ADB BRIEFS

KEY POINTS

- The Sustainable Development Agenda aims to leave no one behind. From a data collection perspective, it entails enhancing current systems to gather adequate information about different population groups, especially the small and vulnerable segments of society. Monitoring the SDGs requires better, more granular data to be available faster. To meet this challenge, national statistical systems should capitalize on new data sources, particularly big data.
- The readiness of national statistical systems to harness big data and other innovative data sources depends on several factors, including hardware and software requirements for storing, examining, and visualizing big data. Official statisticians need to strengthen their skills in analyzing unstructured, unfiltered, and complex collages of data points collected for distinct purposes and which may not have clear target populations.
- Using big data is not just about the actual data, but also about the data ecosystem, i.e., the need for partnerships, frameworks, and communication strategies. Harnessing big data requires research to determine what does and does not work, capacity development in national statistical systems, and stronger partnerships and frameworks to sustain more timely, granular, and meaningful statistics.

Readiness of National Statistical Systems in Asia and the Pacific for Leveraging Big Data to Monitor the SDGs

Jose Ramon Albert
Senior Research Fellow
Philippine Institute for Development Studies

Arturo Martinez Jr.
Statistician
Statistics and Data Innovation Unit
Economic Research and Regional
Cooperation Department

Katrina Miradora
Consultant

Jan Arvin Lapuz
Consultant

Marymell Martillan
Consultant

Criselda De Dios
Economics and Statistics Analyst
Statistics and Data Innovation Unit
Economic Research and Regional
Cooperation Department

Iva Sebastian-Samaniego
Economics Officer
Economic Analysis and Operational
Support Division
Economic Research and Regional
Cooperation Department

INTRODUCTION

In September 2015, the world committed to the 2030 Agenda for Sustainable Development and the Sustainable Development Goals (SDGs), which have broad and interconnected social, economic, environmental, and governance aspects. The SDGs constitute a universal call to action with countries setting forth 169 specific, time-bound, and quantifiable targets for 2030 to leave no one behind (United Nations 2015).

Monitoring the 17 SDGs and the concomitant 169 targets have created new and huge demands on national statistical systems (NSSs). NSSs will have to report on more than 230 indicators in the official SDG indicators framework, with many of these indicators requiring disaggregation by location, sex, age, income, and other relevant dimensions.

ADB

Common instruments used to fund statistical programs (such as bilateral grants and multi-donor trust funds) are not sufficient and do not substantially increase over time (PARIS21 2017). In some cases, either due to the political economy or competing resources, national governments are unlikely to increase resources for statistics development. National statistics offices (NSOs) and other statistics producers in NSSs, particularly in developing economies, then need to find ways of addressing the SDG requirements as well as other data demands using the resources available to them.

In Asia and the Pacific, while disaggregation of statistics by location is available for several SDG indicators, granular data is sparse for other indicators by sex; it is even scarcer, if not absent, for special groups such as disabled persons and indigenous peoples. This is one of the key findings of an Asian Development Bank (ADB) and United Nations Economic and Social Commission for Asia and the Pacific (UNESCAP) survey of NSOs in 2017. The same survey reported that more than half of the 22 reporting NSOs from ADB and/or UNESCAP member countries are utilizing small area estimation (SAE) methods, which entail integrating different data sources, to address disaggregation-related limitations that arise when each data source is used individually and independently.[1] However, SAE methods have limitations too, especially when data integration is confined only to conventional types of data. Related to this, results of the same survey on SDG data compilation conducted by ADB and UNESCAP in 2017 also highlighted that many NSOs acknowledge that the only way that they will be able to meet the disaggregated data requirements of the SDGs is for them to utilize innovative methods and data sources to complement conventional approaches.[2]

BLENDING TRADITIONAL AND INNOVATIVE DATA SOURCES IN THE DATA REVOLUTION

Organizations from both public and private sectors collect data, sometimes as a by-product of an administrative function, and in other cases, through data collection systems designed specifically for generating statistics to inform decision makers. In the private sector, firms either directly collect data and produce statistics for internal use, or subcontract market research organizations to conduct data collection activities and summarize business insights from data collected. In the public sector, NSSs produce official statistics, such as the gross domestic product, poverty rates, consumer price indices, unemployment statistics, net enrollment ratio, agricultural production, and statistics on crime and safety. These statistics serve as inputs in the formulation, implementation, and monitoring and evaluation of public policy. NSSs produce official statistics usually from censuses,

surveys, administrative reporting systems, and other compilations of secondary data following established concepts, definitions, methods, and classification systems. The choice of data source in official statistics production is often guided by considerations on cost and reducing the burden on respondents of surveys and censuses.

