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

STATS BRIEF

April 2021 | Issue No. 29

STATISTICS DIVISION

Big data for population and social statistics

This Stats Brief gives an overview of big data sources that can be used to produce population and social statistics and presents country examples in the use of mobile phone data, sales data, Earth Observation data, social media and other online data to produce mobility-related statistics, labour statistics, health statistics and SDG indicators.

This 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, Gemma Van Halderen, Director, ESCAP Statistics Division, Tanja Sejersen, Statistician, UNESCAP Statistics Division, and, with valuable inputs from Petra Nahmias, Chief of Population and Social Statistics Section, ESCAP Statistics Division and Afsaneh Yazdani, Statistician, UNESCAP Statistics Division[i]

What are population and social statistics?

Population and social statistics underpin evidence-based policy decisions that have an enormous impact on our well-being and quality of life. From health to education, from employment to crime, population and social statistics encompass every facet of our lives. And as new technology exponentially increases the amount of data available and our ability to analyze and disseminate them, it is likely that National Statistics Systems (NSS) will be required to produce information about a wider and wider range of social issues at greater levels of detail.

Unlike economic and environmental statistics, there is no globally agreed framework of the themes or topics which incorporate population and social statistics. Population statistics are widely referred to as demographic statistics and include vital statistics such as those compiled from Civil Registration and Vital Statistics systems. Four population and social statistic themes are identified on the UN Statistics Division homepage - gender statistics, time use statistics, migration statistics and disability statistics. Regionally, the ESCAP Committee on Statistics have agreed on a basic range of population and social statistics comprising 11 domains: population; health; income,

wealth and expenditure; employment; education and training; housing and infrastructure; information and communication; crime and justice; family and community; culture and leisure; and governance.[1]

What big data sources can support the production of population and social statistics?

Traditionally, population statistics are compiled through civil registration and vital statistics (CRVS) systems, censuses and other surveys of individuals and households. In most countries, a population and housing census is conducted every 10 years, in others – at irregular intervals of time. The census is also used to produce social statistics, but household surveys are usually a key source of data for these statistics in many countries. However, there is usually a lag, often substantial, in releasing data and statistics from census and surveys whereas policy making usually requires timely evidence. Furthermore, crises, such as the latest COVID-19 pandemic, often require real-time data, particularly on the location, density and movements of a population, while preserving personal data privacy and protection.

---

1   UN ESCAP, Core set of population and social statistics for Asia and the Pacific, E/ESCAP/CST(3)/5/Add.1, https://undocs.org/en/E/ESCAP/CST(3)/5/Add.1.

National Statistical Offices (NSO) increasingly rely on the use of data from population registers and other administrative sources in the production of demographic statistics.[2] On their own or when linked with other relevant administrative data, population registers population registers can also be useful as a supplement or replacement to a traditional Census to varying extents, can replace or complement surveys or be used as a sampling frame.

Combining administrative data with non-traditional sources could provide timelier and more granular data and even generate new insights. Many NSOs in Asia and the Pacific region are increasingly exploring non-traditional data sources for additional details to survey and administrative data or to generate social and demographic indicators in cases where important data are missing. The rapid advancement of modern technologies and their wide adoption by the public generate digital footprints, which, when repurposed in privacy-protecting and ethical ways, could provide valuable, timely and granular insights on populations and their movements.

For example, **mobile phone data** are used in the production of mobility-related statistics, such as migration, commuting, urbanization and tourism. One can distinguish between high resolution data, such as signaling data, and call detail records (CDR). *Signaling data* provide data on mobile device position updates every time the cell tower changes, and 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 mobile network operator (MNO).[3]

The UN 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***[3] that provides examples of mobile phone data uses for population statistics including migration and commuting statistics, and employment monitoring and information on partnership models with telecom operators.[4]

As an alternative to the MNO's data, aggregated **mobile app geolocation data** could be used to track population mobility. The user's consent to share geolocation makes these data more appealing to use for estimating human mobility in cases where user consent poses a significant challenge.

Other examples of big data used for population and social statistics, include **online job vacancies** to provide insights into employment, labour mobility and market demand for labour, **Earth Observation data** to provide insights into urbanization dynamics, **social media data** to provide insights into population density, mobility and health issues, and **other internet data** detailing users' demographics which can complement and refine official statistics to support and inform public policies across a wide range of issues.

The COVID-19 pandemic stressed the need for timely data about population movements to understand mobility patterns and to evaluate public measures to contain the spread of the virus. This gap encouraged governments and the NSOs to explore private sector data, mostly from MNOs and internet companies, such as Facebook's Data for Good initiative[4] (Disease Prevention Maps, Population Density Maps, Movement Range Maps, Survey on Gender Equality at Home) or Google's Community Mobility Index[5], and online job portals.

---

2   UNESCAP Stats Brief, October 2020 (Issue no. 26): Population Registers: A Key Resource for Producing Vital Statistics, https://www.unescap.org/resources/stats-brief-october-2020-issue-no-26-population-registers-key-resource-producing-vital

3   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

4   Facebook, Data for Good, https://dataforgood.fb.com/

5   Google, COVID-19 Community Mobility Reports, https://www.google.com/covid19/mobility/

How do the National Statistical Systems in Asia-Pacific use big data for producing population statistics?

Much of the big data-related research and experimentation for population and social statistics have been conducted and documented by academia, civil society and international organizations, including in countries in Asia-Pacific. However, there are far fewer documented examples of experimentation and use of big data for official population and social statistics from NSOs.

This Stats Brief addresses this gap and provides specific examples of NSO-led big data projects. Case studies from eight countries (Australia, Indonesia, Georgia, Russian Federation, New Zealand, Republic of Korea, China, and Bangladesh) highlight the experience of NSOs in using big data to produce population and social statistics. Examples are provided across four areas of statistics: mobility-related statistics, including commuting, internal migration and urbanization and mobility patterns during COVID-19; labour statistics; health statistics and SDG indicators. In these case studies, multiple big data sources are explored, such as mobile phone data, mobile app geolocation data, sales data, online data from search engines, social media, job portals and Earth Observation data.

These case studies are complemented with insights into other regional and global efforts, discussion of various data partnership models with telecom operators, and a presentation of the main challenges impeding the exploration and integration of big data sources into the production of population and social statistics.

Insights in this Stats Brief have been obtained through desk research and consultations with representatives of the NSOs.

1. Mobility-related statistics

Mobile phone activities, geolocation-enabled mobile applications, electronic transactions and other electronic or online activities usually contain a time stamp of the activity/transaction and the location of the user or the activity/transaction. Thus, location can be tracked over time revealing mobility patterns of individual users and entire populations. While accessing an individual's data and mobility patterns without explicit consent is considered a breach of privacy in many countries, aggregation of data at a certain privacy-preserving level may reveal mobility patterns significant enough to inform policy.

Mobility insights could inform several social and economic trends such as commuting and labour mobility, migration, urbanization, tourism, and post-disaster and pandemic movements, such as during the COVID-19 crisis.

NSOs in Asia and the Pacific, such as Australia, Indonesia, China, Georgia, Republic of Korea, Bangladesh and New Zealand have been actively exploring alternative data sources to produce mobility-related social and economic indicators, particularly data from MNOs, third party aggregators of mobile apps data geolocation and sales data. Whereas some of the NSOs leveraged the already existing data partnerships for understanding mobility patterns during the COVID-19 pandemic, others, as in the case of a2i (Access to Information)[6] in Bangladesh and the Australian Bureau of Statistics (ABS) developed new data partnerships or agreements with private companies during the pandemic for additional timely insights and response to the pandemic.

Besides the examples of mobile phone data, other data sources can help identify human mobility patterns. For example, the experience of Statistics New Zealand reveals the potential of sales data in showing mobility patterns, particularly after a disaster, and the ABS and National Bureau of Statistics of China used aggregated geolocation data from mobile apps to understand mobility patterns during the COVID-19 pandemic and urbanization trends.

The following case studies discuss individual country experiences in accessing and using big data for mobility-related statistics. This section covers the uses of mobile phone data, mobile app geolocation data, and

---

6   a2i (Access to Information) is a special programme ran from the ICT Division of the Ministry of ICT of Bangladesh and supported by UNDP that catalyzes citizen-friendly public service innovations and access to public sector information.

EO data to produce commuting, internal migration, urbanization and COVID-19 mobility-related statistics.

1.1   Commuting statistics

Commuting statistics include information about people's travel to and from work, such as distance, travel duration and modes of transport. Policy makers and planners use these statistics to make decisions about transportation infrastructure. Whereas traditionally commuting statistics are produced based on survey data, NSOs are gradually testing big data

sources. Mobile phone data and transport data are among the most commonly explored sources for commuting statistics.

● Commuting statistics using mobile positioning data in Indonesia

BPS Statistics Indonesia is among the most active NSOs in the region in exploring the potential of various big data sources for official statistics. In the area of population statistics, BPS has conducted several pilot projects using Mobile Positioning Data (MPD).

