this help document again, just go there to have a brief reminder of this tool's features.

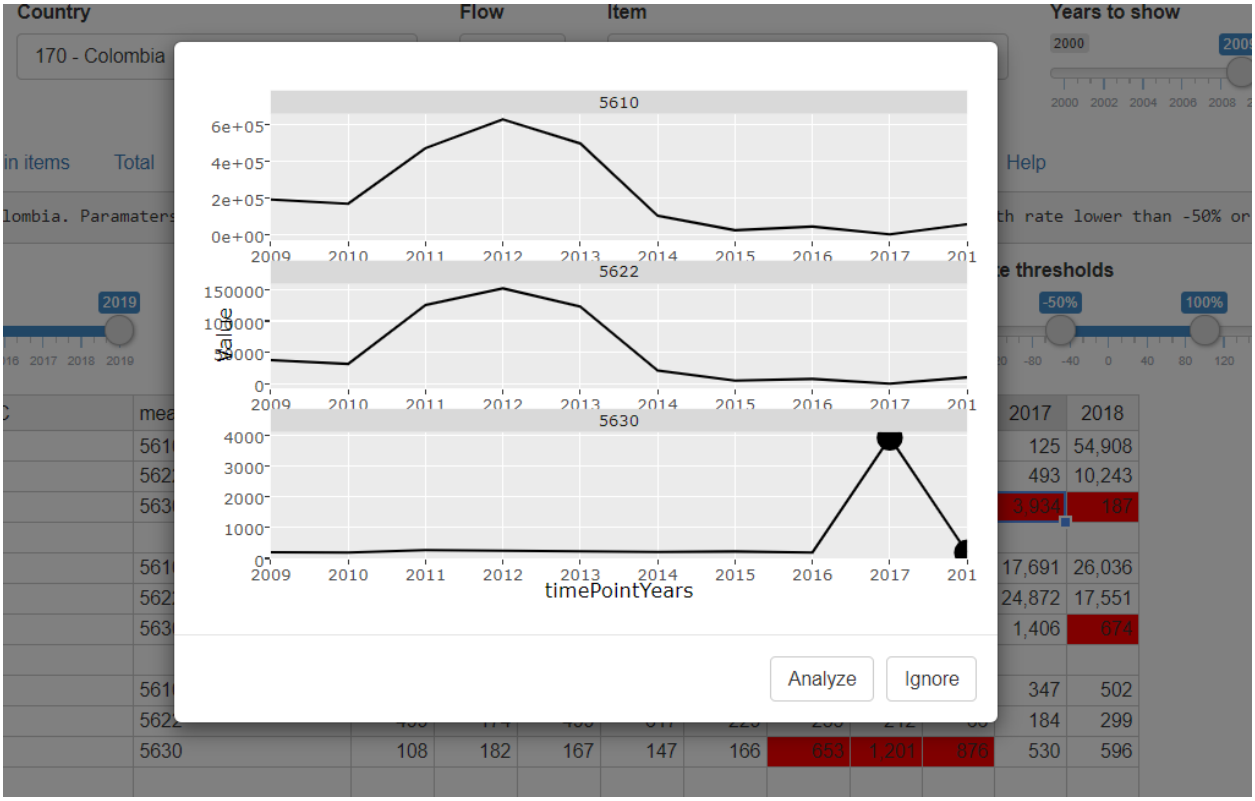
4.2.2.2 Outliers

When you select this tab, total imports and exports for **all** items will be downloaded and the outliers routine will be applied to them. A table with the flow/items that display at least one outlier will be displayed. The outlier(s) will be higlighted in red.



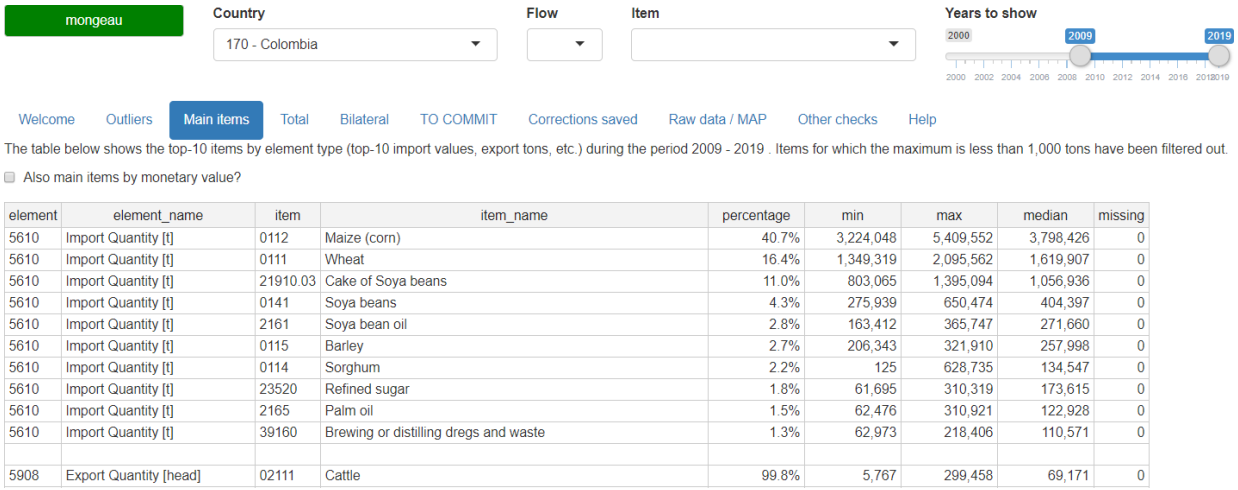
measuredItemCPC	measuredElementTrade	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018
0114 - Sorghum	5610	191,264	167,691	471,296	628,735	496,872	101,404	23,135	43,078	125	54,908
0114 - Sorghum	5622	37,756	31,643	125,330	151,681	123,052	21,197	5,128	7,980	493	10,243
0114 - Sorghum	5630	197	189	266	241	248	209	222	185	3,934	187
01252 - Green garlic	5610	28,263	23,728	23,725	30,254	28,418	26,245	21,271	19,826	17,691	26,036
01252 - Green garlic	5622	13,100	28,907	22,790	25,942	24,390	21,282	23,050	36,392	24,872	17,551
01252 - Green garlic	5630	464	1,218	961	857	858	811	1,084	1,836	1,406	674
01322 - Lemons and limes	5610	4,639	956	2,993	4,194	1,382	360	177	75	347	502

When you click on a cell, a popup window will display the value / quantity and utnit value of that flow/item, and the outlier will be shown with a black point. You can select to "Ignore" the item (this will simply close the popup), or "Analyze". If you select the latter, you will be redirected to the "Total" tab (see below).



4.2.2.3 Main items

Here you will see the top-10 items for the country, grouped by flow (imports and exports) and type of item (tons, heads, 1,000 heads).



element	element_name	item	item_name	percentage	min	max	median	missing
5610	Import Quantity [t]	0112	Maize (corn)	40.7%	3,224,048	5,409,552	3,798,426	0
5610	Import Quantity [t]	0111	Wheat	16.4%	1,349,319	2,095,562	1,619,907	0
5610	Import Quantity [t]	21910.03	Cake of Soya beans	11.0%	803,065	1,395,094	1,056,936	0
5610	Import Quantity [t]	0141	Soya beans	4.3%	275,939	650,474	404,397	0
5610	Import Quantity [t]	2161	Soya bean oil	2.8%	163,412	365,747	271,660	0
5610	Import Quantity [t]	0115	Barley	2.7%	206,343	321,910	257,998	0
5610	Import Quantity [t]	0114	Sorghum	2.2%	125	628,735	134,547	0
5610	Import Quantity [t]	23520	Refined sugar	1.8%	61,695	310,319	173,615	0
5610	Import Quantity [t]	2165	Palm oil	1.5%	62,476	310,921	122,928	0
5610	Import Quantity [t]	39160	Brewing or distilling dregs and waste	1.3%	62,973	218,406	110,571	0
5908	Export Quantity [head]	02111	Cattle	99.8%	5,767	299,458	69,171	0

Not necessarily the flow/items here contain outliers, but given that these flow/items represent the most important ones for the country, you will need to have a look in any case.

As for the "Outliers" tab, when you select a cell, you will be redirected to the "Total" tab (in this case, no popup window will appear, you will be redirected right away).

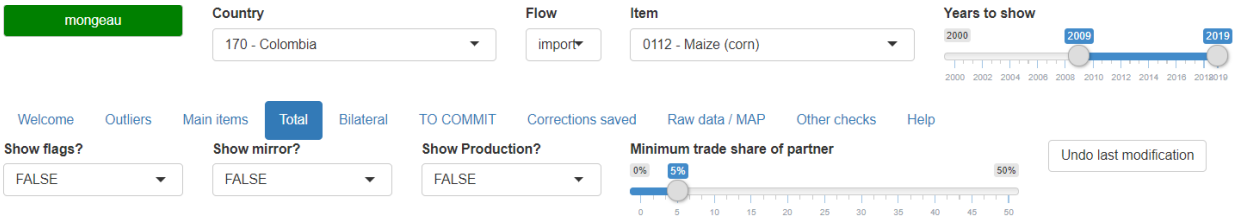
By default, only the main flow/items by quantity are shown. If you want that also the main items by value appear, you need to click on the box "Also main items by monetary value?".

4.2.2.4 Total

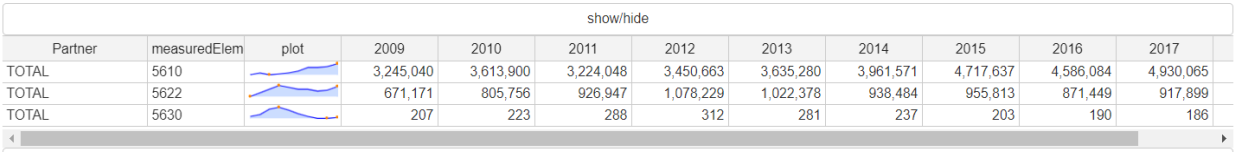
This tab is split into four components. The first one contains options and an "undo last modification" button. The second one is a table with values, quantities and unit values for the selected flow/item at the **total** trade level. The third one shows by default the plot of the series in the tab above (and eventually other information). The fourth one (under the plot), shows by default some bilateral flows of the selected total.

Below the components are shown separately (for spacing issues), but they are in the same page.

First component (plus general header):

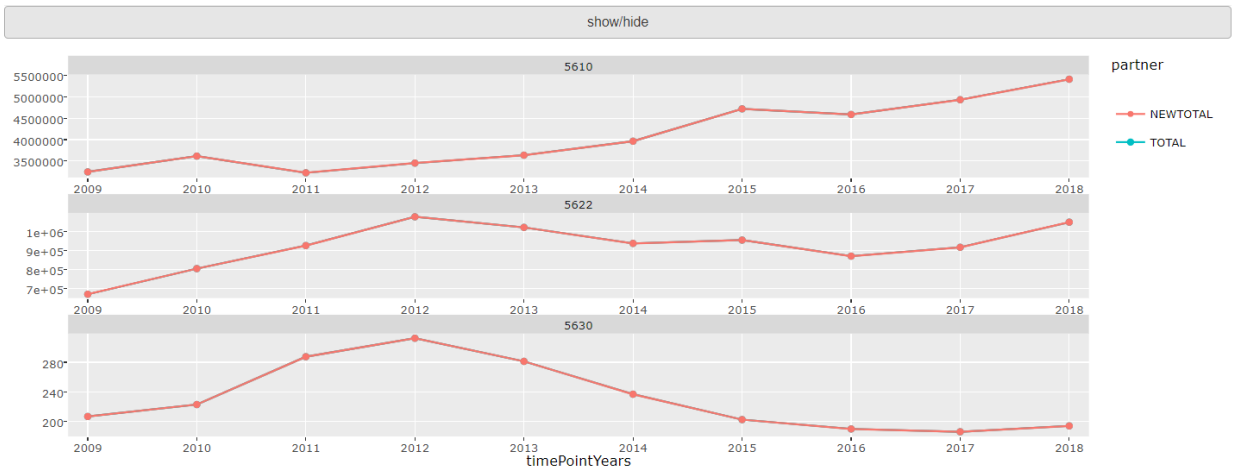


Second component:



Partner	measuredElem	plot	2009	2010	2011	2012	2013	2014	2015	2016	2017
TOTAL	5610		3,245,040	3,613,900	3,224,048	3,450,663	3,635,280	3,961,571	4,717,637	4,586,084	4,930,065
TOTAL	5622		671,171	805,756	926,947	1,078,229	1,022,378	938,484	955,813	871,449	917,899
TOTAL	5630		207	223	288	312	281	237	203	190	186

Third component:



Fourth component:

show/hide

Partner	measuredElem	plot	2009	2010	2011	2012	2013	2014	2015	2016	2017
32 - Argentina	5610		925,203	2,208,855	1,812,038	2,481,334	2,060,398	101,824	11,036	17,250	12,360
32 - Argentina	5622		191,328	481,774	595,819	757,200	610,044	32,806	5,484	9,425	6,623
32 - Argentina	5630		207	218	329	305	296	322	497	546	536
76 - Brazil	5610		884,055	626,909	380,759	456,100	858,172	1,194	1,534	2,253	2,751
76 - Brazil	5622		178,543	155,428	122,202	148,097	211,273	4,427	8,278	13,407	17,746
76 - Brazil	5630		202	248	321	325	246	3,707	5,395	5,950	6,451
840 - United Stat...	5610		1,225,032	703,781	552,011	176,988	645,755	3,856,485	4,703,708	4,565,430	4,914,294
840 - United Stat...	5622		249,573	146,862	183,740	70,169	180,333	897,764	933,370	841,327	889,176
840 - United Stat...	5630		204	209	333	396	279	233	198	184	181

The options (available in the first component) are:

- "Show flags?" will show flags of the data if TRUE, hide them if FALSE (by default this is FALSE, so that you can focus on the data; if you need to understand if data are official, estimates, etc., set it to TRUE).
- "Show mirror?" will download all bilateral flows of the partners for the selected flow/item, and will aggregate them so that you can see what the sum of other countries say about that item for the selected country. By default this is FALSE, mainly because the data that needs to be downloaded is quite large, so it will slow down the app. Set it to TRUE if you are ready to wait.
- "Show Production?" will show production data taken from the "SUA unbalanced" dataset. This is probably more useful for exports, as countries usually export less than what they produce (actually, you should consider also imports and eventualy stocks, so that X <= P + M - S (exports are less or equal availability), but in any case having a look also to production data can be useful). By default this is FALSE, if you set it to TRUE the time required to get the data is not that much.
- "Minimum trade share of partner" is a number X between 0% and 50% (by default 20%) that will filter out from the bilateral table countries that never accounted for X% in the total flow for the selected item of the partner. For instance, if you set this to 30%, you will see all partners in the table at the bottom of the page that in at least one year accounted for at least 30%. This is useful so that you have a look either to main partners (those that usually account a lot for the reporter) or countries that because an outlier show a high percentage over the partner's flow. You will notice that even if the default is 20%, sometimes it scales automatically to a lower number: this happens because no partner accounted for 20%, so the maximum share will be selected automatically.

There is an "Undo last modification" button, but what it does will be explained later.

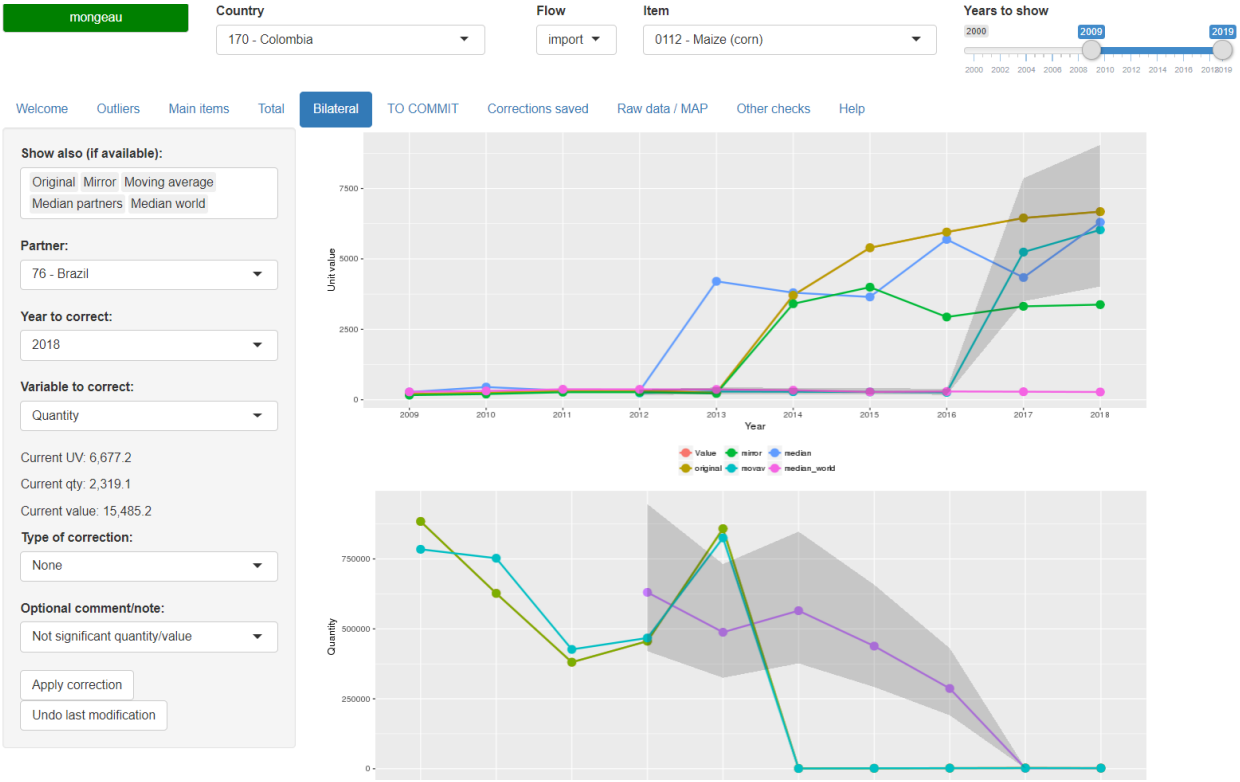
The first table contains data for the total flow for the selected flow/item/years: values, quantities, and unit values. When you do corrections in the "Bilateral" tab, this table will be updated with corrected data.

The plots under the total table, shows the data in that table. It will also show mirror and/or production data if these options are set to TRUE. Moreover, if you click on one cell of the table that contains bilateral flows (under the plots), you will see that partner series appear also in the plots.

In the bottom table, you see bilateral data for the selected flow/item. As explained above, by default you will see only the "main" partners (as per "Minimum trade share of partner" option). If you really need to see **all** partners you can set the "Minimum trade share of partner" to 0 (zero). Be aware that if you do so, the browser will be slowed down a bit. In any case, if you click on a cell, if it contains data, the series for the partner you selected will also be shown in the plots.

4.2.2.5 Bilateral

In this tab you can see bilateral flows, with some useful information, and can correct them, if required.



Select a partner in the "Partner" menu. Note that: 1) only available partners for the selected flow/item are shown, so you don't need to figure out what countries to choose; 2) the partner shuould be automatically populated if you select it in the bilaterals that appear in the "Total" tab.

If there are years that can be corrected (corrections can be done just after 2014, included), then available years will be displayed in the "Year to correct" menu.

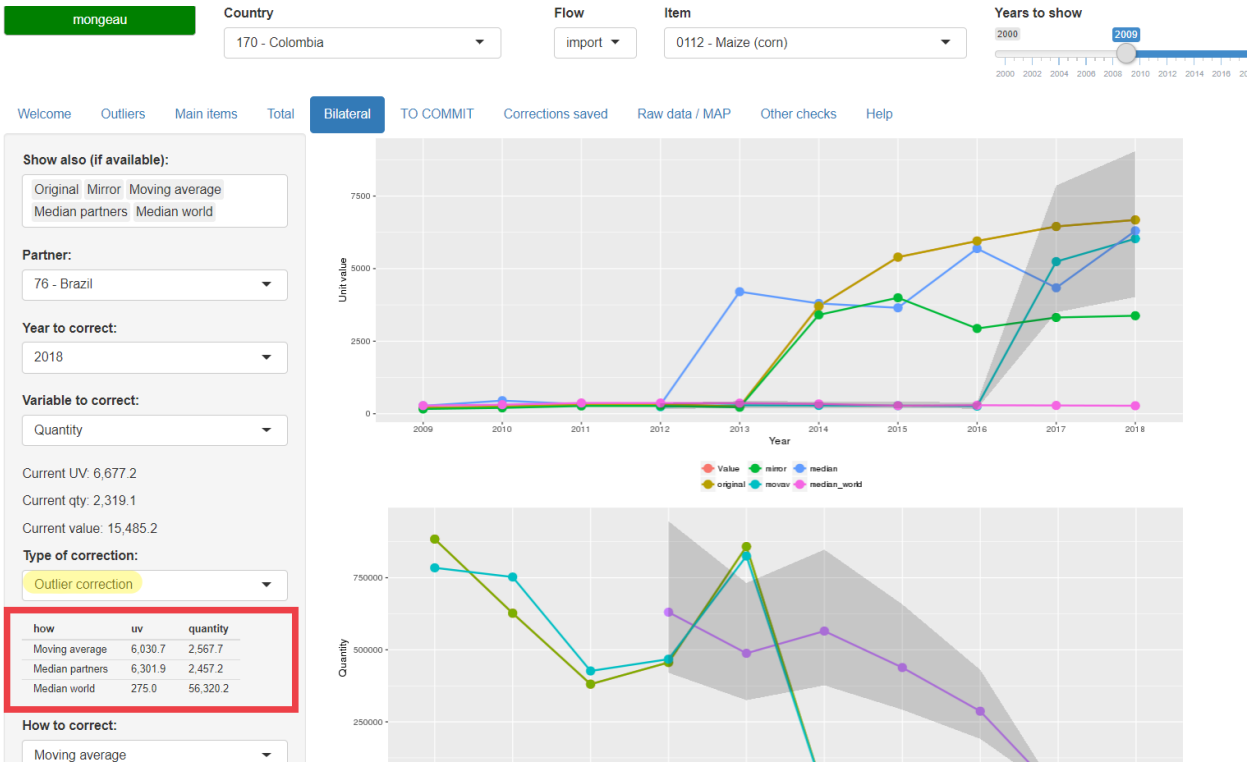
You can select what variable to correct in the "Variable to correct" menu (it can be either "Quantity" or "Value").

There are different types of corrections that can be applied in the "Type of correction" menu:

- **None**: This should be used to confirm that an outlier is OK and should not be corrected. It can seem weird to correct by not applying any correction, but this may be thought more as a signal for future analysts (or yourself) that even if the observation seems an outlier, it was confirmed not to be so and who will again find this figure will not lose any time by revalidating it (is they read the metadata).

- **Measurement factor**: This can be useful if the quantity or value may have been reported with a wrong measurement unit, by a factor of 10. If you choose this option, it will be displayed below if the ratio of the unit value to the median unit value is "near" a factor of 10, and will suggest which factor is the most likely. More technically, the ratio between the unit value and the median unit value is computed; if this ratio is greater than or less than (with a 10% approximation) $10^{(}round(log10(ratio)))$ then it means that the transaction was likely reported with a unit of measurement (of a factor of 10) issue.
- **Mirror flow**: Use the flow of the partner, if it exists (if it does not, you will get a feedback indicating that).
- **Outlier correction**: It is possible to use three different types of correction (in all cases, the values of these figures will be printed, if available, so that you can choose the most appropriate):
  - **Median partners**: median with respect to all partners, that is, the median for the reporter's trade (import or export specific) in the year of interest of the analysed commodity to all its partner is used.
  - **Median world**: median of the unit value of all reporters to all partners, that is, the median of all the reporters that traded (import or export specific) the analysed commodity in the year of interest is used.
  - **Moving average**: the three year-moving average (starting from the year previous to the one analysed) of the unit value of the trade flow of the commodity of the analysed reporter.
- **Publication/website**: when you find data on a publication or website.
- **Expert knowledge**: literature or otherwise justified figure, in other words, this is a figure that you set based on your own estimate.

When you select one type of correction, appropriate submenus will be shown if required, and there is some information on what would be the effect of applying that correction. For instance, in the example below, the "Outlier correction" was chosen and the unit values and quantities for the various options appear.



When you wan to actually see the effect of a correction both at the bilateral level and at the total level, click on "Apply correction". This will make the corrected data to be sent to the plots in this tab, to the bilateral table in the "Total" tab, and consequently to the table with the totals in the "Total" tab.

If you do not want that correction to be saved, click on the "Undo last modification" button. The button is replicated both in this tab and in the "Total" tab. When you do so, the status of the tables and plots will be brought back to the status they were just after you clicked on "Apply correction".

If you are happy with your correction, you can send it to the list of corrections that will be saved by clicking on "Confirm correction". Note that this will accumulate the correction in a pool of corrections that will be saven in the "TO COMMIT" tab. More on this later.

4.2.2.6 TO COMMIT

This tab contains a list of corrections that are ready to be saved. To do so, you need to click on "Send to SWS datasets". What this action will do is that all corrections will be stored in the "correcitons table", which is a dataset of already saved corrections, and will send corrected data to SWS.

You will need to click on "Send to SWS datasets" for each combination of flow/items that you corrected. If you do not do it, and select another flow, item or years, you will lose all your corrections.

4.2.2.7 Corrections saved

This tab contains a list of existing corrections.

You can remove corrections that are **already** saved, by clicking on "Delete selected correction". Again, this wil remove corrections that were saved (those shown in the "Corrections saved" tab), not the corrections that you have not committed yet (those shown in the "TO COMMIT" tab).

