UNITED NATIONS ESCAP
Economic and Social Commission for Asia and the Pacific

STATS BRIEF

March 2021 | Issue No. 28

STATISTICS DIVISION

# Big data for economic statistics

This Stats Brief gives an overview of big data sources that can be used to produce economic statistics and presents country examples of the use of online price data, scanner data, mobile phone data, Earth Observations, financial transactions data and smart meter data to produce price indices, tourism statistics, poverty estimates, experimental economic statistics during COVID-19 and to monitor public sentiment.

The Brief is part of ESCAP's series on the use of non-traditional data sources for official statistics.

*This Stats brief is prepared by Irina Bernal, Consultant, UNESCAP Statistics Division, and Tanja Sejersen, Statistician, UNESCAP Statistics Division, with valuable inputs from Rikke Munk Hansen, Chief of Economic and Environment Statistics, UNESCAP Statistics Division, and Alick Mjuma Nyasulu, Statistician, UNESCAP Statistics Division.*[i]

## Introduction

Economic statistics describe the state of an economy and span several domains and sectors, such as business statistics, macroeconomic statistics, economic accounts, and energy, industry, financial, tourism, labour and trade statistics. Traditionally, much economic data are collected through surveys of individuals, households, and firms; other data are collected from administrative systems such as business registers and tax and customs records.

Technological progress and digital activities generate a wealth of data. These new data sources can provide complementarity to traditionally collected data, bringing greater granularity, higher frequency or even new insights. Among such data sources are scanner data and online price data for generation of price indices, mobile positioning data (MPD) for tourism statistics, Earth Observations for poverty estimation, smart meter data for electricity consumption, financial transactions such as point of sale (POS) data for consumption and economic trends and online job postings for labour market estimates.

Big data has been identified as a priority by the ESCAP Committee on Statistics. National Statistical Offices (NSO) in Asia and the Pacific have been using and experimenting with the use of various big data sources to produce official statistics, including economic statistics. The COVID-19 provided an added impetus for the exploration of big data sources, leading to new partnerships and collaboration among public and private institutions and to the production of statistics. While the use of big data for economic statistics remains at an exploratory phase for many NSOs, some big data sources are already fully integrated into the production of economic statistics by some NSOs.

This brief highlights examples of the use of big data sources for generating economic statistics from the national statistical systems, mostly the National Statistical Offices (NSO), in Asia and the Pacific. Examples include the compilation of **price indices**, such as consumer price index (CPI) and residential property price index (RPPI) using online price data and scanner data, **tourism statistics** using mobile network operator (MNO) data, **poverty estimates** using Earth Observation (EO) data, and **experimental economic statistics** during COVID-19 using various big data sources. Following the presentation of country examples, the brief discusses the main challenges NSOs are confronted with when accessing or using big data sources and provides recommendations on how to address them.

## Big data sources that could support the calculation of economic statistics

NSOs globally are collaborating and exploring various big data sources to either supplement or generate economic statistics. Through the UN Committee of Experts on Big Data and Data Science for Official Statistics with several Task Teams centered on specific big data sources, such as Earth Observation data, mobile phone data, scanner data and AIS (Automatic Identification System) data, NSOs are collaboratively exploring big data and supporting each other in addressing technical, legal, and methodological issues.

In Europe, the European Commission leads the ESSnet Big Data which has 11 work packages covering big data sources such as online job vacancies, smart energy, tracking ships, financial transactions data, Earth Observation data. ESSnet Big Data provides a platform for knowledge exchange among its members.[1] More importantly, the countries' methodologies for individual data sources are published as open source and shared in the public domain.

Unlike the European counterparts, NSOs in Asia-Pacific have so far been using or exploring the use of big data on their own or as part of global collaboratives. NSOs, such as BPS Statistics Indonesia and the Australian Bureau of Statistics, have been documenting and actively sharing their experience in the use of big data for official statistics. Still, multiple NSOs in the region have no experience in the use of big data but have big data efforts planned in their national statistical programmes and are determined to research new data sources and methods.

Big data for price statistics is actively explored by NSOs in the region. The two main big data sources used are scanner data and online price data. **Scanner data** are detailed data on the sales of consumer goods obtained from scanned bar codes of individual products at the electronic points of sale (POS) in retail outlets.[2] Scanner data contain information about the characteristics of the products, their prices and quantities sold. **Online price data** are data collected electronically from retailers' websites. Due to easier accessibility, web scraped data from online marketplaces are more widely used than scanner data. At the global level, the potential of big data for the production of the CPI is explored by the International Working Group on Price Statistics, also known as the Ottawa Group, and the UN Global Task Team on Scanner Data, which counts Australia, New Zealand, Thailand, and Turkey among its members.

**Mobile phone data** are used in the production of tourism statistics and migration statistics. One can distinguish between high resolution data, such as signaling data, and call detail records (CDR). *Signaling data* include data from position updates every time the cell tower changes, or active positioning data, where the mobile phone is pinged and its location is determined through the help of A-GPS (assisted GPS). *CDR* contain information such as time, duration, source number, destination number and approximate location of communications about a telephone call or other telecommunication transaction (i.e. text message) that passes through the device and is registered by the telecom operator.[3] In addition to tourism statistics, social and demographic indicators, such as urban mobility, domestic migration, and population characteristics can also be derived from mobile phone data.

The UN Global Task Team on Mobile Phone Data, which counts Georgia, Indonesia, Philippines, and the Republic of Korea among its members, developed the ***Handbook on the use of Mobile Phone data for Official Statistics*** which explains the use of mobile data in the production of official statistics, such as tourism and events statistics, population and migration statistics, commuting statistics, traffic flow and employment monitoring, and also provides examples of partnership models with telecom operators.[4] BPS

[1] Eurostat, ESSnet Big Data II, https://ec.europa.eu/eurostat/cros/content/essnet-big-data-0_en

[2] OECD, Glossary of Statistical Terms, "Scanner Data", https://stats.oecd.org/glossary/detail.asp?ID=5755#:~:text=Scanner%20data%20constitute%20a%20rapidly%20expanding%20source%20of,being%20used%20increasingly%20for%20purposes%20of%20hedonic%20analysis.

[3] Positium, "How Can Mobile Positioning Data Be Used for Mobility Studies?", https://positium.com/blog/how-can-mobile-positioning-data-be-used-for-mobility-studies

[4] UN Global Working Group on Big Data for Official Statistics, Handbook on the Use of Mobile Phone Data for Official Statistics, September 2019, https://unstats.un.org/bigdata/task-teams/mobile-phone/MPD%20Handbook%2020191004.pdf

Statistics Indonesia and the NSO of Georgia (Geostat) have developed methodologies for using mobile phone data for tourism statistics and BPS contributed to the development of the Handbook.