All over the world, innovations in information and communications technologies (ICT) have led to a data revolution: more data is being captured, produced, stored, accessed, analyzed, archived, and re-analyzed, and at an exponential pace (Independent Expert Advisory Group on a Data Revolution for Sustainable Development 2014). The resulting hyperconnectivity that connects persons to persons, people to machines, and machines to machines has led to a deluge of digital data or big data, characterized by what Hilbert (2013) called the three V's—volume, velocity, and variety. Big data can be categorized largely into three main sources: (i) human-sourced information (e.g., social networks); (ii) process-mediated data (e.g., search engines, commercial transactions), and (iii) machine-generated data (e.g., mobile phone location). All of these voluminous, fast-paced, and complex data, however, are often by-products of transactions from hyperconnectivity, and as such, are often unstructured and do not necessarily relate to a target population, unlike traditional data sources of official statistics.

While the private sector has been the one to primarily engage in harnessing big data, big data statistics are emerging and complementing official statistics compiled by the NSSs from traditional data sources to monitor and analyze the public sector's development targets. For instance, official statistics on flu incidence released by the US Centers for Disease Control and Prevention (CDC) have been found to correlate strongly with the number of Google searches from the US on the term "flu" (Ginsburg, et al. 2009). Twitter conversations in Jakarta, Indonesia on rice prices have also been reported to be a reasonable means of monitoring actual prices of rice in the Indonesian capital (Letouzé 2012). Through the Open Transport Partnership, near real-time traffic data and statistics, including speeds, flows, and delays at intersections, that are sourced from anonymized global positioning system (GPS) data of ridesharing drivers are being used to examine critical areas in traffic management in the Philippines and other developing countries (Krambeck, et al. 2015).[3] More granular data on poverty have been sourced from anonymized call detail records and other information on the behavior of mobile users (Smith, et al. 2013). Digital traces of mobile phone usage have also been used to track population movements and examine people's behavior during disaster events. (Liang, et al. 2014). Nighttime luminosity maps with high-resolution daytime satellite images have been used to yield estimates of household consumption and assets (Jean, et al. 2016). More details on previous and ongoing efforts for enhanced compilation of statistics for development-related purposes are provided below.

---

[1]   The SAE method is a class of statistical techniques that are designed to enhance the reliability of direct survey estimates for small areas (or small subpopulations) without increasing the survey's sample size and by integrating auxiliary information (such as that from census records).

[2]   "Big data" is an umbrella term used to describe the constantly increasing flow of data.

[3]   These big data have the same use as travel time surveys, the traditional method of collecting traffic congestion data, but there is much more data that can be gathered in near real-time from ridesharing drivers at nearly no cost. Data from travel time surveys are still required to analyze different vehicle types, such as motorcycles and trucks, but traffic congestion information from big data can validate statistics from traditional sources.

INITIATIVES ON BIG DATA FOR DEVELOPMENT

Big Data Work of International Development Organizations

Recognizing the need for a typology of development data that will be required to comprehensively measure all elements identified in the SDG agenda, development partners have thus begun to support countries in exploring the rich potential of big data. This section discusses some of the global, regional, and national initiatives on making use of big data that have been undertaken or are underway.

The UN Global Pulse, an information initiative launched by the Executive Office of the United Nations Secretary-General, upscaled its "now-casting" of Twitter data for monitoring rice prices to those of other commodities (beef, chicken, onion, and chili) in Indonesia. The initiative also analyzed GPS-stamped tweets in Jakarta and anonymized data from GPS navigation smartphone apps (such as Waze) to investigate commuting patterns in relation to near-real-time traffic conditions in the Indonesian capital (Pulse Lab Jakarta 2017). Global Pulse also examined anonymized mobile phone data (particularly, call details records and airtime credit purchases) to produce a set of proxies for education and household characteristics in Vanuatu and examined anonymized financial records from four financial service providers in Cambodia to determine factors affecting savings and loans mobilization, with a focus on gender disaggregation. Through its Vulnerability Analysis Monitoring Platform for Impact of Regional Events platform, Global Pulse has also been examining satellite imagery to map locations with climate and rainfall anomalies and to provide climate data visualizations and early warning alerts to policymakers and the general population.

