

Partnership

To mobilize the means required to implement this agenda through a revitalized Global Partnership for Sustainable Development, based on a spirit of strengthened global solidarity, focused in particular on the needs of the poorest and most vulnerable and with the participation of all countries, all stakeholders, and all people.



Snapshots

- The majority of the member economies experienced higher volumes of remittances in proportion to total GDP over the past 15 years, while 27 economies had an increase that exceeds 0.05 percentage points per year.
- Debt service as a proportion of exports of goods and services declined between 2000 and 2015 in 23 economies including India, Kazakhstan, the Kyrgyz Republic, and Papua New Guinea.
- Within Asia and the Pacific, the highest net official development assistance in 2014 were provided to Afghanistan, Viet Nam, Pakistan, India, and Bangladesh.
- In 19 out of 35 economies for which data are available in Asia and the Pacific, national statistical plans are fully funded and under implementation.

Attaining the SDGs by 2030 requires a strong commitment to global partnership and cooperation among all players. For low-income economies, official development assistance (ODA) will continue to be a major resource given their limited capacities to raise public resources domestically. Furthermore, ODA could be catalytic in crowding in other sources and building capacities. This section presents available data on ODA and other indicators of SDG 17 for ADB regional member economies.

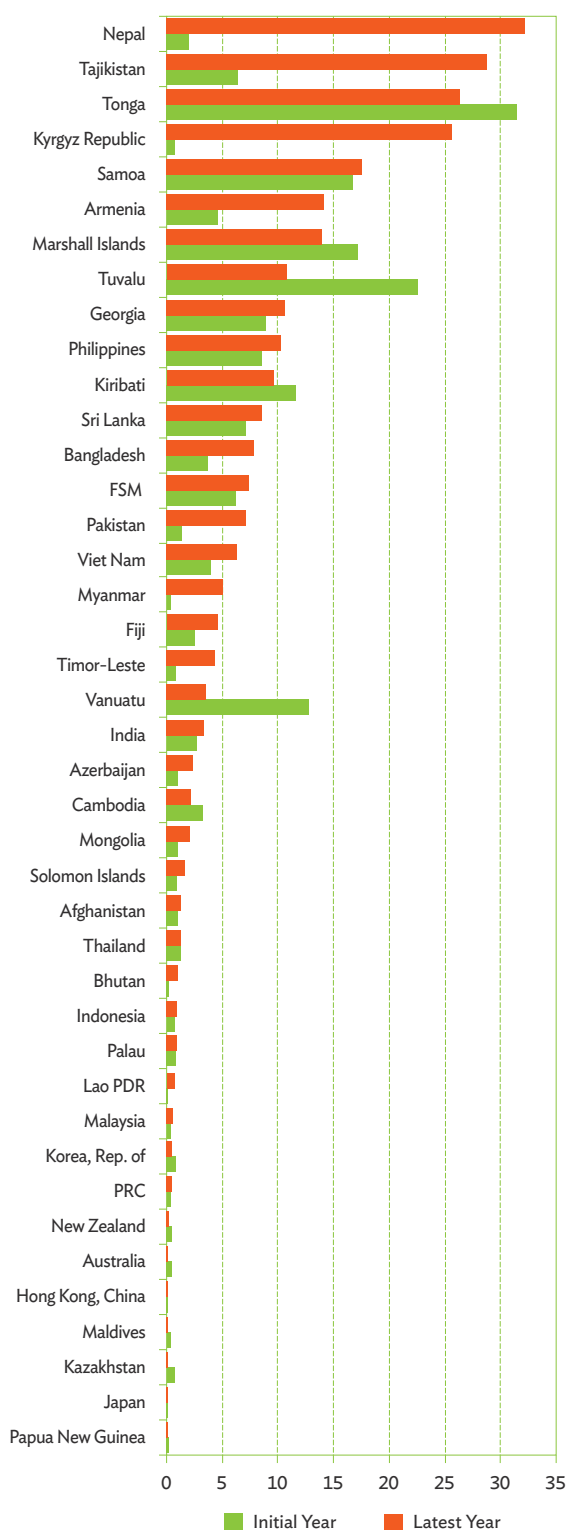
SDG 17: Strengthen the Means of Implementation and Revitalize the Global Partnership for Sustainable Development

Volume of remittances as a proportion of total GDP. Latest data available show that the volume of remittances as a proportion of total GDP is highest in Nepal (32.2%), Tajikistan (28.8%), Tonga (26.3%), the Kyrgyz Republic (25.7%), and Samoa (17.6%) (Figure 6.1). The largest increase within the 15-year period (anytime between 2000 and 2015) was noted

in Nepal (2.0 percentage points per year), followed by Tajikistan (1.7 percentage points per year), and the Kyrgyz Republic (1.7 percentage points per year). On the other hand, the volume of remittances in proportion to GDP declined in Australia, Cambodia, Kazakhstan, Kiribati, the Republic of Korea, the Marshall Islands, New Zealand, Papua New Guinea, Thailand, Tonga, Tuvalu, and Vanuatu. Among the top 10 economies with data in 2014 or 2015, three economies registered a reduction in the volume of remittances: Tuvalu by 1.3 percentage points per year, the Marshall Islands by 0.36 percentage points per year, and Tonga by 0.4 percentage points per year.