Other stakeholders, such as international organizations, for example Global Pulse and its regional lab in Asia-Pacific – Pulse Lab Jakarta; non-government organizations, such as Flowminder and LIRNEasia; and private companies, like Positium, have been supporting governments with insights from mobile network operators' (MNO) data and providing technical guidance on methodologies and privacy-preserving techniques.

In addition to scanner data, online price data and mobile network operator data, other big data sources, which the NSOs in the region are also exploring, can contribute to the production of economic statistics. **Earth observations** (EO) from satellites, aircrafts and UAV (unmanned aerial vehicles, also referred to as "drones"), such as nighttime light images, can provide data relevant to estimating economic activity. **Smart meter data** on electricity consumption and production from individual households and companies could supplement other statistics, such as energy statistics of businesses, tourism seasonality or impact on the environment.[5] **Financial transactions data** obtained either from the Central Bank, the Tax Authority or directly from the payment operators can be indicative of the health of the economy and its trends. Vessel traffic data, or Automatic Identification System **(AIS) data** can be indicative of the marine human activity and provide a proxy for maritime trade.

## Experience of the NSOs in Asia and the Pacific in using big data for economic statistics

Of the NSOs in Asia Pacific that use big data sources for economic statistics, most do so experimentally, with only a few moving towards big data integration into the regular production of statistics. The most promising and ripe area for big data integration is price statistics. Furthermore, the COVID-19 pandemic increased demand for timely and granular statistics which led to the production of new experimental economic statistics.

This section presents the experiences of 12 countries (Azerbaijan, Australia, Georgia, Indonesia, Japan, Malaysia, Mongolia, New Zealand, Philippines, Russian Federation, Thailand, and Viet Nam) in using big data to produce price indices, tourism statistics, poverty estimates, experimental economic statistics during COVID-19, and sentiment analysis. The big data sources explored are online price data, scanner data, mobile phone data, Earth Observations, financial transactions data, tax data, payroll systems, smart meter data, and social media data.

### 1. Compilation of price indices (CPI and RPPI) using big data sources

The lockdowns imposed across multiple countries during the COVID-19 pandemic limited the NSOs' ability to compile price statistics through traditional surveys. The lockdowns have also increased online purchases through the e-commerce platforms. Thus, NSOs received two simultaneous incentives for exploring online price data, particularly in countries with high internet and e-commerce penetration rates.

Most of the NSO examples of integrating big data in the production of economic statistics concentrate on price indices using scanner data, online price data or tax data. As the nature of these data sources differs, so do country experiences in accessing them.

Access to scanner data requires negotiation and agreement with retailers which can be challenging. Whereas automated and regular access to online price data may also require agreements with website owners, these are generally easier to obtain. The NSOs of Australia, Georgia, Japan and New Zealand have access to scanner data and online price data. Most other NSOs in Asia and the Pacific are exploring online price data. As the use of scanner data and web scraped data gain traction in the production of the CPI and as payments become increasingly digitized and e-commerce grows, more countries in the region are considering these non-traditional data sources.

[5] Eurostat, ESSnet Big Data, https://webgate.ec.europa.eu/fpfis/mwikis/essnetbigdata/index.php/ESSnet_Big_Data

### 1.1 Compiling Consumer Price Index (CPI) using scanner data

The NSOs of Australia, Japan, and New Zealand have integrated scanner data into the compilation of price changes of some products in the CPI basket, while the NSO of Georgia is developing the methodology for this data source.[6]

The **Statistics Bureau of Japan (SBJ)** was among the first NSOs to implement scanner data into the CPI. The SBJ implemented a hedonic model using scanner data to measure price change of personal computers in 2000 and of digital cameras in 2003. Price indices of TV and mobile phones using scanner data were among the more recent trial calculations.[7]

The **Australian Bureau of Statistics (ABS)** has been seeking ways to utilize big data for compilation of economic statistics since 2011. It undertook extensive testing and experimentation before introducing scanner data in the production of the official CPI in a phased approach in 2014. Scanner data from retail transactions currently accounts for over 25% of the weight of the Consumer Price Index.[8]

**Statistics New Zealand (Stats NZ)** started developing methods for measuring price change for goods and services using scanner data in 2011. Among the first categories considered were consumer electronic goods as they experience significant technological changes over a short period of time, which is challenging to measure with traditional data methods. By 2014, Stats NZ implemented scanner data for a range of consumer electronic products into the official CPI.[9]

**The National Statistics Office of Georgia (Geostat)** concluded agreements with 2 supermarket networks for regular data access in 2020. The COVID-19 pandemic accelerated the switch from physical price data collection to electronic data transmission by retail networks. Together with item descriptions, price and quantity data, supermarket chains provide corresponding internal classification of items, which is then used by Geostat to create an EAN (International Article Number) – COICOP (Classification of Individual Consumption According to Purpose) convertor. The convertor is updated every month after data transition. Geostat is considering extending the partnership to other supermarkets.

**The Federal State Statistics Service of the Russian Federation (Rosstat)** has also been exploring alternative data sources for compiling consumer price statistics. By 2018, Rosstat had been negotiating for several years with various owners of prices data, both in the private sector and the government, on the possible use of their information to compile the CPI.[10] Rosstat concluded agreements with several retailers for obtaining scanner data. However, the codification of products in the receipts varies across the systems of different retailers and Rosstat is currently exploring how Artificial Intelligence can address this issue. Furthermore, Rosstat has also attempted to access the point of sale (POS), which the Federal Tax Service collects online from economic agents across the country. However, obtaining access to the same data collected by the Federal Tax Service from retailers would require changes to the legislation.

The abovementioned statistical offices gained access to data in different ways. The **ABS** and **Geostat** obtained scanner data directly from the retailers, which required building a good relationship with them and assuring confidentiality through a memorandum of understanding. The **SBJ** and **Stats NZ**, on the other hand, started by acquiring data from market research companies. Several other NSOs have also tried to negotiate access to scanner data with retail networks

[6] Detailed information on the experience of the statistics offices of Australia, Japan and New Zealand can be found in the ESCAP's report Incorporating Non-traditional Data Sources into Official Statistics: The case of consumer price indexes. Lessons and experiences from Australia, Japan, and New Zealand, 2020 https://www.unescap.org/sites/default/files/incorporating_non_traditional_sources_CPI.pdf

[7] Statistics Bureau of Japan, "Improving the accuracy of the CPI by using big data in Japan", https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/Improving%20the%20accuracy%20of%20the%20CPI%20poster.pdf

[8] Australian Bureau of Statistics, Web Scraping, https://www.abs.gov.au/websitedbs/d3310114.nsf/home/web+scraping+in+the+CPI

[9] ESCAP, Incorporating Non-traditional Data Sources into Official Statistics: The case of consumer price indexes. Lessons and experiences from Australia, Japan, and New Zealand, 2020 https://www.unescap.org/sites/default/files/incorporating_non_traditional_sources_CPI.pdf

[10] Federal State Statistics Service (Rosstat), Consumer price statistics in Russia, 2018 https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Russian_Federation_ENG_ppt.pdf

but are currently limiting their experimentation to web scraped data.

