Experiences and best practice in developing new production processes for statistical products: Price statistics – the case of Slovenia

Session1:
Genovefa Ružić, Statistical Office of the Republic of Slovenia

2018 DGINS Conference
Bucharest, 10-11 October 2018

Main objectives

❖ Modernize the **data collection** methods

   ➢ New data sources:

      ✓ scanner data

      ✓ web-scraped data

❖ Explore new technical and methodological solutions for **data compilation**

   ➢ New tools for data processing

   ➢ New methodology for HICP compilation

Obtaining scanner data

❖ Studying the practices of other countries
   ➢ Reviewing, examining, analysing documents of other MSs
   ➢ Study visits at statistical offices of Norway, Sweden and the Netherlands

❖ Detailed market research
   ➢ List of the most important and biggest retailers in Slovenia

❖ Negotiations
   ➢ Invitation letters to the meetings
   ➢ Meetings at the top and expert level
   ➢ Special contract for each retailer

Processing scanner data

❖ **Challenges**

- ➢ Enormous amount of data
- ➢ Gradual data transmission (over a period of more than three years)
- ➢ Commitment to retailers on inclusion of scanner data into production ASAP

❖ **Solutions**

- ➢ Developing our own application for processing scanner data
- ➢ Using scanner data only as a substitute for traditionally collected data
- ➢ Preparing new methodology for more comprehensive usage of scanner data

New production process

❖ New methodology for food, beverages and tobacco
  ➢ In line with Eurostat's guidelines for processing supermarket scanner data
  ➢ Catalogue for each retailer (scanner data linked to the ECOICOP)
  ➢ New index type (Jevons index)
  ➢ New weights at the lowest level (based on retailers' turnover)
  ➢ New system of data control and data processing
  ➢ Integrating all the innovations into an automated production process

❖ Introduced in the HICP in 2018
  ➢ Detailed information for users published on our website
  ➢ Presentation at the press conference

Web scraping

❖ Target product groups 1:
  ➢ Websites with data on profit rents and second hand motor cars
  ➢ Owners: web scraping not allowed (their products)

❖ Target product groups 2:
  ➢ Retailers' websites with data on computer equipment

❖ Testing online applications for web scraping
  ➢ Import.IO tool (freely available)
  ➢ Data Scraping Studio license (renamed Agenty)

❖ In 2018 including the scraped data into HICP compilation

Conclusion and future plans

❖ Conclusion

➢ A big time input to obtain the data and to develop new tools for data processing and to integrate all the innovations into a new production process

➢ Important to maintain good relations with retailers (we will regularly provide them with specific analyses of the movements of their prices)

➢ Never ending work in progress

❖ Future plans

➢ Continue to work on improving the methodology and data processing

➢ Obtain scanner data from new retailers (specialized stores)

➢ Continue to work on web scraping (develop our own custom-made tools)