Data processing in UN Comtrade

United Nations Statistics Division
September 2017

UN Comtrade Overview

- Official trade statistics of almost **200 countries/areas**
- Most comprehensive trade database with more than **4 billion records**
- Annual trade in goods since **1962,** monthly data since **2010**
- A**nnual trade in services** since 2000

https://comtrade.un.org/

UN Comtrade Current Data Elements

- Trade Flow
- Partner
- Commodity
- Trade values
  - CIF for Imports
  - FOB for Exports

- Net weight
- Supplementary Quantity and units
- Estimation Flag

Period	Trade Flow	Reporter	Partner	Commodity Code	Trade Value (US$)	Netweight (kg)	Qty Unit	Qty	Flag
2016	Import	China	World	7201	$50,122,175	190,291,180	Weight in kilograms	190,291,180	6
2016	Export	China	World	7201	$27,967,757	133,061,303	Weight in kilograms	133,061,303	0

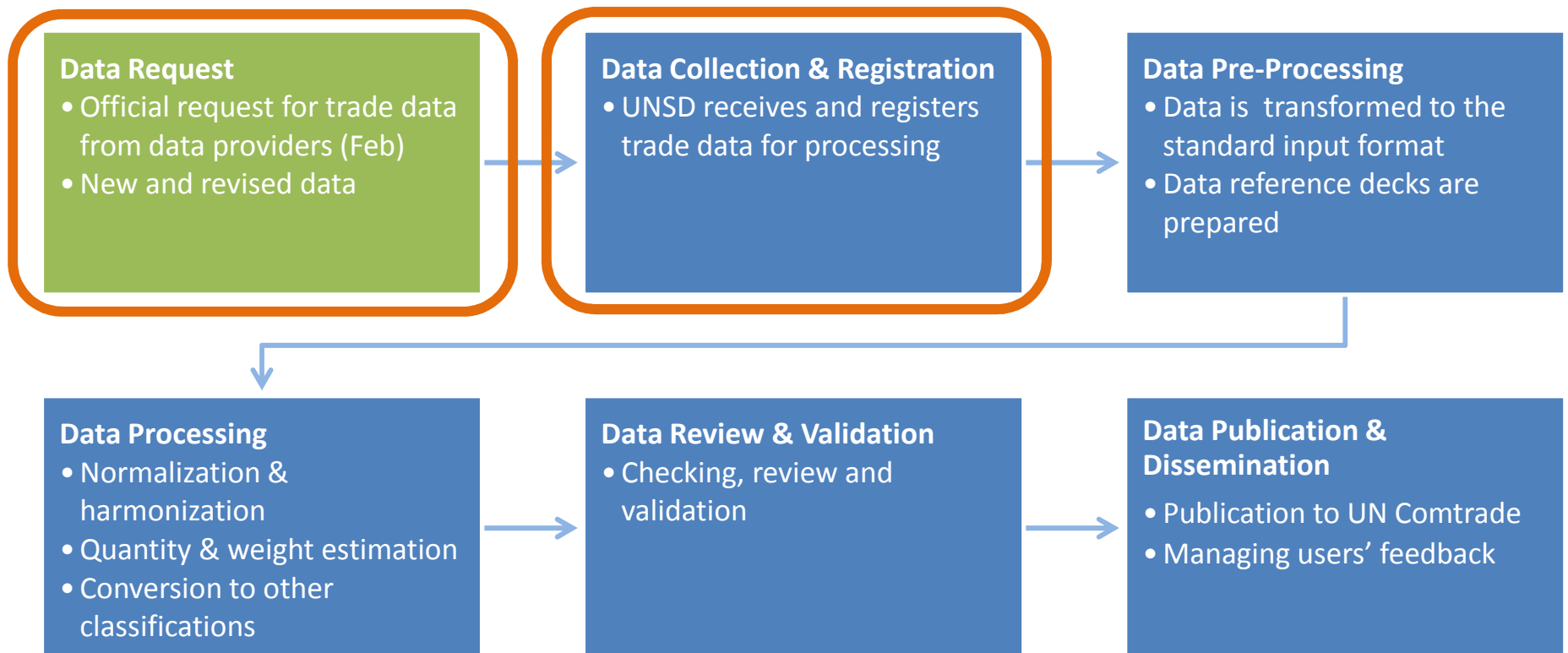# New Data Items (IMTS 2010)