The World Bank launched a new program called Innovations in Big Data and Analytics for Development. The program kickstarted the Big Data Innovation Challenge, which aims to scale early pilots into projects that solve significant development challenges, and to establish best practices for using big data analytics to steer evidence-driven development. In its pilot launch in September 2014, the program received 131 innovative proposals and awarded 14 with funding and technical expertise to enable big data analytics in their projects. One of the winners of the challenge looked at nighttime lights satellite images to improve the monitoring of electricity provision to rural areas across villages in India. Another winner explored the use of indicators derived from high-resolution satellite data—such as identification of built-up areas, building and car density, roof type, and road type—to predict geographic variations in poverty in Pakistan and Sri Lanka.

The Food and Agriculture Organization (FAO) of the United Nations employed the use of remote sensing data, which enables the observation of areas with limited and difficult accessibility, in mapping coastal aquaculture and fisheries structures in the Philippines to provide relevant baseline data for the planning and development of these structures. The FAO's Land and Water Division also collaborated with Pakistan's Space and Upper Atmosphere Research Commission to develop models using satellite imagery to predict crop yield and to provide reliable agricultural information and statistics to the Government of Pakistan.

Acknowledging the opportunities for using big data to accelerate development outcomes and potentially close data gaps in monitoring sustainable development, UN has also been exploring linking various datasets (geospatial data, especially from satellite imagery, census and household survey data, and administrative records), focusing on the need for granular data on multidimensional data on poverty, population dynamics (including movements, urbanization, and migratory status), and disaster risk reduction. Furthermore, there are several tools (e.g., The Partnership in Statistics for Development in the 21st Century's Advance and Data Planning Tool [ADAPT], UNESCAP's Every Policy Is Connected to People, Planet, and Prosperity [EPIC] tool) which can be utilized by NSOs to assist in providing information to monitor and achieve the SDGs and to tailor fit their data production to the needs of policy makers.

Efforts on Big Data in Asia and the Pacific

Several countries in Asia and the Pacific have already started using big data in monitoring socioeconomic indicators and improving delivery of public goods and services. The UN Global Working Group on Big Data developed and is maintaining an inventory of big data projects. The following table provides an overview of some of these initiatives in the region.

A big data-complemented production of official statistics has been one of the more popular big data projects among the countries in Asia and the Pacific. Social media content such as the aforementioned Twitter conversations used in Jakarta is not the only useful innovative data source for monitoring price movements. Scanner data from supermarket chains and other retailers as well as online prices obtained from web scraping are now being used to generate price indices in the People's Republic of China, Japan, the Republic of Korea, and Malaysia. Access to mobile phone records and satellite imagery is now facilitating more efficient and accurate population mapping and population movement analysis. For instance, the integration of high-resolution satellite imagery and GPS helped in delineating census enumeration area maps for the Dili metropolis in Timor-Leste (Taiwo 2004, as cited in UNSD 2004). Mobile phone data that are complemented with secondary data (such as land use and transportation networks) and primary data (from surveys) can be used to yield information on population movement with high granularity and high frequency, as implemented in Bangladesh and soon in Sri Lanka.

The appreciation for the value of incorporating big data in the work programs of data producing agencies is becoming more apparent in some developing countries. The National Statistical Office of Mongolia (NSO Mongolia) has developed a geospatial statistical framework that enables the tracking of highly mobile units of enumeration (i.e., the herders) prior to the conduct of censuses and surveys (Chimeddamba 2017). The same framework, which makes use of satellite imagery, has also been used in the conduct of the NSO Mongolia's by-Census of Agriculture to aid in the identification of crop types and estimation of production. The Statistics Big Data Analytics project of the Department of Statistics Malaysia works

