UN Task Team on Scanner Data – 2023 update

Tanya Flower

Chair of the UN Task Team on Scanner Data

Meeting of the Group of Experts on Consumer Price Indices

7th – 9th June 2023

UN Committee of Experts on Big Data and Data Science for Official Statistics

UNBigData

▶ Previously the Global Working Group on Big Data for Official Statistics

▶ Mandated to give direction to the use of Big Data for Official Statistics

Objectives of UN Task Teams

01

Draft methodological guidance (handbooks/notes)

02

Develop method/code library supporting guidance

03

Develop training material

UN Task Team on Scanner Data

AIM: to increase the use of new data sources (web scraped and scanner data) in consumer price statistics...

Task Team homepage

Task Team objectives

- ▶ Workstream 1: Publish a e-handbook summarising best practice and guidance for using these new data sources in consumer price statistics; make code available to NSIs to test out different methods

- ▶ Workstream 2: Develop new guidance material and code for the process of classifying these new data sources for consumer prices

- ▶ Workstream 3: Develop a new training package using the guidance material to promote the use of these new data sources and methods

Contents of e-handbook

Initial considerations
An overview of these new data sources, how to select categories and retailers, IT system requirements and quality assurance

Data acquisition
How to acquire/access scanner, web scraped and other alternative data sources, as well as tips for monitoring quality

Data preparation
Standardising the data, defining "products", treatment of discounts and refunds, deriving proxy weights for web scraped data

Classification
Different approaches to classifying these data sources from manual to more advanced methods

Contents (2)

Data filtering
Methods to identify and remove outliers; filtering out "dump" products and low sales

Price index methods
An overview of methods available to calculate price indices using these new data sources along with some discussion on how to choose an appropriate method

Aggregation
How to integrate these new data sources into the existing classification hierarchy

Implementation
Tips on how to implement these data sources in practice, including managing unexpected gaps in data supply

Classification

- Conducted a survey of Task Team members to identify common techniques used by NSOs

- Completing guidance material and code examples for each technique (to be included within the e-handbook)

- First look at the content to be provided at the two workshop sessions on 8th June

Overview of 5 common methods NSOs tend to use

Focus for session 1

- Method 0: Manual labelling or validation of predicted labels
- Method 1: Attribute based classification method
- Method 2: Pattern matching classification method
- Method 3: Recommendation / Machine-assisted classification

- Method 4: Machine Learning classification method

- Blending classification methods

Focus for session 2

Training

▶ The first course "What is Scanner Data?" has been finalised and launched. Courses are placed onto the UN Learning Management System (UN LMS) which is held on the UN Global Platform. You will need to create an account to log in but all quick and easy to do and once on you can click through the content.

▶ The second course "Obtaining Scanner Data" is near completion and will be previewed at the first workshop session on 7th June

▶ Additional courses are also in development to align to a new extended curriculum which will include support on all new data sources, not just scanner data, in line with the e-handbook

What is Scanner data?

Scanner data (intro)

Competencies

Grades

Overview

Welcome

Course material

Quiz

Feedback

Certificate

Home

Dashboard

Calendar

Private files

My courses

Scanner data (intro)

Module 1 - What is scanner data and how to obtain it?

➡

What is scanner data and how to obtain it

Click here for transcript
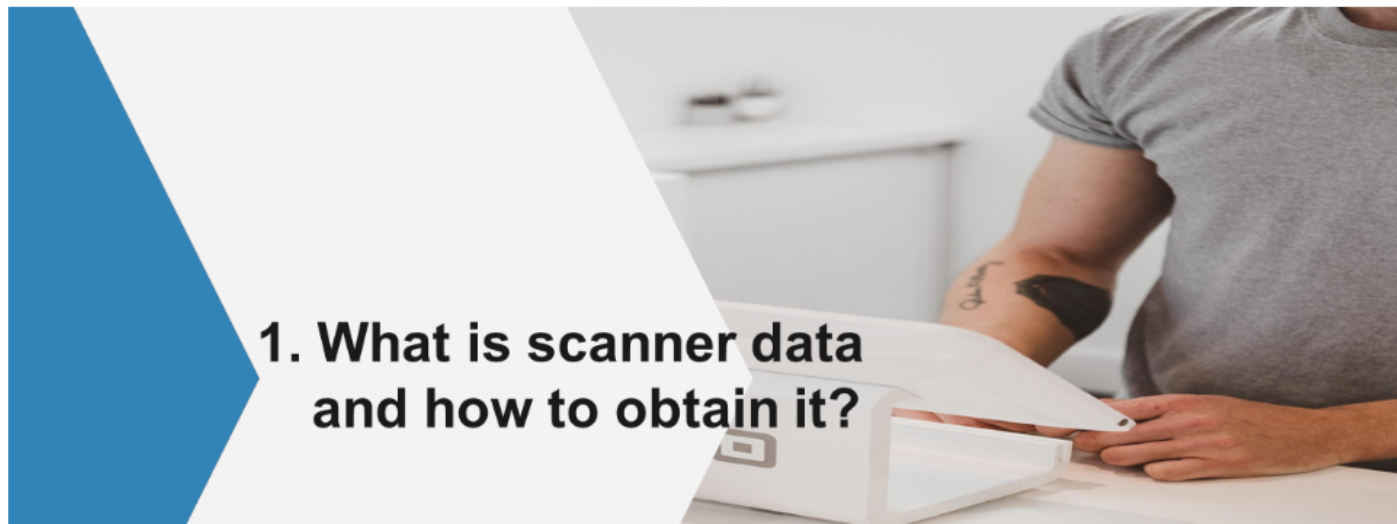
1. What is scanner data and how to obtain it?