- ## Second Partner

  - Country of Consignment

- ## Trade values

  - FOB for Imports (as well as CIF)

- ## Mode of transport

- ## Customs procedure codes

Data Request and Collection

Data Request
- Official request for trade data from data providers (Feb)
- New and revised data

Data Collection & Registration
- UNSD receives and registers trade data for processing

Data Pre-Processing
- Data is transformed to the standard input format
- Data reference decks are prepared

Data Processing
- Normalization & harmonization
- Quantity & weight estimation
- Conversion to other classifications

Data Review & Validation
- Checking, review and validation

Data Publication & Dissemination
- Publication to UN Comtrade
- Managing users' feedback

Pre-Processing Objectives

- **Transform** national data into the input format required by the internal processing system

- **Ensure** that national code references are available so that we are able to map national codes to standard codes correctly during processing stage

- **Handle** unique and special cases

- **Maintain** and level of information originally submitted

- **Ensure that the necessary parameters/fields are available**

  - Trade flow, commodity at detailed 6-digit level, partner code, trade value, weight, quantity and quantity unit code

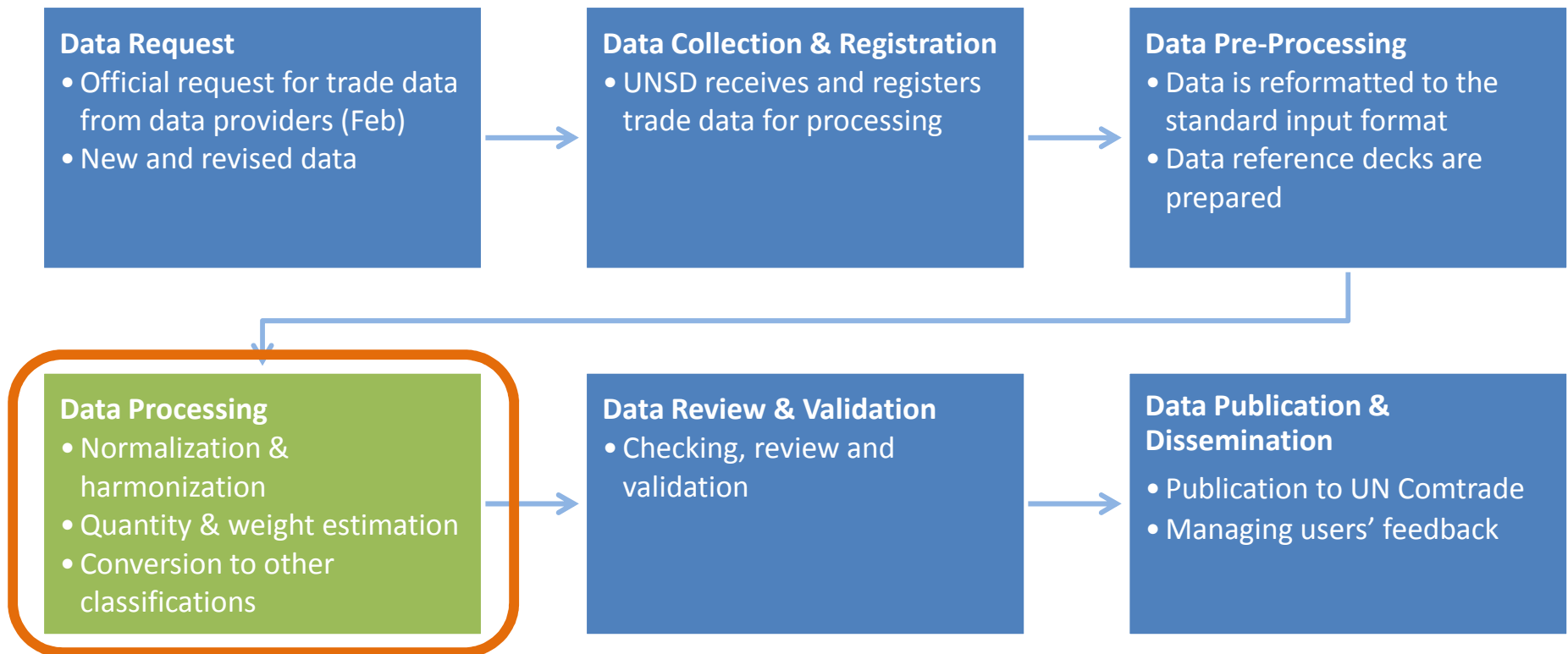- **Produce an output** file in .mdb or tab-delimited format

