

faoswsTrade

Marco Garieri

Food and Agriculture Organization
of the United Nations

Abstract

This vignette provides a detailed description of the various data sources and procedures used in the trade modules.

Keywords: Agricultural Trade, Tariff Line, Eurostat, Mirroring.

DRAFT

The trade module is divided in two submodules: `complete_tf_cpc` and `total_trade_CPC`. Each module is year specific. This means that, at the time being, the trade module run independently for each year. In order to run the `total_trade_CPC`, the output of `complete_tf_cpc` is needed.

1. Complete tf cpc

1.1. Data

Raw data are provided by the SWS Team (subunit of Team F) for both UNSD Tariffline and Eurostat Data. The data have been already prefiltered:

- Eurostat**
 - code of reporter (declarant) just numeric (letters are not allowed)
 - code of partner (partner) just numeric (letters are not allowed)
 - code of CN8 (product_nc) just numeric (letters are not allowed)
- UNSD**
 - code of HS (comm) just numeric (letters are not allowed)

The module downloads only records of commodities of interest for Tariffline Data. The HS chapters are the following: 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, 42, 43, 50, 51, 52, 53. In the future, if other commodity are of interest for the division, it is important to include additional chapter in the first step of the downloading. For Eurostat Data no filtering is applied.

Raw Data - Data content assessment

The pre-analyses of the assessment of the data has to be integrated in the module. In the vignettes folder, a sample of pre-analysis is given, but not integrated in the main module (file name: `preanalysis_2009.Rmd`). The pre-analysis script calculate the total number of records for both Eurostat and UNSD Tariffline datasets and the distribution of length of the commodity HS codes (for UNSD Tariffline) and CN8 for Eurostat is performed. For each country we report if data includes imports, exports, re-exports and re-imports at all possible length.

All records with hs-length (for UNSD Tariffline) or CN8-length (for Eurostat) less than 6 are removed.

The pre-analysis script produces a html file that can be read in any browser, but a suitable solution for the SWS is needed.

This script should be part of the validation steps within the module.

1.2. Process

Mapping UNSD Tariffline and Eurostat data

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

UNSD Tariffline UNSD Tariffline data reports area code with M49 standard. The area code is converted in FAO country code using a specific conversion table provided by Team ENV. Area codes not mapping to any FAO country code or mapping to code 252 (which correspond not defined area) are separately saved and removed from further analyses.

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 was

performed using a specific package (`hsfclmap` developed by Alexander Matrunich), but it is now incorporated in the module. In this step, all the mapping between HS and FCL code is stored. The algorithm tries to map at all possible length (i.e. if a reporting country has a record with hs code at 12 digits and in the package, for the same reporting country, a HS-to-FCL mapping is available at a lower level, for example 10 digits, the algorithm will include in the mapping all the records having the same 10 digits).

An additional method has to be added in the future: the algorithm should try to trim the code not mapped and try to mapped 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.

If a country is not included in the package of the mapping for that specific year, all the records for the reporting country are removed.

All records filtered out are saved in specific variables, but not used in the output at the time being. Further develop of the module Information of the FCL units is added at the end of this step.

Just for UNSD Tariffline data conversion of units of measurements are applied to meet FAO standards, where all weights are reported in metric tonnes, animals in heads or 1000 heads and for some commodity just the value is provided.

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

Distinction between Exports and Re-exports / Imports and Re-imports

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.

Eurostat Eurostat data reports area code with geonomenclature standard. The area code is converted in FAO country code using a specific conversion table provided by Team B/C. Area codes not mapping to any FAO country code or mapping to code 252 (which correspond not defined area) is reported are the records for these area codes are removed. 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 package (`hsfclmap`) as for UNSD Tariffline. 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 package of the mapping for that specific year, all the records for the reporting country are removed.

The possible solution for the missing links in the future using Natural language processing routines to read the metadata.

Information of the FCL units is added at the end of this step.

Values are converted from EUR to USD using the table with average currency for each year provided by Team B/C.

Eurostat data are already provided in the correct units of measurements and do not need further conversions.

Unified Official Trade Flows Dataset

UNSD Tariffline and Eurostat datasets are ready to be merged.