Table: Select Big Data-Related Initiatives in Asia and the Pacific

Economy	Institute or Department	Big Data Project
Australia	UN Global Pulse	Estimating migration flows using online search data
Bangladesh	World Bank Group	Predicting vulnerability to flooding and enhancing resilience using big data
China, People's Republic of	National Bureau of Statistics	Using web scraping price data for the price index of e-commerce
		Crop survey by farmland: using satellite and aerial remote sensing to help estimate agricultural statistics
		Comparison of data of interbank transactions with retail sales: credit card data for use in verifying retail sales
		Application of big data for highway and waterway transport statistics
		Online price changes of means of production
		Big data enterprise statistical indicator
	World Bank Group	Using big data analytics to discover patterns of medical insurance utilization for medical cost monitoring in the People's Republic of China
	UNDP and Baidu	Using big data to support e-waste management in the People's Republic of China
Japan	Ministry of Internal Affairs and Communications	Web scraping and scanner data for price statistics
Korea, Republic of	Statistics Korea of the Republic of Korea	Online price index
		Daily migration of population: using mobile call detail record data for daily migration data
India	World Bank Group	Tracking light from the sky version 2.0 or monitoring rural electrification from space
		Real-time forecasting of skills demand and supply: analytics of big data from BabaJob in India
	UN Global Pulse	Understanding immunization awareness and sentiment through analysis of social media and news content
	Ministry of Agriculture and Farmers Welfare and Indian Space Research Organization	Forecasting Agricultural output using space, agro-meteorology, and land-based information
Indonesia	World Bank Group	Big data for freight transport and logistics policy making
		Using mobile phone data for national, subnational, and geo-coded average prices
		Using big data to predict student achievement in low-income school settings
	UN Global Pulse	Understanding public perceptions of immunization using social media
		Mining citizen feedback data for enhanced local government decision making
	ILO and UN Global Pulse Lab Jakarta	Using social media to track workplace discrimination against women in Indonesia
Pakistan	World Bank Group	Using high-resolution satellite imagery and detection algorithms to better track poverty in Pakistan
Philippines	World Bank Group	OpenRoads Philippines: improved real-time decision making of infrastructure investments for the Philippines by linking geospatial road network data with rich geo-tagged social data collected through mobile phones
Singapore	Department of Statistics	Integrated environment system: using environmental sensing systems and data analytics for real-time environmental information
		Population estimates: using administrative data from many sources for population estimates

Table *continued*

Economy	Institute or Department	Big Data Project
Sri Lanka	World Bank Group LIRNEasia	Enabling up-to-date and accurate authoritative country mapping with crowdsourced geospatial data
		Potential of mobile network big data as a tool in Colombo's transportation and urban planning
Viet Nam	World Bank Group	Using big data to predict student achievement in low-income school settings

Notes: ILO = International Labour Organization, LIRNEasia = Learning Initiatives on Reforms for Network Economies Asia, UN = United Nations, UNDP = United Nations Development Programme.

Sources: United Nations Global Working Group on Big Data Project Inventory and United Nations Economic and Social Commission for Asia and the Pacific, as cited in ADB. 2016. *Key Indicators for Asia and the Pacific 2016.* Manila.

on constructing a big data infrastructure that will implement the following key project components: (i) integration of a business registry with a trade database for the identification of characteristics of enterprises engaged in the international market, (ii) adoption of webscraping techniques to improve the quality of the consumer price index, and (iii) assessment of public feedback on the quality of official statistics produced through social media.

Big data also continues to be a valuable resource in the strategic formulation and implementation of government projects and programs. In the Philippines, the Metro Manila Development Authority (MMDA) partnered with data science consultancy firm Thinking Machines to develop solutions in easing the worsening traffic situation in Metro Manila, by analyzing data from the GPS navigation software Waze. Disaster mitigation and response has vastly improved with the use of big data in tracking and planning activities. The Government of Indonesia worked with UN Global Pulse for the development of a crisis analysis tool, the Haze Gazer, which provides real-time information on fire and haze hotspots as well as the locations of vulnerable members of the population. The tool, which utilizes satellite data, baseline population information, and citizen-generated data published in social media and national complaint system LAPOR!, is expected to enhance the disaster management capacity of the government by enabling the formulation and implementation of well-informed response strategies. A slightly similar approach is also being undertaken in the People's Republic of China, where the Ministry of Environmental Protection has been providing a platform to compile different big data sources in tackling the country's severe air pollution problem. Satellite data, drones, and data from citizen reports are used for prompt prediction and measurement of air quality, and the resulting information is being used to trigger early warning systems for potentially severe smog incidences (Zhang and Hughes 2017).

