

May 5<sup>th</sup>, 2020

# World Development Report 2021

Data for Better Lives

## Concept Note

May 2020

“The digital revolution is far more significant than the invention of writing or even printing.” – Douglas Engelbart, inventor of the computer mouse.

“You can have data without information, but you cannot have information without data.” – Daniel Keys Morton, science fiction author.

“When we exclude half of humanity from the production of knowledge we lose out on potentially transformative insights.” – Caroline Criado Pérez, *Invisible Women: Data Bias in a World Designed for Men*

“The goal is to turn data into information, and information into insight.” – Carly Fiorina, former president and chair of Hewlett-Packard Co.

## Introduction

The omnipresence of data in the daily lives of most people in the world gives rise and support to the view that data will change the world. With the unprecedented rate of data creation, and the increasing role data plays in most of our lives, it is easy to assume that the digital revolution could be the most important life-changing event of this era. But, what does it mean for the more than 700 million people in extreme poverty, living on less than \$1.90 per day? Or, the more than 900 million living without access to electricity?<sup>1</sup> Will their lives be changed despite largely not participating in the digital economy? Can data really help improve the lives of farm households in Nigeria, who are consuming much of what they produce and if they sell some, are doing so in ways that leave no digital trails?

This report will examine the enormous potential of the changing data landscape to improve the lives of poor people but which also opens back doors that can harm people, businesses and societies. It will begin by describing the ways in which data may be used more effectively to improve development outcomes through better public policies, program design and service delivery, in addition to improved market efficiency and job creation through more private sector growth. The focus of the report will be on how data can benefit poor people in poor countries. A central message will be that much of the value of data is untapped, waiting to be realized. The second part of the report will focus on issues of governance, law, policy and infrastructure that can deliver the potential benefits while safeguarding against harmful outcomes.

Most of the new and fascinating ways in which data affects the lives of many of us are linked to people being able to leverage greater value from the data they produce. The data produced by a person from their digital life can be used in innovative ways to help them. But it is not necessary for someone to be the producer of data to benefit from the data revolution. In fact, data is often collected from a small sample of people and inferences from these selected individuals can help shape policy to improve the lives of a vastly larger population, whether they were part of the sample or not. At the turn of the 19<sup>th</sup> century, the English sociologist

Seebom Rowntree interviewed a sample of families with the objective to better understand the poverty experienced not just by those he interviewed, but everyone in the town of York.<sup>2</sup> The findings from this work changed preconceptions by revealing that poverty was pervasive outside of London, and demonstrating that people cycled in and out of poverty over the course of their lives.

The findings and recommendations in this report are the fruits of an impressive information base, including knowledge gained from academic research, international development agency reports, commercial innovation and experiences. Some of the key underpinnings of it are the 2016 WDR on *Digital Dividends*,<sup>3</sup> World Bank reports on *Information and Communications for Development: Data-Driven Development*<sup>4</sup> and *Data for Development: An Evaluation of World Bank Support for Data and Statistical Capacity*,<sup>5</sup> as well as many others. In comparison to these previous reports, this WDR will focus on how data itself rather than digital technology adoption can improve the lives of poor people. It will provide a more comprehensive treatment of data issues that encompasses both traditional public intent data as well as data for private intent.

In part, this WDR will update and reinforce previous messages on how to leverage data for development purposes. But, because the world of data is changing at an incredibly fast pace, this report will also reflect new lessons learned since previous reports. The pace of change in the digital world is astounding and each day brings the world new stories of amazing advances in commerce, communication, well-being and many other aspects of our lives. Like the General Purpose Technologies (GPT) such as electricity or the steam engine, innovations emerging from the data revolution have the potential to touch all aspects of the economy. And, with our lives becoming increasingly more intertwined in the digital world, each day also brings new concerns about personal data protection, misinformation, attacks on software, networks, and data systems. Through a discussion of the numerous ways in which data can help economic development, this report aims to describe the challenges to realizing these gains, offer guidance on how to attain them, and propose safeguards for protecting citizens.

## Conceptual Framework

***This WDR centers on data. It poses two fundamental questions*** (Figure 1). First, how can data contribute to development? Part I of the report will identify the multiple channels through which data can support or inhibit the development process, providing a clear conceptual framework, together with concrete illustrations and examples from recent experience in less developed and emerging countries. Second, what kind of environment is needed to support the creation and reuse of data? Part II of the report will identify important elements that need to be in place, with particular focus on systems that safeguard data while enabling its reuse.

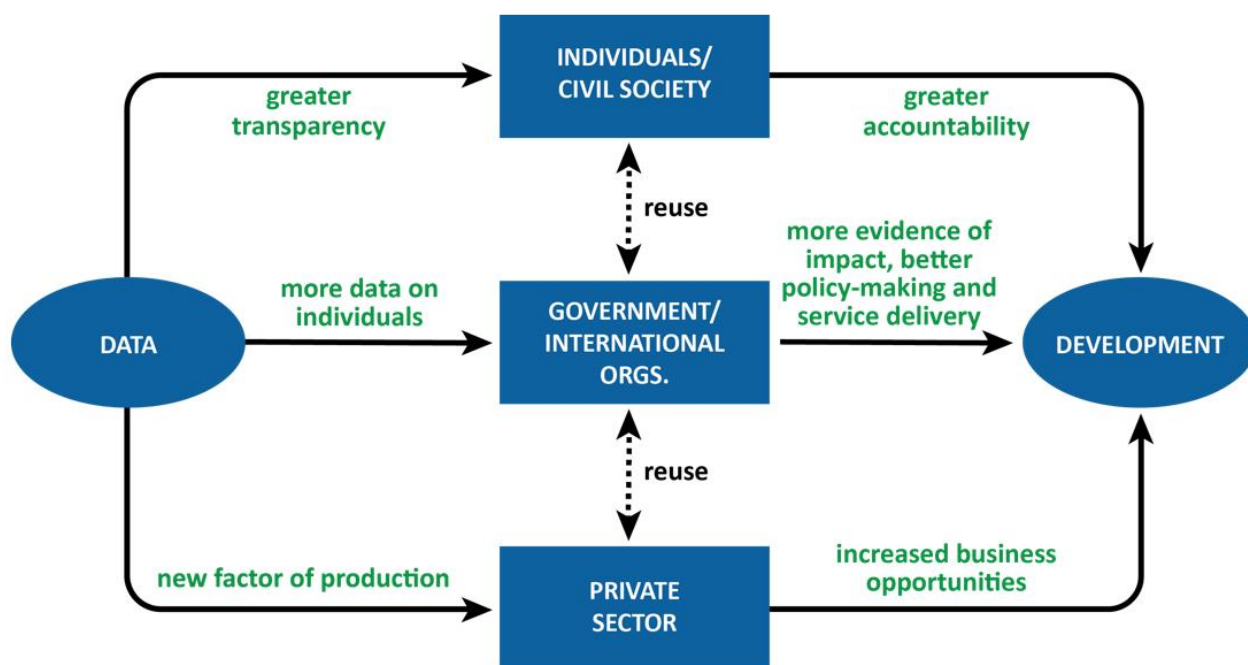
Figure 1: Creating an Environment that Supports Data for Development



***Data can contribute to development by improving the lives of the poor through multiple channels.*** The WDR will introduce a guiding conceptual framework built on three pathways through which data could foster development (Figure 2): data generated by or received by governments and international organizations to support evidence-based policymaking; data made available to civil society to monitor the effects of government policies and to individuals to empower them and enable them to access public and commercial services that are tailored to their needs; and data generated by private firms, which is a factor of production that fuels growth, but could also be mobilized and repurposed to support development objectives. It will also consider how to foster data flows across public, private and civil society channels for development impact.

***First, data enable governments to understand the impact of policies and improve service delivery.*** For traditional data types such as household and firm surveys, national accounts, and administrative data, governments (or agents authorized by governments) have been central to collecting them. Data have been collected typically for specific purposes, often intended to improve policies and foster development. However, without strong data systems in place, much of the potential for data to improve outcomes is left unrealized. Trained staff, budgetary autonomy for agencies that collect data, adequate installations, connected databases, and international partnerships are important factors in shaping successful national data systems. These resources are often scarce in low-income countries, leaving them as the least equipped to collect and effectively use the data necessary to assess and understand the scope and nature of development problems and make inroads to solving them. Enhancing the statistical capacity of client countries therefore has, and will continue to be, a point of emphasis for the World Bank Group and of this report.

Figure 2: How Data Impacts Development – Potential Positive Linkages



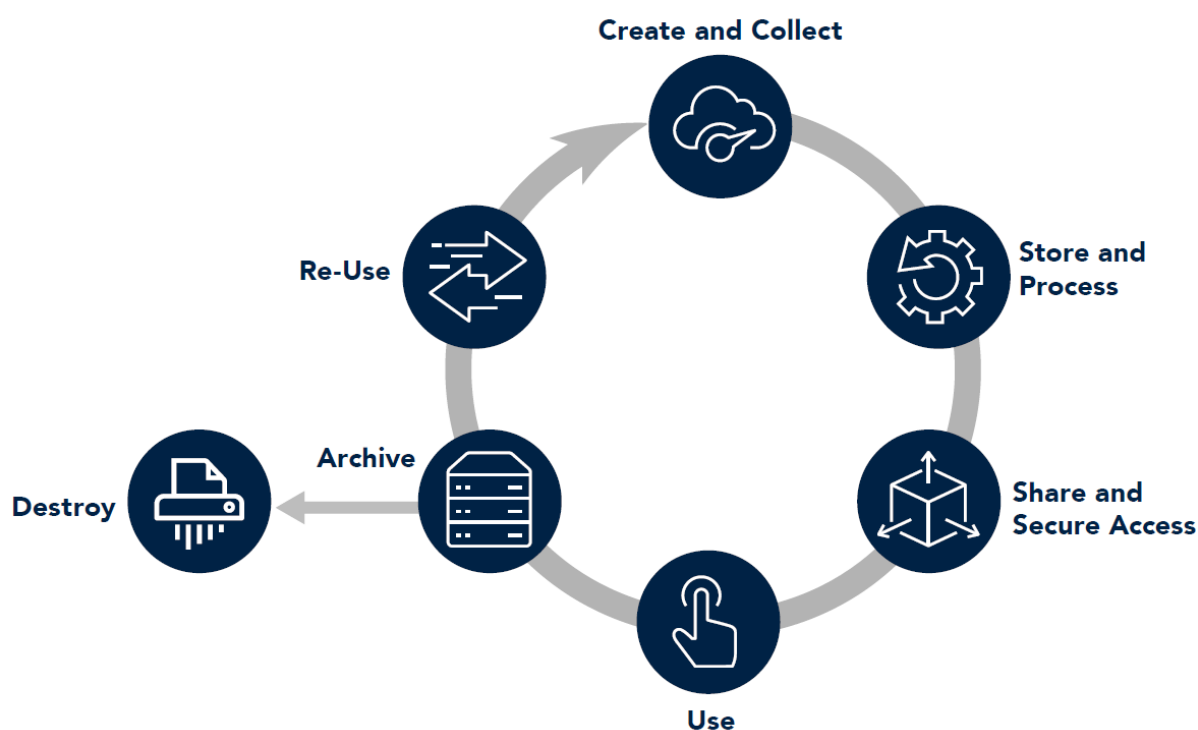
**Second, making data widely available enables civil society to hold governments accountable for policy choices.** Inputs from civil society provide a feedback mechanism through which policies can be adapted and improved. For example, by crowdsourcing information to facilitate more responsive governance, and by giving citizens the ability to conduct their own surveys and analyze the data over large numbers of people, data can foster voice, government accountability, and transparency. Simply providing individuals with better access to their own data is another way to enable citizens to advocate for themselves and improve their lives. For example, since 2010 veterans of the U.S. armed services have access to their own medical records which they can give to health care providers. By increasing transparency and control over their health data, veterans can communicate their health histories and needs to service providers more effectively than in the past, thus improving the care they receive.

**Third, data generated by the private sector holds promise for improving the lives of the poor.** With its proliferation, data have become an important factor of production for firms. Indeed, the business models of some of the world's largest firms (Amazon, Google, Facebook) are predicated on data. There are also examples of important platform business models emerging in some of the middle-income countries of the world (such as Grab in Indonesia or Mercado Libre in Latin America), and these can greatly expand market access opportunities for small and medium-sized enterprises. An example of how data-based private solutions can improve the lives of poor people is digital credit, often applied for via cell phones, which fosters financial

inclusion. Applications in healthcare that were developed by private firms such as tele-radiology, which reduces the required time to take and interpret images, especially for those in remote locations, is another. Women, especially in less developed countries, could stand to benefit most from greater financial inclusion and improved healthcare.

***There is significant potential to extract further development impact from the proliferation of data***, given that data is “non-rival,” meaning that, for example, a person’s call data records, location history, internet usage or medical records, can be used by many firms and governments at the same time and for different purposes. This reality can be illustrated through the data lifecycle (Figure 3). But to realize the potential of this attribute, data need to be made safely accessible across a wide array of users, both public and private (see Box 1 for an example of private companies’ use of public data.)

**Figure 3: The Data Lifecycle**



At the same time, ***data are not a pure public good in that they are also excludable***. And the value of data is increasing in scale. Thus, the accumulation of data by one actor, and the information asymmetries that this creates for other actors, leads to a concentration of power; whether economic or political. These inter-relationships will be explored and discussed throughout the report. Moreover, just because data, especially data collected for commercial purposes, can be reused and repurposed to inform policy and improve development outcomes, there are no guarantees that they will be.