Standardization, editing and outlier detection

- **Application of Notes** Perennial and yearly specific notes are mdb files provided by the Team B/C already saved in a R friendly dataset within the package. These mainly regard quantity adjustments.
- **Self Trade Analysis** For all the records having the same reporter and partner an analysis is performed. The sum of the value is computed for both countries and commodity, in order to spot out the countries reporting massive self trade and which are the main commodity reported as self trade. Summary statistics are computed worldwide.
- **Unit Values computation** For each record having both quantity and value (thus excluding all commodity reported just as value), the unit of value (u_v) is computed as following:

$$u_v = \frac{qty}{value} \quad (1)$$

- **Outlier Detection and Imputation** The outlier are calculated based on the distribution of the unit of value for the same country, year and flow at the HS level. The reason to identify the outlier at the HS level is due to the fact that, under the same FCL code, different commodity might fall (i.e. maize seed and seed). The records with wrong order of magnitude are detected. The outlier are detecting using the Tukey's procedure:
 - The Tukey's five number summary are calculated: minimum (m), lower-hinge (lh), median (med), upper-hinge (uh) and maximum (M).
 - The coefficient for the outlier detection is set up as suggested by Tukey to 1.5 ($coef$).
 - For each value is calculated a specific distance from the lower or the upper-hinge in the following way:

$$x \text{ is outlier if } \begin{cases} x < lh - coef * iqr, & \text{lower outlier,} \\ x > up + coef * iqr, & \text{upper outlier.} \end{cases} \quad (2)$$

where iqr is the interquartile range.

The outlier are then corrected using the corresponding value and dividing it by the median unit of value of that specific commodity, country, flow and year. In this way only few official data are corrected.

- **Missing Quantities Detection and Imputation** For records in which 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 the median of the units of value of the corresponding commodity (HS level/country/flow/year)

Mirroring and Balancing

The flows at this point are aggregated by reported over partner countries to a single total trade for each unique FCL commodity code.

The module produce the list of non-reporting countries: these are the countries present as partners but absent as reporters. For this countries the mirroring process is applied: the corresponding trade of the non-reporting countries are extracted from the partners inverting the flows.

Upon availability of time-series data, a check of the FCL-based unit values across the time series is performed.

A check is performed to account of the CIF/FOB differences [around 12%?] (still to be performed among time series).

2. Total trade CPC

2.1. Unbalanced World Trade Matrix

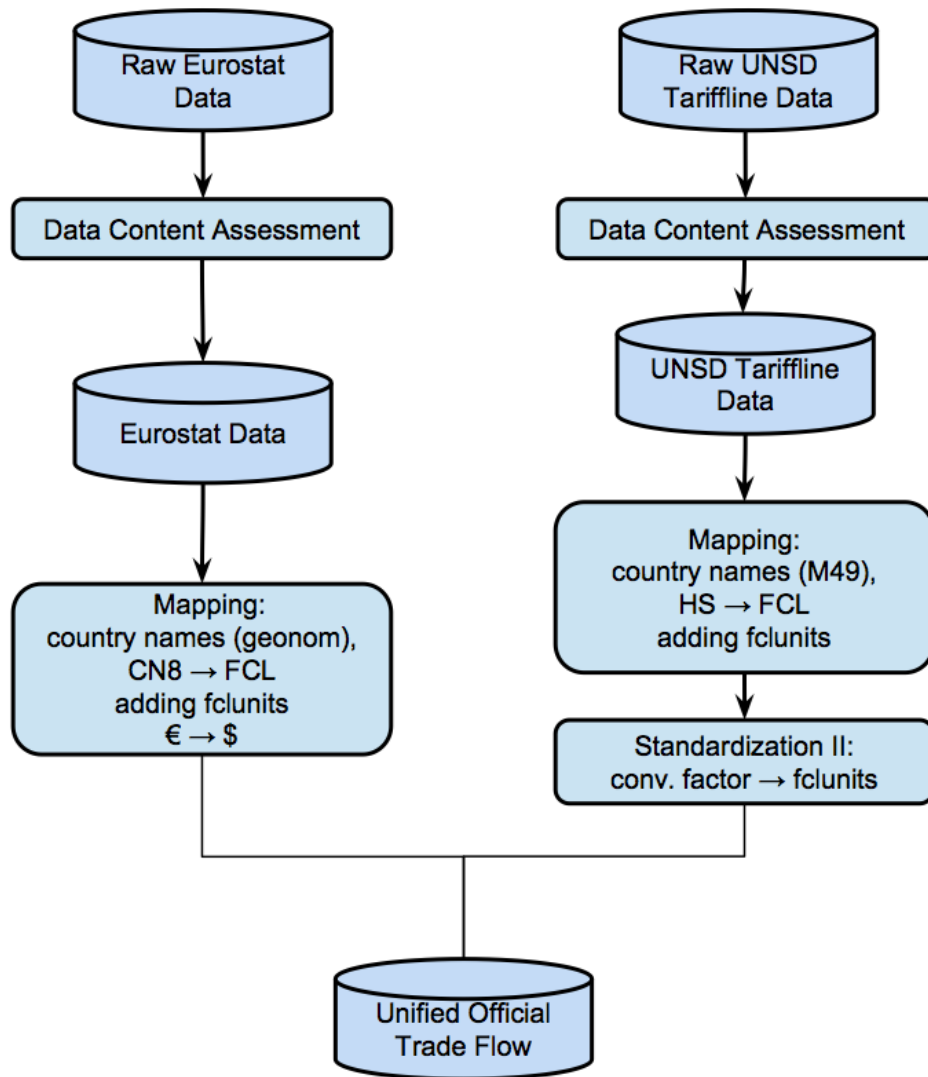
The commodity balance check is computed, which will lead to the "Trade Imbalances" report (not yet implemented).