## 1.2 Compiling CPI using online price data

Statistical offices in the region are progressively exploring web scraping of online price data to produce CPI.

Since 2017, the **ABS** has been incorporating web scraped data in the production of CPI in sub-indexes for clothing[11], hardware, alcoholic beverages, and small appliances.[12] In late 2018, the ABS began research into using automated methods to process web scraped data for clothing and footwear products. The ABS estimates there is the potential to use web scraped data for up to 20-25 per cent of the CPI.[13] The ABS conducts online data scraping in-house to retain control of the process.

**Stats NZ** has been conducting online data collection trials since 2017, concluding that more data was available for goods than for services.[14] As a trial, Stats NZ also created a digital Food Price Index from web-scraped data from two supermarket chains.[15] The main issue was the chain drift in an unweighted series. Since then, Stats NZ has been approaching supermarket chains directly, to negotiate the supply of data. Stats NZ has been using a mix of obtaining web-scraped data from third-party providers, as well as developing its own in-house capacity.

**SBJ** examined the calculation of airplane fares by using web-scraped data obtained from a third-party data provider. Data of all the flights in each route (price, departure date, route, etc.) were collected every day from the reservation websites of the major Japanese airline companies. Prices for the same airline tickets were collected for three time-intervals to register different discount rates. Web scraped data was then verified against the information on sales obtained through interviews with the main airline companies. Among the insights of this experiment was the need to obtain companies' approval for regular data collection from their website as well as obtaining their sales amount to update the weights.[16]

**Department of Statistics Malaysia (DOSM)** has developed an internal portal for Price Intelligence (PI) with the aim of modernizing data collection tools for improving the quality of the CPI. DOSM adopted web crawling and web scraping techniques for prices from the relevant websites. The PI Dictionary was developed based on the best match to Malaysia COICOP (6/7 digits) based on item descriptions using text matching scores. During the COVID-19 pandemic, DOSM used online data for domestic flights airfare prices and for 20 main CPI products from retailer websites during the Movement Control Order. DOSM also uses online House Rental Price data for the International Comparison Program (ICP). For the ICP2017 submission, 45.6% of the Housing Rental Data was based on online price data with 19 housing specifications.[17]

In addition to the CPI compilation from web scraped data, DOSM also performs internal analysis related to predictive analytics for market basket analysis, forecasting price change effects on indices, CPI

[11] https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/Experimental%20clothing%20indexes%20pres.pdf

[12] ESCAP, Incorporating Non-traditional Data Sources into Official Statistics: The case of consumer price indexes. Lessons and experiences from Australia, Japan, and New Zealand, 2020 https://www.unescap.org/sites/default/files/incorporating_non_traditional_sources_CPI.pdf

[13] Australian Bureau of Statistics, "Web scraping in the Australian CPI", April 29, 2020, https://www.abs.gov.au/articles/web-scraping-australian-cpi

[14] Statistics New Zealand, presentation "Towards a big data CPI for New Zealand", Ottawa Group 2017, https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/4a256353001af3ed4b2562bb00121564/1ab31c25da944ff5ca25822c00757f87/$FILE/Towards%20a%20big%20data%20CPI%20for%20New%20Zealand%20-Alan%20Bentley,%20Frances%20Krsinich%20-%20Presentation.pdf

[15] Statistics New Zealand, "Creating a digital Food Price Index from web-scraped data", https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/Creating%20a%20digital%20Food%20Price%20Index%20poster.pdf

[16] Statistics Bureau of Japan, "Improving the accuracy of the CPI by using big data in Japan", https://www.ottawagroup.org/Ottawa/ottawagroup.nsf/home/Meeting+16/$FILE/Improving%20the%20accuracy%20of%20the%20CPI%20poster.pdf

[17] Department of Statistical Malaysia, Mazliana Mustapa's presentation "Leveraging online price data from web crawling", ESCAP Stats Café Series, November 30, 2020 https://www.unescap.org/sites/default/files/Leveraging_online_price_data_from_web_crawling_Malaysia_Stats_Cafe_30Nov2020.pdf

simulation and forecasting by group, state, and location, estimation of duration of price and specification changes and prediction of weather and exchange rate effects on CPI.[18] DOSM also provides a Dataset Generator service, which allows internal users to download data on a specific product for a selected time frame.

The **Philippine Statistics Authority (PSA)** has begun the experimental web scraping of prices from selected online stores since February 2020. The prices of more than 500 commodities in the market basket of the Consumer Price Index (CPI) of the National Capital Region are being regularly monitored. Movements of prices are currently being observed and compared with offline prices.

The **General Statistics Office of Viet Nam (GSO)** piloted the integration of data scraped from online marketplaces in the production of the CPI.[19] The GSO collaborated with the E-commerce Department and Ministry of Industry and Trade to identify the websites for price collection. GSO managed to collect daily data for more than 400 consumer goods, including food and food services, garments, hats and shoes, household goods and services, from 43 websites. In 2019, GSO conducted a conference on using web scraping for the CPI. However, the results are not published, and the data source is not yet integrated into the statistical business process. Among the challenges impeding integration of online price data are a legal framework limiting data sources that can be used to produce official statistics and no policy guiding access to non-traditional data sources, as well as the need for big data storage infrastructure, qualified human resources and high initial investment.

**Statistics Indonesia (BPS)** conducted several experiments in producing CPI from data scraped from e-commerce platforms. Guided by the Indonesian E-Commerce Association (IDEA) 2016 e-Commerce Award, the BPS team selected 14 e-commerce platforms. Out of 167 commodities explored, 163 could be calculated for the CPI. The lack of information on the quantity of commodities sold constituted a limitation in calculating the index. Another limitation is that the websites keep changing over time, impeding automated data collection. Therefore, BPS only focuses on several main online shops and marketplaces ensuring steady data flow. However, the use of online price data remains at the proof of concept (PoC) phase and has not been integrated with the production of the official statistics yet.

BPS is currently also investigating other big data sources for price statistics, such as crowdsourcing, and is planning to seek data collaboration with Internet companies. The 2019 Government Regulation on Trade through Electronic Systems in Indonesia obliges domestic and foreign operators trading through the Electronic Systems in Indonesia to regularly submit their data to government agencies for statistical purposes and BPS could access data on all digital transactions.