**Enabling data reuse is therefore central to our theory of change.** Because they are likely to be at the center of efforts to re-use and share data, governments' efforts to use data to improve development outcomes are in the middle of our theory of change. The two-way arrow between private firms and government is used to indicate reuse of data originally collected for commercial purposes for public policy (and reuse of public intent data by firms). Similarly, the two-way arrow between individuals/society and governments indicates sharing and reuse of data between those parties.

**The evidence base for sharing and re-using data to improve development outcomes will be example-driven.** We know of no large, systematic bodies of data and evidence that summarize cases of re-using, repurposing, and combining different types of data to improve development outcomes at the macro level. To illustrate the experience of (and scope for) such reuse, a number of compelling examples at the micro or sector level will be developed early on in the report. Candidate examples include:

- Using geospatial location data from mobile phones, mobile call detail records (CDR) or social media (Facebook) and online search (Google) data for predicting and tracking disease outbreaks. A point of emphasis will be on how these approaches are being used to monitor and slow the spread of COVID-19, including how geospatial location data can be used to assess the effectiveness of shelter-in-place and social distancing efforts.<sup>6</sup>
- Combining private and public data sources to promote road safety and alleviate congestion. For example, 'hot-spots' can be identified by cross-referencing administrative records on accidents with crowd-sourced data on accidents/delays. Data from private driving companies such as Uber can be overlaid to understand how speed contributes to accidents, and traffic patterns can be adjusted accordingly through well-placed traffic lights, stop signs, and traffic calming.
- Using online media and user-generated content to map water/flood events in real time for water management and food security.
- Combining satellite imagery data from private and public sources to monitor crop yields and forecast malnutrition.<sup>7</sup>
- Combining satellite imagery data with traditional public data sources (census, Living Standards Measurement Surveys) for more frequent and granular estimates of poverty using small area poverty estimation techniques.
- Using georeferenced, anonymous call detail records from the months before and after an earthquake (or other similar disaster) to study disaster-induced population movements and better target service delivery.<sup>8</sup>
- Scraping social media posts for seismological activity to produce low-cost, but reliable impact profiles of seismic damage, again to better target service delivery.<sup>9</sup>

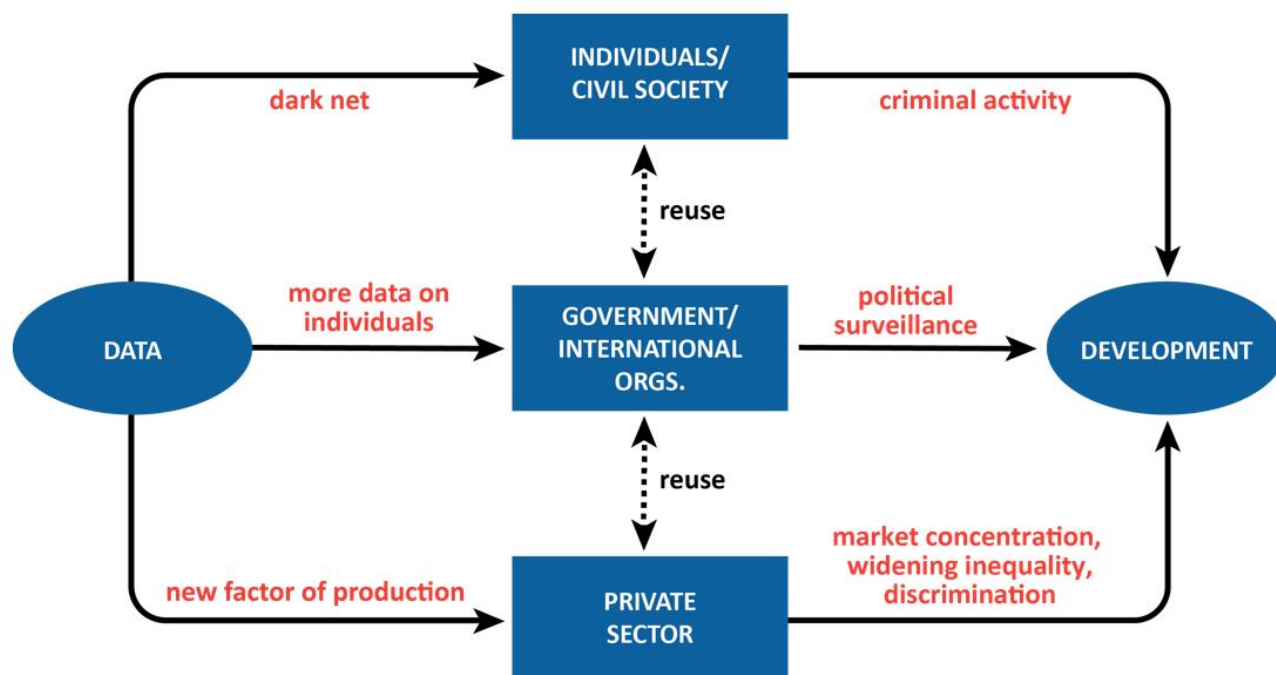
In Chapter 4, which covers synergies in public and private intent data use, we will therefore try to distill lessons from successful reuse cases and identify impediments to reuse.

***The conceptual framework is a means of organizing our thinking, but it is not intended to be a formulaic, supply-side approach to how data affect development outcomes.*** The report will therefore emphasize that there are important demand-side factors that also impede the use of data for development. For example, according to GSMA, 3.3 billion people (44% of the global population) live in areas covered by mobile broadband networks but do not use mobile internet. Understanding why they remain unconnected, and what can be done to get them connected, will be a focus of the report. There are also important institutional impediments that prevent data from positively affecting development outcomes. These, too, will be a focus of the report.

***While use of data offers great prospects for fostering development, it simultaneously poses significant risks that must be managed to avoid negative development impacts.*** Data have the potential to contribute positively to development whether through public, private or civil society channels. Some concrete (though by no means exhaustive) illustrations of such positive impacts were provided in green on Figure 1. However, it is important to recognize that the misuse of data also poses significant development risks that can manifest themselves once again through public, private and civil society channels. Again, some concrete (though by no means exhaustive) illustrations of such negative impacts are provided in red on Figure 4. Governments can potentially abuse citizens' data for political ends, such as election-rigging or politically motivated surveillance or discrimination. Individuals and organized groups can inflict considerable harm through cybercrime that steals and manipulates sensitive information. Private firms can potentially abuse consumers' data through anti-competitive practices such as algorithmic collusion. More generally, key risks that private data sources and digital applications may pose are related to market concentration and the potential for widening inequalities between advanced and developing economies. The WDR aims to give balanced consideration of both the potential positive and negative contributions that data can make to development.



Figure 4: How Data Impacts Development - Potential Adverse Outcomes



**While data can be used to improve development outcomes, the challenges differ across data types.** To fix concepts for readers and better understand those challenges, we sort data types using a two-dimensional framework (Table 1). In one dimension we think of data “stewardship” being linked to the sources of data. Data stewardship is the management and oversight of an organization's data assets to provide users with high-quality data that is easily accessible in a consistent manner. These organizations can be either public or private.<sup>10</sup> The other dimension distinguishes between “traditional” and “new” sources of data. Public stewards are typically associated with traditional data types such as censuses and surveys, though newer sources of data (e.g., from satellite imaging or e-government platforms) have become more prevalent. Private stewards are typically associated with new sources of data from digital tools and applications, though some private firms continue to conduct surveys using traditional methods.

**Table 1. Data Stewardship and Types of Data**

Types of Data		Data Stewardship	
		Public	Private
	Traditional	Census; National Accounts; Household Surveys; Enterprise Surveys; Labor force surveys; Surveys of personal finance; administrative data or records	Any survey conducted by private entities, including public opinion surveys deployed by private entities (e.g., Gallup)
New	Location data from satellite imaging; Digital identification; Face recognition from public cameras; Public procurement data from e-gov platforms	Just-in-time digital data on individual behavior/choices from digital platforms in the private sector	

Source: Adapted from inputs provided by the MENA Chief Economist Office.

***Relative to traditional public data, new private data sources offer improved timeliness, frequency, and granularity of data, but may not be representative in coverage.*** New private data can contribute significantly to addressing public sector development challenges. As alluded to above, private data collected through cell phones, internet usage, satellites, remote sensors, and other sources provide information about individuals and geographic locations that traditional surveys simply cannot. These data are collected cost-effectively, with high frequency (potentially continuously), and at fine levels of granularity (Figure 5).

***By design, traditional data collection efforts by governments are for public purposes and used to inform policy.*** But because the collection of public data via traditional methods tends to be relatively costly,<sup>11</sup> surveys are performed infrequently,<sup>12</sup> and they frequently lack the granularity necessary to make meaningful inferences about sub-populations of interest. At the same time, traditional public data offers key advantages over new private data in terms of its coverage of the population, and thus its potential to benefit more people, and its format, which makes it amenable to inferential analytics by researchers and government officials. In contrast, though it may contain a tremendous amount of information, private data that are not curated are less amenable to analysis.

**Figure 5: Characteristics of Public and Private Data**

	<b>Public/traditional</b>	<b>Private/new</b>
<b>PURPOSE</b>		
<b>Objective</b>	Typically collected for the public good, specific development objectives	Typically collected for private/commercial purposes, not for development objectives
<b>Equity</b>	Benefits spread widely	Benefits spread more narrowly
<b>Representativeness</b>	Of the population (Caveats, hard to reach: Rich, FCV, slums)	Of digital users (Caveats, hard to reach: Poor, digitally disconnected)
<b>Analytics</b>	Inferential, typically human analyst	Predictive, often machine-based (artificial intelligence, machine learning)
<b>DATA CAPTURE</b>		
<b>Timeliness</b>	Long lags	Frequently instantaneous
<b>Frequency</b>	Infrequent	Potentially continuous
<b>Granularity</b>	Coarse	Fine
<b>Collection method</b>	Often face to face/PAPI*/ higher marginal cost	Mobile broadband/remote sensing/CAPI*/ lower marginal cost
<b>*PAPI = pen-and-paper interviewing, CAPI = computer-assisted personal interviewing</b>		

**Throughout the report, we classify data based on whether the original intent was for public or commercial purposes.** Because new (and traditional) private data are collected for commercial purposes, we refer to them as data for *private intent*, recognizing this is a shorthand that may result in imperfect classification in some cases. And, similarly, we refer to data originally collected for public purposes as *public intent* data.

**Any simple framework used to classify data types carries limitations.** While public stewardship data has long been collected using traditional methods, those methods are being updated and adapted; new methods will increasingly supplement or replace traditional methods, and so the vertical dimension in Table 1 is likely to become less relevant. The distinction between public and private stewardship of data may also not be a salient one in some cases. For example, citizen-generated data – data that people or their organizations produce to directly monitor, demand or drive change on issues that affect them – can be produced through crowdsourcing mechanisms or citizen reporting initiatives, and is often organized and managed by civil society groups. The data often reside with a private steward but are clearly collected for public purposes.

***Improved inter-operability of public intent data could increase its development impact.***

Despite its advantages in coverage, suitability for some types of analysis, and potential for informing and improving policy, public intent data is often stored within countries in different government agencies and formatted in different ways. Fragmentation and incompatibilities thus limit governments' scope to use their data to the fullest extent to improve policies, service delivery, and targeting. Fostering interoperability across public intent data sources is therefore another point of emphasis throughout the report.

***Public and private intent data therefore have inherent complementarities that could be exploited to foster development.*** Combining the two types of data can advance evidence-based policy through more precise and timely official statistics, that are produced more cheaply, while preserving the representativeness that is characteristic of public intent data. For example, building on the well-established infrastructure for socioeconomic surveys collected by governments, satellite data and call data records from mobile phones offer new opportunities for updating poverty estimates for small areas more frequently. More generally, the high frequency of data collected from alternative sources for commercial purposes holds promise for producing better estimates of current socioeconomic conditions when large-scale, costly surveys such as censuses or living standards measurement surveys (LSMS) are not scheduled for several years.

***Throughout the report, separate but parallel treatment will be given to issues posed by personal and non-personal data.*** It is very important to distinguish systematically between personal and non-personal data, as these are typically generated, used and treated in very different ways. Personally identifiable data are data that in some form convey information that is specific to a known (or knowable) individual. Within the category of personal data, some types of data – for example, health histories or banking transactions – may be more sensitive than others – such as shopping records. Non-personal data are data that are generated about non-human subjects, including institutions or machines. In practice, the boundary between personal and non-personal data is becoming increasingly blurred as personal characteristics may be inferable from non-personal data, such as mobile phone records. Advances in artificial intelligence also make the anonymization of personal data more challenging and make it possible to make personal inferences from combining multiple sources of non-personal data.