Upcoming governance changes

▶ Discussions are currently ongoing between the UN Statistical Division and the members of the Intersecretariat Working Group on Price Statistics (IWGPS) to look at options for supporting the Task Team longer term in order to ensure the legacy of this handbook and training material

▶ In particular, this will ensure the material can continue to be maintained, updated, and used in wider training programs

▶ This has unfortunately resulted in delays to publication of our handbook materials while these discussions are ongoing but we will continue to provide updates when available

What's coming up?

▶ Agree longer term governance structure for the Task Team to ensure the material can continue to be maintained, updated, and used in wider training programs

▶ Update Terms of Reference for the Task Team to cover this structure

▶ Publish e-handbook and continue development of training material

▶ Set up a new workstream to develop the code library on the UN Global Platform, using published code packages to create example notebooks

▶ As new approaches or methods arise, the Task Team can also commission new workstreams who can coordinate input across all three areas. This ensures the materials can be maintained but also expanded as necessary

▶ Engage with international colleagues to identify areas where we can add the most value, for example working with UN Regional Hubs on rolling out the training material to areas with high demand

How can you get involved?

We're always on the look out for new members!

Please contact any of the steering group members:

▶ Tanya Flower (Chair and Workstream 1 lead) tanya.flower@ons.gov.uk

▶ Serge Goussev (Workstream 2 lead) serge.goussev@statcan.gc.ca

▶ Federico Polidoro (Workstream 3 lead) fpolidoro@worldbank.org

▶ Benson Sim (UN Secretariat) simb@un.org

What does membership of the UN Task Team entail?

- You will be allocated to one of the workstreams (partially based on time zone but can take into account preferences as well!)

- Workstreams tend to meet monthly or as needed to discuss progress and tasks

- It depends on individual circumstances but the time commitment expected is around 1 day per month

- *"Taking part in the task team activities has been an excellent experience to get a deeper and broader understanding of the problems, methods and opportunities involved in this rich area while contributing to the development and dissemination of materials that aim to benefit NSOs around the world that are facing the challenges of the adoption of big data sources for the production of price statistics.*

- *Also, participating in an international collaboration with experts with different views and expertise from all over the world has been a rich personal working experience that allows a better integration with the international community of price statistics and knowledge sharing by means of the discussions in the working sessions. "*

Vladimir Goncalves Miranda, Brazilian Institute of Geography and Statistics (IBGE)

Useful links

- [Task Team homepage](#)

Many NSOs have used either Python or R to write any new systems required for processing these data sources due to their open source nature. For examples of how these methods can be implemented, there are some available packages namely:

- R: [PriceIndices](#)
- R: [IndexNumR](#)

Thanks for listening! And a reminder of our workshops…

Session	Presenters	Time/ Date
Introduction to the acquisition of scanner data for CPI	Kristiina Nieminen, Statistics Finland, and Federico Polidoro, World Bank	Session 1, Wednesday 7 June, 14.00 –14.30
Scanner data processing and new index formulas in the latest version of the *PriceIndices* package	Jacek Białek, University of Lodz/Statistics Poland	Session 2, Wednesday 7 June, 14.30 –15.00
Getting started with focus on simpler methods for classifications	Serge Goussev, Statistics Canada; Kjersti Hov, Statistics Norway; and Jens Mehrhoff, IMF	Session 3, Thursday 8 June, 14.00 –14.30
Advanced classification methods: use of machine learning	Serge Goussev, Statistics Canada; Kjersti Hov, Statistics Norway; and Jens Mehrhoff, IMF	Session 4, Thursday 8 June, 14.30 –15.00