faoswsTrade: `complete_tf_cpc` and `total_trade_CPC` modules

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

17 November 2017

Abstract

The trade module is divided in two submodules: `complete_tf_cpc` and `total_trade_CPC`. The `complete_tf_cpc` module uses the most disaggregated bilateral transactions between countries (obtained from Eurostat for European countries, and from UNSD for the remaining countries), filters the agriculture-related flows of interest for the Organisation, standarises units of measurement, maps item and country codes to a standard set of codes, imputes missing quantities, and computes unit values. The `total_trade_CPC` aggregates bilateral trade (i.e., the results of `complete_tf_cpc`) at reporter level. Each module is year specific, i.e, the trade modules run independently for each year (not necessarily in a chronological order). To run the `total_trade_CPC` module, the output of the `complete_tf_cpc` module is required. In this document, an overall narrative of what the modules do is presented. A detailed step-by-step illustration of the modules is given in a separate document.

1 Complete tf cpc

1.1 Data

Raw data are harvested and provided by the SWS Team (subunit of Team F) for both UNSD Tariff line and Eurostat Data. The dataset is prefiltered by downloading only chapters of interest, which are:

```
01, 02, 03, 04, 05, 06, 07, 08, 09, 10, 11, 12, 13, 14, 15, 16, 17,
18, 19, 20, 21, 22, 23, 24, 33, 35, 38, 40, 41, 43, 50, 51, 52, 53.
```

In the future, if other commodities become of interest for the division, it is important to include additional chapters in the first step of the downloading.

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

year	reporter	partner	flow	hs	value	weight	qty	qunit	chapter

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

Besides filtering by chapters, only some HS codes inside the chapters are considered. These codes are given in the `hs6faointerest` datatable:

Table 3: HS-6 codes of interest

hs6_code
180520
020576
090838
110243
150379
530211
020892
210546
010525
510425

Both Eurostat and Tariff line data are given the same variable names and data types before being processed further.

1.1.1 Eurostat

- only `stat_regime` equal to 4 is kept.

  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.

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

- only numeric codes of reporters and partners are kept (letters are not allowed; basically this removes the "EU" total).

- only numeric CN8 codes (`hs`) are kept (letters are not allowed).

1.1.2 UNSD

- only numeric HS (`hs`) codes are kept (letters are not allowed).

1.2 Process

1.2.1 Aggregate UNSD Tariff line individual Shipments

The tariffline data from UNSD contains multiple rows with identical combination of `reporter` / `partner` / `commodity` / `flow` / `year` / `qunit`. Those are transactions registered separately, thus rows containinig non-missing values and quantities can be aggregated. Missing variables of the same type are also aggregated if they are *all* missing, as they will produce a missing aggregated value for missing disaggregated values while correctly summing the remaining variables.

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

The results of the aggregation of the previous example are shown in the following table. In this case, even if there is a unique combination of `reporter` / `partner` / `flow` / `year` / `hs`, the rows are two: indeed, one aggregated the transactions where the `weight` variable was not missing, and the other aggregated those cases where it was missing.

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

After quantity conversions are made or missing quantities get imputed (see further) the two different rows in the previous table in the table will be aggregated into a single transaction.

1.2.2 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.

1.2.2.1 UNSD Tariff line

1. UNSD Tariff line data reports area code with Tariff line M49 standard codes (which are different for official M49). The area code is converted in FAO country code using a specific conversion table provided by Team ENV.

Table 6: Subset of the M40 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

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).

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

code	description	selectionOnly	type	startDate	endDate
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

3. European countries (as reporters) already in Eurostat data are removed.

4. Area codes not mapping to any FAO country code are separately saved and removed from further analyses. All countries mapping to code 252 (which correponds to undefined areas) are mapped to the 896 M49 code ("Other nei").

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

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. Set (extend if required) all HS codes to the maximum length (by reporter / flow) found in the HS-FCL mapping table valid for the reporter in that year (see below).