Debt service as a proportion of exports of goods and services. Figure 6.2 shows the distribution of debt service relative to exports of goods and services. In the majority of the economies with available data, the numbers show that the proportion of debt service declined in the past 15 years. The annual reduction exceeded 0.5 percentage points in India, Kazakhstan, the Kyrgyz Republic, and Papua New Guinea. On the

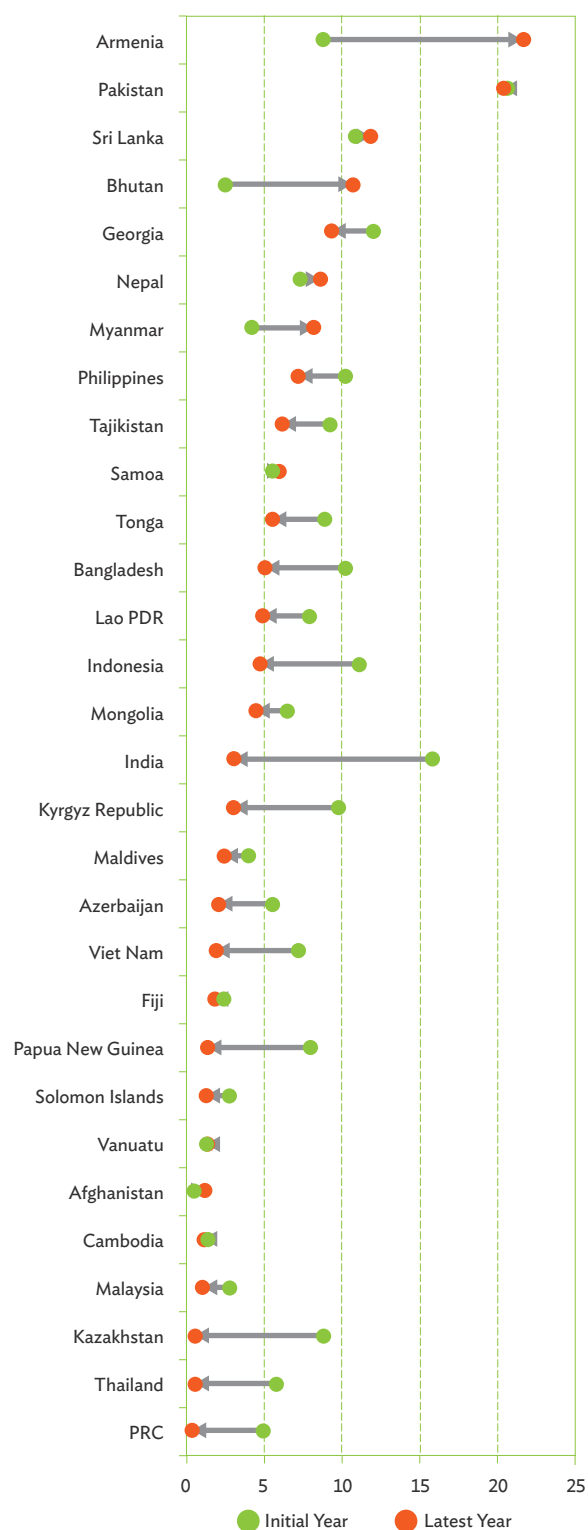
Figure 6.1: Volume of Remittances as a Proportion of Total Gross Domestic Product (%)



FSM = Federated States of Micronesia, Lao PDR = Lao People's Democratic Republic, PRC = People's Republic of China.
Source: Table 6.1.

[Click here for figure data](#)

Figure 6.2: Debt Service (% of exports of goods and services)



Lao PDR = Lao People's Democratic Republic, PRC = People's Republic of China.
Source: Table 6.1.

[Click here for figure data](#)

other hand, it increased by more than 0.5 percentage points per year in Armenia and Bhutan. The five economies with highest proportion of debt service as a proportion of exports are Armenia (21.6%), Pakistan (20.3%), Sri Lanka (11.9%), Bhutan (10.7%), and Georgia (9.4%).

Net official development assistance. Within Asia and the Pacific, the highest net official development assistance in 2014 were provided to Afghanistan, Viet Nam, Pakistan, India and Bangladesh.³

Availability of National Statistical Plan. National statistical plans provide a strategy for an integrated development of a national statistical system. In about half of the 35 economies for which data are available from UNSD's SDG Indicators Global Database for 2015, national statistical plans were fully funded and under implementation. For a few economies, no such plan existed in 2015 while for others, national statistical plans have either expired or were awaiting adoption. Table 6.3 summarizes the availability of national statistical plans in economies of Asia and the Pacific based on latest data available.

Equity and Other Issues

While ODA flows from developed to developing countries comprised the majority of the development assistance flows during the MDG era, ODA flows between developing countries are likely to increase in the coming years. This type of ODA flow could play a key role in crafting better public policies for social inclusion that are contextualized to developing economies. Nevertheless, the sustainable development agenda still needs to be financed from a more innovative and diverse range of sources combining public, private, and joint financing that raise funds both internally and externally.

Data Gaps

Indicators for various targets under the theme of partnership are not available; when they are, they are sparse and not regularly updated. Difficulty in monitoring progress with respect to SDG 17 may also arise due to the lack of quantitative targets in some areas.

³ Details are provided in Regional Table 4.16.

Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

Table 6.1: **Selected Indicators for SDG 17 - Development Financing**
Mobilize additional financial resources for developing countries from multiple sources

Assist developing countries in attaining long-term debt sustainability through coordinated policies aimed at fostering debt financing, debt relief and debt restructuring, as appropriate, and address the external debt of highly indebted poor countries to reduce debt distress

Regional Member	17.3.2 Volume of Remittances in US Dollars (as a proportion of total GDP)		17.4.1 Debt Service (as a proportion of exports of goods and services)	
	2000	Latest Year	2000	2013
Developing Member Economies				
Central and West Asia				
Afghanistan	1.0 (2008)	1.3 (2014)	0.5 (2005)	1.2
Armenia	4.6	14.1 (2015)	8.8	21.6
Azerbaijan	1.1	2.4 (2015)	5.5	2.1
Georgia	9.0	10.6 (2015)	12.1	9.4
Kazakhstan	0.7	0.1 (2014)	8.8	0.6
Kyrgyz Republic	0.7	25.7 (2015)	9.8	3.0
Pakistan	1.5	7.2 (2015)	20.8	20.3
Tajikistan	6.4 (2002)	28.8 (2015)	9.2 (2002)	6.2 (2012)
Turkmenistan
Uzbekistan
East Asia				
China, People's Rep. of	0.4	0.4 (2015)	4.9	0.3
Hong Kong, China	0.1	0.1 (2015)
Korea, Rep. of	0.9	0.5 (2015)
Mongolia	1.1	2.1 (2014)	6.5	4.5
Taipei, China
South Asia				
Bangladesh	3.7	7.9 (2015)	10.2	5.1
Bhutan	0.3 (2006)	1.0 (2015)	2.5 (2006)	10.7
India	2.7	3.4 (2014)	15.8	3.1
Maldives	0.4	0.1 (2015)	4.0	2.5
Nepal	2.0	32.2 (2015)	7.3	8.6
Sri Lanka	7.1	8.5 (2015)	10.9	11.9
Southeast Asia				
Brunei Darussalam
Cambodia	3.3	2.2 (2014)	1.4	1.1
Indonesia	0.7	1.0 (2014)	11.1	4.7
Lao PDR	0.0	0.8 (2015)	7.9	4.9
Malaysia	0.4	0.6 (2015)	2.8	1.1
Myanmar	0.4 (2012)	5.0 (2015)	4.2	8.2
Philippines	8.6	10.3 (2015)	10.2	7.2
Singapore
Thailand	1.3	1.3 (2015)	5.8	0.5
Viet Nam	4.0	6.3 (2011)	7.2	1.9
The Pacific				
Cook Islands
Fiji	2.6	4.6 (2014)	2.4	1.8
Kiribati	11.6 (2006)	9.6 (2014)
Marshall Islands	17.2 (2005)	14.0 (2014)
Micronesia, Fed. States of	6.3 (2009)	7.3 (2014)
Nauru
Palau	0.8 (2005)	0.9 (2014)
Papua New Guinea	0.2	0.1 (2014)	8.0	1.4 (2012)
Samoa	16.7	17.6 (2014)	5.5 (2004)	6.1
Solomon Islands	1.0	1.6 (2015)	2.8	1.3
Timor-Leste	0.8 (2006)	4.4 (2015)
Tonga	31.5 (2001)	26.3 (2014)	8.9 (2001)	5.6 (2012)
Tuvalu	22.6 (2005)	10.7 (2014)
Vanuatu	12.7	3.5 (2014)	1.4	1.2
Developed Member Economies				
Australia	0.5	0.2 (2015)
Japan	0.0	0.1 (2015)
New Zealand	0.5	0.2 (2015)

... = data not available at cutoff date, 0.0 = magnitude is less than half of unit employed, Lao PDR = Lao People's Democratic Republic, SDG = Sustainable Development Goal.

Sources: United Nations. Sustainable Development Goals Indicators Global Database. <http://unstats.un.org/sdgs/indicators/database/> (accessed 21 July 2016); World Bank. World Development Indicators. <http://databank.worldbank.org/data/reports.aspx?source=world-development-indicators> (accessed 3 September 2016).

Table 6.2: **Selected Indicators for SDG 17 - Access to Technology and Data Communications**
Fully operationalize the technology bank and science, technology and innovation capacity-building mechanism for least developed countries by 2017 and enhance the use of enabling technology, in particular information and communications technology

Regional Member	17.6.2 Fixed Internet Broadband Subscriptions per 1,000 Inhabitants ^a	
	Initial Year	2015
Developing Member Economies		
Central and West Asia		
Afghanistan	0.01 (2004)	0.05
Armenia	0.00 (2001)	95.78
Azerbaijan	0.12 (2002)	197.60
Georgia	0.09 (2001)	146.35
Kazakhstan	0.07 (2003)	130.49
Kyrgyz Republic	0.01 (2002)	37.06
Pakistan	0.09 (2005)	9.53
Tajikistan	0.00 (2003)	0.70
Turkmenistan	0.02 (2008)	0.56
Uzbekistan	0.11 (2003)	35.66
East Asia		
China, People's Rep. of	0.02 (2000)	185.61
Hong Kong, China	65.02 (2000)	319.36
Korea, Rep. of	84.17 (2000)	402.50
Mongolia	0.02 (2001)	71.17
Taipei, China	10.44 (2000)	242.60
South Asia		
Bangladesh	0.30 (2007)	24.10
Bhutan	2.98 (2008)	35.55
India	0.05 (2001)	13.35
Maldives	0.67 (2002)	64.74
Nepal	0.04 (2006)	10.64
Sri Lanka	0.02 (2001)	31.00
Southeast Asia		
Brunei Darussalam	5.59 (2001)	79.95
Cambodia	0.00 (2002)	5.33
Indonesia	0.02 (2000)	10.89
Lao PDR	0.00 (2003)	5.19
Malaysia	0.17 (2001)	89.50
Myanmar	0.00 (2005)	3.50
Philippines	0.13 (2001)	33.99
Singapore	17.61 (2000)	264.50
Thailand	0.03 (2001)	92.42
Viet Nam	0.01 (2002)	81.38
The Pacific		
Cook Islands	- (2001)	130.90 (2013)
Fiji	8.51 (2005)	14.26
Kiribati	3.56 (2005)	1.11
Marshall Islands	24.39 (2013)	18.87
Micronesia, Fed. States of	0.06 (2003)	31.37
Nauru	...	94.76 (2010)
Palau	3.74 (2004)	57.49
Papua New Guinea	0.46 (2008)	1.97
Samoa	0.18 (2004)	11.00
Solomon Islands	0.44 (2004)	2.43
Timor-Leste	0.01 (2003)	0.88
Tonga	0.11 (2002)	18.89
Tuvalu	5.18 (2004)	100.85
Vanuatu	0.08 (2003)	16.28
Developed Member Economies		
Australia	6.30 (2001)	278.52
Japan	6.80 (2000)	304.87
New Zealand	1.21 (2000)	315.46

