
# Data as a force for public good

## Main messages

1. Public intent data, a foundation of public policies, can play a transformative role in the public sector. However, gaps in the availability, quality, and usability of these data are pervasive, particularly in low-income countries—the countries that stand to benefit most from improving public intent data.

2. Lack of resources, technical capacity, and data governance hamper the production of useful data for public policy. Lack of data literacy and demand for data limits their use for public policy.

3. These problems can be addressed through the high-level prioritization of data, including long-term financing, investments in human capital, and laws conducive to the safe production, exchange, and use of data. Some investments in better data have paid for themselves.

4. Ensuring a political commitment to and predictable government financing for the production of public intent data remains a central struggle in lower-income countries. The political will to prioritize funding for data systems can be stimulated by boosting the demand for data.

## The central role of public intent data

Suppose a woman walks into a doctor's office and is given a diagnosis without examination by the doctor: no measurement of her heart rate, no recording of her symptoms, and no review of her medical history. The doctor just prescribes a medication. Such an approach, and such a world in which crucial data are not gathered, analyzed, and acted on, would not be welcome, to say the least.[1]

Yet all too often governments make decisions affecting people's well-being without understanding or even taking into account essential data. Designing policies without data is akin to a shot in the dark.[2] This problem is particularly acute in the poorest countries, where gaps in both the availability and the use of data are severest.[3]

Just as data gathered by a doctor can help improve a patient's diagnosis and ultimate well-being, data gathered by governments, international organizations, research institutions, and civil society can improve societal well-being by enhancing service delivery, prioritizing scarce resources, holding governments accountable, and empowering individuals. These data serve as the foundation for core functions of governments and their endeavors to reduce poverty. The data a doctor gathers often take the form of a conversation or some other means of communicating information between patient and doctor. In the same way, data gathered with the intent of informing public policy should enrich the policy dialogue and allow for systematic flows of information and communication among governments, their citizens, and commerce.

Such flows of information and communication require long-term investments in statistical capacity, infrastructure, data governance, data literacy, and data safeguards. These investments depend on one another. Failure in one area jeopardizes the value that data bring to development. Too often these investments are not made in the poorest parts of the world, contributing to data deprivations and poverty.

How should such deprivations be addressed? This chapter discusses the pathways through which data for public policy generate value for development, the obstacles to safe realization of value, and how those obstacles can be overcome.

## Public intent data and development: Three pathways for adding value

Public intent data—data collected with the intent of serving the public good by informing the design, execution, monitoring, and evaluation of public policy, or through other activities—are a prerequisite for many government functions. For that reason, government agencies are the primary producers of public intent data through censuses, surveys, and administrative data, among other things. Citizens, civil society organizations (CSOs), nongovernmental organizations (NGOs), academic institutions, and international organizations also contribute critically to the production of public intent data through surveys, crowdsourcing platforms, and other means. Data from firms can also be used for public policy—a topic that will be covered in chapter 4.[4] This chapter distinguishes between six types of public intent data that all serve the public good (box 2.1).

The discussion that follows uses country examples to describe three important pathways through which public intent data can bring value to development by (1) improving service delivery, (2) prioritizing scarce resources, and (3) holding governments accountable and empowering individuals. But these are not the only pathways. Others include regulating the economy and markets, fostering public safety and security, and improving dispute or conflict resolution.

The country examples reveal several conditions that should be in place to maximize the value of public intent data. The data need to be (1) produced with adequate spatial and temporal coverage (complete, timely, and frequent); (2) high in quality (granular, accurate, and comparable); (3) easy to use (accessible, understandable, and interoperable); and (4) safe to use (impartial, confidential, and appropriate)—see figure 2.1.[5] With these features, development-related data have the *potential* to transform development outcomes. For this potential to be realized, the data must be used explicitly to generate public good, including through the three pathways summarized in the following sections.

### Pathway 1: Improving service delivery

*Increasing access to government services.* One of the fundamental ways in which public intent data can improve livelihoods is by increasing access to government services. More access often requires data representative of all residents. Use of administrative data, particularly foundational identification (ID) systems such as national IDs and civil registries as well as digital identification, ensures that all persons are covered and access is equitable. In Thailand at the turn of the century, only 71 percent of the population was covered by a public health insurance scheme that was intended to be universal. Yet the country had a near-universal foundational ID and population

## Box 2.1 Six types of public intent data

*Administrative data*—such as birth, marriage, and death records and data from identification systems; population, health, education, and tax records; and trade flow data—are generated by a process of registration or record keeping, usually by national authorities. Administrative data also include data used by governments to run projects, programs, and services. The digital revolution has created new types of administrative data—for example, when education and health inspectors' use of smartphone apps channels data to a central register.

*Censuses* aim to systematically enumerate and record information about an entire population of interest, whether individuals, businesses, farms, or others. Most prominently, population and housing censuses record every person present or residing in a country and provide essential information on the entire population and their key socioeconomic conditions.

*Sample surveys* draw on a smaller, representative sample of the entire population, typically from censuses, to collect detailed information more frequently. These surveys cover many domains such as household surveys, farm surveys, enterprise surveys, labor force surveys, and demographic and health surveys. Key official statistics, such as unemployment and national accounts, rely on survey data, often in combination with administrative data and census data.[a]

*Citizen-generated data* are produced by individuals, often to fill gaps in public and private sector data or when the accuracy of existing data is in question. These data, which can have an important monitoring and accountability function, contribute to solving problems that citizens face.[b] Examples include HarassMap, an Egyptian tool that maps cases of sexual harassment based on citizen reports, and ForestWatchers, a platform through which citizens monitor the deforestation of the Amazon.

By contrast, *machine-generated data* are automatically generated by a sensor, application, or computer process without human interactions. An example is the sensors that monitor air pollution. These data emerge when devices are embedded with sensors and other technologies, allowing them to transfer data with each other, a system known as the Internet of Things.

*Geospatial data* relate multiple layers of information based on their geographic locale. Public intent geospatial data include satellite imagery of the Earth such as that provided by the US National Aeronautics and Space Administration's Landsat program and the European Space Agency's Copernicus program; weather data; and cadastral (property and land record) data.[c]

These data types are neither exhaustive nor mutually exclusive. For example, all data sources can be georeferenced and thus can be used in geospatial applications, and some administrative data and geospatial data can be machine-generated. Data sources are interoperable when they can be linked across and within these types though common numeric identifiers for persons, facilities, or firms; geospatial coordinates; time stamps; and common classification standards.

a. Sample surveys also include the surveys that are implemented by social media companies and target a sample of users who are active on their platforms. Examples include the Future of Business and Gender Equality at Home surveys conducted on the Facebook platform.
b. Meijer and Potjer (2018).
c. Such data sources are discussed in greater detail in chapter 4.

## Figure 2.1 Certain data features can maximize the value of public intent data

| Ensuring the data have adequate coverage | Ensuring the data are of high quality | Ensuring the data are easy to use | Ensuring the data are safe to use |
|---|---|---|---|
| • Completeness<br>• Timeliness<br>• Frequency | • Granularity<br>• Accuracy<br>• Comparability | • Accessibility<br>• Understandability<br>• Interoperability | • Impartiality<br>• Confidentiality<br>• Appropriateness |

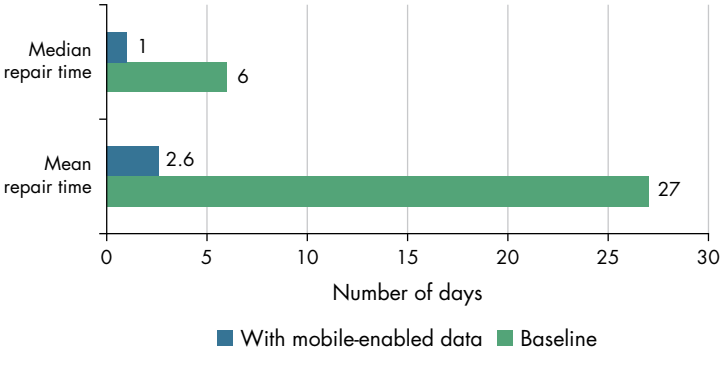
*Source:* WDR 2021 team, drawing on Jolliffe et al. (forthcoming).

registration system in which citizens and residents were issued a personal ID number when they were born or when their households were registered for the first time. Leveraging this register and the personal ID information from the existing public insurance scheme, the government was able to identify the population not covered and so was able to increase health insurance coverage from 71 percent to 95 percent.[6]

Machine-generated data also have the potential to markedly improve access to services such as water. In Kenya, sensors on water hand pumps, which are inoperable in one-third of rural Africa, provide real-time data on their functionality. This system helped reduce the average time to repair a broken pump from 27 days to three days and the median time from six days to one day (figure 2.2).[7]

*Better preparing for and responding to emergencies.* Public intent data can also lead to a better emergency response when disasters hit, whether environmental, financial, health, or conflict related. For example, weather data, especially weather forecasts, can help people anticipate and prepare for extreme events. The value of such data was revealed by two intense cyclones in the Bay of Bengal 14 years apart. The 1999 cyclone caught the Indian state of Odisha by surprise, causing massive devastation, killing more than 10,000 people, and destroying housing and public infrastructure. Since then, the Odisha State Disaster Management Authority and the government of Odisha have invested in weather forecast data and disaster response measures. When another cyclone hit in 2013, nearly 1 million people were evacuated to cyclone shelters, safe houses, and inland locations, and only 38 people died during and immediately after the storm.[8] These impressive results would not have been possible without the weather data that gave sufficient advance warning of the cyclone.

**Figure 2.2 Improving access to water: Using real-time sensor data to reduce repair time for broken hand pumps in Kenya**




*Source:* SSEE 2014. Data at http://bit.do/WDR2021-Fig-2_2.

Mobile technologies have the potential to speed up emergency responses. In Uganda, a health reporting program that provides beneficiaries, health professionals, and the Ministry of Health with real-time health data by using text messaging was able to cut the response time to outbreaks of disease by half. The technology was used after the 2012 Ebola outbreak to help implement quarantines and other protective measures.[9] As these examples demonstrate, timely data can contribute to quick reactions to a crisis.

*Generating useful knowledge.* Data generated and used by academic institutions, think tanks, and international organizations play a vital role in ensuring that policies are evidence-based. Impact evaluations of reforms and development projects are frequently used to assess whether past policies have had the intended consequences and to improve program design. In the last few decades, numerous field experiments have tested policies in a real-life setting under strict statistical conditions that allow cause and effect to be ascertained. Findings from such experiments have been used to implement new policies and scale up existing programs. One estimate suggests that the new policies and programs built on the research findings have reached more than 400 million people worldwide.[10] In Brazil, evidence from 2,150 municipalities found that many mayors are willing to pay to learn the results of impact evaluations, and that informing mayors about research on a simple and effective policy increases the probability by 10 percentage points that their municipality implements the policy.[11]

Research also plays an important role in ensuring the accuracy of the data collected by governments, which is critical to preventing policy recommendations based on inaccurate or misleading data.[12] The World Bank's Living Standards Measurement Study (LSMS) program, while supporting the production of household survey data in 106 countries between 2011 and 2020,[13] has also drawn attention to the importance of research on survey methodologies and the role of better measurement in eliminating systematic measurement errors in self-reported survey data that otherwise bias empirical analyses and policy conclusions.[14] Much of the methodological research led by the LSMS is carried out in partnership with national statistical offices (NSOs), in turn facilitating the adoption of improved methods in downstream national surveys.

### Pathway 2: Prioritizing scarce resources

*Targeting resources and reaching marginalized populations and areas.* When public intent data are granular—that is, they are tied to an individual or a specific location—they can help target resources and foster inclusion. In Croatia, data from the population census were

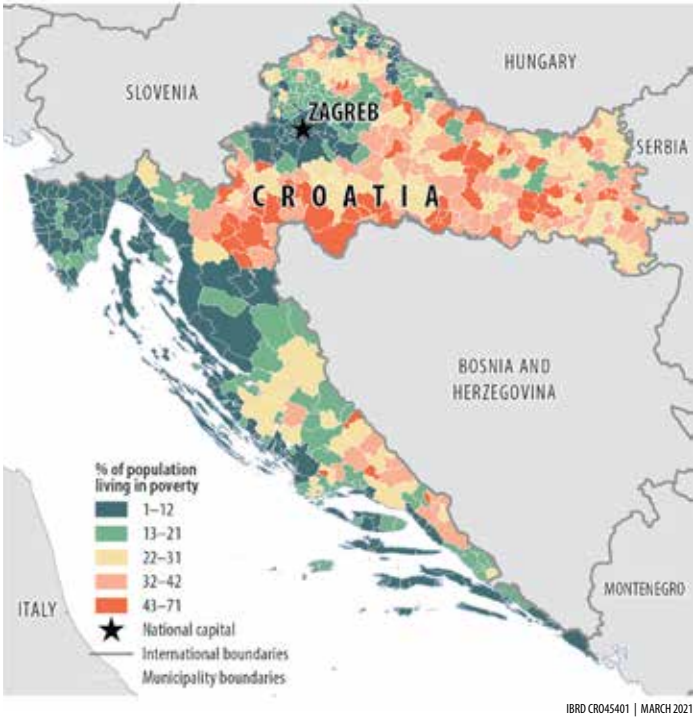
combined with household survey data and administrative data to create detailed maps of poverty and deprivations (map 2.1).[15] The maps revealed large differences in living standards across municipalities and within the territorial boundaries used for allocating funds from the European Union (EU).[16] More than one-third of the EU's annual budget—equivalent to more than €50 billion—is dedicated to investments in infrastructure, such as hospitals and schools, in less economically developed areas. Because the allocation of funds depends on an area's gross domestic product (GDP) per capita, poor municipalities situated in non-poor regions may not receive funding. Armed with the poverty map, Croatia responded with proposals for new geographical subdivisions that concentrate EU funds in the poorest areas.[17] This reordering, thanks to better data and analysis, has the potential to reduce inequality and pockets of poverty in Croatia.

A long-running and rich example of the value of granular data are the Demographic and Health Surveys, which cover topics such as HIV/AIDS and gender-based violence (see spotlight 2.1). Over the last few decades, data from 82 of these surveys, disaggregated by sex, have been used as inputs for developing laws banning domestic violence, developing HIV education programs, and more.[18] In Vietnam, a survey on gender-based violence revealed that more than half of women have experienced physical, sexual, or emotional abuse; that nearly half of these had physical injuries as a result; and that seven in eight did not seek any help. These data spurred a public discussion about the topic, informed the National Strategy on Gender Equality, and introduced counseling, health, legal, and shelter services for women subject to violence at home.[19]

*Saving money and resources.* Interoperability between geospatial data and government records can help governments save resources. Incomplete and out-of-date property and taxpayer records are an important reason that taxes remain uncollected in many low- and middle-income countries. In Tanzania, the government introduced a Geographic Information System for tax reporting and revenue collection. The system identified buildings via satellite imagery, collected and digitized data on their characteristics, and provided a comprehensive, up-to-date record of taxable properties. Using this new method, the city government of Arusha identified 102,904 buildings—nearly five times more than with earlier databases. One year after the system was introduced, the eight participating cities increased their revenue collection by 30 percent on average.[20]

Interoperable administrative data have also been used to increase efficiencies and save costs in public welfare systems. For example, in Argentina the government identified ineligible beneficiaries across various social programs using the country's system of unique taxpayer ID numbers. The exercise generated estimated savings of US$143 million over eight years.[21] More generally, investments in better data systems have been shown to pay for themselves.[22]

**Map 2.1 Reducing poverty: Mapping pockets of poverty in Croatia allowed better targeting of antipoverty funds**




*Source:* World Bank 2016a. Data at http://bit.do/WDR2021-Map-2_1.

*Monitoring progress and determining priorities.* Public intent data can also help prioritize resources by monitoring progress on key indicators and deliverables over longer periods of time. Such monitoring is vital for creating and tracking national and international development goals. The Sustainable Development Goals (SDGs), for example, rely heavily on public intent data.[23] If the data needed to measure one of the targets were collected only every 10 years, tracking progress would become challenging.

