TERMS OF REFERENCE

Task Team on Scanner Data

Global Working Group on Big Data for Official Statistics

Introduction

The statistical community has the obligation of exploring the use of new data sources, such as scanner data, to meet the expectation of society for enhanced products and improved and more efficient ways of working. The task team on scanner data is created as a separate team, since scanner data is one of the Big Data sources which is used more and more in national statistical systems for the calculation of price indices. Many of the price measurement issues and methods for scanner data from supermarket chains and other retailers apply also to other big data sources. For example, online prices obtained from webscraping and unit values obtained from basic Customs transactions. The deliverables of the task team will also be of use for these other big data sources for price index measures.

Besides the terms mentioned in this document, the Terms of Reference of the Global Working Group (GWG) (see E/CN.3/2015/4) serve as general reference.

Objectives

The objectives of the task team are:

1. the delivery of an open source application with an associated Application Program Interface (API), which can be shared among all partners in the statistical community. This application will take cleaned and classified[1] scanner data[2] (i.e. Analysis Ready Data (ARD)) and will apply a range of analysis and monitoring processes before enabling a range of methods[3] to be used for estimation of price indexes. The exact method can be specified by the user.[4]
2. the development of training and instructional material on the use of the application
3. the development of accompanying methodological guidance material which will a) summarise the relevant literature on methods, b) point to internationally-agreed

---

[1] The EU expert group on scanner data aims to develop open-source tools for the classification of Prices big data from theirprimary sources in unprocessed format, i.e. raw data

[2] Or other big-data such as web-scraped online data, or customs data.

[3] Including 'multilateral' methods that make full use of a set of backdata along with traditional methods such as chained-Jevons and chained Tornqvist, for comparison purposes

[4] i.e. combination of 1. multilateral index method (eg GEKS/CCDI, ITGEKS, FE/TPD, TDH, GK, CLIP), 2. Extension method (eg movement splice, window splice, half-window splice, mean-splice, and fixed base-month extension methods) and 3. Estimation window length.

recommendations on which methods are appropriate in which situations and c) catalogue existing and intended practice across NSIs in the use of Prices big data

The accompanying material 3 will be a living document, and will be circulated to international groups such as the IWGPS, the Ottawa Group and the EU scanner data expert group for review and input.

Recommendations on methodology will be out of scope. The development of an open source API which enables applications of any method will contribute to the empirical and theoretical research still underway.

In other words, the task team will develop documentation, tools and applications, which can be used by the statistical community.

Deliverables and Time Schedule

Within the course of one year, this task team plans to achieve the following:

1. Delivery of an open source application for analysis, monitoring and index estimation from cleaned and classified Prices' big data
2. Accompanying training and instructional material on the use of the application
3. Accompanying methodological guidance material summarising and referencing to literature, recommendations and cataloging good practice