Concept Note

Regional Workshop on Web-scraping for Consumer Price Statistics

16-20 September 2024

Background

The ESCAP Committee on Statistics decided at its 7th Session decided to "feature big data for official statistics in its future work, with an emphasis on sharing country research, experiences and good practices and facilitating capacity development" and "strengthening legislative provisions and institutional mechanisms to enable national statistical systems to take full advantage of new and innovative technologies while respecting the Fundamental Principles of Official Statistics." In implementing this decision, ESCAP is running a capacity development project entitled *the 2030 Data Decade - Strengthening the institutional capacity of national statistical offices in Asia and the Pacific to use innovative, new and big data sources for official statistics in support of the 2030 Agenda for Sustainable Development* (the Big Data Project) amongst others. One component of this project involves capacity development in the area of web scraping prices focusing on consumer price indices in Azerbaijan, Cambodia, Fiji, Lao PDR, Maldives, Nepal, Thailand, Tonga and Uzbekistan.

Web scraping and Consumer Price Indices

Globally, e-commerce accounts for a growing proportion of all retail sales, a trend that was supercharged during the COVID-19 pandemic. This trend was accompanied by a rapid rise in online retail activities. Production of price statistics such as consumer price indices (CPI) may be adapted by collecting online prices in addition to data collection in outlets and markets to ensure data remains representative. This adaptation has several potential additional advantages, including reducing data collection costs and response burdens.

From May 2024, ESCAP has organized a series of training webinars on web scraping for consumer price indices. The webinars have covered sharing of experiences, challenges of web scraping, ethical issues in web scraping, the development and application of ICT tools for web scraping in Python. This training process has also included a set of homework exercises. Participants have also benefited from the expertise mentors to enhance their knowledge.

ESCAP is organizing a follow-up in person workshop from the webinars in Bangkok, Thailand on 16-20 September 2024, and which is the focus on this concept note. The overall aim of the workshop is to strengthen web scraping skills acquired through a series of webinars undertaken since May 2024.

Objectives and Expected Outcomes

The in-person workshop's overall aim is to embed the knowledge and skills delivered to participants via the webinars May through August 2024. By the end of the workshop, participants will:

- Have embedded Python coding knowledge and skills for web scraping prices data and applied this to their chosen website.
- Understand how to process and clean web-scraped data, ready for inclusion into the CPI calculation process.
- Have discussed further and explored coding to automate a pipeline for incorporating web scraped prices data into the NSO's processes.
- Have a deeper appreciation of the methodological challenges for incorporating web scraped prices into CPI calculations, and how to overcome these.

Target Participants and Travel Arrangements

ESCAP invites two participants from each of the following countries: Azerbaijan, Cambodia, Fiji, Lao PDR, Maldives, Nepal, Thailand, Tonga and Uzbekistan. Nominated experts must have attended all webinars, submitted and passed homework exercises, have engaged actively with their mentor and contributed in the online "Slack" platform. Nominations should be sent to ESCAP by **15 August 2024** by email to stat.unescap@un.org with a copy to nyasulua@un.org and nannapas.sukwattanipaat@un.org. ESCAP will provide return economy air tickets by the most direct route and pay a daily subsistence allowance in line with United Nations rules and regulations. Nominated participants are expected to make their own arrangements regarding travel or medical insurance. Further information will be provided to nominees that have been accepted.

Tentative Training Programme

Mon 16 Sep – Fri 20 Sep 2024				
Mon 16 Sep **Web scraping for CPI – Day 1**	**Morning** (09:00 – 12:30)	**Welcome and Intro** • Meet and greet everyone in person – round the room intro • Remind of objectives of the course – what are we setting out to achieve this week **Session** • Recap Python programming concepts from virtual sessions • Recap on methodologies • Discuss country specifics & issues encountered • Independent work by country teams, mentorship provided on specific cases	NSO participants	Frances Krsinich & other experts
	Afternoon (13:30 – 17:00)	**Session** • Processing of scraped data, cleaning and database insertion • Independent work by country teams, mentorship provided on specific cases	NSO participants	Frances Krsinich & other experts
Tue 17 Sep **Web scraping for CPI - Day 2**	**Morning** (09:00 – 12:30)	**Session** • Independent work by country teams, mentorship provided on specific cases • Presentation preparation	NSO participants	Frances Krsinich & other experts
	Afternoon (13:30 – 17:00)	**Session** • Independent work by country teams, mentorship provided on specific cases • Presentation preparation	NSO participants	Frances Krsinich & other experts
Wed 18 Sep **Web scraping for CPI – Day 3**	**Morning** (09:00 – 12:30)	**Session** • Reproducible Analytical Pipelines • Hands-on practical session in Python	NSO participants	Christophe Bontemps & Frances Krsinich

	Afternoon (13:30 – 17:00)	**Session** • Independent work by country teams, mentorship provided on specific cases • Presentation preparation	NSO participants	Christophe Bontemps & Frances Krsinich
Thu 19 Sep **Web scraping for CPI – Day 4**	**Morning** (09:00 – 12:30)	*NSO presentations of results & next steps – 10 mins max* • Share results of web scraping your chosen site • Challenges & desired improvements • Next steps • Q&A	NSO participants	Frances Krsinich & other experts
	Afternoon (13:30 – 17:00)	*NSO presentations of results & next steps – 10 mins max* • Share results of web scraping your chosen site • Challenges & desired improvements • Next steps • Q&A	NSO participants	Frances Krsinich & other experts
Fri 20 Sep **Web scraping for CPI – Day 5**	**Morning** (09:00 – 12:30)	• Further NSO presentations if more time required • Discussion around other applications for web scraped data at the NSO • Evaluation of training programme • Next steps	NSO participants	Frances Krsinich & Ceri Regan