Cross-country comparable composite indexes—often created by think tanks, research institutions, and international organizations—allow countries to benchmark their performance over time and against peers and to decide on priorities. These data can induce countries to respond with reforms in areas where they are lagging. Multidimensional poverty indexes, which measure poverty at the household and individual levels, track certain indicators in countries

over time, helping countries decide on areas of focus. Costa Rica issued a presidential directive calling for use of such an index for budgetary planning and as an official measure for allocating resources and monitoring and evaluating social programs. The country has used the index to modify its budget allocation, which helped accelerate poverty reduction during a period of austerity without an increase in budget.[24]

### Pathway 3: Holding government accountable and empowering individuals

*Fostering transparency and increasing government accountability.* CSOs and individuals are frequent producers and users of public intent data. Their demand for data can encourage transparency through data analysis and data feedback systems. In China, media and watchdog organizations in Beijing noted inconsistencies between official government data on air quality and data from independent air quality monitoring systems. Heightened concerns about air quality have fueled a dramatic expansion in publicly available, real-time data from thousands of air quality monitoring locations.[25] The central government launched a US$275 billion plan to improve air quality throughout the country, and the Beijing municipal government promised an additional US$160 billion toward that goal.[26]

Good data can also encourage transparency in and improve public procurement. Too often, public projects are not implemented adequately due to poor procurement such as inflated costs, corruption, or ghost contracts. Because 12 percent of global GDP is spent on public procurement, this finding matters tremendously for development outcomes.[27] In Uganda, in an attempt to improve procurement outcomes, local government entities made administrative procurement data from the bidding process down to the level of execution of contracts available to certain CSOs. These CSOs trained community members to understand the information in the contracts and conduct site checks to verify it. The findings revealed mismanagement of resources by contractors and government officials and a high dependence on noncompetitive contracts. Not only did Uganda undertake reforms to ensure that contracts were complying with national procurement standards, but the national public procurement agency also upgraded its procurement portal in line with international open contracting data standards, making Uganda the first African country to do so.[28]

Government accountability can also be enhanced through e-governance.[29] In Pakistan, a smartphone app that equips government health inspectors with real-time data on rural public health clinics led to a 74 percent increase in clinic inspections. In turn, doctor attendance rose by 18 percentage points, thereby improving health care services.[30]

*Empowering individuals.* Disadvantaged groups are sometimes left out of government efforts to collect data because governments fail to acknowledge inclusion of those groups as a policy objective. Citizens must then often collect the data needed to empower themselves. That data, such as on harassment and early warning systems, can help fill a gap that neither the public sector nor the private sector can fill. The map-based mobile app Safetipin allows users to report mobility and safety issues in cities related to lighting, walk paths, visibility, public transport, and security. Beyond informing citizens where it is safe to be in their city, these data can be used to conduct citywide audits. In Bogotá, Colombia, the city government wanted to use this tool to map safety around bike paths. The biking community helped collect images along 230 kilometers of bike paths in the city, which were then analyzed by Safetipin (map 2.2). This analysis supported the authorities in understanding where to improve lighting and add closed-circuit TV cameras.[31]

Public intent data can also empower individuals to make better choices through more information and knowledge. The digital revolution has greatly increased the accessibility of data, as well as how easily information can be spread. One example is providing smallholder farmers with agricultural information digitally, often through text messages, to increase their productivity. Such data transmission can improve on extension services, which rely on in-person agricultural advice and are more costly to sustain and whose quality is more difficult to ensure. A meta-analysis suggests that providing agricultural information increases yields by 4 percent and farmers' probability of increasing productivity-enhancing agrochemical inputs.[32] With more than 2 billion people living on smallholder farms, these numbers can have major effects on global poverty and shared prosperity.

Many of the features of public intent data that increase their value for development can also increase their potential for harm. Data may be misused for political surveillance and control or discrimination and exclusion, or they may inadvertently expose sensitive information about individuals.[33] For example, in República Bolivariana de Venezuela, a digital biometric fingerprint system was introduced initially for voter registration and identification, but it has since been integrated with other registers. Identification

with the digital fingerprint has become mandatory to purchase basic goods such as food and medicine, which has led to numerous cases of stores refusing to sell young people, foreigners, and LGBTQI individuals such goods.[34] To avoid data being harmful in this and other ways, certain prerequisites must be put in place, notably robust data protection laws, independent oversight, and legal and technological solutions to safeguard the confidentiality of individuals and prevent misuse of data.

## Gaps in the coverage, quality, and usability of public intent data

Despite the demonstrated value of public intent data, gaps in their availability, quality, and usability persist, particularly in poor countries. This section documents these gaps, drawing on the World Bank's Statistical Performance Indicators (SPI), described in box 2.2, as well as two other prominent indexes rating public intent data availability and quality.[35]

### When the coverage of data is inadequate: Lack of timeliness, frequency, and completeness

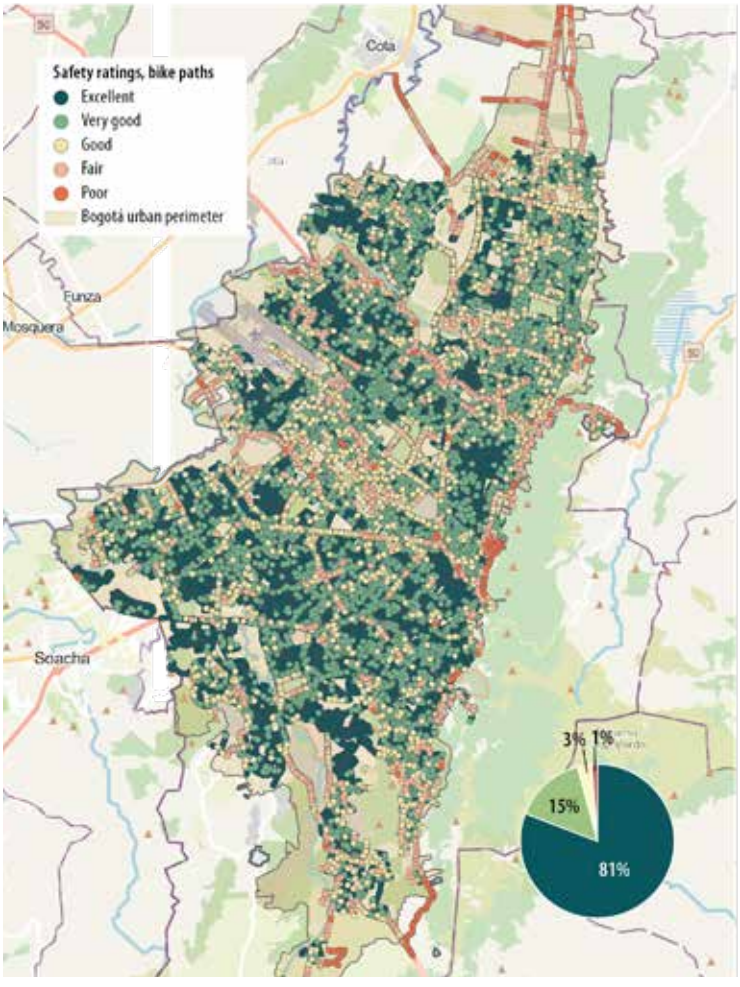
*Lack of timely and frequent data* remains an issue in many thematic areas and across all types of public intent data. Timeliness is particularly an issue with survey and census data because long lags commonly occur between their collection and their release. For example, according to the Statistical Performance Indicators, half of low-income countries have not undertaken a population and housing census in the last 10 years, and 18 percent have not done so in the last 20 years.[36] The census has a foundational function in any statistical system and is critical for political representation and resource allocation. The costs of allowing the census to become outdated are demonstrable.[37] Monthly or quarterly industrial production indexes, which are important to track current economic activity, are available in only 9 percent of low-income countries, compared with 40 percent of lower-middle-income countries, 48 percent of upper-middle-income countries, and 64 percent of high-income countries.[38]

Ground-based sensors, deployed in Internet of Things systems, can measure some outcomes, such as air pollution, climatic conditions, and water quality, on a continual basis and at a low cost. However, adoption of these technologies is still too limited to provide timely data at scale, particularly in low-income countries.[39]

*Lack of completeness* is often less of a problem in census and survey data because they are designed to cover the entire population of interest. For administrative data, the story is different. Civil registration and vital statistics systems (births and deaths) are not complete in any low-income country, compared with completeness in 22 percent of lower-middle-income countries, 51 percent of upper-middle-income countries, and 95 percent of high-income countries.[40] These gaps leave about 1 billion people worldwide without official proof of identity.[41] More than one-quarter of children overall, and more than half of children in Sub-Saharan Africa, under the age of five are not registered at birth.[42]

Although population and housing censuses are designed to represent all individuals at the time of

**Map 2.2 Improving public safety: The use of citizen-collected data in Bogotá led to greater safety around bike paths**




IBRD COL45414 | MARCH 2021

*Source:* Safetipin 2016.

*Note:* Safety ratings of poor to excellent for Bogotá bike paths are based on safety scores.

## Box 2.2 The World Bank's Statistical Performance Indicators

The World Bank's Statistical Performance Indicators (SPI) measure statistical performance across 174 countries.[a] The indicators are grouped into five pillars: (1) data use, which captures the demand side of the statistical system; (2) data services, which looks at the interaction between data supply and demand such as the openness of data and quality of data releases; (3) data products, which reviews whether countries report on important indicators; (4) data sources, which assesses whether censuses, surveys, and other data sources are created; and (5) data infrastructure, which captures whether foundations such as financing, skills, and governance needed for a strong statistical system are in place. Within each pillar is a set of dimensions, and under each dimension is a set of indicators to measure performance. The indicators provide a time series extending at least from 2016 to 2019 in all cases, with some indicators going back to 2004. The data for the indicators are from a variety of sources, including databases produced by the World Bank, International Monetary Fund (IMF), United Nations (UN), Partnership in Statistics for Development in the 21st Century (PARIS21), and Open Data Watch—and in some cases, directly from national statistical office websites. The indicators are also summarized as an index, with scores ranging from a low of 0 to a high of 100.

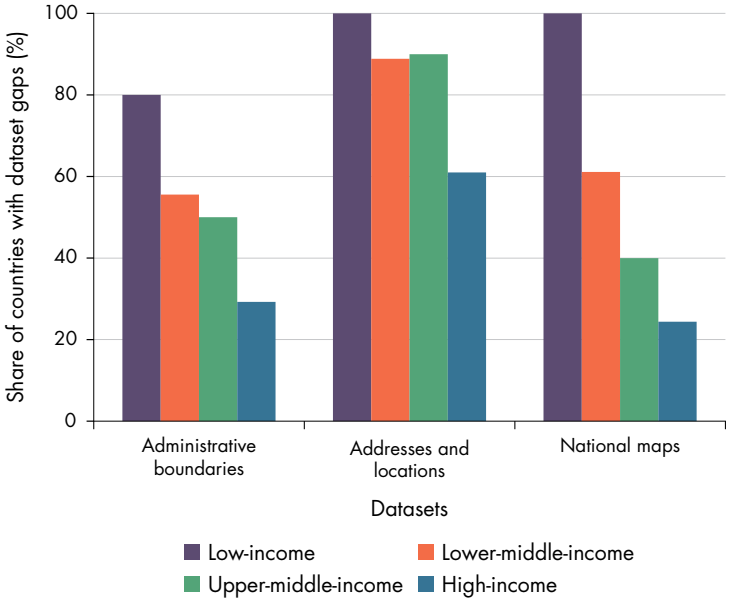
a. World Bank, Statistical Performance Indicators (database), http://www.worldbank.org/spi; Dang et al. (2021a, 2021b).

the census, they can leave out some of the poorest and most vulnerable. Many vulnerable groups are hard to count in the first place, especially when census enumeration focuses on residence and the concept of the household. These groups include the displaced, the homeless, slum inhabitants, nomads, migrants, young children, and the disabled.[43] The extent of undercounting is difficult to measure systematically, but in 2013 it was estimated that globally between 170 million and 320 million people were missing from population census frames, with the poorest more likely to be missed.[44] As noted, in many countries the census determines the allocation of resources and political representation. Thus these omissions have real consequences and can disenfranchise vulnerable populations.[45] They also affect the representativeness of household surveys that use census-based sampling frames.[46]

Lower-income countries also are susceptible to coverage gaps in geospatial data, especially in some of the geospatial reference datasets such as administrative boundaries, postal codes, and maps. The Global Open Data Index of the Open Knowledge Foundation assesses the availability and openness of three such geospatial datasets in 94 countries: administrative boundaries, addresses and locations, and national maps. The assessment reveals that all three datasets are often incomplete in lower-income countries (figure 2.3).

Similarly, the road network coverage of the open mapping platform OpenStreetMap is complete in many high-income countries, but less so in lower-income countries. OpenStreetMap is a citizen-generated geospatial application that relies on its users to digitize the location of roads and other infrastructure. Its coverage disparities reflect the barriers to making this type of data work for the poorest countries. In India, by 2015 only 21 percent of the road network had been digitized.[47]

**Figure 2.3 Gaps in geospatial datasets are especially large in lower-income countries**




*Source:* WDR 2021 team calculations, based on data of Open Knowledge Foundation, GODI (Global Open Data Index) (database), https://index.okfn.org/. Data at http://bit.do/WDR2021-Fig-2_3.

### When data quality is poor: Lack of granularity, accuracy, and comparability

*Lack of granularity* can occur when data are not available at the desired level of disaggregation. The gaps in data on women and girls are particularly severe. Only 10 of the 54 gender-specific indicators (19 percent) in the SDGs are widely available, based on international standards for measurement, and only 24 percent of the available gender-specific indicators are from 2010 or later.[48] Gaps in sex-disaggregated data related to the COVID-19 pandemic are also pervasive, causing knowledge of the gender impacts of the pandemic to be incomplete (box 2.3).

## Box 2.3 Gender data and the COVID-19 pandemic

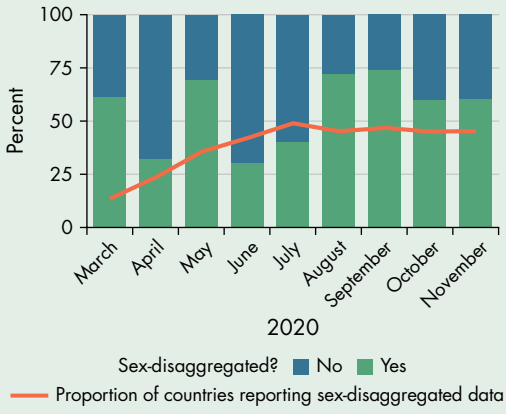
The COVID-19 pandemic was not gender-blind; it affected men and women differently and may have exacerbated gender inequalities.[a] Yet knowledge of the gender impacts of COVID-19 is incomplete because of data gaps across all dimensions of well-being. At the most basic level, data are lacking on COVID-19 infections and deaths among men and women. In March 2020, only 61 percent of reported COVID-19 cases were disaggregated by sex, and these data were provided by 26 countries. By November 2020, reporting had grown to 80 countries, but the proportion still stood at 60 percent. The reporting was irregular throughout 2020, as shown in figure B2.3.1.

Understanding the gender dimensions of the COVID-19 impacts extends well beyond case and mortality data. The data systems in place prior to the pandemic had notable gender data gaps that hampered the ability to track impacts and inform policy. For example, monitoring impacts on jobs requires regular and timely data on informal employment where women predominate. However, only 41 percent of low-income countries (LICs) and lower-middle-income countries (LMICs) report data on informal jobs disaggregated by sex. And in seven of the 10 countries where the recent economic contraction is severest, less than 38 percent of Sustainable Development Goal economic opportunity indicators are available by sex.[b] Furthermore, preexisting biases in face-to-face household survey design and implementation bled into phone surveys implemented during the pandemic, limiting measurement of the gender-related impacts of the crisis. These biases include designing phone surveys aimed at household heads and lack of survey content on time use.

There are also notable gaps in the gender data needed to inform policy design and effectiveness. Although the expansion of social protection programs is arguably the largest policy response to offset the economic impacts of the crisis, comparable sex-disaggregated measures of social protection coverage are largely unavailable. Data on personal identification cards and mobile phone ownership should inform program design decisions, especially as countries scale up digital platforms. Yet data on gender differences in ownership of personal identity cards are missing for more than a third of countries. Less than a quarter of LICs and LMICs report data on mobile phone ownership by women.[c]

Even though the pandemic created new demands for statistics, it also interrupted the supply. More than half of LICs and LMICs reported that the COVID-19 pandemic affected national statistical offices' ability to produce socioeconomic statistics.[d] This problem requires immediate attention, but building effective, gender-aware data systems will require sustained financial and human capital investments.

**Figure B2.3.1 Proportion of COVID-19 cases reported with sex-disaggregated data for 190 countries**




*Sources:* Global Health 50/50, University College London, COVID-19 Sex-Disaggregated Data Tracker (database), November 30, 2020, data release, https://globalhealth5050.org/the-sex-gender-and-covid-19-project/; Global Change Data Lab, University of Oxford, Our World in Data, Coronavirus Pandemic (COVID-19) (database), https://ourworldindata.org/coronavirus; calculations of Open Data Watch, Washington, DC. Data at http://bit.do/WDR2021-Fig-B2_3_1.

*Sources:* Mayra Buvinic (Center for Global Development), Lorenz Noe (Data2x), and Eric Swanson (Open Data Watch), with inputs from the WDR 2021 team.

a. UN Women (2020).
b. Buvinic, Noe, and Swanson (2020).
c. Buvinic, Noe, and Swanson (2020).
d. UNSTATS and World Bank (2020).

Although data disaggregated at the individual level are central to understanding and addressing conditions that uniquely affect the lives of women, men, children, adults, the elderly, and persons with disabilities, the required data are not being sufficiently produced. For example, survey data on ownership of physical and financial assets have traditionally been collected at the household rather than the individual level, limiting their usefulness in understanding women's relative wealth, rights, and decision-making power in their families.[49] Monetary poverty estimates are also based on household-level measures of resources, and "poor individuals" are identified based on the poverty status of their entire households, regardless of differences within households among women, men, and children in access to and use of resources.[50] Meanwhile, gaps remain in the adoption and proper implementation of the survey questions developed by the Washington Group on Disability Statistics—questions that are critical for obtaining internationally comparable estimates on disabilities and for disaggregating relevant SDG indicators by disability status.[51]

Finally, despite the enormous potential of geographically granular data for targeting policies effectively, such disaggregated data are rarely available comprehensively. According to the 2020 Open Data Inventory, about 90 percent of official statistics, even when they are available, are not consistently reported at the regional level (first administrative division), and almost none are consistently reported at the district level (second administrative division).