... = data not available at cutoff date, 0.00 = magnitude is less than half of unit employed, - = magnitude equals zero, Lao PDR = Lao People's Democratic Republic, SDG = Sustainable Development Goal.

a The original indicator refers to "Fixed Internet Broadband Subscriptions per 100 Inhabitants".

Sources: United Nations. Sustainable Development Goals Indicators Global Database. <http://unstats.un.org/sdgs/indicators/database/> (accessed 21 July 2016); International Telecommunication Union. World Telecommunication/ICT Indicators Database. <http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx> (accessed 6 June 2016).

Strengthen the means of implementation and revitalize the Global Partnership for Sustainable Development

Table 6.3: **Selected Indicators for SDG 17 - Availability of National Statistical Plan**

Enhance capacity-building support to developing countries, including for least developed countries and small island developing States, to increase significantly the availability of high-quality, timely and reliable data disaggregated by income, gender, age, race, ethnicity, migratory status, disability, geographic location, and other characteristics relevant in national contexts

Regional Member	17.18.3 Availability of National Statistical Plan
Developing Member Economies	
Central and West Asia	
Afghanistan	A (2015)
Armenia	A (2015)
Azerbaijan	A (2011)
Georgia	D (2015)
Kazakhstan	...
Kyrgyz Republic	D (2015)
Pakistan	B (2015)
Tajikistan	A (2015)
Turkmenistan	D (2015)
Uzbekistan	E (2015)
East Asia	
China, People's Rep. of	D (2012)
Hong Kong, China	...
Korea, Rep. of	...
Mongolia	A (2015)
Taipei, China	...
South Asia	
Bangladesh	A (2015)
Bhutan	D (2015)
India	A (2015)
Maldives	A (2015)
Nepal	A (2015)
Sri Lanka	A (2015)
Southeast Asia	
Brunei Darussalam	...
Cambodia	A (2015)
Indonesia	D (2015)
Lao PDR	A (2015)
Malaysia	...
Myanmar	E (2015)
Philippines	A (2015)
Singapore	...
Thailand	A (2012)
Viet Nam	A (2015)
The Pacific	
Cook Islands	...
Fiji	E (2015)
Kiribati	A (2015)
Marshall Islands	D (2015)
Micronesia, Fed. States of	D (2015)
Nauru	...
Palau	...
Papua New Guinea	E (2015)
Samoa	A (2015)
Solomon Islands	E (2015)
Timor-Leste	A (2015)
Tonga	D (2015)
Tuvalu	E (2015)
Vanuatu	A (2015)
Developed Member Economies	
Australia	...
Japan	...
New Zealand	...

... = data not available at cutoff date, Lao PDR = Lao People's Democratic Republic, SDG = Sustainable Development Goal.

- A National statistical plan is fully funded and under implementation
- B National statistical plans are completed and awaiting adoption
- C National statistical plans are expired or without a plan and are currently designing or planning
- D National statistical plans are expired
- E National statistical plans do not exist

Sources: United Nations. Sustainable Development Goals Indicators Global Database. <http://unstats.un.org/sdgs/indicators/database/> (accessed 22 October 2016).



What Is Big Data?

Big data generally refer to the type of data arising from people's digital transactions with computers, social media, mobile phones, photos, satellite images, sensors, and other types of digital technology. There are three main sources of big data: human-sourced information, process-mediated data, and machine-generated data. The human-sourced information includes data coming from social networks, personal documents, search engines, videos, mobile data content, user-generated maps, and e-mail, among others. Process-mediated data are those coming from traditional business systems, e.g., produced by public agencies (including medical records); and those that are produced by business (commercial transactions, banking records, e-commerce, credit cards, and loyalty cards). Machine-generated data may include fixed sensors (home, weather, traffic, scientific, security, and surveillance) and mobile sensors (mobile phone location, data from computer systems like logs and weblogs). In all of these contexts, big data are characterized in terms of five Vs: volume, velocity, variety, veracity, and variability.

Compilation of big data has been growing at a very fast pace; in fact, it is bounded by the storage capacities of various entities that collect the data. However, with the continuous development of new information technology and the dramatic increase of devices at the periphery of the network including embedded sensors, smartphones, and tablet computers, data extraction and storage capacity are becoming less of a constraint in the compilation of big data (Villars, Eastwood, and Olofson 2011). Similarly, big data analytics is flourishing. In fact, the existing literature offers a wide array of analytical tools such as regularized regression, model selection and validation, classification, and dimension reduction that can be used for examining big data (Wu and Kumar 2009).