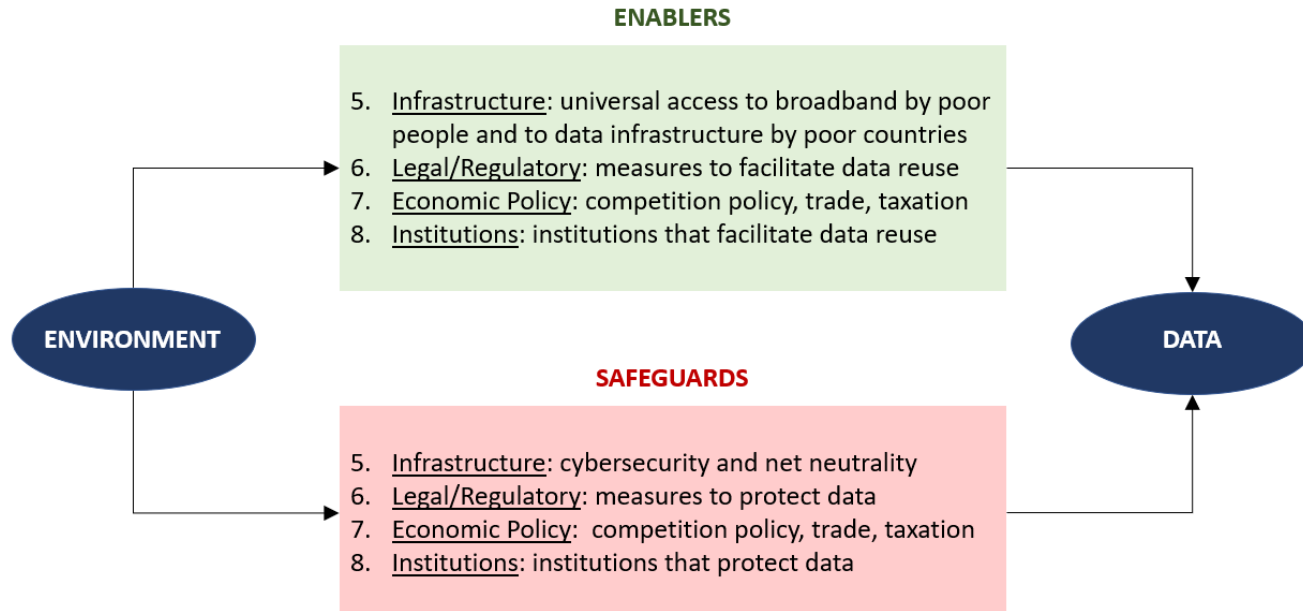
***New data sources reduce information symmetries (between firms and customers, governments and citizens), but they can also lead to asymmetries in power and opportunity.***

The timeliness, frequency, and granularity of new data mean that firms have an abundance of information about customer preferences and behavior. They can better tailor product offerings, enabling a range of transactions that would not have otherwise occurred, thus enhancing efficiency. Similarly, new data collected by governments can improve the targeting and reduce the costs of delivering services. But, because there are increasing returns to scale in data, a handful of firms with data-driven business models have grown to be among the world's largest. Such concentration of personal information in a handful of entities raises concerns about

market power and discrimination. Personal information is also concentrated in governments, which can be used to amass and maintain political power, discourage voice, and even discriminate against some population segments. A key theme throughout the report will therefore be on balancing the efficiency gains that new data bring against equity concerns.

***At the heart of these challenges is the need to develop a supportive environment for data creation and reuse; one that suitably balances “enablers” and “safeguards”.*** Harnessing the full development potential of data entails its repeated reuse to extract a wide range of different insights. This in turn rests on a transaction between the data provider and the data user that is founded on trust. Without adequate “safeguards”, the provider may lack the confidence that data can be shared without potential abuses. Such “safeguards” include data protection regulations, including the right of consent on the part of the data provider and a series of obligations on the part of the data user. They may also entail provisions for the protection of cybersecurity, or for the transparency of algorithms. At the same time, without adequate “enablers” it may become prohibitively difficult to transfer data among different providers in an agile and seamless manner.

***Establishing data governance frameworks that safeguard individual data while expanding its development benefits to many stakeholders is challenging*** (Figure 6). A first fundamental problem is the lack of data infrastructure in many less developed countries, which would enable them to efficiently collect, share, store and process data, and provide neutral access to it for all their citizens. Infrastructure design and operation is also critical to maintaining safeguards, such as cybersecurity. Second, legal and regulatory frameworks for data are seriously under-developed. These need to provide the essential foundation for data protection that safeguards rights associated with personally identifiable information, as well as intellectual property for non-personally identifiable data. They also need to incorporate provisions such as transparency, interoperability and data portability that facilitate data reuse. Third, the proliferation of data-driven business models is posing new economic policy challenges. As competition agencies grapple with the market power of globally dominant technology firms, tax authorities struggle to collect revenues from virtual companies that can readily shift corporate profits across international borders, and authorities strive to balance the demands of data protection with the opportunities for international trade in data services. Fourth, effective implementation and enforcement of safeguards and enablers requires that a suitable institutional ecosystem be in place. For each of these challenges, approaches adopted by developed countries may require considerable adaptation to fit the varying policy objectives and weaker institutional environments of less developed countries of the world.

**Figure 6: How Environment Affects Data Creation and Reuse**

The numbers 5,6,7,8 in the above image denotes chapters in the part 2 of the report.

While it would be premature to list a full set of key messages, there are a few linked themes we will weave throughout the report. One will be that ***the value of data is largely untapped***. Just as with ideas, data have the economic attribute that their value is not diminished with use and can therefore be used many times, by different people, for different purposes and not get used up.<sup>13</sup>

In addition to expanding access to data, there are also ***significant returns to improving linkages across data types and sources***. Allowing one data source to overlay another, or to make them interoperable, can add substantial value that goes beyond using each of the data sources independently. Moreover, interoperability is relevant both within and across private and public sectors. For example, a ministry of health will be able to form better public policy if it can connect its health data with data from other ministries such as education, labor, and planning. And, a private firm will be able to operate more effectively if it can link its data with other sources of information, such as satellite data on population density and socioeconomic data on wealth and wellbeing.

Another theme of the report is that ***scale matters***. While the returns to the first few bits of data are essentially zero, there is a point where the returns from additional data, and from improvements in the systems supporting these data, are substantial and increasing.<sup>14</sup> An implication of this for private-intent data is that market forces are likely to lead to data agglomeration and market concentration potentially necessitating interventions on the part of government. And indeed, the powerful economies of scale that exist in the collection and

utilization of data have led to a marked concentration of market power in data-driven businesses, which may preclude entry by small firms. The report will address how anti-trust policy needs to be modernized to address the complex competition challenges posed by platform businesses, and what kinds of data remedies may be effective in doing so. Another important aspect of benefit/revenue sharing is the issue of fair taxation of cross-border data-driven businesses, which will also be included in the report.

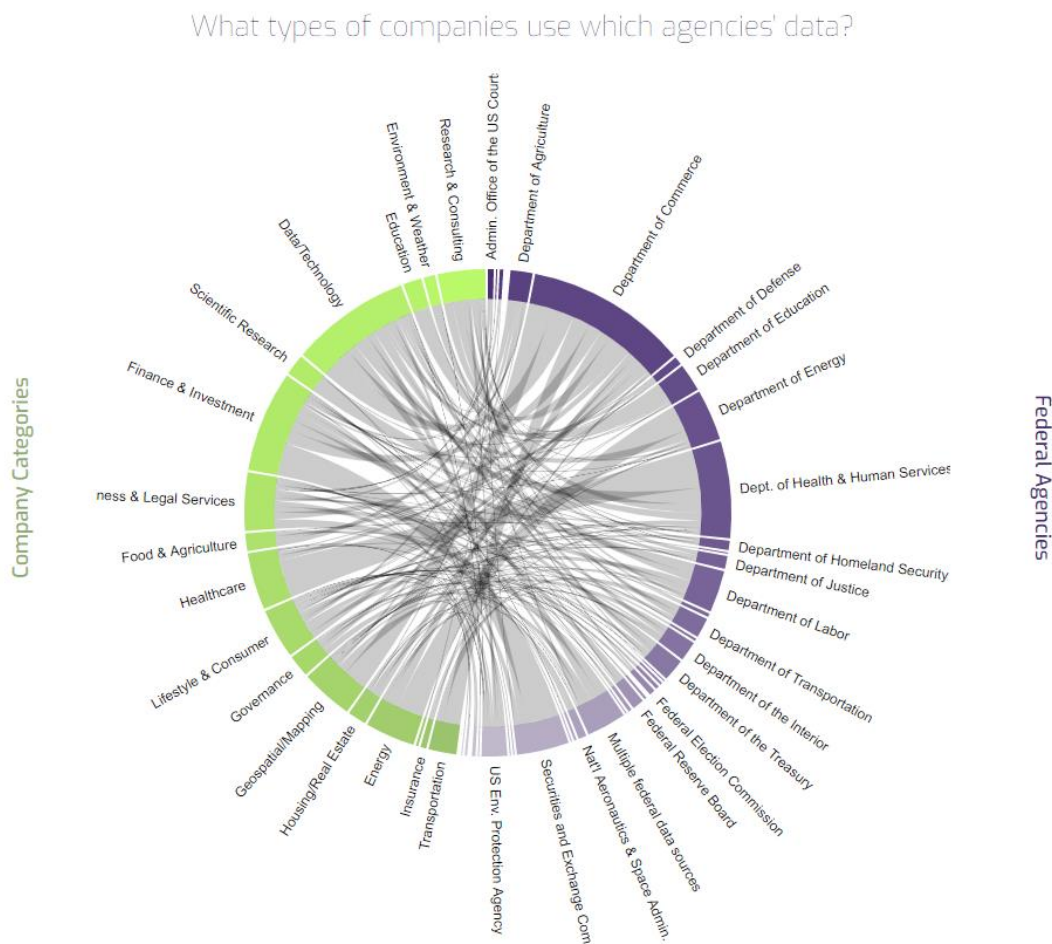
Scale also matters for public intent data, and government failures can lead to both an under-supply of data and missed opportunities from data agglomeration. This will be particularly the case for lower-income countries with relatively weak data systems and limited interoperability. Realizing the gains from agglomeration of public intent data will require a new vision for national data systems where data flow readily across Ministries and to the public for improved policy and program design. Realizing this vision will require relatively large investments and restructuring of national data systems to place data at the center of policy. Additionally, realizing the potential of public intent data may also necessitate the development of national data centers and potentially cross-border cooperation on matters of data.

But once scale is achieved, a separate set of government failures can surface related to concentration of personal information since, as was noted above, governments can potentially abuse citizens' data for political ends, such as election-rigging or politically-motivated surveillance or discrimination.

A final theme is, therefore, the ***importance of trust***. Reaping the full development benefits of data typically involves the exchange of data between parties, and a willingness to engage in this requires the parties to *trust* the security of the exchange. Without adequate security, there is always the risk that data may be misused to the detriment of development outcomes. Achieving a good level of trust depends on putting in place a framework of safeguards for the protection of data, providing rights to data providers while at the same time imposing obligations on data users. In addition, the regulatory framework, as well as the hard and soft infrastructure used to exchange and store data, needs to be secure against cybercrime intrusions.

While these concerns apply particularly to private intent data, trust is also crucial for use of public intent data. Here the concerns relate not only to potential misuse of personal data for surveillance of the population and political ends, but also to transparency. At the most basic level, documentation of sources and collection and aggregation methods is crucial for data quality and for inspiring trust among users of data. But transparent documentation is not a priority in all countries, and data opacity may be a conscious choice on the part of some governments, which significantly undermines public trust and could contribute to social unrest. In short, data policy options discussed in the report will need to take account of political economic constraints.

**Box 1:** A study<sup>15</sup> of 500 firms based in the United States examined private sector use of publicly available government data. The figure below reveals widespread, cross-sector use of Open Data. The grey lines emanating from the purple-shaded part of the circle show which type of private sector firms used data from which government departments. The size of the semi-circle for each Department reflects the count of firms using their data. For example, significantly fewer firms were using Open Data from the Department of Agriculture relative to the Department of Commerce (home of the US census and many other important data) as evidenced by a smaller portion of the circle allocated to the Department of Agriculture. But, even in the case of Agriculture, firms from seven distinct sectors (Data/Technology, Finance/Investment, Legal Services, Food and Agriculture, Lifestyle/Consumer, Geospatial, Housing/Real Estate) used their Open Data. Similarly, the grey lines emanating from the private sector reveal from which departments they use data. For example, the Finance and Investment private sector used Open Data from 19 different Departments and Agencies.



Source: Open Data 500, The GovLab, New York University. <http://www.opendata500.com/us/>.



## Report Outline

The two parts of the WDR will map directly to the conceptual framework outlined above. The first part will focus on the linkage between data and development, making sense of the data landscape and identifying associated development opportunities and risks. The second part will focus on the contextual elements that need to be in place to create a suitable environment for data creation and sharing (Figure 7). Examples and case studies will be used to illustrate both how data can be better enabled to further development objectives and the importance of establishing safeguards to prevent the misuse of data harming development objectives. In particular, **COVID-19 examples will be used throughout to illustrate many of the issues addressed in the report**, including the deficiencies of public sector data systems, the complementarities between private and public intent data, and the governance challenges posed by accessing private intent data for public purposes. COVID-19 examples will be showcased across the different chapters both through boxes and narratives.<sup>16</sup>

**Figure 7: Report Structure Mapping to Conceptual Framework**



### Part 1: Landscape and Opportunities

This part of the report will provide a description of the current data landscape, both in terms of data created primarily with public intent and that created primarily for private purposes. The landscape will highlight where progress has been made, and where work needs to be done, in order to improve data’s usability. With the landscape in place, the discussion will shift to the potential for each of the data types to be used more effectively to improve the wellbeing of poor people and people in poor countries. Chapter 1 will lay out the conceptual framework described above, motivate the need to develop a supportive environment for data creation and reuse that suitably balances “enablers” and “safeguards,” and describe the report’s key themes. The remainder of the report will be organized as follows.

#### **Chapter 2: Data for Public Policy**

Around the world, governments, international organizations and research institutions collect data on people, companies and the environment with ***the intent of designing, executing and evaluating public programs and policy***. As such, these kinds of data provide the foundation for a core function of governments.