## 1.3 Compiling CPI using other data sources

Whereas scanner data and online price data are the mostly used big data sources for the production of the CPI, some NSOs in the region are obtaining private sector data in other ways, either through a different model of negotiation with the retailers, as in the case of the State Statistical Committee of the Republic of Azerbaijan, or through collaboration with the Tax Authority, as in the case of the NSO of Mongolia.

**The State Statistical Committee of the Republic of Azerbaijan** uses data from the electronic databases of trade networks for statistical purposes. In 2018, the State Statistical Committee (SSC) concluded an agreement with 10 trade networks with 264 supermarkets, accounting for 20% of retail trade turnover in Baku. The trade networks submit to the SSC the requested information (name of product, barcode, unit of measurement, retail price, trademark,

[18] Department of Statistics Malaysia, STATSBDA, 28-31 August 2017, https://www.unescap.org/sites/default/files/Session5.2.2_Malaysia_StatsBDA_PSS_RSG_28-31Aug2017.pdf

[19] Thuy Nguyen Van, Hoan Nguyen Cong, The practical experiences of collecting big data to compiling consumer price indices of the General Statistics Office of Viet Nam (GSO), 2018 https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Viet_Nam_ppt.pdf and Thuy Nguyen Van, Hoan Nguyen Cong, presentation “The Practical Experiences of Collecting Big Data to Compiling Consumer Price Indices of the General Statistics Office of Viet Nam (GSO)”, UNECE, 2019 https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2018/Viet_Nam_ppt.pdf?fbclid=IwAR38F_JpNEkwXDqIprK7IP1Blnu3rGzFQ9ew2KhPJt_D7NTbj4gREJ643lk

specification and producing country) in the electronic form every10 days. The SSC developed a special software for the collection of this information and its integration with its own database. The difference between the trade network database data and scanner data is that the former does not necessarily contain information on the quantity of products sold. From mid-2019, the method was extended to 10 other cities and regions in the country.[20]

**The National Statistics Office of Mongolia (NSO)** made great efforts in linking its statistical business register with statistical reports of the economic and social sectors, national accounts, and other national estimations and surveys to the annual, quarterly and monthly taxation report and Enterprise income tax reported from General Department of Taxation and data on electronic payment system of Value-added tax from Information Technology Center for Customs, Tax and Finance. The NSO of Mongolia advocates the importance of linked databases to administration organizations.

### 1.4 Compiling Residential Property Price Index using online price data

Another area of price statistics is Residential Property Price Index (RPPI). The National Statistical Offices of Georgia and Viet Nam, and the Statistics Department of Bank Indonesia explored the use of web scraping for computing residential property price index (RPPI). Geostat and the Viet Nam GSO received technical assistance from the IMF on incorporating big data into the RPPI.

**The Viet Nam GSO** discovered web data collection as a potential source of data for RPPI, reducing collection and processing costs, while waiting for the administrative data source on transaction prices to be developed.

**Geostat** received technical guidance, including methodological support and training on using R for web scraping from the IMF to compile experimental indices from two websites for the RPPI.[21] Geostat signed an agreement with the website owners, permitting automated web scraping. It started accumulating data in 2019 and has been developing the right infrastructure since then. It considers publishing the experimental statistics in April 2021. Building on the experience obtained through web scraping for the RPPI, Geostat is also considering using web scraped data for tracking price changes of used cars and electronics and is currently designing the project.

The **Statistics Department of Bank Indonesia** also attempted to address the issue of traditional data source reliability but also accessibility, as data on declared property transactions such as administrative data from the land registry or property tax records, are difficult to acquire. Bank Indonesia collected data from online property advertisements from the two biggest property websites over the period of two and a half years between 2016-2018 and employed a direct hedonic approach to calculate robust property price indices.[22] While the timeliness of online asking (listed) price is the main benefit, Bank Indonesia judged, however, that the difference between the asking and actual transaction price may produce misleading estimates.

## 2. Compiling tourism statistics with mobile positioning data

The compilation of tourism statistics is guided by the International Recommendations for Tourism Statistics 2008 developed by the World Tourism Organization (UNWTO), United Nations Statistics Division (UNSD) and the International Labour Organization (ILO).[23] With the support of UNSD, UNWTO launched the initiative Towards a Statistical Framework for Measuring the Sustainability of Tourism (MST) to measure tourism's role in sustainable development.[24]

[20] The State Statistical Committee of the Republic of Azerbaijan, "Price statistics: application of a new approach to data collection, https://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.22/2020/vebinar_Aze.pdf

[21] IMF Country Report No. 20/172. Georgia, Technical Assistance Report—Residential Property Price Indices

[22] Arief Noor Rachman, "An alternative hedonic residential property price index for Indonesia using big data: the case of Jakarta," February 2019, https://ec.europa.eu/eurostat/cros/system/files/euronaissue2-2019-article3.pdf

[23] United Nations, International Recommendations for Tourism Statistics 2008, https://unstats.un.org/unsd/publication/Seriesm/SeriesM_83rev1e.pdf

[24] UNWTO, On measuring the sustainability of tourism: MST, https://www.unwto.org/standards/measuring-sustainability-tourism

Both the United Nations and governments are exploring big data sources for compilation of tourism statistics. For example, Eurostat developed a taxonomy of big data sources with potential for measuring tourism and provides a detailed explanation of how each data source could contribute to the production of tourism statistics.[25] Eurostat identified the following data sources as potentially relevant for tourism statistics: **communication systems** - mobile network operator data, smart mobile devices data, social media posts; **World Wide Web** - web activity, dynamic and static websites; **business process generated data** - flight booking systems, stores cashier data, financial transactions; **sensors** - traffic loops, smart energy meters, vessel radio identification, satellite images; **crowdsourcing** - Wikipedia contents, picture collections.

ESCAP Statistics Division is also exploring the opportunities and uses of big data for official statistics and has recently launched a study on COVID-19 tourism using big data sources focusing on Maldives, Bali (Indonesia) and Phuket (Thailand). An integrated analysis is being conducted using web scraped data from online platforms, such as Agoda, Google Trends and Trip Advisor, combined with satellite imagery, international flight data and data from official tourism websites of these countries. The results of the study will be available by the end of the first quarter of 2021.

In Asia and the Pacific, the NSOs are exploring mobile network operator (MNO) data for tourism statistics, with two prominent examples from BPS Indonesia and Geostat.