Figure 1. The use of Telkomsel's mobile positioning data to estimate human mobility indicators in Indonesia
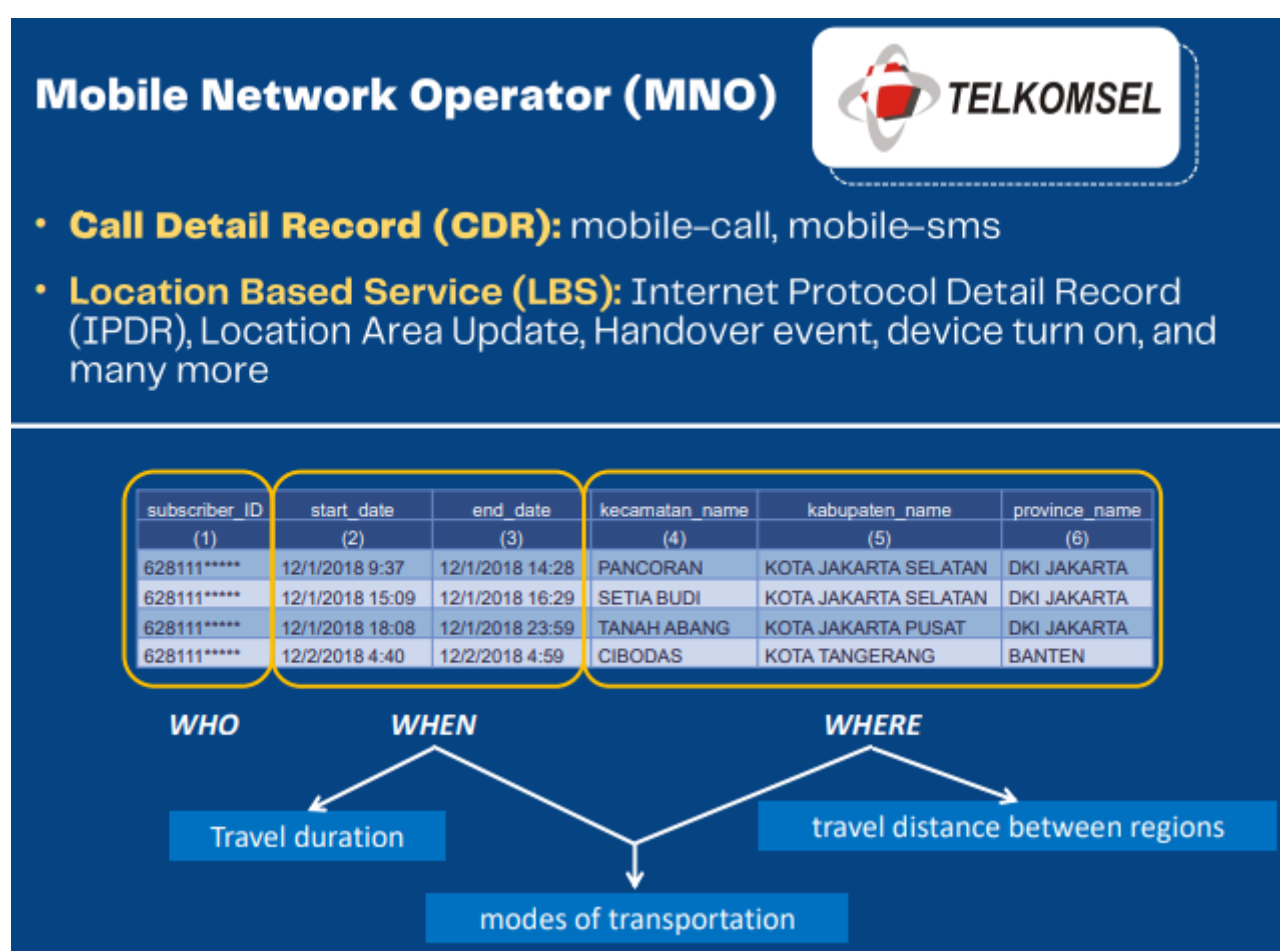
Source: BPS Statistics Indonesia

BPS uses MPD in two domains. The first is population mobility statistics, such as inbound statistics at the border, domestic tourist statistics, national tourist statistics, commuter statistics, and non-permanent mobility statistics. The second is observation point statistics, such as visitors to the 2018 Asian Games, visitors to the 2018 IMF Annual Event, statistics for the expansion of the Greater Bandung Metropolitan Area

in 2019, and visitor statistics for eight "super priority" tourist destinations that the Ministry of Tourism decided to revive in 2020 following the pandemic.

With 355 million active telephone users for a population of 268 million people in Indonesia, which is an equivalent of a 132% mobile penetration rate, mobile phone data could serve as an alternative data

source complementing and refining existing population statistics.

To explore the promising potential of mobile network data, BPS partnered with Telkomsel, the largest mobile operator in the country with more than 170 million customers, for generating mobility and migration statistics. Mobile data for each subscriber is obtained from their activity using the phone in the form of Call Detail Record (CDR) and Location Based Service (LBS), such as Internet protocol detail record (IPDR), location area update, handover event, device turn on, etc. The collected data are used to generate information about travel duration, travel distance between regions and modes of transportation, which in turn allows to distinguish between commuting, domestic tourism and migration, as depicted in Figure 1.

Data processing relies on the sandboxing method. MPD is stored and processed in Telkomsel's system, without any data transfer to BPS. However, BPS can access a 5% MPD data sample in the database using a token from the MNO for data exploration and algorithm development. Based on these data, BPS provides the definition and the flow concept defining commuters, tourists, and migrants. The MNO then builds and applies the script to the entire dataset and provides tabulation results to BPS. With the support of Positium, the methodology and algorithms have been refined over time, following the additional measures taken for quality assurance.[7] Travel diaries of volunteers confirming mobility were used to validate MPD results.

Figure 2. Mobile positioning data vs. traditional survey results in Indonesia
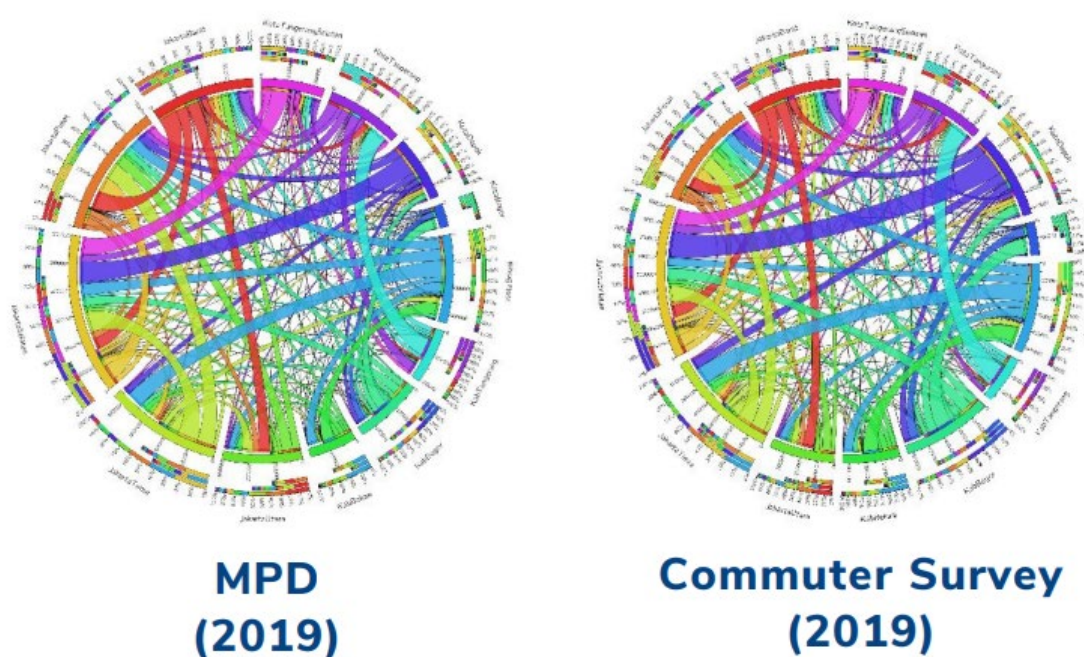
Source: BPS Statistics Indonesia

When compared to traditional survey outcomes for producing commuting statistics, MPD generated similar results, as depicted in Figure 2, but with several advantages. MPD proved to be cost effective by requiring only one eighth of the traditional survey budget and is used to increase the coverage of data as

it provides more granularity at the municipality and sub-district levels than the traditional survey. It also provides more frequent, timely and up to date data, and could be deployed monthly as opposed to once every four years and only in different metropolitan areas.[8]

---

7   Siim Esko's presentation "Before working with MPD," at Mobile Phone Data Training, 11 June 2019,
     https://unstats.un.org/bigdata/events/2019/jakarta/presentations/03%20MPD%20-%20Methodological%20framework.pdf

8   Edi Setiawan, presentation "Big data for social and demographic statistics," ESCAP Stats Café Series, November 2, 2020,
     https://www.unescap.org/sites/default/files/BPS-Big_Data_for_Social_and_Demographics_Statistics_Stats_cafe_2Nov2020.pdf

BPS also used MPD to improve sub-district level population mobility statistics [9] In 2019, BPS introduced the Metropolitan Statistical Area (MSA) as an alternative to the delineation of the Metropolitan Areas (MA) used by the Ministry of Agrarian Affairs and Spatial Planning. MPD was used for the pilot Cekungan Bandung Area to produce commuter mobility indicators between the central city and surrounding areas at the sub-district level in the Cekungan Bandung Area.[10] Among the benefits of using MPD identified in this pilot are an increased coverage and level of granularity of data at sub-district level, reduced costs vis-a-vis traditional survey, timeliness, and increased frequency.[11]

Among the main challenges BPS encountered in using MDP is building trust and collaboration with the private sector as well as data access and privacy issues. As any NSO, BPS strives for access to the 'raw' or primary data for analysis. The MNO, on the other hand, abides to corporate privacy rules and guidelines in the absence of a national law regulating big data exchange between private companies and government. In this context, BPS started building a trusted relationship with the MNO by signing a commercial contract and a Memorandum of Understanding (MoU), by closely collaborating with it on the methodology development and allowing BPS data scientists to access and process sample data in the sandbox. The sustainability of the data access remains a concern, as the contract with the MNO is renegotiated on a yearly basis and a yearly budget should be allocated to this specific data source. Nevertheless, with the support from the Ministry of Planning, BPS is exploring collaboration with other MNOs.

● Tracking transient population dynamics using mobile positioning and transport data in the Republic of Korea

KOSTAT conducted its first population movement analysis based on mobile phone data in 2017. It partnered with KT, Korean telecommunications operator, signing an MoU. The analysis was conducted on data from KOSTAT and KT on KT's Big Data Analysis Platform at one of the KT's branches. This pilot led to the establishment of institutional standards for mobile data analysis.