4.2.2.8 Raw data / MAP

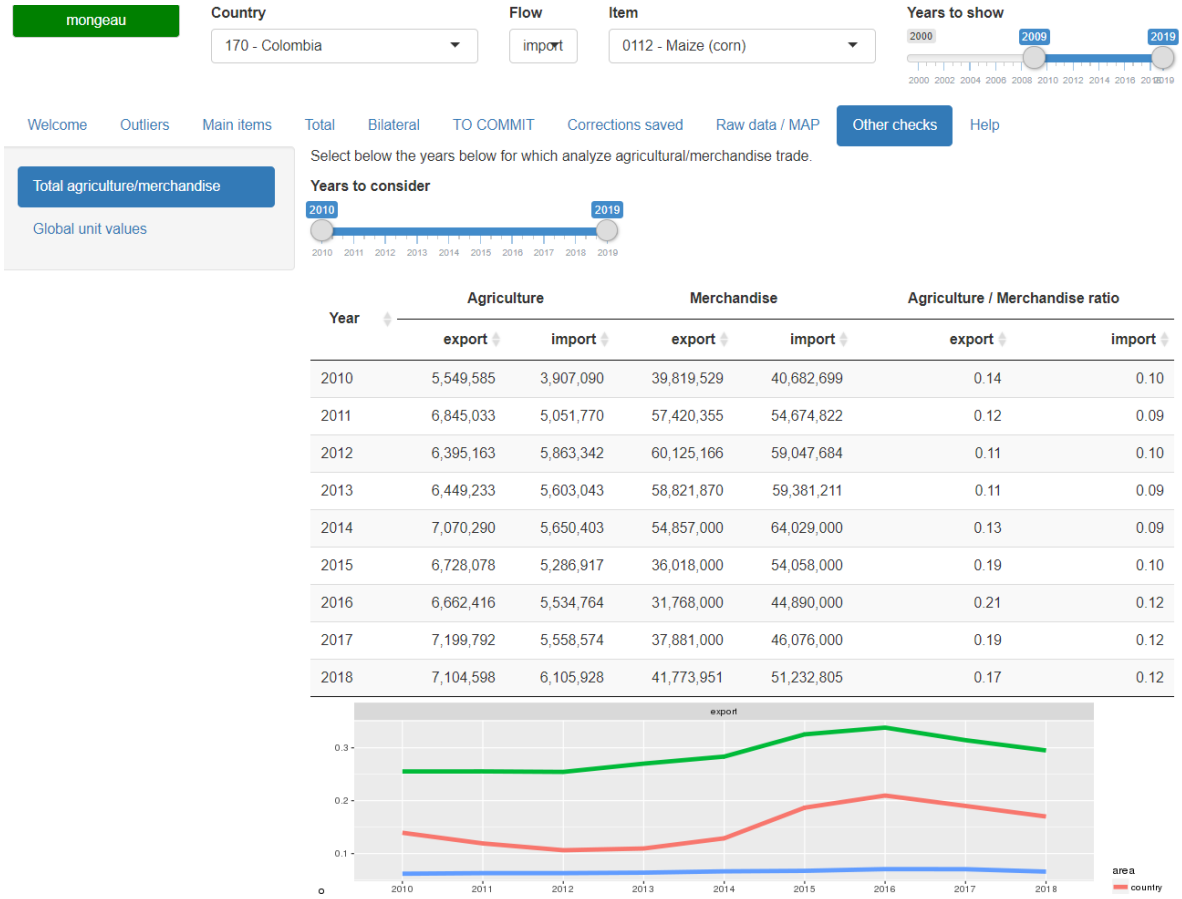
This was imported from the previous Shiny tool, but probably will require some re-writing because it seems not to have been used much after all (likely it was not made clear how this should be used).

Changes to the trademap need to be cleared by officers, so contact one of them if you need to change a map entry.

4.2.2.9 Other checks

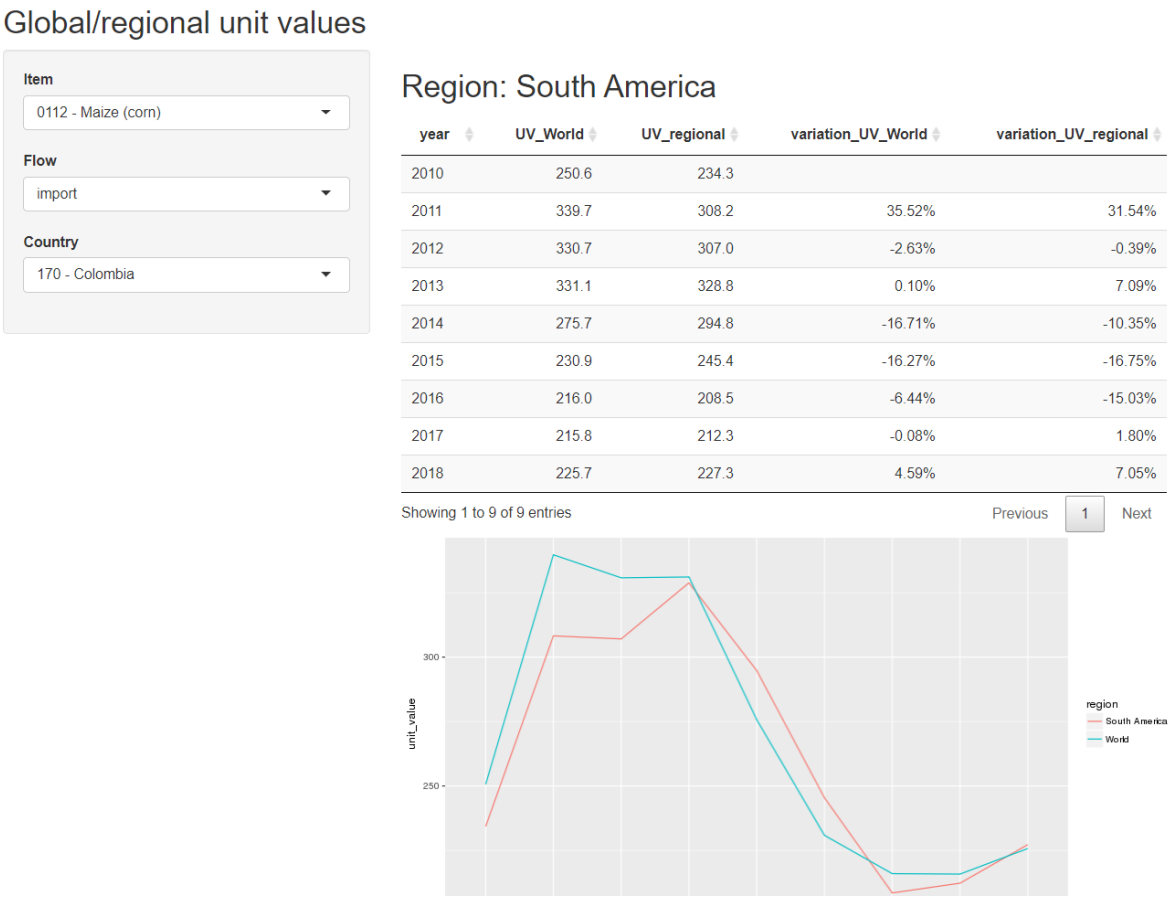
In this tab you find other checks useful for validation:

- Total agriculture/merchandise: comparison of total agricultural trade vs total merchandise trade. You will need to select the years that you want to analyse in the "Years to consider slider"



Year	Agriculture		Merchandise		Agriculture / Merchandise ratio	
	export	import	export	import	export	import
2010	5,549,585	3,907,090	39,819,529	40,682,699	0.14	0.10
2011	6,845,033	5,051,770	57,420,355	54,674,822	0.12	0.09
2012	6,395,163	5,863,342	60,125,166	59,047,684	0.11	0.10
2013	6,449,233	5,603,043	58,821,870	59,381,211	0.11	0.09
2014	7,070,290	5,650,403	54,857,000	64,029,000	0.13	0.09
2015	6,728,078	5,286,917	36,018,000	54,058,000	0.19	0.10
2016	6,662,416	5,534,764	31,768,000	44,890,000	0.21	0.12
2017	7,199,792	5,558,574	37,881,000	46,076,000	0.19	0.12
2018	7,104,598	6,105,928	41,773,951	51,232,805	0.17	0.12

- Global unit values: contains import and export unit values for all items and countries, along with their world and regional averages, as well as the growth rates.



4.2.2.10 Help

Contains a copy of this documentation (the part regarding the tool).

4.2.3 Final notes / recommendations

Please consider that the tool was designed so that you can validate one country at a time, and do so by going by flow and item. For instance, let's say you need to validate Colombia, then after you do the authentication steps, you go by flow/item combinations, e.g., you first validate all imports of wheat, then validate all export of bananas, then... you got it. "Validate all" means that, e.g., you check if the total imports of wheat is an outlier, then if it seems so you go and validate all partners that make sense to validate (usually the "main" partners) with the tools available in the tool (e.g., apply a measurement factor, a mirror flow, etc.) and once all your corrections to bilateral flows of imports of whats are done, save corrections and send the data to SWS. When you **end** with imports of wheat, you do the same for exports of bananas. When you finish with exports of bananas... you got it. When you finish analysis and (eventually) corrections for all flows and all items, you can conclude your validation by finally going to SWS and confirming that data is to be saved in the two sessions you created.

There is a technical reason why you should validate one flow/item at a time: each time you select another flow or item (and for that matter, even if you change the years), the tool will make a new query and you will lose all your corrections.

You can do any query you like within the tool, but for validation purposes, limit yourself to the flow/items that are shown in the "Outliers" tab and in the "Main items" tab, as they will show the flow/items that may be outliers and those that are most important for the country, respectively.

The tool allows you to do "multiple corrections" to correct ALL bilateral flows in one go. Please, use this very cautiously. If you find, e.g., that ALL partners's quantities are to be divided by 1,000, this may be better handled at the source of the data. It could be useful to liase with UNSD so that in a coordinated manner these flows get fixed. All agencies will benefit from this. Also, another reason why this needs to be used conservatively is that it will be difficult to "undo" all corrections (after they are saved) if they were made by mistake (it happens: for instance you selected 100 when indeed you wanted to select 1,000). Finally, it can be that you just spotted that the main partner needed to be fixed by a factor, then assumed also all other partners needed such adjustment, but later find that seen from the partner side all other corrections do not make sense.

4.2.4 Technical information

The tool is a Shiny app[18]. Its source code is publicly available at:

https://github.com/SWS-Methodology/TRADEvalidation

When a modification of the app is required, after the modification is done, the developer needs to update the version in the server. The path in the server where the application is stored is:

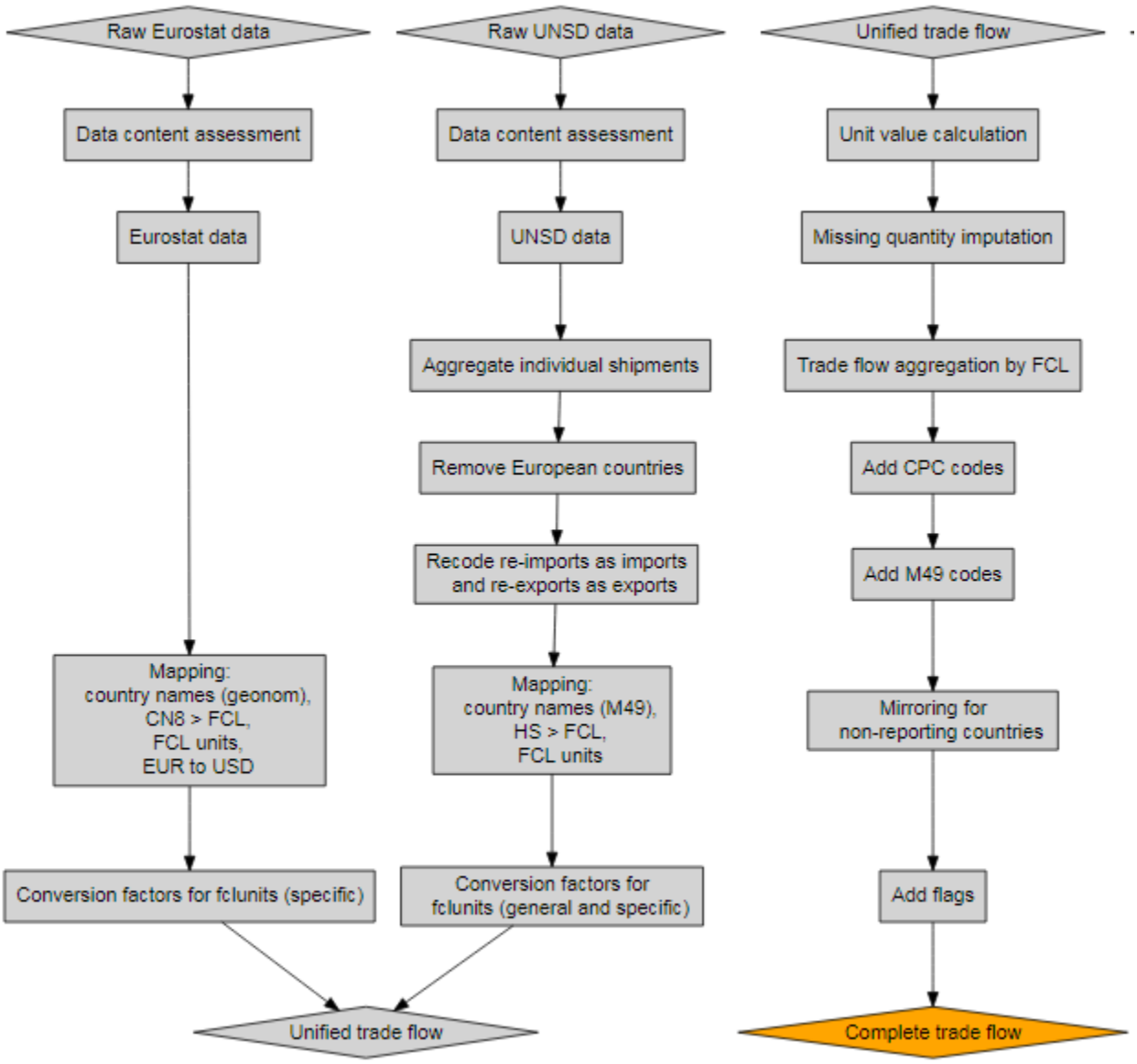
`/srv/shiny-server/TRADEvalidation_test`

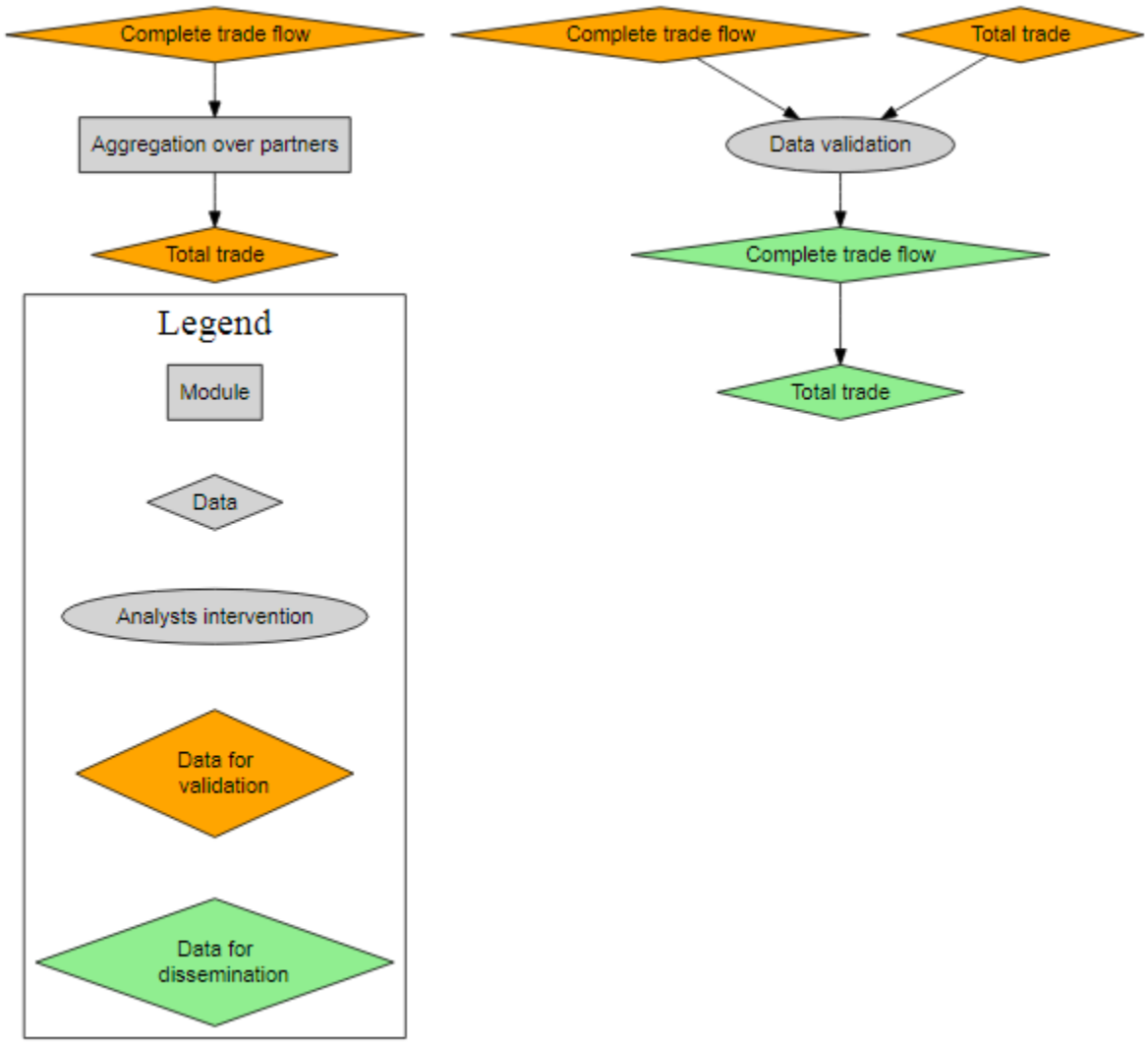
5 Workflow

5.1 Flow chart

The whole process is displayed is depicted graphically in the next flow chart.

[18] https://shiny.rstudio.com/

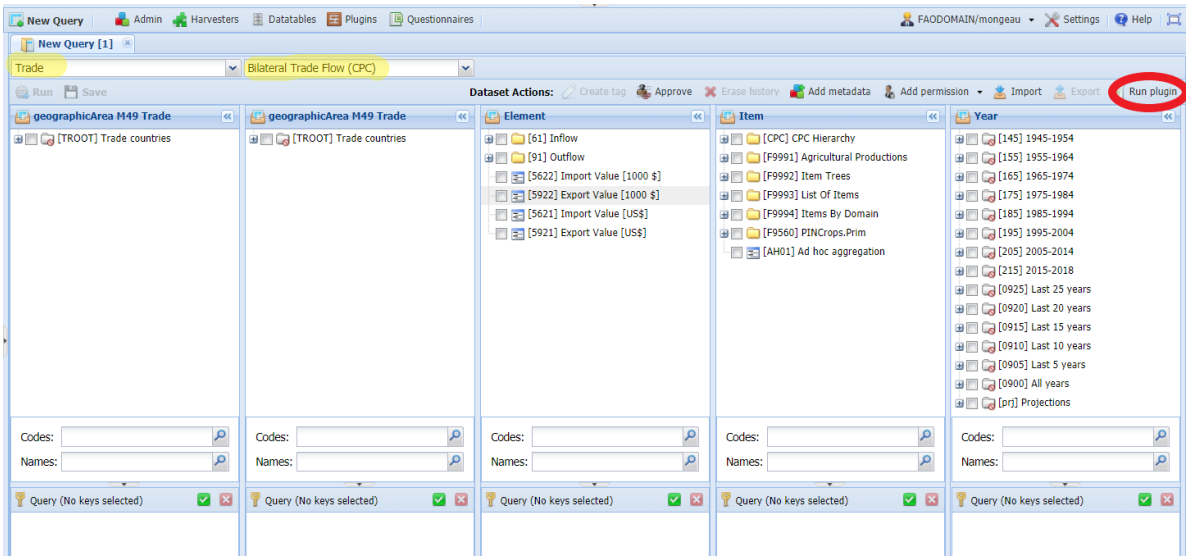

Raw Eurostat data
Data content assessment
Eurostat data
Mapping:
country names (geonom),
CN8 > FCL,
FCL units,
EUR to USD
Conversion factors for fclunits (specific)
Raw UNSD data
Data content assessment
UNSD data
Aggregate individual shipments
Remove European countries
Recode re-imports as imports
and re-exports as exports
Mapping:
country names (M49),
HS > FCL,
FCL units
Conversion factors for
fclunits (general and specific)
Unified trade flow
Unified trade flow
Unit value calculation
Missing quantity imputation
Trade flow aggregation by FCL
Add CPC codes
Add M49 codes
Mirroring for
non-reporting countries
Add flags
Complete trade flow



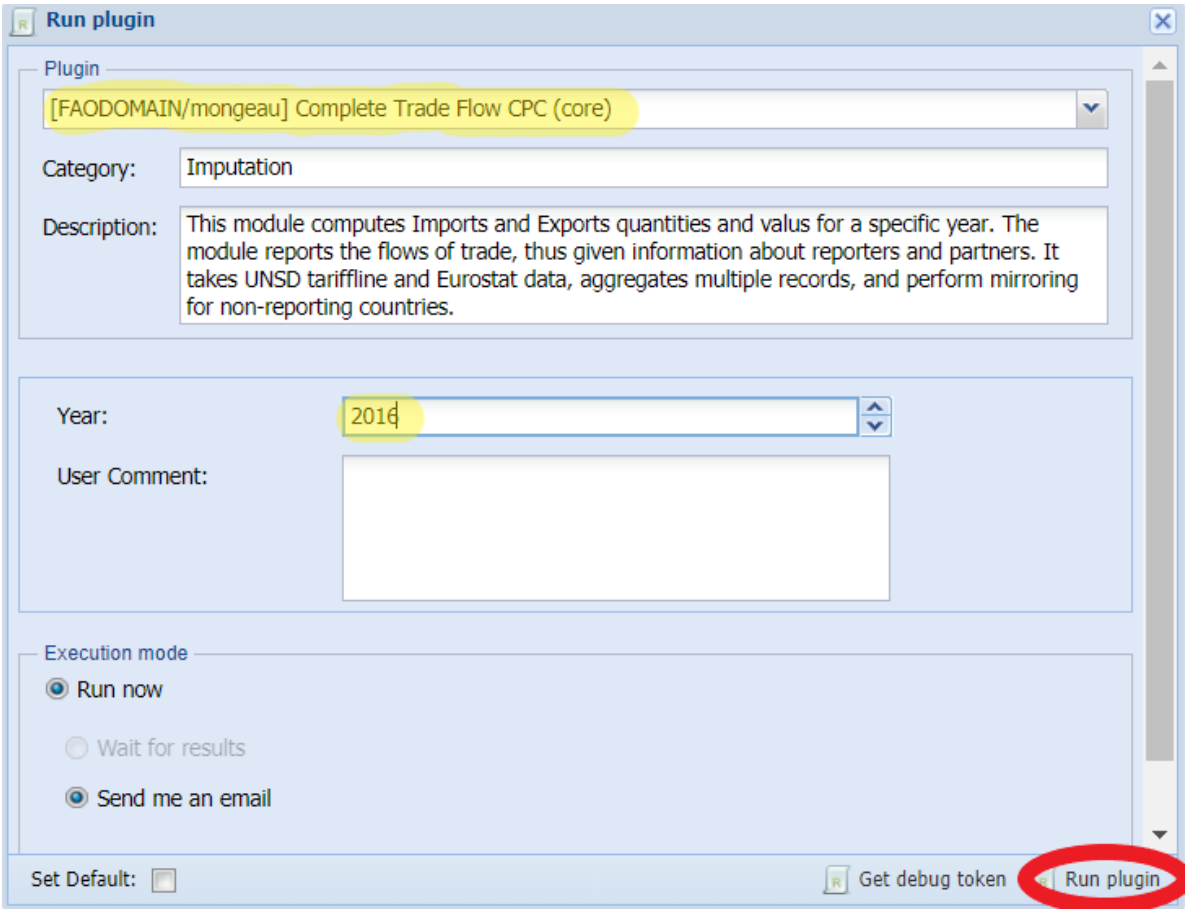
5.2 Plugin flow

The steps required to complete one full round of the trade process so to save data in the appropriate trade datasets are the following (validation will be covered in the "Data validation" section):

1. Open a query in the "Bilateral Trade Flow (CPC)" dataset of the "Trade" domain (dimensions should not be set, as this is a *core* plugin), and click on "Run plugin":

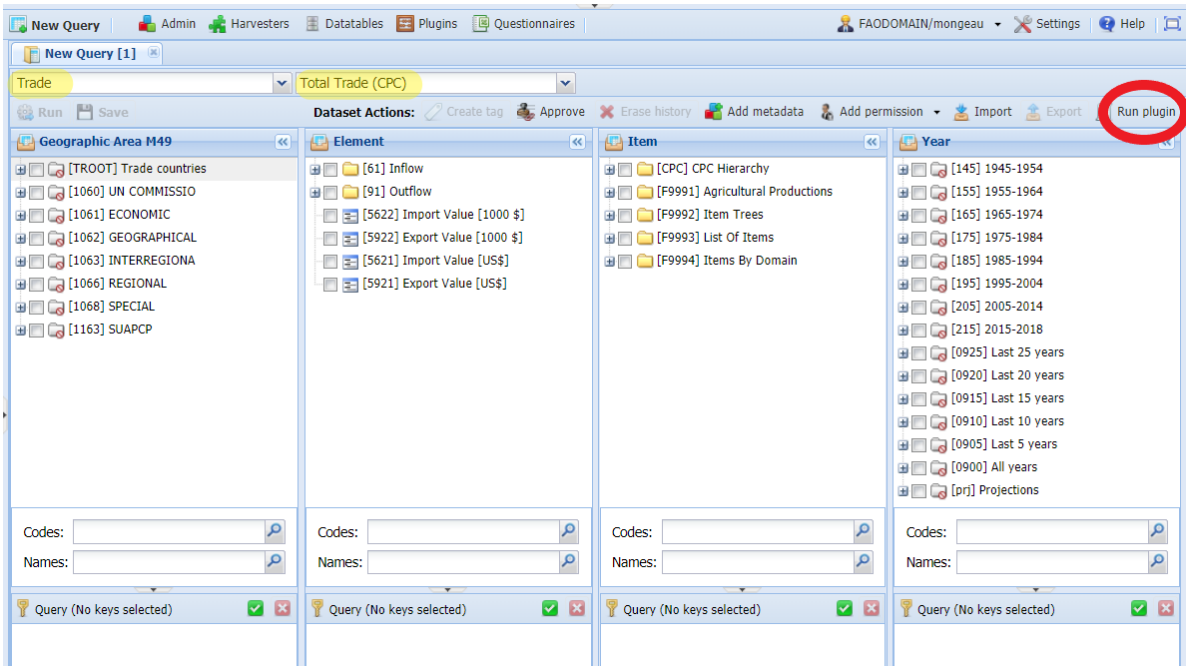


2. Select the "[FAODOMAIN/mongeau] Complete Trade Flow (CPC) (core)" plugin, set the year the plugin needs to process data, and click "Run plugin":

