

## Poster Session: Alternative Data Sources for Price Statistics

Statistics Finland presented a poster about the main outputs of a project which goals are to study, develop and test the capabilities to utilize scanner data and web scraping in statistics production. During the poster session participants' questions and remarks focused on the following topics:

### Scanner/sales data

- Many participants were interested in the level of detail of data. For instance the data on daily consumption goods is aggregated on a day level for each product and the data delivery is agreed to be monthly.
- Scanner/sales data may involve some additional costs to NSOs (e.g. setting-up costs). However, that is not always the case, because it can be beneficial from the informants' point of view to have an automated data transfer instead of manual work.
- Pharmaceutical data is interesting also for the use of SPPI and NA because it is possible to calculate a margin from that data. It is possible that other countries have similar kind of data on pharmaceutical products because those are subject to many laws and regulations.

### Web Scraper

- There are many challenges in web scraping. For example, administrators can block web scrapers, web sites change irregularly and it is difficult to observe quality changes. To avoid problems, "netiquette" should be observed not to burden the web sites unnecessarily. It is also a good practice to tell an informant (e.g. enterprise) about web scraping in advance. It is also possible to get an alternative interface to the informant's data instead of public website. All in all, it is good to co-operate with the informant about the web scraping of their websites.
- The weighting of data from web scraping was discussed. If there is not any other sources available, the product level weights (e.g. by CPA) may need to be asked directly from the informant.
- It was discussed that it can be difficult to obtain the prices that are actually paid by the buyer. For instance, airline companies usually have corporate programs for business travelers which affect the prices.

### Obtaining alternative data sources

- We discussed also our findings in the project. When starting to look for alternative data sources and negotiations with possible data providers, it is preferable to try to obtain scanner/sales data when possible because usually it is comprehensive. If it is not available, then an interface with the data provider's data is one option. In addition, a web scraper is also an alternative to retrieve data.
- Index compilation needs to be adjusted when using different kinds of data, like scanner data. It is not anymore possible to do quality adjustment by observation. In the project, also an index compilation methodology for utilizing the new data acquired was designed and coded. The index compilation method is based on the principle of pair-comparison.