The term “Big Data,” on the other hand, refers to a wider ecosystem that includes various actors who play different roles in the generation, storage, retrieval, analytics, and usage of big data. Although the private sector has been the major user of big data, big data are expected to play a more significant role in complementing the traditional data sources for official statistics in the coming years. Hence, it is important for all players to understand the potential benefits as well as the constraints in using these new data sources. The following section discusses this issue.

Big Data and Official Statistics

There are several specific examples that showcase how big data can be useful for compiling official statistics. For instance, Statistics Netherlands uses location data through mobile phones to generate proxy measures for daytime population and tourism statistics. In addition, they use data from social media messages to develop a proxy indicator of consumer confidence. They also calculate inflation based on price information extracted from the web.⁴ Furthermore, Statistics Netherlands uses traffic loop detection data in measuring the volume of vehicles and traffic density.⁵ On the other hand, the Australian Bureau of Statistics has been examining the possible applications of big data in the development of sampling frames or registers, full or partial data substitution, imputation of missing data items, and data validation (Tam and Clarke 2015). There are various activities using big data and official statistics in Asia and the Pacific as well. For instance, web scrapping data is used in the development of price statistics (including price indexes) in the People's Republic of China (PRC), Japan, and the Republic of Korea. On the other hand, crowd-sourced data are used in decisions on infrastructure investments in the Philippines and in managing urban growth in Sri Lanka. Call detail records (CDRs) from mobile phones are used in monitoring daily migration in the Republic of Korea. CDR data are

⁴ For details, please see Struijs, Braaksma, and Daas (2014).

⁵ For details, please see Daas et al. (2015).

Table 7.1: Data List of 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 Rep. of	National Bureau of Statistics	Using web scraping price data for 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, Rep. of	Statistics 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
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 (IES): using environmental sensing systems and data analytics for real-time environmental information
		Population estimates: using administrative data from many sources for population estimates
Sri Lanka	World Bank Group	Enabling up-to-date and accurate authoritative country mapping with crowdsourced geospatial data
	LIRNEasia	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

ILO = International Labour Organization, UN = United Nations, UNDP = United Nations Development Programme.

Sources: United Nations Global Working Group on Big Data Project Inventory; United Nations Economic and Social Commission for Asia and the Pacific.

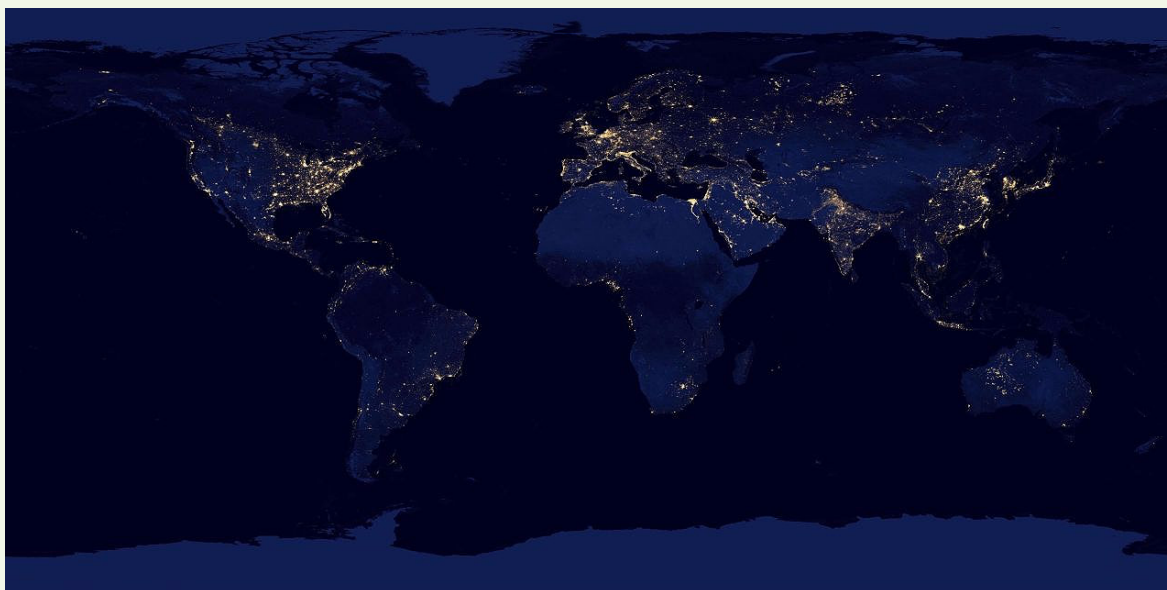
also used in price monitoring and in freight and logistics decision-making in Indonesia. Satellite imagery and remote sensing images are used in crop estimates in the PRC, in poverty tracking in Pakistan, in monitoring rural electrification in India, and monitoring pollution levels in Singapore. Box 7.1 provides another example of how data from satellite images, particularly nighttime lights, can be used to develop proxy measures for different types of social and economic indicators. In summary, there are numerous ongoing initiatives that

explore the potential uses, limitations, and constraints in using big data in the generation of official statistics. To further advance these initiatives and explore the other frontiers of big data, the United Nations (UN) created a Global Working Group (GWG) on Big Data for Official Statistics in collaboration between the World Bank and the UN Statistics Division. Currently, the GWG maintains a detailed catalog of big data-related projects that are relevant for different types of official statistics through its GWG Big Data Inventory.