## 2.1 Compilation of cross-border and domestic tourism statistics using mobile phone data in Indonesia

BPS Statistics Indonesia collaborates with the Indonesian Ministry of Tourism, Positium – an Estonian company, and Telkomsel – the biggest telecom operator in the country with more than 170 subscribers, to gain access to and develop an effective methodology for using mobile positioning data (MPD) to produce official tourism statistics. BPS uses MPD for measuring cross-border[26] and domestic tourism, and it used MPD to measure visitors at two events in 2018: the ASIAN Games and the IMF Meeting. [27]

BPS started using MPD in 2016 to accurately capture and increase the coverage of international visitor arrivals in cross-border areas.[28] Until the start of the project, Telkomsel had not been saving and storing signaling data. Only after entering in partnership with BPS and the Ministry of Tourism, it realized the value of using these data and started storing them over longer periods of time. In 2017, the Ministry of Tourism and BPS started collaborating with Positium, and in 2018, it began upgrading tourism statistics according to the new methodology handbook and quality assurance framework developed in collaboration with the company. Positium developed a methodology and an automated system for processing MPD to compile

[25] Eurostat, Tourism statistics: Early adopters of big data? 2017 Edition https://ec.europa.eu/eurostat/documents/3888793/8234206/KS-TC-17-004-EN-N.pdf/a691f7db-d0c8-4832-ae01-4c3e38067c54 page 9

[26] BPS Statistics Indonesia, Rifa Rufiadi's presentation "Mobile Positioning Data Make sub Region Tourism Visible," https://unstats.un.org/unsd/bigdata/conferences/2019/presentations/conference/day2/session1/Rifa%20-%20Indonesia%20-%20MPD%20make%20sub%20region%20tourism%20visible.pdf https://bit.ly/3l1Seno and Titi Kanti Lestari, Siim Esko, Srapono, Erki Saluveer, Rifa Rufiadi, "Indonesia's Experience of using Signaling Mobile Positioning Data for Official Tourism Statistics," http://www.15th-tourism-stats-forum.com/pdf/Papers/S3/3_2_Indonesia's_Experience_of_using_Signaling_MPD_for_Official_Tourism_Statistics.pdf

[27] Statistics Indonesia, "The Use of Mobile Positioning Data for Official Statistics: Indonesia's Examples," https://unstats.un.org/bigdata/events/2019/tbilisi/presentations/Session%200/The%20Use%20of%20MPD_Human%20Mobility%20Project%201.0.pdf . Additional information on Indonesia's experience can be found at: Statistics Indonesia, Winida Albertha's presentation "Indonesia's Experience on International Migration Statistics," United Nations Expert Group Meeting on Improving Migration Data, New York, USA, 1-3 July 2019, https://unstats.un.org/unsd/demographic-social/meetings/2019/newyork-egm-migration/4%20Indonesia.pdf and Statistics Indonesia, presentation "The Use of Mobile Positioning Data, Indonesia's Experiences," International Symposium on the Use of Big Data for Official Statistics, Hangzhou, China, 16-18 October 2019, https://unstats.un.org/bigdata/events/2019/hangzhou/presentations/day2/6.%20The%20Use%20of%20Mobile%20Positioning%20Data%20Indonesia%E2%80%99s%20Experiences.pdf

[28] Titi Kanti Lestari, Siim Esko, "Lessons for Effective Public-Private Partnerships (PPPs) from the Use of Mobile Phone Data on Indonesian Tourism Statistics," Asia-Pacific Economic Statistics Week 2019, https://communities.unescap.org/system/files/8._lessons_for_effective_public-private_partnerships_from_the_use_of_mobile_phone_data_in_indonesian_tourism_statistics.pdf

cross-border tourism statistics within the premises of Telkomsel.

The next step was generation of commuting and domestic tourism statistics with Telkomsel. At that stage, Positium only consulted BPS and Telkomsel, as the two actors gained sufficient experience and expertise to carry the work forward on their own. Nevertheless, BPS signed a long-term Memorandum of Understanding (MoU) with Positium in 2019 to collaborate on further research and capacity building.

The method currently used for data processing is called Sandboxing. While MPD are stored and processed in the mobile network operator's system, and data are not transferred into the BPS database, BPS can access the database using a token from the MNO. However, access is limited to a 5% data sample for the purpose of data exploration and algorithm development. In this setting, BPS provides the definition and the flow concept used to define commuters and tourists, whereas the MNO builds a script to implement it on the data and then shares the tabulation of the results with BPS. Furthermore, the Ministry of Planning is negotiating with the other MNOs access to their data to complete the data source for MPD.

In the future, BPS plans to extend the use of MPD in other statistics, such a poverty, unemployment, socio-economic activities and others.

### 2.2 Producing domestic tourism statistics in Georgia

Like BPS Indonesia, **Geostat** uses mobile positioning data to produce Georgian Tourism Statistics. However, this is part of a larger project, launched in 2018 in partnership with UNSD, ITU, Georgian National Communication Commission (GNCC) – the telecommunication regulator, and Positium. It focuses on tourism statistics, migration statistics, and information society statistics. Previously, ITU conducted a pilot project on measuring information society (including an indicator on inbound roaming subscriptions per foreign tourists) in Georgia, in partnership with the mobile network operators.[29]

Unlike BPS Indonesia and several other NSOs that partnered directly with the MNOs to access MPD data or MPD-derived insights, Geostat partnered with the telecommunications regulator - GNCC. GNCC then negotiated access to data from three mobile network operators (MNOs), specifying the format and time for the requested data. The national legislation facilitates GNCC's free access to anonymized mobile phone data. Furthermore, the Personal Data Protection Office plays an important role in this collaborative by providing guidance on data anonymization, storage, and access. While Geostat does not have access to the primary data, it collaborates with GNCC on the development of the methodology that is applied on the anonymized data that GNCC collects from the MNOs. Several divisions in Geostat used MDP to produce tourism and migration statistics.

Geostat supplements the ongoing surveys with insights from MPD. MPD allows for validation of the sampling frame for the domestic tourism survey and provides detailed information on sub-national levels of tourism and migration.[30]

## 3. Estimating poverty with Earth Observation (EO) data

The **Philippines Statistics Authority** (PSA) and **Thailand National Statistical Office** in collaboration with Asian Development Bank (ADB) conducted exploratory poverty mapping through integration of Earth Observation data and using Artificial Intelligence.[31] The objective of the project was to provide updated and more granular statistics on poverty while keeping the associated costs at a manageable level.