Since then, the regulatory framework has evolved with three new data-related laws adopted. However, clear guidelines on the integration of various data sources, including those from the private sector, remain to be established.

Following the first pilot, KOSTAT signed an MoU with SK Telecom (SKT), another Korean telecommunications operator, in May 2019. Due to high competition in the telecommunications market, KOSTAT could only partner with one MNO for the research and SKT was the partner of choice. KOSTAT has been conducting joint research with SKT under stricter privacy rules, providing guidelines for statistical analysis and performing data validation, while SKT conducts data analysis and shares the results with its partner. No other data than the research results are exchanged between the partners.

KOSTAT conducted a series of pilot analyses of transient populations based on MPD, mapping movements of workers along with the time spent commuting and time spent at work. Based on these data, KOSTAT provides a data visualization service on population movements (inflows and outflows) on 3D maps for each city, county, and district across the country on a weekday, weekend, and monthly basis. The map also shows the actual number of residents in

9   Noviyanti, Isnaeni, Prabawa, Panca D., Sari, Dwi Puspita, Koswara, Ade, Lestari, Titi Kanti, Fahyuananto, M. Hanif, Setiawan, Edi, Towards big data as official statistics: Case study of the use of mobile positioning data to delineate metropolitan areas in Indonesia, Statistical Journal of the IAOS, vol. 36, no. 4, pp. 943-954, 2020 https://content.iospress.com/articles/statistical-journal-of-the-iaos/sji200750

10   Panca D. Prabawa1, Hamim T. Soblia, Yudi F. Amin, Winida Albertha, Edi Setiawan, "The Use of Mobile Positioning Data (MPD) to Delineate Metropolitan Area in Indonesia: Case Study in Cekungan Bandung", 2020 Asia-Pacific Statistics Week, June 2020, https://www.unescap.org/sites/default/files/42_The_Use_of_Mobile_Positioning_Data_to_Delineate_Metropolitan_Area_in_Indonesia.pdf

11   Edi Setiawan, presentation "Big data for social and demographic statistics," ESCAP Stats Café Series, November 2, 2020, https://www.unescap.org/sites/default/files/BPS-Big_Data_for_Social_and_Demographics_Statistics_Stats_cafe_2Nov2020.pdf

Seoul based on MPD, as well as registered residents based on administrative statistics. The service was

expanded to include demographic analysis for each business district in Seoul.

Figure 3. Estimation of population inflow and outflow using mobile phone data in the Republic of Korea

Source: KOSTAT, http://data.kostat.go.kr
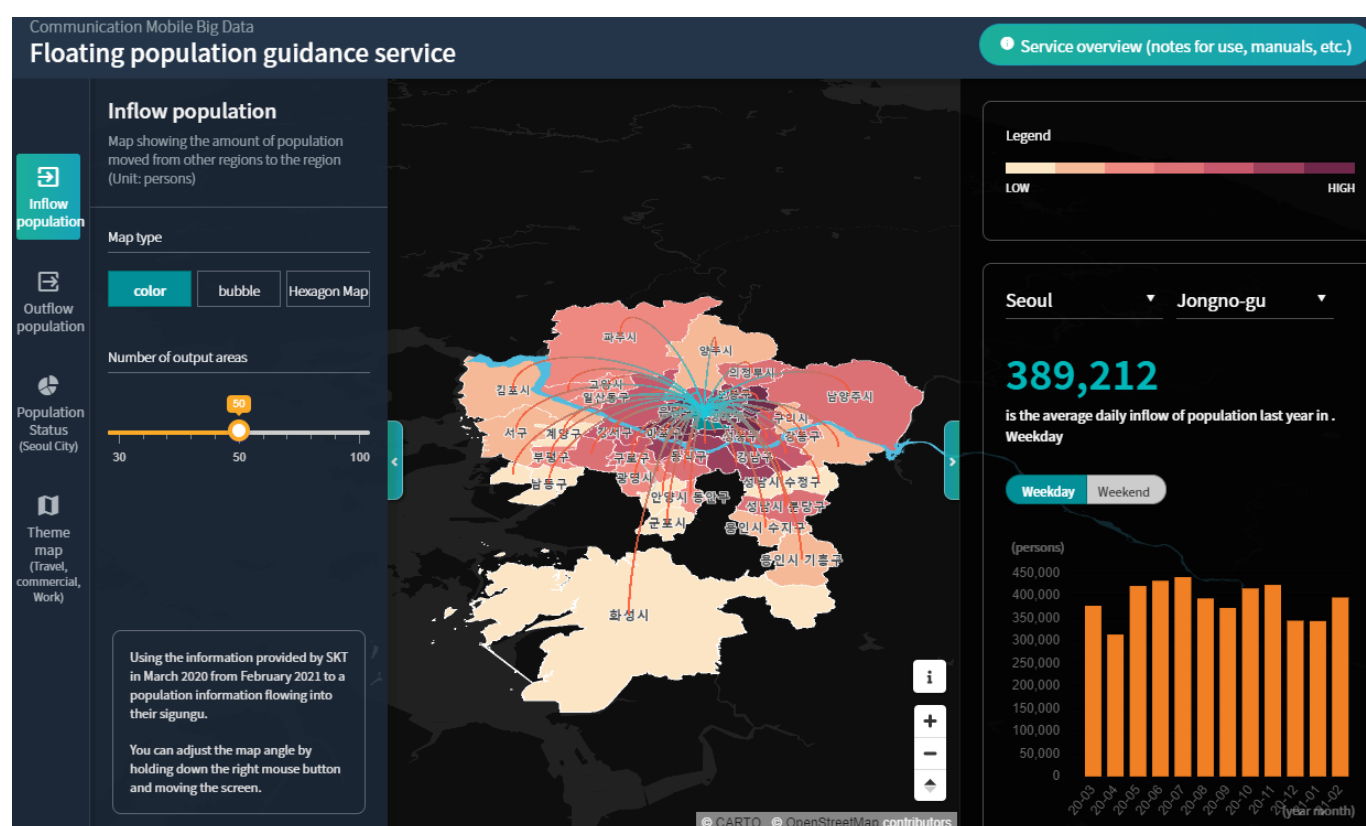
In addition to MPD, KOSTAT is collaborating with a public organization holding transport data for a project that will begin in 2021. KOSTAT will apply the same partnership model used with SKT. Similarly, KOSTAT will provide topics for analysis, specific methodology and output images and the data processing methods will be reviewed by the public and private organizations.

1.2   Internal migration and urbanization

Internal migration is defined as a person's change of usual residence within a country. It can be influenced by a range of reasons, such as search for better economic opportunities, flight from violence or political repression, natural disasters and many others. Multiple big data sources can be indicative of internal migration. For example, professional networking websites or social media networks, where users update their location, could provide information about user's relocation within the country or outside. However,

these data sources are mostly explored by international development partners. Mobile phone data, on the other hand, remains the most explored big data source for migration. Another example is sales data that can be reflective of internal migration following natural disasters.

● Estimating internal migration numbers from mobile positioning data in Georgia

Like BPS Statistics Indonesia, Geostat has been exploring the use of mobile phone data for population mobility and tourism statistics. While Geostat conducts tourism surveys, there is a lack of information on internal migration and a population register is yet to be established. The use of mobile phone data for producing migration statistics is part of a larger project, launched in 2018 in partnership with UNSD, ITU, Georgian National Communication Commission (GNCC) – the telecommunications regulator, and

Positium - an Estonian-based company, and focuses on tourism statistics, migration statistics, and information society statistics.[12]

Unlike BPS Statistics Indonesia and several other NSOs that partnered with the MNOs directly for access to data or insights, Geostat partnered with the telecommunications regulator. GNCC then negotiated access to data from three MNOs: Magticom, Geocell and Mobitel, specifying the format and time for the requested data. GNCC obtains free access to anonymized mobile phone data through legislation. Furthermore, the Personal Data Protection Office guides data anonymization, storage, and access. Geostat does not have access to the primary data. Instead, it collaborates with GNCC on the development of the methodology to be applied on the anonymized data that GNCC collects from the MNOs. While the methodology has been developed, its current testing on actual MNO data has been put on hold until GNCC resolves the IT infrastructure issue related to data storage and its full compliance with the national legislation. The project is still ongoing, but it highlights shortcomings related to technology and legal compliance that may arise during project development.

● Estimating natural disaster-related human mobility using sales data in New Zealand

After the Kaikoura earthquake in New Zealand in November 2016, Statistics New Zealand (Stats NZ) explored sales tax data to estimate human mobility.[13] From the tax data Stats NZ estimated sales increases and decreases for the period of October-November 2015 and October-November 2016. The results of the analysis provided insights into people's dispersion following the November 2016 Kaikoura earthquake, which was evident in the transaction load across different routes throughout New Zealand.

Figure 4. Sales increases and decreases for the period October-November 2015 and the same period in 2016 indicative of population displacement following the Kaikoura Earthquake

Source: Stats NZ

---

12   Paata Shavishvili, presentation "Cooperation with UNSD in the field of migration statistics", 2019, https://unstats.un.org/bigdata/events/2019/tbilisi/presentations/Session%200/Projects%20on%20Migration.pdf

13   Gary Dunnet, presentation "Improving economic statistics with big data", ESCAP Stats Café Series, November 30, 2020, https://www.unescap.org/sites/default/files/Improving_economic_statistics_with_big_data_Stats_NZ_30Nov2020.pdf

● **Estimating resident population characteristics and monitoring migration and urbanization in China**

The National Bureau of Statistics of China (NBS) is cooperating with private companies on research and exploration of the application of big data methods for population statistics. With 1.59 billion mobile users registered in China in mid-2019, mobile phone data can contribute to dynamic monitoring of population mobility.