*Poor accuracy* of data can limit their usefulness. For those collecting individual-level data through household surveys, a concern is the choice of survey respondents. Relying on proxy respondents to elicit individual-level information—a common cost-saving mechanism in large-scale household surveys—has been shown to produce wrong estimates of gender differences in asset ownership, labor market outcomes, decision-making, and control of income.[52] Reported levels of income, wages, and firm profits vary, depending on the length of the period over which they are recalled by survey respondents.[53] The length of recall also matters for the accuracy of survey data on agricultural production, health, and labor.[54]

Accuracy is also a concern for administrative data. One reason for the proliferation of survey data is the perception that administrative records are unreliable and incomplete.[55] A study of multiple African countries found overreporting of vaccination rates in health information systems by 5 percent of countries and of primary enrollment rates in education management systems by a third. This data inflation appears to be connected to making aid flows conditional on results, creating an incentive to misreport.[56]

Data quality concerns and methodological challenges also characterize data produced by the Internet of Things. For example, the quality of data generated by low-cost commercial sensors used for air pollution monitoring has been found to vary widely when benchmarked against reference measurements.[57] Sensors must be calibrated to the specific conditions in which they are used to yield accurate results, but the calibration process remains expensive and time-consuming.[58]

*Lack of comparability* is particularly a concern among low-income countries. Only 40 percent of low-income countries, 20 percent of countries in fragile and conflict-affected situations (FCS), and 40 percent of countries in Sub-Saharan Africa (figure 2.4) have at least three comparable estimates of extreme poverty.[59] It is therefore difficult to understand changes in living standards over time and design policies to eradicate poverty. Recent innovations in data collection in these countries suggest a slightly more optimistic picture for the future.[60] It is also important to note that some lack of comparability over time is necessary, particularly when adopting new global standards.

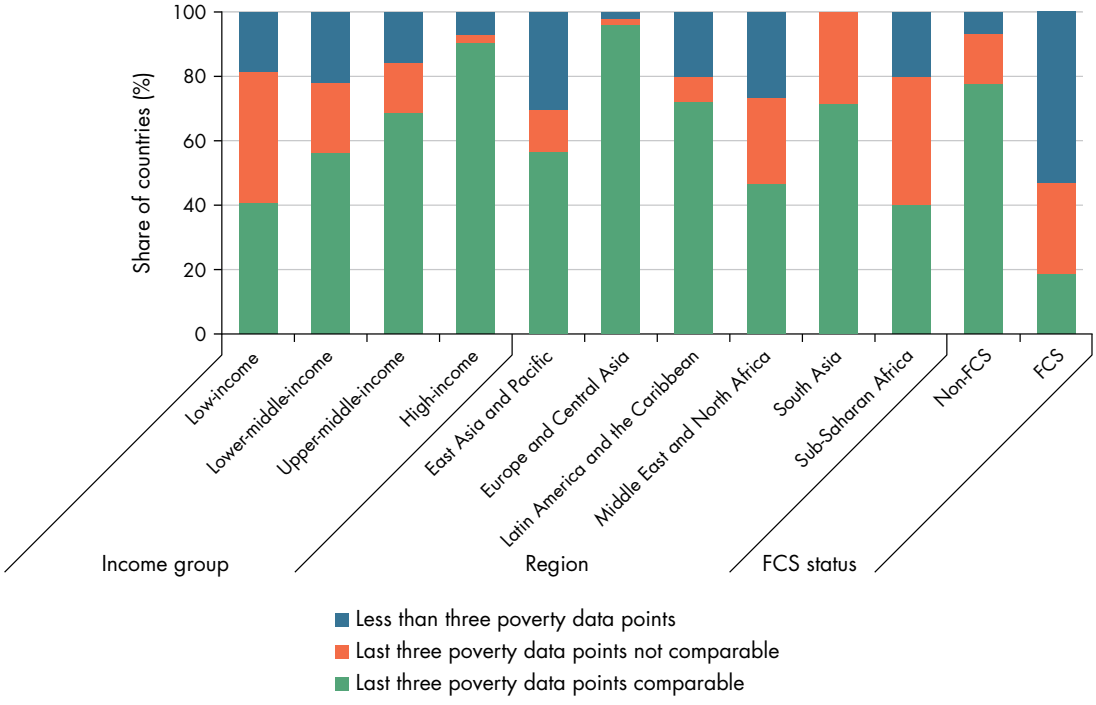
### When data are not easy to use: Lack of accessibility, understandability, and interoperability

*Lack of data accessibility* prohibits actors from using data. According to an assessment of the Open Data Inventory, lower-income countries lag far behind in overall data openness (table 2.1), although even high-income countries have mediocre openness scores. Only 11 percent of low-income countries consistently make data available with a license classifiable as open, compared with 19 percent of lower-middle-income countries, 22 percent of upper-middle-income countries, and 44 percent of high-income countries.

The Open Data Inventory assessment also reveals some limitations to machine readability. To the extent that governments publish official statistics, only 37 percent of low-income countries make at least some of these available in machine readable formats, compared with 51 percent of lower-middle-income countries, 61 percent of upper-middle-income countries, and 81 percent of high-income countries.

One reason for lack of data accessibility is that data systems in the public sector can be very fragmented. The health sector, for example, often has many different health information systems because of its tendency to have many different service providers. These include many private providers whose data are often

**Figure 2.4** Lower-income countries, especially those affected by fragility and conflict, have less comparable poverty data than other country groups




*Source:* WDR 2021 team calculations, based on World Bank, PovcalNet: Data (database), http://iresearch.worldbank.org/PovcalNet/data.aspx. Data at http://bit.do/WDR2021-Fig-2_4.

*Note:* Only those economies with at least one international poverty estimate are included. FCS status refers to the World Bank's "Classification of Fragile and Conflict-Affected Situations" (World Bank 2020a).

**Table 2.1** Assessment of the openness of data, by country income group

| Indicator | Low-income | Lower-middle-income | Upper-middle-income | High-income |
|---|---|---|---|---|
| Openness score (0–100) | 38 | 47 | 50 | 66 |
| Available in machine readable format (%) | 37 | 51 | 61 | 81 |
| Available in nonproprietary format (%) | 75 | 85 | 81 | 84 |
| Download options available (%) | 56 | 68 | 68 | 78 |
| Open terms of use/license (%) | 11 | 19 | 22 | 44 |

*Source:* WDR 2021 team calculations, based on 2020/21 Open Data Inventory indicators (Open Data Watch, ODIN [Open Data Inventory] [database], https://odin.opendatawatch.com/), also used as part of the World Bank's Statistical Performance Indicators database, http://www.worldbank.org/spi.

*Note:* The openness score is the average by country income group on a scale of 0–100. All other indicators are the percentage of published data averaged by country income group.

unavailable to the Ministry of Health. In Ethiopia, a study of the health sector found 228 different digital health information applications, of which only 39 percent sent data to the Ministry of Health.[61] Administrative data, in particular, are too often siloed in different systems, prohibiting their effective use for monitoring and policy design. Although data coordination within agencies is often limited, the challenge of siloed systems is even greater across government agencies.[62]

*Lack of understandability* prevents even those data that are accessible from generating value. To be understandable, data must be well disseminated, backed up with sufficient metadata, responsive to user needs, and, for certain purposes, summarized and visualized for the user. A majority of countries have data portals and provide metadata for their published data—practices that facilitate wider data use.[63] Low-income countries perform comparatively well in the data portal and metadata categories, but even here they lag. A larger gap remains in terms of advance release calendars, which commit government units to release data on a predetermined timetable. Only

**Table 2.2** Data dissemination practices and openness, by country income group

| Indicator | Low-income | Lower-middle-income | Upper-middle-income | High-income |
|---|---|---|---|---|
| NSO uses advance release calendar | 30 | 75 | 92 | 98 |
| NSO has data portal | 84 | 91 | 95 | 92 |
| NSO has conducted user satisfaction survey | 10 | 20 | 19 | 33 |
| NSO makes metadata available | 63 | 91 | 97 | 100 |

*Source:* Cameron et al. 2019.

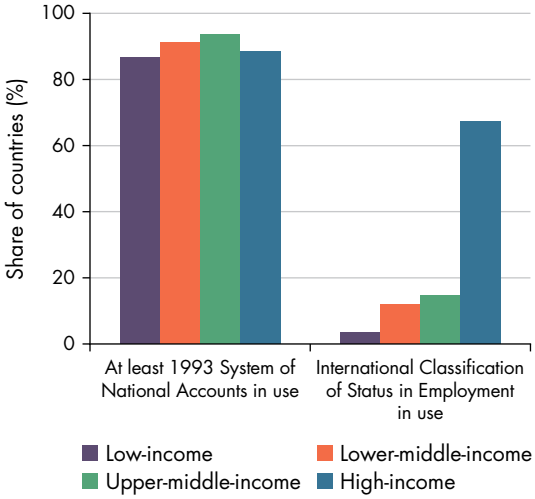
*Note:* Data are for 2019. The percentages reflect the proportion of the population in each income group whose national statistical office (NSO) has the listed attribute.

30 percent of NSOs in low-income countries publish such calendars, compared with almost all high-income countries. Across the board, only a few NSOs utilize user satisfaction surveys, which could play an important role in gauging and understanding data demand (table 2.2).

*Limitations to interoperability.* The use of common standards, methodologies, and classifications across public intent data sources ensures interoperability and enables data integration. Common and unified identification is needed across producers of public intent data for geographic divisions below the national level, such as regions, states, and districts. There is significant scope for expanding the use of georeferencing in censuses, surveys, and collection of administrative data, particularly in low-income settings. The use of common and unified personal identifiers to match data across multiple data sources is more contentious because of privacy and equity concerns, and robust data protection legislation is a prerequisite for their use.[64] Personal identification also requires trust and comprehensive civil registration and vital statistics systems, which have so far been elusive in the poorest countries. The use of tokenized identifiers in line with privacy by design principles is a potential solution.[65]

Adhering to set methodologies and standards in line with international best practices greatly increases the interoperability and usability of public intent data. The World Bank's Statistical Performance Indicators capture this aspect of public intent data systematically. Under the indicator on data infrastructure, standards related to systems of national accounts, employment status, consumption, consumer price indexes, and government finance statistics, among others, are assessed. The indicator shows a strong income gradient in the adherence to international best-practice standards and methodologies.[66] For example, the International Classification of Status in Employment is being used in two-thirds of high-income countries but in only 7 percent of low-income countries (figure 2.5). By contrast, a large share of all countries globally is using at least the 1993 international standards for the System of National Accounts (SNA 1993).

**Figure 2.5** Lower-income countries are less likely than other countries to adhere to international best-practice statistical standards and methodologies




*Source:* WDR 2021 team calculations, based on World Bank, Statistical Performance Indicators (database), http://www.worldbank.org/spi. Data at http://bit.do/WDR2021-Fig-2_5.

### When data are not safe to use: Lack of impartiality, confidentiality, and appropriateness for development

Gaps also remain in the safety of data. These can occur when data are not immune to influence from stakeholders, when they are not stored securely, or when they are not properly deidentified. For example, Greece's debt statistics appear to have deliberately misrepresented the country's financial situation in

the lead-up to the 2009 euro crisis, and data breaches are all too common in government and private sector databases.[67]

Similarly, deidentifying individuals has not always proved to be enough to maintain confidentiality. In the 1990s, the governor of Massachusetts in the United States approved making deidentified medical records of state employees available for researchers. Although key identifiers such as name and address were removed from the data, by triangulating the information available with other public information a researcher was able to identify the medical records of the governor and other individuals (see chapter 6 for more details).[68] One way to minimize these concerns is to ensure that only appropriate data are produced—data that measure concepts of interest, have a clear policy purpose, and are not produced from attempts to collect excessive information or surveil individuals. Such data, of course, can still be misused and mishandled.

## Why data gaps persist: The political economy of public intent data

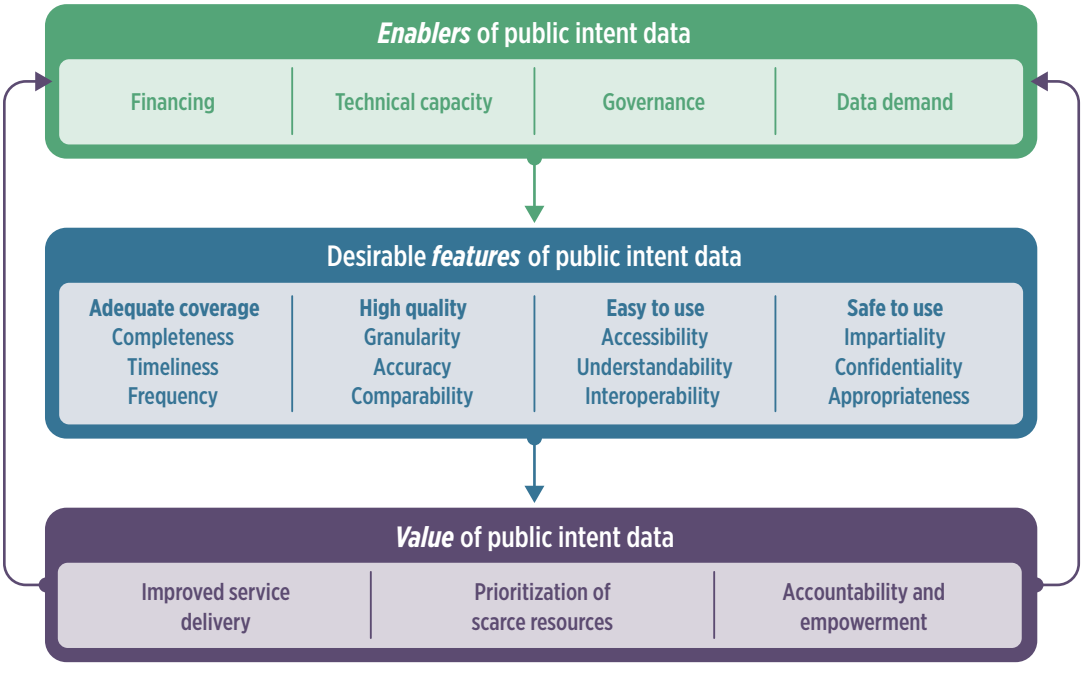
The previous two sections describe how public intent data can yield great value for development, yet gaps in public intent data are severe, particularly in low-income countries—the countries that stand to benefit most from the data. Why do these data gaps persist? This section answers that question, complementing existing data sources with structured interviews with NSOs across all income groups and geographical regions.[69] This approach requires digging one level deeper and understanding the main roadblocks on the pathways to data for public policy, or conversely, the enablers of public intent data. The main roadblocks identified are lack of financing, technical capacity, data governance, and demand for data (figure 2.6).

A common reason for these roadblocks is lack of understanding of and commitment to the use of data for policy making. In a positive feedback loop, realizing the value of public intent data increases understanding of the potential of the data, leading to a commitment to the further production and use of public intent data. To spearhead such commitments, SDG Target 17.18 calls for increasing the availability of high-quality, timely, and disaggregated reliable data, and SDG Target 17.19 calls for developing measurements of progress related to statistical capacity building.

### Deficiencies in financing

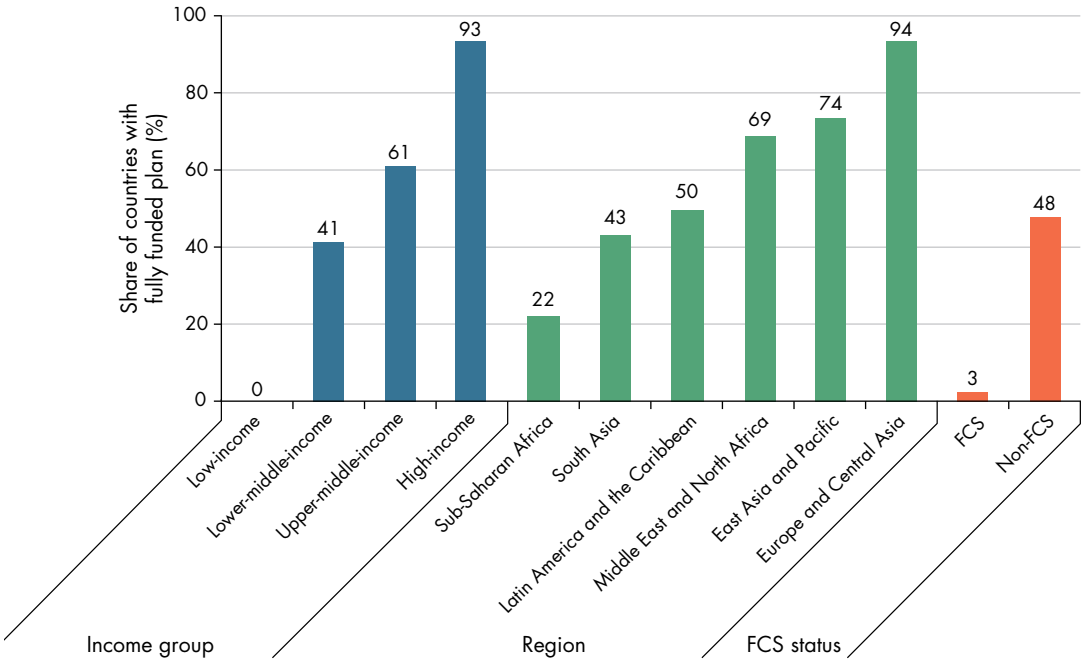
Underinvestment and misaligned investment priorities are perpetuating data gaps.

**Figure 2.6** A positive feedback loop can connect enablers and features of public intent data with greater development value




*Source:* WDR 2021 team.

**Figure 2.7** Most countries do not fully fund their national statistical plans




*Source:* WDR 2021 team calculations, based on indicators collected by the Partnership in Statistics for Development in the 21st Century (PARIS21) that are also used as Statistical Performance Indicators (World Bank, http://www.worldbank.org/spi). Data at http://bit.do/WDR2021-Fig-2_7.

*Note:* Having a fully funded national statistical plan under implementation is Sustainable Development Goal Indicator 17.18.3. FCS = fragile and conflict-affected situations.