The NSOs used publicly available Landsat 8 and Sentinel 2 imagery and the off-the-shelf convolutional neural network (CNN) algorithm ResNet34, a Deep Learning algorithm, along with the open-source

[29] ITU, Big Data for Measuring the Information Society: Country Report – Georgia, https://www.itu.int/en/ITU-D/Statistics/Documents/bigdata/Georgia.pdf

[30] Geostat, Aleksandre Ambokadze's presentation "Perspectives of using MPD in Georgia," International Symposium on the Use of Big Data for Official Statistics, Hangzhou, China, 16-18 October 2019, https://unstats.un.org/bigdata/events/2019/hangzhou/presentations/day2/7.%20Perspectives%20of%20using%20MPD%20in%20Georgia.pdf

[31] https://www.unescap.org/events/asia-pacific-stats-caf-series-using-remote-sensing-data-accelerate-global-development

analytical platforms. The NSOs used the estimated proportion of population living beyond the national poverty line that they compiled through the small area estimation (SAE) technique. The other primary data source used was the night lights data compiled by the Visible Infrared Imaging Radiometer Suite (VIIRS). The intensity levels were categorized into discrete groups using combination of Gaussian Mixed Models and heuristic methods. The intensity of nigh light was used to estimate poverty rates following an approach proposed by Stanford University.

The results showed a good performance of the algorithms, registering a slight performance difference between the Philippines and Thailand, which could be caused by different night light values in these countries. The case study showed that data granularity could be achieved through integration of alternative data sources and not necessarily through the redesign of existing data collection systems, which would incur significant costs. In addition, satellite imagery can be used to forecast poverty for the years when the benchmark estimates are not available.

Additional information about this project can be found in the ADB publication Mapping Poverty through Data Integration and Artificial Intelligence: A Special Supplement of the Key Indicators for Asia and the Pacific.[32]

## 4. Using big data to produce experimental economic statistics during the COVID-19 pandemic

The COVID-19 pandemic has stressed the need for timely and detailed data about the state of the economies to support policy response. Some NSOs accepted this call by exploring new timely data, either from administrative records or the private sector. The ABS has been the most active in exploring big data sources for the development of experimental economic statistics during the COVID-19 pandemic. Australia's experience is presented below along with examples from Stats NZ and BPS Statistics Indonesia.

The examples from the ABS and Stats NZ highlight the importance of data partnerships among various public authorities as well as with the private sector. However, while these partnerships were forged during the pandemic, requiring immediate action, and since most of these new indicators are experimental in nature, the sustainability of data supply and partnership arrangements will require formalization following the pandemic. Nevertheless, these examples highlight the potential and the timeliness of big data in providing insights into the state of national economies.

### 4.1 Estimating business sales, household consumption and GDP from bank transactions data

At the onset of the COVID-19 pandemic, the ABS negotiated with the Australian Banking Association (ABA) access to bank transactions data – aggregated, de-identified transactions data from major banks, to inform official ABS estimates of business sales, household consumption and Gross Domestic Product (GDP), and to assist in understanding the evolution of the Australian economy. The ABS obtained a sample of inflows and outflows of household and business accounts, with significant limitations, such as exclusion of international transactions. The ABS could classify transactions by area and expenditure class, identifying differences in expenditure patterns across different geographical areas. The ABS is in the process of negotiating provision agreements with the banks, as data provision and the underlying conditions following the crisis remain uncertain.

### 4.2 Using smart meter electricity data to understand COVID-19 impact and to complement economic statistics in Australia.

The ABS conducted an exploratory analysis of electricity usage patterns in Melbourne households and businesses for January to April 2020, compared to the same period in 2019, to understand COVID-19 impact.[33] It analyzed only anonymized data from two energy companies. The address and other identifiers relating to the meter servicing the premise were

[32] Asian Development Bank, Mapping Poverty through Data Integration and Artificial Intelligence, September 2020, http://dx.doi.org/10.22617/FLS200215-3

[33] Australian Bureau of Statistics, "Using electricity data to understand COVID-19 impacts, 2020," October 10, 2020, https://www.abs.gov.au/ausstats/abs@.nsf/Latestproducts/4661.0Main%20Features12020?opendocument&tabname=Summary&prodno=4661.0&issue=2020&num=&view=

removed and replaced with an anonymized identifier, preventing the ABS from identifying the exact location of any meter in the dataset.

The results of the analysis showed that electricity usage patterns were indicative of a trend of a higher proportion of households staying home during the crisis. While the business electricity usage fell, the fall was more substantial for particular industry groups.

The advantages of using electricity consumption data are the coverage and ability to show timely changes. Analysis of electricity consumption data can complement other household and economic statistics in supporting evidence-based policy making in the current uncertain environment.

However, there were several challenges in using and interpreting data generated by smart meters. These include the different ways that changes in business conditions and behavior can be reflected in the data, variety both between and within the industries, the impact of weather on energy consumption, the difference between electricity purchasing and its usage, the absence of data on self-generated electricity form solar panels, and data quality related to business locations, in cases where several businesses in different industries share the same meter.

Also, the data in this pilot project only covered inner Melbourne and the results cannot be generalized Australia-wide, since unlike Victoria, other States and Territories do not have a universal smart meter penetration, which may present a methodological challenge if the ABS acquires data from beyond Victoria in the future.

As a next step, the ABS plans to compare smart meter data to survey response data to examine whether electricity usage trends can be leveraged to improve the robustness of economic statistics. The ABS also plans to use weather data to help account for these effects, and to explore how the data can be used to inform the treatment of electricity usage in the ABS Environmental Accounts and its distribution across industries and households.

### 4.3 Estimating Australian jobs and wages using payroll data

The ABS has also tapped into the administrative big data from the Australian Taxation Office (ATO)'s Single Touch Payroll (STP) system to publish fortnightly information about Australian jobs and wages.[34] As the crisis required fast and reliable information on the developments in the labour market, the ABS developed with the Tax Office a quick index, which is being further improved. This dataset provides new high frequency information on changes in total jobs and total wages paid for all employing businesses that report to the ATO through the Single Touch Payroll system.[35] Access to this data source was accelerated during the pandemic. However, following the crisis, the procedures guiding data exchange will require formalization.

### 4.4 Near-real-time indicators from COVID-19 – New Zealand Activity Index (NZAC)

To address the need for timely data during the COVID-19 pandemic, particularly related to the economic sector, Stats NZ collaborated with several government authorities to develop near-real-time economic statistics.

Stats NZ collaborated with the Treasury and the Reserve Bank to launch the New Zealand Activity Index (NZAC) in June 2020. The NZAC summarizes several monthly indicators of economic activity, including consumer spending, unemployment, job vacancies, traffic volumes, electricity grid demand, business outlook, and manufacturing activity and is intended to be interpreted as a broad measure of economic activity. The Treasury releases the NZAC monthly in the Treasury's Weekly Economic Update.[36]

[34] Australian Bureau of Statistics, "ABD exploring new data sources to inform official statistics in response to COVID-19," 22 May, 2020, https://www.abs.gov.au/websitedbs/d3310114.nsf/home/ABS%20Media%20Statements%20-%20ABS%20exploring%20new%20data%20sources%20to%20inform%20official%20statistics%20in%20response%20to%20COVID-19

[35] Australian Bureau of Statistics, "Measuring the impacts of COVID-19, Mar-May 2020," 11 June, 2020, https://www.abs.gov.au/articles/measuring-impacts-covid-19-mar-may-2020

[36] The Treasury, New Zealand Activity Index, https://www.treasury.govt.nz/publications/research-and-commentary/new-zealand-activity-index

The NZAC and its component indicators are released in the COVID-19 Data Portal.

