Can mobile phone data be used for official statistics? Asia and the Pacific says yes

The Statistical Offices of Bangladesh, Cambodia, Georgia, Indonesia, Malaysia, Mongolia, Nepal, the Philippines, Thailand and Viet Nam, ten in all, met in Jakarta, Indonesia, 11-14th June 2019 to share experiences on why and how to use mobile phone data for official statistics at a Regional Workshop on the use of Mobile Phone Data for Official Statistics.

Warmly welcomed by Dr Kecuk Suhariyanto, Chief Statistician of BPS Indonesia, himself a champion for innovation and responding to increasing demands for timely and relevant official statistics, and representatives from the National Statistical Offices were joined by Dr Wayne Wobcke from the University of New South Wales School of Computer Science and Engineering and resource persons from UN ESCAP, UN DESA, UN Pulse Lab Jakarta and Positium, to explore what mobile phone data means to the production of official statistics.

Mobile phone data and the UN Fundamental Principles of Official Statistics

The workshop started and ended with a participative discussion on the ten UN Fundamental Principles of Official Statistics. Participants explored a range of questions such as: if we are to use mobile phone data for official statistics, will the statistics be produced on an impartial basis (fundamental principle one)? Can we ensure that the methods and procedures for storing big data has been decided based on strictly professional considerations (principle two)? How do legal mechanisms for collecting and using mobile phone data look like (principle three) and how can we guarantee that individual data and statistics are strictly confidential (principle six)? The general consensus was, however, that with careful considerations of these questions, users can continue to trust official statistics also when it make use of mobile

phone data, either as a direct data source or indirectly through validation and quality assurance processes.

Mobile phone data use in Asia and the Pacific

The use of mobile phone data in Asia and the Pacific can be considered as in its infancy but interest is strong. The workshop therefore benefitted from the experiences from Estonia, which has made extensive use of mobile phone signal data. The workshop consultant from the Estonian company Positium, shared his country's experience in why and how to use mobile phone signal data. In the case of Estonia, the data is used various use cases, including population, tourism and human mobility statistics.

In Asia and the Pacific, BPS Indonesia has started using signaling data from mobile phones in the production of its cross-border inward tourism statistics for those cross-border areas where there is no immigration office collecting data on cross-border tourists. This work was then expanded into all cross-border inward tourism to validate official statistics produced using traditional data collection methods. These experiences highlight how mixed modes can improve statistics quality substantially. BPS Indonesia are looking to expand to other types of tourism statistics such as domestic tourism, particularly as their confidence in the tourism statistics based on signaling data grows.
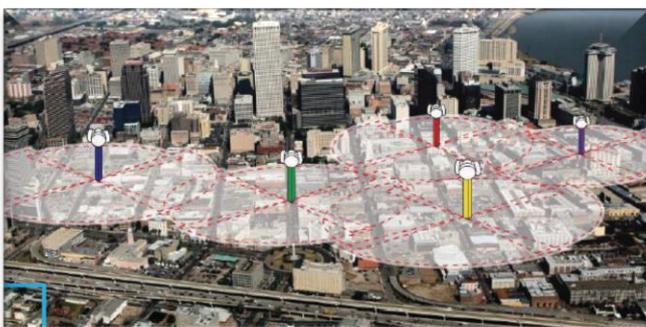
Why are countries choosing to use mobile phone data?

Over the course of the four-day workshop, countries shared their thoughts about the potential benefits of using mobile phone data in their own context. Ten key reasons were identified:

1.   Much more detailed statistics, through Sub-national estimates and granularity

There was broad consensus that mobile phone data can provide much better granularity in data than traditional use of surveys. Georgia shared two examples where sub-national statistics may be possible if mobile phone data was used. Producing sub-national population estimates by using mobile phone data to produce internal migration statistics, and by using mobile phone data to produce sub-national tourism statistics, which are currently not possible due to sample size limitations with the existing tourism surveys.

2.   Sampling frame validation

Georgia spoke about using mobile phone data to validate the household survey sampling frames used for their general social survey and their domestic tourism survey sampling frame. Both sampling frames are currently only updated every 10 years after the Census.