Pre-processing: Data validation

- **Check trade values against metadata**. Trade values may sometimes be broken down into value, insurance and freight.
- **Check if there are negative values**
- **Accumulated/Aggregated Totals**
  - Aggregate detailed data to get totals by trade flow, and compare with official or reported totals.
  - New vs Revised datasets: We compare aggregated totals in detail to determine if we should re-process revised data
- Verify that related **oil trade data** is reported
- **Confidential values**
  - Suppressed data under pseudo headings
    - '9999AA', where 'AA' reflects chapter
  - Reported negative values under unknown commodity
- **High-level totals** must be greater than the aggregated totals of detailed records

Processing & Harmonization

Data Request
- Official request for trade data from data providers (Feb)
- New and revised data

Data Collection & Registration
- UNSD receives and registers trade data for processing

Data Pre-Processing
- Data is reformatted to the standard input format
- Data reference decks are prepared

Data Processing
- Normalization & harmonization
- Quantity & weight estimation
- Conversion to other classifications

Data Review & Validation
- Checking, review and validation

Data Publication & Dissemination
- Publication to UN Comtrade
- Managing users' feedback

Processing & Harmonization

Commodity Normalization

- **Code Normalization**
  - National data items are mapped to standard codes (M49, WCO, Revised Kyoto Convention Codes)
  - Currency conversion to USD
  - Output: Tariffline data

- **Commodity Normalization**
  - Removes monetary gold [HS: 711890, 710820] from dataset
  - Non-standard codes are addressed
  - Special codes are mapped (i.e. pseudo-codes)

Quantity Information

- UN Comtrade includes **net weight** and, when available, **supplementary quantity**

- Quantity units are standardized according to World Customs Organization recommended units to provide comparable data

- **Non-standard quantities** are converted using specific conversion factors or FAO conversions

- UN Comtrade **estimates missing quantity** at the most detailed reported level when possible using the reporter's reported weighted unit value within the same 6-digit commodity flow or median standard unit value of all reporters for the same 6-digit commodity flow the previous year, after elimination of outliers

- Detect and remove extreme outliers in quantity based on:
  - Historical Unit Value Check
  - Standard Unit Value Check
 - Historical Value Check

Conversion to other classifications

- Data reported in latest HS classification is converted to earlier versions of HS

- Use of WCO provided correlation tables between latest HS and previous HS version

- Our practice is to convert most detailed level in original classification into one single code in the target classification, without splitting.