*Underinvestment by governments.* Underinvestment in public intent data systems is widespread. Only half of countries had a national statistical plan that was fully funded in 2019 (figure 2.7).[70] Lack of national funding for statistics is especially a struggle for fragile and conflict-affected countries, countries in Sub-Saharan Africa, and low-income countries. Whereas 93 percent of high-income countries have a fully funded national statistical plan, not a single low-income country has one. A recent review of public financing of statistics found that seven of 10 low- and middle-income countries analyzed funded less than half of their respective national statistical plans, with country contributions ranging from 9 percent to 77 percent.[71]

This problem is more pressing in low-income countries with less government revenue to spend on multiple priorities. However, the cost of public data systems is modest relative to that of other government functions. Decision-makers in budget offices may not fully understand how much funding is needed to produce high-quality data or lack the incentives to prioritize data. How well public data systems are funded is thus also a matter of high-level government officials recognizing the value of public intent data and offering leadership to encourage collection of them.[72] A key factor in such an effort is the perceived relevance and credibility of public intent data and its producers.[73]

Another reason for lack of funding for data is the absence of a benchmark guiding how much governments should spend, unlike for other areas of government spending. For example, the Education 2030 Framework for Action urges countries to allocate at least 4–6 percent of GDP or at least 15–20 percent of their total public expenditure to education. The Abuja Declaration urges countries to spend at least 15 percent of their annual budget to improve the health sector.[74] No similar guidelines are found on data.

*Underinvestment by donors.* Donors also invest relatively little in public intent data. The share of total official development assistance devoted to statistics has ranged between 0.35 percent and 0.4 percent in recent years, or US$693 million in 2018.[75] The combination of national and donor contributions leaves a funding gap of between US$100 million and US$700 million a year globally to upgrade public intent data systems, depending on the scope of improvements.[76]

*Misalignment of investment priorities.* Beyond the size of investments in public intent data, how donors invest matters as well. With insufficient government funding of data and with donors stepping in to fill needs, the risk is that donor priorities will be funded

at the expense of national priorities and that donors, instead of national stakeholders, will become the main clients of NSOs.[77]

Because investments in data tend to be small, donors have limited incentives to make longer-term commitments that strengthen data systems such as technical capacity, research and development, infrastructure, or recording of administrative data. Instead, many investments prioritize the production of new data or specific survey efforts such as a one-off survey on a specific topic.[78] In particular, donor priorities skew toward monitoring and international reporting.[79] Although most national governments subscribe to international reporting, there is arguably a more immediate need for frequent and highly geographically disaggregated data and strong administrative data systems for the effective day-to-day functioning of government.[80]

Within the development community, lack of donor coordination can undermine public intent data systems, leading to duplication of and parallel systems for data collection. Each project uses its own set of indicators to report results instead of relying on and strengthening country data systems.[81] Such situations can arise if donors need to fulfill their internal reporting requirements or are suspicious of the accuracy of government-reported data.

Lack of funding is also an issue for citizen-generated data. Interviews with representatives from NGOs in Argentina, Kenya, and Nepal revealed that lack of funding can constrain the collection of citizen-generated data.[82] Similarly, although the cost of sensors has steadily fallen over the last few years, the costs of equipment, deployment, and transmission, as well as the lack of off-the-shelf tools for environments facing resource constraints, are still major barriers to the generation and use of machine-generated data, especially in smallholder agriculture.[83]

### Deficiencies in technical capacity

Data gaps are also persisting because of underqualified, understaffed, and underpaid data producers and lack of technology and infrastructure.

*Lack of qualified staff, proper staff renumeration, and career incentives.* The gaps in public intent data also stem from limited technical capacity, especially in lower-income countries—a result in part of the limited and misaligned resources previously discussed. A shortage of skilled data scientists, statisticians, and economists across public data systems is a critical constraint on the performance of the data producers and the production of data, especially at a time when data from digital sources are becoming more important. The absence of key personnel in strategic positions who have a commitment to data is especially costly because of the importance of relationships between ministries and NSOs and with civil society as a catalyst for the flow of data and information.[84]

According to a global survey of NSOs conducted by PARIS21, after a shortage of funds the biggest obstacle to countries' successful development of capacity is lack of skilled staff to implement programs.[85] In a list of 15 goals for capacity development, 86 percent of African NSOs selected strengthening human resources as one of their five most important goals, higher than any other category. It is particularly difficult for NSOs to recruit new staff with the skills needed to achieve their objectives. When reporting the most frequent methods of human resource development, only 7 percent of NSOs reported recruitment of staff with new skill sets, and most of these NSOs were in high-income countries.[86]

Recruitment and retention of skilled staff are difficult without competitive pay scales and career tracks.[87] Consultations with NSOs revealed that differences in pay scales across government entities especially make it difficult for NSOs to recruit skilled staff. In Ethiopia, the Central Statistical Agency follows civil service rules and regulations for remuneration of staff, whereas research institutes and universities have their own rules and regulations.

A common challenge for other government agencies that produce data is that they lack designated data scientists or statisticians. This is particularly problematic when other agency staff may lack the time and capacity to make better use of the data collected within their institution.[88]

*Lack of technology, software, and infrastructure.* Even when producers of public intent data have staff with the skills needed to collect, process, and disseminate those data, they often lack the technological infrastructure to be effective in their work. Constraints in technology and information technology (IT) infrastructure compound constraints in technical capacity. For example, as part of the Global COVID-19 Survey of NSOs, many NSOs in low- and middle-income countries noted their need for software to collect data remotely to meet new data demands.[89] In the PARIS21 survey, the option most selected to achieve priorities for a national statistical system in the medium term is acquiring up-to-date technology and infrastructure.[90] Technological shortcomings also constrain the ability of individuals to produce data themselves because many types of citizen-generated data rely on phone or web technologies.[91]

### Deficiencies in governance

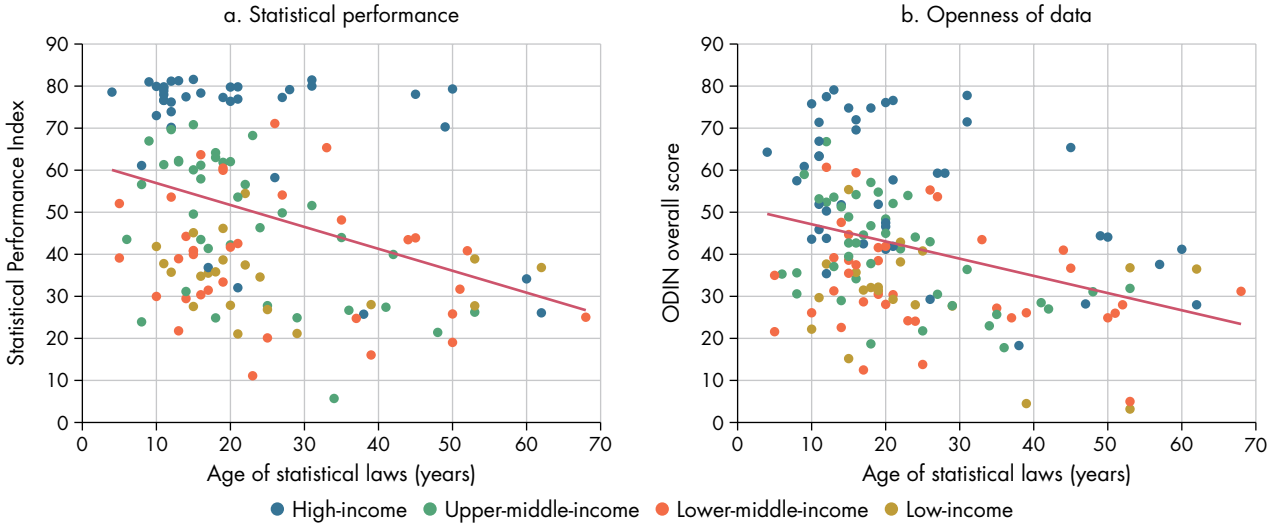
In addition to shortages of skills and funding, various failures and problems with data governance impede the potential of public intent data from being realized. At the national level, clear institutional mandates and good coordination among the data-producing agencies are critical for the exchange, interoperability, and timely publication of data.[92] In practice, exchanges of data across ministries and between ministries and NSOs and beyond are rare, even in well-resourced and high-capacity environments.[93] The absence of clear mandates, responsibilities, and incentives to effectively coordinate data production and data exchanges can obstruct collaboration and lead to duplication of data-gathering efforts.[94]

*Deficiencies in the legal framework.* The legal framework governing data production and data exchanges is a common barrier. Outdated statistical laws can make it difficult for NSOs and data-producing agencies to operate and collaborate effectively in light of recent changes in the data landscape, such as the proliferation of new data types, sources, and producers. In Chile, the National Institute of Statistics (INE) has had difficulties accessing key data from other public institutions in a timely fashion, primarily because the national statistical law is not sufficiently clear in authorizing INE's access to statistical information. When the law was passed in 1970, data exchanges were not a concern. Although a process to modernize the law has been at the forefront of political discussions for a decade, a revised version has yet to be formally implemented. This issue is a concern more generally because the older the national statistical law, the lower is statistical performance in general and data openness in particular at any country income level (figure 2.8).

Other important elements of the legal framework are regulations governing data protection and the right to information. When these safeguards are lacking or weak, data exchanges can entail serious risks to data protection.[95] Lack of comprehensive data protection regulations is a problem in many parts of the world.[96] A review of African countries found that only 28 percent had procedures in place to ensure deidentification of data before publication.[97] Without a requirement to share data and guidance on how to treat confidential information, any risk-averse government employee would face few incentives to share data, especially confidential data, considering the possibly high costs should confidentiality be breached. The absence of comprehensive data protection legislation can also facilitate misuse of data such as for political control or discrimination.[98]

*Independence of the NSO.* The legal, financial, and institutional independence of the NSO is an important element of a successful public intent data system, especially its data quality and openness.[99] The

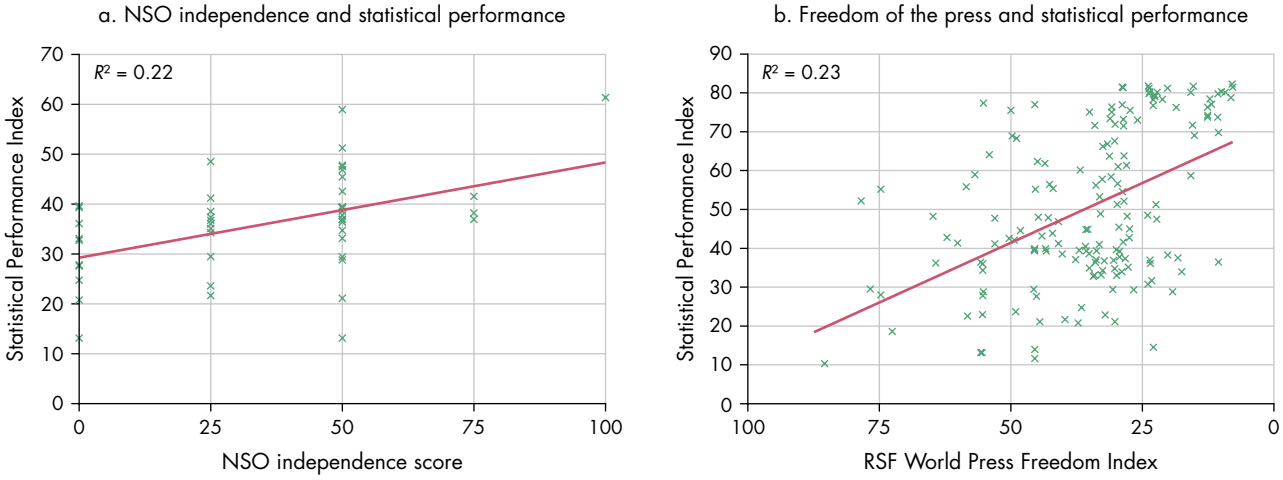
**Figure 2.8 The older a country's statistical laws, the lower is its statistical performance and the less open are its data**




*Sources:* WDR 2021 team, based on UNSTATS (Statistics Division, Department of Economic and Social Affairs, United Nations), UNSTATS (database), https://unstats.un.org/unsd/dnss/cp/searchcp.aspx; Partnership in Statistics for Development in the 21st Century (PARIS21), https://paris21.org/knowledge-database?keyword=&type%5B%5D=Statistical-Legislation-Country-Documents&date-from=&date-to=&page=; World Bank, World Development Indicators (database), https://databank.worldbank.org/source/world-development-indicators. Data at http://bit.do/WDR2021-Fig-2_8.

*Note:* In panel a, the regression coefficient on age, controlling for GDP per capita, is −0.48, $p < .01$; in panel b, −0.39, $p < .01$. For the Statistical Performance Indicators, see World Bank, Statistical Performance Indicators (database), http://www.worldbank.org/spi. For the Open Data Inventory (ODIN), see Open Data Watch, https://odin.opendatawatch.com/.

## Figure 2.9 Greater NSO independence and freedom of the press are positively correlated with better statistical performance




*Sources:* NSO independence score: Mo Ibrahim Foundation, Ibrahim Index of African Governance (database), http://mo.ibrahim.foundation/iiag/; World Press Freedom Index: Reporters Without Borders, 2020 World Press Freedom Index (database), https://rsf.org/en/ranking_table. Data at http://bit.do/WDR2021-Fig-2_9.

*Note:* The x's represent countries. Panel a shows only African countries, and panel b shows all countries with data available. The NSO independence score ranges from 0 to 100. The World Press Freedom Index ranges from 100 to 0—lower values imply greater press freedom. For the Statistical Performance Index, see World Bank, Statistical Performance Indicators (database), http://www.worldbank.org/spi. NSO = national statistical office; RSF = Reporters Without Borders.

independence of producers of public intent data also reinforces the credibility of and trust in the data and its producers, which encourages data use in both government and civil society.[100]

An indicator capturing the independence of NSOs in all African nations is included in the Ibrahim Index of African Governance.[101] The indicator measures the institutional autonomy and financial independence of an NSO. A perfect score indicates that an NSO is able to publish data without clearance from another government branch and has sufficient funding to do so. A higher score on the NSO independence indicator is highly correlated with statistical performance as captured by the World Bank's SPI (figure 2.9, panel a). In 2019 the average score on NSO independence was 34 out of 100, with low-income African countries scoring below average. These findings illustrate that NSO independence is precarious, particularly in lower-income countries. Anecdotes of attacks on NSO independence around the world suggest that fragile NSO independence is not limited to the African context.[102] For example, in 2007 the Argentine government began interfering with the independence of Argentina's NSO, the National Institute of Statistics and Censuses (INDEC). The effort initially focused on the consumer price index and later expanded to other official statistics, casting doubt especially on reported inflation statistics. Recognizing the harmful effects of these measures, by 2015 a new government had undertaken efforts to rebuild the institute, and INDEC resumed the delivery of trustworthy statistics with transparency and complete adherence to international principles.[103]

A government's interest in having an independent national statistical system can be affected by several competing factors. On the one hand, a government may have a vested interest in curtailing statistical independence and the production and dissemination of reliable data, fearing these could expose poor policy decisions and performance, dilute power, and increase public scrutiny and pressure.[104] In this case, lack of independence and the availability of reliable data would make it harder to hold governments accountable.[105] On the other hand, an independent statistical system producing reliable data in a transparent fashion best informs government decision-making and increases citizens' trust in government data and public institutions in general.[106] Such transparency can also facilitate favorable capital market and investment conditions and foster GDP growth.[107] Finally, international cooperation can boost statistical independence and data transparency when adherence to standards of data quality and the independence of their producers is required for accession to international organizations or agreements. An example is Colombia's successful bid to join the Organisation for Economic Co-operation and Development (OECD).[108]

Civil society performs a vital function in demanding transparency and holding government accountable. Citizen-generated data can be used to challenge official statistics when their accuracy or impartiality are in question. A free and empowered press is a

critical check on government power in general and on government interference with statistical independence and data transparency in particular. Greater press freedom, as measured in the World Press Freedom Index compiled by Reporters Without Borders,[109] is highly correlated with statistical performance as well as with statistical independence, regardless of a country's size or income level (figure 2.9, panel b).

### Deficiencies in data demand

Even when high-quality data are available and accessible, they must be put to an appropriate use to have an impact on development. As such, lack of data use is blocking the path to development.

*Low levels of data literacy.* Several barriers to data use remain. Low levels of data literacy among both policy makers and civil society are one barrier.[110] Potential data users need to have both a conceptual understanding of how data can inform policy questions and the technical skills to extract the relevant information from data. An analysis of the use of statistics in news articles in 32 countries in four languages revealed considerable scope for journalists to improve their critical engagement with statistics—and that finding is likely to apply to civil society at large.[111] For policy makers as well, data literacy is frequently identified as a barrier to data use.[112] Among the general population, comparably low literacy and numeracy rates in lower-income countries fundamentally diminish the pool of potential data users.[113]

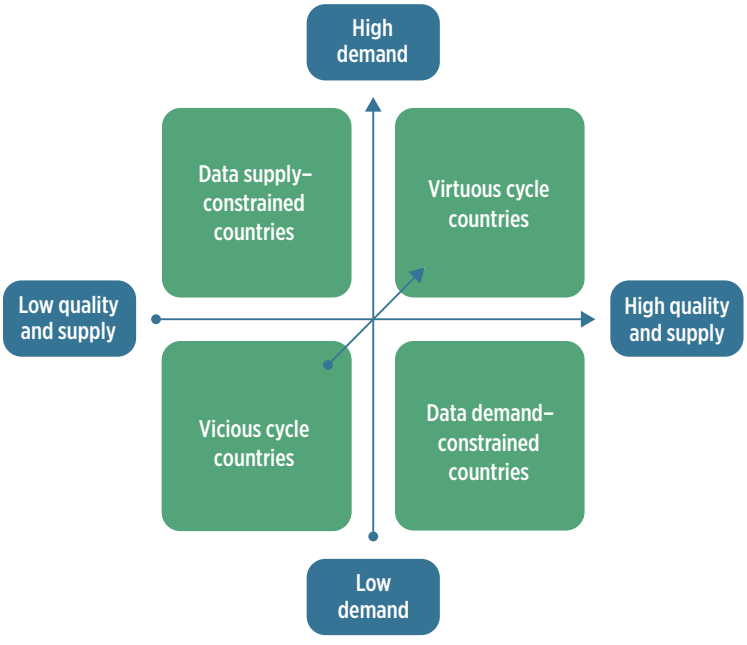
*Lack of incentives for and interest in data use.* Even when policy makers have the skills to use data, they may not be interested in exercising those skills because they do not attach value to data. Accordingly, another major factor affecting demand for public intent data is lack of incentives to use the data.[114] When political leaders exhibit a commitment to data use, they can generate expectations for civil servants to rely on data more frequently and create incentives for accountability. "Political champions," as well as changes in administration or individual government officials, often create opportunities for data-driven policy making.[115] A data-literate society plays a major role in creating these political commitments to data use by demanding—and rewarding—the justification of policy decisions with data.