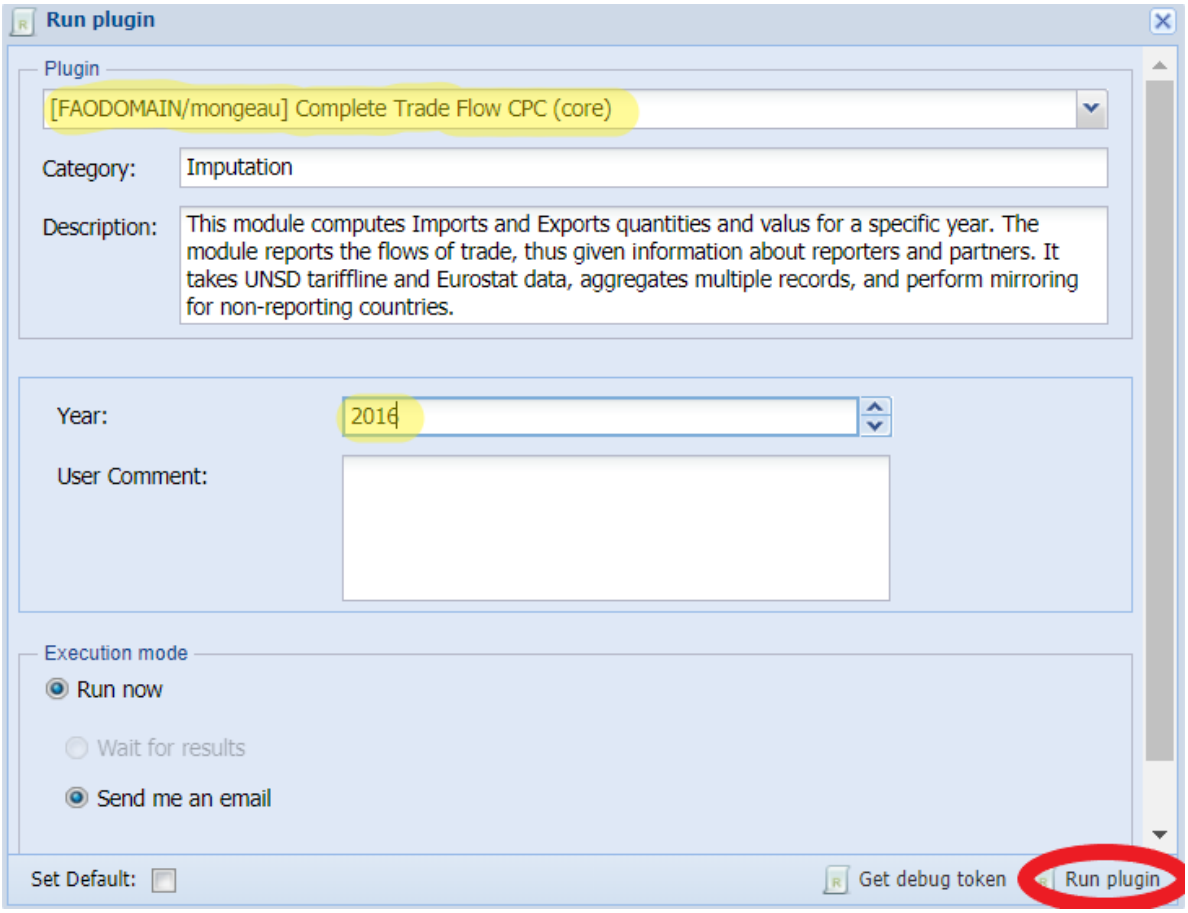


When the module completes, the person that launched the plugin will get an email from SWS with subject "FAO SWS - R plugin execution requested": the body says whether it "has completed" or if it "has failed". Moreover, if the plugin completes successfully, all users in the `ess_trade_people_in_charge` datatable of the `trade-reference-files` will receive an email with subject "Bilateral trade plugin (year *YEAR*) ran successfully" (where *YEAR* is the year set as parameter).

3. Once the "Complete Trade Flow (CPC)" plugin has completed successfully, open a query in the "Total Trade (CPC)" dataset of the "Trade" domani (as before, no dimensions need to be set given this is a *core* plugin), and click on "Run plugin":



4. Select the "[FAODOMAIN/mongeau] Complete Trade Flow (CPC) (core)" plugin, set the year the plugin needs to process data, and click "Run plugin":



As for the bilateral plugin, SWS will send an email on the completion status of the plugin ("FAO SWS - R plugin execution requested") and the users in the `ess_trade_people_in_charge` datatable will receive an email with subject "Total trade plugin (year *YEAR*) ran successfully" (where *YEAR* is the year set as parameter) if the plugin completes successfully.

5. Steps 1-4 must to be repeated for all years that need to be processed.

6. Validation by country analysts starts, as described in the next subsection.

As of this writing (June 2020), the "Complete Trade Flow (CPC)" plugin and the "Total Trade (CPC)" plugin take around 25 and 7 minutes on average, respectively (the amount of time depends on the SWS server performance, the amount of data, and other techincal factors).

5.3 Validation flow

Once the data has been saved for all years, as described in the previous subsection, analysts will start the data validation process. It differs on whether the validation needs to be done for reporting countries, or non-reporting countries. Also, validation of commodity tables need to be taken into account. Validation occurs in the following order:

1. reporting countries
2. non-reporting countries
3. total merchandise trade
4. commodity tables

At the beginning of the data validation excercise, a "monitoring table" will be compiled by the trade officer. It contains the distribution of countries for the first 3 validation phases (reporting countries, non-reporting countries, and total merchandise trade), and the distribution of commodities (for commodity tables). Given that the process is sequential, not necessarily the distribution for all steps will be available at the very beginning of the cycle.

In some of the steps involved in validation, the analyst will need to check for external sources of data. Some examples of where to find additional data are:

- TRADEMAP: https://www.trademap.org/
- FAO-AMIS: https://app.amis-outlook.org/#/market-database/supply-and-demand-overview
- USDA: https://apps.fas.usda.gov/psdonline/app/index.html#/app/advQuery
- National Statistics Offices
- Publications available on the web or in FAO library (e.g., "Oil World")

Documents with more details on the process will be circulated by the trade officer at the beginning of the cycle.

5.3.1 Reporting countries

This step will show the analyst outliers at the total level, by running the plugin:

`Tradeoutlier_most_recent_version`

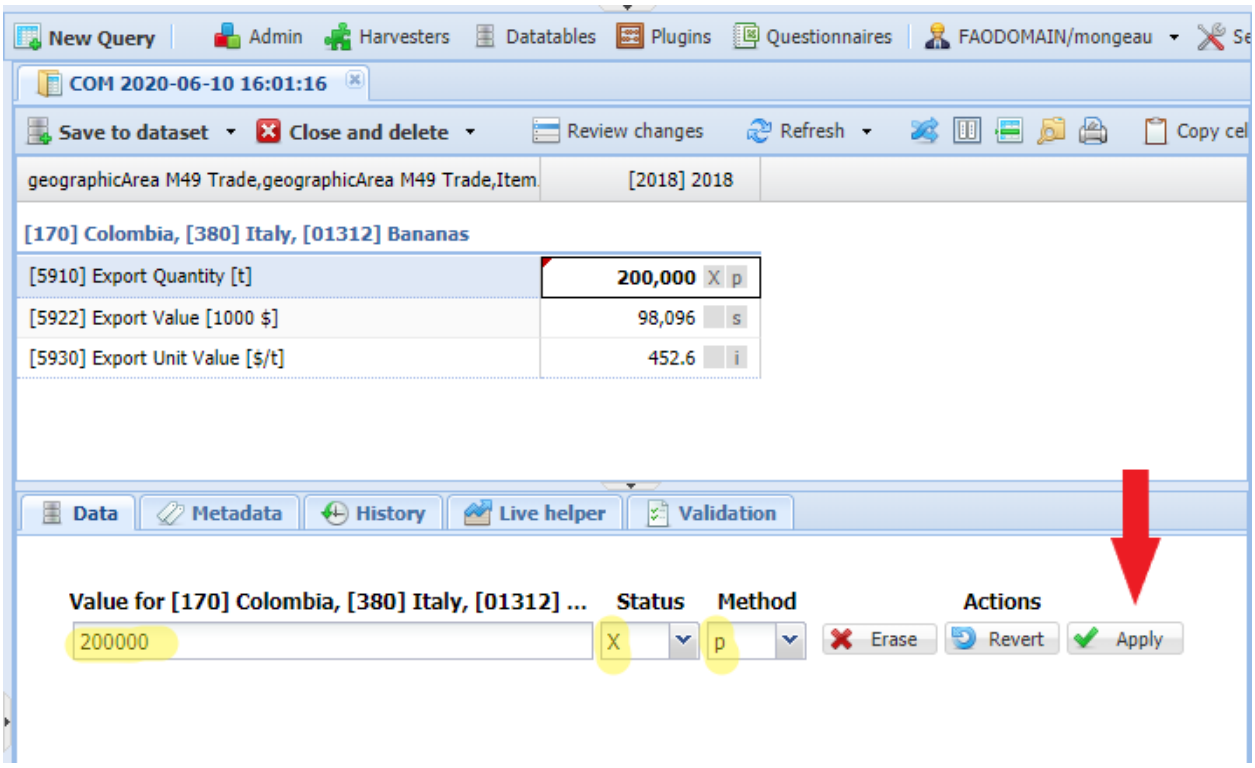
when it completes, and email will be sent with a list of outliers. The analysts is required to have a look at the historical time series of all items contained in the email to check whether the identified outlier is a true positive (i.e., an actual outlier) and, thus, it deserves attention. For doing this, there are two options:

- open a session in the "Bilateral Trade Flow (CPC)" dataset with the required reporter/item/flow/years combination, making sure to include all partners (as for years, it would be useful to have data for at least five years previous of the identified outlier).
- open that Shiny tool, as explained in the "Interactive tool for data validation (Shiny tool)" Section.

Given that in any case the actual validation of bilateral data happens through the Shiny tool, the second option is preferred.

The analyst is required to use the tools available in the Shiny tool to make corrections to bilateral data, and keep comments on the actions made. These can be saved in the Excel file sent by the `Tradeoutlier_most_recent_version` plugin. Taking notes will help post-validation checks by officers.

If, for some valid reason, bilateral flows cannot be reconciled with the sum of bilateral flows even after these are validated, the analyst have the option to override the aggregated total obtained from bilateral data, and input some exteranl figure, considered to be more in line with the country historical trend. In this case, the analyst need to write the appropriate number in the corresponding cell on the SWS session on the total dataset, and change flags in an appropriate way (e.g., if the data comes from a national statistical office, the flag combination needs to be changed to X,p), and click on "Apply":



Once this operation is done, a red triangle will appear, indicating it is "session data". The analyst will need to click on th "Save to database" button (visible in the previous screenshot in the upper-right part, third row).

5.3.2 Non-reporting countries

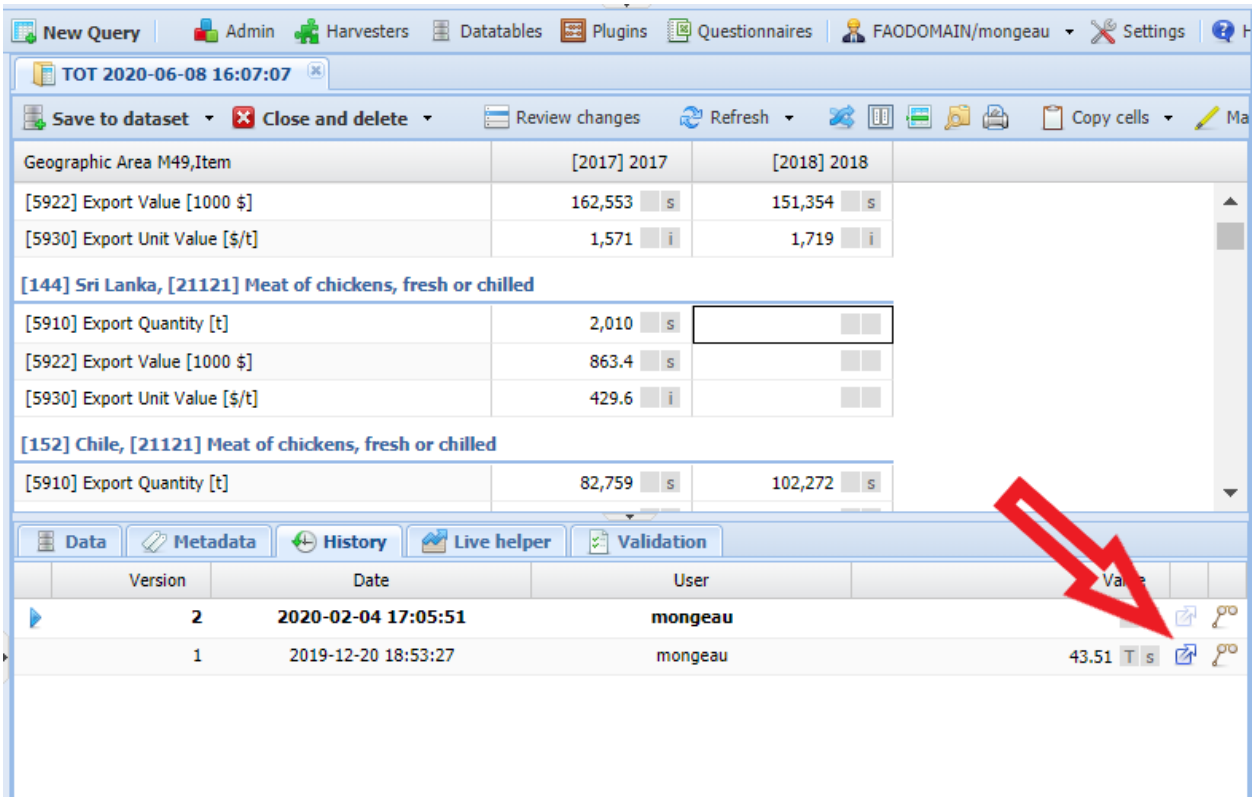
As the validation of reporting countries is completed, the analysts will be required to validate non-reporting (thus, *mirrored*) countries. The following plugin:

`top_commodities_selection_routine`

will read data for the required non-reporting and will send an email with a list of deleted flows (because they were considered not to match historical time series for that partner, see the "Discarding incompatible mirror data ("TP criterion") Section"), or flows that are relatively small, compared to previous data.

Analysts will need to find external data sources for **all deleted flows** (highlited with red cells in the email sent by the plugin) or, though in rare occasions, estimate themselves what the flow should be.

It may happen that external sources confirm the removed figure (e.g., a drought affected seriously the country and the exports are much lower compared to its standards). If this is the case, the analysts will need to "restore" a figure, i.e., selecting the deleted figure and selecting "History" in the lower part of the SWS UI, then the appropriate history entry, then click the restore symbol, as shown in the following Figure:



After data has been restored, a red triangle will appear, indicating it is "session data". The analyst will need to click on th "Save to database" button (visible in the previous screenshot in the upper-right part, third row).

Please note that non-reporting countries cannot be validated in the Shiny tool. This is because the tool is meant to be used in the validation of reporters.

Some outliers may have been missed by analysts that worked on reporting countries or, more likely, what is not an outlier (or important one) for a reporter, may be so for a non-reporting country. Thus, a check for the presence of outliers will be required by using the following plugin:

`Tradeoutlier_most_recent_version`

Once outliers are identified, the analyst will need to check if the bilateral flow that creates the outlier at the total level is an outlier also for the country that originally reported that flow, or if it is an outlier only for the non-reporting country that is being analysed. In the former case, the analyst need to contact the colleague that had that country under his/her responsibility and ask for this to be re-validated. In the latter case, the analyst will need to contact the trade officer and agreen on a solution.

5.3.3 Total merchandise trade

`monetary_values_ranked_commodities`

5.3.4 Commodity tables

`trade_commodity_tables`

6 Future work

6.1 Streamline raw data updates

As of June 2020 the process of checking whether a country that in a previous round had no official data, or any other major data update, is done manually. The person in charge looks at the "Data availability" comtrade page at:

https://comtrade.un.org/data/da

and then decides based on the needs, what data needs to be updated. This process has at least a couple of drawbacks:

1. being a manual check, it is time consuming and can be error prone
2. it relies on public comtrade data

As for point 1), a routine should be developed so that it queries the comtrade "Data availability" page and compares the status of country/year combinations, in order to obtains insight of the status of the data. This depends on the data made available by comtrade, and this brings to point 2): the ESS trade processing system uses UNSD raw Tariffline data and the date when they are disseminated in comtrade may be different. Probably, UNSD and FAO should agree on having a shared table or communication mechanism that says when Tariffline data are updated.

6.2 Outlier identification/imputation at the bilateral level

Outliers were identified in a previous version of the "Complete Trade Flow CPC" module and imputed automatically by using the median unit value with a specific-to-generic median unit value calculation, i.e., with the method used for estimating mising quantities. Results were found to be unsatisfactory, thus automatic imputation was switched off. Said strategy could be supplemented by using information of neighbour or similar countries (e.g., the median unit value of Asian countries for a detailed HS level can be used for imputing an outlier for an Asian country instead of going up to the HS8 or HS6 level for the country itself as attempted in the specific-to-generic approach).

6.3 CIF/FOB

The CIF/FOB correction for mirroring is, at the time being, set up to 12%. This has been suggested by the CLFS Team. Additional work might be done in order to assess if this estimate is appropriate, but logic suggests that this is a very crude approximation. Indeed, there are different range of percentages for different type of countries and by distance between reporters and partners (e.g., the cost of transportation of a given commodity is definitely different if it is between France and Italy or France and Australia). A study can be conducted on available records on both sides: this means records for which the commodity is reported by the reporter and by the partner.

To have an idea of the different CIF/FOB margins by countries, refer to the following dataset (by OECD):

https://stats.oecd.org/Index.aspx?DataSetCode=CIF_FOB_ITIC

A working paper[19] describing the methodology used is available at:

https://www.oecd-ilibrary.org/economics/estimating-transport-and-insurance-costs-of-international-trade_8267bb0f-en

[19]Miao, G., Fortanier, F., (2017), "Estimating Transport and Insurance Costs of International Trade", OECD Statistics Working Papers 2017/04.

6.4 Self trade analysis

- A script within the vignette folder, named selftrade.R, has been used to perform some simple analyses on the self trade. The script filter all records for which the reporter and the partner are the same. The script compute the sum of all value across all commodities per country (Figure 1), or the sum of all the value for each commodity across all countries (Figure 2). In this way we can spot out the countries reporting massive self trade as well as which are the main commodities reported in self trade.

  This is an example of the graphical output (still part of the script).

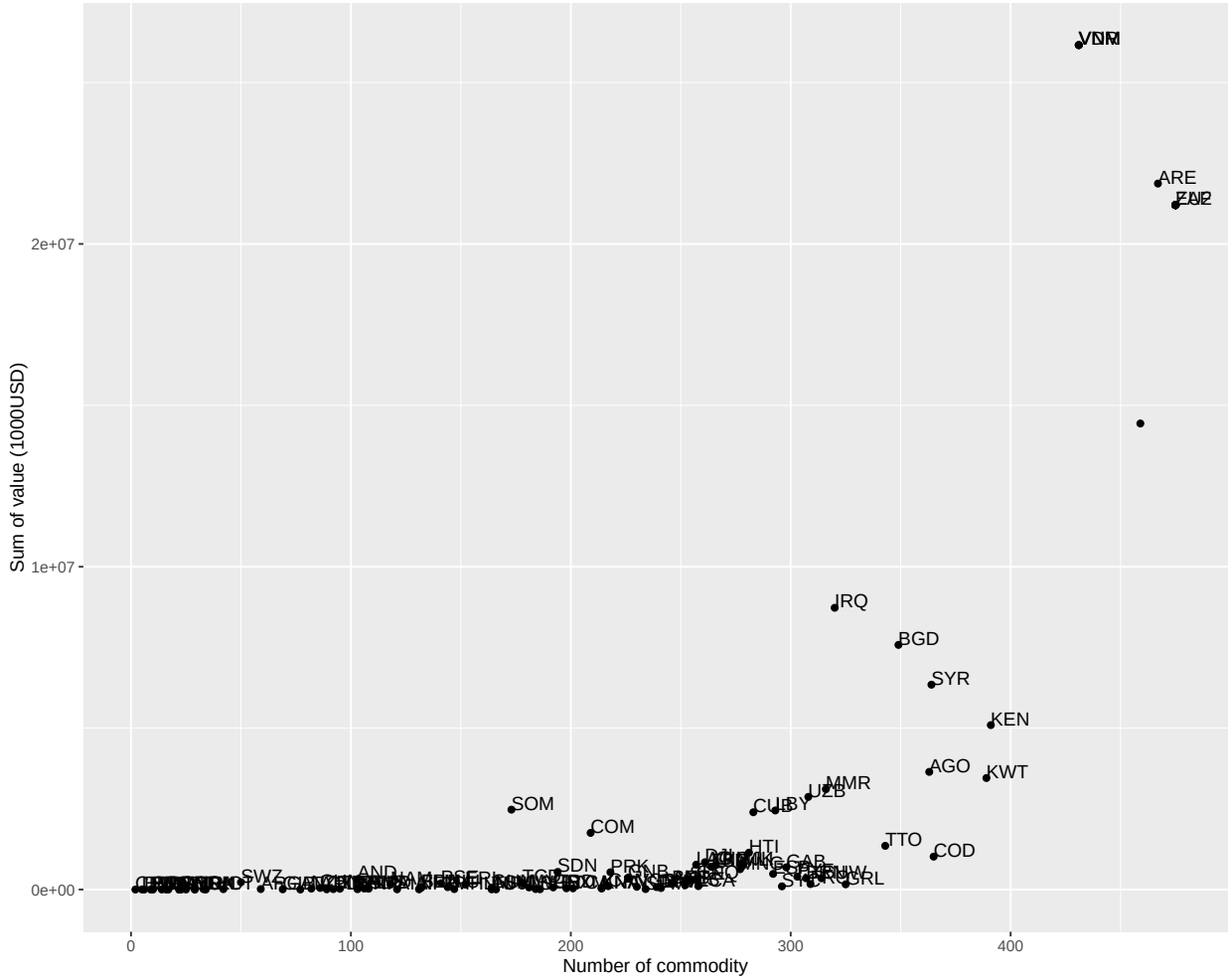


Figure 1: Sum of all self trade records by country.

- This might be incorporated in the module and might produce suitable output within the SWS. More documentation is needed.

6.5 Mapping from HS to FCL/CPC

In the module for commodities we have two different mappings. From HS to FCL, using a mapping table produced by the CLFS Team and then from FCL to CPC 2.1. In the future direct mapping from HS to CPC has been asked from management. A possible solution, where adding the column with the one-to-one CPC codes has been sent to Carola (09.06.2016), but anyway this needs revision (link)

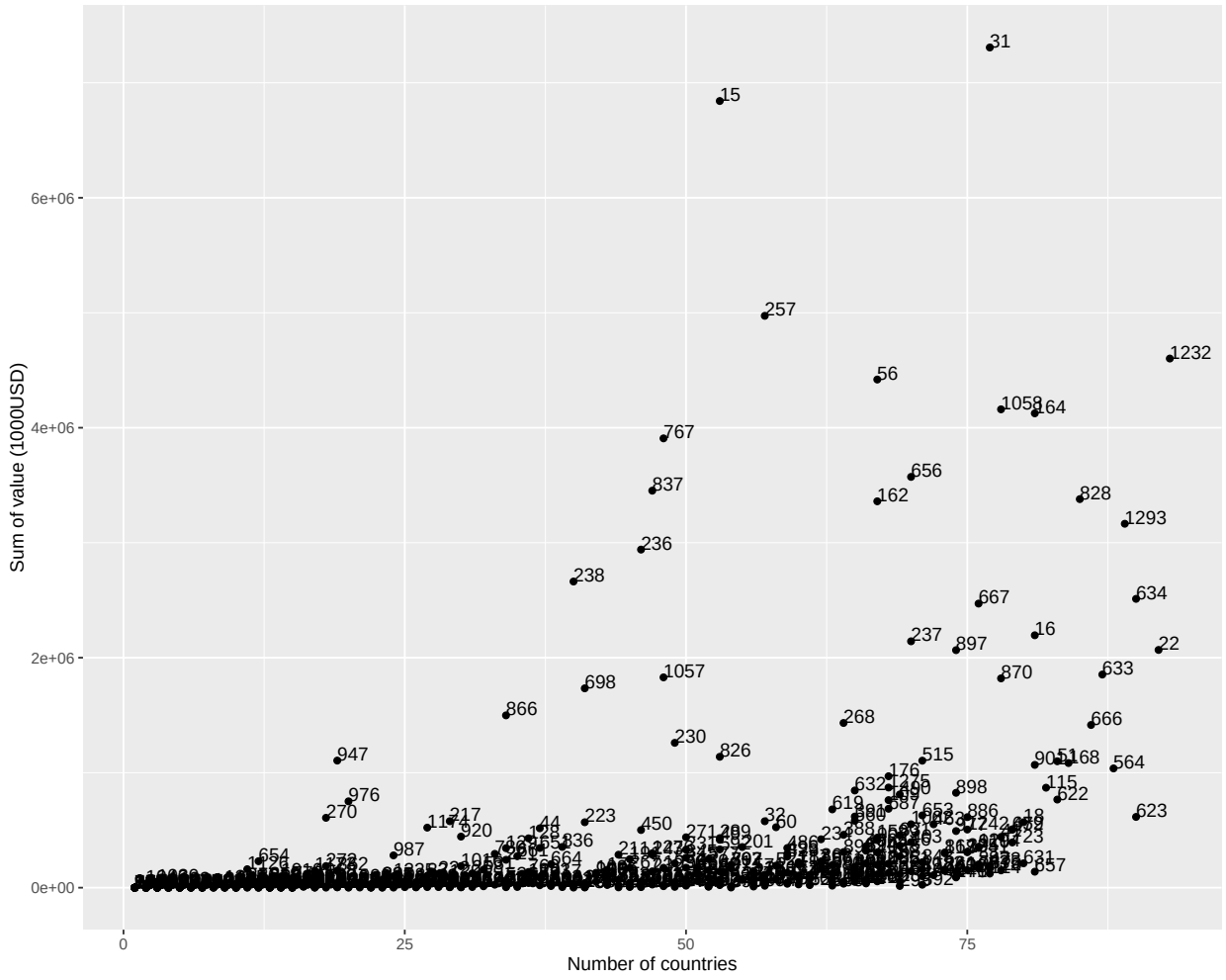


Figure 2: Sum of all self trade records by commodity. Codes in FCL.