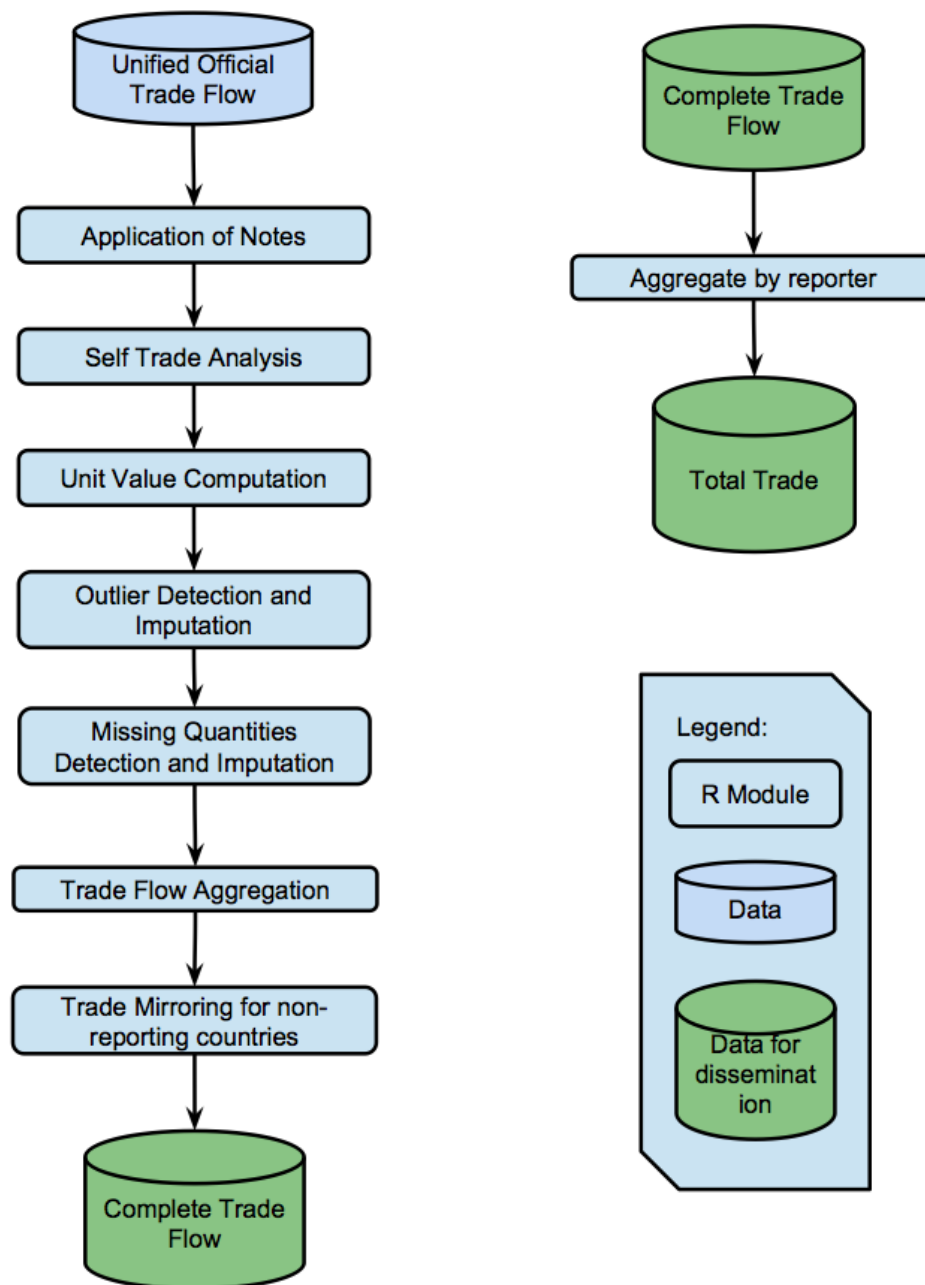
2.2. Total Trade

The conversion from FCL to CPC version 2.1 expanded is added and the final output tables of trade by commodity is provided.

3. Flow Chart Process

Description of the entire flow





4. Future work

Specific analyses are needed for the time series in the future, but these are not yet implemented.

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.

This paper is dynamically generated on July 26, 2016 and is subject to changes and updates.

Affiliation:

Marco Garieri

Economics and Social Statistics Division (ESS)

Economic and Social Development Department (ES)

Food and Agriculture Organization of the United Nations (FAO)

Viale delle Terme di Caracalla 00153 Rome, Italy

E-mail: marco.garieri@fao.org

URL: <https://github.com/SWS-Methodology/faoswsTrade>

DRAFT