A view central to this report is that, across all the paths in which data can improve development outcomes, there are numerous missed opportunities for realizing more value from existing data. This chapter will set out a vision for a new government-based data architecture that unifies the archiving, protection, documentation, exchange and dissemination of public intent data based in a government agency that possesses both autonomy and fiscal resources to carry out its mandate. In many cases, this role could be carried out by a revitalized and empowered National Statistical Office (NSO), though this need not necessarily be the case. This notion of centrally positioning a unified data system is intended to break down barriers across ministries in the use and collection of data, and to centralize the archiving and dissemination of de-identified data as well as the archiving and protection of confidential data. The aim of this unified national data system would be to facilitate the sharing of data across ministries, improve the dissemination and documentation of public access data, improve coordination of data collection activities, facilitate common protocols for design, fieldwork and enumerator training, implement data protection protocols such as de-identification, and centralize and enhance the pool of government statisticians, survey experts, and data scientists.

The chapter will outline the types of data collected with public intent, how they can lead to better development outcomes, and the risks to personal data protection associated with this kind of data.<sup>17</sup> It will also **document gaps in existing public intent data and discuss why such data often are undersupplied**. These data provide systematic flows of information and communication between governments, commerce and civil society, with the aspiration of not leaving any of them out of the conversation. This chapter will cover how effective these information flows are in improving the well-being of citizens with a focus on challenges for ensuring that the poor benefit, and challenges that are particularly pronounced in poor countries. It will end with recommendations on how public intent data can maximize its development impact, drawing on initiatives to improve the stock, quality, infrastructure, literacy and use of this kind of data.

This chapter will distinguish between four main types of public intent data for use in monitoring and designing public policy that are foundational to an effective data infrastructure: censuses, sample surveys, administrative data, and geospatial data.<sup>18</sup> **Data for public policy can lead to better development outcomes through multiple channels** and all three pathways between data and development from the theory of change – governments, private sector and civil society – are relevant for public intent data. The report will elaborate and provide examples on these pathways and draw lessons for how this kind of data can maximize its impact.

Data collected by government entities are most typically used to improve public policy flowing through the middle channel in Figure 2. They are used to **improve program design**, particularly in the targeting of resources.<sup>19</sup> In Croatia, for example, a detailed map of deprivations across 556 municipalities relying on household survey data and census data has been used to ensure that EU funds are more concentrated in the poorest municipalities (Figure 8).<sup>20</sup> Public intent data are also used to **improve monitoring of progress in development outcomes**. For example,

tracking the United Nation’s Sustainable Development Goals (SDGs) heavily relies on public data. SDG 1 on poverty primarily relies on household survey data while SDG 6 on clean water and sanitation relies on a mixture of household surveys, population and housing censuses, as well as administrative data.<sup>21</sup> The data gathered in surveys, censuses and through administrative records are compiled to produce key economic statistics, such as growth rates, inflation rates, and unemployment rates, which can **improve and align macroeconomic policy with development objectives**.

Public intent data are used for development through the other channels noted in Figure 2. The private sector frequently uses government statistics to gauge market trends, adjust prices in different markets, and in general **improve business decisions**, which spurs better development outcomes through competition and better use of resources. Public intent data can also foster better development outcomes through civil society by **improved transparency, accountability and empowerment**. For example, data feedback systems such as hotlines and warning systems (e.g. air quality, food price spikes, and disaster alerts) can empower citizens and improve the communication flow between them and the government. Similarly, these data are used to create cross-country indicators of performance such as the Human Development and Doing Business indices, which then help citizens to understand where they are lagging behind their peers. In Togo, for example, inspired by Rwanda’s achievements in terms of Doing Business, the authorities made several reforms in the areas of getting credit, registering property and starting a business.<sup>22</sup>

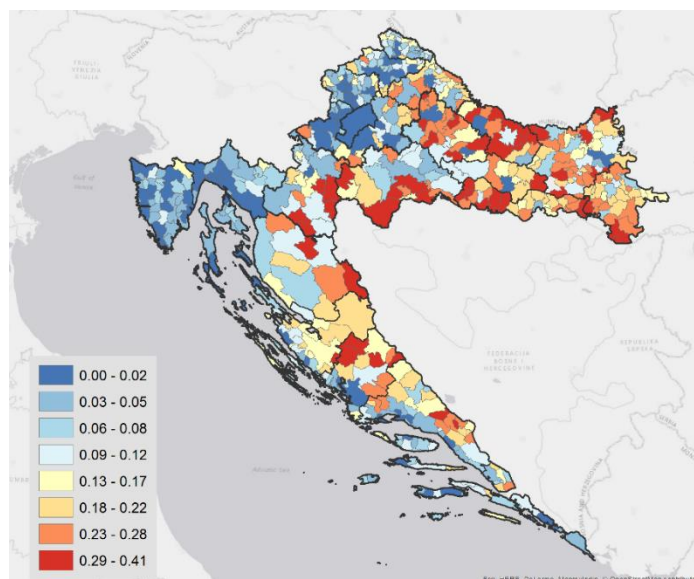
Despite the demonstrated value of public intent data for development, there is a persisting **need to improve the availability, quality, timeliness and policy relevance of public data** across the globe, and particularly in countries with greater development challenges. Tracking progress towards the goals laid out in the national development plans and the 2030 Agenda places an unprecedented burden on the national statistical systems to generate the required data and statistics. Yet many of these systems remain unequipped from infrastructure, technical and financial standpoints to deal with these challenges. For example, as of 2017, only 56 percent of all national statistical plans across the globe was fully funded.<sup>23</sup> Cross-country comparisons using the Statistical Performance Index reveal that the poorest countries have the longest way to go in terms of building statistical capacity.<sup>24</sup> Even for countries with strong statistical capacity, the ability to disaggregate findings by important sub-populations -- women, children, the disabled, the displaced – is limited.<sup>25</sup> Several initiatives are underway to address these gaps. The World Bank is supporting statistical capacity building, particularly in IDA-eligible countries, through the Data for Policy package aiming at reducing gaps in the availability of core data, stimulating data utilization for evidence-based policymaking, and more.<sup>26</sup>

In discussing how data collected for public policy purposes help form the foundation of well-functioning data systems and help support economic development, we anticipate the emergence of several messages, including:

## Potential Messages

- Realizing the potentials of public intent data will require a new vision of a unified, national data system – one that places data centrally in the decision-making process and one that improves the collection, protection and exchange of data. While it is true that there are important gaps in the stock of foundational data, the more important issue is the need to develop and improve the architecture that supports the curation and safe use of these data within governments, research communities and the general public.
- Ensuring political commitment to, and predictable government financing for, public data production remains a central struggle in lower-income countries. Public data will continue to serve as a foundation for public policies but weaknesses in the availability and quality of these data remain widespread, limiting their value. In many less developed countries, there is a clear vision of the national statistical infrastructure, but stable financing and staffing are inadequate to realize this vision. There is a need for greater national focus on data and for stronger international coordination.

**Figure 8: Index of deprivations for municipalities and cities**



*Source:* Croatian Bureau of Statistics (2016), Corral et al. (2019) and the Ministry of Regional Development and EU Funds.

*Notes:* This map produced by the Government of Croatia in collaboration with the World Bank based on data collected for public purposes, provides policy makers with a relatively high level of geographic resolution for addressing poverty.

### **Chapter 3: Data for Private Purposes**

This chapter will focus on productivity and development impacts of data originally created for private intent. The recent explosion of new data has largely come from private sources, such as digitization of firm operations, mobile phone usage by individuals, digital transactions, social media interactions, satellites, and remote sensing. Such data are collected at exceedingly high frequency and can provide a comprehensive picture of user behavior, preferences, and decision-making. Private intent data therefore has unprecedented promise for development impact, but it also requires careful attention to issues of personal data protection and ownership, limits of anonymization, equitable distribution of benefits, governance capabilities and frameworks, and other potential side effects.

**Pathway to Development I: Private data can lead to productivity gains and growth.** The origins of private-intent data trace back to the rapid digitization of transactions and decisions within firms and organizations. Research evidence on the impact of such digital technologies on productivity comes primarily from developed economies. Nevertheless, there are lessons and implications that apply to less developed countries also. In fact, the 2016 World Development Report (WDR) focused on digital dividends due to technological advancement and implications for development policy. The 2016 WDR message revolved around the impact of digital technology on economic growth as being mediated through three mechanisms, namely inclusion, efficiency, and innovation. **Inclusion** refers to more firms being able to compete through digital marketplaces; **efficiency** refers to gains due to faster processing and digitization of previously manual tasks; and **innovation** refers to new products and processes being invented as a by-product of greater online competition.

The chapter will revisit and reinforce those messages, and briefly summarize the literature on digital technology adoption and productivity from the perspective of less developed economies. Key themes revolve around complementarities and heterogeneous impacts of digitalization across firms. For example, studies from advanced economies have shown that productivity gains can be driven by complementarities between digital technologies and organizational capital and management skills,<sup>27</sup> R&D and intangible investments,<sup>28</sup> human capital and ICT-related skills,<sup>29</sup> and a regulatory environment that facilitates efficient reallocation of resources.<sup>30</sup> Heterogeneity arises because high-productivity firms tend to adopt digital technologies,<sup>31</sup> and the productivity effects of technology adoption tend to be larger for manufacturing than service firms, and more broadly in industries involving a high share of routine tasks.<sup>32</sup> Given the types of firms in less developed countries and potentially limited scope for exploiting complementarities, gains in terms of inclusion, efficiency, and innovation may be harder to achieve in those environments.

While this subsection will briefly summarize these impacts, it will not be the focus of the chapter (the 2016 WDR already covered this topic in depth). Our focus is on how data itself (rather than digital technology adoption) can improve the lives of the poor and help less

developed economies advance. The chapter therefore focuses on more direct pathways through which data generated with private intent can improve development outcomes.

***Pathway to Development II: Private data can directly improve development outcomes.***

According to recent estimates, the reach of mobile phone networks has expanded to almost universal coverage worldwide,<sup>33</sup> which helps generate data on billions of people,<sup>34</sup> though unfortunately a large share of the poor in less developed countries remain unconnected. Similarly, there has been a recent push to apply the targeting and matching algorithms used in consumer and social media advertising to applications related to development. These data are also increasingly being fed into machine learning processes that better enable targeting of development needs. In this section, the report will focus on examples where private intent data has demonstrated its potential to improve development outcomes. These include specific applications in the finance and health sectors, and cross-sectoral applications related to survey methodology.

***Innovations in Finance: Digital credit can relax credit constraints and deepen financial inclusion.***

Companies and products such as Lenddo, Cignifi, and M-Shwari have begun utilizing individuals' digital footprints, Call Detail Records (CDR), and Transaction Detail Records (TDR) to generate credit scores for applicants. Additional private data sources are proving valuable for assessing creditworthiness. For example, the app used by Tala, a digital lender, is reported to process over 500 unique user variables in calculating credit scores. Lenddo assesses creditworthiness using social media accounts to score applicants on their online behavior and the strength of their connections. Members within a social network can also vouch for one another and exert social pressure for loans to be repaid. Using this approach, Lenddo's default rates are comparable to the industry average for the microcredit sector yet entail lower transaction costs.<sup>35</sup> Another example comes from behavioral signatures in mobile phone data, where research has found predictive power for credit worthiness in phone usage patterns, periodicity, and mobility.<sup>36</sup>

These examples are not exhaustive of digital applications in finance. For example, conversational interfaces are opening a new set of possibilities for understanding customers through their digital interactions with chatbot content on smartphones and basic phones.<sup>37</sup> And psychometric tools, which are used to assess the abilities, attitudes and personality traits of individuals have shown promise in identifying creditworthy borrowers among those who lack a credit history.<sup>38</sup>

***Digital credit therefore has potential to promote financial inclusion and could help close the gender gap in usage of financial services.*** However, credit-scoring tools that integrate many sources of data and data points on an individual, most of which are collected without consumer knowledge, create serious transparency challenges. And ready access to credit could lead to over-indebtedness, especially among those who lack financial literacy.

***Innovations in health: Data can expand access, reduce costs, improve quality, and create new products and services.*** In so doing, data can help promote equitable and affordable access to health services. And the potential impact of these innovations comes across different functional areas in health: population health, individual care, health systems, and pharmaceuticals and medical technologies.<sup>39</sup> Patient care is often ultimately through public providers, but the applications discussed in this section were developed by private companies for commercial purposes. For example, using mobile phone data and mapping technology to geospatially validate vaccinator attendance in Pakistan has drastically improved attendance and geographical coverage and boosted vaccination rates.<sup>40</sup> Similarly, Indonesia is currently using artificial intelligence and big data analytics to monitor and predict disease burdens.<sup>41</sup>

Data and data analytics are also transforming how individual health care is delivered. Virtual health assistants provide the opportunity to assist patients directly in their own care and wellness through data-driven analytics with care recommendations. Similarly, innovations in data analytics give healthcare workers tools to triage and diagnose patients and can provide clinical decision support. Augmenting and providing specialized expertise to frontline health care workers and generalist physicians located outside of health centers or without access to specialized medical expertise can help address important constraints in access, cost, and quality of health care. Likewise, advances in image-based diagnostic support, such as for radiologists and pathologists, has been shown to increase their productivity by enabling faster and more accurate diagnoses.<sup>42</sup>

Finally, health systems can use data and data analytics to better plan and predict resource usage and to improve quality assurance and training. In this vein, pharmaceutical and medical technology companies are leveraging data and data analytics to enable faster, more accurate, and less costly research and development.<sup>43</sup> While direct evidence on impacts for less developed countries is scarce, these advances in data are likely to have trickle-down effects in the form of cheaper healthcare access, affordable drugs, and better trained care providers.