Also, during the pandemic, Stats NZ launched the experimental COVID-19 data portal, sourcing data from Stats NZ and a number of other government agencies and private sector organizations. The portal reports on economic, health, income support, and social aspects of COVID-19's impact on New Zealand and its recovery.[37]

### 4.5 Estimating human mobility and employment indicators using online data in Indonesia

BPS Statistics Indonesia used several online data sources to track changes in population mobility and employment during the COVID-19 pandemic, such as Google Mobility Index, flight tracker and an online job portal.[38]

Changes in population mobility were tracked through the Google Mobility Index. Google mobility data provided insights into new mobility patterns at work and at home, as well as across retail and recreation, grocery and pharmacy, parks and transits, along with changes compared to pre-pandemic level.

The air transport activity was tracked through the flight tracker, registering a substantial drop in the air transport activity in its 5 busiest airports during the pandemic.

The employment situation was analyzed using online job vacancies advertisements on jobs.id portal. Online data showed a decline in online job vacancy advertisements across all sectors during the period January-May 2020.

## 5. Sentiment analysis around economic statistics

Several NSOs in Asia and the Pacific, such as New Zealand, Republic of Korea and Viet Nam, use online data to monitor population's sentiment towards economic issues.

**The Reserve Bank of NZ** started developing indicators measuring economic news sentiment in February 2020 in response to the COVID-19 pandemic. It used data from GDELT (Global Database of Events, Language, and Tones), which contains article-level information on sentiment, themes, locations and metadata, and applied natural language processing (NPL) and machine learning (ML) algorithms to produce the economic news sentiment. The weekly indicators are published on the national COVID-19.

Similar to the Reserve Bank of NZ, **KOSTAT** also produces a "consumer sentiment index" on economic activities based on online news data. However, KOSTAT is experiencing limitations in its ability to introduce latest technology, such as NPL, in a timely manner, due to the growing computing infrastructure requirements.

The **Viet Nam GSO** performed sentiment analysis on public opinion about the planned GDP revision for the period 2010-2017 following the results of the 2017 economic census and data from the General Department of Taxation.[39] GSO used Artificial Intelligence, such as ML and Deep Learning, for the analysis of content collected from the website, social media posts on personal, fan, and group pages, as well as forums, magazines' websites, and Youtube. The period covered in the analysis is the date of GSO's first mentioning of the GDP revision (16 August, 2019) to one month following the publication of the revision results (31 January 2020). This social listening tool informed the GSO about public's reactions to its decision, offering recommendations on its future transparent information sharing with the public.

[37] Statistics New Zealand, COVID-19 data portal, https://www.stats.govt.nz/experimental/covid-19-data-portal

[38] BPS Statistics Indonesia, Ali Said's presentation "Current Progress of Big Data Utilization for Official Statistics in Indonesia", ESCAP Stats Café series, 17 August 2020, https://www.unescap.org/sites/default/files/Current_Progress_of_Big_Data_Utilization_for_Official_Statistics_in_Indonesia_11th_Stats_Cafe_17Aug2020.pdf

[39] Nguyen the Hung, "Listening the public opinion? An approach from big data with the case of revision GDP in the period 2010-2017 in Vietnam," 2020 Asia-Pacific Statistics Week, https://www.unescap.org/sites/default/files/APS2020/35_Listening_the_public_opinion-big_data_revision_GDP_2010-2017_GSO_Viet_Nam.pdf

## Challenges and Recommendations

Big data remains an exploration arena for most national statistical offices and they are confronted with a multitude of challenges related to using them including entry barriers to specific data sources and regulatory and practical challenges.

As the experiences of the NSOs in the region of using big data to produce economic statistics vary, so do the types of challenges that they confront. The NSOs that have not started the exploration of big data for economic or other official statistics often cite a restrictive legal framework, privacy concerns and lack of relevant skills, technical infrastructure, and financial resources for big data exploration as impediments to the use of new and non-traditional data sources for official statistics. While the NSOs that are actively researching the potential of big data may also confront similar challenges across legal, technical, technological and financial areas, they also face data source-specific challenges, such as ensuring data privacy, sustainability of data access, and data quality and representatives, among others.

Based on the experiences of the NSOs in the region, including the feedback received from the NSOs with no experience in the use of big data, a common set of major challenges is drawn and presented below. The order of their presentation does not reflect the level of their importance.

### 1. Legal and regulatory framework

The national legal frameworks of countries in the Asia-Pacific region often pose challenges to the use of big data in official statistics. The range of regulatory issues is broad. In some countries, the Law on Statistics limits the data sources that can be used in the production of statistics. Therefore, the integration of big data would require the revision of the Law on Statistics. In most countries, the regulatory framework does not address conditions of data access from the private sector, including the cost of access, and sustainability of access to private sector data. The experience of some countries shows that data access is regulated at the level of individual data sources (e.g. mobile phone data in Georgia or online transaction data in Indonesia), rather than for “big data” more generally.

### 2. Data privacy and citizen perception

Personal data protection poses many challenges to the use of big data for official statistics. Privacy issues usually arise at the level of data access. In most cases, NSOs have access to anonymized or aggregated data, as data providers follow strict internal privacy protection protocols. But in addition to ensuring data privacy at access and analysis levels, the NSOs are also confronted with public’s perception about the use of personal data. This obliges the NSOs to address the public concerns through awareness raising and transparent communication on how personal data are used and how privacy and anonymity are ensured throughout the statistical production process.

### 3. Data access, sustainability and partnerships

The experience of the NSOs in the region shows different models of accessing private sector data, depending on the data source and the country. While some NSOs purchase samples of aggregated data covering a short period of time and a specific geographic area to test the relevance of the data source or they just outsource both data collection and analysis to third parties, others negotiate access to data directly with data providers or national regulatory bodies, building partnerships and co-designing methodologies. In some cases, NSOs obtain access to anonymized and aggregated data that they analyze internally, while in others they only access the insights that were produced from applying their methodologies and algorithms to dataholders’ raw data.

As in multiple instances access to private sector implies costs, the sustainability of private sector data remains under question in the context of tightening budgets and in the absence of legislation regulating financial modalities of external data access. Also, as several governments and statistical organizations have negotiated access to big data during the pandemic to produce economic and mobility-related statistics, uncertainties related to data access and partnerships models remain to be addressed post-COVID-19.