6.6 Mapping from Comtrade M49 and Geonomenclature directly to M49

The country codes, as the commodity ones, have two steps of mapping. This results in higher risk of data loss due to unsolved mapping. A direct map from Comtrade M49 (Tariff line UNSD) to M49 and from Geonomenclature (Eurostat) to M49 would be ideal.

6.7 Food-aid

This has to be incorporated also to understand the trend in a time series analysis. This needs special study to understand if we can get the data just from the exports not reported as imports in the partner.

6.8 Other tasks

The GitHub repo of the faoswsTrade module contains a list of "issues" with different degrees of urgency, so that future developers can improve the module:

https://github.com/SWS-Methodology/faoswsTrade/issues

Some of the issues are feature requests. For instance, the latest issue as of this writing (June 2020) is issue #176: it is suggested to merge the "Complete Trade Flow CPC" and "Total Trade CPC" plugins into a single plugin. The reason is that the two are usually run together in two different scheduled runs, and for the reasons exaplained in the issue, it seems sensible to have a single plugin for the process. As an additional comment, the "Total Trade CPC" may be recreated as a function that can be called as "Complete Trade Flow CPC" ends or from a standalone small plugin that just calls this function to re-calculate total flows if that is seen as a requirement.

Appendix: Statistical Working System

According to the definition given in the ESS website, the "Statistical Working System is a corporate platform used in FAO for the processing and storage of statistical datasets, providing the framework needed to use the same methods, standards, classifications and approaches within statistical processes."[20]

In a nutshell (and for what it is useful as far as trade processing in concerned), the SWS is composed of three different elements: a user interface (UI), a datataset backend, an R server.

There are two SWS environments: *QA* (for *Quality Assurance*) and *LIVE* (known more frequently by CIO and other divisions as *production*).[21] The two are conceptually the same, the main difference being that QA is used during the development phase, from both statistical (by ESS and other techincal units), and information techonology (by CIO) points of view. These are the URL from accessing the two servers, from the web and R-API:

- QA:
  - web: https://hqlqasws1.hq.un.fao.org:8443/sws/
  - R-API: https://hqlqasws1.hq.un.fao.org:8181/sws
- LIVE (production):
  - web: https://sws.fao.org/
  - R-API: https://sws.fao.org:8181

Describing the SWS UI is out of the scope of this document. In the following sections, the frontend side of the database backend, which stores datasets and datatables, and the R-server will be reviewed.

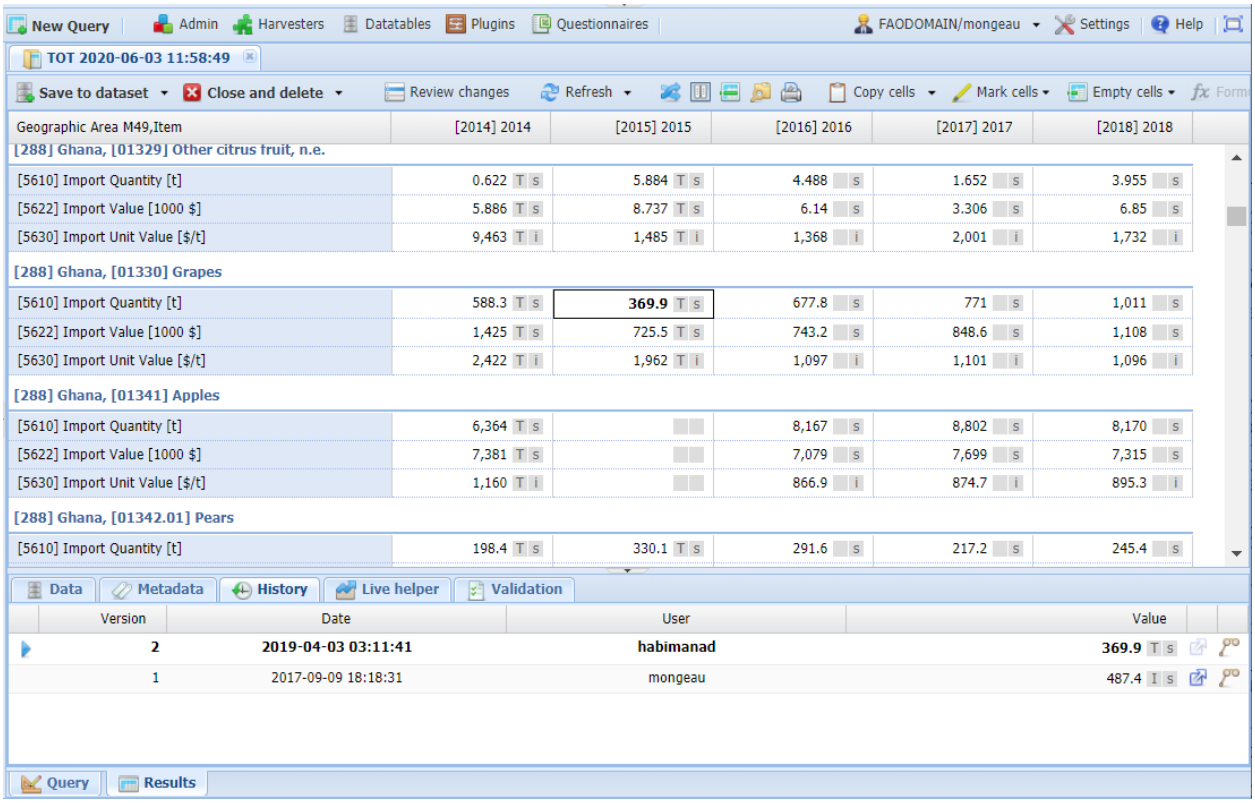
[20] Source: http://www.fao.org/economic/ess/ess-home/ess-about/statistical-methods/en/

[21] The CLFS team decided to call the *production* environment as *LIVE* in order to avoid confusion with the *production* domain (i.e., the domain where agricultural production data is stored).

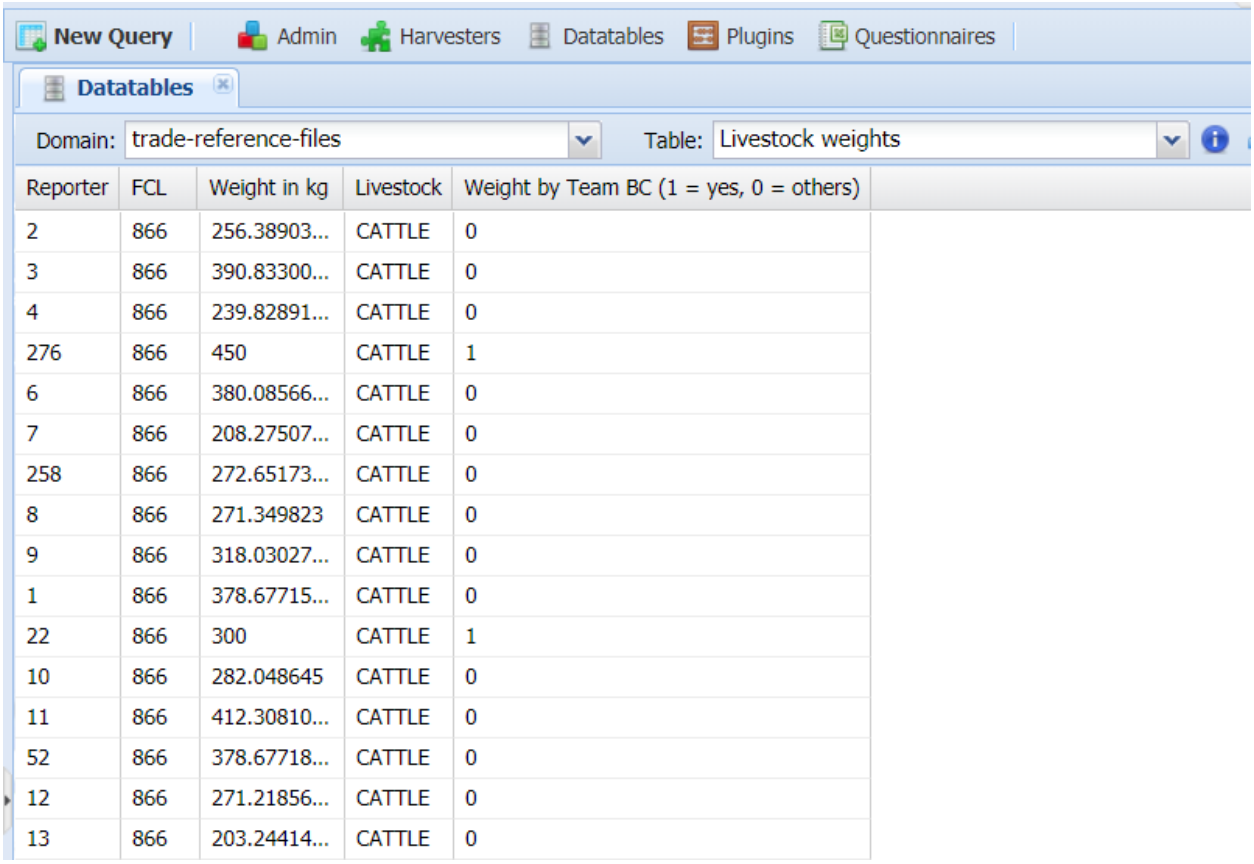
6.9 Datasets and datatables

The SWS database server contains two different kind of objects for storing data: datasets and datatables. The two differ in some aspects, the main ones being datasets have a history log for each cell (i.e., when the number was written, by whom, with which plugin, etc.), and cell and block metadata, while datatables do not have any of this two features. Being datatables more simple objects, they are usually used for data that is not meant to be changed often, like reference tables or raw data, as in the case of the raw bilateral trade data used by the trade plugins.

An example of an SWS dataset and its history log is given below. It shows results of a query in the "Total Trade (CPC)" dataset of the "Trade" domain. It shows also the "history" feature: the selected observation (import quantity of grapes by Ghana in 2015) was written twice by two different users. The observation figure/flags can be changed in the "Data" tab. Metadata can be obtained by clicking on the "Metadata" tab. A plot of the selected row(s) is available in the "Live helper" tab.



An example of reference data in an SWS datatable is given below. The table contains livestock weight used by the "Complete Trade Flow CPC" plugin when imputation of heads is required (i.e., when the data contains information only on total weight).



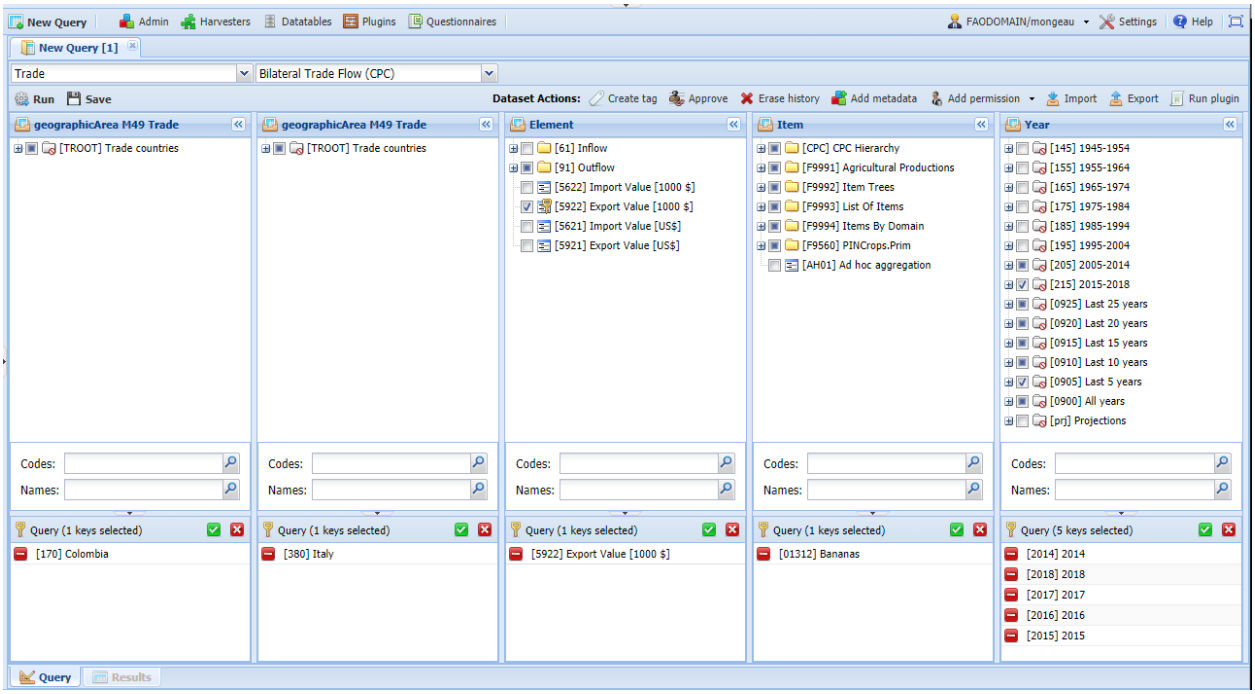
Reporter	FCL	Weight in kg	Livestock	Weight by Team BC (1 = yes, 0 = others)
2	866	256.38903...	CATTLE	0
3	866	390.83300...	CATTLE	0
4	866	239.82891...	CATTLE	0
276	866	450	CATTLE	1
6	866	380.08566...	CATTLE	0
7	866	208.27507...	CATTLE	0
258	866	272.65173...	CATTLE	0
8	866	271.349823	CATTLE	0
9	866	318.03027...	CATTLE	0
1	866	378.67715...	CATTLE	0
22	866	300	CATTLE	1
10	866	282.048645	CATTLE	0
11	866	412.30810...	CATTLE	0
52	866	378.67718...	CATTLE	0
12	866	271.21856...	CATTLE	0
13	866	203.24414...	CATTLE	0

6.9.1 List of datasets used by the trade module

There are two datasets relative to the trade process stored in the "Trade" domain: "Bilateral Trade Flow (CPC)" and "Total Trade (CPC)". As the names suggest, the former contains bilateral transactions, while the latter contains total flows (computed as aggregation of bilateral transactions).

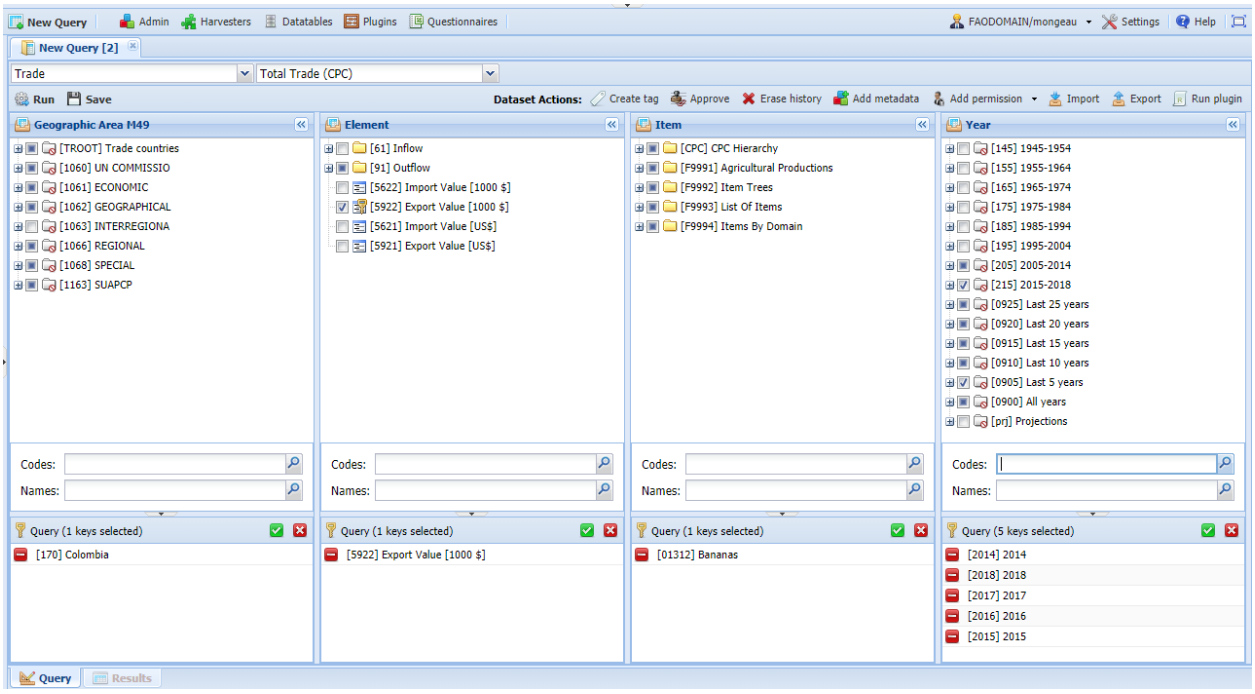
6.9.1.1 Bilateral Trade Flow (CPC)

This dataset (SWS id: `completed_tf_cpc_m49`) is defined by five dimensions: two "Geographic Area M49 Trade" (one for reporters, the other for partners, shown on the left and right, respectively), "Element", "Item", and "Year". These dimensions correspond to the following SWS codelists: `geographicAreaM49Reporter`, `geographicAreaM49Partner`, `measuredElementTrade`, `measuredItemCPC`, and `timePointYears`, respectively.



6.9.1.2 Total Trade (CPC)

This dataset (SWS id: `total_trade_cpc_m49`) is defined by four dimensions: "Geographic Area M49", "Element", "Item", and "Year". These dimensions correspond to the following SWS codelists: `geographicAreaM49`, `measuredElementTrade`, `measuredItemCPC`, and `timePointYears`, respectively.



6.9.2 List of datatables used by the trade module

There are three different domains that contain datatables related to the trade process:

- `trade-input-data`: this domain contains raw bilateral data from UNSD and Eurostat.
- `trade-reference-files`: reference files used by the trade plugins are stored in this domain.
- `trade-reports`: the output or the "Pre-Processing Report" plugin are saved here.

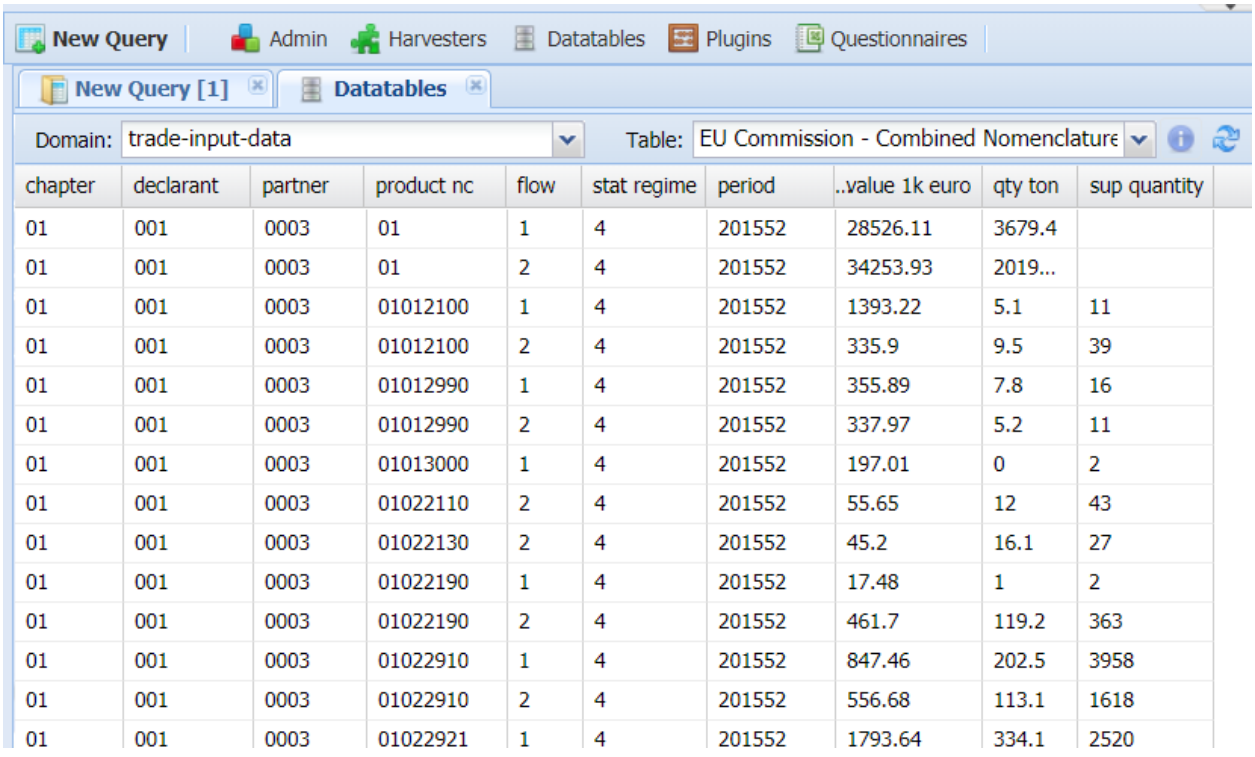
6.9.2.1 `trade-input-data`

6.9.2.1.1 EU Commission - Combined Nomenclature *YEAR*

These datatables (year specific, e.g., "EU Commission - Combined Nomenclature 2018")[22] contain the Eurostat raw trade data.

The variables contained are:

- `chapter`: the HS chapter (i.e., the first two digits of `product_nc`)
- `declarant`: reporting country (geonomenclature code)
- `partner`: partner country (geonomenclature code)
- `product_nc`: HS (CN8) code
- `flow`: flow (1 = imports; 2 = exports)
- `stat_regime`: statistical regime [23]
- `period`: year+week combined (e.g., 201852)
- `value_1k_euro`: monetary value in thousands of euros (FOB for exports, CIF for imports)
- `qty_ton`: weight in metric tons
- `sup_quantity`: supplementary quantity



chapter	declarant	partner	product nc	flow	stat regime	period	..value 1k euro	qty ton	sup quantity
01	001	0003	01	1	4	201552	28526.11	3679.4	
01	001	0003	01	2	4	201552	34253.93	2019...	
01	001	0003	01012100	1	4	201552	1393.22	5.1	11
01	001	0003	01012100	2	4	201552	335.9	9.5	39
01	001	0003	01012990	1	4	201552	355.89	7.8	16
01	001	0003	01012990	2	4	201552	337.97	5.2	11
01	001	0003	01013000	1	4	201552	197.01	0	2
01	001	0003	01022110	2	4	201552	55.65	12	43
01	001	0003	01022130	2	4	201552	45.2	16.1	27
01	001	0003	01022190	1	4	201552	17.48	1	2
01	001	0003	01022190	2	4	201552	461.7	119.2	363
01	001	0003	01022910	1	4	201552	847.46	202.5	3958
01	001	0003	01022910	2	4	201552	556.68	113.1	1618
01	001	0003	01022921	1	4	201552	1793.64	334.1	2520

[22] There is a datatable called "EU Commission - Combined Nomenclature All Years" that is supposed to have all "EU Commission - Combined Nomenclature *YEAR*" tables combined into one, but it is not maintained as it is not practically used (years available are 2000-2016).