NBS uses mobile signaling data and mobile app data from mobile phones for population migration and monitoring in some regions, but not in other domains.[14] Mobile app data are collected through a software development kit (SDK) that gathers mobile user data, such as terminal location and time information. While both data sources provide geolocation information, they register several differences related to the coverage area, location mode and positioning characteristics. For example, mobile signaling data is collected from the base stations, while mobile app data - over Wi-Fi and GPS, which results in a wide coverage with small area difference for mobile signaling data and a narrow coverage with big area differences between urban and rural for mobile app data.

Mobile signaling data provide mobile users' position update every time the cell tower changes. These data are used to monitor the residence time, location, and length of residence of mobile users in each administrative region at regular time intervals. NBS analyses the increase and decrease of mobile users compared with the base period.

NBS uses geolocated data from users' mobile apps for dynamic monitoring of population migration and urbanization. From mobile apps data, NBS estimates total population and models population projection. The number of mobile devices serves as the basis of calculation, filtering out multiple devices per person, considering the applications' market share and the elderly without smartphones, and then calculating the proportion of people and children. These data are also used to support the calculation of monthly resident population and analysis of the user flow between a certain municipality in China and various provinces and cities. Also, the distribution and mobility of urban and rural population are analyzed from geolocated data and the boundaries of urban and rural areas. With these example, NBS China uses big data for dynamic monitoring of population migration across the regions of the country as well as the urbanization trends.

1.3 Estimating mobility patterns during COVID-19

The response to the COVID-19 pandemic demanded timely and disaggregated data. Several governments and NSOs in the region, such as BPS Statistics Indonesia, KOSTAT and Stats NZ leveraged existing partnerships with the private sector, primarily with the MNOs, to analyze population mobility and predict virus propagation and hotbeds. Others, like a2i (Access to Information) in Bangladesh negotiated access to MNO data at the beginning of the pandemic. In other instances, NSOs explored other location-based data, such as the Australian Bureau of Statistics (ABS) using aggregated geolocated app data from mobile devices.

● **Understanding the before and after COVID-19 population movements in the Republic of Korea and New Zealand with the help of mobile phone data**

KOSTAT has been conducting *population movement analysis before and after COVID-19*, on a weekly basis since February 2020 based on the MNO data.[15] The bias towards SKT users (the telecommunications operator that KOSTAT partnered with on this research, and whose share of the domestic mobile communicate market is around 42%) in the analysis of population movements was corrected with the census data by applying weights by region, sex and age. The analysis was conducted on population flows by sex, age, location type, region, and weekday/weekend to identify changes in population movements caused by COVID-19.

---

14 Department of Population and Employment Statistics, NBS, presentation "Applications of Big Data in Demographic Statistics", October 2019, https://unstats.un.org/bigdata/events/2019/hangzhou/presentations/day2/5.%20Applications%20of%20Big%20Data%20in%20Demographic%20Statistics.pdf

15 http://data.kostat.go.kr

In New Zealand, **Data Ventures**, a subsidiary of **Stats NZ**, collaborates with major telecommunication companies in New Zealand and obtains aggregated and anonymized mobile phone data for analysis. In the context of COVID-19 pandemic, Data Ventures developed the *Report on COVID-19 impact on Local Councils' central business districts (CBD) Population*.[16] The report provides the changes in population mobility for eight central business districts compared across the last month and before COVID-19. It also regularly publishes the *COVID-9 Population Report*[17] and shares the script for estimating the amount of mobility within regional councils, or based on categories such as recreation, retails, and workplace areas from anonymized mobile phone data in a public repository.[18] These telecommunication data has also flowed through to International Travel statistics produced by Stats NZ, forming part of additional insight during this COVID-19 period about estimates of international visitor regional location in New Zealand. The estimated regional breakdown of international visitor numbers is published weekly as part of a regular provisional release.

- ● Internet data to estimate COVID-19 mobility in Indonesia

BPS Statistics Indonesia used several online data sources, such as Google Community Mobility Report, Apple Mobility Trend Report, flight tracker, and night-time light (NTL), to nowcast changes in population mobility and economic activity during the COVID-19 pandemic.[38]

Changes in population mobility were tracked through the Google Community Mobility Report. 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 heat map of population mobility for retail and recreation is illustrated in Figure 5. The air transport activity was tracked through the flight tracker, registering a substantial drop in the air transport activity in its five busiest airports during the pandemic.

BPS Statistics Indonesia relied on MDP to estimate *internal migration and commuters before and after the COVID-19* pandemic,[19] identifying mobility trends in the metropolitan areas during the pandemic.

- ● Using aggregated mobile app geolocation data to understand changes in population movements, business activity and travel in Melbourne, Australia
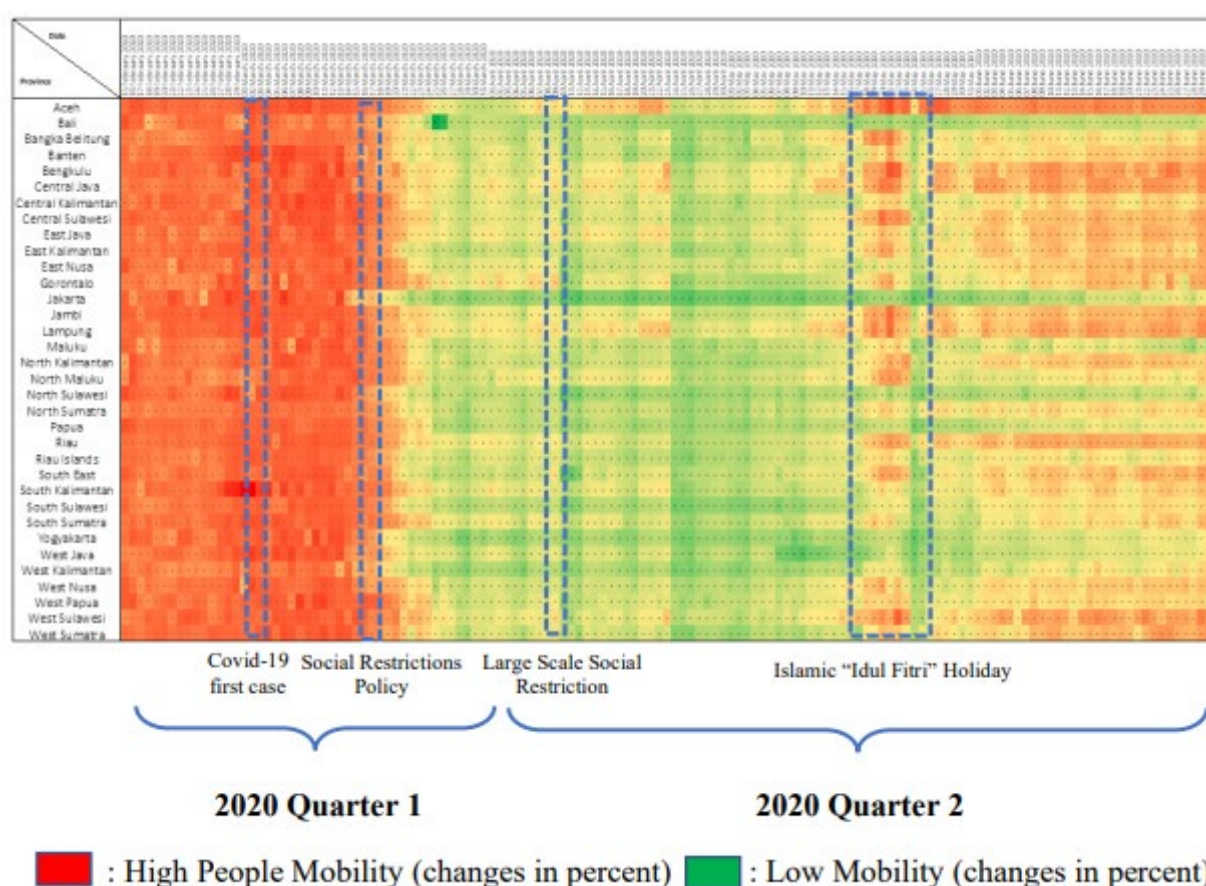
Unlike the NSOs tapping into mobile phone data, such as signaling data or call detail records (CDR) to track population mobility during COVID-19, the **ABS** collected geolocated data of mobile devices from providers of aggregated de-identified mobility data to understand changes to populations movements, business activity and reductions in travel. The aim was to identify areas of close contacts and measure foot traffic between two points of interest, including businesses, to understand the level of activity.