7. Commodity codes are reported in HS codes (*Harmonized Commodity Description and Coding System*). The codes are converted in FCL (*FAO Commodity List*) codes. This step is performed using a table incorporated in the SWS that was built starting from the MDB files used in the previous system (known as *Shark/Jellyfish*). In this step, all the mapping between HS and FCL code is stored. If a country is not included in the mapping table for that specific year, all the records for the reporting country are removed. All records without an FCL mapping are filtered out and saved in specific variables.

Table 8: 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

Some codes can be unmapped in the previous table, i.e., no FCL code can be assigned to some HS codes. In this case, the module generates a list of these codes that is sent to Team B/C so that they can update the table by including these missing codes.

mapped_by	year	reporter_fao	reporter_name	flow	hs_chap	hs	hs_extend	fcl	details
K	2002	33	Canada	1	33	3301230000	1.631024e-314	753	GTIS TL de
2	2014	219	Tonga	1	7	7099100	3.507421e-317	0	NA
K	2015	107	C?te dIvoire	1	8	80261	3.965420e-315	234	Generic HS2
1	2015	146	Moldova Rep	2	8	80830901	3.993577e-316	0	NA
2	2014	299	Palest, OT	1	11	110510	5.459919e-319	0	NA
2	2015	230	Ukraine	1	8	80261	3.965420e-315	0	NA
1	2014	68	France	2	3	3081100	1.522266e-317	0	FISHERY C
2	2014	299	Palest, OT	1	19	190240	9.399105e-319	0	NA
2	2015	299	Palest, OT	1	10	100199	4.950488e-319	0	NA
std-HS6	2014	7	Angola	1	2	2075200	0.000000e+00	1073	Standard_H
2	2014	230	Ukraine	2	2	20752	1.025285e-315	0	NA
T	2014	299	Palest, OT	1	22	220600	1.089909e-318	517	Generic HS2
1	2014	153	NewCaledonia	1	10	10085000	4.982652e-317	0	NA
2	2014	180	Palau	1	9	9011200	4.452124e-317	0	NA
C	2013	169	Paraguay	1	3	3083000000	1.523204e-314	0	FISHERY C

Moreover, some of the original codes in the HS-FCL mapping can be better mapped to another FCL code: the mapping table has a `correction_fcl` (not shown in the previous example in order to save space) that can be used to override the original `fcl` variable. This feature was requested by Team B/C that is responsible for these corrections.

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

Table 10: Subset of FCL units

fcl	fclunit
1096	heads
1068	1000 heads
521	mt
1083	1000 heads
1150	heads

fcl	fclunit
44	mt
840	mt
1098	mt
1181	number
1126	heads

9. 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.

10. Commodity specific conversions are added. This is done by taking the ratio of the reported weight and quantity (divided by 1000), computing the median of this ratio by FCL and the originally reported unit of measurement, and finally multiplying it by the reported quantity. 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. [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/1000 ratio). The first-best approach, in any case, would be to have speficic tables derived from external sources. This is currently under investigation.]

1.2.2.2 Eurostat

1. Eurostat classifies areas in their "geonomenclature". These codes are converted in FAO country codes using a specific conversion table, stored in the SWS, provided by Team B/C.

Table 11: 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

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

2. Commodity codes are reported in CN8 codes (*Combined Nomenclature 8 digits*). The 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. If a specific record has a CN8 code not mapping to any specific FCL code, then the record is reported and removed. If a country is not included in the mapping table for that specific year, all the records for the reporting country are removed.

3. Information of the FCL units is added. This step is straighforward 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: 1057, 1068, 1072, 1079, 1083, 1140, 1181.

5. Values are converted from EUR to USD using the table, stored in the SWS, with the average EUR/USD exchange rate for each year provided by Team B/C.