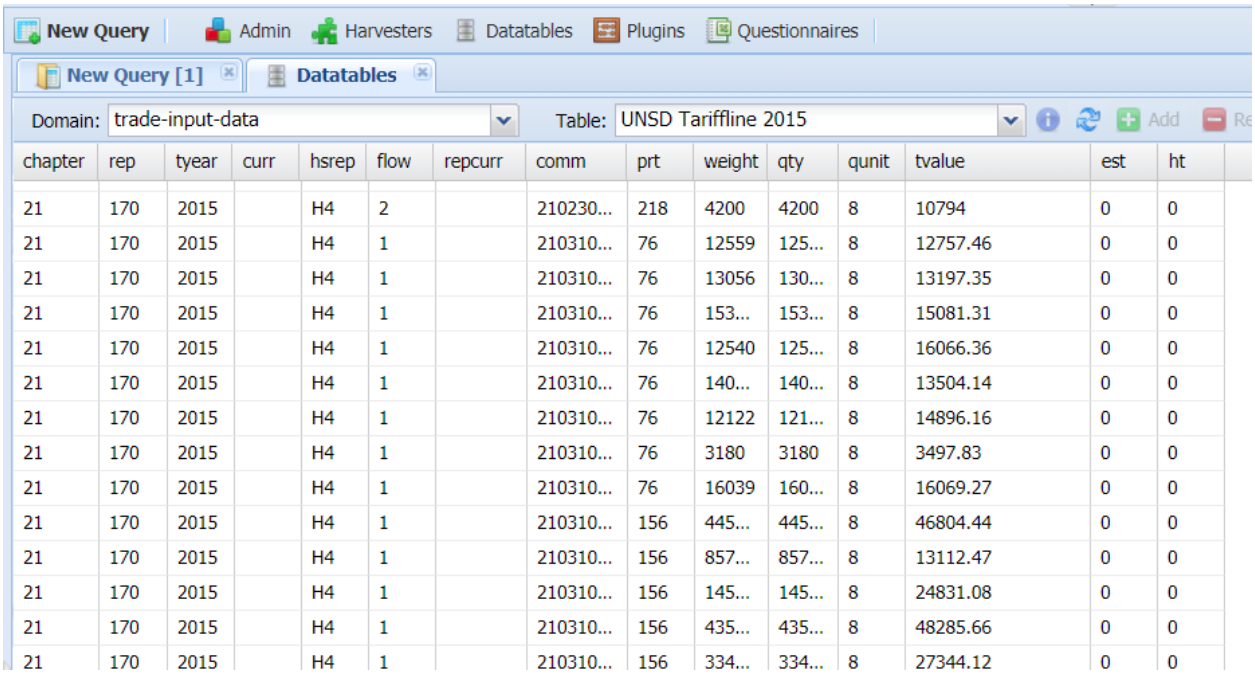
[23] See http://trade.ec.europa.eu/doclib/docs/2013/may/tradoc_151348.pdf.

6.9.2.1.2 UNSD Tariffline *YEAR*

These datatables (year specific, e.g., "UNSD Tariffline 2018")[24] contain UNSD trade data. These are also called "legacy format" as UNSD changed the format of how they disseminate Tariff line data (see the "UNSD Tariffline v2 *YEAR*" tables for "new" format). The legacy format was the one with which the main part of the trade processing system was developed.

The variables contained are:

- `chapter`: the HS chapter (i.e., the first two digits of `comm`)
- `rep`: reporting country (M49 code)
- `tyear`: year
- `curr`: currency
- `hsrep`: HS version
- `flow`: flow (1 = imports; 2 = exports; 3 = re-exports; 4 = re-imports)
- `repcurr`: currency of reporter
- `comm`: HS code
- `prt`: partner country (M49 code)
- `weight`: net weight in kilograms
- `qty`: quantity
- `qunit`: unit of measurement of `qty` (see also discussion on `comtradeunits`)
- `tvalue`: monetary value in US dollars (FOB for exports, CIF for imports)
- `est`: ??? (this variable is not used)
- `ht`: ??? (this variable is not used)



chapter	rep	tyear	curr	hsrep	flow	repcurr	comm	prt	weight	qty	qunit	tvalue	est	ht
21	170	2015		H4	2		210230...	218	4200	4200	8	10794	0	0
21	170	2015		H4	1		210310...	76	12559	125...	8	12757.46	0	0
21	170	2015		H4	1		210310...	76	13056	130...	8	13197.35	0	0
21	170	2015		H4	1		210310...	76	153...	153...	8	15081.31	0	0
21	170	2015		H4	1		210310...	76	12540	125...	8	16066.36	0	0
21	170	2015		H4	1		210310...	76	140...	140...	8	13504.14	0	0
21	170	2015		H4	1		210310...	76	12122	121...	8	14896.16	0	0
21	170	2015		H4	1		210310...	76	3180	3180	8	3497.83	0	0
21	170	2015		H4	1		210310...	76	16039	160...	8	16069.27	0	0
21	170	2015		H4	1		210310...	156	445...	445...	8	46804.44	0	0
21	170	2015		H4	1		210310...	156	857...	857...	8	13112.47	0	0
21	170	2015		H4	1		210310...	156	145...	145...	8	24831.08	0	0
21	170	2015		H4	1		210310...	156	435...	435...	8	48285.66	0	0
21	170	2015		H4	1		210310...	156	334...	334...	8	27344.12	0	0

6.9.2.1.3 UNSD Tariffline v2 *YEAR*

These datatables (year specific, e.g., "UNSD Tariffline v2 2018")[25] contain UNSD trade data. These are also called "new format" as UNSD changed the format of how they disseminate Tariff line data (see the "UNSD

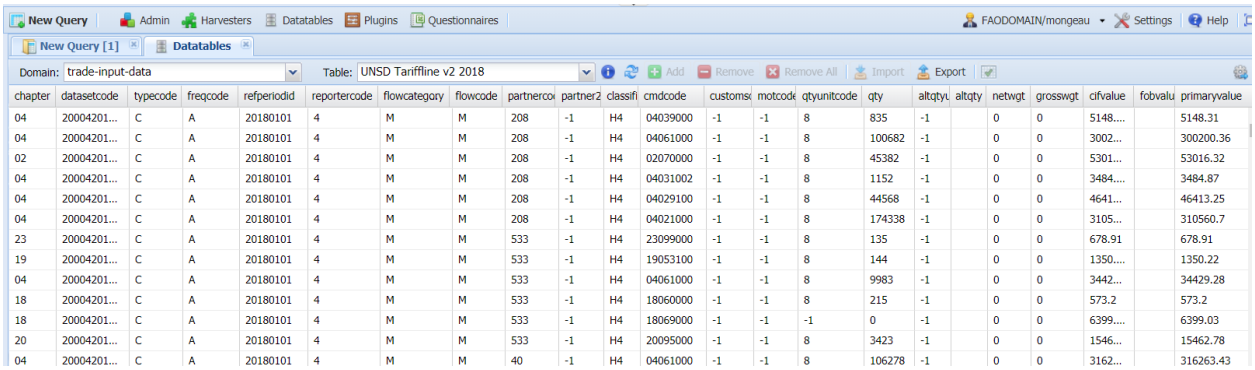
[24]There is a datatable called "UNSD Tariffline All years" that is supposed to have all "UNSD Tariffline *YEAR*" tables combined into one, but it is not maintained as it is not practically used (years available are 2000-2016).

[25]There is a datatable called "UNSD Tariffline All years" that is supposed to have all "UNSD Tariffline *YEAR*" tables combined into one, but it is not maintained as it is not practically used (years available are 2000-2016).

Tariffline *YEAR*" tables for "legacy" format). Note that given that the legacy format was the one with which the main part of the trade processing system was developed, the new tables are converted to "legacy" by the "Complete Trade Flow CPC" plugin.

The variables contained are:

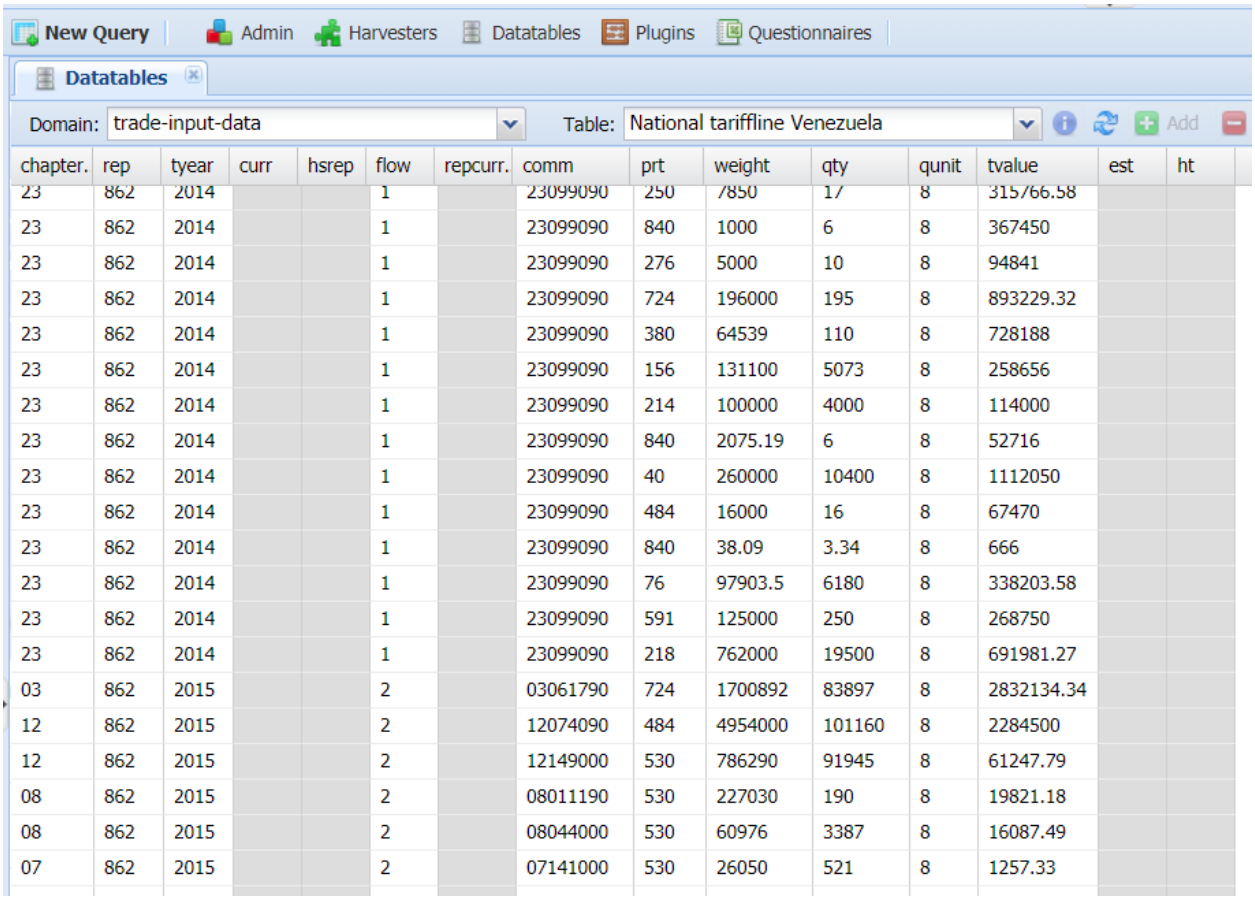
- `chapter`: the HS chapter (i.e., the first two digits of `comm`)
- `datasetcode`: combination of several keys to identify specific datasetCode
- `typecode`: product type (goods or services)
- `freqcode`: the time interval at which observations occur
- `refperiodid`: the period of time to which the measured observation is intended to refer
- `reportercode`: the country or geographic area to which the measured statistical phenomenon relates
- `flowcategory`: simplified trade flow (exports or imports)
- `flowcode`: trade flow or sub-flow (exports, re-exports, imports, re-imports, etc.)
- `partnercode`: the primary partner country or geographic area for the respective trade flow
- `partner2code`: a secondary partner country or geographic area for the respective trade flow
- `classificationcode`: indicates the product classification used and which version (HS, SITC)
- `cmdcode`: product code in conjunction with classification code
- `customscode`: customs or statistical procedure
- `motcode`: the mode of transport used when goods enter or leave the economic territory of a country
- `qtyunitcode`: unit of primary quantity
- `qty`: value of primary quantity
- `altqtyunitcode`: unit of secondary quantity
- `altqty`: value of secondary quantity
- `netwgt`: net weight
- `grosswgt`: gross weight
- `cifvalue`: trade value in CIF
- `fobvalue`: trade value in FOB
- `primaryvalue`: primary trade values (taken from CIF or FOB values)



chapter	datasetcode	typecode	freqcode	refperiodid	reportercode	flowcategory	flowcode	partnerco	partner2	classifi	cmdcode	customs	motcode	qtyunitcode	qty	altqtyu	altqty	netwgt	grosswgt	cifvalue	fobvalu	primaryvalue
04	20004201...	C	A	20180101	4	M	M	208	-1	H4	04039000	-1	-1	8	835	-1		0	0	5148....		5148.31
04	20004201...	C	A	20180101	4	M	M	208	-1	H4	04061000	-1	-1	8	100682	-1		0	0	3002...		300200.36
02	20004201...	C	A	20180101	4	M	M	208	-1	H4	02070000	-1	-1	8	45382	-1		0	0	5301...		53016.32
04	20004201...	C	A	20180101	4	M	M	208	-1	H4	04031002	-1	-1	8	1152	-1		0	0	3484....		3484.87
04	20004201...	C	A	20180101	4	M	M	208	-1	H4	04029100	-1	-1	8	44568	-1		0	0	4641...		46413.25
04	20004201...	C	A	20180101	4	M	M	208	-1	H4	04021000	-1	-1	8	174338	-1		0	0	3105...		310560.7
23	20004201...	C	A	20180101	4	M	M	533	-1	H4	23099000	-1	-1	8	135	-1		0	0	678.91		678.91
19	20004201...	C	A	20180101	4	M	M	533	-1	H4	19053100	-1	-1	8	144	-1		0	0	1350....		1350.22
04	20004201...	C	A	20180101	4	M	M	533	-1	H4	04061000	-1	-1	8	9983	-1		0	0	3442...		34429.28
18	20004201...	C	A	20180101	4	M	M	533	-1	H4	18060000	-1	-1	8	215	-1		0	0	573.2		573.2
18	20004201...	C	A	20180101	4	M	M	533	-1	H4	18069000	-1	-1	-1	0	-1		0	0	6399....		6399.03
20	20004201...	C	A	20180101	4	M	M	533	-1	H4	20095000	-1	-1	8	3423	-1		0	0	1546...		15462.78
04	20004201...	C	A	20180101	4	M	M	40	-1	H4	04061000	-1	-1	8	106278	-1		0	0	3162...		316263.43

6.9.2.1.4 National tariffline *COUNTRY*

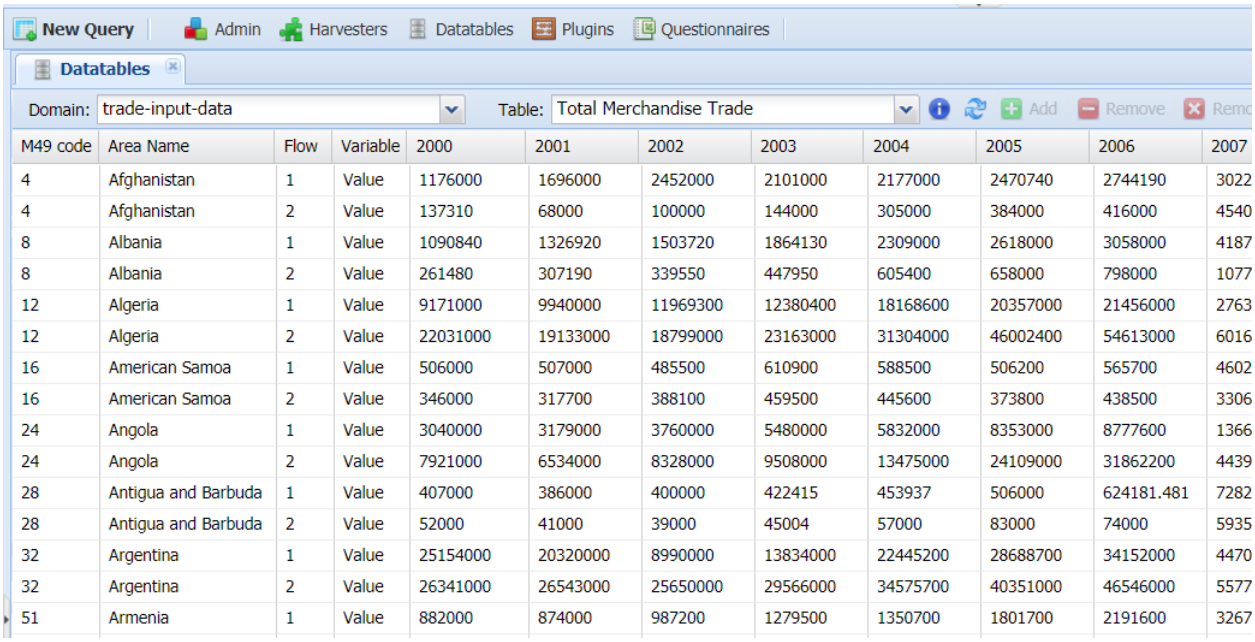
These datatables contain rawdata for `COUNTRY`, where `COUNTRY` is a country for which UNSD has no data or, for any reason, the UNSD data cannot be used, and FAO has access to country files. The table is structured as the "UNSD Tariffline YEAR" datatables, except that it may contain data for different years: given that countries do not generally share the files with this format, the original files are converted into the UNSD format. The only country for which such datatable exists, so far, is Venezuela ("National tariffline Venezuela"). In the example below a portion of the table is shown, where it can be seen that 2014 and 2015 are avaiable.



chapter.	rep	tyear	curr	hsrep	flow	repcurr.	comm	prt	weight	qty	qunit	tvalue	est	ht
23	862	2014			1		23099090	250	7850	17	8	315766.58		
23	862	2014			1		23099090	840	1000	6	8	367450		
23	862	2014			1		23099090	276	5000	10	8	94841		
23	862	2014			1		23099090	724	196000	195	8	893229.32		
23	862	2014			1		23099090	380	64539	110	8	728188		
23	862	2014			1		23099090	156	131100	5073	8	258656		
23	862	2014			1		23099090	214	100000	4000	8	114000		
23	862	2014			1		23099090	840	2075.19	6	8	52716		
23	862	2014			1		23099090	40	260000	10400	8	1112050		
23	862	2014			1		23099090	484	16000	16	8	67470		
23	862	2014			1		23099090	840	38.09	3.34	8	666		
23	862	2014			1		23099090	76	97903.5	6180	8	338203.58		
23	862	2014			1		23099090	591	125000	250	8	268750		
23	862	2014			1		23099090	218	762000	19500	8	691981.27		
03	862	2015			2		03061790	724	1700892	83897	8	2832134.34		
12	862	2015			2		12074090	484	4954000	101160	8	2284500		
12	862	2015			2		12149000	530	786290	91945	8	61247.79		
08	862	2015			2		08011190	530	227030	190	8	19821.18		
08	862	2015			2		08044000	530	60976	3387	8	16087.49		
07	862	2015			2		07141000	530	26050	521	8	1257.33		

6.9.2.1.5 Total Merchandise Trade

This table contains data on total merchandise imports and exports for all countries, expressed in US dollars.



M49 code	Area Name	Flow	Variable	2000	2001	2002	2003	2004	2005	2006	2007
4	Afghanistan	1	Value	1176000	1696000	2452000	2101000	2177000	2470740	2744190	3022
4	Afghanistan	2	Value	137310	68000	100000	144000	305000	384000	416000	4540
8	Albania	1	Value	1090840	1326920	1503720	1864130	2309000	2618000	3058000	4187
8	Albania	2	Value	261480	307190	339550	447950	605400	658000	798000	1077
12	Algeria	1	Value	9171000	9940000	11969300	12380400	18168600	20357000	21456000	2763
12	Algeria	2	Value	22031000	19133000	18799000	23163000	31304000	46002400	54613000	6016
16	American Samoa	1	Value	506000	507000	485500	610900	588500	506200	565700	4602
16	American Samoa	2	Value	346000	317700	388100	459500	445600	373800	438500	3306
24	Angola	1	Value	3040000	3179000	3760000	5480000	5832000	8353000	8777600	1366
24	Angola	2	Value	7921000	6534000	8328000	9508000	13475000	24109000	31862200	4439
28	Antigua and Barbuda	1	Value	407000	386000	400000	422415	453937	506000	624181.481	7282
28	Antigua and Barbuda	2	Value	52000	41000	39000	45004	57000	83000	74000	5935
32	Argentina	1	Value	25154000	20320000	8990000	13834000	22445200	28688700	34152000	4470
32	Argentina	2	Value	26341000	26543000	25650000	29566000	34575700	40351000	46546000	5577
51	Armenia	1	Value	882000	874000	987200	1279500	1350700	1801700	2191600	3267

6.9.2.2 `trade-reference-files`

6.9.2.2.1 Adjustments

This datatable contains the corrections to raw data that were applied in the previous trade processing system (*Shark/Jellyfish*), and is kept for reference. The current trade module does not use this table.

6.9.2.2.2 Aggregation *TYPE*

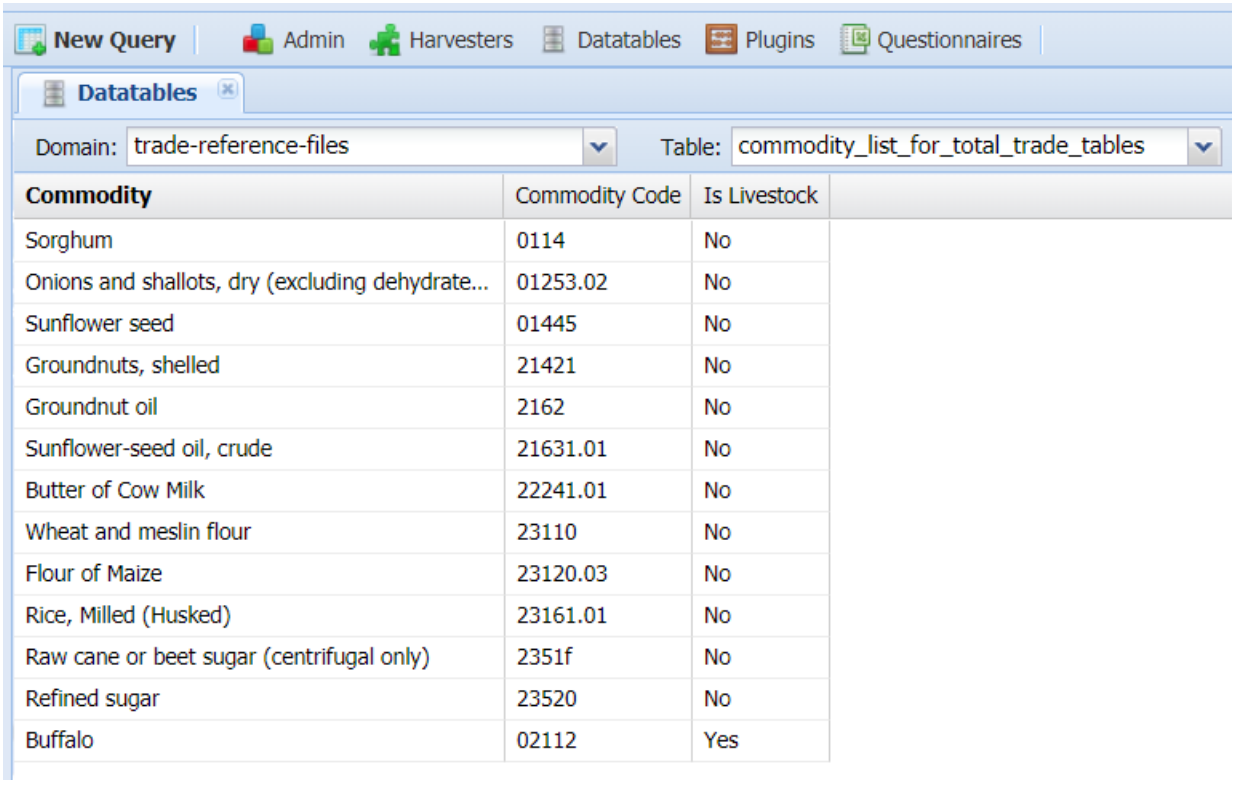
There are three different datatables, for these *TYPE*s: AreaGroups, AreaYears, ItemGroups. These datatables are used by a plugin that carries out aggregation, not by trade-related plugins.

6.9.2.2.3 commodity_list_for_total_trade_tables

This datatable contains the items used by the `trade_commodity_tables` plugin for doing the "Commodity tables".

