Deliverable 4.3: Initial Physical Dataset

12 February 2021

Contents

1. Description

The initial physical dataset was generated for the EU-28 countries plus the Rest of the World region for the year 2011. It contains highly detailed data on products that were produced in these countries, the international trade, and the households expenditures (all in tonnes), and the composition (in weight fraction) of two elements (tantalum and copper) of said products. Additionally, for each of these data points, a flag was appointed to indicate the quality and the level of transformation applied to the original dataset to be converted into tonnes and/or to the Panorama classification system.

The product of the necessary procedures for conversion and harmonization (see Deliverable 3.2 and 5.2) is presented in the structure that is to be used in the Deliverable 5.3 (Roll out the approach for a number of materials/products). This dataset can be found in the Annex as a JSON file. There are 5 main objects in the file:

- code
- label
- dim
- val
- flag

Code and label are used as unique IDs and description of the data, and they have 5 nested objects each:

- fin – representing only household expenditure.
- elm – elements Ta and Cu.
- reg – for the 28 individual countries in the EU-28 plus one Rest of the World region.
- sin – single; void data only to be used to artificially increase the dimension of the data structure making it uniform across different objects.
- pro – for the 629 products covered in the first trial in which Ta and Cu are covered.

Dim are the dimensions of each of the nested objects in val, and by extension, flag. This object has 8 nested objects in it:

- dom_min – representing the dimensions of the minimum value (based on conversion factors) of the domestic production. This means that the data in val is given as a 3D object (reg x pro x sin). Since sin (single) is used, the reader can imagine that the corresponding val object comprises of a matrix of regions by products.

- dom_med – same as dom_min, but in val, mediums values are displayed based on conversion factors.
- dom_max - same interpretation but with maximum values.
- imp – representing total imports for each country in a 3D object (reg x sin x pro).
- exp – representing total exports for each country in a 3D object (reg x pro x sin).
- trad – representing the bilateral trade amongst the 29 region in a 3D object (pro x reg x reg).
- cmp – representing the composition of the 629 products in terms of tantalum and copper in a 2D object (pro x elm).
- fin – representing the household expenditure of each of the 629 products per region in a 3D object (reg x pro x fin).

For each of the nested objects in the dim object, there is a corresponding one in the val object. Apart from dom_min, dom_med, and dom_max, which are reduced to dom, the same applies to the flag. This means that the quality flags in the dom_min, dom_med and dom_max are the same and represented under dom in flag. In the val and flag nested objects, the values and flags are displayed according to the dimensional order given in dim. Additionally, the order of the codes and labels are important, as they are used as indexes in the val and flag objects.

It is important to notice, that contrary to the other regions' original micro-data, the RoW production and trade data was obtained in the end using the existing RoW production and consumption data from Exiobase itself. It has been then disaggregated into PAN level of product detail using the corresponding detailed RoW trade micro-data.

As an example, there is a data point in trad under val and flag with indexes [56,2,4] of 339.4 tonnes and 1 respectively. As dim states, trad is a 3D object (pro x reg x reg). So the first index 56 – corresponds to the product, the second index 2 to a region (in this case the source) and the third index 4 to another region (in this case target). In order to know which product, source region, and target region these correspond to, the reader can look into the code and label corresponding to the products and regions. According to the order in code and label, the product with index 56 is C_GARM ("Wearing apparel; furs (18)"), the source region with index 2 is BE (Belgium), and the target region with index 4 is CY (Cyprus). In other words, Belgium exported to Cyprus 339.4 tonnes of wearing apparel; furs in the year 2011, and the assigned quality flag to this value is 1. Figure 1 illustrates this description.
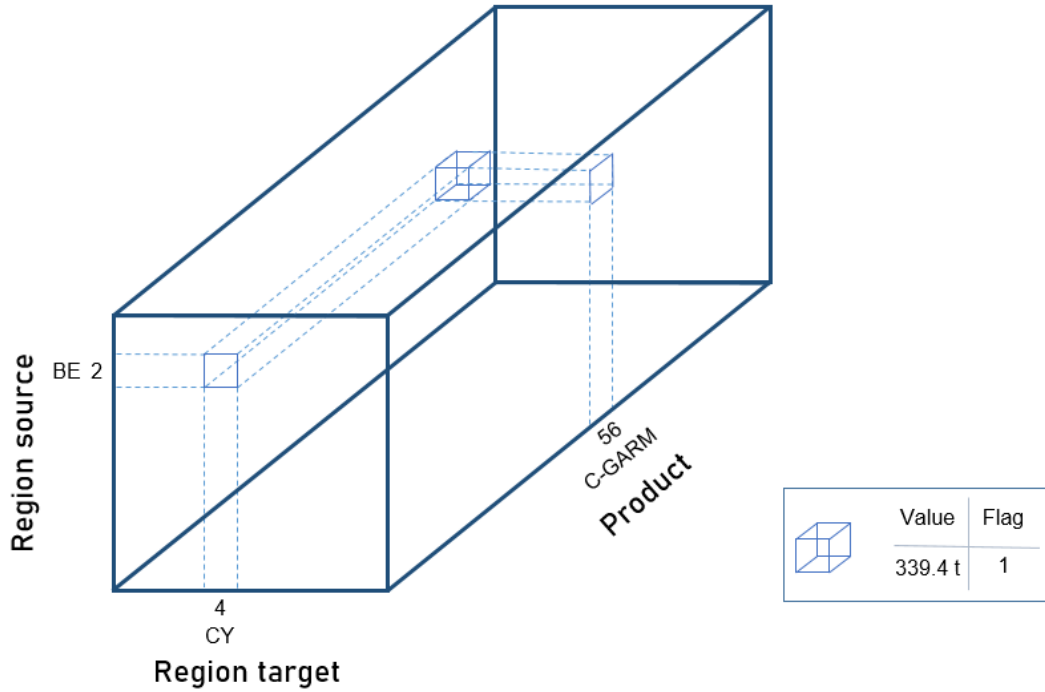
Figure 1: Example of data point within the data structure for bilateral trade.

2. Appendix

Initial physical dataset

See the file in the appendix: **4.3_physical dataset.json**