Box 7.1: Using Nighttime Lights to Measure Social and Economic Indicators

Data on nighttime lights is a good example of a novel source of information that is increasingly being used in ongoing studies that showcase the application of big data for monitoring the Sustainable Development Goals (SDGs).

Box Figure 7.1: Nighttime Lights of the World



Source: National Aeronautics and Space Administration.

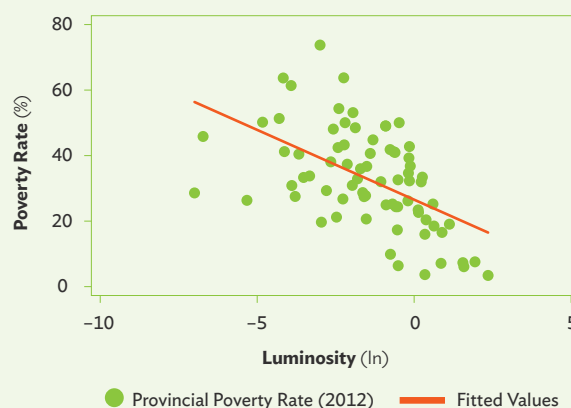
What are Nighttime Lights?

The Defense Meteorological Satellite Program Operational Linescan System of the United States National Oceanic and Atmospheric Administration collects satellite images of nighttime lights of every corner of the world within the 65° south and north latitude between 8:30 p.m. and 10 p.m. local time. Each pixel of an image represents a square kilometer of ground area, while a digital number ranging from 0 to 63 is used to gauge the intensity of the lights. The raw data are reprocessed to remove the noise caused by cloud cover, snow, and ephemeral lights caused by fire. Compiled annual data are available through the National Geophysical Data Center's website from <http://ngdc.noaa.gov/eog/index.html>.

Deriving Proxy Measures of Socioeconomic Indicators Using Data on Nighttime Lights

In principle, nighttime light is an important input in many economic production and consumption activities such as transportation of goods and people, outdoor lighting, illumination of houses and buildings, and consumption of mass media (Pinkovskiy and Sala-i-Martin 2015). Hence, it is not surprising to note that several studies find that nighttime lights or luminosity data correlate well with measures of economic activity such as gross domestic product (GDP) as well as other non-GDP-based socioeconomic indicators of population size, employment, and poverty (e.g., Chen and Nordhaus 2010, 2011; Gosh et al. 2010; Pinkovskiy and Sala-i-Martin 2015).

Box Figure 7.2: Correlation between Provincial Poverty Rates and Nighttime Lights Index Values



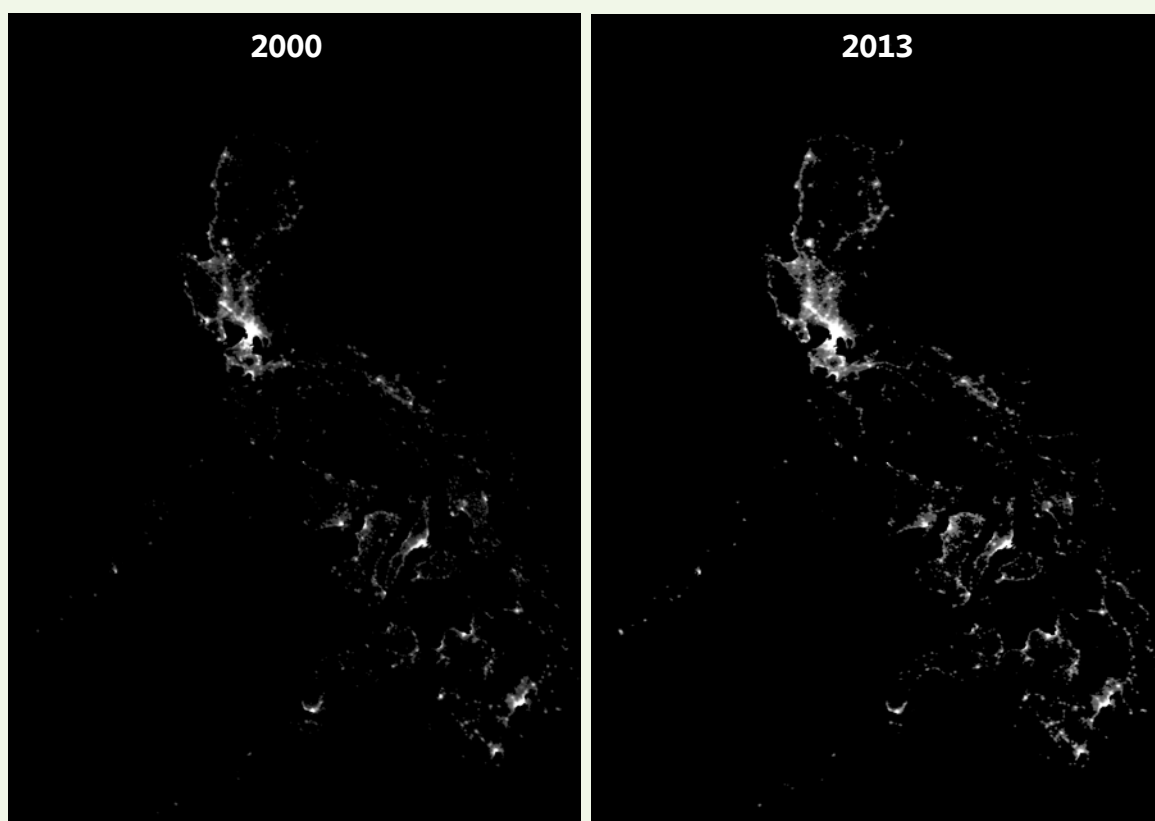
Source: ADB estimates based on poverty numbers compiled by the Philippine Statistics Authority and nighttime lights data.

continued.