Ways Forward with ADB's Knowledge and Support Technical Assistance Project

ADB is aware of the changing landscape in data and the need for granular data for the SDGs and has started a Data for Development

knowledge and support technical assistance project that aims to strengthen the capacity of NSOs in augmenting the limitations of conventional data sources with innovative data sources in official statistics production. These efforts are in support of the principle of the SDGs to leave no one behind, and its concomitant granular data requirements. This ADB project aims to keep tabs of relevant initiatives on using big data within Asia and the Pacific so that countries and their stakeholders can have a more nuanced understanding on the scalability of such initiatives.

READINESS OF THE OFFICIAL STATISTICS COMMUNITY FOR BIG DATA

Opportunities to use big data for monitoring sustainable development are growing as evinced by an increasing number of applications and case studies showcasing how big data can potentially enhance the monitoring of progress with respect to the SDGs (Table). However, there are several considerations that need to be taken into account before the official statistics community can fully capitalize on big data. The survey on SDG data compilation conducted by ADB and UNESCAP identifies some of these challenges:

Access to Big Data

Access to the big data sources, sets, and streams is the most cited challenge by 7 out of 16 NSOs in ADB member countries that participated in the 2017 ADB–UNESCAP survey. Even when big data are publicly available, these may only be a minute portion of the actual data. For instance, only a very small subsample of social media data is available publicly for free (while the entire data set has to be purchased for use), and there are questions on whether the actual sample made available for public use is representative (Fan and Bifet 2012). Digital traces can be incomplete and proprietorial rights over more comprehensive data make it difficult or expensive for NSSs to obtain access.

Technological Requirements

Retrieving and examining big data streams require adequate technological infrastructure, in terms of both hardware and software. Many current data mining tools are neither suitable nor efficiently used for large datasets using the conventional sequential computers that NSOs currently have. NSOs that intend to routinely use big data will thus need better ICT infrastructure and ample bandwidth to download these big data sources, as well as to catalog, organize, and process the complex collage of data in a sufficiently timely manner. A recent practice in big data analytics is utilizing a cluster of computers running a framework tool such as Hadoop-MapReduce, and/or cloud computing and processing.[4] The availability of interfaces by some statistical packages has, however, significantly contributed to the use of big data analytics. Further, the cloud has also emerged as an ideal computing environment for big data (Agrawal, et al. 2011). On the infrastructure side, cloud computing provides options for accessing and managing very large data sets as well as for supporting powerful infrastructure elements at a relatively low cost. Further, an increasing number of software held in a hybrid cloud are also capable of performing the processing and data integration tasks.

A related technological issue is the curation of big data. The big data sources result in a messy collage of data points. Various tools have to be used to assess the veracity of big data. Bias does not necessarily disappear in voluminous big data. While statistics using big data may not be completely accurate, they are often viewed as "good enough" and at near real-time. The gains in velocity (and cost) in yielding statistics from big data sources, as well as the complexity and the sheer size of big data, however, require different types of data processing and analytic tools from those used for "small data" to yield statistics that are fit for use.

Capacity

Analytics on big data require new skill sets. While NSOs have had experience in curating data from traditional data sources, they often have no data scientists who are strong in both data and computational focus. Analysis of big data can also be burdened with methodological challenges regarding their veracity. In January 2013, the Google Virus Trends estimate (11%) of flu levels in the United States was nearly double the official estimate (6%) from the CDC (Butler 2013). Whether or not digital traces can represent information on an entire population is crucial to the reliability of big data.

Credibility is fundamental in official statistics (Fellegi 1996). While the quality of statistics involves several criteria, the production of official statistics often focuses on precision and accuracy over timeliness and other quality features. Unlike traditional data sources of official statistics that are designed to produce precise and accurate statistics that estimate parameters of populations, many types of big data do not

have clear target populations. Big data sources are usually produced as a by-product in the course of some other activity (e.g., making a call on a mobile phone or taking a photograph and sharing it). Furthermore, while big data can enlighten, it can also obscure information, especially if the limitations of such data are poorly understood and if the data are examined inadequately with bias and with malice.