***Innovations in Surveys:*** A final application of private intent data is not an innovation in data methods or sources, but rather a refinement in individual-level survey methodology. Gallup, Inc. is a global analytics company known for its public opinion polling. The Gallup World Poll has been used to understand the lives of people around the world by researchers, private companies and international organizations, including the World Bank.<sup>44</sup> When the World Bank in 2011 launched the Global Findex database, the world's most comprehensive database on how adults save, borrow, make payments, and manage risks, it did so by leveraging the Gallup World Poll. The database has become a mainstay of global efforts to promote financial inclusion and is used to track progress towards the World Bank goal of Universal Financial Access by 2020 and the United Nation's SDGs (goal 8.10).

Other organizations have also partnered with Gallup, Inc. to create public use data that is used to track progress towards the United Nations Sustainable Development Goals.<sup>45</sup> Similar to the Global Findex example, in these cases the lines between data collected for private and public

intent become blurrier. And the potential for combining public and private intent data to improve development outcomes becomes greater. Exploiting synergies between the two types of data is, therefore, the focus of the next chapter.

***Risks and Challenges of using private data for development purposes:*** The chapter closes with assessments of specific risks and challenges in using private data to improve development outcomes. These include (a) issues of equity stemming from uneven coverage of some data types such as online and social media data; (b) the skill levels of development practitioners and regulators in aggregating diverse and oblique data to derive meaningful insights; (c) manipulation and gaming new data if individuals become aware that their mobile habits and engagement patterns are being used to target benefits; and (d) the inadequacy of legal and regulatory frameworks for protecting personal data and ensuring transparency in how data are being collected and used.

### ***Potential Messages***

- Innovations in the use and application of private intent data are opening doors to development impact that were previously unimaginable. Such applications and uses hold great potential for empowering the poor and women, improving service delivery to them, and enabling them to participate more fully in the economy.
- At the same time, however, the way forward warrants a cautious approach to avoid unintended consequences of private data proliferation and requires a conducive and enabling environment that takes seriously concerns about equity, personal data protection, gaming, durability, and regulation.

### ***Chapter 4: Data Synergies***

Development problems are typically complex, multi-dimensional issues with links to societal, cultural, economic, environmental, demographic and many other factors. Policy design based on data covering just one factor will be incomplete, and sometimes ill-advised. Drawing information from multiple sources can help to mitigate this concern and lead to better development outcomes. This chapter highlights potential synergies from using both public and private intent data to enhance economic development through each of the channels outlined in Figure 2 – more specifically: better policy and program design, improved public service delivery and transparency, and private sector growth.

In recent years, the spread and pace of diffusion of mobile phone technology, sensors, and high-resolution satellite imagery has been unprecedented. Multi-sectoral applications that leverage these new data sources have raised considerable interest as they aspire to generate insights at scale, with extraordinary spatial and temporal resolution. Through key examples, this chapter will first illustrate that ***combining different types of data can advance evidence-based policy by producing more timely, representative, and precise official statistics*** – more cheaply – than using foundational data alone.



In a rapidly changing world, **timeliness** of development data is critical for decision makers. Yet most data collected for public policy purposes is laboriously gathered, entered and processed; frequently aging many months or years before it can even be used for analysis. Inferences made from these data are based on how the world looked when the data were collected, rather than today. By complementing public data with more frequent and timely private data, policy inferences can be made based on reflections of the world today. A couple of examples of real-time analysis based on the combined usage of public and private data include the prediction of traffic conditions based on data from road maps, video cameras, GPS devices in taxis and sensors embedded in the streets,<sup>46</sup> and predicting displacement with cellphone call detail records from the months before and after the earthquake in Haiti (validated with household survey data) to assist relief organizations.<sup>47</sup> The increasing availability of data from satellite imagery, mobile phones, and internet records raises the possibility of “nowcasting” economic indicators such as poverty and gross value added by combining measures of change from these alternative source data with past survey data to generate more timely estimates.<sup>48</sup>

Integrated data applications can generate accurate estimates by increasing the **spatial resolution** of public statistics allowing, for example, greater refinement in targeted policies and actions. Examples include (i) impacts of floods and droughts on household welfare and agricultural outcomes by merging georeferenced household survey data with publicly available geospatial data on rainfall, elevation and flooding and privately held, georeferenced administrative data on aid flows;<sup>49</sup> (ii) population density by combining publicly available, low-resolution, aggregated census counts with high-resolution building footprint estimates;<sup>50</sup> (iii) poverty estimates by complementing socioeconomic surveys and census data with satellite and call data records,<sup>51</sup> and (iv) plot-level crop yields by overlaying satellite imagery (Sentinel-2) and georeferenced agricultural survey data.<sup>52</sup>

**Synergies from combining data types can also serve to improve service delivery and transparency** in a manner that enhances the flow of information between the civil and public sectors. The improved transparency allows for feedback to play a greater role in holding governments accountable and also allows governments to more effectively disseminate important public service information. Recent advances in monitoring deforestation provide an example of how data synergies can enable greater accountability for governments and companies by increasing **transparency**. A case study from Ethiopia demonstrated the scalability of localized monitoring by indigenous groups, combined with open-access satellite imagery, cloud computing, and publicly available property maps as an effective tool for rapid detection of deforestation.<sup>53</sup> In addition, through social media and tools like Global Forest Watch,<sup>54</sup> the international community can better assist community-based groups to hold governments accountable in achieving national sustainable development commitments.<sup>55</sup>

Improved transparency with accountability mechanisms in place can serve to improve service delivery of government programs. Data synergies can also **improve service delivery** directly. As one example, combining sample survey data on public health with data from internet searches

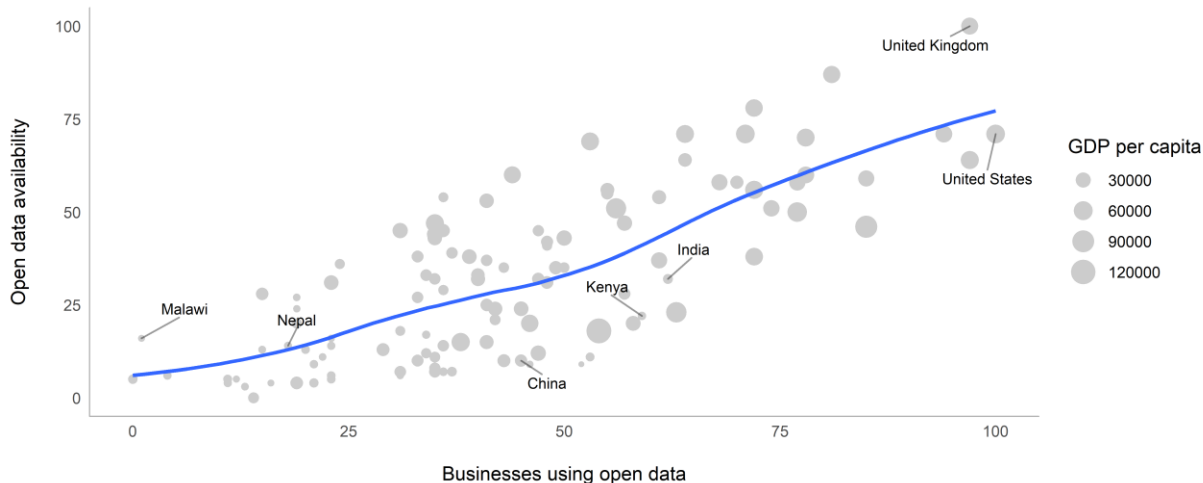
and social media posts can help governments deliver better targeted public health interventions in a significantly more timely manner.<sup>56</sup> Similar methods of combining data sources have also aided governments in a more effective delivery of services in response to disasters.<sup>57</sup>

***Making public intent data accessible can spur private sector growth and innovation.*** When public data is openly available, companies can incorporate it into their own data streams without bearing the time and financial costs of data collection. Examples of private sector use of public intent data are widespread in advanced economies, particularly the United States. For example, private healthcare firms use the US Medical Expenditure Panel Survey combined with proprietary firm data to shape their business decisions and lobbying efforts; and businesses in general combine the American Community Survey with their information on customer preferences to regionally customize store inventories.<sup>58</sup> The annual value of open data to the private sector is estimated to be in the trillions of dollars.<sup>59</sup>

To some extent, the private sector in lower-income countries also benefits from access to open data. For example, satellite and remote sensing data is at the core of the business models of crop index insurance providers in African agriculture. In Ghana, agriculture-technology companies combine government meteorological and administrative data with proprietary data to provide advice to farmers.<sup>60</sup> Given the significantly lower levels of access to, and use of, open data in lower-income countries (Figure 9), there is likely ***much scope for the private sector in lower-income countries to benefit substantially more from increased access to open data.***

***Potential Messages:***

- Existing data can be better used to tackle development issues in locations and for populations that previously were not accessible.
- The development benefits of combining public, foundational data with newer types of private and citizen-generated data could be much greater if they were made interoperable and accessible from the design phase, while being attentive to consent and personal data protection concerns.

**Figure 9: Access to Open Data and Private Sector Usage**

Source: <https://opendatabarometer.org/4thedition/>.

*Note:* The relationship between businesses using open data and the availability of open data by country in 2016. The estimates are constructed by subject matter experts based on primary and secondary country-level data and literature searches. “Open Data availability” assesses the availability of data across 12 sectors, from health to land ownership. The axes are normalized indices and data points should be interpreted in relative terms to the other countries.

## Part 2: Realizing the Opportunities

The discussion in part 1 illustrates the scope for data to greatly improve the lives of poor people. Part 2 tempers slightly the enthusiasm by highlighting the many building blocks that need to be in place to realize the opportunities. One such block is a digital infrastructure that allows countries to collect, share, and combine data for evidence-driven development. In addition to improving the physical data infrastructure, it is also critical to improve the human capital in terms of data literacy and skills to reap the potential returns (Chapter 5). Improved data infrastructures facilitate the efficient exchange of valuable information in ways that can be used for positive and negative development outcomes. Ensuring that opportunities are realized while protecting the rights of citizens and firms requires innovative thinking on the appropriate legal framework (Chapter 6) and government policies (Chapter 7) to find this balance. Even with all the building blocks, there is still a need to think through how these blocks should be placed, and what holds them together, in a way that helps to ensure that the world’s poor share in the benefits of our digitally transforming world (Chapter 8).

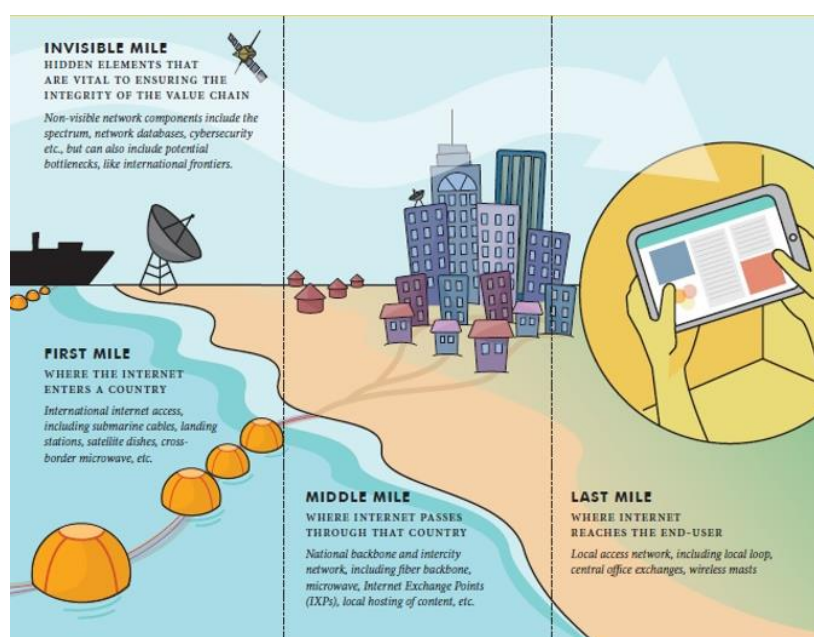
### **Chapter 5: Infrastructure Challenges**

Internet traffic has experienced dramatic growth from around 100 GB of traffic *per day* in 1992, to 45,000 GB *per second* in 2017 and is forecast to more than triple between 2017 and 2022;

equivalent to every person on the planet consuming 50 GB per month.<sup>61</sup> Such exponential growth in data traffic is challenging data infrastructure to keep pace along the full length of the network: including first, middle and last miles (Figure 10).

Collection, transmission, storage and processing infrastructure comprise the underlying foundation of the data economy. Mobile devices, computers, cameras and sensors capture data at source. Fiber optic cables and wireless networks, supplemented by microwave and satellite, get data to its destination. Servers host the data providing users remote access through curated databases. Cloud computing platforms offer software for processing both structured and unstructured data. Ensuring an adequate balance of investment along this supply chain depends on well designed regulatory frameworks that provide confidence to private investors, while creating competitive pressures to contain costs and enhance the quality of service delivery. **Complementary to this physical infrastructure (both hard and soft) are the skilled human resources** that are needed to develop and operate the system but which are in scarce supply in many countries.

Figure 10: Basic architecture of data networks



Source: <http://blogs.worldbank.org/digital-development/how-wdr16-policy-framework-applied-union-comoros> Kelly (2016).