Correlation and conversion

Correlation and Conversion of "Automatic data processing machines (SITC 752)"
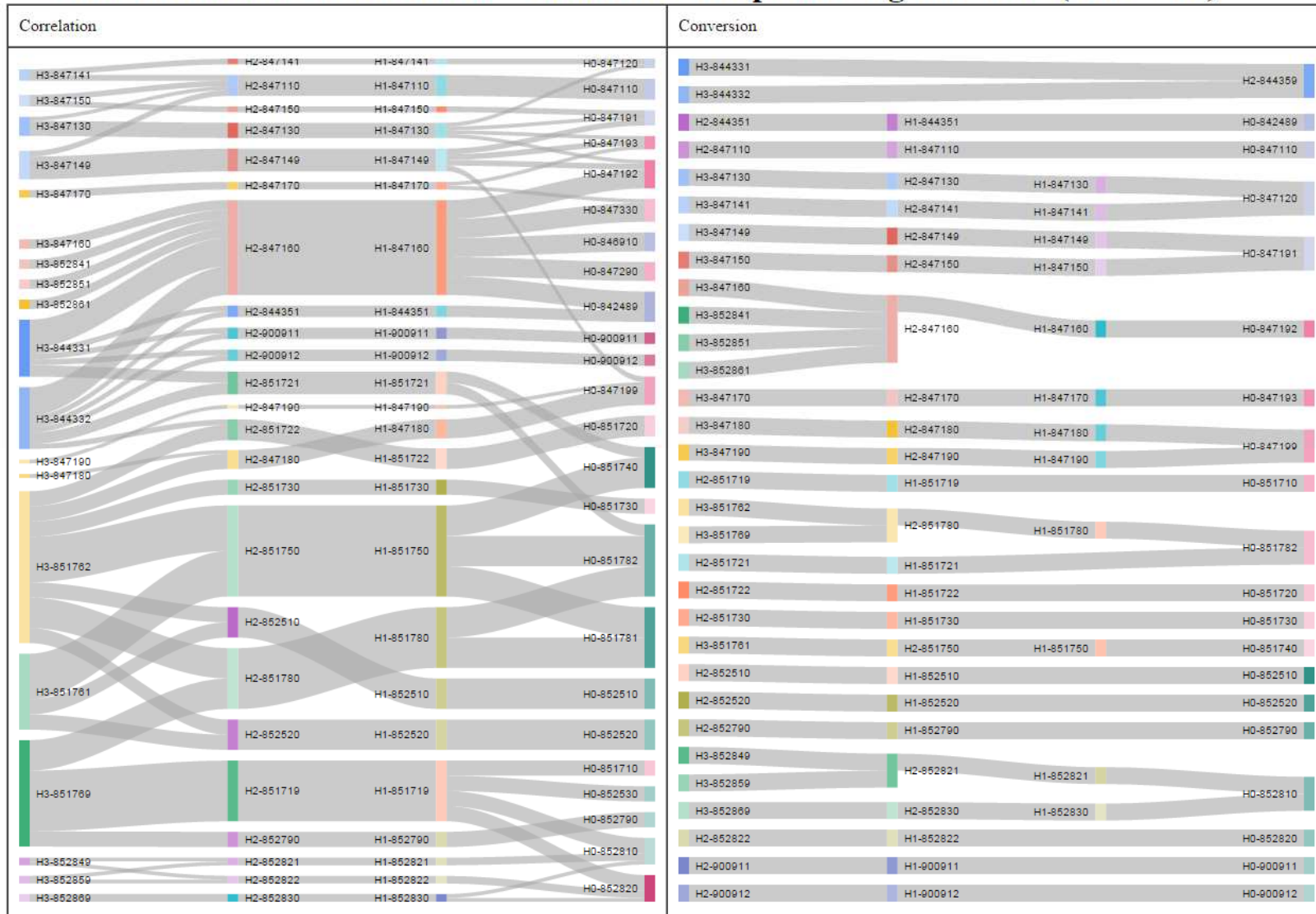
The complex relationships (left) are simplified (right)

Conversion to other classifications

Conversion and Correlation Tables

▦ Download Complete correlations among HS, SITC and BEC   ▦ Download Conversion Methodology General Note (2017)

FROM / TO	HS 2012	HS 2007	HS 2002	HS 1996	HS 1992	SITC 4	SITC 3	SITC 2	SITC 1	BEC 4
HS 2017	▦	▦	▦	▦	▦	▦	▦	▦	▦	▦
HS 2012	-	▦	▦	▦	▦	▦	▦	▦	▦	▦
HS 2007	-	-	▦	▦	▦	▦	▦	▦	▦	▦
HS 2002	-	-	-	▦	▦	-	▦	▦	▦	▦
HS 1996	-	-	-	-	▦	-	▦	▦	▦	▦
HS 1992	-	-	-	-	-	-	▦	▦	▦	-
SITC 4	-	-	-	-	-	-	-	-	-	-
SITC 3	-	-	-	-	-	-	-	▦	▦	▦
SITC 2	-	-	-	-	-	-	-	-	▦	▦
SITC 1	-	-	-	-	-	-	-	-	-	-

https://unstats.un.org/unsd/trade/classifications/correspondence-tables.asp

Data Review

- Consolidated report contains result of each processing phase
- Reviews both by editor and supervisor
- Check overall totals against official sources, time series, trends
- Check level of memorandum items (overall value of non-standard commodity codes)
- Verify the mapping of partners, commodities and quantity units are correct
- Check integrity reports (possible wrong classification, conversion factors changed more than 1% from previous processing, export to itself, general trade system but data has free zones, partner world vs sum of individual partners)