---

16   Data Ventures, "COVID-19 Impact on Local Councils' CBD Population," October 12, 2020, https://reports.dataventures.nz/bespoke/local-council-cbd-patterns/20201012-covid-19-council-cbd-behaviour-october-2020.html

17   Data Ventures NZ, COVID-19 Population Report, 6 July 2020, https://reports.dataventures.nz/population/20200706-covid-19-population-report-6th-july-2020.html

18   Github, dataventuresnz/mobility-index, Mobility index, https://github.com/dataventuresnz/mobility-index

19   https://www.unescap.org/sites/default/files/BPS-Big_Data_for_Social_and_Demographics_Statistics_Stats_cafe_2Nov2020.pdf

Figure 5. Heat map of people's mobility among 34 provinces for Retail and Recreation in Indonesia

2020 Quarter 1   **2020 Quarter 2**

■ : High People Mobility (changes in percent)   ■ : Low Mobility (changes in percent)

Source: BPS Statistics Indonesia[20]

Anonymized sample data for Melbourne, Australia was obtained through a commercial agreement with a local private company that collected such data. Whereas data was aggregated and did not permit tracking of individual position, it provided valuable insights into population mobility during the pandemic. While the ABS has experience in analyzing aggregated and anonymized CDR data from a 2016 trial project, in the context of COVID-19 the ABS preferred accessing anonymized mobility data shared and collected through the user's explicit consent. The ABS also explored other data sources from the public records and private sector to inform social and economic statistics in response to COVID-19 and publicly communicated about the different sources it used.[21]

● **Identifying high-risk COVID-19 zones based on population density and mobility in Bangladesh**

In **Bangladesh,** the **a2i** team leads data innovation efforts in the government and has done so during the COVID-19 pandemic. Recognizing the need for timely and location-based data for a rapid and targeted policy response and infection tracking during the pandemic, a2i turned to the private sector and other government agencies for new data. It collected the self-reported syndrome data and analyzed call detail records from the largest mobile phone operator in Bangladesh, with over 64 million subscribers, to track the spread of the outbreak in near real-time.[22]

20   Rendra Achyunda Anugrah Putra, Silvia Arini, Measuring the Economics of a Pandemic: How People Mobility depict Economics: As Evidence of People's Mobility Data towards Economic Activities, https://www.imf.org/-/media/Files/Conferences/2020/8th-stats-forum/paper-rendra-putra-and-silvia-arini.ashx

21   Australian Bureau of Statistics, "ABS responds to COVID-19", https://www.abs.gov.au/abs-responds-covid-19

22   Ramiz Uddin, presentation "Achieving the SDGs in a time of COVID-19: Bangladesh Perspective", 6th International Conference on Big Data for Official Statistics, https://unstats.un.org/unsd/bigdata/conferences/2020/presentations/day2/session4/Part%20B-1.%20Ramiz%20Uddin.pdf

The self-reported syndrome data (cough, fever, and shortness of breath) along with information about contacts with someone with symptoms or COVID-19 positive, were collected from several sources: 1) a hotline number equipped with an Interactive Voice Response (IVR) system of National helpline and health helpline, 2) several internet and mobile applications and 3) an unstructured supplementary service data (USSD) based messaging system. The responses were geolocated based on the nearest cell phone tower of the respondent and mapped to upazila (sub-district) level. The Ministry of Telecommunications facilitated access to anonymized data from the MNO.

Figure 6. COVID-19 zoning map for Dhaka, Bangladesh highlighting hotspots
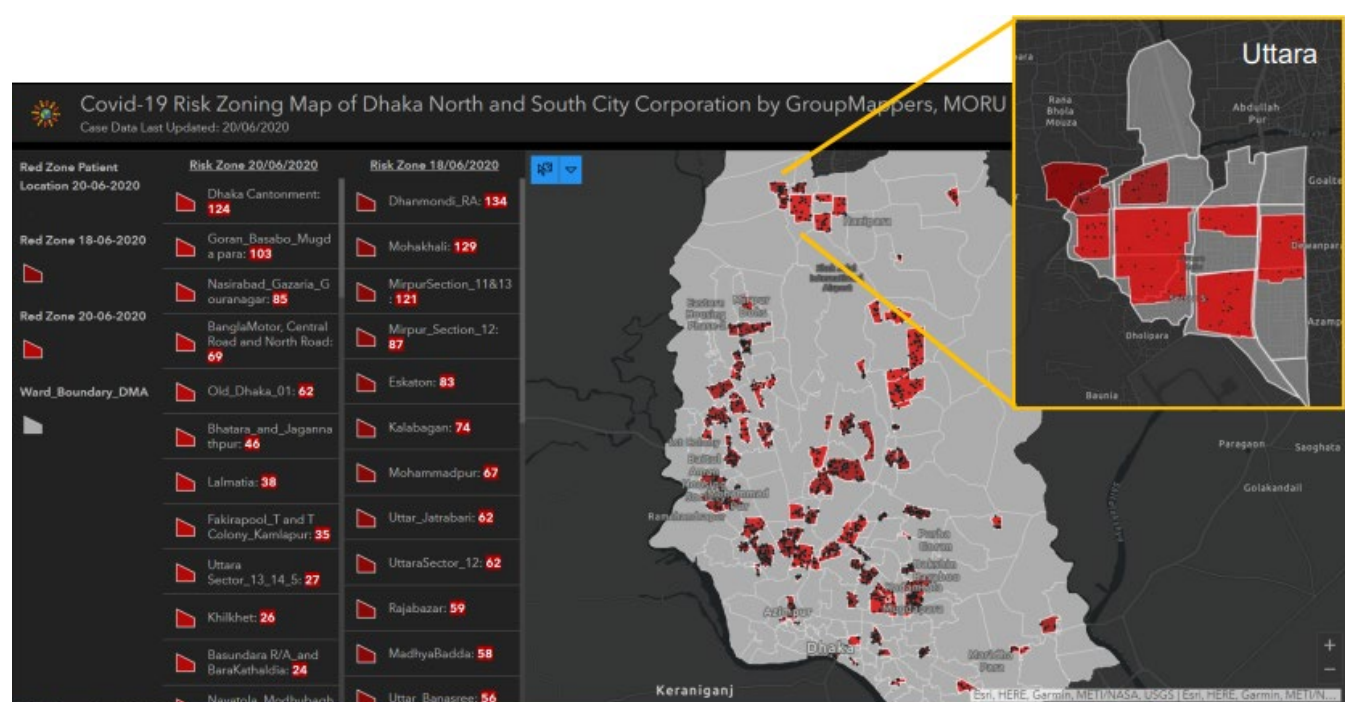
Source: Ramiz Uddin, UNDP

A Task Force including the Ministry of Health and a2i was created to analyze the data. A team of data scientists and epidemiologists from several leading universities provided support to the Task Team in generating maps classifying the risk zones across the country at the district and upazila levels. Despite the success in accessing data and generating risk maps in record time, the a2i team recognizes the need for greater internal capacity, data management, data security and privacy protection, and more data partnerships.

2. Labour statistics

Labour statistics refer to statistics about jobs and joblessness and in addition to labour supply, they provide data on working conditions, poverty and inequality, competitivness, and industrial relations. At the global level, International Labour Organization is the leading source of labour statistics. At the country level, NSOs usually compile labour statistics from employers' reports, taxation offices, the population census and household and business surveys. During the COVID-19 pandemic, several NSOs tapped into both administrative and internet data to understand labour dynamics during the crisis.

- Estimating Australian jobs and wages during COVID-19 using payroll data

The ABS tapped into private sector data, such as data on financial transaction from the Australian Banking Association, to estimate the economic impact of COVID-19 on the economy. However, to understand the jobs and wages dynamic during the same period, the ABS explored administrative big data from the Australian Taxation Office (ATO)'s Single Touch Payroll (STP) system.[34] As the crisis required fast and

reliable information on the developments in the labour market, the ABS developed with the Australian Tax Office a quick index, as presented in Figure 7, 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.

Figure 7. Payroll jobs and total wages indexed to the week ending 14 March 2020 in Australia
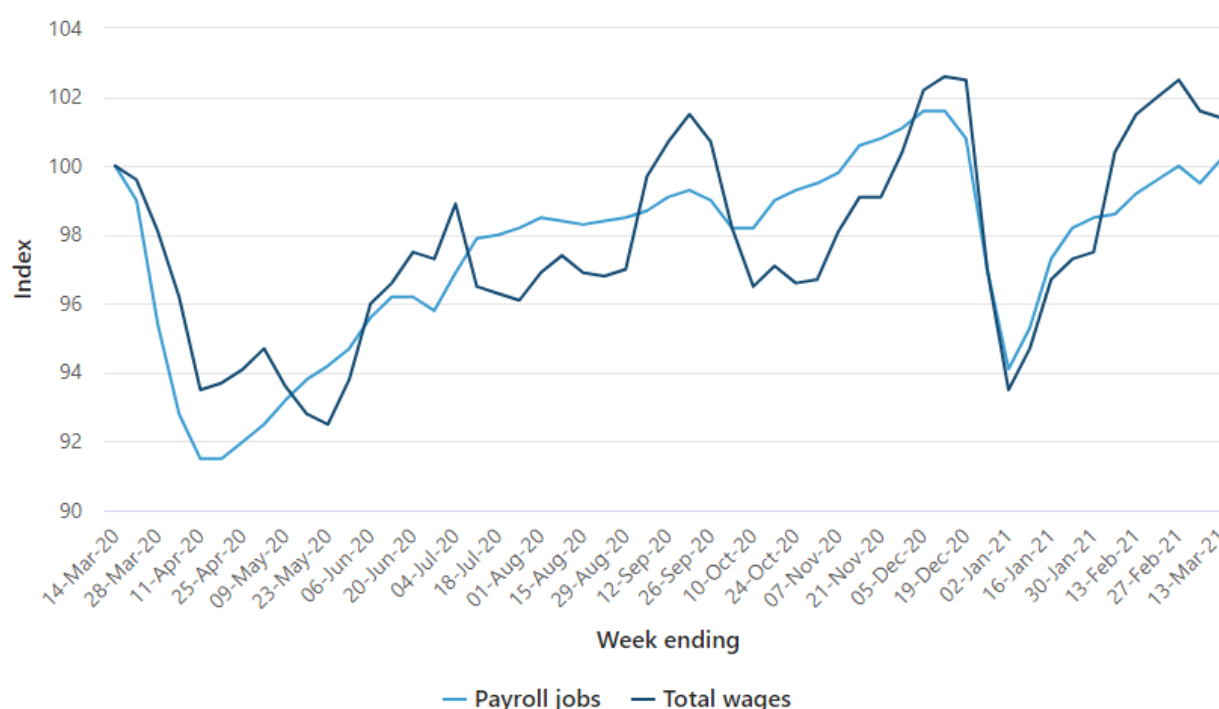
Source: Australian Bureau of Statistics (ABS)

● Estimating employment indicators using online data in Indonesia

During COVID-19 pandemic, **BPS Statistics Indonesia** analyzed data for the online vacancies advertisements on the job portal www.jobs.id to assess the employment situation. Online data showed a decline in online job vacancy advertisements across all sectors during the period January-May 2020.