**The development of modern data infrastructure raises equity concerns, both within countries as well as across countries**, as evidenced by a large and widening digital divide. For instance, in early 2019, the average Finnish mobile user consumed 17 GB of data per month, almost seven times more than the average user in Peru.<sup>62</sup> Many developing nations lag behind in their

capacity to generate, transmit, process and analyze data. Indeed, around 100 economies lack the necessary infrastructure to keep the internet functioning in the event that international connectivity is interrupted, with potential for significant economic disruption.<sup>63</sup> The reality is that data centers and cloud services, in particular those owned by large information technology multinationals, are geographically concentrated in a handful of developed countries.<sup>64</sup> Even within countries, access to data infrastructure remains quite skewed towards higher income groups in urban areas. For example, in South Africa, although 99.5% percent of the population lives within range of a mobile broadband signal, only 60% access the internet due to a range of demand-side barriers, including costs of devices and services, as well as limitation of digital literacy and availability of local content.<sup>65</sup> The gap between rural and urban (metro) access in South Africa is notable; 45% of rural dwellers use a mobile phone to access the internet compared to 67% in urban areas. Even within geographical areas there is emerging evidence of a gender divide in access to broadband.<sup>66</sup>

In several respects, this chapter will go beyond coverage of digital infrastructure discussed in earlier reports.<sup>67</sup> First, it will focus particularly on **data** infrastructure (internet exchange points, data centers, cloud computing, software) as opposed to narrower **communications** technologies, and address the latest technological innovations (including 5G, -IoT, satellites, balloons, fixed wireless). Second, it will evaluate recent industry trends; including declining average revenue trends due to the growth of Over-the-Top applications<sup>68</sup> that use infrastructure for free,<sup>69</sup> as well as the emerging vertical integration between content and infrastructure providers. Third, given that 95% of the world's population is already covered by mobile broadband networks,<sup>70</sup> discussion of access will be broadened to include consideration of how the cost and **capability of handsets**, as well as the speed and **affordability of networks**, affects users' ability to take advantage of data infrastructure once it is in place. Fourth, the chapter will undertake original empirical analysis to characterize the patterns of demand for data in less developed countries of the world and the anatomy of cross-border data flows, as well as the availability of modern data infrastructure and the economics underlying its adoption in developing country contexts.

This chapter will examine barriers to core data infrastructure investment in emerging markets including market size, regulatory challenges, awareness, security and electricity, particularly in the context of 5G.<sup>71</sup> One of the central issues to be explored in the chapter will be the case for **developing national-level data infrastructure for storage and processing versus relying on international facilities**. From a country's perspective – beyond the potential for the data infrastructure industry to contribute to local economic growth – there are certain micro-economic tradeoffs that need to be understood. For instance, the price of international bandwidth to access overseas servers abroad needs to be balanced against the lower costs for cloud services in international markets.<sup>72</sup> There will be situations where storing data overseas makes sense, such as hosting websites aimed at the international market. On the other hand, it may be more cost-advantageous to have national infrastructure to exchange locally destined traffic and to host government services. There may, in addition, be non-financial considerations

associated with the protection and control of access to sensitive personal data. From the perspective of large global providers of data services, there may also be an economic case to host content closer to users in less developed countries through peering at local Internet Exchange Points. Such considerations have motivated some global technology companies to vertically integrate into data infrastructure development, with potential implications for net neutrality and costs.<sup>73</sup> To conclude, data storage facilities may also have a significant impact on the environment, mainly driven by the underlying technology of the site and its geographical characteristics.

**Potential Messages:**

- As mobile coverage is almost worldwide, efforts to achieve universal access must increasingly shift to addressing demand-side barriers that prevent uptake of services once these become available. Critical factors include: the cost and speed of broadband services; the affordability of handsets; digital literacy; and the availability of relevant local content.
- Given the exponential growth of data traffic, there is a need to ensure that the development of data infrastructure keeps pace. An important strategic choice for less developed countries regards striking the right balance between the development of domestic data infrastructure versus reliance on international service providers.

**Chapter 6: Legal & Regulatory Challenges**

The increasing volume of data about individual behavior, and their potential integration across sources and subsequent reuse, creates opportunities to enhance decision-making capabilities and drive socio-economic progress through innovation.<sup>74</sup> However, there are also **mounting concerns regarding misuse of personal data**, based on various forms of individual surveillance combined with the application of personalized algorithms. Such manipulation of citizens may be motivated either by political ends (including undue electoral influence) or commercial considerations (as with perfect price discrimination).

This chapter will examine the legal foundations for the regulation of data flows to maximize the development benefits of data while minimizing the associated risks to data holders. Central to the chapter is the notion of a **data transaction** – whereby data are exchanged between two parties – which are critical to the functioning of the data-driven economy. The two central pillars of the chapter are the legal and normative **“enablers” and “safeguards”** that relate to the collection and use of data in any particular transaction. These will be enumerated and explained in detail, and the range of possible legal approaches will be identified (a high-level view is presented in Figure 11).

Figure 11: Overview of conceptual framework

<p><b>ENABLERS:</b> <i>facilitate sharing of data</i></p> <ul style="list-style-type: none"> <li>• Transparency</li> <li>• Openness <ul style="list-style-type: none"> <li>• Interoperability</li> <li>• Portability</li> <li>• Access to information</li> </ul> </li> </ul>	<p><b>SAFEGUARDS:</b> <i>prevent abuses of data sharing</i></p> <ul style="list-style-type: none"> <li>• Accountability</li> <li>• Data Protection <ul style="list-style-type: none"> <li>• rights (consent)</li> <li>• obligations (use limitations)</li> </ul> </li> <li>• Cybersecurity</li> <li>• Algorithmic Regulation</li> </ul>
--	---

Legal enablers affect the usability of data and the ease with which it can be shared across parties. **Transparency** ensures that the existence and location of data is known. **Interoperability** facilitates the combination and cross-referencing of different data sources, while **portability** empowers users to transfer their own data across platforms. The principle of access to information avoids discrimination in the availability of data. An important example of this is net neutrality, which ensures that users do not face undue restrictions on their ability to access specific types of data available on the web.

Legal safeguards refer to the **creation of trust around the collection and use of data**; without which the supply of data may ultimately be prejudiced. An important principle is that data collectors and users remain accountable to data providers. This requirement can be particularly challenging in the case of algorithms that are often opaque and dynamic in nature and may lead to unforeseen consequences or embody unintended biases. The central issue of data protection involves the creation of both legal rights for data providers and legal obligations for data users. This involves the creation of agency over data through mechanisms such as consent, rights of use of data, and regimes that allow reuse of data without consent for legitimate purposes. Safeguards also encompass how data is secured and protected, covering the various obligations of those collecting, processing or using data to take certain precautions to ensure the integrity and adequate protection of the data. A related issue is the substantive and procedural provisions in law that criminalize unlawful or illegal access or use of infrastructure, systems and data. This is known as cybercrime.

Finally, the effective implementation and enforcement of enablers and safeguards needs to be underpinned by the legal establishment of suitable institutions, whose governance will be further developed in Chapter 8. Examples of legal responses to these include the creation of entities such as Computer Emergency Response Teams and Data Protection Authorities.

Within the construct of enablers and safeguards, the chapter will explore the interplay of **several legal and economic issues related to the creation, collection and use of personal and nonpersonal data** for public and private intent. One issue is anonymization of personally identifiable information as a form of protecting privacy. The EU's General Data Protection Regulation (GDPR) has detailed provisions about anonymization and reverse-engineering of identity, while In India, for example, the proposed Data Protection Bill, 2019, takes the extraordinary step of criminalizing reidentification. A second issue is data localization, which restricts cross-border data flows for a variety of economic and political reasons that merit deeper analysis. A third issue is the conundrum of ownership of data, including assessing its relevance and associated trade-offs, and providing alternative normative frameworks for data governance. Solutions to these issues may be technical as much as legal, hence the chapter will explore how privacy-by-design and privacy-enhancing technologies, can also be used to achieve data protection, working hand-in-glove with legal and regulatory regimes or even outside of them.

While grounded in a **rights-based legal approach to data protection** that identifies the rights and obligations of different actors in data transactions, the chapter will aim to integrate economic considerations, by surfacing trade-offs between and within the different legal enablers and safeguards, and, where appropriate, by adopting a cost-benefit framework for assessing regulatory impacts. Some research indicates that regulatory restrictions of data flows tend to reduce productivity and economic output for the data-intensive industries,<sup>75</sup> while other analysts suggest that it is the very protections around data that create incentives for data sharing and therefore increase the value going into the data lifecycle.<sup>76</sup>

The chapter will consider how both hard and soft law impact the data discussion. **Hard law** includes domestic, regional and international law, as well as case law and statutory law that originate from sources of tort, contract and competition law. Some of the issues found in domestic law have their origins in well-hewn and commonly agreed standards already enshrined in international laws, conventions and treaties. Whereas hard law is shaped by the state, **soft law** includes standards, terms and conditions of use, norms and even codes of conduct and other voluntary frameworks used by non-state actors. In this sense, soft law plays as important a role as hard law in how data is collected, stored and used throughout the data value chain. The chapter will also identify emerging areas of law that may apply to data, such as trust and competition law, that could serve as a legal and/or normative framework to support the governance of data for public and private intent.

**The development of legal and regulatory frameworks for data is generally less advanced across the world's less developed countries.** There is considerable public debate regarding how legal frameworks may need to be adapted to those environments, and whether it makes sense to transplant regulations from developed countries, since differing policy objectives, as well as varying institutional capacities, may both have a bearing on the design of the optimal legal and



regulatory framework. At the same time, cross-border spillovers from data regulation in larger economies may affect degrees of freedom for the determination of national systems.

The conceptual framework for data regulation provided in this chapter will underpin the further elaboration of economic policy implications that is to come in Chapter 7 as well as the design of institutional ecosystems to be developed in Chapter 8. In addition, the chapter will provide a landscape survey illustrating the prevalence of legal and normative measures to both enable and safeguard data collection and use across less developed countries of the world, drawing upon case study material from a range of countries with advanced or otherwise distinctive legal and normative frameworks for regulating the collection and use of data.

***Potential Messages:***

- A suitable legal and regulatory environment for data entails a balanced development of both safeguards that provide trust to underpin data transactions, and enablers that facilitate the flow of data across parties. Design of the legal framework should also be informed by a sound understanding of the potential economic trade-offs associated with different aspects of the regulatory framework for data.
- In less developed countries that face weak regulatory environments, the design of suitable safeguards and enablers may need to be carefully adapted to local priorities and capacities.

***Chapter 7: Economic Policy Challenges***

***Data collected for private intent has increasingly become a driver of economic prosperity.*** It is estimated that the digital economy currently accounts for 10% of global GDP.<sup>77</sup> Data-oriented companies are now among the largest globally by market capitalization. In addition to their role as a new factor of production, data have also become a new tradeable commodity. While trade in goods has slowed in the last decade, cross-border flows of data have surged dramatically. Data also facilitate the trading of traditional goods, since according to some estimates electronic-commerce platforms reduce the cost of distance in trade by 60%.<sup>78</sup>

This chapter will focus on three key areas of economic policy that are deeply affected by the shift towards data-driven value creation: they are ***competition policy; trade policy; and taxation policy.***<sup>79</sup> In each of these cases, data creates important economic opportunities, while at the same time posing substantial risks.

First, data-driven platform business can ***enhance competition through more effective market intermediation***, which allows smaller enterprises (as well as lagging regions and populations) to participate more widely in national and global markets and supports greater product and process innovation. However, the powerful network externalities associated with digital platforms create a propensity for market power to be concentrated in a few large (often multinational) firms, making it difficult for start-ups to gain a foothold in the market.

Second, data plays a key role in ***opening up new areas of international trade***, whether as a directly traded commodity embedded in digital services,<sup>80,81</sup> or as a facilitator of Global Value

Chains<sup>82</sup> and electronic commerce.<sup>83</sup> However, national efforts directed at data protection may in turn constrain cross-border exchanges of data, with important and sometimes unintended repercussions for international trade.

Third, data is potentially a powerful tool that can **improve the efficiency of tax administration**, by bringing smaller and more informal businesses into the tax net. However, the ability of digital enterprises to have economic scale in a certain jurisdiction without any physical presence greatly complicates the collection of tax revenues.

**Competition policy.** Specific features of data-driven markets render competition issues more complex and raise the propensity for concentration of market power.<sup>84</sup> In particular, **data is likely to have increasing returns to scale and dimensionality, while possession of data leads to a competitive advantage** and allows leveraging of (direct and indirect) network effects (such as optimizing product development, attracting more valuable advertising, improving search, reducing switching).<sup>85</sup> While anti-trust authorities around the world have begun investigating cases of anti-competitive practices by digital platforms, there are increasing concerns that the traditional tools of competition policy are not well-suited to addressing competition issues in the digital economy.<sup>86</sup> For example, the price impact of mergers may not be as relevant in sectors where services are provided free of charge, while non-price dimensions affecting quality of service (such as the degree of data protection provided) are becoming more important. Moreover, the scope for tacit collusion between firms may be greatly increased through the use of algorithms.<sup>87</sup> In such a fast-moving segment of the economy, lengthy ex-post review of mergers may come too late, while predatory acquisitions of small technology start-ups by dominant players could have anti-competitive consequences well beyond their negligible impact on market share. Possible intertwined issues that are being widely considered, include mandatory portability of personal data, data interoperability, data sharing or data pooling, as well as limitations on the structure and operations of large data-driven firms.<sup>88, 89</sup> However, there is limited evidence of the efficacy of such actions and important questions remain, including whether data is an essential facility, and the extent to which such remedies raise costs and stifle incentives to collect data. This chapter will review the relevance of various policies to safeguard competition, building upon a global database of anti-trust investigations of digital platforms, set against the broader landscape of their market structure and merger dynamics in less developed countries of the world.