Note that that this mapping also includes:

- the "European Union" aggregate, which includes 28 countries EU countries
- the "Europe" aggregate, which excludes 28 countries that are member of European Union
- the "Asia" aggregate, which excludes one country (Cyprus - 196) that is also a member of "Europe"

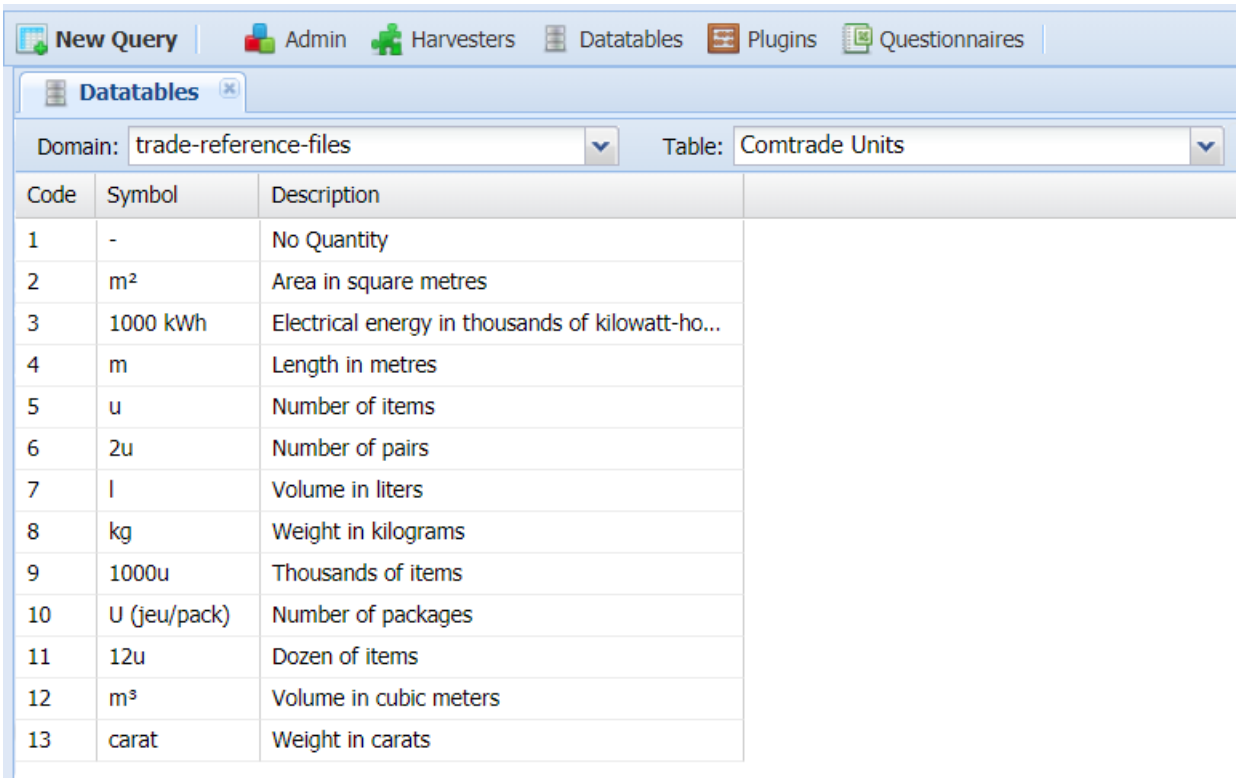


Commodity	Commodity Code	Is Livestock
Sorghum	0114	No
Onions and shallots, dry (excluding dehydrate...	01253.02	No
Sunflower seed	01445	No
Groundnuts, shelled	21421	No
Groundnut oil	2162	No
Sunflower-seed oil, crude	21631.01	No
Butter of Cow Milk	22241.01	No
Wheat and meslin flour	23110	No
Flour of Maize	23120.03	No
Rice, Milled (Husked)	23161.01	No
Raw cane or beet sugar (centrifugal only)	2351f	No
Refined sugar	23520	No
Buffalo	02112	Yes

6.9.2.2.4 Comtrade Units

The `qunit` codes found in UNSD Tariffline data are reported in this table, along with their correspondences to the standard measurement units suggested by the World Customs Organization. See the introduction section of the "Recommendation of the customs co-operation council on the use

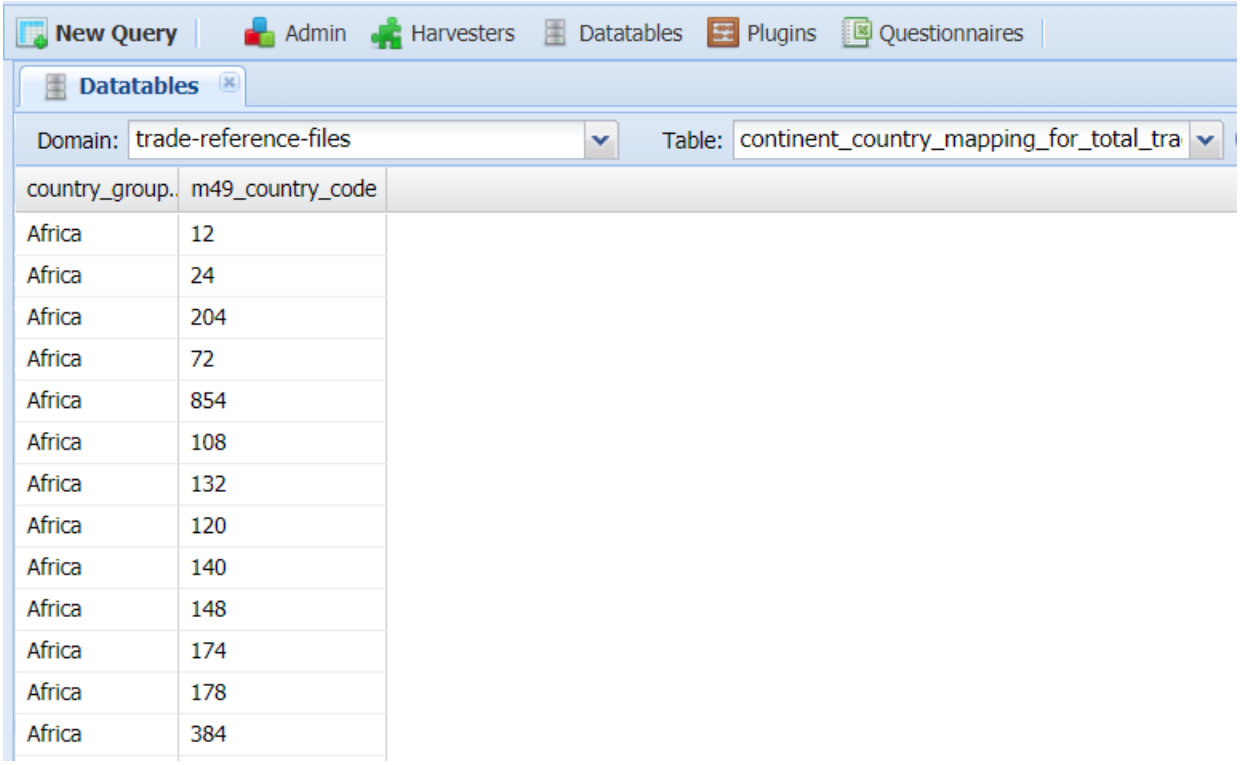
of standard units of quantity to facilitate the collection, comparison and analysis of international statistics based on the harmonized system" available at http://www.wcoomd.org/en/about-us/legal-instruments/recommendations/~/media/WCO/Public/Global/PDF/About%20us/Legal%20Instruments/Recommendations/HS/Recommendation_HS2012_UnitsOfQuantity.ashx.



Code	Symbol	Description
1	-	No Quantity
2	m²	Area in square metres
3	1000 kWh	Electrical energy in thousands of kilowatt-ho...
4	m	Length in metres
5	u	Number of items
6	2u	Number of pairs
7	l	Volume in liters
8	kg	Weight in kilograms
9	1000u	Thousands of items
10	U (jeu/pack)	Number of packages
11	12u	Dozen of items
12	m³	Volume in cubic meters
13	carat	Weight in carats

6.9.2.2.5 continent_country_mapping_for_total_trade_tables

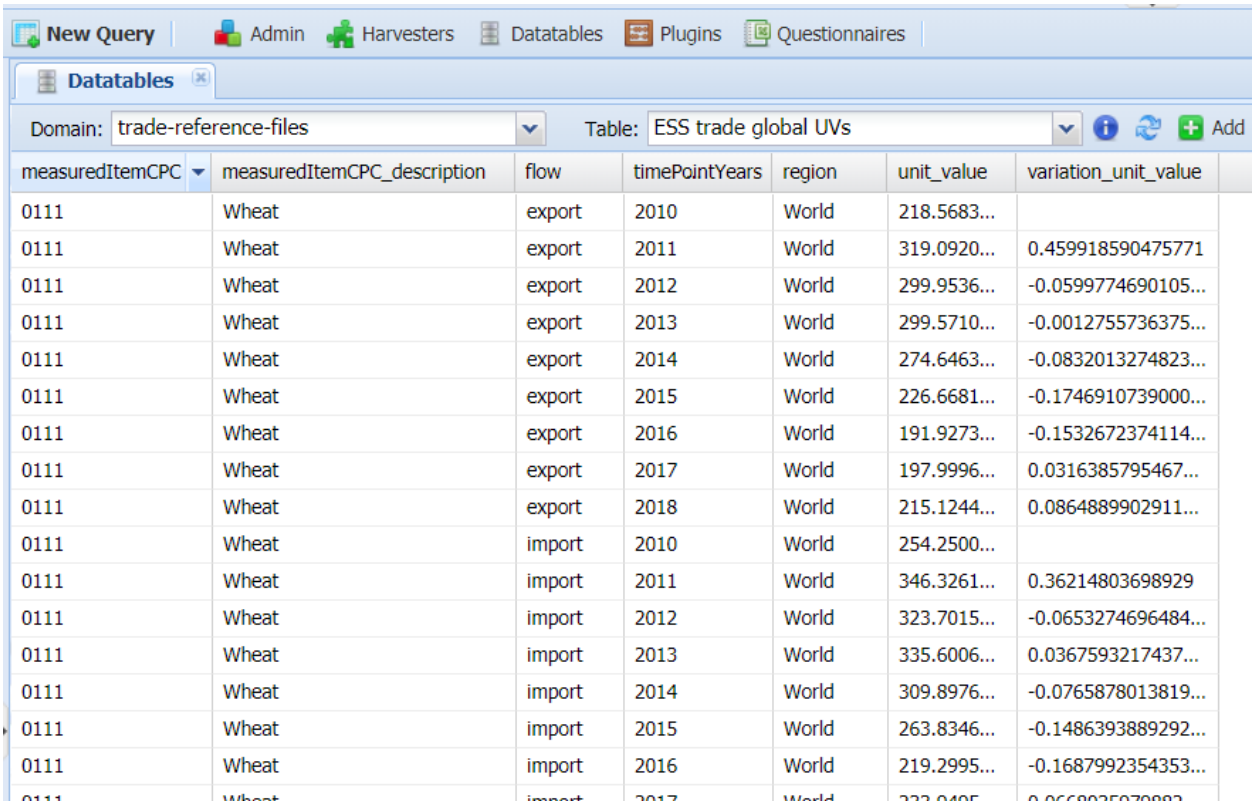
This datatable contains a list of country-continent correspondences.



country_group..	m49_country_code
Africa	12
Africa	24
Africa	204
Africa	72
Africa	854
Africa	108
Africa	132
Africa	120
Africa	140
Africa	148
Africa	174
Africa	178
Africa	384

6.9.2.2.6 ESS trade global UVs

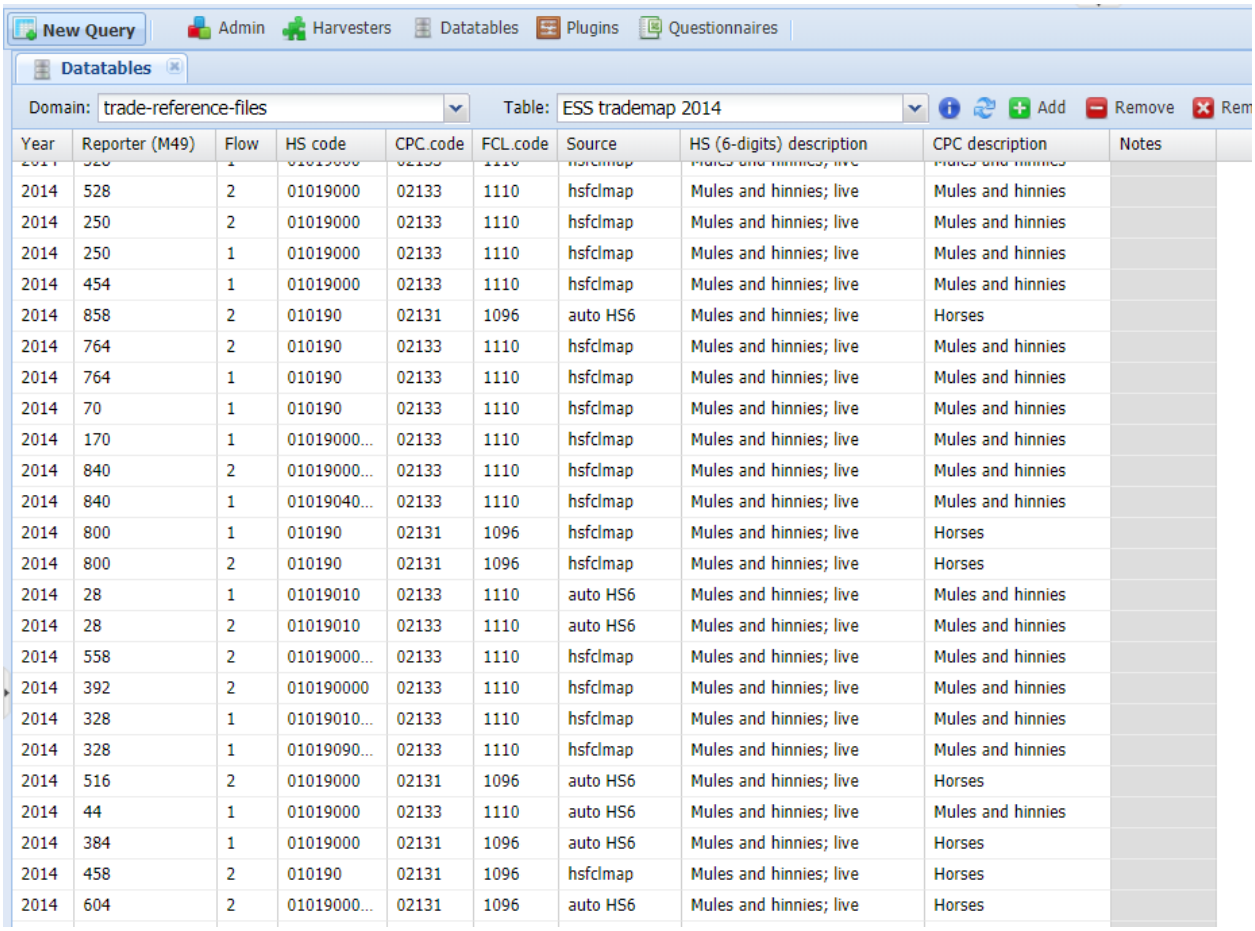
Unit values (`unit_value`) at the total trade level are calculated and saved in this datatable, along with their variation rates (`variation_unit_value`). The region the country belongs to is also reported, as this information is useful to summarize unit values by region.



measuredItemCPC	measuredItemCPC_description	flow	timePointYears	region	unit_value	variation_unit_value
0111	Wheat	export	2010	World	218.5683...	
0111	Wheat	export	2011	World	319.0920...	0.459918590475771
0111	Wheat	export	2012	World	299.9536...	-0.0599774690105...
0111	Wheat	export	2013	World	299.5710...	-0.0012755736375...
0111	Wheat	export	2014	World	274.6463...	-0.0832013274823...
0111	Wheat	export	2015	World	226.6681...	-0.1746910739000...
0111	Wheat	export	2016	World	191.9273...	-0.1532672374114...
0111	Wheat	export	2017	World	197.9996...	0.0316385795467...
0111	Wheat	export	2018	World	215.1244...	0.0864889902911...
0111	Wheat	import	2010	World	254.2500...	
0111	Wheat	import	2011	World	346.3261...	0.36214803698929
0111	Wheat	import	2012	World	323.7015...	-0.0653274696484...
0111	Wheat	import	2013	World	335.6006...	0.0367593217437...
0111	Wheat	import	2014	World	309.8976...	-0.0765878013819...
0111	Wheat	import	2015	World	263.8346...	-0.1486393889292...
0111	Wheat	import	2016	World	219.2995...	-0.1687992354353...

6.9.2.2.7 ESS trademap *YEAR*

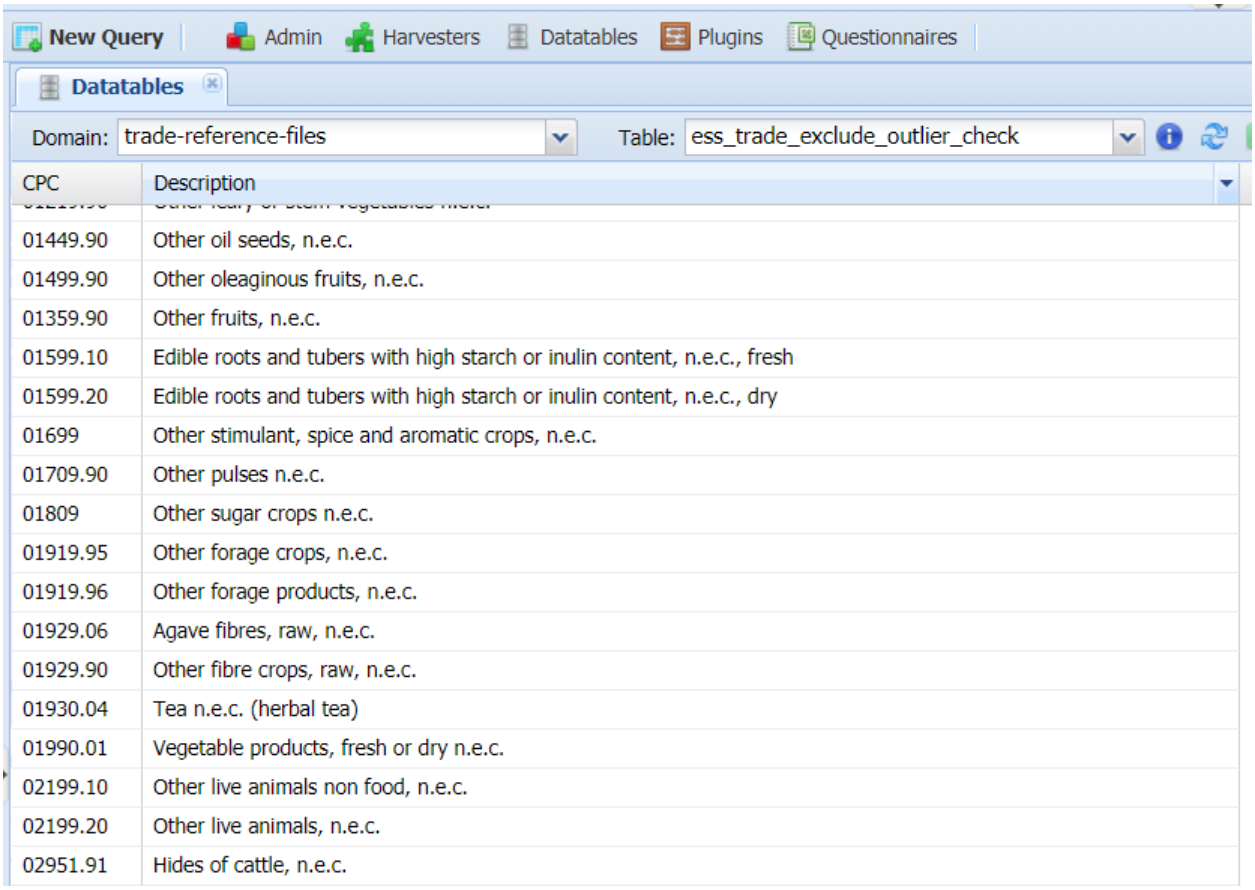
These datatbales have already been presented in the Section "HS-FCL-CPC code mapping". They contain the correpondences of country HS codes to the FCL and CPC, by *YEAR*.



Year	Reporter (M49)	Flow	HS code	CPC code	FCL code	Source	HS (6-digits) description	CPC description	Notes
2014	528	2	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	250	2	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	250	1	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	454	1	01019000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	858	2	010190	02131	1096	auto HS6	Mules and hinnies; live	Horses	
2014	764	2	010190	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	764	1	010190	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	70	1	010190	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	170	1	01019000...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	840	2	01019000...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	840	1	01019040...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	800	1	010190	02131	1096	hsfclmap	Mules and hinnies; live	Horses	
2014	800	2	010190	02131	1096	hsfclmap	Mules and hinnies; live	Horses	
2014	28	1	01019010	02133	1110	auto HS6	Mules and hinnies; live	Mules and hinnies	
2014	28	2	01019010	02133	1110	auto HS6	Mules and hinnies; live	Mules and hinnies	
2014	558	2	01019000...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	392	2	010190000	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	328	1	01019010...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	328	1	01019090...	02133	1110	hsfclmap	Mules and hinnies; live	Mules and hinnies	
2014	516	2	01019000	02131	1096	auto HS6	Mules and hinnies; live	Horses	
2014	44	1	01019000	02133	1110	auto HS6	Mules and hinnies; live	Mules and hinnies	
2014	384	1	01019000	02131	1096	auto HS6	Mules and hinnies; live	Horses	
2014	458	2	010190	02131	1096	hsfclmap	Mules and hinnies; live	Horses	
2014	604	2	01019000...	02131	1096	auto HS6	Mules and hinnies; live	Horses	

6.9.2.2.8 ess_trade_exclude_outlier_check

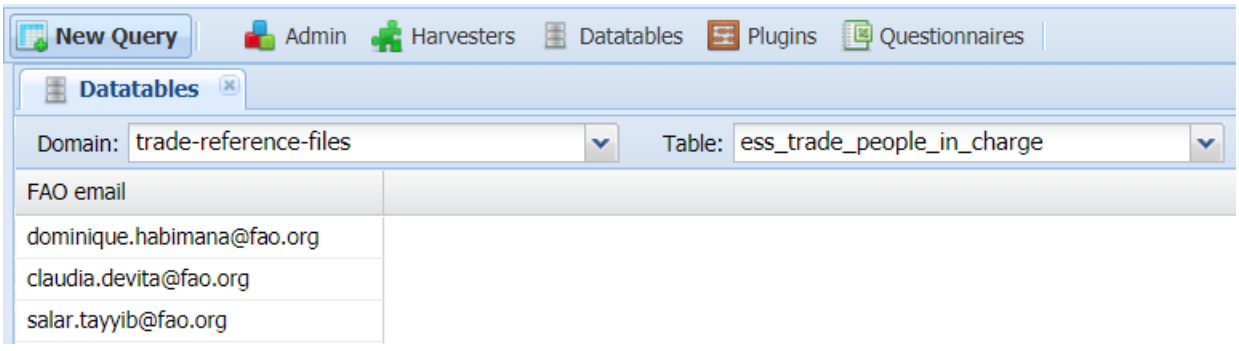
Items with volatile unit values because they are aggregation of heterogeneous items are excluded from the outlier check. The list with these items is contained in this datatable.



CPC	Description
01449.90	Other oil seeds, n.e.c.
01499.90	Other oleaginous fruits, n.e.c.
01359.90	Other fruits, n.e.c.
01599.10	Edible roots and tubers with high starch or inulin content, n.e.c., fresh
01599.20	Edible roots and tubers with high starch or inulin content, n.e.c., dry
01699	Other stimulant, spice and aromatic crops, n.e.c.
01709.90	Other pulses n.e.c.
01809	Other sugar crops n.e.c.
01919.95	Other forage crops, n.e.c.
01919.96	Other forage products, n.e.c.
01929.06	Agave fibres, raw, n.e.c.
01929.90	Other fibre crops, raw, n.e.c.
01930.04	Tea n.e.c. (herbal tea)
01990.01	Vegetable products, fresh or dry n.e.c.
02199.10	Other live animals non food, n.e.c.
02199.20	Other live animals, n.e.c.
02951.91	Hides of cattle, n.e.c.

6.9.2.2.9 ess_trade_people_in_charge

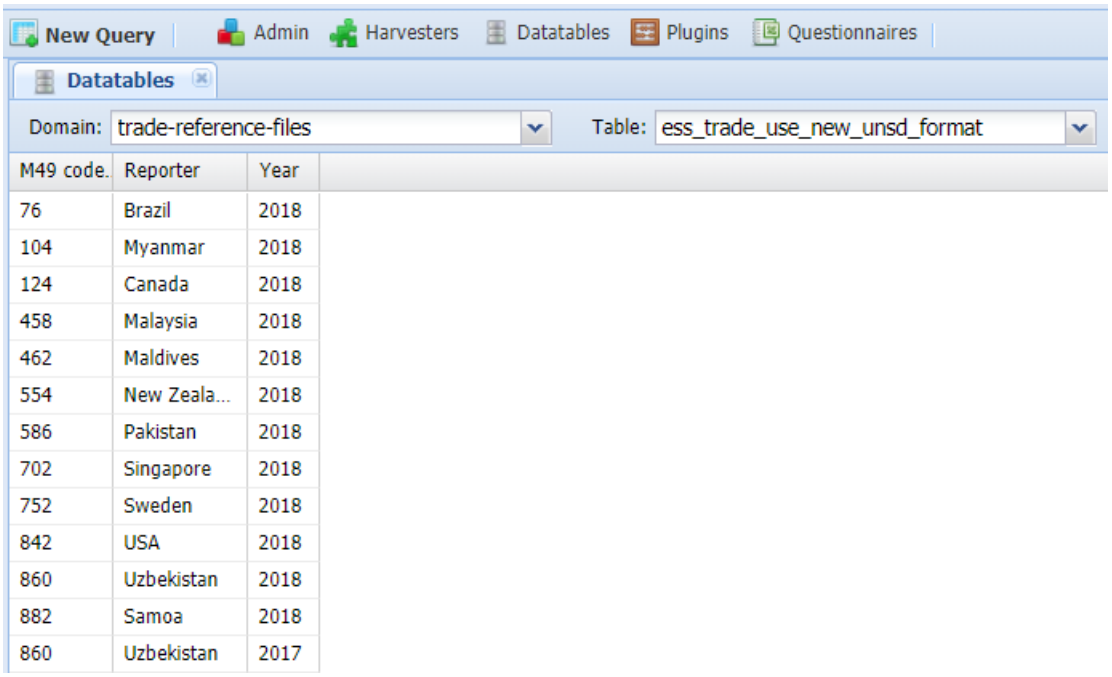
People in charge of the trade process get email from the trade plugins at various stages, mainly as confirmation of the successful operations carried out. This table contains the list of these users (and their emails).