Table 12: 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

1.2.3   Pre-processing reports

The module generates various indicators/statistics on the raw data that are combined into different pre-processing reports (PPR). The following PPR are available in the "trade-reports" SWS domain:

1. Reporters by year
2. Non-reporting countries
3. Number of records by reporter/year
4. Missing data by report
5. Check qty and value included
6. Import and export content check

For a detailed explanation of what the PPR contain and how to update them, see the document "Trade Module: Plug-in Pre-Processing Report".

1.2.4   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.

1.2.5   Standardization, editing and outlier detection

1.2.5.1   Unit Values computation

For each record having both quantity and value (thus excluding all commodity reported just as value), the unit value (uv) is computed as following:

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

1.2.5.2 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 specific 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 (whithout taking into account any commodity code).

1.2.5.3 Outlier Detection and Imputation

In the currect 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 overall 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. In order to have more information on this topic, please see the validation tool documentation.

1.2.6 Mirroring

The module generates the list of non-reporting countries: these are the countries 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 conversion.

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.

1.3 Flags

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, the different kind of flags, and the conditions that should be met in order to assign them, are reported in the document "Flag Management in the Trade module".

An observation can have multiple Observation Status flags and Method flags associated with it. The final flag is the "weakest" flag: the `flagWeightTable` table contains the weights that should be assigned to all flags, and the one with the lowest value prevails over the other flags. For instance, if two official transactions at the HS level are aggregated in one CPC code, then the "s" (for aggregation) Method flag will prevail over the two "h" Method flags of each transaction.

1.4 Convertion 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 into CPC ones, 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` dataset, within the `trade` domain.