**Trade policy.** The boom of data-driven businesses creates opportunities for new entrants in global trade. However, **data protection and data localization regulations, often motivated by domestic policy concerns, will simultaneously affect the flow of data across borders** and hence impact trade in digital services. To date, three broad approaches are increasingly competing to set a global regulatory standard for cross-border data flows: a “liberalizing” approach championed by the United States and the APEC Privacy Rules; a “regulating” model based on the European Union’s General Data Protection Regulation; and a “mercantilist” approach led by China’s Cybersecurity Law and Personal Information Standards of 2018.<sup>90</sup> This chapter will

analyze the implications of different data regulations for international trade and consider prospects for international harmonization of provisions regarding cross-border flow of data. The costs of implementation of such data regulations will be considered, as well as the evidence of their impact on flows of international trade.<sup>91</sup>

**Taxation Policy.** *Where digital goods and services are provided across international borders, it is challenging to determine where value creation takes place*, and therefore where the liability for income tax should arise. Under current rules, non-resident virtual providers can even offer goods and services to consumers without being deemed resident for the purposes of corporation tax, even while taking advantage of the infrastructure network and legal framework in the jurisdiction of consumption. The growing digitalization of the economy and the importance of intangible assets exacerbates the perennial challenge of tax-base erosion and profit shifting.<sup>92</sup> This trend is placing increasing strain on the international consensus around taxation rules and undermines the basis for fair competition between multinational firms and their domestic rivals. A possible response is to shift the burden of revenue mobilization towards indirect taxes, which are already central for revenue mobilization in less developed countries of the world.<sup>93</sup> However, this presents policy and administrative challenges of its own. For example, with respect to VAT, failure to subject to taxation the consumption of goods and services purchased from foreign suppliers disadvantages domestic competitors. The chapter will assess current and future policy proposals to (better) tax the digital economy,<sup>94</sup> review and complement empirical work on the associated revenue potential,<sup>95</sup> and discuss implications for equal treatment between online and offline business as well as between domestic and foreign entities.

**Potential Messages:**

- The shift toward a data-driven economy further complicates traditional policy challenges associated with competition, taxation and trade, leaving countries with weaker institutional environments particularly exposed. In particular, the way in which a country develops its legal framework of enablers and safeguards for data will have important policy implications for competition, taxation and trade that deserve full consideration when such regulations are being designed.
- The data-driven economy is forcing a rethink of major international policy agreements relating to taxation and trade. Any adaptation of these rules should be sensitive to the specific needs and interests of the world's poorer countries.

**Chapter 8: Data Governance Challenges**

The proliferation of data is raising a host of new issues about institutional roles and responsibilities. The role of traditional data institutions, such as National Statistical Offices (NSOs), may need to adapt to improve the flow and use of data. And ***new institutions will be needed to serve emerging functions, such as that of enforcing data regulations, or facilitating citizens' curation of their own data.*** At the same time, the global character of data flows leads

to significant cross-border regulatory spillover effects. Frameworks need to be developed to support international regulatory harmonization on a mutually beneficial basis. The interests of poorer nations should be safeguarded in international negotiations on data issues where they may have limited voice.

For the purposes of this chapter, ***data governance is defined as the confluence of policies, platforms, and state and non-state institutions for the effective creation, collection, storage, management, sharing, use and destruction of data***. To expand on the above definition, good data governance aims to maximize the positive linkages between data and development outcomes through well-functioning pathways described in chapter 1 (Figure 2). To illustrate more concretely how good governance can foster linkages between data and development outcomes, the chapter will build on the examples described in chapters 1 through 4. This will include the type of agreements and standards which facilitated the data flow described, and will analyze the relevant enablers and safeguards, and identify key challenges and how they were overcome.

The 2017 WDR on “Governance and the Law”<sup>96</sup> unpacks key governance principles, which are applicable to this conception of data governance: a comprehensive and effective policy environment; an integrated architecture underpinning the data lifecycle (Figure 3)<sup>97</sup>; and strong institutions to solve collective development problems and balance power asymmetries. Given the rapid pace of innovation, the roles and functions of these key components are evolving.

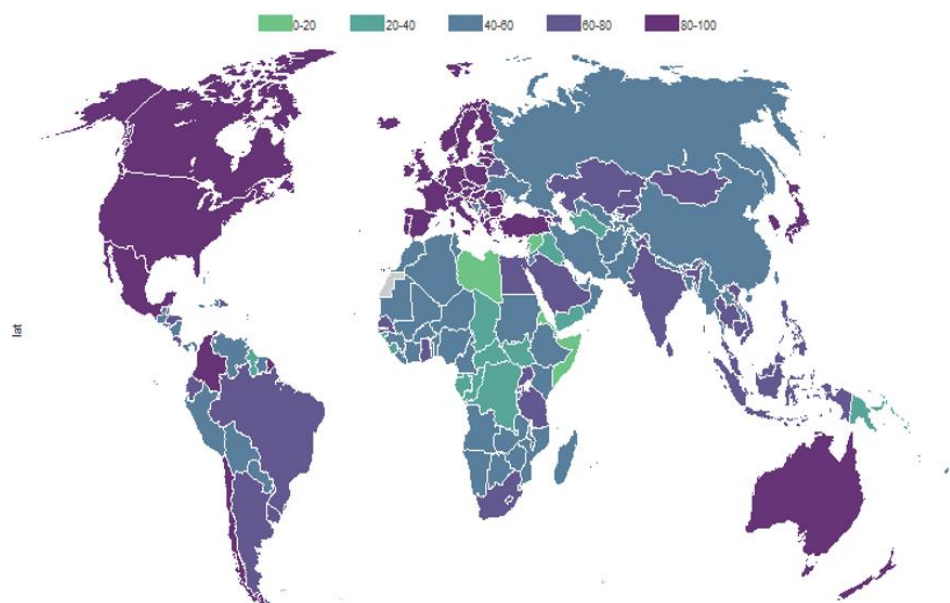
Following the 2016 WDR on “Digital Dividends”, ***this chapter will highlight the importance of strengthening the “analog complements” of data governance***, including implementing a strong legal and regulatory enabling environment, exploring the role and utility of different skills and capacity development models, and strengthening the accountability and effectiveness of relevant institutions.

Focusing initially at the national level, the chapter will ***map out the different functions needed to enable and safeguard data transactions along the entire data life cycle***. It will go on to consider which institutions are needed or best suited to perform the key actions identified, including both traditional actors (such as National Statistical Offices) as well as potential new actors (such as data regulators or data aggregators). Due consideration will be given to the avoidance of conflicts of interest, as well as the adequate provision of checks and balances. The report will also consider the role of data intermediaries, including – among others – data brokers and data markets. The chapter will take a holistic view of data governance, encompassing both formal and informal institutional arrangements, public and private sector roles, as well as the supply and demand-side of governance (including civil society participation). An institutional mapping exercise will analyze how different data governance frameworks have been implemented across a variety of innovating countries, drawing on ongoing World Bank research.

**With the rise in cross-border data transfers, some of the challenges associated with enabling and safeguarding** the use of this personal and non-personal data, including the ethical use of these data by artificial intelligence programs, **may be more appropriately addressed at the regional or international level.**<sup>98</sup> According to a survey conducted by the Pathways to Prosperity Commission in 2019,<sup>99</sup> policymakers in developing country contexts emphasized that areas where “global efforts” were most needed included the areas of (i) taxation; (ii) cybercrime and cybersecurity; (iii) privacy and data protection; (iv) market competition; (v) Intellectual Property (IP); and (vi) data sharing and interoperability. Within these priority areas, international cooperation and coordination was considered most needed to support the development of regulatory and technical standards. International regulatory cooperation in the above areas can have positive externalities, including improving regulatory predictability, reducing compliance burdens, reducing the risks of regulatory arbitrage, and potentially encouraging investment flows. Such international cooperation can also play an important role in enabling convergence in the development of high-level principles<sup>100</sup> to guide the design and implementation of national level governance frameworks.

The chapter will **explore different multi-stakeholder governance models**, including standards that enable data access to the different stakeholder groups (such as Open Data), as well as data partnerships that facilitate bilateral or multilateral data flows (such as Data Collaboratives,<sup>101</sup> Open Mobility Foundation,<sup>102</sup> Humanitarian Data Exchange,<sup>103</sup> and Data Trusts.<sup>104</sup>)

The evaluation of alternative data governance models will be informed by the institutional constraints that are typically faced in less developed countries of the world, including limited statistical capacity (Figure 12) and weak capability for enforcing regulations. Moreover, **limited data literacy of the wider population** is a serious challenge that cuts across all efforts to build effective and participatory institutions. On this matter, some insights can be drawn from the parallel literature on financial literacy,<sup>105</sup> which suggests that integrating digital literacy into the school curriculum may ultimately be more effective than targeted adult education;<sup>106</sup> although the long-run effects are not yet known.<sup>107</sup>

**Figure 12: Statistical Performance Index, 2018<sup>108</sup>**

Source: Based on early assessments of the Statistical Performance Index (SPI), 2018 data.

Notes: SPI measures the statistical performance of countries over time, helps identify the strengths and weaknesses of national statistical systems and areas of potential improvements. SPI is planned to be launched as part of WDR 2021.

### ***Potential Messages:***

- When choosing a governance model, it is important to keep in mind that it needs to be fit for purpose. In a developing country context, this entails being aware of the uneven levels of institutional maturity and technical capacity. This chapter will propose a “maturity model” approach to developing governance arrangements, highlighting the depth of institutional and governance reforms needed both at national and international levels.
- But to keep pace with the fast-changing world, good governance should be dynamic and inclusive. Data flows between various institutions and stakeholders and their roles will continue to evolve in parallel with the digitization and technology innovations. New and existing models need to be flexible and designed to adapt to these changes.

### **Institutional Roadmap on the Data agenda for the World Bank Group (WBG)**

This report’s ambition is to appeal to a wide range of stakeholder groups (governments, private sector, academia, civil society and individuals). The discussions and recommendations in the report aim to inform, educate and spur debates and positive actions among all these stakeholder groups. In order to inform and guide the Bank’s position and work in the data agenda and its operational activities, the team will produce a policy paper that will outline the recommendations for the Bank as an additional output along with the WDR2021 report.

## Consultations, partnerships, and timetable

### Team

Robert Cull, Vivien Foster and Dean Jolliffe are the co-directors of the report and Malarvizhi Veerappan is the report manager. The core and extended team comprise of staff from various Bank units and will bring diverse expertise to inform the range of issues discussed in this report. The core and extended team include Adele Moukheibir Barzelay, Miriam Bruhn, Rong Chen, Niccolo Comini, Hai-Anh H. Dang, Cem Dener, Samuel Paul Fraiberger, Craig Hammer, Talip Kilic, Jan Loeprick, Daniel G. Mahler, David Medine, Martin Molinuevo, David Newhouse, Sara Nyman, Vincent Francis Ricciardi III, David Satola, Dorothe Singer, Philip Wollburg and Bilal Husnain Zia. Kenneth Zaul Moreno Sermeno and Marcelo Buitron serve as research analysts; Chisako Fukuda will coordinate the communications aspects of the WDR; Michael Minges, Rory Macmillan and Zia Mehrabi as expert consultants; and Selome Missael Paulos provides administrative support to the team.

The report is sponsored by the Development Economics Vice-Presidency (DEC). Aart Kraay, the acting Chief Economist will oversee the production of the report. The report will be developed in close partnership with Makhtar Diop, Vice President, Infrastructure.

### Internal Consultations

The WDR team has begun holding internal consultations with various stakeholders. The team has held informal discussions with numerous colleagues on the outline and specific planned chapters of the Report. It has made presentations to the Chief Economists Council, Human Development (HD) Global Leads Team and bilateral discussions with several Senior Leadership Team members. The team has also held informal consultations with several Executive Directors and advisors, and over the coming weeks will meet with others. Throughout the preparation of the Report and beyond, the team will continue to work closely with the Development Data Group, Director, Haishan Fu. A Brains Trust (see Annex 3) comprised of staff from various Bank units has been formed to ensure there is continuous engagement and feedback throughout the report production process. The team will also reach out to several Bank units to seek country case studies that can be used as illustrative examples in the report.

### External Consultations and Partnership

Since the announcement of the topic, the team has engaged with several external experts. The team also participated in a High-Level Roundtable on 'Data for Development' during the 2019 Annual Meetings, which was chaired by Pinelopi Goldberg and included participants who were senior representatives from several bilateral aid agencies, foundations, and several government leaders from World Bank's client countries, including key data champions.

Considering the wide appeal of this topic and the interest to the various stakeholder groups discussed in the report, the team plans to conduct consultations to cover the following stakeholder groups during the coming months:

1. Governments, International Organizations and bilateral development partners
2. Leading researchers
3. Civil society organizations and citizen groups
4. Private sector

To ensure the report is informed by these varied stakeholder groups, the team will also establish a high-level WDR 2021 Advisory Panel and Technical Board comprised of senior government officials, national statistical agencies, competition authorities, private sector leaders, platform companies, civil society members and leading researchers and policy makers. A tentative list of members being considered for the Advisory Panel and the Technical Advisory Board are listed in Annex 1 and Annex 2 respectively.

#### Timetable

Following the Bank-wide review of the Concept Note (CN) on February 18th, 2020, the team will present the CN for board discussion on March 31st, 2020. The Bank-wide review of the Report's yellow cover is planned for September 2020 (tbc), and the Board discussion of the Report's yellow cover is planned for October 2020 (tbc). The WDR 2021 will be launched in January 2021 and the team is considering various key events for the launch.



