



UN Task Team on Scanner Data update

Tanya Flower

Chair, UN Task Team on Scanner Data

Systems Architecture Survey webinar

21st January 2025

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



- ▶ Previously the Global Working Group on Big Data for Official Statistics
- ▶ Mandated to give direction to the use of Big Data for Official Statistics
- ▶ Other Task Teams include mobile phone data, earth observation data etc
- ▶ <https://unstats.un.org/bigdata/index.cshtml>

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 – summer 2024 refresh



AIM: enable and expand the practical usage of alternative data sources (including scanner) in consumer price statistics. Future work will look to develop the wider usage of these data (for example, in household expenditure statistics)

Task Team [homepage](#)

Refreshed objectives

- ▶ Previous objectives were tied closely to the three workstreams. As part of the work to plan out the additional workstreams, we have also taken the opportunity to refresh our objectives to ensure they are better representative of our longer-term aim
- ▶ We have also introduced two principles to inform the delivery of our objectives:
 - *Timeliness*. Ensure that any content produced by the Task Team in line with the linked objectives remain up to date with current discussions in the field. For this purpose, the Task Team will be supported by a Secretariat function which will be responsible for maintaining the existing published content.
 - *Informed choice*. Provide a range of options for these topics, so we should not recommend a single “best” approach - instead colleagues will be encouraged to review the options and decide what is best for their particular situation.

Refreshed objectives

Aligned to the mandate of the [UN Committee of Experts on Big Data and Data Science for Official Statistics](#), the specific objectives of the Task Team cover the following:

- ▶ *Repository of information.* Provide guidance on the end-to-end methods and processes of using these data. The guidance should reference existing literature where possible.
- ▶ *New guidance material.* Provide methodological guidance for those topics where existing literature does not exist or does not fully capture the different range of possible options for consideration.
- ▶ *Practical implementation.* Provide practical guidance on the implementation of these methods and processes in a live production setting.
- ▶ *Training.* Provide a full suite of training materials to supplement the guidance produced above.

Workstreams

- ▶ The Task Team has been split into several workstreams that are aligned to the objectives.
- ▶ We have one workstream that has been completed this year (handbook).
- ▶ The 4 in progress workstreams are as follows:
 - Classification workstream
 - Training workstream
 - System architecture workstream
 - FAIR workstream
- ▶ As workstreams complete their deliverables, in future we will be able to open up more workstreams in line with our ongoing objectives.

Handbook workstream

- ▶ Status: **COMPLETE**
- ▶ Deliverables:
 - Produce an e-handbook (wiki) on using alternative data sources (ADS) to produce consumer price statistics.
 - This was delivered in May 2024: <https://unstats.un.org/wiki/display/GWGSD/>
- ▶ Timeline:
 - No further deliverables planned for this workstream specifically. New handbook material and updates will be coordinated by the Secretariat, in particular we would like to focus initial efforts to review the price index chapters (specifically the multilateral and extension methods pages) to incorporate more recent research from the annual round of conferences.

Classification workstream

- ▶ Status: **in progress**
- ▶ Deliverables:
 - Publish guidance on potential methods available for classifying scanner data to produce data ready for price index compilation - via chapters in the handbook
 - Code notebooks to demonstrate each method
- ▶ Timeline:
 - Presentations have been made on the progress to date at the annual Consumer Prices conferences (UNECE and Ottawa). Links are available in the summary page here: <https://unstats.un.org/wiki/display/GWGSD/Classification>
 - April 2025 (UNECE meeting): remaining content ready for publication in the handbook, and code notebooks if there is time

Training workstream

- ▶ Status: **in progress**
- ▶ Deliverables:
 - Curriculum of training requirements, aligned to the handbook material (**COMPLETE**)
 - End to end training programme of online courses, along with recorded voiceovers and exercises to test a participants understanding, to be published on the UN learning platform
- ▶ Timeline:
 - April 2025 (UNECE meeting): material published for data acquisition and index methods courses

System architecture workflow

- ▶ Status: **in progress**
- ▶ Deliverables:
 - Summarise existing practices on how countries are setting up their IT infrastructure in order to be able to incorporate alternative data sources into consumer price statistics
 - Provide guidance for future development activity linked to given transformation objectives
- ▶ Timeline:
 - Winter 2024/25: run survey for NSOs to collect existing experience
 - April 2025 (UNECE meeting): present on results of survey, linked to guidance for future development activity and specific case studies
 - 25/26: write up content and publish via handbook, consider possible training materials to support this guidance

FAIR workstream

FAIR data is data which meets the principles of findability, accessibility, interoperability, and reusability (FAIR)

- ▶ Status: **in progress**
- ▶ Deliverables:
 - Develop a data catalogue tool for open datasets, to include publishing several open datasets (for example, Dominick's data)
 - Produce some high level guidance for researchers on how the FAIR principles can be used in the field of consumer price statistics and alternative data sources
- ▶ Timeline:
 - April 2025 (UNECE meeting): present proof of concept data catalogue and draft guidance
 - 25/26: add additional datasets to the data catalogue, publish guidance on the handbook and flush out process on FAIR in the discipline, consider possible training materials to support this guidance

Longer term planning

- ▶ The deliverables listed here are medium priority for the Task Team - it is unlikely we can take these forward until some of the other workstreams complete unless we can identify some additional capacity
 - Discussion forum - we would like to identify a way to enable communication between practitioners in this field in order to share knowledge and seek assistance
 - Seminar series - again, to enable communications between international colleagues, a seminar series could be another way to share results and feedback between the annual conference cycle
 - Code notebooks - following the model of the classification workstream, we would like to provide additional notebooks covering a wider range of applicable methods to use alternative data sources for consumer price statistics
 - Other uses of these prices alternative data - these data can be used for other statistics not just consumer prices (for example, scanner data for household expenditure). We would like to start collating guidance on this usage.

Actions for you!

- ▶ Please complete the Task Team questions at the end of the survey so we can get a better idea on topics that would be helpful to focus on/ key priorities etc
- ▶ If you would like to join any of the Task Team workstreams or have other suggestions, please get in touch!

Tanya Flower (Chair) tanya.flower@ons.gov.uk

Serge Goussev (Classifications and FAIR lead) serge.goussev@statcan.gc.ca

Federico Polidoro (Training lead) fpolidoro@worldbank.org

Collin Brown (Systems architecture lead) collin.brown@statcan.gc.ca

Vladimir Goncalves Miranda (Secretariat lead) vladimir.miranda@ibge.gov.br

Thanks for listening!
Any questions?

The background features a series of overlapping, semi-transparent geometric shapes, primarily triangles, in various shades of blue and green. These shapes are positioned on the right side of the slide, creating a modern, abstract design. The colors range from light, airy blues to deeper, more saturated greens and blues.