*Low trust in the quality of public intent data.* Another reason for lack of data use is the often low trust in the quality of public intent data. Although data users can check for signs of internal coherence, the accuracy of data cannot be inferred from the data alone, and incorrect statistics can take years to be detected, if they are detected at all.[116] A survey of data producers and users in 140 countries found that NSO officials have much greater confidence in the quality of national statistics than ministry officials have.[117]

*Lack of infrastructure to access and use the data.* A final reason for lack of data use is related to the infrastructure needed to access and use data. For example, internet access is key to obtaining data, but penetration rates are lower in poorer countries. The exclusive sharing of data via online channels may exclude large shares of potential data users who are hampered by limited internet connectivity.[118] And certain users may be unaware that data are available for use.[119] Lack of internet connectivity, reliable power, and data centers are also major challenges in the use of Internet of Things systems and sensor data.[120]

Use of public intent data by a diverse group of actors often translates into greater demand for high-quality data. The rise in demand can drive investment in data and capacity, setting off a virtuous cycle of increasing data demand and supply (figure 2.10). For example, government ministries' reliance on and demand for high-quality data have been associated with NSOs in Latin America exhibiting higher capacity.[121] In the same region, demand for and interest in accurate and high-quality statistics in civil

**Figure 2.10 Data supply and demand can generate either virtuous or vicious cycles of data production and use**




*Source:* Adapted from Sanga (2013).

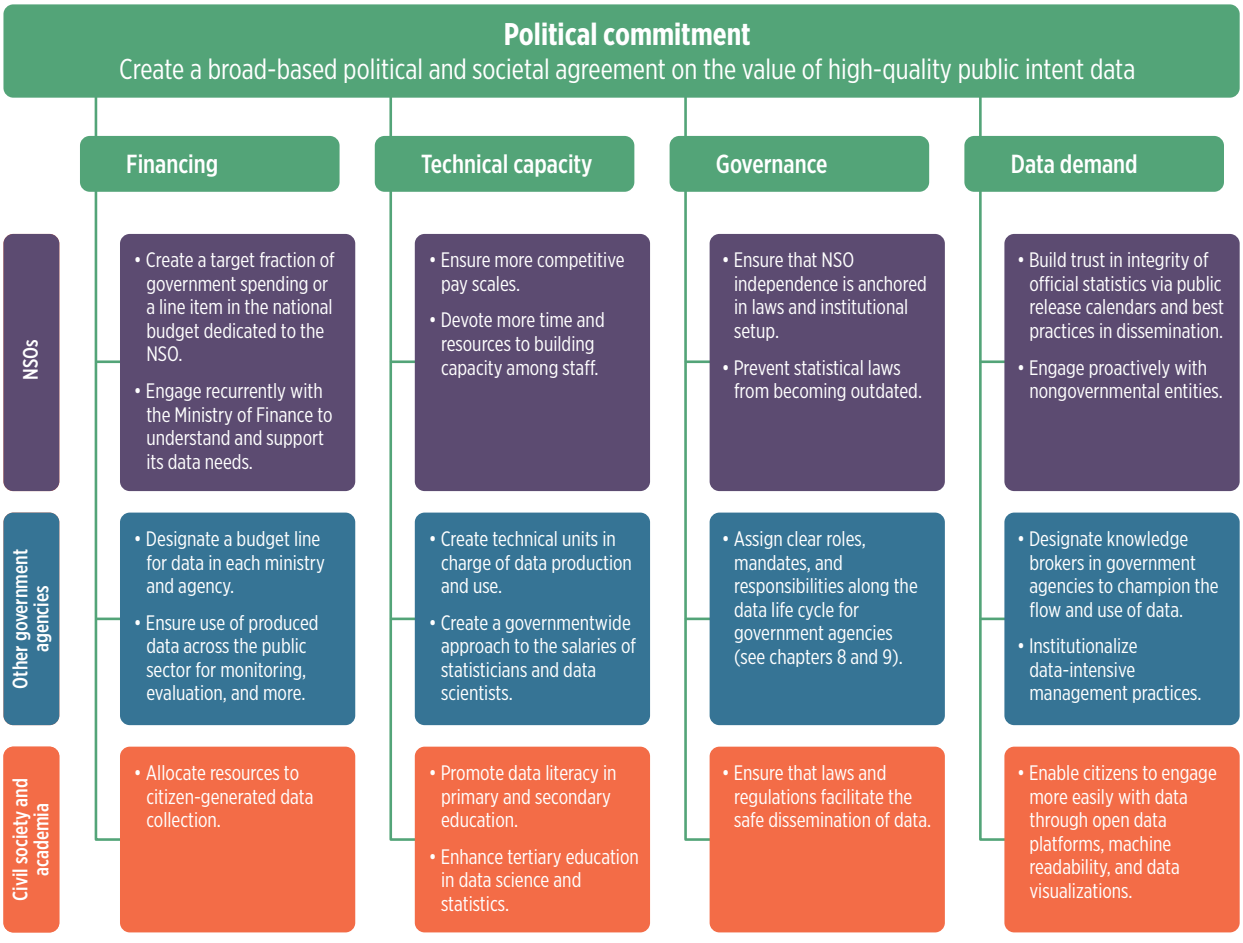
society, academia, the media, and the private sector have led to better funding, autonomy, and capacity in national statistical systems.[122] Conversely, countries with a low supply of data are likely to use data less, creating a vicious cycle of data production and use. In general, countries can benefit from assessing whether their constraints are primarily on the supply side or the demand side for data. They can then use such an assessment to prioritize data-related policies and maximize their return on development.[123]

## Realizing the potential of public intent data

To maximize the impact of public intent data on development, governments need to address the financing, technical capacity, governance, and data demand roadblocks. This section describes policies to overcome these foundational challenges. Figure 2.11 summarizes some of the main policies governments can enact, categorized by the actors and barriers they primarily address. International organizations also have a role to play, and spotlight 2.2 discusses how they can contribute to addressing the key roadblocks.

Chapter 9 builds on the analysis in this section, specifically in the domain of data governance, laying out a bold vision for an integrated national data system. Such a system can transform the role the public sector plays in the data modernization agenda by incorporating public intent data alongside private intent data, integrating the users and producers of both, and enabling safe data exchanges.

**Figure 2.11** Policies to realize the potential of public intent data




*Source:* WDR 2021 team.

*Note:* The figure summarizes policies governments can enact, categorized by the actors and barriers the policies are primarily addressing. Many policies span several actors and barriers but are placed into one box here for simplification. The role of the private sector in realizing the potential of public intent data is discussed in chapter 4. The role of international organizations is examined in spotlight 2.2. NSO = national statistical office.

A common reason for the four roadblocks on the pathways to data for public policy just described is the lack of a political understanding and appreciation of the value of data for policy making. Achieving high-quality production and use of public intent data requires an unequivocal high-level political commitment to data for development, even when data do not yield politically convenient insights. A broad-based political and societal agreement on the value of public intent data is the most effective way to ensure a robust political commitment to data. Such a social contract for data can build the trust of all participants that they will not be harmed in the production, exchange, and use of data. Actors from across the public sector, private sector, civil society, and academia can play an important role in demanding and encouraging agreement. One mechanism for formulating such broad agreement and formalizing a commitment to data is confirming the importance of data in countries' national development plans. Another mechanism is formulating a national data strategy—a topic discussed in greater detail in chapters 8 and 9.

### Financing needs: Strengthening and sustaining financial resources for data producers

Most low-income and lower-middle-income countries severely underspend on data. Securing sustainable financing is an enduring struggle for data producers and users. To reap the full value of data for development, governments must raise current spending levels drastically. At the same time, it is painfully hard to obtain and benchmark how much governments are spending on data. Thus one priority is to improve the statistics on government spending on data.

One way to increase the priority given to financing of data is to establish a target (percentage) for the government expenditure on the national statistical system. Such a target can be derived with a view toward the resources needed to fully fund the national statistical plan or be based on the spending of peer countries that have achieved sufficient funding. If a government commits to such a target through a national development plan or through other means, it arms data producers during later budget negotiations.

Another way to implement stable and transparent government financing is to insert a line item in the national budget dedicated to the NSO. The absence of such a budget line has been a problem for even high-income countries. For example, the European Union's statistical agency, Eurostat, recently saw its budget line merged into an overarching digitization and modernization budget, raising fears that funding for statistical needs could be at risk. Conversely, one of the biggest steps in ensuring the independence of the United Kingdom's Office for National Statistics in 2007 was giving the office authority over how it uses its budget. Similarly, data-producing ministries and other government agencies could each receive a designated budget line for the production, processing, management, and safe sharing of the administrative data they produce. Other investment priorities should be closing existing coverage gaps in vital statistics and other registers and including populations that are hard to reach.

The Ministry of Finance has a special role to play as the most influential actor in budget negotiations for government-financed producers of public intent data. Recurrent engagement with, and consequently systematic use of, public intent data and official statistics by the Ministry of Finance is also likely to improve the funding for data producers and the NSO in particular.[124] As documented in the examples earlier in this chapter, it is important that the Ministry of Finance understand that investing in data may improve budgets through increased revenue collection and elimination of duplication in beneficiaries, among other things.

Stable government financing can also be secured by ensuring that data play a role in government programs and projects. When government projects have numerical targets, data management and data analysis are a must. Where relevant, the legislature could require that government program budgets be supported or justified by evidence, necessitating the use of data and therefore funding for data. Although linking funding for data to monitoring government targets may also create disincentives in producing accurate data, resisting such disincentives must be at the core of an NSO's mission to ensure credibility of and trust in official statistics.

Sometimes the financing for data is sufficient, but the resources need to be better spent. Government funding of citizen-generated data, for example, can complement that of other public intent data and be a less costly alternative. But doing so requires that civil society data platforms have sufficient capabilities and resources for community outreach, coordination, monitoring of data collection, and quality assessments of the data.[125]

### Technical capacity needs: Investing in human capital for production of public intent data

Once more and better funding is provided, investment in technical capacity is a top priority. Such an

effort should start with human capital: investing in statisticians, data scientists, and applied economists across the public sector and in data literacy in the population at large. These investments would promote demand for public intent data and bolster the credibility of and trust in public intent data producers. These goals could be achieved through a combination of education and training initiatives.

Meanwhile, the public sector at large and the NSO in particular should seek qualified statisticians, applied economists, and data scientists. One way of doing so is to create a governmentwide approach to the formulation of salary scales and renumeration of the positions across the public sector, including in the NSO, the central bank, and line ministries. Such an approach could minimize differentials in salary scales within and across government agencies and subsequently create an opportunity to adopt more competitive salary scales to attract and retain talent.

A strategic area in which NSO capabilities in low- and middle-income countries could be strengthened is research on the development of improved methods and standards for data production. The capacity to conduct such methodological research is critical to improving the availability, quality, and usability of public intent data. NSOs could establish a business line on experimental statistics, which may serve as an avenue for participating in cutting-edge, multidisciplinary research efforts centered on integrating public intent and private intent data. Low-capacity NSOs, however, will have to strengthen, and in certain cases create, capabilities in data science and geographic information systems. Twinning arrangements between NSOs with established programs on experimental statistics and those beginning to build these capabilities may be one way to accelerate progress. These activities are also aligned with the call for international organizations to sustain investments in the search for improved methods of data collection, curation, and analysis (see spotlight 2.2).

Beyond NSOs, data-related capabilities in ministries and other government agencies are often insufficient. They could remedy the situation by first creating technical units in charge of data production, processing, management, and dissemination to improve data quality. These units could also develop ministry-specific action plans for capacity building, and should be empowered by receiving the financial, technological, and human resources they need to fulfill their mandated roles in the national data system. Their goal would be delivery of high-quality knowledge disseminated in accordance with a ministry-specific public release calendar.

Capacity building should also be pursued in a country's education system.[126] In line with the aspirations of SDG Target 4.6, primary and secondary educational institutions should elevate foundational numeracy and statistical literacy skills so that, like general literacy, they are part of the fundamental curricula. These skills would empower an informed public of data users and create a pool of potential candidates for specialized data professions. In tertiary education and data-driven academic fields, advanced education on statistics should be enhanced in ways that equip future technocrats with data skills that meet policy makers' demands.

An example at the country level is Politeknik Statistika, a highly selective university established by Statistics Indonesia in 1958. Politeknik Statistika awards bachelor's degrees, with an emphasis on applied training in official statistics, in preparation for statistical careers at Statistics Indonesia and the public sector at large. Examples at the regional level include the Ecole Nationale Supérieure de Statistique et d'Economie Appliquée (ENSEA) in Côte d'Ivoire and the Eastern Africa Statistical Training Center (EASTC) in Tanzania.

Degree and certificate programs with a data science theme, including those offered online, can facilitate development of statistical capacity in techniques that cut across statistics and computer science, such as artificial intelligence and machine learning. A noteworthy example is the Think Data Science Program that was launched in 2019 by the Palestinian Central Bureau of Statistics (PCBS), in partnership with the Arab American University in West Bank and Gaza. As part of this program, students have to complete a graduation project at the PCBS, which gets accredited by the Ministry of Higher Education.

Finally, investments in human capital should be accompanied by investments in physical infrastructure, IT platforms, and software capabilities (see chapter 5).

### Governance needs: Making laws and regulations conducive to production and use of quality data

Effective use of public intent data depends on having in place a governmentwide national data strategy or another high-level document that outlines the roles, responsibilities, and mandates of various government agencies. Such arrangements are discussed in detail in chapters 8 and 9.

The NSO must be truly independent, impartial, and nonpolitical. Its independence should be

anchored in laws and an institutional setup that curtails political interference in official statistics and other public data products.[127] Debatable is whether placement of the NSO under the executive branch of government leaves it open to attacks on its independence. On the one hand, it is important that the NSO be positioned to inform public debate and policy. But this may be difficult to achieve if the NSO is administratively separated from other parts of the government and does not maintain a close relationship with influential ministries such as the Ministry of Finance or Treasury and the Ministry of Commerce or Industry. On the other hand, reporting to a specific ministry or an individual as part of the executive branch leaves the NSO vulnerable to being questioned, pressured, or otherwise influenced in its involvement with politically sensitive statistical activities.

Another way to safeguard against the politicization of data is by making deidentified public intent datasets publicly available and accessible. Ensuring the creation and dissemination of deidentified public intent datasets is partly a political task and partly a technical one.

On the political front, the NSO and other government agencies must promote open data for development. These agencies should ensure that statistical laws and regulations permit the public dissemination of deidentified public intent data—both aggregated data and microdata. They should also actively engage with data users to cultivate a shared understanding of the value of reusing open data for research and for design and evaluation of public policy. Administrative data in particular are often not accessible beyond the ministry collecting the data.

On the technical front, safeguarding the confidentiality of subjects of public intent data production is an unconditional requirement. Confidential data include both personally identifiable information and the geographic coordinates of data subjects, including communities, households, facilities, and establishments. Although best practices, standards, and tools for microdata deidentification are available,[128] the risk of disclosure is increasing with enhancements in the interoperability of public intent data. These trends call for continued improvement of deidentification techniques. Building capabilities within technical units of ministries and NSOs in the use of analytical tools to remove sensitive information, spatially deidentify microdata, and deal responsibly with the risk of disclosure will also foster a better culture of open data.

### Data demand needs: Expanding the use of public intent data

The precondition for the widespread use and reuse of data is greater data literacy among the citizenry at large and government decision-makers. The integrity of and public trust in official statistics are also critical to the demand for data. The integrity of official statistics is closely tied to the perceived independence and trustworthiness of the NSO. Existing best practices can ensure integrity of and trust in the computation of official statistics and the timing of their release, even in the face of political pressures. A first set of practices centers on effective outreach and communication about NSO products. These practices include publishing a release calendar and providing a public explanation of potential deviations from release dates, as well as publicly disseminating meticulous documentation and metadata allowing findings to be replicated. Other best practices include refraining from participating in national politics and carrying out periodic outreach efforts to cultivate public understanding and acceptance of the importance of an independent statistical agency.

NSOs could also increase use of and demand for their data by engaging proactively with and listening to stakeholders in government, academia, the private sector, CSOs, and the media.[129] These engagements may have multiple objectives such as disseminating statistical outputs, understanding and responding to user needs, exploring links between NSO data products and other data, and strengthening statistical literacy. Statistics Canada, Statistics Indonesia, and Mexico's National Institute of Statistics and Geography (INEGI) have engaged in recurrent training of journalists in print, radio, television, and digital media outlets on official statistics. INEGI has expanded its work program on data and statistics related to crime and victimization in Mexico in response to the growing demands from policy makers and data users. Elsewhere, the Palestinian Central Bureau of Statistics regularly disseminates official statistics on important international observances, such as International Workers' Day and International Women's Day.

Closely involving civil society in the use and production of data is critical. This involvement can be achieved by establishing advisory boards composed of independent technical experts who can help prepare national statistical strategies in view of the needs of all users—not only the needs of various government agencies.