FAO email
dominique.habimana@fao.org
claudia.devita@fao.org
salar.tayyib@fao.org

6.9.2.2.10 ess_trade_use_new_unsd_format

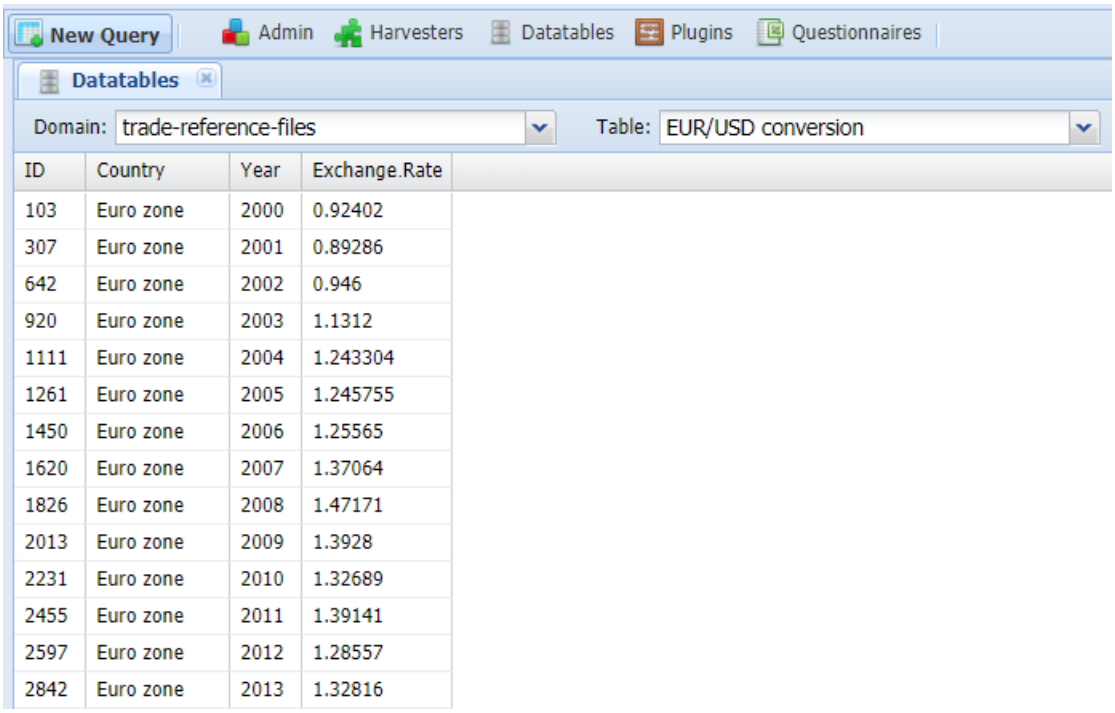
This table contains the country/year combinations for which the new UNSD data format need to be used.



M49 code.	Reporter	Year
76	Brazil	2018
104	Myanmar	2018
124	Canada	2018
458	Malaysia	2018
462	Maldives	2018
554	New Zeala...	2018
586	Pakistan	2018
702	Singapore	2018
752	Sweden	2018
842	USA	2018
860	Uzbekistan	2018
882	Samoa	2018
860	Uzbekistan	2017

6.9.2.2.11 EUR/USD conversion

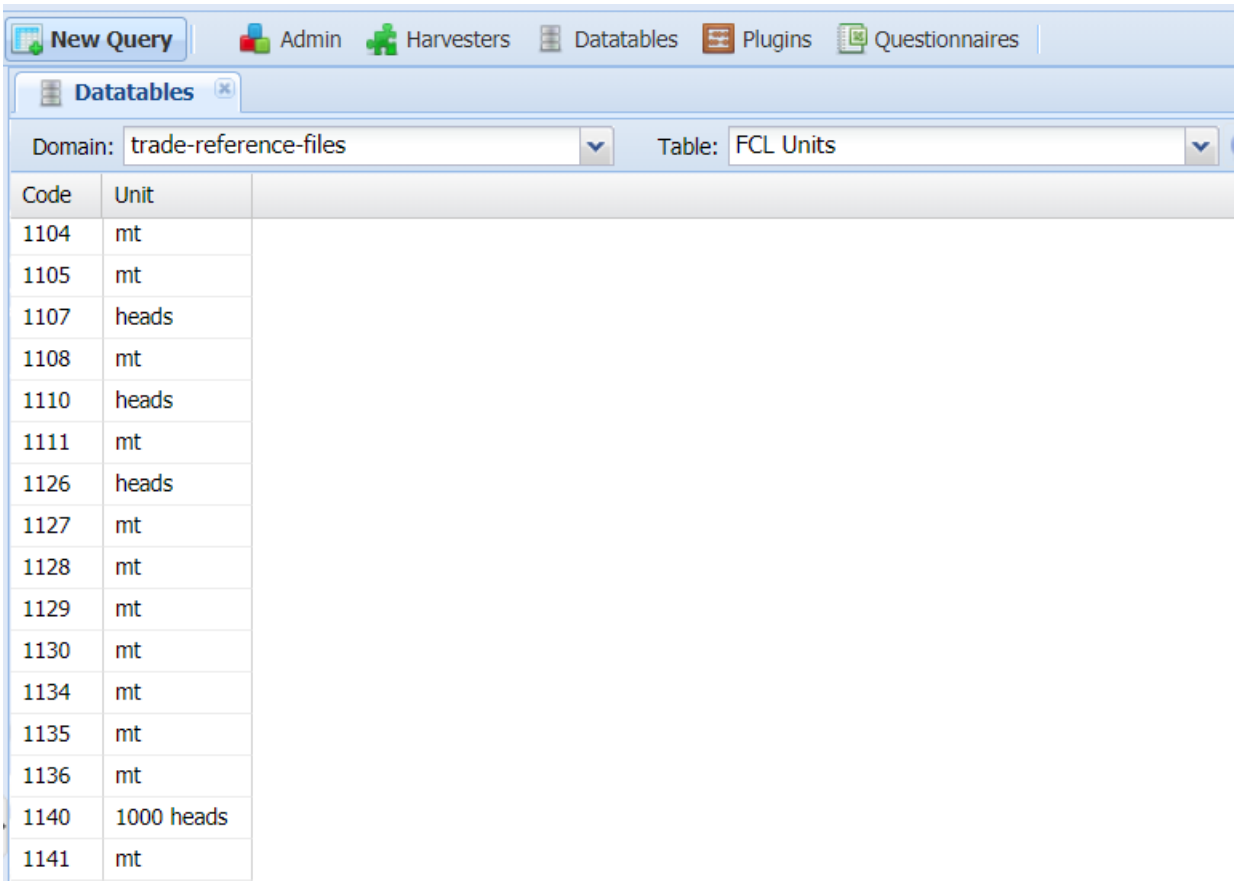
The EUR/USD average yearly nominal exchange rate is reported in this table.



ID	Country	Year	Exchange.Rate
103	Euro zone	2000	0.92402
307	Euro zone	2001	0.89286
642	Euro zone	2002	0.946
920	Euro zone	2003	1.1312
1111	Euro zone	2004	1.243304
1261	Euro zone	2005	1.245755
1450	Euro zone	2006	1.25565
1620	Euro zone	2007	1.37064
1826	Euro zone	2008	1.47171
2013	Euro zone	2009	1.3928
2231	Euro zone	2010	1.32689
2455	Euro zone	2011	1.39141
2597	Euro zone	2012	1.28557
2842	Euro zone	2013	1.32816

6.9.2.2.12 FCL Units

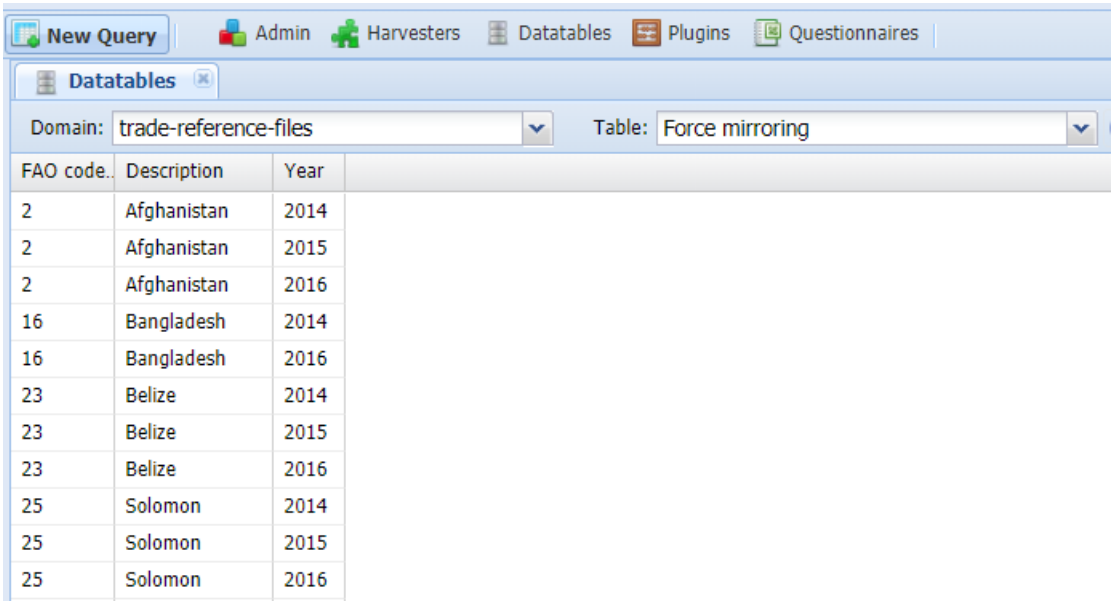
Unit of measurement for items coded with FAO codes are reported in this table. There are five types of possible units: `mt` (for metric tons), `heads`, `1000 heads`, `number`, and `$ value only` (for transaction for which no quantity is required).



Code	Unit
1104	mt
1105	mt
1107	heads
1108	mt
1110	heads
1111	mt
1126	heads
1127	mt
1128	mt
1129	mt
1130	mt
1134	mt
1135	mt
1136	mt
1140	1000 heads
1141	mt

6.9.2.2.13 Force mirroring

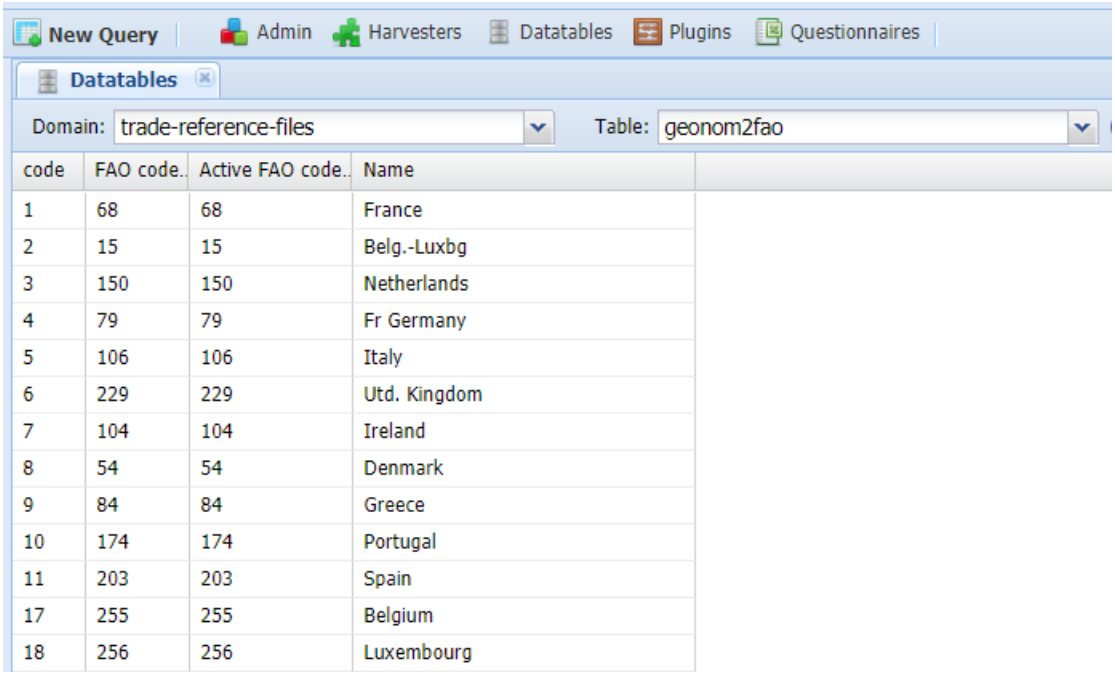
This table contains countries for which the mirroring mechanism is applied, even though they appear as reporters in raw data. The main reason for which a country appears in this list is because data has been considered of low quality or incomplete.



FAO code..	Description	Year
2	Afghanistan	2014
2	Afghanistan	2015
2	Afghanistan	2016
16	Bangladesh	2014
16	Bangladesh	2016
23	Belize	2014
23	Belize	2015
23	Belize	2016
25	Solomon	2014
25	Solomon	2015
25	Solomon	2016

6.9.2.2.14 geonom2fao

The correspondences betwee FAO area codes and geonomenclature (Eurostat country codes) are reported in this table.



code	FAO code..	Active FAO code..	Name
1	68	68	France
2	15	15	Belg.-Luxbg
3	150	150	Netherlands
4	79	79	Fr Germany
5	106	106	Italy
6	229	229	Utd. Kingdom
7	104	104	Ireland
8	54	54	Denmark
9	84	84	Greece
10	174	174	Portugal
11	203	203	Spain
17	255	255	Belgium
18	256	256	Luxembourg

6.9.2.2.15 HS to FCL mapping (old)

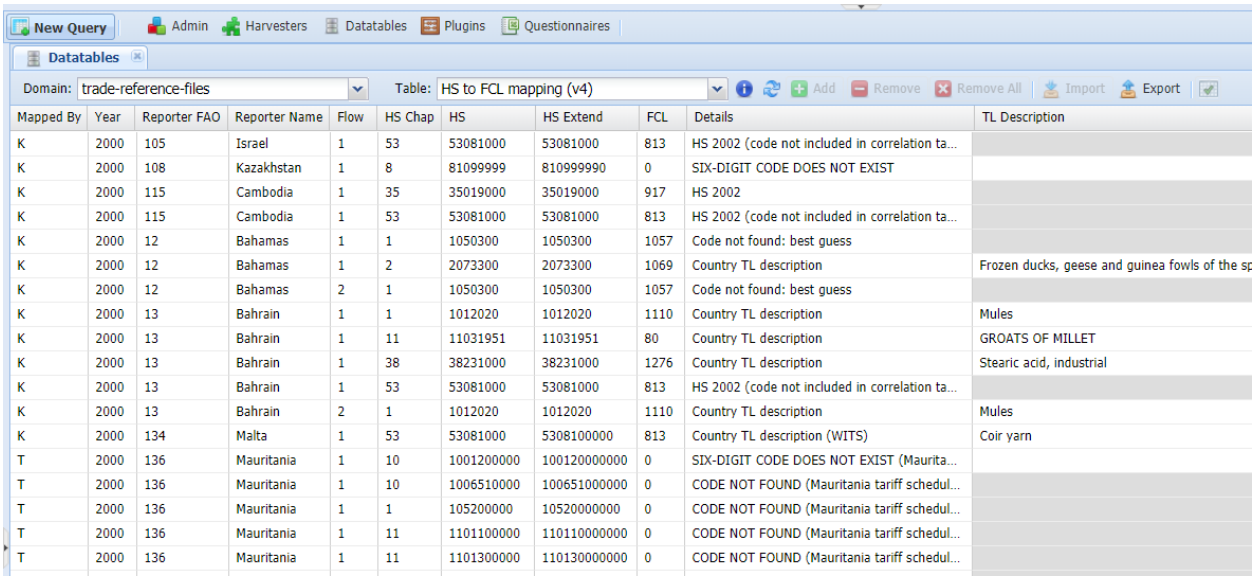
This table can be ignored: it is a previous HS-FCL mapping table, kept for reference. See “HS to FCL mapping (v5)”.

6.9.2.2.16 HS to FCL mapping (v3)

This table can be ignored: it is a previous HS-FCL mapping table, kept for reference. See “HS to FCL mapping (v5)”.

6.9.2.2.17 HS to FCL mapping (v4)

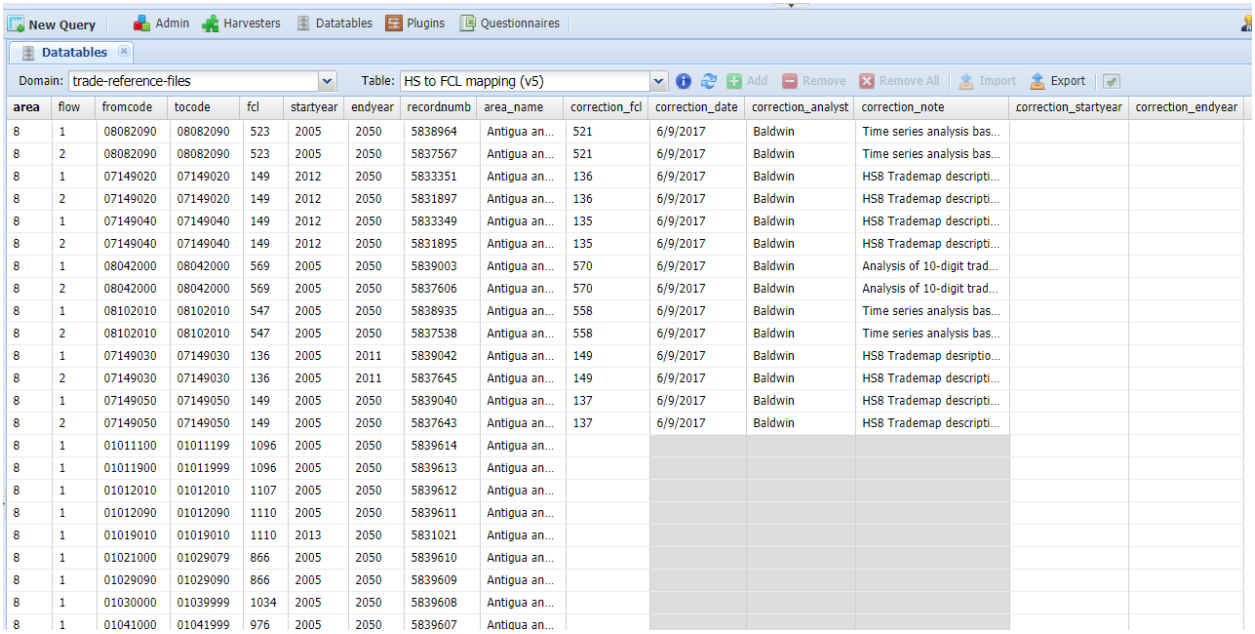
Additions to the “HS to FCL mapping (v5)” table (see next table and the“HS-FCL-CPC code mapping” Section) are reported in this table.



Mapped By	Year	Reporter FAO	Reporter Name	Flow	HS Chap	HS	HS Extend	FCL	Details	TL Description
K	2000	105	Israel	1	53	53081000	53081000	813	HS 2002 (code not included in correlation ta...	
K	2000	108	Kazakhstan	1	8	81099999	810999990	0	SIX-DIGIT CODE DOES NOT EXIST	
K	2000	115	Cambodia	1	35	35019000	35019000	917	HS 2002	
K	2000	115	Cambodia	1	53	53081000	53081000	813	HS 2002 (code not included in correlation ta...	
K	2000	12	Bahamas	1	1	1050300	1050300	1057	Code not found: best guess	
K	2000	12	Bahamas	1	2	2073300	2073300	1069	Country TL description	Frozen ducks, geese and guinea fowls of the sp
K	2000	12	Bahamas	2	1	1050300	1050300	1057	Code not found: best guess	
K	2000	13	Bahrain	1	1	1012020	1012020	1110	Country TL description	Mules
K	2000	13	Bahrain	1	11	11031951	11031951	80	Country TL description	GROATS OF MILLET
K	2000	13	Bahrain	1	38	38231000	38231000	1276	Country TL description	Stearic acid, industrial
K	2000	13	Bahrain	1	53	53081000	53081000	813	HS 2002 (code not included in correlation ta...	
K	2000	13	Bahrain	2	1	1012020	1012020	1110	Country TL description	Mules
K	2000	134	Malta	1	53	53081000	5308100000	813	Country TL description (WITS)	Coir yarn
T	2000	136	Mauritania	1	10	1001200000	100120000000	0	SIX-DIGIT CODE DOES NOT EXIST (Maurita...	
T	2000	136	Mauritania	1	10	1006510000	100651000000	0	CODE NOT FOUND (Mauritania tariff schedul...	
T	2000	136	Mauritania	1	1	105200000	10520000000	0	CODE NOT FOUND (Mauritania tariff schedul...	
T	2000	136	Mauritania	1	11	1101100000	110110000000	0	CODE NOT FOUND (Mauritania tariff schedul...	
T	2000	136	Mauritania	1	11	1101300000	110130000000	0	CODE NOT FOUND (Mauritania tariff schedul...	

6.9.2.2.18 HS to FCL mapping (v5)

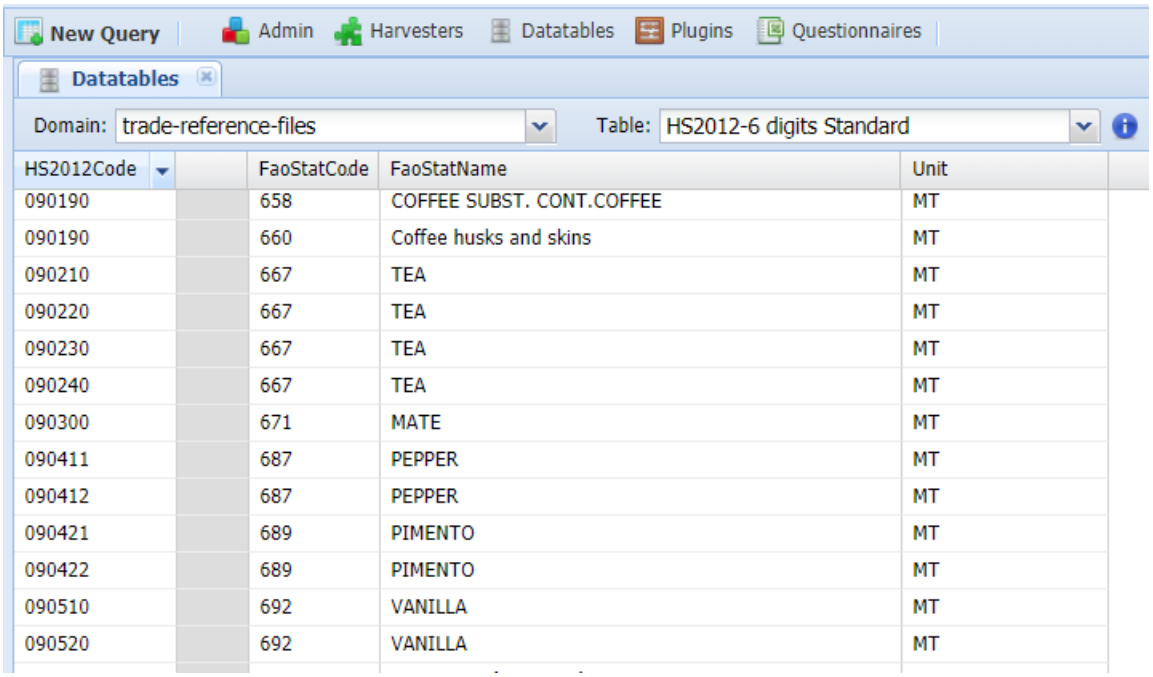
The contents of this table were already shown in the "HS-FCL-CPC code mapping" Section. It contains the latest version of the "historical" HS to FCL correspondence table. Any correction to this table is done in the "correction_X" column, where X is the variabole that needs to be corrected (e.g., "correction_startdate" will correct the `startdate` variable).



area	flow	fromcode	tocode	fcl	startyear	endyear	recordnumb	area_name	correction_fcl	correction_date	correction_analyst	correction_note	correction_startyear	correction_endyear
8	1	08082090	08082090	523	2005	2050	5838964	Antigua an...	521	6/9/2017	Baldwin	Time series analysis bas...		
8	2	08082090	08082090	523	2005	2050	5837567	Antigua an...	521	6/9/2017	Baldwin	Time series analysis bas...		
8	1	07149020	07149020	149	2012	2050	5833351	Antigua an...	136	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	2	07149020	07149020	149	2012	2050	5831897	Antigua an...	136	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	07149040	07149040	149	2012	2050	5833349	Antigua an...	135	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	2	07149040	07149040	149	2012	2050	5831895	Antigua an...	135	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	08042000	08042000	569	2005	2050	5839003	Antigua an...	570	6/9/2017	Baldwin	Analysis of 10-digit trad...		
8	2	08042000	08042000	569	2005	2050	5837606	Antigua an...	570	6/9/2017	Baldwin	Analysis of 10-digit trad...		
8	1	08102010	08102010	547	2005	2050	5838935	Antigua an...	558	6/9/2017	Baldwin	Time series analysis bas...		
8	2	08102010	08102010	547	2005	2050	5837538	Antigua an...	558	6/9/2017	Baldwin	Time series analysis bas...		
8	1	07149030	07149030	136	2005	2011	5839042	Antigua an...	149	6/9/2017	Baldwin	HS8 Trademap desriptio...		
8	2	07149030	07149030	136	2005	2011	5837645	Antigua an...	149	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	07149050	07149050	149	2005	2050	5839040	Antigua an...	137	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	2	07149050	07149050	149	2005	2050	5837643	Antigua an...	137	6/9/2017	Baldwin	HS8 Trademap descripti...		
8	1	01011100	01011199	1096	2005	2050	5839614	Antigua an...						
8	1	01011900	01011999	1096	2005	2050	5839613	Antigua an...						
8	1	01012010	01012010	1107	2005	2050	5839612	Antigua an...						
8	1	01012090	01012090	1110	2005	2050	5839611	Antigua an...						
8	1	01019010	01019010	1110	2013	2050	5831021	Antigua an...						
8	1	01021000	01029079	866	2005	2050	5839610	Antigua an...						
8	1	01029090	01029090	866	2005	2050	5839609	Antigua an...						
8	1	01030000	01039999	1034	2005	2050	5839608	Antigua an...						
8	1	01041000	01041999	976	2005	2050	5839607	Antigua an...						