## 4. Capacity and skills

Using big data requires a broad range of additional skills to traditional data sources used for statistics. The necessary skills vary by data sources. NSOs that are experimenting with big data are developing in-house capacity for a greater autonomy, even if they were previously outsourcing some of the tasks. As big data is rapidly evolving, it also requires keeping up to date with new methods and methodologies. Nevertheless, the amount and availability of publicly available trainings and resources is also growing. The Statistical Institute for Asia and the Pacific (SIAP) and the UN Committee of Experts on Big Data and Data Science for Official Statistics are addressing this need by developing capacity building trainings.

## 5. Technological infrastructure

Integration of big data into official statistics requires a technological infrastructure capable of storing and analyzing large volumes of data and ensuring data privacy and security. Such infrastructure requires considerable upfront investments that many NSOs cannot afford, even if in the long term such infrastructure coupled with new methods of statistical production may result in financial savings and other benefits. In addition, as big data technologies and data volumes are evolving at a growing speed, some NSOs are confronted with growing computer infrastructure requirements.

## 6. Data source-specific challenges

Individual big data sources also present specific challenges, with some presented below:

***Online price data.*** Online price data (web scraped) from the e-commerce platforms may not cover the full list of goods or services that the NSOs rely on in the compilation of the CPI. Technical issues include frequent changes in the website structure, the need to update crawlers or develop separate crawlers for different websites, the possibility of automatic blockage of high frequency web scraping, which calls for collaboration and partnership with web site owners. Another important challenge is the quality of data. For example, KOSTAT did not reflect web scraped data in the official statistics due to quality issues and concerns.

***Scanner data.*** Access to scanner data requires partnerships with retailers or with third party data providers. In some countries, like in the Russian Federation which has over 100 retail networks, reaching an individual agreement with most of the retailers may not be feasible, while in others, the retailers are less open for negotiations. However, the experience of Geostat provides a successful example of collaboration with retailers, who in addition to data sharing, align the data with the COICOP classification.

***Mobile phone data.*** A few countries in the region managed to obtain access to the MNO data to produce tourism and/or migration statistics. Due to data privacy concerns, access to this data source has seen multiple challenges across the countries in the region. Differences in legal frameworks result in different partnerships and data access models across the countries that managed to negotiate access. Some statistical organizations negotiate and access data directly from the MNOs, such as the case of Statistics Indonesia and KOSTAT, while others collaborate with the national telecommunications regulator or the ITC ministry that regulate data access and underlying conditions, as in the case of Georgia.

Building on the successful experiences and lessons learned from some of the NSOs in the region, several recommendations could be provided to the NSOs that have not yet embarked on the big data journey but are willing to do so. These recommendations are not limited to economic statistics and can be considered in other areas of statistics.

While some may experience challenges with a restrictive legal framework in the production of statistics, they can still identify priority areas for the use of big data in economic statistics and pilot small projects, based on which, they could build the case for regulatory changes.

Building trusted partnerships with the private sector for data access and experimentation is another important issue to be considered. However, this effort may take time, depending on country, data source and data provider. Developing partnerships with academia can also be helpful, especially when the NSOs have limited internal capacity for big data processing and analysis.

Also, being transparent about the way the NSO uses personal data and how it ensures privacy is crucial for building trust among the public.

As multiple NSOs in the region and elsewhere are experimenting with big data and more information on their experiences in available, newcomers can rely on that wealth of information as a starting point for their big data experimentation. In this regard, ESCAP Statistics Division has been actively supporting regional knowledge sharing on the use of big data for official statistics through dedicated efforts. Several Stats Cafés have addressed the uses of big data in official statistics, with one session focused on the production of economic indicators using big data. Following one of the Stats Cafes, the NSOs of Iran, Bhutan and India expressed interest in learning about the use of smart meter data and ESCAP facilitated an informal knowledge sharing with Statistics Denmark on that matter. ESCAP has also launched the Data Integration Community of Practice (DI CoP), which explores the integration of alternative data sources, both administrative and private sector data, into official statistics. Among other initiatives is the compilation of examples from countries in the region of the uses of big data for environment, economic and social statistics, which this Stats Brief is part of.

## Conclusion

There is a multitude of big data projects in the area of economic statistics undertaken by the NSOs across Asia-Pacific. However, most of the projects remain experimental in nature. The main exception is price statistics where several NSOs have already integrated scanner data and/or online data in the production of the CPI for certain products. In some areas, scanner data or online price data can replace fully or partially the traditional survey data collection, in most other cases, big data sources are used to complement conventional data, by adding timeliness, granularity, or new insights.

Based on the NSO examples provided in this brief and the interest expressed by other NSOs in the region, scanner data, online price data and mobile phone data are the most promising big data sources for economic statistics. Therefore, regional efforts should concentrate on capacity building, knowledge sharing and partnership development with the private sector to support the integration of big data into economic statistics.

For more information regarding ESCAP's work in statistics development please visit: http://www.unescap.org/our-work/statistics

Previous issues of Stats Brief: http://www.unescap.org/resource-series/stats-brief

Contact us: stat.unescap@un.org

[i] Additional inputs to this Stats Brief have been received from Anders Holmberg, Chief Methodologist, the Australian Bureau of Statistics; Ric Clarke, Director of Emerging Data and Methods, the Australian Bureau of Statistics; Revaz Maisuradze, Chief Specialist, Department of Planning, Coordination and Communication, National Statistics Office of Georgia; Toghrul Ajalov, Deputy head, Department of Statistical Works Coordination and Strategic Planning, State Statistical Committee of the Republic of Azerbaijan; Gary Dunnet, Deputy Chief Methodologist, Statistics New Zealand; Richard Evans, General Manager, Economic and Environment Insights, Statistics New Zealand; Chris Ball, Reserve Bank of New Zealand; Ali Said, Director, Statistical Analysis and Development, BPS Statistics Indonesia; Setia Pramana, Big Data Specialist, BPS Statistics Indonesia; Jamaliah Jaafar, Director, Integration and Data Management, DOSM Malaysia; Lkhagva Myagmarsuren, Director, Information Technology Department, National Statistics Office of Mongolia; Nguyen Tri Duy, Deputy Director, Institute of Statistics of Viet Nam, Le Thuy Tuen, Principle Statistician of Statistical Methodology and Quality Management Department, General Statistics Office of Vietnam, Nguyen The Hung, Principle Statistician of Statistical Data Collection and Information Technology Application Department, General Statistics Office of Vietnam, Le Manh Quan, Statistical of Price Statistics Department, General Statistics Office of Vietnam; Benjie Navarro, Director, International Cooperation Unit, the Philippines Statistics Authority. This brief also benefited from presentations and discussions at Asia-Pacific Stats Cafés on 16 November and 30 November 2020.