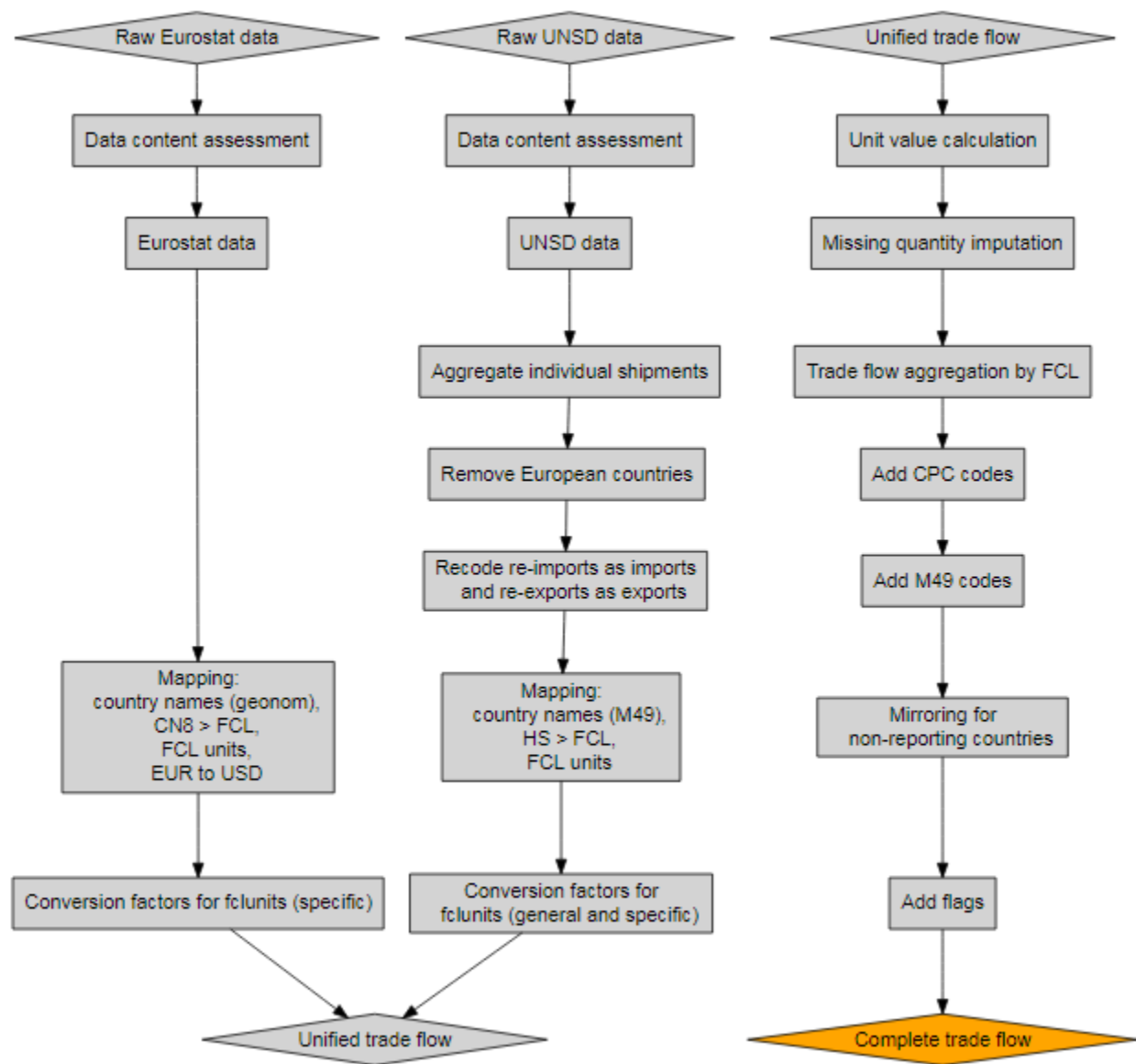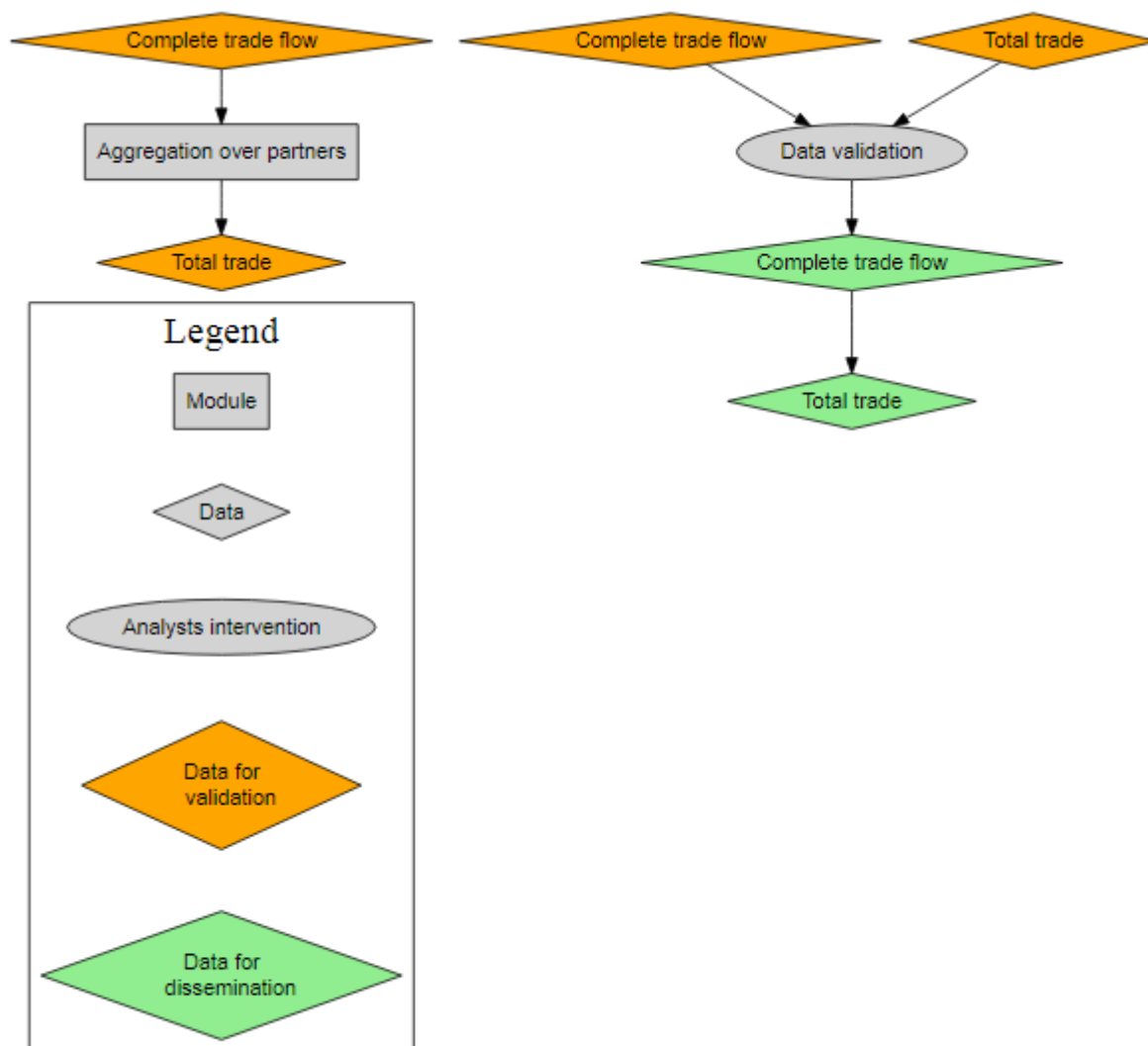# 2 Total trade CPC