Box 7.1: (continued)

Access to nighttime lights data allows countries with weak statistical systems to impute the indicators identified above at regular time intervals. In addition, data on nighttime lights are also potentially useful for spatial analysis, as they allow for estimation at disaggregated levels since each pixel represents a sufficiently small area. For example, Lo (2001) concluded that nighttime light data produced reasonably accurate estimates of urban population at the provincial, country, and city levels in the People's Republic of China. Nighttime luminosity data also serve as a validation tool when socioeconomic indicators that are supposed to correlate well with each other manifest inconsistent trends. For example, Pinkovskiy and Sala-i-Martin (2015) argue that nighttime luminosity data are an effective validation tool when survey-based estimates of income conflict with GDP-based measures. According to the authors, “the strength of the correlation between nighttime lights and measured income is directly related to the strength of the correlation between the given income measurement and the true income it is trying to measure.”

Box Figure 7.3: Images of Nighttime Lights in the Philippines, 2000 and 2013



Source: ADB calculations based on data downloaded from National Oceanic and Atmospheric Administration. Version 4 DMSP-OLS Nighttime Lights Time Series. <http://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html> (accessed 18 January 2016).

In an ongoing study, staff of the Economic Research and Regional Cooperation Department of the Asian Development Bank are exploring the feasibility of using nighttime lights data to measure the impact of infrastructure projects. To accomplish this, the authors are analyzing the increase in the illumination values within a 5-, 10-, and 15-kilometer radius surrounding the areas of a randomly selected set of road projects before, during, and after project implementation. Preliminary results suggest that the increase in the luminosity of areas with road projects was significantly higher than the observed increase in the illumination values of a preselected control group consisting of “similar” areas that did not have a road project during the same reference time period. Measures of GDP, poverty, and unemployment can then be estimated before and after project implementation by using the correlation between these measures and the illumination values.

continued.

Box 7.1: (continued)

Despite the advantages of using nighttime light data to derive proxy measures of conventional socioeconomic indicators, this approach is not without limitations. For example, since old satellites are replaced by new ones, there might be inconsistent readings from year to year. In addition, since the distribution of illumination values is right-censored, it is not possible to estimate economic growth or temporal changes in other socioeconomic measures for an area that has already reached the maximum digital number value of 63. Nevertheless, examining the relationship between nighttime lights and socioeconomic indicators serves as a good building block for policy in terms of using novel data sources to complement traditional sources that are used to compile official statistics.

Sources:

- X. Chen and W. Nordhaus. 2010. The Value of Luminosity Data as a Proxy for Economic Statistics. *NBER Working Paper* 16317. Cambridge MA: National Bureau of Economic Research (NBER).
- . 2011. Using Luminosity Data as a Proxy for Economic Statistics. *Proceedings of the National Academy of Sciences of the United States of America* 108(21): 8589–8594.
- T. Ghosh, R. L. Powell, C. D. Elvidge, K. E. Baugh, P. C. Sutton, and S. Anderson. 2010. Shedding Light on the Global Distribution of Economic Activity. *The Open Geography Journal* 3: 148–161.
- C. P. Lo. 2001. Modeling the Population of China Using DMSP Operational Linescan System Nighttime Data. *Photogrammetric Engineering & Remote Sensing* 67: 1037–1047.
- M. Pinkovskiy and X. Sala-i-Martin. 2015. Lights, Camera, ... Income!: Estimating Poverty Using National Accounts, Survey Means, and Lights. *NBER Working Paper* 19831. Cambridge MA: National Bureau of Economic Research (NBER).

Although most of the studies that examine the viability of using big data to enhance official statistics compilation provide encouraging results, there are various issues that need to be considered before a decision to fully scale up such initiatives can be made. First, there should be a careful consideration of cost implications and sustainability of statistical outputs. Second, self-selection bias and representativeness are issues that need to be tackled when using several types of big data such as crowd-sourced and web scraped data since many people are still not connected to the internet. Third, there is also a need to work for codification and production of a metadata system to support the use of big data in official statistics (Ploug 2013). On this front, the UN Statistics Division is leading the development of classification and standards toward the formal definition of concepts related to big data for international comparability. In addition, there are several research areas that need to be examined further. In particular, big data analytics require new statistical methods that can allow inferences that are not heavily dependent on the conventional notion of statistical significance. Big data analytics should also

provide analytical tools that tackle inherent features of big data such as heterogeneity, noise accumulation, spurious correlation, and incidental endogeneity.

In addition to these challenges, there is also a need to address institutional barriers to the use of big data for compilation of official statistics. For instance, specific guidelines on sharing of private sector data holdings need to be carefully developed and examined in close consultation with the public sector, particularly the national statistics offices. Furthermore, there is a need to develop the capacity of national statistics offices in the use of big data and continuous development of new methods that will facilitate the efficient integration of big data into the national statistical systems.

Big Data and Sustainable Development Goals

The previous section has touched on the various applications of big data for official statistics in general. This section summarizes how we can leverage big data for monitoring the SDGs in particular.

It focuses on three themes in which big data can play an important role in addressing the data gaps in the SDGs: disaggregation, timeliness, and development of proxy indicators.