3. Health statistics

Health statistics include information about mortality, morbidity, health systems, health service coverage and risk factors. Traditionally, health statistics are collected from health records, disease surveillance systems, hospitals and household surveys. Whereas big data has been widely used in the healthcare system, advancing research in health, identifying new drugs, and improving the treatment and diagnosis of patients, there is also potential of using big data for health statistics. In the region, BPS Statistics Indonesia piloted a project exploring the use of social media data to measure mental health statistics.

● Harnessing social media data to measure mental health statistics in Indonesia

BPS Statistics Indonesia utilized unstructured Twitter data to experiment producing mental health statistics

form social media data.[23] The NSO conducted the pilot during 1 January – 30 April 2020 collecting data through Twitter API. It manually identified key words related to mental health and classified tweets into four categories: depression, stress, generalized anxiety disorder, and common mental disorders. It then trained new data with machine learning algorithms. As the pilot coincided with the beginning of the COVID-19 pandemic, the results showed a spike in the number of tweets containing the key words of interest. In this pilot project, BPS Statistics Indonesia managed to identify potentially relevant content for mental health indicators. As a next step, the NSO considers exploring other social media platforms such as Facebook, YouTube and Instagram for deeper sentiment analysis.

4. SDG indicators

Multiple NSOs in the region are exploring big data sources to address data gaps in official statistics to monitor the SDG indicators. Whereas UNEP has identified 93 official SDG indicators as environment-related, the authors are not aware of any similar identification process for population and social-related SDGs or for economic-related SDGs. The identification of thematic or domain-related SDG indicators allows for an analysis of which big data sources have the most potential. For example, geospatial information and Earth Observation data have been identified as having the most potential for environment-related SDG indicators.

Examples for big data use for two SDG indicators which could be considered population and social-related SDG indicators are provided below and additional information on the use of big data for all SDG indicators can be found in ESCAP's Working Paper *Big data for the SDGs – country examples in*

compiling SDG indicators using non-traditional data sources.[24]

● SDG indicators 9.1.1 and 7.1.1 - Using geospatial data to track population's access to all-season road in the Philippines and Mongolia and population's access to electricity in Mongolia

The Rural Access Index (RAI) measures the proportion of people who have access to an all-season road within a walking distance of 2 kilometers and is indicative of the extent of access of people living in rural areas. For calculating SDG indicator 9.1.1(proportion of a country's population that is within two kilometers of an all-season road) for the Davao Region, **the Philippines Statistics Authority (PSA)** combined the region's Administrative Maps with Gridded Population map from WorldPop using QGIS software.[25] The indicator showed that only slightly more than half of the rural population of the Davao Region lived within 2-kilometers of an all-season road in 2015. However, there is a striking difference in rural accessibility within the Davao region, particularly between North and South, as depicted in Figure 8. While the use of geospatial information and EO data holds promising results, the methodology has some drawbacks and needs further refinement. For example, in the Philippines, there is no official definition of an "all-season" road. Furthermore, the methodology takes into account the horizontal distance from the all-season road, without considering elevation and presence of water surface that could obstruct access to the road. The UN Task Team on Measuring Rural Access supports the statistical community with improving methodologies and tools to collect necessary data for assessing Rural Access Index.[26]
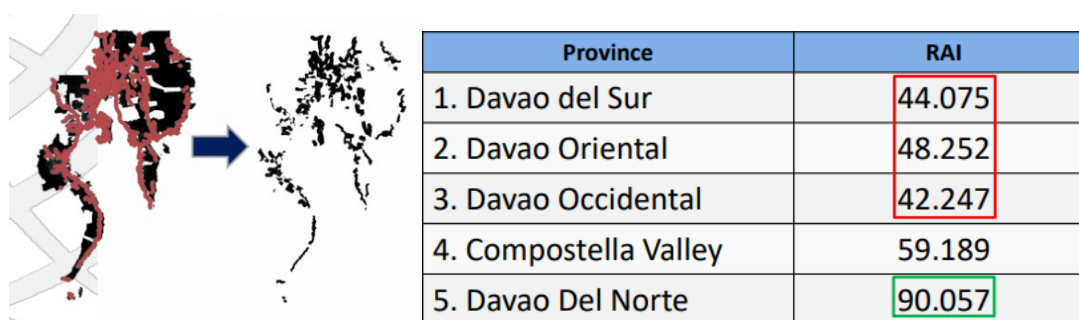
23   Fitri Andri Astuti, Nurul Aunun Nisa, Harnessing Social Media Data to Measuring Mental Health Statistics, 2020 Asia-Pacific Statistics Week, https://www.unescap.org/sites/default/files/49_Harnessing_Social_Media_Data_to_Measuring_Mental_Health_Statistics_Indonesia.pdf

24   UN ESCAP, Working Paper Series (SD/WP/12/January 2021): Big data for the SDGs – country examples in compiling SDG indicators using non-traditional data sources, https://unescap.org/kp/2021/working-paper-series-sdwp12january-2021-big-data-sdgs-country-examples-compiling-sdg

25   Mae Abigail O. Miralles, Joy Angiela H. Garraez, presentation « Measuring the Rural Access Index (SDG 9.1.1) for the Philippines, International Seminar on the Use of Big Data for Official Statistics, October 2019, Hangzhou, China, https://unstats.un.org/bigdata/events/2019/hangzhou/presentations/day2/3.%20RAI%20Presentation%20for%20International%20Symposium_Philippines.pdf

26   UN Task Team in Measuring Rural Access, https://unstats.un.org/bigdata/task-teams/rural-access/index.cshtml

Figure 8. Overlapping rural population map with the map of 2-kilometer radius from an all-season road (on the left) to obtain Rural Access Index for the Davao Region in the Philippines (on the right).

Province	RAI
1. Davao del Sur	44.075
2. Davao Oriental	48.252
3. Davao Occidental	42.247
4. Compostella Valley	59.189
5. Davao Del Norte	90.057

Source: Philippines Statistics Authority

Like PSA, the **NSO of Mongolia** also experimented estimating the SDG indicator 9.1.1 using ArcGIS desktop.[27] The main challenge in the case of Mongolia was missing road data. Following a similar approach, the NSO also experimented with estimating SDG indicator 7.1.1 (proportion of population with access to electricity). Both SDG indicators measure the population's access to basic services.

● SDG indicator 17.19.2 - Using Earth Observation data to estimate the proportion of population and living quarters in remote and inaccessible areas in Malaysia

Department of Statistics Malaysia (DOSM) entered a strategic collaboration with Malaysian Space Agency (MySA) to identify the proportion of population and living quarters (LQ) in remote and inaccessible areas using Earth Observation data. At the initial stage, GIS layer of Enumeration Block (EB) and Building Unit (BU) were published for 3 states with another 3 states in progress. As contribution to calculations of the indicator, MySA provided the latest imagery at 1.5m and 0.5m resolution captured every 14 days, which is more frequently updated that Google Maps, as illustrated in Figure 9. Among the biggest challenge of using EO data in this project is clouds and trees obstructing building units.

Figure 9. Satellite image from MySA Map vs. Google Maps

Source: DOSM

---

27   Bolortuya Jambaldorj, presentation "Eath Observation Data in the Development of Official Statistics for SDGs in Mongolia", Asia-Oceania GEO (AOGEO) symposium, March 2021, https://aogeo.net/2021/public/day1/ss1-panel-bolortuya.pdf

Contributions from development partners

Several development partners and NGOs have been supporting governments in Asia-Pacific in policy making and crisis response using big data sources. Equipped with technical expertise, technology, financial resources, and entering in partnerships with data holders, these actors tapped into various big data sources ranging from mobile network operator data, to satellite imagery, online maps and social media data to produce population estimates or to map and track post-disaster population displacements or, most recently, estimating human mobility during the COVID-19 pandemic.

The **UN Global Pulse** and its regional innovation hub **Pulse Lab Jakarta (PLJ)** have been exploring the power of both mobile phone data and social media data through a set of pilots in countries in Asia-Pacific estimating mobility and migration, displacements, and identifying trends in discrimination against women in the workplace on social media. Through a partnership with Digicel, Pulse Lab Jakarta generated evidence-based estimates on population vulnerability in case of natural disasters and traced population flows during Cyclone Donna in Vanuatu[28] and Cyclone Gita in Samoa.[29] It also mapped population displacement after the earthquake in Papua New Guinea[30] and produced real-time displacement maps after the Palu, Indonesia earthquake, combining Facebook Geoinsights and mobile network operator data.[31] In the context of the COVID-19 pandemic, PLJ used data from Village Potential Statistics (PODES) and Facebook Population Density Map to identify potential hotspots for the virus in West Java, Indonesia.[32] In Turkey, UN Global Pulse partnered with UNHCR to understand and measure refugee integration through the lens of spatial and social segregation using call detail records (CDR) data from Turk Telekom.[33]

The **Flowminder Foundation** has been harnessing the power of mobile phone data for disaster response, socioeconomic analysis and precision epidemiology since 2010. Using anonymized CDR from MNOs, Flowminder mapped migration patterns due to climate change in the Cyclone Mahasen-struck Bangladesh in 2013[34] and provided key information on the largest displacement following the 2015 Gorkha Earthquake in Nepal.[35] Flowminder also collaborated with UNFPA, WorldPop and the National Statistical Office of Afghanistan to integrate GIS data, satellite imagery and survey data (socio-demographic and economic survey (SDES) and census) for producing population estimates in a country (Afghanistan) where the last census was conducted in 1979.[36]

In the area of gender statistics, Flowminder has recently developed a high-resolution sex-disaggregated

28   Idzalika R., Amin I., Riyadi Y., Hodge G., Lee J. G. (2019). Understanding Aggregate Human Behaviour Changes in Response to a Natural Disaster in Vanuatu via Mobile Network Data Analysis. Retrieved from: http://idl.iscram.org/files/rajiusidzalika/2019/1864_RajiusIdzalika_etal2019.pdf

