



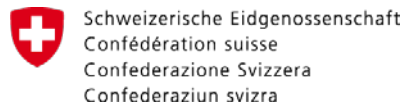
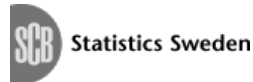
Practical Guidelines for processing scanner data

Dirk-Jan Hoogerdijk, Eurostat
15th Meeting of the Ottawa Group
Eltville am Rhein, 10 – 12 May 2017



Practical Guidelines for stable assortment

Many NSIs in EU doing somewhat the same thing ...





Practical Guidelines for stable assortment

What are GTINs, SKU ...

Comparison with traditional price collection

Pros and cons of using scanner data

Sampling dimensions: geographic, time

Obtaining data



Practical Guidelines for stable assortment

Mapping GTINs to COICOP (has become a separate project)

Processing either through mimicking traditional methods or matched model unweighted Jevons

Integration with traditionally collected data

A draft is done, there are still some loose ends to fix



Mapping GTINs to COICOP – **to be finished in March 2018**

NSIs develop methodologies, often in-house by price statisticians, for mapping GTINs to COICOP.

Idea to develop a methodology centrally, involve all scanner data processing EU countries and experts from the outside. Experts on semantic technologies, machine learning and AI. Using real data.

Focus on supermarket data. Usually NSIs start here, results also useful for web scraped data.

*Prototype for mapping available by **mid-2017**.*



Linking replacing GTINs – to be finished in March 2018

Dealing with replacements is very important (avoidance of downward drift), especially for dynamic assortments.

Dealing with replacements (without physical evidence) relies on the same data as used for mapping + price and quantity information.



HICP and scanner data – ongoing discussion

Current processing of scanner data in Europe is consistent with HICP regulations: mimicking traditional methods or using matched model unweighted Jevons.

In the HICP context comparability between national HICPs is fundamental.

Scanner data offers the opportunity to harmonise methods across countries at a lower level.



Multilateral methods - ongoing discussion

The search for methods that would

- 1) work in dynamic universes*
- 2) utilise as much information as possible
(including using weights at low levels)*
- 3) avoid chain drift*

Multilateral methods are a potential candidate