Data visualization is another way in which NSOs could increase the reach of official statistics and the public's understanding of them. It does little good to achieve greater mastery of advanced analytics without also ensuring that the policy makers designing and enacting interventions that improve lives understand and appreciate the value added by data.[130] Distilling complex phenomena into compelling visuals and narratives for broad audiences is a timeless idea that can effectively influence public debate and policy making (for a pioneering example, see box 1.2 in chapter 1).

From the perspective of government ministries and agencies, one way to jump-start data use in planning and policy making is through the institutionalized adoption of data-intensive management practices. In Rwanda, as part of the nationwide Imihigo performance contracts launched in 2006, mayors commit to setting development targets. Each target is subsequently evaluated and ranked by the national government with respect to its achievement and whether it was monitored appropriately.[131] Management of these contracts not only requires large amounts of data to evaluate performance, but, more important, puts data on development outcomes at the center of the policy discourse.[132]

When low data literacy or appreciation of data are barriers to their use, knowledge brokers can facilitate data use in the public sector. A knowledge broker points policy makers to the relevant data and creates value through collaboration.[133] The important role of knowledge brokers is highlighted by evidence from a survey conducted by AidData: policy makers reported that they learn about sources of data primarily through personal interactions.[134]

The role of knowledge broker can be fulfilled by government officials and by outsiders. Central analytical units and technical staff in line ministries can serve as intermediaries for NSOs seeking to reach senior officials and increase data use, presenting data in both technical and nontechnical ways tailored to the needs of decision-makers.[135] Another useful technique is joint analytical exercises by the government and researchers. Collaboration between external researchers and policy makers is a major facilitator of the use of evidence and data.[136]

If governments address these financing, human capital, governance, and data demand roadblocks, the value of public intent data can be maximized. Chapter 9 discusses sequencing of the required government interventions, placing such activities within an integrated national data system. Another way in which data can lead to better lives is via the private sector. That is the topic of the next chapter.

## Notes

1. Unfortunately, in some contexts this scenario is not too far from reality. For example, Das and Hammer (2007) found that doctors in New Delhi often perform only a fraction of the recommended examinations and tests when patients present with common yet dangerous health conditions.
2. As just one example, in Ethiopia a 2016 study by Rogger and Somani (2018) surveying 1,831 officials of 382 organizations spanning three tiers of government revealed officials' significant lack of knowledge about their area of work. Half thought that their district's population was at least 50 percent larger or smaller than it was. Government staff in the educational sector were on average 38 percent off when estimating primary enrollment figures.
3. Cameron et al. (2019).
4. See chapter 1 for more information on the distinction between public intent data and private intent data, chapter 3 for a discussion of private intent data, and chapter 4 for a discussion of how both kinds of data can be repurposed.
5. See Jolliffe et al. (forthcoming) for a lengthier discussion of these 12 features of public intent data and examples of how they can generate value for development.
6. World Bank (2018d).
7. SDSN TReNDS (2018b); SSEE (2014).
8. Hallegatte et al. (2017).
9. SDSN TReNDS (2018a).
10. J-PAL (2018).
11. Hjort et al. (2019).
12. Arezki et al. (2020), for example, show that imprecise definitions of employment in the Middle East and North Africa blur the lines between unemployment and informality and distort the role of women and rural areas in national labor markets.
13. The World Bank LSMS team provided the number of countries in which LSMS-supported survey data production took place from 2011 to 2020.
14. Abay et al. (2019); Arthi et al. (2018); Carletto, Gourlay, and Winters (2015); Carletto, Savastano, and Zezza (2013); Carletto et al. (2017); De Weerdt, Gibson, and Beegle (2019); Desiere and Jolliffe (2018); Dillon et al. (2019); Gaddis et al. (2019); Gourlay, Kilic, and Lobell (2019); Kilic et al. (2017, 2018).
15. World Bank (2016a, 2017a).
16. The EU uses a Nomenclature of Territorial Units for Statistics, known as NUTS, for the purpose of allocating funds. Many EU countries have a hierarchy of three NUTS levels. The second level, NUTS 2, is used for allocation of funds. In the accompanying text, NUTS 2 is referred to as areas.
17. Government of Croatia (2019).
18. Open Data Watch (2015b).
19. Data2x (2019).

20. McCluskey and Huang (2019) and unpublished notes shared with the WDR 2021 team. The 30 percent refers to own-source revenue collection—that is, the part of the revenue collection that the cities themselves oversee in contrast to revenue they receive from national authorities and more.
21. World Bank (2018c). See also World Bank (2020b).
22. Roseth, Reyes, and Amézaga (2019) and references cited therein provide evidence of an up-to-date census generating savings to the government many times its cost. The value of public intent data to the private sector is discussed in spotlight 3.1 and elsewhere.
23. The SDG on clean water and sanitation relies on a mix of household surveys, population and housing censuses, and administrative data (SDSN 2015). Earth observation data are used for the SDGs on sustainable cities and communities, life below water, life on land, and more (Anderson et al. 2017). Citizen-generated data are often used when government data are missing and to verify government data (Lämmerhirt et al. 2018). In the Philippines, for example, the NSO identified more than 80 relevant SDG indicators where data are missing and CSOs could provide inputs through community-based monitoring systems (PARIS21 and PSA 2020).
24. MPPN (2017).
25. Yin et al. (2020). See World Air Quality Index Project, World's Air Pollution: Real-Time Air Quality Index (database), https://waqi.info/, and OpenAQ, OpenAQ (database), https://openaq.org/, for publicly available, real-time data from air quality monitoring stations around the globe, including those in China.
26. Open Data Watch (2015a).
27. Bosio and Djankov (2020).
28. AFIC (2018); GPSA (2020).
29. World Bank (2017c).
30. Callen et al. (2019). Petrov, Gurin, and Manley (2016) and Verhulst and Young (2017) contain many other examples and channels through which open data may lead to better development outcomes.
31. Safetipin (2016).
32. Fabregas, Kremer, and Schilbach (2019).
33. For example, in several high-profile cases researchers have been able to reidentify individuals from publicly available microdata, even though the data had been published in a deidentified fashion (Heffetz and Ligett 2014).
34. Díaz (2018); Fundación Reflejos de Venezuela (2016); Privacy International (2019). LGBTQI stands for lesbian, gay, bisexual, transgender, queer (or questioning), intersex.
35. Open Data Watch, ODIN (Open Data Inventory) (database), https://odin.opendatawatch.com/; Open Knowledge Foundation, GODI (Global Open Data Index) (database), https://index.okfn.org/.
36. WDR 2021 team calculations based on 2019 Statistical Performance Indicators (World Bank, Statistical Performance Indicators [database], http://www.worldbank.org/spi). As of December 2014, 21 countries had not completed a census during the 2010 round of the population and housing census (Statistics Division, Department of Economic and Social Affairs, United Nations, World Population and Housing Census Programme [database], http://mdgs.un.org/unsd/demographic-social/census/index.cshtml).
37. See estimates in, for example, Roseth, Reyes, and Amézaga (2019) and references cited therein.
38. WDR 2021 team calculations based on 2019 Statistical Capacity Indicators (World Bank, Statistical Capacity Indicators [database], https://datatopics.worldbank.org/statisticalcapacity/SCIdashboard.aspx).
39. López-Vargas, Fuentes, and Vivar (2020).
40. WDR 2021 team calculations based on 2019 Statistical Performance Indicators (World Bank, Statistical Performance Indicators [database], http://www.worldbank.org/spi).
41. Desai, Diofasi, and Lu (2018); World Bank, Global ID4D Dataset (Identification for Development Global Dataset) (database), https://datacatalog.worldbank.org/dataset/identification-development-global-dataset.
42. United Nations (2019b).
43. Carr-Hill (2013); Randall (2015); Seltzer and Walker (2020); Toulemon (2017).
44. Carr-Hill (2013).
45. Jerven (2019).
46. Where vital registration systems function well, administrative records can be used to update census population counts. But these systems are weak in lower-income countries. Gaps in registration will likely leave out more vulnerable people than the census, especially seasonal migrants and the displaced (Dunning, Gelb, and Raghavan 2014). The use of administrative records from nongovernment actors can supplement official records.
47. Maron (2015).
48. UN Women (2018). Gender-specific SDG indicators are those that explicitly call for disaggregation by sex or that refer to gender equality as the underlying objective.
49. Doss, Kieran, and Kilic (2020). Administrative recording of land titles can serve the function of documenting asset ownership at the individual level. However, land and property ownership registries are among the less developed administrative recording systems globally. According to the Global Open Data Index, these data are available in less than one-quarter of countries, even in high-income nations, and are rarely openly available.
50. World Bank (2017b). Advances have been made in intrahousehold poverty estimation based on structural models and existing household survey data—that is, clothing expenditures for women, men, and children (Lechene, Pendakur, and Wolf 2019). The predictions provided by these models, however, have yet to be validated in the context of randomized survey experiments that collect detailed, individual-disaggregated consumption data that can, in turn, be used to compute observed estimates of intrahousehold poverty among women, men, and children. These observed estimates can, in turn, be compared with predictions stemming from structural models, based on the data elicited through prevailing approaches to household survey data collection.

51. Tiberti and Costa (2020); UN Women (2018). Similarly, individual-disaggregated data on time use are required to monitor SDG Target 5.4. Yet of the 84 countries known to have conducted time use surveys in the past, only 24 percent of them have collected data since 2010.
52. Ambler et al. (2020); Bardasi et al. (2011); Chen and Collins (2014); Deere, Alvarado, and Twyman (2012); Fisher, Reimer, and Carr (2010); Jacobs and Kes (2015); Kilic and Moylan (2016); Kilic, Moylan, and Koolwal (2020); Kilic et al. (2020).
53. See de Mel, McKenzie, and Woodruff (2009); de Nicola and Giné (2014); Gibson and Kim (2010).
54. Arthi et al. (2018); Das, Hammer, and Sánchez-Paramo (2012); Deininger et al. (2012); Gaddis et al. (2019); Kilic et al. (2018); Wollburg, Tiberti, and Zezza (2020).
55. Sandefur and Glassman (2015).
56. Sandefur and Glassman (2015).
57. Karagulian et al. (2019).
58. Antony et al. (2020); Morawska et al. (2018).
59. Similar findings were reported in Beegle et al. (2016).
60. See Hoogeveen and Pape (2020) for more information on such innovations. The last two poverty data points are comparable in 60 percent of countries in FCS and in 75 percent of low-income and Sub-Saharan African countries.
61. FMOH (2018).
62. CTO (2018).
63. Custer and Sethi (2017); Kiregyera (2017).
64. However, under secure circumstances authorized third-party researchers can be allowed to match individual-level records across multiple data sources to generate insights that rely on individual-level matching.
65. Privacy by design refers to proactively embedding privacy considerations in the design of information technology and data systems. See examples from Austria, Estonia, and India covered in *ID4D Practitioner's Guide: Version 1.0* (World Bank 2019b).
66. WDR 2021 team calculations based on the 2019 Statistical Performance Indicators (World Bank, Statistical Performance Indicators [database], http://www.worldbank.org/spi).
67. Katsimi and Moutos (2010).
68. Heffetz and Ligett (2014).
69. In particular, the team had discussions with the NSOs of Canada, Chile, Ethiopia, India, Indonesia, Mexico, the United Kingdom, and West Bank and Gaza.
70. See United Nations (2019b) for similar findings.
71. Calleja and Rogerson (2019). McQueston (2013) found similar results.
72. Dargent et al. (2020); OECD (2017).
73. United Nations (2019b).
74. UNESCO (2016); WHO (2011).
75. PARIS21 (2020).
76. Calleja and Rogerson (2019).
77. Sethi and Prakash (2018).
78. Calleja and Rogerson (2019). National governments also tend to prioritize covering ongoing expenses for collecting data over onetime investments in systems.
79. Lange (2020).
80. Calleja and Rogerson (2019); Sandefur and Glassman (2015); World Bank (2018a).
81. Sanna and McDonnell (2017).
82. Piovesan (2015).
83. Antony et al. (2020); Hosman (2014); López-Vargas, Fuentes, and Vivar (2020); Pham, Rahim, and Cousin (2016).
84. Allard et al. (2018).
85. PARIS21 (2018).
86. PARIS21 (2018).
87. Dargent et al. (2020).
88. Allard et al. (2018); Johnson, Massey, and O'Hara (2015).
89. Fu and Schweinfest (2020); UNSTATS and World Bank (2020).
90. PARIS21 (2018).
91. Lämmerhirt et al. (2018).
92. OECD (2019).
93. Allard et al. (2018).
94. Calleja and Rogerson (2019); Khan, Wales, and Stuart (2015).
95. OECD (2019).
96. Amnesty International (2019); Privacy International (2013); United Nations (2019a).
97. Van Belle et al. (2018).
98. Amnesty International (2019); Privacy International (2013); United Nations (2019a).
99. Independence of the national statistical system underpins the UN Statistical Commission's 1994 Fundamental Principles of Official Statistics. The commission highlighted its concern for independence in its 2015 *United Nations Fundamental Principles of Official Statistics: Implementation Guidelines* (UNSTATS 2015).
100. Childs et al. (2019); Taylor (2016).
101. Mo Ibrahim Foundation, Ibrahim Index of African Governance (IIAG) (database), http://mo.ibrahim.foundation/iiag/.
102. Bodin (2011); Todesca (2017); Trewin (2018); von Oppeln-Bronikowski et al. (2015).
103. Todesca (2017).
104. Hoogeveen and Nguyen (2019); Taylor (2016); World Bank (2016b, 2017c).
105. Desiere, Staelens, D'Haese (2016); Jerven (2014).
106. Brackfield (2012); World Bank (2018a).
107. Arezki et al. (2020); Cady (2005); Cady and Pellechio (2006); Kubota and Zeufack (2020).
108. Dargent et al. (2020).
109. See Reporters Wthout Borders, 2020 World Press Freedom Index (database), https://rsf.org/en/ranking_table.
110. World Bank (2016b, 2018a).
111. Klein, Galdin, and Mohamedou (2016).
112. Custer and Sethi (2017); Kiregyera (2017).
113. WDR 2021 team based on information in World Bank, "Literacy Rate, Adult Total (% of People Ages 15 and Above)," https://data.worldbank.org/indicator/se.adt.litr.zs.
114. World Bank (2017c).
115. Manning, Goldman, and Hernández Licona (2020).
116. Hoogeveen and Nguyen (2019).
117. Sethi and Prakash (2018).

118. Custer and Sethi (2017); World Bank (2018a).
119. Custer and Sethi (2017); Kiregyera (2017).
120. ITU (2016); López-Vargas, Fuentes, and Vivar (2020); Pham, Rahim, and Cousin (2016).
121. Dargent et al. (2020).
122. Dargent et al. (2020).
123. Scott (2005).
124. World Bank (2019a).
125. Lämmerhirt et al. (2018).
126. OECD (2017).
127. Bodin (2011); Todesca (2017); Trewin (2018); von Oppeln-Bronikowski et al. (2015).
128. For more information on resources and tools related to the anonymization of microdata, see World Bank and PARIS21 Consortium, Microdata Anonymization (database), International Household Survey Network, PARIS21 Consortium, https://ihsn.org/anonymization.
129. Snorrason (2018).
130. Ashby (2019).
131. World Bank (2018b).
132. Krätke and Byiers (2014).
133. Head (2016); Manning, Goldman, and Hernández Licona (2020).
134. Masaki et al. (2017).
135. Sethi and Prakash (2018).
136. Oliver et al. (2014).

## References


Abay, Kibrom A., Gashaw T. Abate, Christopher B. Barrett, and Tanguy Bernard. 2019. "Correlated Non-Classical Measurement Errors, 'Second Best' Policy Inference, and the Inverse Size-Productivity Relationship in Agriculture." *Journal of Development Economics* 139 (June): 171–84. https://doi.org/10.1016/j.jdeveco.2019.03.008.

AFIC (Africa Freedom of Information Center). 2018. "Eyes on the Contract: Citizens' Voice in Improving the Performance of Public Contracts in Uganda." 2nd Monitoring Report, AFIC, Kampala, Uganda. https://africafoicentre.org/download/eyes-on-the-contract-citizens-voice-in-improving-the-performance-of-public-contracts-in-uganda/.

Allard, Scott W., Emily R. Wiegand, Collen Schlecht, A. Rupa Datta, Robert M. Goerge, and Elizabeth Weigensberg. 2018. "State Agencies' Use of Administrative Data for Improved Practice: Needs, Challenges, and Opportunities." *Public Administration Review* 78 (2): 240–50.

Ambler, Kate, Cheryl Doss, Caitlin Kieran, and Simone Passarelli. 2020. "He Says, She Says: Exploring Patterns of Spousal Agreement in Bangladesh." *Economic Development and Cultural Change.* Published ahead of print, November 16. https://doi.org/10.1086/703082.

Amnesty International. 2019. "New Technologies and Their Impact on the Promotion and Protection of Human Rights in the Context of Assemblies: Submission to the Office of the United Nations High Commissioner for Human Rights," Amnesty International, London. https://www.amnesty.org/download/Documents/IOR4012842019ENGLISH.pdf.

Anderson, Katherine, Barbara Ryan, William Sonntag, Argyro Kavvada, and Lawrence Friedl. 2017. "Earth Observation in Service of the 2030 Agenda for Sustainable Development." *Geo-Spatial Information Science* 20 (2): 77–96.

Antony, Anish Paul, Kendra Leith, Craig Jolley, Jennifer Lu, and Daniel J. Sweeney. 2020. "A Review of Practice and Implementation of the Internet of Things (IoT) for Smallholder Agriculture." *Sustainability* 12 (9): 3750.