3.   Validating survey statistics

Both Indonesia and Georgia highlighted how mobile phone data can be used to validate and support findings from their survey-based tourism statistics, particularly domestic tourism.

4.   Improving accuracy

Accuracy is also a recognized driver for using big data, including mobile phone data. The accuracy of traditional data collection methods, particularly sample surveys, needs to be considered when assessing various data sources for official statistics. The fifth UN Fundamental Principle of Official Statistics calls on statistical agencies to assess the quality, cost, timeliness and respondent burden when choosing data sources. Sample sizes for surveys may not support statistical priorities, such as sub-regional statistics. Mobile phone data, in contrast, could be a complete census of subscribers, and provide scope for more accurate sub-national statistics.

Philippines shared an example of exploring the use of mobile phone data to improve the accuracy of de-facto or daytime population estimates. Philippines currently use the census, conducted every ten years, to derive de-facto estimates based on a sample to capture school age population and those in the labour force. The mobile phone data offers a potentially alternative method to produce the daytime population during the inter-census period, which are often requested by their users at local government units, especially in metropolitan areas.

5.   Reducing cost

Cost is a recognized driver by many for using big data, including mobile phone data. The cost of traditional data collection methods, such as household surveys, is an increasing reason why national statistical offices are looking at alternative data sources. As one participant put it, using mobile phone data may not replace the need for household surveys, but may reduce the need to conduct household surveys on a monthly basis. In Indonesia, the frequency of tourism surveys may be able to be reduced from monthly or quarterly, or even six monthly, with the use of mobile phone data in the intervening months which are then ground-trusted and calibrated with the survey estimates.

> **Ten reasons for using mobile phone data for official statistics:**
>
> 1.   Much more detailed statistics, through Sub-national estimates and granularity
> 2.   Sampling frame validation
> 3.   Validating survey statistics
> 4.   Improving accuracy
> 5.   Reducing cost
> 6.   Improving timeliness
> 7.   Reduce respondent burden
> 8.   To meet unmet demand
> 9.   To fill data gaps with the SDG indicators
> 10.   Building new business models
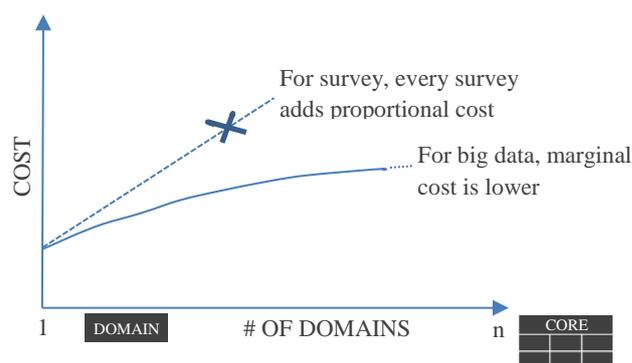
6.   Improving timeliness

Timeliness is also a recognized driver for using big data, including mobile phone data. The timeliness of traditional data sources, including censuses, surveys and administrative data, are all issues which big data may be

able to address. Participants discussed the ten year frequency of a Population Census as the main source for population estimates, and the potential for mobile phone data to provide a data source for inter-censal population estimates.

7.   Reduce respondent burden

Another issue, which could potential be alleviated by using mobile phone data is the increasing burden on respondents. With the growing demand for information and new surveys, respondents are sometimes asked to participate in multiple surveys lasting several hours. Fewer or less frequent surveys mean less respondent burden. Most participants said they came to the mobile phone data workshop with the hope that this data source can help reduce respondent burden.

8.   To meet unmet demand

Bangladesh shared their plan to use mobile phone signal data to address an unmet demand for a better understanding of the patterns of traffic jams in Dhaka. Quoting a World Bank study which showed traffic jams result in 32 million hours of lost working time for individuals, participants demonstrated how the national statistical office could produce statistics which meet the practical test of utility (fundamental principle one) with the major beneficiary being the transport users in Dhaka.

Malaysia shared an example of exploring the use of mobile phone data to produce de-facto or daytime population estimates, for transportation planning and internal migration. There was agreement that the use of big data should be driven by user needs rather than data availability.