As pointed out earlier, the “*leave no one behind*” principle that the SDGs espouse requires the statistical indicators to be broken down or disaggregated by subpopulation groups. However, the existing data collection mechanism even for Tier 1 indicators do not usually allow such disaggregation. Hence, there is a need to explore how new data sources could complement conventional data collection strategies in methodologically robust ways to facilitate disaggregation. Big data can potentially provide a more granular social and geospatial breakdown and reduce the cost of collecting such data. For instance, CDR and process-mediated data can complement official statistics in providing a finer disaggregation of poverty indicators. With an appropriate data mining algorithm, gender and time–location indicators can be generated from these databases and can be used subsequently as auxiliary information in the estimation of poverty incidence and other indicators at various disaggregation levels. Similarly, mobile technology can also be used to oversample marginalized groups that are harder to reach through conventional data collection methods.

Complementing the conventional data sources used to compile the SDG indicators with big data can also potentially improve the timeliness of the release of the statistics. Even with Tier 1 indicators, regular updating has been a challenge due to the frequency of surveys and censuses. Several types of big data like those generated from traditional business systems can complement official statistics as indicators of themes under prosperity. Model-based estimation with big data as exogenous factors can be used in updating indicators between business survey and/or census years. Similar modeling approaches can be

used in the establishment of early warning systems to monitor progress toward the achievement of SDG targets on a more regular basis (e.g., annually). As an example, Google Trends data are used in predicting influenza prevalence (Yang, Santillana, and Kou 2015). A similar method may also be useful in predicting HIV or tuberculosis prevalence rates to complement the scarce data available among the countries in the region.

Big data are expected to play a key role in developing proxy measures for SDG indicators classified as Tier 3. For instance, the use of fixed censor data (weather) may be explored for the development of some indicators for SDG 13 (climate action) in combination with global weather indicators such as El Niño Southern Oscillation data. Similarly, mobile censor data can be mined for latent indicators that can be used to monitor SDG 12 (responsible consumption and production). In particular, data from different sources can be combined in the generation of indexes wherein the process associated with the targets can be viewed to be the latent factors from various indicators available. These latent factors can be extracted through principal components analysis or sparse principal components analysis.

Summary

This section has identified several applications of big data in compiling official statistics and SDG indicators. Overall, the results are encouraging and highlight that big data’s applications are wide-ranging. Nevertheless, there are some important lessons that are worth pointing out. “Big Data” is not a panacea to all data gaps that exist in official statistics and SDG indicators. In fact, big data are not always the right data because, in some cases, they can even introduce additional sources of bias and spurious correlations that could yield misleading conclusions. Secondly, sophisticated technology and data mining algorithms are not sufficient to fully understand

the results from analysis of big data. It requires an effective combination of sectoral expertise and distinct hardware and software capabilities.

As the development community increasingly recognizes the advantages of using big data to enhance the relevance and timeliness of official statistics, it should also make a conscious effort to address issues surrounding data quality and methodology, development of skills needed to work with big data, technological requirements, and the legal framework for sharing and principles of use of big data. To accomplish this, the development community needs to continue tracking relevant initiatives using big data so that we can have a more nuanced understanding on the scalability of such initiatives. Furthermore, the development community needs to work closely with various stakeholders including the private sector and government, particularly the national statistics bureaus.

References

- Asian Development Bank. 2015. *Key Indicators for Asia and the Pacific 2015*. Manila.
- P. Daas, M. Puts, B. Buelens, and P. van den Hurk. 2015. Big Data as a Source of Official Statistics *Journal of Official Statistics* 31(2): 249–262.
- T. Hak, S. Janouskova, and B. Moldan. 2016. Sustainable Development Goals: A Need for Relevant Indicators. *Ecological Indicators* 60: 565–573.
- N. Ploug. 2013. New Forms of Data for Official Statistics. In *Proceedings of the 59th World Statistics Congress of the International Statistical Institute*, 25–30 August, Hong Kong, China. The Hague, Netherlands: International Statistical Institute.
- J. Sachs. 2012. From Millennium Development Goals to Sustainable Development Goals. *The Lancet* 379: 2206–2211.
- P. Struijs, B. Braaksma, and P. Daas. 2014. Official Statistics and Big Data. *Big Data and Society*, April-June 2014: 1–6.
- Sustainable Development Solutions Network (SDSN). 2016. Measuring Peaceful, Just and Inclusive Societies: SDG16 Data Initiative.
- S. Tam and F. Clarke. 2015. Big Data, Official Statistics and Some Initiatives by the Australian Bureau of Statistics. *International Statistical Review* 83(3): 436–448.
- United Nations Development Programme (UNDP). 2013. *Strategy Paper: Sustainable and Inclusive Urbanization in Asia Pacific*. New York.
- United Nations High Commissioner for Refugees (UNHCR). 2014. *Global Trends 2013*. Geneva.
- R. Villars, M. Eastwood, and C. Olofson. 2011. *Big Data: What It Is and Why You Should Care*. IDC White Paper. Framingham, MA: International Data Corporation.
- World Health Organization (WHO). 2015. *Health in 2015: From MDGs to SDGs*. Geneva.
- X. Wu and V. Kumar. 2009. *Top 10 Data Mining Algorithms*. New York: Chapman and Hall/CRC Press.
- S. Yang, M. Santillana, and S. Kou. 2015. Accurate Estimation of Influenza Epidemics Using Google Search Data via ARGO. *Proceedings of the National Academy of Sciences* 112: 14473–14478.