Arezki, Rabah, Daniel Lederman, Amani Abou Harb, Nelly Youssef, Louis William El-Mallakh, Rachel Yuting Fan, Asif Mohammed Islam, et al. 2020. "Middle East and North Africa Economic Update, April 2020: How Transparency Can Help the Middle East and North Africa." World Bank Other Operational Studies 33475, World Bank, Washington, DC.

Arthi, Vellore, Kathleen Beegle, Joachim De Weerdt, and Amparo Palacios-López. 2018. "Not Your Average Job: Measuring Farm Labor in Tanzania." *Journal of Development Economics* 130 (January): 160–72.

Ashby, Deborah. 2019. "Pigeonholes and Mustard Seeds: Growing Capacity to Use Data for Society." *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 182 (4): 1121–37.

Bardasi, Elana, Kathleen Beegle, Andrew Dillon, and Pieter Serneels. 2011. "Do Labor Statistics Depend on How and to Whom the Questions Are Asked? Results from a Survey Experiment in Tanzania." *World Bank Economic Review* 25 (3): 418–47.

Beegle, Kathleen, Luc Christiaensen, Andrew L. Dabalen, and Isis Gaddis. 2016. *Poverty in a Rising Africa.* Africa Poverty Report. Washington, DC: World Bank.

Bodin, Jean-Louis. 2011. "How to React When the Independence of Statisticians and the Integrity of Statistics Are Endangered?" *Statistical Journal of the IAOS* 27 (1–2): 59–69.

Bosio, Erica, and Simeon Djankov. 2020. "How Large Is Public Procurement?" *Let's Talk Development* (blog), February 5, 2020. https://blogs.worldbank.org/developmenttalk/how-large-public-procurement.

Brackfield, David. 2012. "OECD Work on Measuring Trust in Official Statistics." *Bulletin of the ISI 58th World Statistics Congress of the International Statistical Institute, 2011* (December 2012): 3721–26. The Hague, Netherlands: International Statistical Institute.

Buvinic, Mayra, Lorenz Noe, and Eric Swanson. 2020. "Understanding Women's and Girls' Vulnerabilities to the COVID-19 Pandemic: A Gender Analysis and Data Dashboard of Low- and Lower-Middle Income Countries." Data2x, United Nations Foundation, Washington, DC.

Cady, John. 2005. "Does SDDS Subscription Reduce Borrowing Costs for Emerging Market Economies?" *IMF Staff Papers* 52 (3): 503–17.

Cady, John, and Anthony J. Pellechio. 2006. "Sovereign Borrowing Cost and the IMF's Data Standards Initiatives." IMF Working Paper WP/06/78, International Monetary Fund, Washington, DC.

Calleja, Rachel, and Andrew Rogerson. 2019. "Financing Challenges for Developing Statistical Systems: A

Review of Financing Options." PARIS21 Discussion Paper 14, Partnership in Statistics for Development in the 21st Century, Paris.

Callen, Michael, Saad Gulzar, Ali Hasanain, Muhammad Yasir Khan, and Arman Rezaee. 2019. "Data and Policy Decisions: Experimental Evidence from Pakistan." King Center on Global Development Working Paper 1055, Stanford University, Stanford, CA.

Cameron, Grant James, Hai-Anh H. Dang, Mustafa Dinc, James Stephen Foster, and Michael M. Lokshin. 2019. "Measuring the Statistical Capacity of Nations." Policy Research Working Paper 8693, World Bank, Washington, DC.

Carletto, Calogero, Sydney Gourlay, Siobhan Murray, and Alberto Zezza. 2017. "Cheaper, Faster, and More Than Good Enough: Is GPS the New Gold Standard in Land Area Measurement?" *Survey Research Methods* 11 (3): 235–65.

Carletto, Calogero, Sydney Gourlay, and Paul Winters. 2015. "From Guesstimates to GPStimates: Land Area Measurement and Implications for Agricultural Analysis." *Journal of African Economies* 24 (5): 593–628.

Carletto, Calogero, Sydney Savastano, and Alberto Zezza. 2013. "Fact or Artifact: The Impact of Measurement Errors on the Farm Size–Productivity Relationship." *Journal of Development Economics* 103 (July): 254–61.

Carr-Hill, Roy. 2013. "Missing Millions and Measuring Development Progress." *World Development* 46 (June): 30–44.

Chen, J. Joyce, and LaPorchia A. Collins. 2014. "Let's Talk about the Money: Spousal Communication, Expenditures, and Farm Production." *American Journal of Agricultural Economics* 96 (5): 1272–90.

Childs, Jennifer Hunter, Aleia Clark Fobia, Ryan King, and Gerson Morales. 2019. "Trust and Credibility in the US Federal Statistical System." *Survey Methods: Insights from the Field,* February 22. https://surveyinsights.org/?p=10663.

CTO (Office of the Chief Technology Officer, US Department of Health and Human Services). 2018. "The State of Data Sharing at the US Department of Health and Human Services." US Department of Health and Human Services, Washington, DC. https://www.hhs.gov/sites/default/files/HHS_StateofDataSharing_0915.pdf.

Custer, Samantha, and Tanya Sethi, eds. 2017. "Avoiding Data Graveyards: Insights from Data Producers and Users in Three Countries." AidData, Global Research Institute, College of William and Mary, Williamsburg, VA.

Dang, Hai-Anh, Mustufa Dinc, Juderica Diaz, Hiroko Maeda, John Pullinger, Umar Serajuddin, Brian Stacy, et al. 2021a. "Measuring the Statistical Performance of Countries: An Overview of Updates to the World Bank Statistical Capacity Index." World Bank, Washington, DC.

Dang, Hai-Anh, John Pullinger, Umar Serajuddin, and Brian Stacy. 2021b. "Statistical Performance Index: A New Tool to Measure Country Statistical Capacity." Policy Research Working Paper, World Bank, Washington, DC.

Dargent, Eduardo, Gabriela Lotta, José Antonio Mejía-Guerra, and Gilberto Moncada. 2020. "Who Wants to Know? The Political Economy of Statistical Capacity in Latin America." Inter-American Development Bank, Washington, DC.

Das, Jishnu, and Jeffrey Hammer. 2007. "Money for Nothing: The Dire Straits of Medical Practice in Delhi, India." *Journal of Development Economics* 83 (1): 1–36.

Das, Jishnu, Jeffery Hammer, and Carolina Sánchez-Paramo. 2012. "The Impact of Recall Periods on Reported Morbidity and Health Seeking Behavior." *Journal of Development Economics* 98 (May): 76–88.

Data2x. 2019. "Data Breaks the Silence on Violence against Women: A Case Study of Vietnam." Gender Data Impact Case Study, United Nations Foundation, Washington, DC. https://data2x.org/wp-content/uploads/2020/02/Impact-Case-Studies-Vietnam-4P.pdf.

Deere, Carmen Diana, Gina E. Alvarado, and Jennifer Twyman. 2012. "Gender Inequality in Asset Ownership in Latin America: Female Owners vs Household Heads." *Development and Change* 43 (2): 505–30.

Deininger, Klaus, Calogero Carletto, Sara Savastano, and James Muwonge. 2012. "Can Diaries Help in Improving Agricultural Production Statistics? Evidence from Uganda." *Journal of Development Economics* 98 (May): 42–50.

de Mel, Suresh, David J. McKenzie, and Christopher M. Woodruff. 2009. "Measuring Microenterprise Profits: Must We Ask How the Sausage Is Made?" *Journal of Development Economics* 88 (1): 19–31.

de Nicola, Francesca, and Xavier Giné. 2014. "How Accurate Are Recall Data? Evidence from Coastal India." *Journal of Development Economics* 106 (January): 52–65.

Desai, Vyjayanti T., Anna Diofasi, and Jing Lu. 2018. "The Global Identification Challenge: Who Are the 1 Billion People without Proof of Identity?" *Voices* (blog), April 25, 2018. https://blogs.worldbank.org/voices/global-identification-challenge-who-are-1-billion-people-without-proof-identity.

Desiere, Sam, and Dean Mitchell Jolliffe. 2018. "Land Productivity and Plot Size: Is Measurement Error Driving the Inverse Relationship?" *Journal of Development Economics* 130 (January): 84–98. https://doi.org/10.1016/j.jdeveco.2017.10.002.

Desiere, Sam, Lotte Staelens, and Marijke D'Haese. 2016. "When the Data Source Writes the Conclusion: Evaluating Agricultural Policies." *Journal of Development Studies* 52 (9): 1372–87.

De Weerdt, Joachim, John Gibson, and Kathleen Beegle. 2019. "What Can We Learn from Experimenting with Survey Methods?" LICOS Discussion Paper 418, LICOS Center for Institutions and Economic Performance, Faculty of Economics and Business, Katholieke Universiteit Leuven, Leuven, Belgium.

Díaz, Marianne. 2018. "El Cuerpo Como Dato." @Derechos-Digitales América Latina, Santiago, Chile. https://www.derechosdigitales.org/wp-content/uploads/cuerpo_DATO.pdf.

Dillon, Andrew, Sydney Gourlay, Kevin McGee, and Gbemisola Oseni. 2019. "Land Measurement Bias and Its Empirical Implications: Evidence from a Validation Exercise." *Economic Development and Cultural Change* 67 (3): 595–624.

Doss, Cheryl Renee, Caitlin Kieran, and Talip Kilic. 2020. "Measuring Ownership, Control, and Use of Assets." *Feminist Economics* 26 (3): 144–68.

Dunning, Casey, Alan Gelb, and Sneha Raghavan. 2014. "Birth Registration, Legal Identity, and the Post-2015 Agenda." CGD Policy Paper 046, Center for Global Development, Washington, DC.

Fabregas, Raissa, Michael M. Kremer, and Frank Schilbach. 2019. "Realizing the Potential of Digital Development: The Case of Agricultural Advice." *Science* 366 (6471): eaay3038.

Fisher, Monica, Jeffrey J. Reimer, and Edward R. Carr. 2010. "Who Should Be Interviewed in Surveys of Household Income?" *World Development* 38 (7): 966–73.

FMOH (Federal Ministry of Health, Ethiopia). 2018. *eHealth Apps Inventory.* Addis Ababa: FMOH.

Fu, Haishan, and Stefan Schweinfest. 2020. "COVID-19 Widens Gulf of Global Data Inequality, While National Statistical Offices Step Up to Meet New Data Demands." *Data Blog,* June 5, 2020. https://blogs.worldbank.org/opendata/covid-19-widens-gulf-global-data-inequality-while-national-statistical-offices-step-up.

Fundación Reflejos de Venezuela. 2016. "El drama de ser transgénero e intentar comprar en un supermercado." Fundación Reflejos de Venezuela, Caracas.

Gaddis, Isis, Gbemisola Oseni, Amparo Palacios-López, and Janneke Pieters. 2019. "Measuring Farm Labor: Survey Experimental Evidence from Ghana." Policy Research Working Paper 8717, World Bank, Washington, DC.

Gibson, John, and Bonggeun Kim. 2010. "Non-Classical Measurement Error in Long-Term Retrospective Recall Surveys." *Oxford Bulletin of Economics and Statistics* 72 (5): 687–95.

Gourlay, Sydney, Talip Kilic, and David B. Lobell. 2019. "A New Spin on an Old Debate: Errors in Farmer-Reported Production and Their Implications for the Inverse Scale–Productivity Relationship in Uganda." *Journal of Development Economics* 141 (November): 1–35. https://www.sciencedirect.com/science/article/pii/S0304387818306588.

Government of Croatia. 2019. "Gov't Launches Changes to Country's Statistical Subdivision." News release, January 23. https://vlada.gov.hr/news/gov-t-launches-changes-to-country-s-statistical-subdivision/25178.

GPSA (Global Partnership for Social Accountability). 2020. "Making Public Contracts Work for People: Experiences from Uganda." GPSA, World Bank, Washington, DC. https://www.thegpsa.org/stories/making-public-contracts-work-people-experiences-uganda.

Hallegatte, Stéphane, Adrien Vogt-Schilb, Mook Bangalore, and Julie Rozenberg. 2017. *Unbreakable: Building the Resilience of the Poor in the Face of Natural Disasters.* Climate Change and Development Series. Washington, DC: World Bank.

Head, Brian W. 2016. "Toward More 'Evidence-Informed' Policy Making?" *Public Administration Review* 76 (3): 472–84.

Heffetz, Ori, and Katrina Ligett. 2014. "Privacy and Data-Based Research." *Journal of Economic Perspectives* 28 (2): 75–98.

Hjort, Jonas, Diana Moreira, Gautam Rao, and Juan Francisco Santini. 2019. "How Research Affects Policy: Experimental Evidence from 2,150 Brazilian Municipalities." NBER Working Paper 25941, National Bureau of Economic Research, Cambridge, MA.

Hoogeveen, Johannes, and Nga Thi Viet Nguyen. 2019. "Statistics Reform in Africa: Aligning Incentives with Results." *Journal of Development Studies* 55 (4): 702–19.

Hoogeveen, Johannes, and Utz Pape, eds. 2020. *Data Collection in Fragile States: Innovations from Africa and Beyond.* Cham, Switzerland: Palgrave Macmillan.

Hosman, Laura. 2014. "Emerging Markets: Top ICT Hardware Challenges." Inveneo, San Francisco.

ITU (International Telecommunication Union). 2016. "Harnessing the Internet of Things for Global Development." ITU, Geneva. https://www.itu-ilibrary.org/science-and-technology/harnessing-the-internet-of-things-for-global-development_pub/80d1ac90-en.

Jacobs, Krista, and Aslihan Kes. 2015. "The Ambiguity of Joint Asset Ownership: Cautionary Tales from Uganda and South Africa." *Feminist Economics* 21 (3): 23–55.

Jerven, Morten. 2014. "The Political Economy of Agricultural Statistics and Input Subsidies: Evidence from India, Nigeria, and Malawi." *Journal of Agrarian Change* 14 (1): 129–45.

Jerven, Morten. 2019. "The Problems of Economic Data in Africa." In *Oxford Research Encyclopedia of Politics*. Oxford, UK: Oxford University Press. https://doi.org/10.1093/acrefore/9780190228637.013.748.

Johnson, David S., Catherine Massey, and Amy O'Hara. 2015. "The Opportunities and Challenges of Using Administrative Data Linkages to Evaluate Mobility." *Annals of the American Academy of Political and Social Science* 657 (1): 247–64.

Jolliffe, Dean, Talip Kilic, Daniel Gerszon Mahler, and Philip Randolph Wollburg. Forthcoming. "Under What Conditions Are Data Valuable for Development?" WDR 2021 background paper, World Bank, Washington, DC.

J-PAL (Abdul Latif Jameel Poverty Action Lab). 2018. "Annual Report 2018 and a Look Ahead to 2019." J-PAL, Massachusetts Institute of Technology, Cambridge, MA. https://www.povertyactionlab.org/sites/default/files/2018-annual-report-web-ready.pdf.

Karagulian, Federico, Maurizio Barbiere, Alexander Kotsev, Laurent Spinelle, Michel Gerboles, Friedrich Lagler, Nathalie Redon, et al. 2019. "Review of the Performance of Low-Cost Sensors for Air Quality Monitoring." *Atmosphere* 10 (9): 506.

Katsimi, Margarita, and Thomas Moutos. 2010. "EMU and the Greek Crisis: The Political-Economy Perspective." *European Journal of Political Economy* 26 (4): 568–76. https://doi.org/10.1016/j.ejpoleco.2010.08.002.

Khan, Amina, Joseph Wales, and Elizabeth Stuart. 2015. "Country Priorities for Data Development: What Does History Tell Us?" Report, Overseas Development Institute, London. https://www.odi.org/sites/odi.org.uk/files/odi-assets/publications-opinion-files/9695.pdf.

