# Outcome

A csv dataset containing time series of trade in value and in volume (base 100 = 2017), and price index.

## Variables :

**aggregation\_level** : aggregation level at which the price index was computed. (« Production stage » for instance)

**year**: year

**production\_stage** : production stage with 5 categories (Primary, semi-finished, parts and components,

**yearly\_change\_price\_index** : change in price index (ratio between current and previous year).

**Price\_index\_base\_100** : Price index (base 100 = 2017)

**Trade\_value\_dollars** : Trade value (in M$)

**Trade value\_base\_100** : Trade value (base 100 = 2017)

(for now start date is 2017 but in the end will be 2000)

# From HS to ISIC

*https://ec.europa.eu/eurostat/web/international-trade-in-goods/methodology/classifications*

*Eurostat manages correspondence tables enabling the transposition of data collected according to the Combined Nomenclature into other classifications like the Standard International Trade Classification (SITC), the Classification of Products by Activity (CPA), the Broad Economic Activities (BEC) or the classification of goods for transport statistics NSTR/Rev.1. All classifications and correspondence tables are available on Eurostat’s metadata server RAMON.*

Good news : out of 5383 HS5 codes, only 40 are linked to 2 ISIC\_2d and 1 to 3 ISIC\_2d.

# Roadmap

1. Create a dataset at the same level of aggregation than the price index to count the number of observations. Helps determining whether price index makes sense.
2. Filter outliers
   1. Impute missing uv ? Not done for now
   2. Provide various thresholds ?
   3. Filter at different aggregation level ? (for now only HS 6-digits).
3. Compute total trade at the right level of aggregation.
4. Average price indices at the right level of aggregation.
   1. Year
   2. Year x ISIC\_2d
   3. Year x BEC
   4. Year x ISIC\_2d x BEC
5. Compute evolution over time of trade in value and in volume (base 100 = 2017)

Total trade and weights for the price index are NOT computed on the same sample, because price index uses only the subsample of i-j-k present both in current and previous period.

# Programs

01\_prepare\_data : creates the correspondence tables and a version of BACI aggregated by ISIC\_2d

02\_compute\_price\_indices : compute time variation of the price index

03\_ plot\_contents : make some graphs

04\_export\_clean\_data : produce the csv dataset that we share with the world.

# Filtering rule :

Compute distribution of delta\_ln\_uv at the year x product level and remove the observations for which delta\_ln\_uv is not between the 5th and 95th percentile.

# Which aggregation level should we choose ?

The ISIC\_2d needs to have enough observations for the price index to be correctly computed : we aggregate all sectors representing less than 1% of trade flows into an aggregated « NED » sector.