This second submodule uses as input the output of the previous submodule. It aggregates total trade flow by reporting country for partners countries to a single total trade for each unique CPC commodity code.

The module saves the ouput into the dataset `total_trade_cpc_m49`, within the `trade` domain.

3 Flow Chart Process

The whole process is displayed in the next flow chart.

```
┌─────────────────────┐      ┌─────────────────────┐      ┌─────────────────────┐
  ◇ Raw Eurostat data ◇        ◇  Raw UNSD data   ◇        ◇ Unified trade flow ◇
           │                            │                            │
           ▼                            ▼                            ▼
  ┌──────────────────┐         ┌──────────────────┐         ┌──────────────────┐
  │ Data content     │         │ Data content     │         │ Unit value       │
  │ assessment       │         │ assessment       │         │ calculation      │
  └──────────────────┘         └──────────────────┘         └──────────────────┘
           │                            │                            │
           ▼                            ▼                            ▼
  ┌──────────────────┐         ┌──────────────────┐         ┌──────────────────┐
  │  Eurostat data   │         │   UNSD data      │         │ Missing quantity │
  └──────────────────┘         └──────────────────┘         │ imputation       │
                                        │                    └──────────────────┘
```

- 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

4 Future work

4.1 Validation Steps

This section represents the most high priority task for the trade module.

4.2 Outlier identification/imputation

Outliers were identified in a previous version of the `complete_tf_cpc` module and imputed automatically by using the median unit value with a specific-to-generic median unit value calculation, as explained above. 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 detailde 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).

4.2.1 Destination Table

The `complete_tf_cpc` module produces output for all the records passing all the routines and not filtered out. The module does not check if any commodity is missing. A possible solution would be to have a destination table with all the commodities of interest and the module should fill the destination table. In this way the output validation step should be achieved.

4.3 CIF/FOB

The CIF/FOB correction for mirroring is, at the time being, set up to 12%. This has been suggested by team B/C. 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.

4.4 Re-import and Re-export

All re-imports and re-exports are considered as, respectively, imports and exports. More study might be conducted in order to identify countries more prone to report re-imports and re-exports.

4.5 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).

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

4.6 Pseudo-automatic mapping of commodities

An additional method could be added in the future: the algorithm should try to trim the code not mapped and try to map them with shorter HS codes. If any of shorter codes (from right to left) are then not mapped, we can definitely discard the record. If a specific record has a HS code not mapping to any specific FCL code, then the record is reported and removed.