Sample Data Processing Reports

Table.6. Partner Value Comparison

Flow	Partner		Value.A	Value.B	%diff	diff	%share
1	0	AAA	$611,364,435,458.00	$611,364,435,458.00	0.00 %	$0.00	100.00 %
1	4	AFG	$2,506,231.00	$2,506,231.00	0.00 %	$0.00	0.00 %
1	8	ALB	$23,238,224.00	$23,238,224.00	0.00 %	$0.00	0.00 %
1	10	ANT	$16,470.00	$16,470.00	0.00 %	$0.00	0.00 %
1	12	DZA	$4,627,196,986.00	$4,627,196,986.00	0.00 %	$0.00	0.76 %
1	16	ASM	$32,941.00	$32,941.00	0.00 %	$0.00	0.00 %
1	20	AND	$20,263,962.00	$20,263,962.00	0.00 %	$0.00	0.00 %
1	24	AGO	$2,370,503,388.00	$2,370,503,388.00	0.00 %	$0.00	0.39 %
1	28	ATG	$2,972,890.00	$2,972,890.00	0.00 %	$0.00	0.00 %
1	31	AZE	$2,260,091,970.00	$2,260,091,970.00	0.00 %	$0.00	0.37 %
1	32	ARG	$807,654,225.00	$807,654,225.00	0.00 %	$0.00	0.13 %
1	36	AUS	$1,610,338,536.00	$1,610,338,536.00	0.00 %	$0.00	0.26 %
1	40	AUT	$5,339,687,175.00	$5,339,687,175.00	0.00 %	$0.00	0.87 %

Table.8. Commodity Value Comparison

Flow	Code		Value.A	Value.B	%diff	diff	%share
1	TOTAL	All Commodities	$611,364,435,458.00	$611,364,435,458.00	0.00 %	$0.00	100.00 %
1	01	Live animals; animal prod	$352,104,872.00	$352,104,872.00	0.00 %	$0.00	0.06 %
1	02	Meat and edible meat offa	$4,818,724,643.00	$4,818,724,643.00	0.00 %	$0.00	0.79 %
1	03	Fish and crustaceans, mol	$4,164,666,970.00	$4,164,666,970.00	0.00 %	$0.00	0.68 %
1	04	Dairy produce; birds' egg	$3,463,250,314.00	$3,463,250,314.00	0.00 %	$0.00	0.57 %
1	05	Products of animal origin	$316,305,297.00	$316,305,297.00	0.00 %	$0.00	0.05 %
1	06	Live trees and other plan	$1,429,273,637.00	$1,429,273,637.00	0.00 %	$0.00	0.23 %
1	07	Edible vegetables and cer	$3,023,608,529.00	$3,023,608,529.00	0.00 %	$0.00	0.49 %
1	08	Edible fruit and nuts; pe	$4,163,397,384.00	$4,163,397,384.00	0.00 %	$0.00	0.68 %
1	09	Coffee, tea, maté and spi	$1,388,285,917.00	$1,388,285,917.00	0.00 %	$0.00	0.23 %
1	10	Cereals	$910,063,821.00	$910,063,821.00	0.00 %	$0.00	0.15 %
1	11	Products of the milling i	$380,426,930.00	$380,426,930.00	0.00 %	$0.00	0.06 %
1	12	Oil seeds and oleaginous	$862,062,497.00	$862,062,497.00	0.00 %	$0.00	0.14 %

- **Table6: Partner Value Comparison**
  - Check major differences by partners
- **Table8: Commodity Value Comparison**
  - Check major differences by commodities

New UN Comtrade interface

UN Comtrade Database | Extract data | Bulk download *new* | Data Availability ▾ | Metadata ▾ | Reference ▾ | Knowledge base | API portal

1. Type of product & Frequency

Type of product
◉ Goods ◯ Services

Frequency
◉ Annual ◯ Monthly

2. Classification

HS
◉ As reported ◯ 92 ◯ 96 ◯ 02 ◯ 07 ◯ 12

SITC
◯ As reported * ◯ Rev. 1 ◯ Rev. 2 ◯ Rev. 3 ◯ Rev. 4

BEC
◯ BEC

3. Select desired data

Periods (year)

× 2016

All or a valid period. Up to 5 may be selected.