29   Pulse Lab Jakarta, "Informing National Statistics and Managing Disaster Risks." The Medium, October 2, 2019, https://medium.com/pulse-lab-jakarta/informing-national-statistics-and-managing-disaster-risks-b9b1e4f16a51

30   Prahara P., Zahara A., Rheza M., Khaefi M. R., Alkarisya D., Riyadi Y., Izdalika R., Hodge G. (2019). Comparing population displacement estimates from mobile network data and other sources https://www.internal-displacement.org/global-report/grid2019/downloads/background_papers/Hodge_FinalPaper.pdf

31   Pulse Lab Jakarta, "Real Time Data for Faster Decision Making in Times of Crisis." 2019, https://www.unglobalpulse.org/project/real-time-data-for-faster-decision-making-in-times-of-crisis/

32   Pulse Lab Jakarta, "Alternative Use of Traditional Data in Times of COVID-19." The Medium, September 28, 2020, https://medium.com/pulse-lab-jakarta/alternative-use-of-traditional-data-in-times-of-COVID-19-558914f23f7f

33   "Using Mobile phone call detail records to understand refugee integration in Turkey." Migration Data Portal, 2020, https://migrationdataportal.org/data-innovation/using-mobile-phone-call-detail-records-understand-refugee-integration-turkey

34   Xin Lu, David J. Wrathall, Pål Roe Sundsøy, Md. Nadiruzzaman, Erik Wetter, Asif Iqbal, Taimur Qureshi, Andrew Tatem, Geoffrey Canright, Kenth Engø-Monsen, Linus Bengtsson, Unveiling hidden migration and mobility patterns in climate stressed regions: A longitudinal study of six million anonymous mobile phone users in Bangladesh, Global Environmental Change, Volume 38,2016, Pages 1-7, ISSN 0959-3780, https://doi.org/10.1016/j.gloenvcha.2016.02.002

35   Monitoring population movements post disaster: 2015 Nepal Earthquakes, Flowminder, https://web.flowminder.org/case-studies/nepal-earthquake-2015

36   Presentation "High-resolution population mapping in Afghanistan", Remote Sensing Technology Complementing Official Statistics, February 2, 2017, https://unfpa.org/sites/default/files/event-pdf/FINAL-Afghanistan_RS_Project_EB_mtg_2_Feb_17-ajt.pdf and Donna Clarke, presentation "High-resolution population mapping with incomplete data: Applications in Afghanistan," The United Nations World Geospatial Information Congress, http://ggim.un.org/unwgic/presentations/1.4-Donna_Clarke.pdf

dynamic mapping combining traditional survey data sources with digital data (including geolocated survey data, satellite imagery, and MNO data) to provide a more detailed understanding of women's and girls' lives in Nepal.[37]

The Sri Lanka-based NGO **LIRNEasia** has also been actively exploring the potential of big data for development. Using mobile phone data, LIRNEasia estimated population density and movement in Sri Lanka.[38] In collaboration with the University of Tokyo's Spatial Data Commons,[39] LIRNEasia informed urban planning in Colombo using MNO data.[40]

Multiple organizations are actively engaged in addressing privacy issues in the age of big data. The United Nations Sustainable Development Group released the *Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda*,[41] while GSMA, which works with governments, regulators and the mobile industry, encourages responsible privacy governance practices and has developed a set of privacy principles and requirements.[42] Other organizations have also developed guidelines and software for data

anonymization and production of aggregate statistics. For example, for mobile phone data-derived statistics, Flowminder provided guidelines on producing mobility indicators[43] and aggregates from CDR data[44] in a privacy preserving way abiding to the GDPR regulation. It also developed FlowKit[45] that extends CDR data analysis beyond billing to disaster response, precision epidemiology, transport, and mobility with privacy protection in mind. University of Tokyo developed the open-source Mobipack,[46] enabling mobility analysis and visualization using de-identified telecom data. And Positium, the Estonia-based company supporting NSOs in Europe as well as in Georgia and Indonesia, developed the Positium Data Mediator.[47]

At the global level, the **UN Committee of Experts on Big Data and Data Science for Official Statistics** (UN-CEBD) includes several thematic Task Teams[48] focused on specific big data sources with members from several countries in Asia-Pacific. These thematic task teams develop handbooks and methodologies and provide trainings at the global and regional levels on the uses of various big data sources in official statistics. Furthermore, UN-CEBD developed the Big Data

37   "Towards High-Resolution Sex-Disaggregated Dynamic Mapping," 2019, https://data2x.org/resource-center/towards-high-resolution-sex-disaggregated-dynamic-mapping/

38   LIRNEasia, presentation "Mobile Network Big Data for Urban Development", September 2019, http://lirneasia.net/wp-content/uploads/2016/09/Samarajiva_UoM_Sep2016-v2.pdf and Rohan Samarajiva, Sriganesh Lokanathan, Kaushalya Madhawa, Gabriel Kreindler, Danaja Maldeniya, "Big Data to Improve Urban Planning", Review of Urban Affair, May 20, 2015, Vol I. NO 22, http://lirneasia.net/wp-content/uploads/2013/09/Big_Data_to_Improve_Urban_Planning.pdf

39   Spatial Data Commons, https://sdc.csis.u-tokyo.ac.jp/what-we-do/

40   Lokanathan, Sriganesh & Kreindler, Gabriel & de Silva, Nisansa & Miyauchi, Yuhei & Dhananjaya, Dedunu. (2014). Using Mobile Network Big Data for Informing Transportation and Urban Planning in Colombo. SSRN Electronic Journal. 10.2139/ssrn.2526642

41   United Nations Sustainable Development Group, "Data Privacy, Ethics and Protection: Guidance Note on Big Data for Achievement of the 2030 Agenda," (2017), https://unsdg.un.org/resources/data-privacy-ethics-and-protection-guidance-note-big-data-achievement-2030-agenda

42   GSMA, Data Protection & Privacy, https://www.gsma.com/iot/knowledgebase/data-protection-privacy/#:~:text=GSMA%20is%20committed%20to%20working%20with%20stakeholders%20to,GSMA%E2%80%99s%20Mobile%20Policy%20Handbook%2C%20which%20is%20updated%20yearly.

43   Flowminder, "Enabling Mobile Network Operators to produce aggregates from CDR data," https://covid19.flowminder.org/cdr-aggregates

44   Flowminder, "Enabling analysts to produce mobility indicators from CDR aggregates," https://covid19.flowminder.org/mobility-indicators

45   FlowKit, https://flowkit.xyz/

46   Spatial Data Commons, Analysis Tools – Mobipack, https://sdc.csis.u-tokyo.ac.jp/software-name/

47   Positium, Positium Data Mediator, https://positium.com/research/positium-data-mediator#:~:text=Methodology%20The%20Positium%20Data%20Mediator%20%28PDM%29%20is%20a,include%20population%20statistics%2C%20mobility%20studies%20and%20tourism%20data

48   UN Committee of Experts on Big Data and Data Science for Official Statistics (UN-CEBD)includes the following Task Teams: AIS Data, Big Data and the SDGs, Earth Observation Data, Measuring Rural Access, Mobile Phone Data, Privacy-Preserving Techniques, Scanner Data, Trainings, Competencies and Capacity Development. More information available at https://unstats.un.org/bigdata/task-teams/index.cshtml

Marketplace showcasing big data (mobile phone data, social media data, satellite imagery, road sensor data, etc.) projects relevant to population and social statistics, including projects on daytime population estimates, daily migration, mobility, happiness index, health care satellite account and many others.

Besides UN-CEBD and its thematic Task Teams, individual development organizations, which are also members of those task teams, are exploring the potential of big data sources with most of these efforts focused on human mobility and migration. For example, the International Organization for Migration's (IOM) **Global Migration Data Analysis Center** (GMDAC)[49] has been exploring the big data potential for understanding global migration. According to IOM, big data sources used in migration-related studies can be divided into three categories: *mobile-phone-based* – call detail records or mobile money transfers; *internet-based* – social media data and geo-located social media activity, IP addresses from e-mails; and *sensor-based* – Earth Observation Data. GMDAC explores the potential of big data for migration statistics and compiles examples from around the world in the Data Innovation Directory[50] on the migration data portal. Furthermore, the IOM-led Big Data for Migration (BD4M)[51] initiative seeks to accelerate the responsible and ethical use of new data sources and innovative methodologies to inform migration policies and programmes.

Another global initiative is Development Data Partnership.[52] This initiative is led by the World Bank and addresses primarily to development organizations. This partnership brings together international organizations and 25 data partners including companies such as Google, Esri, LinkedIn, Twitter, Waze that share data for development and humanitarian projects. This partnership has projects in several countries in Asia-Pacific. Several of the latest projects focus on human mobility during the COVID-19 pandemic, such as analyzing spatial accessibility of Health Facilities in

Indonesia and the Philippines using Facebook and Mapbox Data[53] and tracking the second COVID-19 wave in Viet Nam using Facebook Mobility Data.[54]

Challenges in using big data sources to produce population and social statistics

Big data sources bring considerable advantages for the production of population and social statistics such as increased coverage, level of granularity, frequency, timeliness and reduced costs. However, multiple challenges persist.

There are two particular challenges to the use of big data sources for population and social statistics which are not as evident as for economic and environment statistics. The first big challenge is the data collection unit – population and social statistics are mainly based on personal data and personal data and are, more often than not, subject to national privacy laws in addition to statistical laws. The second challenge is the criticality of population statistics to so many aspects of good governance such as parliamentary representation, social protection mechanisms and financial distributions across federal, state and local government arrangements. The robustness and quality of population statistics is therefore of paramount concern to Governments, societies and economies.

With few exceptions, most of the big data use cases in population and social statistics are for population statistics and remain experimental in nature. Successful big data pilots do not necessarily lead to integration of the data sources into statistical production for a variety of reasons, not least because long-term data access for integration requires different commitments and arrangements.