9.   To fill data gaps with the SDG indicators

Cambodia shared an example of exploring the use of mobile phone data in Cambodia to fill data gaps with the SDG beside traditional surveys. In Cambodia, mobile phone data can also help verify survey results.

10. Building new business models

Lastly, participants briefly discussed new business models which big data may offer. For example, a national statistical office may wish to supplement their official statistical data series, such as official population and migration estimates, with other, lower 'priority' statistics such as human mobility statistics to/from events, as part of a commercial service offering.

In addition to these examples, countries also considered the use of metadata on mobile phone usage for ICT statistics. Thailand faces a challenge of producing internet user and behavior statistics, with frequent dissemination and for particular groups such as youth. Mobile phone usage data and other phone log and internet usage data offer the potential to address these data challenges.

Why are others using mobile phone data and other big data?

In addition to discuss amongst themselves about why to use mobile phone data, participants visited Pulse Lab Jakarta and heard about many exiting big data experiments from across the region including.

- Mobile phone data being used for post-earthquake displacement studies in Papua New Guinea and Vanuatu

- Relative wealth index calculations based on mobile phone data in Papua New Guinea

- MIND or Managing Information for Natural Disasters, an automated, open source platform that integrates multiple non-traditional data sets to aid logistics planning and information management following natural disasters.

- VAMPIRE or Vulnerability Analysis Monitoring Platform for Impact of Regional Events, an integrated map-based visualisation tool to track the impact of drought for vulnerable populations.

- After Dark: Encouraging Safe Transit for Women Traveling at Night, research focused on understanding the unique experiences of women who travel at night using public transportation

- Improving traffic safety through video analysis, a video analysis pipeline for the purpose of improving traffic safety in Jakarta. The pipeline transforms raw traffic video footage into databases that are ready to be used for traffic analysis

Participants found these examples very interesting and were already thinking about how similar projects could be undertaken in their countries.

Technical challenges in using mobile phone data?

Participants spent about 50% of their time working with mobile phone data. Using a single phone record of data over three years, participants learned about data structures, how to pre-process the data, how to extract the necessary information to detect necessary information in order to turn the data into a statistic. Hands-on training was an invaluable part of the workshop. Participants saw first hand some of the methods they would need to apply when using mobile phone data.

There was recognition of the work of the UN Global Working Group on Big data's efforts to produce standards, methods and guidelines to support countries in their efforts in using new data sources for producing official statistics.

Big data and modernization

Participants heard that the use of big data is just one piece of a bigger modernization effort. Estonia's experience, for example, demonstrated that the use of mobile phone data requires considerable modernization to IT infrastructure if the NSO decides to transfer the data into a central location. The mobile phone data is large – it's big data after all. It doesn't fit into standard NSO data warehouses. For one year, in Indonesia there are over 330 million subscribers, 1500 billion records, and over 500 Terabytes of data.

Next steps

Participants developed draft roadmaps for the use of mobile phone data in their countries. The roadmaps outlined the problem to be solved, how to solve it, data access, cooperation models needed, stakeholders to involve, legal requirements, etc.

Ensuring public acceptance of the use of mobile phone data for official statistics was a common element of all roadmaps, as was engagement with stakeholders across ministries and with data providers. Countries also noted having a clause in law that enables access to big data (as a special case of private sector data) would be a helpful step toward making better use of mobile phone data.

Countries agreed to share the roadmaps with their management as a concrete outcome from the workshop. The UN Secretariat, represented by the Statistics Divisions of UN ESCAP and UN DESA, look forward to following up with countries on how well the road maps were received and on progress made. The Secretariat also looks forward to providing technical assistance to these countries, and shared details of forthcoming opportunity at the ISI World Statistics Congress, 2019, during the Symposium on Data Science and Official Statistics, to be held 15-17th August 2019 in Kuala Lumpur, Malaysia.

"Consultation and technical assistance is priceless in this project[1]"

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

---

[1]*Quote from one of the participants at the Regional workshop on the use of mobile phone data for official statistics.*