In the context of official statistics production, blending traditional and innovative data sources requires a new skill set for all NSO staff, from managers to methodologists to IT staff. There also needs to be paradigm shifts in statistical production among managers and leaders in NSOs and require them to develop the necessary soft skills in building partnerships in big data ecosystem.

Data Privacy

One major concern regarding big data use is related to data privacy, security, and related issues (UNECE 2013). Although various mechanisms to protect privacy are in place, including asking people to opt out of studying the information they give, and anonymization methods such as differential privacy and "space time boxes", these methods are not foolproof. That is, even if the data are anonymized, it is possible to reidentify the sources. While official statisticians have established protocols on data confidentiality which are often articulated in the statistics laws and/or data protection laws of various countries, this is not the case for big data, which are usually in the hands of firms in the private sector which adhere to different principles and rules on data confidentiality.

Ecosystem

Big data are not just about data sources, sets, or streams, but are also about a complex ecosystem (Letouzé 2012). Disruptive technologies have the effect of "disintermediating" producers of official statistics and developing country citizens and the private sector and other organizations (that traditionally supply data to NSSs), since emerging technologies empower citizens to collect and publish their own data.

Technological solutions of using big data in official statistics require strengthening institutions and developing proper skills, a process that requires building trust, which takes time, perseverance, as well as soft skills. NSOs must develop new business models to leverage data resources, human talent, and decision-making capacity. They should develop and enhance their institutional frameworks and arrangements, such as public-private partnerships and linkages with various institutions engaged in data science, to further promote official statistics as a public good, and ensure the constant improvement of the quality of the data, particularly regarding timeliness, disaggregation, and meaningfulness. Countries will require guidance such as statistical standards and knowledge materials. The knowledge materials should

---

4   Hadoop is an open-source software project managed by the Apache Software Foundation targeted at supporting the execution of data-oriented application on clusters of generic hardware. Hadoop MapReduce provides a framework for writing applications which process big data in an efficient manner.

present not only what works but what does not. NSOs will also require legal protocols to access big data holdings for development purposes (without infringing on data privacy), as well as to prevent misuse of big data.

SUMMARY

Recognizing the many barriers and bottlenecks in meeting the data demands for the SDGs, most NSOs in ADB member countries that participated in the 2017 ADB–UNESCAP Survey consider big data as a promising means of addressing data gaps for SDGs. The use of big data can supplement data from traditional data sources and can serve as an additional source for official statistics. Further, the costs associated with traditional data collection activities, and the increasing levels of nonresponses due to the burden associated with primary data collection-even if conducted with advanced modalities such as web surveys-lead to potential losses in data quality. Although a few NSOs in the 2017 ADB–UNESCAP Survey reported having access to aerial photos and satellite imagery, mobile data, web-scraped online price data, and social media data, only a limited number of countries mentioned having any ongoing big data projects. Those that have big data projects are currently working on using satellite imagery, geo-spatial data, and social media data to improve the granularity of statistics on poverty and welfare.

In summary, research and case studies on big data and its applications for SDG monitoring are needed to provide lessons on integrating conventional and innovative data sources. Further, it is essential that these lessons be turned into practical guidelines for NSS on the use of big data before they fully integrate data sources.

REFERENCES

Agrawal, D., S. Das, and A. El Abbadi. 2011. Big Data and Cloud Computing: Current State and Future Opportunities. Proceedings of the 14th International Conference on Extending Database Technology. ACM. https://openproceedings.org/2011/conf/edbt/AgrawalDA11.pdf

Bussay, A. and I. H. Akhtar. 2009. Monitoring of Crops through Satellite Technology. FAO. https://www.researchgate.net/publication/287208777_Monitoring_of_Crops_through_Satellite_Technology.

Asian Development Bank and United Nations Economic and Social Commission for Asia and the Pacific. 2017. Survey on SDG Data Compilation.

Butler, D. 2013. When Google Got Flu Wrong. *Nature*. https://www.nature.com/news/when-google-got-flu-wrong-1.12413.

Chimeddamba, L. 2017. Mongolia's Experience on Integrating Geospatial Information with Administrative Records to Produce Official Statistics. Presented in the Workshop on Integrating Nontraditional Data Sources in the Production of the SDG Indicators. 28–30 November. https://unstats.un.org/unsd/capacity-building/meetings/da10-workshop-2017/4b.2-Mongolisa-EN.pdf.