Nevertheless, despite the specificities and differences across big data sources, several common challenges related to big data use for official statistics have been identified. They mainly refer to 1) legal frameworks

49   Global Migration Data Analysis Center, https://gmdac.iom.int/

50   Global Migration Data Analysis Center, Data Innovation Directory, https://migrationdataportal.org/data-innovation

51   Big Data for Migration Alliance, https://data4migration.org/

52   Development Data Partnership, https://datapartnership.org/

53   Development Data Partnership, "Facebook, Mapbox: Analyzing Spatial Accessibility of Health Facilities in Indonesia and Philippines", https://datapartnership.org/planning-for-equitable-access-to-health-infrastructure/

54   Development Data Partnership, "Facebook: Tackling COVID-19 Second Wave using Facebook Mobility Data in Vietnam", https://datapartnership.org/updates/tackling-covid19-second-wave-using-facebook-mobility-data-in-vietnam/

guiding the production of official statistics and access to non-traditional data sources, 2) data access and partnership negotiations, related to terms of access, privacy protection, and supply continuity, and 3) the internal technical, financial and technological resources of the statistical organizations and the national statistical system in general. These challenges are also relevant to other areas of statistics.

Legal constraints are often referred to as an impediment for using big data for official statistics. Countries facing this issue cite a regulatory framework that limits the data sources that could be used to produce official statistics. However, exploration of new data sources through pilot projects could be relied upon to test and build the case for integration of alternative data sources where relevant and needed.

Data privacy is also an important challenge that needs to be addressed when considering big data sources, particularly personal data, for production of official statistics. National legislative frameworks on privacy and personal data protection vary across countries in the Asia-Pacific region and globally. Furthermore, several NSOs in the region note the absence of a legal framework obliging private companies to share their data with the public sector to produce statistics and inform public policy, or guiding data access from the private sector.

Whereas in some countries statistical legislation guides the production of statistics, meaning that any changes to data production and underlying data sources (e.g. to administrative or private sector sources) would have to undergo legal review, in other countries, statistical legislation can be more generic and include scope for other data sources. Indonesia, for example, includes data collection by "*other ways according to the development of science and technology*", in addition to census, survey and administrative product compilation.[55]

Experiences of countries in the region show that statistical organizations follow **different partnership models to access private sector data**. In the example of the MNO data, organizations like BPS Indonesia, KOSTAT, Stats NZ and the a2i programme in Bangladesh negotiate access directly with data providers. In other countries, like Georgia, access to the MNO data is negotiated indirectly through the national regulatory authority, which impose private sector data sharing, at least for specific projects, through regulatory procedures. Other models include data acquisition from third party providers, as in the case of the ABS and geolocated data from mobile apps. In the case of New Zealand, an additional step was taken - creation of a subsidiary of the statistical office known as Data Ventures. These models differ and are based on national specificities of countries, their regulatory frameworks, project objectives, internal resources, and convenience.

An integral part of data access negotiations and agreements are **conditions of access**. In the above-mentioned examples, some statistical organizations obtained data at no cost, whereas others did so through a commercial agreement. Geostat and BPS Statistics Indonesia, while conducting similar projects with MNO data, obtained access differently: Geostat had free access granted by legal provision, while BPS Statistics Indonesia obtained it through a commercial agreement. The experience of countries in the region indicates that most of the NSOs obtain mobility-related data on a commercial basis. These conditions should also be analyzed in the development stage of the project, whether it is a pilot or scaling up.

The **costs of private sector data** for official statistics remains a subject of active debate. Some participants of the statistics community argue that official statistics are a public good and that data to produce them should be obtained for free or not exceeding marginal costs. Others take into consideration that companies incur costs associated with data storage and processing, such as anonymization and aggregation, for sharing in accordance with internal and national data privacy processes and regulation, and that these costs should be covered. Also, given the new nature of big data sources, some data collecting companies build new business models around the data they collect.

Another element to consider is access to **private sector data** versus access to **private sector insights**. The presented case studies highlight these differences.

---

55   Ali Said, BPS Indonesia, presentation at Stats Café on 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

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, others negotiate direct access with data providers or national regulatory bodies, building partnerships and co-designing data analysis methodologies. In some cases, NSOs obtain access to anonymized and aggregated data that they analyze internally, whereas in others they can only access the insights that were produced from applications of the methodologies and algorithms developed by them. In one case analysis is performed on 'raw' or primary data, while in other cases on highly aggregated data.

Integration of big data sources and methods into the production of official statistics requires investments, particularly for internal capacity building, for development of the necessary IT infrastructure for data collection, storing, processing and analysis, and for continuous data access, especially when data are obtained on a commercial basis. While some countries in the region have been gradually building their internal technological infrastructure to meet the requirements for using big data and Artificial Intelligence, such as the ABS, Stats NZ, NBS China, Geostat, DOS Malaysia and BPS Statistics Indonesia, other countries have recognized their technological unpreparedness, among other challenges, to start exploring alternative sources of data for official statistics. The investments should be weighed against the traditional data collection methods and the associated expenses. A trade-off may be possible, especially by the potential of big data to complement some traditional methods of data collection by providing more timely and granular data for a better-informed evidence-based decision-making.

Recommendations

Building on successful experiences and lessons learned from some of the NSOs in the region, several recommendations could be provided to the NSOs that are either embarking on the big data journey or have been already experimenting with big data for official population and social statistics.

Whereas some NSOs 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 population and social statistics and pilot small projects that if successful, could build the case for ongoing production and regulatory changes if needed.

Building trusted partnerships with the private sector for data access and experimentation is another important issue to be considered. As the interest of the NSOs in MNO data for demographic and social statistics grows, accessing their data requires long negotiations. Therefore, NSOs are encouraged to **negotiate with the private sector early on to explore partnership opportunities** and understand the challenges.

Regulatory authorities should consider developing guidance on proprietary data access by statistical organizations and systems. In the absence of clear guidance on data access, exchange and use, private companies may be unwilling to share data or may refer to strict corporate data protection procedures preventing them from data sharing.

NSOs are encouraged to **seek support and guidance from civil society, specialized private companies or academia**, when designing and implementing big data pilots. Developing partnerships with academia can also be helpful, especially when the NSOs have limited internal capacity for big data processing and analysis. UN-CEBD and its Task Teams are an important source of support, knowledge, and cooperation opportunity.

To address consent, privacy and security issues, NSOs should **collaborate with national personal data protection agencies at the design stage of any project**. Collaboration with these authorities will help mitigate risks associated with personal data use and guide collaboration with data providers. Also, communication and transparency about how personal data are used and protected through data privacy and security is important for building trust with the general public. NSOs should consider engaging in an open dialogue with citizens about the issues of consent and personal data protection, particularly when considering integration of administrative data and private sector data into the census or during a census period.

To support greater awareness of the uses of big data in population and social statistics, NSOs are also encouraged to **document and actively share their experiences** at regional and global forums., Regardless of their experience with big data, NSOs can also rely

on the wealth of information already available. In this regard, ESCAP Statistics Division has been actively supporting regional knowledge sharing around the use of big data for official statistics through dedicated efforts, such as Stats Cafés focused on big data for official statistics, Data Integration Community of Practice (DI CoP), Stats Briefs, Working Papers and blogs highlighting the use cases of big data in official statistics.

Conclusion

In the Asia-Pacific region, statistical organizations have been actively experimenting with the use of big data sources in the production of population mobility and urbanization statistics, but less so for other population statistics such as births and deaths or social statistics. Mobile Network Operator or MNO data are the most researched data in the area of population mobility statistics. Other data sources, such as Facebook and Google data, and public transportation data are also researched, but at a lesser scale and mostly by international development partners.

The case studies presented in this brief provide specific examples of how NSOs in the region integrate big data into the production of population and social statistics. The case studies illustrate different data sources, methods, partnerships, and approaches employed.

They also illustrate big data opportunities, possibilities, and challenges.

Despite a multitude of examples, most of the big data projects in population and social statistics remain at the experimental phase. However, a few exceptions, as in the example of mobile phone data for mobility and tourism statistics in Indonesia, can be noted.

Whether statistical offices are already experimenting or are only considering the use of big data, they face a broad range of challenges, which need to be addressed. These include legal frameworks guiding statistical production, data access and data privacy; partnerships with the private sector; internal technical and technological capacity building and financial resources. This brief provided an overview of the most common challenges, how some NSOs in the region are responding to them, and several recommendations on embarking on or moving forward on the big data journey.

The COVID-19 pandemic has been the impetus for more exploration of big data for timely and high-frequency insights, particularly in the area of human mobility. The pandemic has uncovered important gaps in traditional statistics and showed the value of non-traditional data for timely insights. To remain relevant, NSOs should further explore the opportunities that big data sources and methods can provide to produce official population and social statistics.

---

[i] *Additional inputs to this Stats Brief have been received from Gary Dunnet, Deputy Chief Methodologist, Statistics New Zealand; Jisook Yoon, Director, Big Data and Statistics Division, KOSTAT; Young Ran Kim, Deputy Director, Big Data and Statistics Division, KOSTAT; Ramiz Uddin, Head of Experimentation at Accelerator Lab, UNDP Bangladesh; Sangita Paul, Data Scientist, a2i, Bangladesh; Anders Holmberg, Chief Methodologist and General Manager, Methodology Division, Australia Bureau of Statistics; Ric Clarke, Director of Emerging Data and Methods, Australia Bureau of Statistic; Shorena Tisklauri, Chief Specialist of Methodology and Quality Management Division, National Statistics Office of Georgia; Ali Said Director, Statistical Analysis and Development, BPS Statistics Indonesia; Setia Pramana Big Data Specialist, BPS Statistics Indonesia; Liben Hou, Deputy Director Division of General Affairs, Department of Statistical Design and Management, National Bureau of Statistics of China. The brief also benefitted from discussions and presentations during the Asia-Pacific Stats Café held on November 2, 2020.*