6.9.2.2.19 HS2012-6 digits Standard

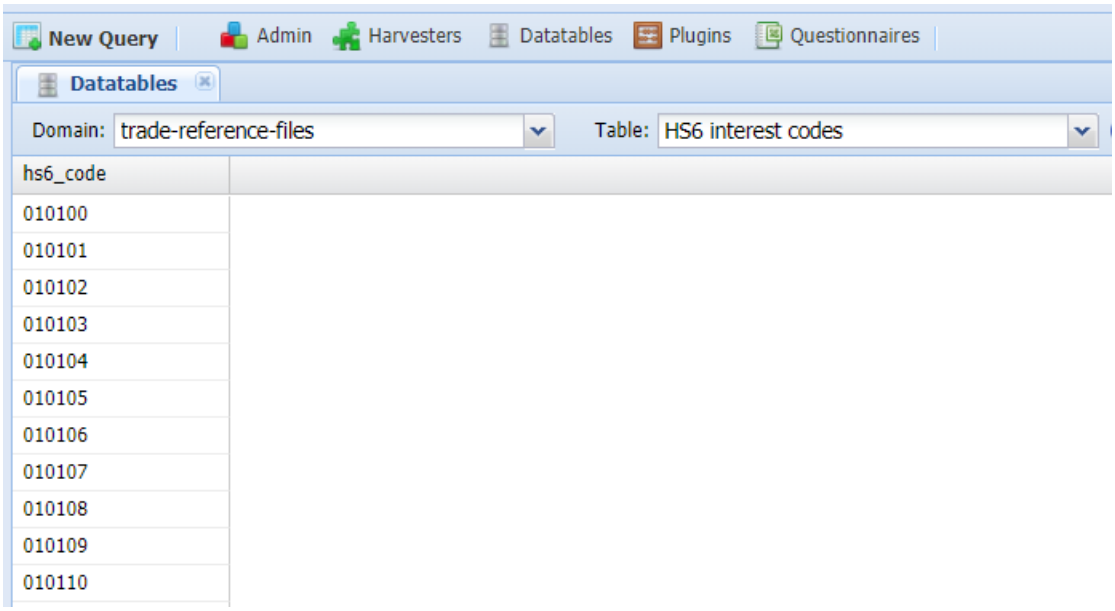
This table contains the correspondences of HS codes at 6 digits to FCL codes. Note that these correspondences are not necessary 1-to-1. Indeed, there are 413 (67%) HS codes that map 1-to-1, while 202 (33%) map many-to-1 (i.e., at least two different HS codes assigned to 1 FCL code).



HS2012Code	FaoStatCode	FaoStatName	Unit
090190	658	COFFEE SUBST. CONT.COFFEE	MT
090190	660	Coffee husks and skins	MT
090210	667	TEA	MT
090220	667	TEA	MT
090230	667	TEA	MT
090240	667	TEA	MT
090300	671	MATE	MT
090411	687	PEPPER	MT
090412	687	PEPPER	MT
090421	689	PIMENTO	MT
090422	689	PIMENTO	MT
090510	692	VANILLA	MT
090520	692	VANILLA	MT

6.9.2.2.20 HS6 interest codes

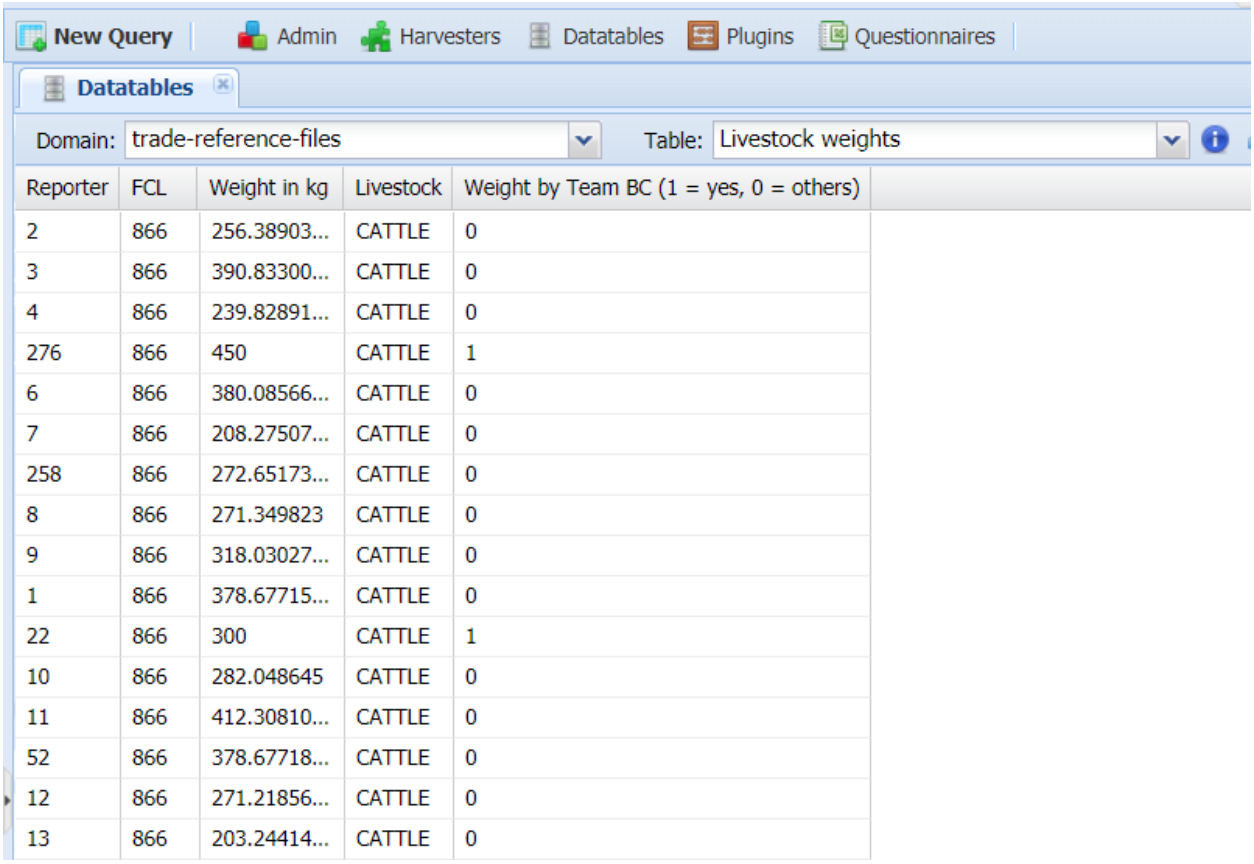
The list of HS codes at 6 digits that are of interest to the processing of the trade process are reported in this table. After the first filtering by chapters, this table is used to further subset the raw data that need to be kept and processed. See Section "Data harvesting".



hs6_code
010100
010101
010102
010103
010104
010105
010106
010107
010108
010109
010110

6.9.2.2.21 Livestock weights

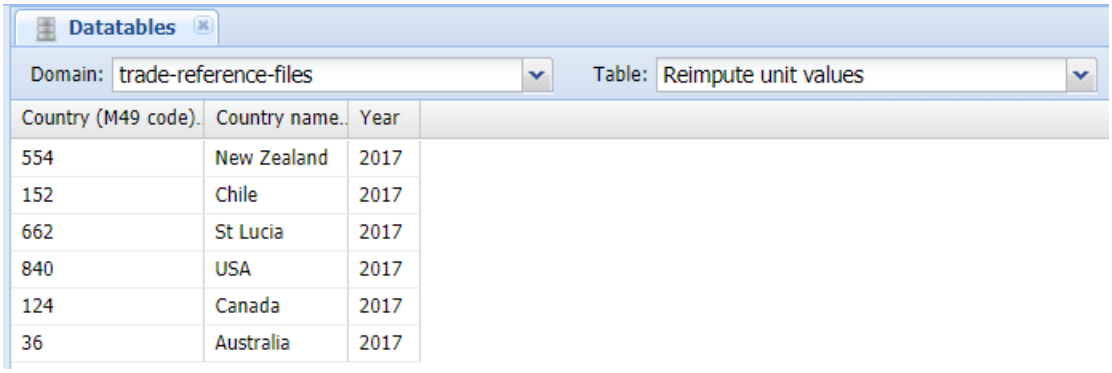
Average livestock weight in kilos is reported in this table. It is used by the "Complete Trade Flow CPC" plugin, when raw data has no information about heads (or 1,000 heads) but give weight in kilos. Dividing reported kilos by the average weight in this tables gives the number of animals.



Reporter	FCL	Weight in kg	Livestock	Weight by Team BC (1 = yes, 0 = others)
2	866	256.38903...	CATTLE	0
3	866	390.83300...	CATTLE	0
4	866	239.82891...	CATTLE	0
276	866	450	CATTLE	1
6	866	380.08566...	CATTLE	0
7	866	208.27507...	CATTLE	0
258	866	272.65173...	CATTLE	0
8	866	271.349823	CATTLE	0
9	866	318.03027...	CATTLE	0
1	866	378.67715...	CATTLE	0
22	866	300	CATTLE	1
10	866	282.048645	CATTLE	0
11	866	412.30810...	CATTLE	0
52	866	378.67718...	CATTLE	0
12	866	271.21856...	CATTLE	0
13	866	203.24414...	CATTLE	0

6.9.2.2.22 Reimpute unit values

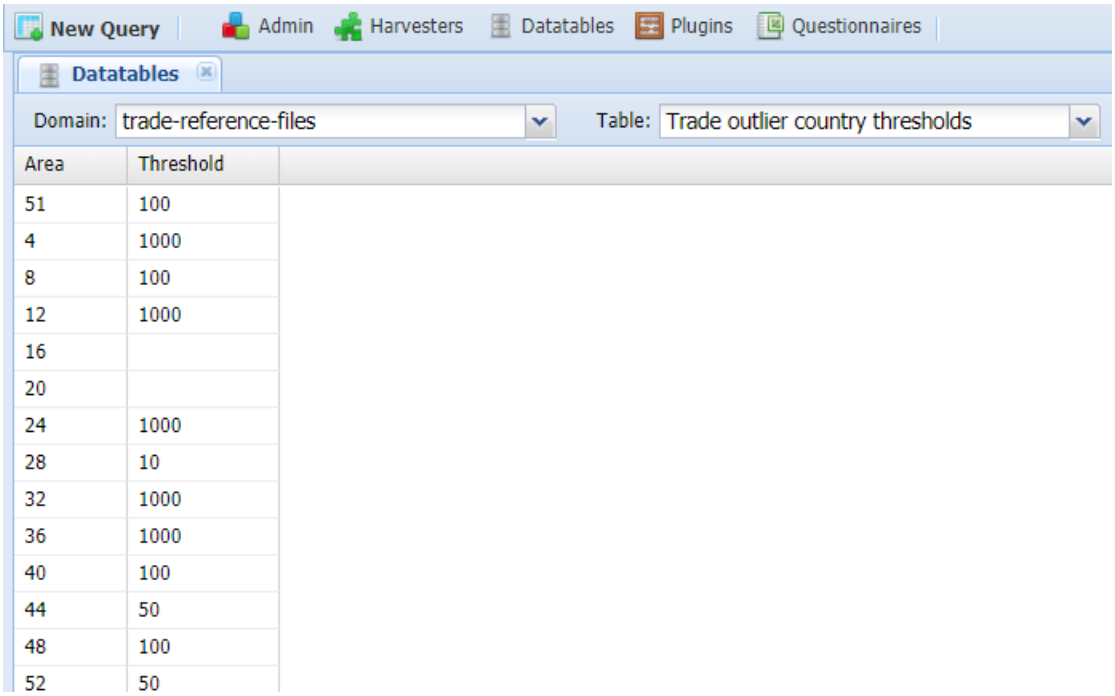
This table contains a list of countries for whith the "re-imputation" of unit values need to be done. For a description of what "re-imputation" is, please see the "Re-imputation of flows" sub-section.



Country (M49 code).	Country name.	Year
554	New Zealand	2017
152	Chile	2017
662	St Lucia	2017
840	USA	2017
124	Canada	2017
36	Australia	2017

6.9.2.2.23 Trade outlier country thresholds

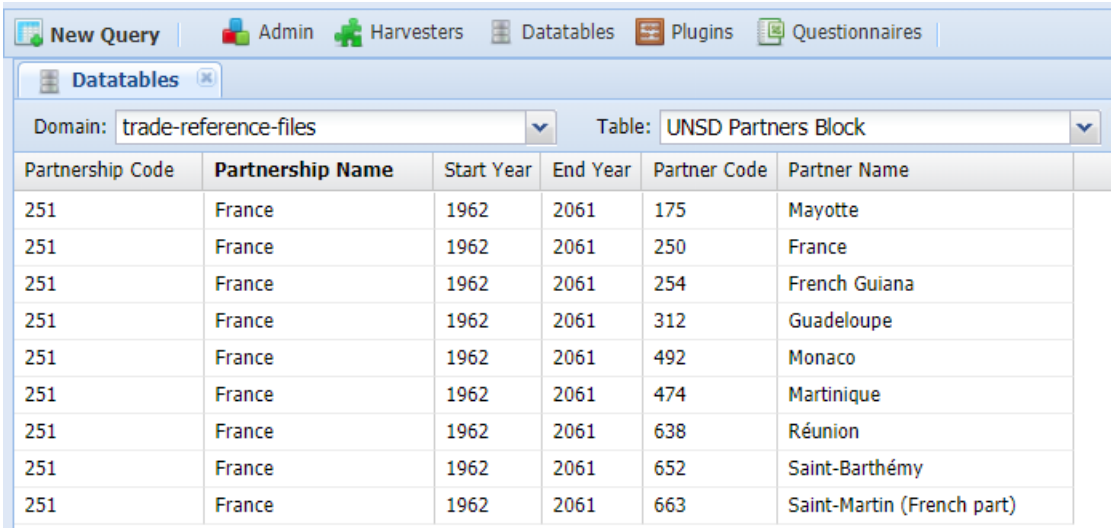
Outliers routines are applied for commodities that exceed some quantity threshold (so to avoid analysing very small quantities). This depends of the country. This table reports these thresholds (countries with no threshold will be assigned the default threshold of 1,000). Country codes are M49 codes.



Area	Threshold
51	100
4	1000
8	100
12	1000
16	
20	
24	1000
28	10
32	1000
36	1000
40	100
44	50
48	100
52	50

6.9.2.2.24 UNSD Partners Block

Tariff line M49 codes (which are different from official M49) are converted in FAO country codes using this conversion table, provided by Team ENV. In a subsequent step these M49 codes are converted to FAO codes.



Partnership Code	Partnership Name	Start Year	End Year	Partner Code	Partner Name
251	France	1962	2061	175	Mayotte
251	France	1962	2061	250	France
251	France	1962	2061	254	French Guiana
251	France	1962	2061	312	Guadeloupe
251	France	1962	2061	492	Monaco
251	France	1962	2061	474	Martinique
251	France	1962	2061	638	Réunion
251	France	1962	2061	652	Saint-Barthémy
251	France	1962	2061	663	Saint-Martin (French part)

6.9.2.3 `trade-reports`

The datatables contained in this domain have already been presented in the "Pre-Processing Report plugin" Section. These are:

- "Check qty and value included"
- "Import and export content check"
- "Missing data by report"
- "Non-reporting countries"

- "Number records by reporter/year"
- "Reporters by years"

There other datatables with the name corresponding to the previous one, but suffixed "OLD": these were used in a first version of the "Pre-Processing Report" plugin, and are stored for reference. Finally, there are a couple of datatables that were used to keep a log of the plugins runs ("Info runs: complete trade" and "Info runs: total trade"), but a filesystem log mechanism was developed, which contains additional information with respect to tsaid datatables.

6.10 R server

The following output shows the results of `sessionInfo()` obtained from a plugin that does not explicitly load any package, in the current (as of June 2020) R server in QA:

```
R version 3.1.2 (2014-10-31)
Platform: x86_64-unknown-linux-gnu (64-bit)

locale:
[1] LC_CTYPE=en_US.UTF-8 LC_NUMERIC=C
[3] LC_TIME=en_US.UTF-8 LC_COLLATE=en_US.UTF-8
[5] LC_MONETARY=en_US.UTF-8 LC_MESSAGES=en_US.UTF-8
[7] LC_PAPER=en_US.UTF-8 LC_NAME=C
[9] LC_ADDRESS=C LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] stats graphics grDevices utils datasets methods base

other attached packages:
[1] faosws_0.9.7

loaded via a namespace (and not attached):
[1] bit_1.1-12 bit64_0.9-5 bitops_1.0-6 chron_2.3-45
[5] curl_2.6 data.table_1.9.6 RCurl_1.95-4.8 RJSONIO_1.3-0
```

6.11 Specific plugins-related information

6.11.1 Standard vs. core plugins

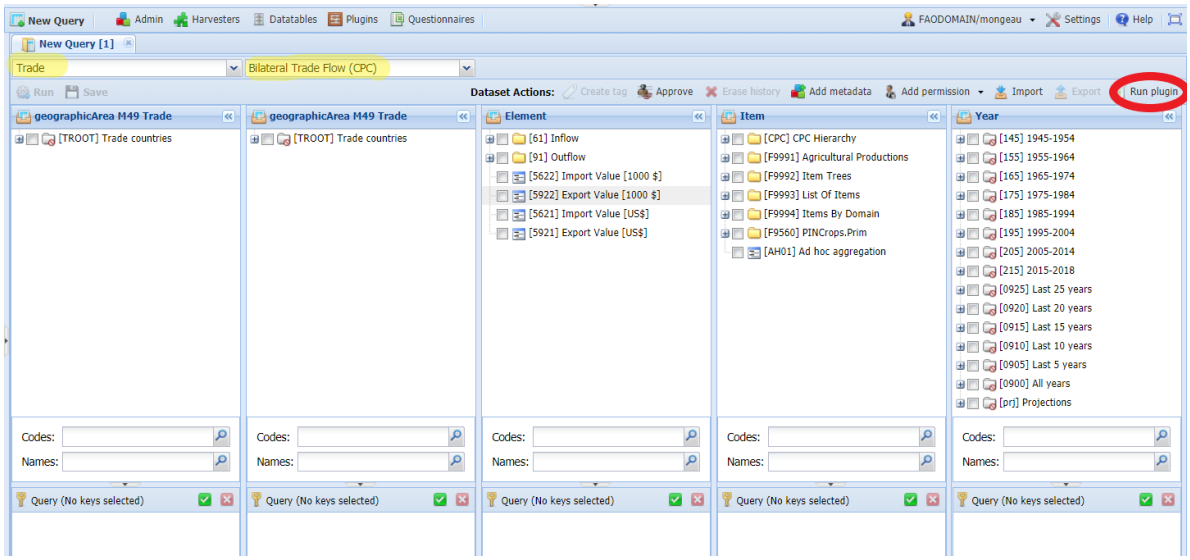
The main difference between standard and core plugins is that the former write data to a user's session, who will then need to save manually the modifications to the dataset, while core plugins write directly to the dataset, with no manual intervention required from the user.

6.11.2 Plugin scheduling

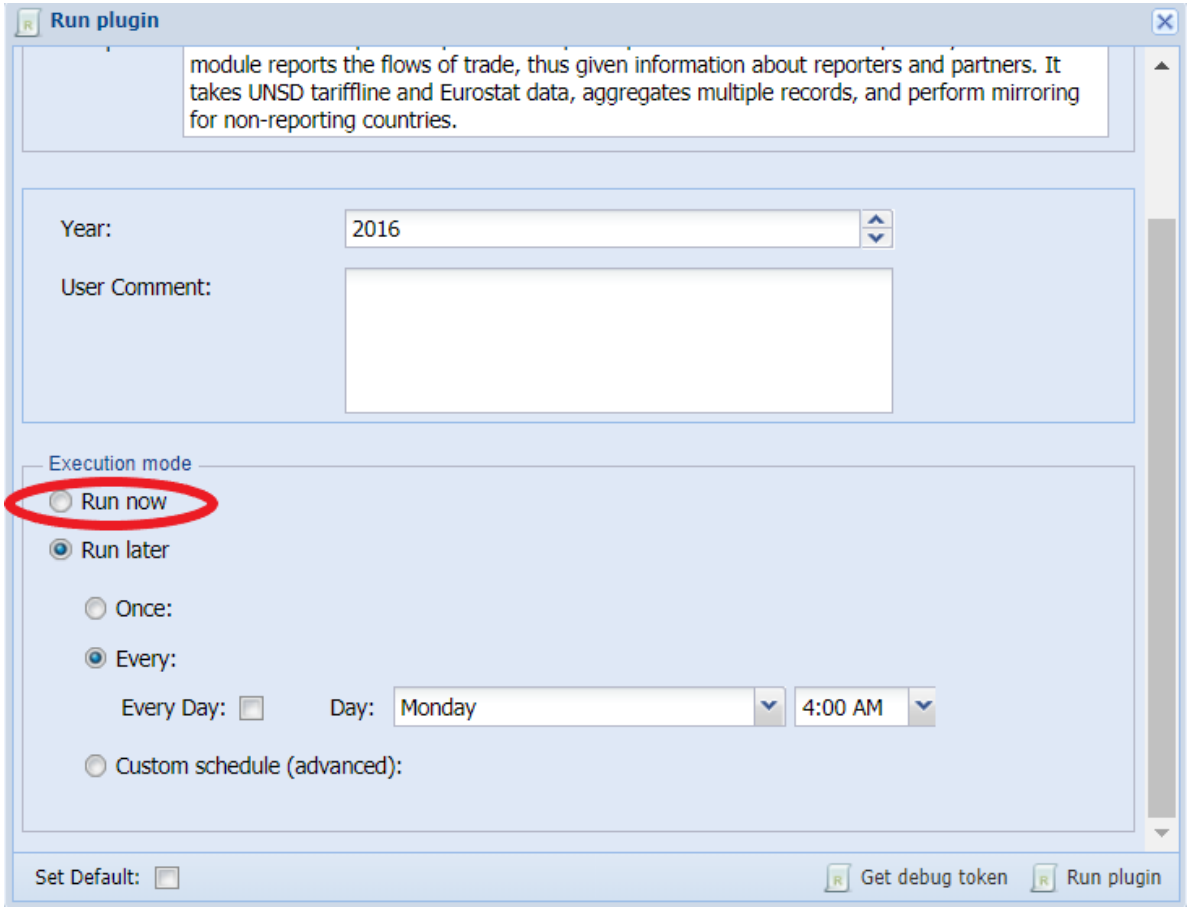
Plugins can be set to run on a specific time in the future. This is called "plugin scheduling". In the context of the trade processing, usually the "Complete Trade Flow CPC" and "Total Trade CPC" plugins are scheduled to run overnight for the years the validation is being carried out.

Below, the steps for scheduling a plugin are explained, by using the case of the "Complete Trade Flow CPC" plugin.

1. Create a new query to the appropriate dataset (i.e., where the plugin runs; in the example it is the "Bilateral Trade Flow (CPC)" dataset of the "Trade" domain) and click on "Run plugin" (notice that no dimension is specified):



2. After selecting the plugin and entering eventual parameters, click on the "Run later option" available in the "Execution mode" section:



3. Choose one of the scheduling options available, namely:

- *Once*: the plugin will run only once at the selected day/time.
- *Every*: the plugin will run each day at the specific time (in the screenshot, it will run each Monday at 4:00 a.m.); if the "Every day" option is selected, the plugin will run each day at the required time (this is how the trade plugins are set to run, every day at a specific time).
- *Custom schedule (advanced)*: this is a more advanced schedulin specification, that allows to set seconds, minutes, hours, day of week, day of month, year (as this option is selected a help page can be consulted by clicking on the "Help" button).

Note that in order to avoid that for any given circumstance that can make a plugin fail, both the "Complete Trade Flow CPC" and "Total Trade CPC" plugins are scheduled to run **twice** per night.

Appendix: Accessing/updating files in the R shared drive

As mentioned in the "Raw data: SWS datatables vs. RDS files" Section, the "Complete Trade Flow CPC" reads raw data from RDS files stored in the *R shared drive*, which is a folder in the SWS server where R plugins can write and read data to/from, and can be accessed also by local PCs in the FAO network by means of a Samba service.

It can be accessed by PCs inside the FAO network at the following addresses:

- QA: \\hqlprsws1.hq.un.fao.org\sws_r_share
- LIVE: \\hqlprsws2.hq.un.fao.org\sws_r_share

The location in the SWS server, where R can read from and write to, is in both cases `/work/SWS_R_Share`.

In what follows, the R shared drive folder will be referenced to simply as "SHARE": the user will need to modify that accordingly if he/she is using a local PC or the SWS server itself.

The UNSD and Eurostat raw tables (for years 2014-2018) are stored in the `SHARE/trade/datatables` folder. For instance, the UNSD raw data file form 2018 is available as:

```
SHARE/trade/datatables/ct_tariffline_unlogged_2018.rds
```

Notice that the name of the file is the same as the corresponding SWS datatable, except the ".rds" extension.

If the "Complete Trade Flow CPC" finds a raw data table in the R shared drive, it will use it, otherwise it will fallback to reading the SWS datatable. For instance, if 2019 data has not been already saved as an RDS file in the shared drive, the plugin will then execute the command:

```
faosws::ReadDatatable("ct_tariffline_unlogged_2019")
```

while if the data was previously saved, it would execute:

```
readRDS("SHARE/trade/datatables/ct_tariffline_unlogged_2019.rds")
```

The latter brings down the time spent to read the data by a factor of 20 times, reducing the time required to load UNSD and Eurostat raw data from 40 minutes less than 2.

The drawback of using the readRDS() method is that the person in charge of the trade-process need to keep the RDS files updated with respect to any modification occurred to the correspondent SWS datatables. However, given that raw data files are not changed frequently, it is suggested to keep using the readRDS() method, as the performance gained with it outweighs the need to do these infrequent synchnronisations.

The following R instructions will download the SWS datatable and will save it to the R shared drive (the examples will use the "ct_tariffline_unlogged_2019" datatable and "SHARE" as name for the shared drive, change both accordingly):

```
library(faosws)

trade_data <- ReadDatatable("ct_tariffline_unlogged_2019")

saveRDS(trade_data, "SHARE/trade/datatables/ct_tariffline_unlogged_2019.rds")
```

Disclaimer

This Working Paper should not be reported as representing the official view of the FAO. The views expressed in this Working Paper are those of the author and do not necessarily represent those of the FAO or FAO policy. Working Papers describe research in progress by the authors and are published to elicit comments and to further discussion.