



UN Task Team on Scanner Data – 2024 update

Tanya Flower, Federico Polidoro, Serge Goussev
Steering group for the UN Task Team on Scanner Data

18th Meeting of the Ottawa Group on Price Indices

13th - 15th May 2024

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

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 NSOs 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

e-handbook

- ▶ This handbook has been developed to provide a useful source of information on using new data sources in the production of consumer price statistics

The handbook is now available to the public!!!

Link here: <https://unstats.un.org/wiki/display/GWGSD/>

More information will be provided in our 1st workshop session

- ▶ Handbook is not prescriptive, instead it aims to provide an overview of relevant information for colleagues to refer to in their own work and decide what is best for the particular environment/project which they are working within
- ▶ References existing literature but also expands some topics where there is less guidance available currently (for example, data acquisition and classification).
- ▶ Given the pace at which this topic evolves, the aim is that this handbook will be updated over time to reflect the most recent analysis and findings
- ▶ The current (live) version of the handbook is still being finalized and updated, with some topics (i.e. classification) still in the works. The material already written should provide a large amount of value to NSOs, hence we are making it available, even as we keep updating and expanding it.

Classification

- ▶ Conducted a survey of Task Team members to identify common techniques used by NSOs
- ▶ Finalizing written guidance on classification methods and key considerations on applying them (to be published as part of the handbook)
- ▶ Looking to develop code notebooks to demonstrate how classification method and their application
- ▶ Join the classification workstream this afternoon!

Content covered in this workshop

- ▶ Quick contextual discussion to review what was covered in 2023 CPI Expert Group workshop
- ▶ Overview of other considerations applicable to classification:
 - How to implement classification in production
 - Blending classifications or the blended classification approach
 - Metadata and replicability best practices for classification step
 - Considerations when dealing with taxonomy changes
- ▶ Discuss the work planned over the coming year from workstream 2

Training

- ▶ The first course “Alternative Data Sources to compile CPI: an overview” has been finalised and will be launching shortly.
- ▶ 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 course is temporarily available here for reference <https://moodle-2.dev.officialstatistics.org/> with login access available soon. Note this is a test environment so user data is not guaranteed to be kept when we have transferred to the live site
- ▶ The second course “Alternative Data Sources acquisition” is near completion and will be made available by end of June. By July that on price indices (bilateral and multilateral) will also be available
- ▶ 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.
- ▶ We will be covering the training curriculum in more detail in our 2nd workshop this afternoon! You will also be able to find up to date links and information on our [handbook training page](#)



UN Big Data Learning

Scanner Data Characteristics

UN-CEBD Task Team
on Scanner Data

Upcoming changes (1/2)

- ▶ A solution to guarantee the continuity of the governance of the e-handbook and of the related materials (training courses, classification tools) is in progress to be implemented and should take place in the next months
- ▶ The solution is based on the establishment of a Secretariat that will support the Task Team in taking care, with the due continuity, of maintenance of the handbook, training materials, applications, as well as of the wiki and the task team's website.
- ▶ The Secretariat should be placed at the Brazilian UN Regional Hub on Big Data and Data Science for Official Statistics, with support of an expert co-funded by the United Nations, the World Bank and the International Monetary Fund (IMF)
- ▶ The Secretariat would operate closely with the TT on the basis of a written agreement, which would be signed by the statistical office of Brazil (IBGE) and the UN Statistics Division on behalf of UNCEBD, as an extension of the existing MOU for the Brazilian Regional Hub.
- ▶ The TT will inform the IWGPS on progress and activities.

Upcoming changes (2/2)

Main tasks of the Secretariat in supporting the TT

- Makes marginal updates to keep evergreen the developed handbook, code notebooks, or training materials as required.
- Consults with the TT on areas where the e-handbook and the training materials have to be revised, updated or enhanced
- Adheres to time commitments and responsibilities, arranges meetings and agendas, and follows up on actions.
- Reviews the materials produced by the TT and provides sign off that these materials are recommended for compilation of consumer price statistics, alongside best practice guidelines and practical advice for implementation.
- Supports the finalization of the Training materials and coordinates how best the current task team-developed training material can be used to support existing training programmes in the use of alternative data sources for CPI that are in place at the international and regional agencies, including in the context of ICP
- Promotes TT work with international agencies with the aim of supporting their goals of capacity building using the pre-prepared Task Team training material.
- Coordinates Training of the TT with other UN Regional Hubs on Big Data and Data Science for Official Statistics in Indonesia, Rwanda and the UAE

What's coming up?

- ▶ In the coming months, we are looking to launch a new workstream on systems and architecture best practices!
 - With the scale of ADS adoption, open-source tools, and cloud versus on-premises hardware options - many of us are asking questions on how to best operationalize ADS effectively and what best practices are.
 - Working with international organizations, we could aim for objectives such as a survey of current NSO approaches with an output report or a discussion and documentation of best practices to achieve the target state. Developed materials could be available in the handbook and would thus also be evergreen, helping NSOs learn from each other. We could look to share interim results at 2025 UNECE CPI Expert Group
 - If this sounds like it's of interest, give us a shout!
- ▶ Other workstreams/additional work:
 - Develop code packages and demonstration notebooks on public data so that NSOs can see the operational aspects to the methodological aspects in the handbook
 - Outline best practices other uses of alternative data, such as for weights, retail sales, PPPs, etc

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](#)
- ▶ [Task Team handbook](#)

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...

This will run in parallel with the poster session (10.30 - 12.30 May 14th) and we have included time to allow attendees to move between the poster session and relevant workshop sessions

Session	Presenters	Time/ Date
Publication launch of the handbook on utilising new data sources in the production of consumer price statistics	Tanya Flower, Office for National Statistics	Session 1, Tuesday 14 May 10.40 - 11.10
An overview of alternative data sources (ADS), including the features of an ADS project	Federico Polidoro, World Bank	Session 2, Tuesday 14 May 11.15 - 11.45
Classifying alternative data for consumer price statistics	Serge Goussev, Statistics Canada	Session 3, Tuesday 14 May 11.50 - 12.20