Kilic, Talip, and Heather G. Moylan. 2016. "Methodological Experiment on Measuring Asset Ownership from a

Gender Perspective (MEXA)." Technical Report, World Bank, Washington, DC.
Kilic, Talip, Heather G. Moylan, John Ilukor, Clement Mtengula, and Innocent Pangapanga-Phiri. 2018. "Root for the Tubers: Extended-Harvest Crop Production and Productivity Measurement in Surveys." Policy Research Working Paper 8618, World Bank, Washington, DC.
Kilic, Talip, Heather G. Moylan, and Gayatri B. Koolwal. 2020. "Getting the (Gender-Disaggregated) Lay of the Land: Impact of Survey Respondent Selection on Measuring Land Ownership." Policy Research Working Paper 9151, World Bank, Washington, DC.
Kilic, Talip, Goedele Van den Broeck, Gayatri B. Koolwal, and Heather G. Moylan. 2020. "Are You Being Asked? Impacts of Respondent Selection on Measuring Employment." Policy Research Working Paper 9152, World Bank, Washington, DC.
Kilic, Talip, Alberto Zezza, Calogero Carletto, and Sara Savastano. 2017. "Missing(ness) in Action: Selectivity Bias in GPS-Based Land Area Measurements." *World Development* 92 (April): 143–57.
Kiregyera, Ben. 2017. "Supporting Implementation of Fundamental Principles of Official Statistics in the African Region." *Statistical Journal of the IAOS* 33 (4): 863–67.
Klein, Thilo, Anaïs Galdin, and El Iza Mohamedou. 2016. "An Indicator for Statistical Literacy Based on National Newspaper Archives." Paper presented at International Association of Statistics Education's 2016 Roundtable Conference, "Promoting Understanding of Statistics about Society," Berlin, July 19–22, 2016. http://iase-web.org/Conference_Proceedings.php?p=Promoting_Understanding_of_Statistics_about_Society_2016.
Krätke, Florian, and Bruce Byiers. 2014. "The Political Economy of Official Statistics: Implications for the Data Revolution in Sub-Saharan Africa." PARIS21 Discussion Paper 5, Partnership in Statistics for Development in the 21st Century, Paris. http://ecdpm.org/wp-content/uploads/DP-170-Political-Economy-Official-Statistics-Africa-December-2014.pdf.
Kubota, Megumi, and Albert Zeufack. 2020. "Assessing the Returns on Investment in Data Openness and Transparency." Policy Research Working Paper 9139, World Bank, Washington, DC.
Lämmerhirt, Danny, Jonathan Gray, Tommaso Venturini, and Axel Meunier. 2018. "Advancing Sustainability Together? Citizen-Generated Data and the Sustainable Development Goals." Global Partnership for Sustainable Development Data, United Nations, New York. http://www.data4sdgs.org/resources/advancing-sustainability-together-citizen-generated-data-and-sustainable-development.
Lange, Simon. 2020. "Key Trends in Development Cooperation for National Data and Statistical Systems." OECD Development Policy Paper 31, Organisation for Economic Co-operation and Development, Paris.
Lechene, Valérie, Krishna Pendakur, and Alex Wolf. 2019. "OLS Estimation of the Intra-Household Distribution of Consumption." IFS Working Paper W19/19, Institute for Fiscal Studies, London.
López-Vargas, Ascensión, Manuel Fuentes, and Marta Vivar. 2020. "Challenges and Opportunities of the Internet of Things for Global Development to Achieve the United Nations Sustainable Development Goals." *IEEE Access* 8: 37202–13.
Manning, Richard, Ian Goldman, and Gonzalo Hernández Licona. 2020. "The Impact of Impact Evaluation." UNU-WIDER Working Paper 2020/20, United Nations University–World Institute for Development Economics Research, Helsinki.
Maron, Mikel. 2015. "How Complete Is OpenStreetMap?" *Mapbox* (blog), November 19, 2015. https://blog.mapbox.com/how-complete-is-openstreetmap-7c369787af6e.
Masaki, Takaaki, Samantha Custer, Agustina Eskenazi, Alena Stern, and Rebecca Latourell. 2017. "Decoding Data Use: How Do Leaders Source Data and Use It to Accelerate Development?" AidData, Global Research Institute, College of William and Mary, Williamsburg, VA.
McCluskey, William, and Chyi-Yun Huang. 2019. "The Role of ICT in Property Tax Administration: Lessons from Tanzania." CMI Brief 6, Chr. Michelsen Institute, Bergen, Norway.
McQueston, Kate. 2013. "Autonomy, Independence, and Capacity of National Statistics Offices." Background paper, Center for Global Development, Washington, DC; African Population and Health Research Council, Nairobi.
Meijer, Albert, and Suzanne Potjer. 2018. "Citizen-Generated Open Data: An Explorative Analysis of 25 Cases." *Government Information Quarterly* 35 (4): 613–21.
Morawska, Lidia, Phong K. Thai, Xiaoting Liu, Akwasi Asumadu-Sakyi, Godwin Ayoko, Alena Bartonova, Andrea Bedini, et al. 2018. "Applications of Low-Cost Sensing Technologies for Air Quality Monitoring and Exposure Assessment: How Far Have They Gone?" *Environment International* 116 (July): 286–99.
MPPN (Multidimensional Poverty Peer Network). 2017. "Using the MPI to Determine National Budgets in Costa Rica." *Dimensions* 4 (August): 14–18, Oxford Poverty and Human Development Initiative, Oxford Department of International Development, University of Oxford, Oxford, UK. https://www.mppn.org/wp-content/uploads/2017/08/Dim_4_ENGLISH_online.pdf.
OECD (Organisation for Economic Co-operation and Development). 2017. *Development Co-operation Report 2017: Data for Development*. Paris: OECD.
OECD (Organisation for Economic Co-operation and Development). 2019. *The Path to Becoming a Data-Driven Public Sector*. OECD Digital Government Studies Series. Paris: OECD. https://doi.org/10.1787/059814a7-en.
Oliver, Kathryn, Simon Innvar, Theo Lorenc, Jenny Woodman, and James Thomas. 2014. "A Systematic Review of Barriers to and Facilitators of the Use of Evidence by Policymakers." *BMC Health Services Research* 14 (January 3), article 2. https://doi.org/10.1186/1472-6963-14-2.
Open Data Watch. 2015a. "Breathe Deep: Air Quality Reform in China." Data Impacts Case Studies, Open Data Watch, Washington, DC. https://dataimpacts.org/project/data-help-china-breath-better/.

Open Data Watch. 2015b. "Disaggregated Data: Impacts of Demographic and Health Surveys." Data Impacts Case Studies, Open Data Watch, Washington, DC. https://dataimpacts.org/project/health-surveys/.

PARIS21 (Partnership in Statistics for Development in the 21st Century). 2018. "Survey Results: New Approaches to Capacity Development and Future Priorities, CD4.0 Survey." PARIS21, Paris. https://paris21.org/capacity-development-40/cd40-survey.

PARIS21 (Partnership in Statistics for Development in the 21st Century). 2020. "Partner Report on Support to Statistics: PRESS 2019." PARIS21, Paris. https://paris21.org/sites/default/files/inline-files/PARIS21_Press%202019_WEB.pdf.

PARIS21 (Partnership in Statistics for Development in the 21st Century) and PSA (Philippine Statistics Authority). 2020. *Use of Citizen-Generated Data for SDG Reporting in the Philippines: A Case Study.* Paris: PARIS21. https://paris21.org/sites/default/files/inline-files/PSA-report-FINAL.pdf.

Petrov, Oleg, Joel Gurin, and Laura Manley. 2016. "Open Data for Sustainable Development." Connections: Transport and ICT, Note 2016-5, World Bank, Washington, DC.

Pham, Congduc, Abdur Rahim, and Philippe Cousin. 2016. "Low-Cost, Long-Range Open IoT for Smarter Rural African Villages." In *2016 IEEE International Smart Cities Conference (ISC2),* edited by Institute of Electrical and Electronics Engineers, 512–17. Red Hook, NY: Curran Associates.

Piovesan, Federico. 2015. "Statistical Perspectives on Citizen-Generated Data." DataShift, Civicus, Johannesburg. http://civicus.org/thedatashift/wp-content/uploads/2015/07/statistical-perspectives-on-cgd_web_single-page.pdf.

Privacy International. 2013. "Biometrics: Friend or Foe of Privacy?" Privacy International, London. https://privacyinternational.org/news-analysis/1409/biometrics-friend-or-foe-privacy.

Privacy International. 2019. "Submission to the Special Rapporteurship on Economic, Social, Cultural, and Environmental Rights of the Inter-American Commission on Human Rights Regarding the Situation of Economic, Social, Cultural, and Environmental Rights in the Region." TEDIC, InternetLab, Derechos Digitales, Fundación Karisma, Dejusticia, Asociación por los Derechos Civiles, and Privacy International, Privacy International, London.

Randall, Sara. 2015. "Where Have All the Nomads Gone? Fifty Years of Statistical and Demographic Invisibilities of African Mobile Pastoralists." *Pastoralism* 5, article 22.

Rogger, Daniel Oliver, and Ravi Somani. 2018. "Hierarchy and Information." Policy Research Working Paper 8644, World Bank, Washington, DC.

Roseth, Benjamin, Angela Reyes, and Karla Yee Amézaga. 2019. "The Value of Official Statistics: Lessons from Intergovernmental Transfers." IDB Technical Note 1682, Inter-American Development Bank, Washington, DC. https://publications.iadb.org/publications/english/document/The_Value_of_Official_Statistics_Lessons_from_Intergovernmental_Transfers_en.pdf.

Safetipin. 2016. "Bogota: A Safety Analysis Report." Safetipin, Gurgaon, India. https://safetipin.com/report/bogota-report-2016/.

Sandefur, Justin, and Amanda Glassman. 2015. "The Political Economy of Bad Data: Evidence from African Survey and Administrative Statistics." *Journal of Development Studies* 51 (2): 116–32.

Sanga, Dimitri. 2013. "The Challenges of the Narrative of African Countries' Development: Data Demand and Supply Mismatches." Paper presented at conference "African Economic Development: Measuring Success and Failure," School for International Studies, Simon Fraser University, Vancouver, Canada, April 18–20, 2013.

Sanna, Valentina, and Ida McDonnell. 2017. "Data for Development: DAC Member Priorities and Challenges." OECD Development Co-Operation Working Paper 35, OECD, Paris.

Scott, Christopher. 2005. "Measuring Up to the Measurement Problem: The Role of Statistics in Evidence-Based Policy-Making." Partnership in Statistics for Development in the 21st Century, Paris. https://paris21.org/sites/default/files/MUMPS-full.pdf.

SDSN (United Nations Sustainable Development Solutions Network). 2015. "Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development." SDSN, New York. https://sustainabledevelopment.un.org/content/documents/2017Data-for-Development-Full-Report.pdf.

SDSN TReNDS (United Nations Sustainable Development Solutions Network–Thematic Research Network on Data and Statistics). 2018a. "Data Sharing via SMS Strengthens Uganda's Health System: A Case Study of mTRAC, Uganda." SDSN TReNDS, New York. http://www.data4sdgs.org/sites/default/files/2018-09/mTRAC%20CaseStudy_FINAL.pdf.

SDSN TReNDS (United Nations Sustainable Development Solutions Network–Thematic Research Network on Data and Statistics). 2018b. "Handpump Data Improves Water Access." SDSN TReNDS, New York. http://www.data4sdgs.org/sites/default/files/2018-12/Smart%20Handpump%20Case%20Study.pdf.

Seltzer, Judith, and Deborah Klein Walker. 2020. "Counting Children in the US 2020 Census: Assure Our Future Is Represented." *NAM Perspectives* Commentary, National Academy of Medicine, Washington, DC. https://doi.org/10.31478/202003d.

Sethi, Tanya, and Mihir Prakash. 2018. "Counting on Statistics: How Can National Statistical Offices and Donors Increase Use?" AidData, Global Research Institute, College of William and Mary, Williamsburg, VA. https://www.aiddata.org/publications/counting-on-statistics.

Snorrason, Hallgrímur. 2018. "Securing the Independence of Official Statistics: Introductory Remarks." *Statistical Journal of the IAOS* 34 (2): 145–47.

SSEE (Smith School of Enterprise and the Environment). 2014. "From Rights to Results in Rural Water Services: Evidence from Kyuso, Kenya." Water Programme Working Paper 1, SSEE, University of Oxford, Oxford, UK. https://www.smithschool.ox.ac.uk/publications/reports/SSEE-rights-to-results_final_March2014.pdf.

Taylor, Matthew. 2016. "The Political Economy of Statistical Capacity: A Theoretical Approach." IDB Discussion Paper IDP-DP-471, Inter-American Development Bank, Washington, DC.

Tiberti, Marco, and Valentina Costa. 2020. "Disability Measurement in Household Surveys: A Guidebook for Designing Household Survey Questionnaires." LSMS Guidebook, Living Standards Measurement Study, World Bank, Washington, DC. http://documents.worldbank.org/curated/en/456131578985058020/Disability-Measurement-in-Household-Surveys-A-Guidebook-for-Designing-Household-Survey-Questionnaires.

Todesca, Jorge A. 2017. "Political Power and the Argentine Statistical System: The Case of INDEC1." *Statistical Journal of the IAOS* 33 (4): 875–83.

Toulemon, Laurent. 2017. "Undercount of Young Children and Young Adults in the New French Census." *Statistical Journal of the IAOS* 33 (2): 311–16.

Trewin, Dennis. 2018. "What Does an Independent Official Statistical Agency Mean in Practice?" *Statistical Journal of the IAOS* 34 (2): 165–69.

UNESCO (United Nations Educational, Scientific, and Cultural Organization). 2016. "Education 2030: Incheon Declaration and Framework for Action." Document ED-2016/WS/28, Institute for Information Technologies in Education, UNESCO, Paris. http://uis.unesco.org/sites/default/files/documents/education-2030-incheon-framework-for-action-implementation-of-sdg4-2016-en_2.pdf.

United Nations. 2019a. "Extreme Poverty and Human Rights: Note by the Secretary-General." Document A/74/493, United Nations, New York. https://undocs.org/A/74/493.

United Nations. 2019b. *The Sustainable Development Goals Report 2019.* New York: United Nations.

UNSTATS (Statistics Division, Department of Economic and Social Affairs, United Nations). 2015. *United Nations Fundamental Principles of Official Statistics: Implementation Guidelines.* New York: United Nations. https://.unstats.org/unsd/dnss/gp/Implementation_Guidelines_FINAL_without_edit.pdf.

UNSTATS (Statistics Division, Department of Economic and Social Affairs, United Nations) and World Bank. 2020. "Monitoring the State of Statistical Operations under the COVID-19 Pandemic: Highlights from the Second Round of a Global COVID-19 Survey of National Statistical Offices." World Bank, Washington, DC.

UN Women. 2018. *Turning Promises into Action: Gender Equality in the 2030 Agenda for Sustainable Development.* New York: UN Women.

UN Women. 2020. "COVID-19 and Its Economic Toll on Women: The Story behind the Numbers." *News and Events,* September 16, 2020. https://www.unwomen.org/en/news/stories/2020/9/feature-covid-19-economic-impacts-on-women.

Van Belle, Jean-Paul. 2018. *Africa Data Revolution Report 2018: Status and Emerging Impact to Open Data in Africa.* With contributions by Danny Lämmerhirt, Carlos Iglesias, Paul Mungai, Hubeidatu Nuhu, Mbongeni Hlabano, Tarik Nesh-Nash, and Sarang Chaudhary. Washington, DC: World Wide Web Foundation. https://webfoundation.org/docs/2019/03/Africa-data-revolution-report.pdf.

Verhulst, Stefaan G., and Andrew Young. 2017. "Open Data in Developing Economies: Toward Building an Evidence Base on What Works and How." Governance Lab, New York University, New York. https://odimpact.org/files/odimpact-developing-economies.pdf.

von Oppeln-Bronikowski, Sibylle, Christine Kronz, Irina Meinke, and Hannah Wirtzfeld. 2015. "How Can Professional and Ethical Frameworks Strengthen Statisticians in Their Practical Work?" *Statistical Journal of the IAOS* 31 (4): 513–22.

WHO (World Health Organization). 2011. "The Abuja Declaration: Ten Years On." WHO, Geneva. https://www.who.int/healthsystems/publications/abuja_report_aug_2011.pdf?ua=1.

Wollburg, Philip, Marco Tiberti, and Alberto Zezza. 2020. "Recall Length and Measurement Error in Agricultural Surveys." *Food Policy.* Published ahead of print, December 1, 2020. https://doi.org/10.1016/j.foodpol.2020.102003.

World Bank. 2016a. *Croatia: Small-Area Estimation of Consumption-Based Poverty (Poverty Maps).* Washington, DC: World Bank. https://razvoj.gov.hr/UserDocsImages//Istaknute%20teme/Kartom%20siroma%C5%A1tva//Croatia%20Small-Area%20Estimation%20of%20Consumption-Based%20Poverty%20(Poverty%20Maps).pdf.

World Bank. 2016b. *World Development Report 2016: Digital Dividends.* Washington, DC: World Bank.

World Bank. 2017a. "Index of Multiple Deprivation: Conceptual Framework for Identifying Lagging Municipalities and Towns in Croatia." World Bank, Washington, DC. https://razvoj.gov.hr/UserDocsImages//Istaknute%20teme/Kartom%20siroma%C5%A1tva//Index%20of%20Multiple%20Deprivation%20-%20Conceptual%20framework_18.06.2019.pdf.

World Bank. 2017b. "Monitoring Global Poverty: Report of the Commission on Global Poverty." World Bank, Washington, DC. https://openknowledge.worldbank.org/bitstream/handle/10986/25141/9781464809613.pdf.

World Bank. 2017c. *World Development Report 2017: Governance and the Law.* Washington, DC: World Bank.

World Bank. 2018a. *Data for Development: An Evaluation of World Bank Support for Data and Statistical Capacity.* Washington, DC: Independent Evaluation Group, World Bank.

World Bank. 2018b. *Improving Public Sector Performance: Through Innovation and Inter-Agency Coordination.* Global Report: Public Sector Performance. Washington, DC: World Bank.

World Bank. 2018c. "Public Sector Savings and Revenue from Identification Systems: Opportunities and Constraints." World Bank, Washington, DC.

World Bank. 2018d. "The Role of Digital Identification for Healthcare: The Emerging Use Cases." Identification for Development (ID4D), World Bank, Washington, DC.

World Bank. 2019a. *IC4D, Information and Communication for Development 2018: Data-Driven Development.* Washington, DC: World Bank. http://documents1.worldbank.org/curated

/en/987471542742554246/pdf/128301-9781464813252.pdf.
World Bank. 2019b. *ID4D Practitioner's Guide: Version 1.0.* Washington, DC: World Bank. http://documents1.worldbank.org/curated/en/248371559325561562/pdf/ID4D-Practitioner-s-Guide.pdf.
World Bank. 2020a. "Classification of Fragile and Conflict-Affected Situations." *Brief,* July 9, World Bank, Washington, DC. https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/harmonized-list-of-fragile-situations.
World Bank. 2020b. "West Africa Unique Identification for Regional Integration and Inclusion Program, Phase 2." Project Appraisal Document, Report PAD3556, World Bank, Washington, DC.
Yin, Peng, Michael Brauer, Aaron J. Cohen, Haidong Wang, Jie Li, Richard T. Burnett, Jeffrey D. Stanaway, et al. 2020. "The Effect of Air Pollution on Deaths, Disease Burden, and Life Expectancy across China and Its Provinces, 1990–2017: An Analysis for the Global Burden of Disease Study 2017." *Lancet Planetary Health* 4 (9): 386–98.