Reporters

× All

All or a valid reporter. Up to 5 may be selected. All may only be used if a partner is selected.

Partners

× World

World , All , or a valid reporter. Up to 5 may be selected. All may only be used if a reporter is selected.

Trade flows

× All

All or select multiple trade flows.

HS (as reported) commodity codes

× TOTAL - Total of all HS commodities

All , Total , AG[X] or a valid code. Up to 20 may be selected. If you know the code number, e.g. 01 - Live animals , type 01 . To search by description type a word, e.g. rice .

4. See the results

Preview » **Download CSV**

Issues opening CSV in Excel? See this Microsoft how-to.

Metadata

Data Release

- Ongoing and continuous release and publication of datasets

- Data availability www.comtrade.un.org/data/da

API:
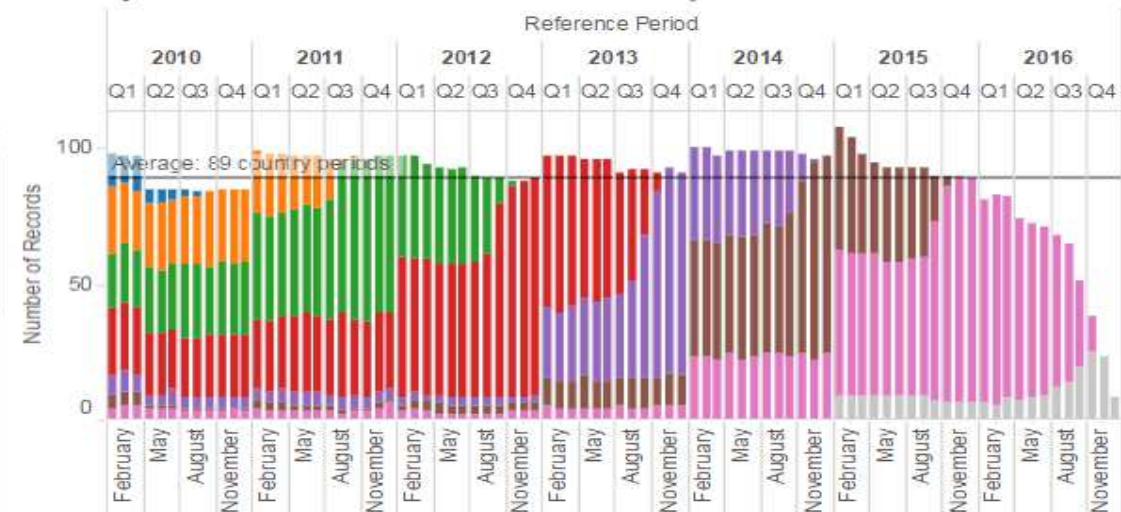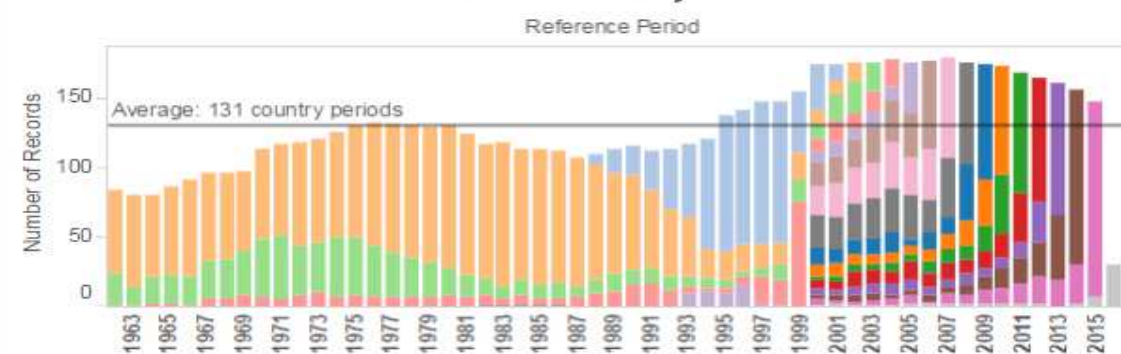https://comtrade.un.
org/data/dev/portal/

User support
mailbox
comtrade@un.org

Monthly Merchandise Trade Data Availability

Annual Merchandise Trade Data Availability

Thank You!