Fan, W. and A. Bifet. 2012. Mining Big Data: Current Status and Forecast to the Future. http://kdd.org/exploration_files/V14-02-01-Fan.pdf.

Fellegi, I. 1996. Characteristics of an Effective Statistical System. *International Statistical Review*. 64(2): pp. 165-197.

Ginsburg, J. et al. 2009. Detecting Influenza Epidemics Using Search Engine Query Data. *Nature*. 457. pp. 1012–1014. https://www.nature.com/articles/nature07634.

Hilbert, M. 2013. Big Data for Development: From Information- to Knowledge Societies. Available at SSRN: https://ssrn.com/abstract=2205145 or http://dx.doi.org/10.2139/ssrn.2205145.

Independent Expert Advisory Group on a Data Revolution for Sustainable Development. 2014. A World That Counts—Mobilising the Data Revolution for Sustainable Development. http://www.undatarevolution.org/wp-content/uploads/2014/11/A-World-That-Counts.pdf

Jean, N. et al. 2016. Combining Satellite Imagery and Machine Learning to Predict Poverty. *Science*. 353, (6301). pp. 790–794. http://science.sciencemag.org/content/353/6301/790.

Krambeck, H. et al. 2015. Open Traffic Big Data Challenge Project Completion Report No: ACS15491. The World Bank. http://pubdocs.worldbank.org/en/513661445369530688/Open-Traffic-Completion-Report-1.pdf.

Letouzé, E. 2012. *Big Data for Development: Challenges and Opportunities*. UN Global Pulse. http://www.unglobalpulse.org/sites/default/files/BigDataforDevelopment-UNGlobalPulseJune2012.pdf.

Liang, G. et al. 2014. Quantifying Information Flow During Emergencies. *Scientific Reports*. 4 (3997). https://www.nature.com/articles/srep03997.pdf.

Pulse Lab Jakarta. 2017. Using Big Data for Statistics to Track the SDGs. 27 October. https://medium.com/pulse-lab-jakarta/tracking-the-sdgs-using-big-data-dad0ad351f2e.

Secretariat of the Partnership in Statistics for Development in the 21st Century (PARIS21). 2017.

Partner Report on Support to Statistics (PRESS) 2017. http://www.paris21.org/sites/default/files/2017-10/PRESS2017_web2.pdf.

Smith C., A. Mashadi, and L. Capra. 2013. Ubiquitous Sensing for Mapping Poverty in Developing Countries. Proceedings of the Third Conference on the Analysis of Mobile Phone Datasets. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.408.9095&rep=rep1&type=pdf.

Travaglia, C. et al. 2004. Mapping Coastal Aquaculture and Fisheries Structures by Satellite Imaging Radar: Case Study of the Lingayen Gulf, the Philippines. *FAO Fisheries Technical Paper 459*. FAO: Rome. http://www.fao.org/docrep/007/y5319e/y5319e00.htm.

United Nations (UN). 2015. A/RES/70/1—Transforming Our World: The 2030 Agenda for Sustainable Development. Resolution adopted by the General Assembly on 25 September 2015. https://undocs.org/A/RES/71/313.

United Nations Economic Commission for Europe (UNECE). 2013. What Does "Big Data" Mean for Official Statistics? https://statswiki.unece.org/pages/viewpage.action?pageId=77170614.

United Nations Statistics Division (UNSD). 2004. ESA/STAT/AC.98/14—Integration of GPS, Digital Imagery and GIS with Census Mapping. United Nations Expert Group Meeting to Review Critical Issues Relevant to the Planning of the 2010 Round of Population and Housing Censuses. 15–17 September. https://unstats.un.org/unsd/demographic/meetings/egm/CensusEGM04/docs/AC98_14.pdf.

White, T. 2012. *Hadoop: The Definitive Guide*. California: O'Reilly Media.

World Bank. 2016. *Big Data Innovation Challenge: Pioneering Approaches to Data-Driven Development*. World Bank: Washington, DC. https://openknowledge.worldbank.org/handle/10986/25102.

Zhang, B. and R. Hughes. 2017. China Deploys Big Data to Clear Smog. *Nature*. 542 (31).