4.7 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 team B/C and then from FCL to CPC 2.1. In the future direct mapping from HS to CPC has
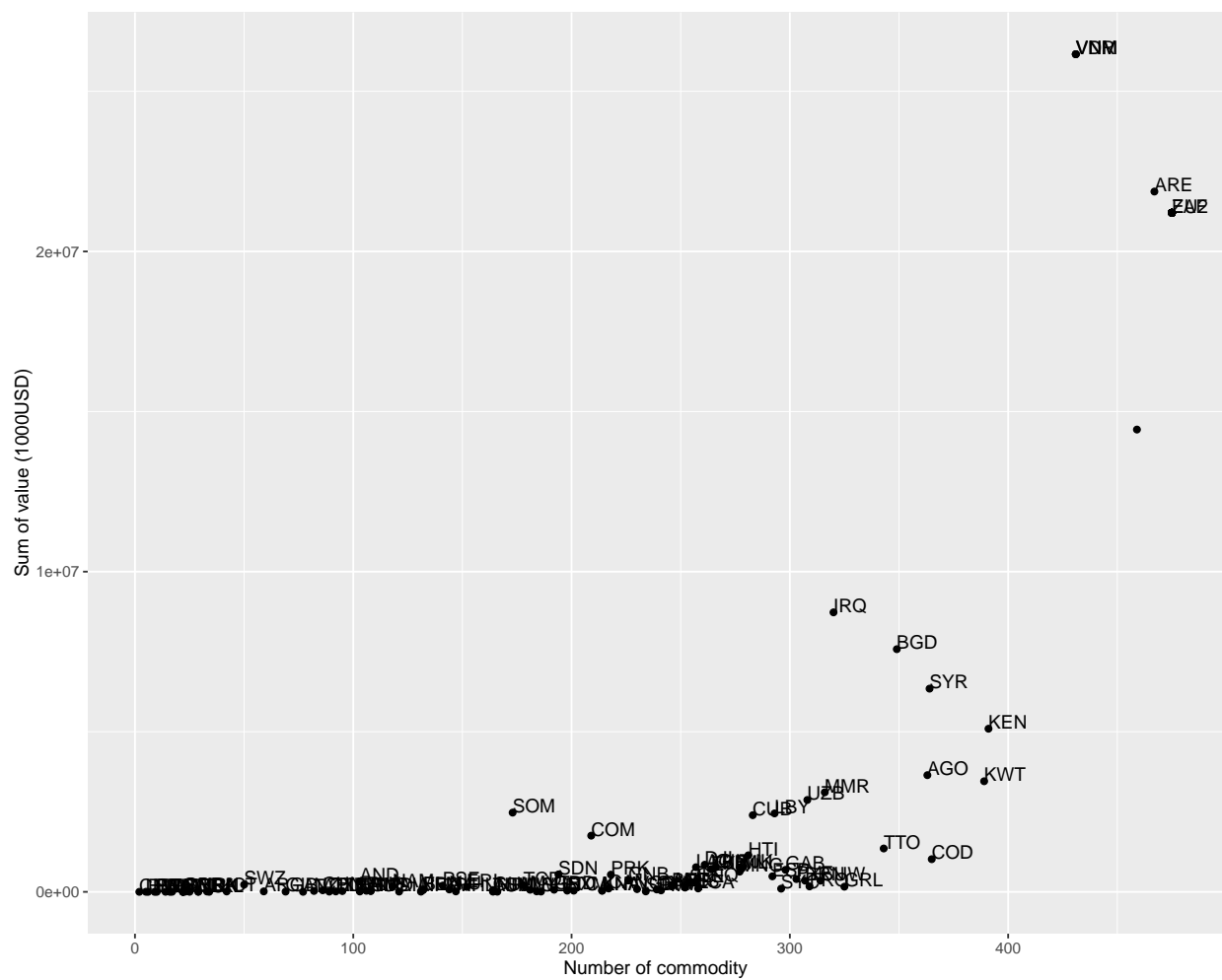
Figure 1: Sum of all self trade records by country.

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

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)

4.8   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.

4.9   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.

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.