### Annex 1: Advisory Panel members (To be confirmed)

*\* The members are being consulted and will be confirmed based on their availability and sector/regional representations. Below table provides an idea of the different types of organizations that are being considered for the Advisory Panel.*

<b>Type</b>	<b>Region</b>
<b>Civil Society Organization</b>	OECD
<b>Competition Agency</b>	Africa
<b>Foundation</b>	OECD
<b>International Organizations</b>	OECD
<b>National Statistics Office</b>	MENA
<b>Politician</b>	OECD
<b>Politician</b>	OECD
<b>Platform Business</b>	LAC
<b>Technology Company</b>	Asia

Annex 2: Technical Board (To be confirmed)

*\*The members are being current consulted, and the list will be updated in the WDR website once their availability is confirmed.*

## Annex 3: Brains Trust members (Bank staff from various units)

<b>Staff</b>	<b>Title</b>	<b>Bank Unit</b>
<b>Andrew L. Dabalen</b>	Practice Manager	EA2PV
<b>Bill Maloney</b>	Chief Economist	GGEVP
<b>Daniel Lederman</b>	Lead Economist	MNACE
<b>Davide Strusani</b>	Principal Sector Economist	CSEDR
<b>Fredesvinda F. Montes Herraiz</b>	Senior Financial Sector Specialist	EFNFI
<b>Gero Carletto</b>	Manager	DECPM
<b>Haishan Fu</b>	Director	DECDG
<b>Joao Pedro Wagner De Azevedo</b>	Lead Economist	HEDGE
<b>Junaid Kamal Ahmad</b>	Country Director	SACIN
<b>Kathleen Beegle</b>	Lead Economist	HGNDR
<b>Kimberly D Johns</b>	Senior Public Sector Specialist	EMNGU
<b>Luis Alberto Andres</b>	Lead Economist	SAFW2
<b>Marelize Gorgens</b>	Senior Monitoring and Evaluation Specialist	HHNGE
<b>Marianne Fay</b>	Country Director	LCC6C
<b>Mary Hallward-Driemeier</b>	Senior Economic Advisor	ETIDR
<b>Michael Ferrantino</b>	Lead Economist	ETIRI
<b>Paolo Verme</b>	Lead Economist, Manager of the Research program on Forced Displacement Lead Economist	GTFSa
<b>Saki Kumagai</b>	Governance Specialist	
<b>Sharada Srinivasan</b>	Young Professional	IDD01
<b>Tania Begazo</b>	Senior Economist	IDDDR
<b>Tim Kelly</b>	Lead Digital Development Specialist	IDD02
<b>Umar Serajuddin</b>	Manager	DECIS
<b>Vyjayanti Desai</b>	Program Manager	IDD03

## Notes

---

<sup>1</sup> Estimate based on tables from the World Bank's World Development Indicators indicating that 86.8% of the world's population has access to electricity in 2015 and based on United Nation's Department of Economic and Social Affairs estimates of global population of 7.38 billion people in 2015 (World Bank, 2015).

<sup>2</sup> Rowntree (2000).

<sup>3</sup> World Bank (2016).

<sup>4</sup> Kelly et al. (2018).

<sup>5</sup> World Bank (2018c).

<sup>6</sup> Wesolowski et al. (2015).

<sup>7</sup> Burke and Lobell (2017), Osgood-Zimmerman et al. (2018).

<sup>8</sup> Blumenstock et al. (2016).

<sup>9</sup> Lu et al. (2012).

<sup>10</sup> An effective data steward maintains agreed-upon data definitions and formats, identifies data quality issues and ensures that users adhere to specified data standards. On the private side, data stewards' responsibilities can also include helping to identify and articulate ways to utilize corporate data to create competitive advantages in the market.

<sup>11</sup> Kilic et al. (2017).

<sup>12</sup> Serajuddin et al. (2015).

<sup>13</sup> This is essentially remarking that data is a good; Coyle et al. (2020).

<sup>14</sup> This is similar to suggesting that there is a nonconcavity in the value of data and information (Radner and Stiglitz, 1984), and is also linked to the point that because ideas are nonrivalrous, they exhibit increasing marginal returns over a range (Romer, 1990).

<sup>15</sup> The firms that participated in the study were not randomly selected nor were they selected from a frame of US firms. The intent of the study is primarily to convey the diverse usage of public data from essentially all facets of government by a wide array of firms from many different sectors.

<sup>16</sup> A series of short notes on data aspects of the COVID-19 crisis will be released ahead of the launch of the main report. This will provide relevant just-in-time material for the global policy debate on COVID response, as well as helping to build interest in the coming report.

<sup>17</sup> While the issue of personal data protection is a critical element in this process, this concern will be addressed in greater detail in second part of this report.

<sup>18</sup> *Censuses* aim to systematically enumerate, and record information about, an entire population of interest, whether individuals, businesses, farms or others; and because they serve as frames, they are the lynchpin to ensuring that everyone is represented in sample survey data. *Sample surveys* typically collect detailed information about a population of interest, whether individuals, firms, or other domains for policy importance. *Administrative data* typically results from the process of registration, usually by national authorities. Administrative data include civil registers of a person's vital events, such as birth, marriage, or death, population records, health and tax records, or trade flows. *Geospatial data* includes satellite imagery of the earth, such as provided by NASA's Landsat and ESA's Copernicus Programme, and weather data often recorded from sampled geographic points.

<sup>19</sup> World Bank (2018a); Bedi et al. (2007).

<sup>20</sup> Corral et al. (2019).

<sup>21</sup> SDSN (2015).

<sup>22</sup> In total more than 1000 reforms have been inspired by Doing Business since 2003. World Bank (2020a).

- 
- <sup>23</sup> United Nations (2018).
- <sup>24</sup> Cameron et al. (2019).
- <sup>25</sup> Doss et al. (2019); Hoogeveen and Pape (2020).
- <sup>26</sup> The fourth pillar of the IDA19 replenishment contains a commitment to improve data for evidence-based policy making through the Data for Policy package, which for a number of IDA countries aims to support a set of core statistics including survey and administrative data, boost the analytical power of existing data, improve the ability to disaggregate statistics by vulnerable groups, stimulate data utilization by governments and citizens and much more (World Bank, 2019b).
- <sup>27</sup> Brynjolfsson and Hitt (2000); Bloom et al. (2012); Aral et al. (2012).
- <sup>28</sup> Corrado et al. (2017); Mohnen et al. (2018).
- <sup>29</sup> Bugamelli and Pagano (2004).
- <sup>30</sup> Gust and Marquez (2004); Bartelsman (2013).
- <sup>31</sup> Gal et al. (2019).
- <sup>32</sup> Akerman et al. (2013); Dhyne et al. (2018).
- <sup>33</sup> ITU World Telecommunication/ICT Indicators database.
- <sup>34</sup> Blumenstock(2018).
- <sup>35</sup> World Economic Forum (2013).
- <sup>36</sup> Bjorkegren and Grissen (2015).
- <sup>37</sup> McCaffrey and Schiff (2017).
- <sup>38</sup> Arraiz et al. (2015).
- <sup>39</sup> USAID (2019).
- <sup>40</sup> Sarwar (2017).
- <sup>41</sup> Singh and Landry (2019).
- <sup>42</sup> USAID (2019).
- <sup>43</sup> USAID (2019).
- <sup>44</sup> Deaton (2008); Falk et al. (2018); for a list of projects that have use Gallup World Poll data see also <https://www.gallup.com/services/170945/worldpoll.aspx> (Gallup Inc, 2020)
- <sup>45</sup> For example, the United Nations Food and Agricultural Organizations has added questions to the Gallup World Poll to collect data for its Food Insecurity Experience Scale starting in 2014 (goal 2: end hunger, achieve food security and improved nutrition, and promote sustainable agriculture). And in 2015 the International Labor Organization and Walk Free Foundation started adding questions that measure the incidence of modern slavery (goal 8.7: eradicate forced labor, child labor, modern slavery and human trafficking).
- <sup>46</sup> Järv et al. (2012).
- <sup>47</sup> Lu et al. (2012); Bengtsson et al. (2011).
- <sup>48</sup> Bhradury et al. (2018); Bok et al. (2018).
- <sup>49</sup> McCarthy et al. (2018)
- <sup>50</sup> Tiecke et al. (2017).
- <sup>51</sup> Wardrop et al. (2018); Lai et al. (2019); Blumenstock et al. (2015); Steele et al. (2017); Engstrom et al. (2017).
- <sup>52</sup> Lobell et al. (2020a); Lobell et al. (2020b).
- <sup>53</sup> Pratihast et al. (2014).
- <sup>54</sup> Hansen et al. (2016).

- 
- <sup>55</sup> Similarly, a recent initiative monitors illegal deforestation in countries from the world's 10 largest palm oil producers that has created a feedback system where civil society monitors the private sector in a way which allows for government action (World Resource Institute, 2019).
- <sup>56</sup> Kraemer et al. (2019); Yang et al. (2017); Milinovich et al. (2014).
- <sup>57</sup> Lu et al. (2012).
- <sup>58</sup> Hughes-Cromwick and Coronado (2019).
- <sup>59</sup> Manyika et al. (2013).
- <sup>60</sup> Adrason and Schalwyk (2017).
- <sup>61</sup> CISCO (2019).
- <sup>62</sup> Tefficient (2019).
- <sup>63</sup> Countries must have root nameservers, country-code top-level domain (ccTLD) nameservers, and Internet exchange points (IXPs) within their borders in order to maintain autonomy and internal connectivity during periods when international cables are damaged. <https://www.pch.net/ixp/summary> (Internet Society, 2019).
- <sup>64</sup> Marsan (2014).
- <sup>65</sup> Statistics South Africa (2019).
- <sup>66</sup> GSMA (2018).
- <sup>67</sup> World Bank (2016); World Bank (2018b).
- <sup>68</sup> Over the Top (OTT) applications are services provided over the internet and not directly by the internet service providers. Examples of these are WhatsApp, Skype, Netflix, etc.
- <sup>69</sup> UNCTAD (2019).
- <sup>70</sup> Ericsson (2019).
- <sup>71</sup> García Zaballos and Iglesias (2017).
- <sup>72</sup> Kende (2017).
- <sup>73</sup> Many Possibilities (2019).
- <sup>74</sup> Argenton and Prüfer (2012).
- <sup>75</sup> Bauer et al. (2016).
- <sup>76</sup> OECD (2019).
- <sup>77</sup> MGI (2016).
- <sup>78</sup> Lendle (2012).
- <sup>79</sup> In practice, there may be important cross-effects between these three areas as creating a level playing field for taxation of digital businesses and opening-up to international trade in digital services will intensify competitive pressures in the affected sectors of the economy.
- <sup>80</sup> Bieron (2015); Ferracane (2019).
- <sup>81</sup> Arnold (2014); Duggan (2013).
- <sup>82</sup> NBT (2015).
- <sup>83</sup> USITC (2017).
- <sup>84</sup> Cremer et al. (2019).
- <sup>85</sup> HM Treasury (2019).
- <sup>86</sup> Stigler Center (2019).
- <sup>87</sup> This will build on the WBG IC Unit's forthcoming policy note on antitrust in the digital economy as well as recent reports on the topic from various jurisdictions including: Stigler Center (2019); BRICS Competition Innovation Law & Policy Centre (2019); Núñez Reyes (2018); Cremer (2019); Indian Ministry of Commerce and Industry (2019); HM Treasury (2019); OECD (2018a).

---

<sup>88</sup> India's prohibition on Amazon and Flipkart selling their own products on their platforms is the most notable recent example. In the Fintech sector: The Malawian antitrust authority advocated with the banking association to share data on credit histories; and in Mexico, the antitrust authority advocated for regulation of terms/fees that banks may charge for sharing customer transaction data.

<sup>89</sup> This section will draw on the Digital Antitrust Database and the Competition Advocacy Database (both global, by WB-FCI-IC).

<sup>90</sup> Hillman (2018).

<sup>91</sup> Ferracane and van der Marel (2019).

<sup>92</sup> Beer and Loerprick (2015); OECD (2018b).

<sup>93</sup> World Bank (2019a).

<sup>94</sup> Less developed countries can be disproportionately affected by international tax avoidance, tax evasion and tax base erosion, as taxes collected from Multinational Enterprises (MNEs) generally constitute a larger share of their total tax revenues compared to developed countries. This is particularly important for countries with significant natural resources (Beer and Loerprick, 2017). While there is still no agreement on direct taxation of the digital economy, a consensus approach in the OECD's International VAT/GST Guidelines can guide developing economy policy makers in imposing VAT on the direct supply to consumers of services and intangibles by foreign suppliers.

<sup>95</sup> Efforts are ongoing to agree on a new basis for allocating profits when MNEs interact digitally with an economy, generating significant profits but without having a taxable presence under current rules. The G20/OECD envisage agreement for a framework of reform amongst the 136 member countries of the 'Inclusive Framework on BEPS' in January/February 2020 and detailed reform proposals to be adopted in June 2020.

<sup>96</sup> World Bank (2017).

<sup>97</sup> Data Lifecycle is the different stages that a unit of data goes through from its initial creation to deletion.

<sup>98</sup> OECD (2013).

<sup>99</sup> Pathways for Prosperity Commission (2019).

<sup>100</sup> CSIS (2019).

<sup>101</sup> GovLab (2020b); Development Data Partnership (2020); World Bank (2020b).

<sup>102</sup> OMF (2020).

<sup>103</sup> Humdata (2020).

<sup>104</sup> Digital Society Lab (2020).

<sup>105</sup> Bruhn et al. (2014).

<sup>106</sup> Bruhn et al. (2016); Frisancho (2018); Lührmann et al. (2018).

<sup>107</sup> Entorf and Hu (2018).

<sup>108</sup> The Statistical Performance Index (SPI) is a revised and improved measure from the previous Statistical Capacity Indicator (SCI) measure (World Bank, 2020c).

## References

- Adrason, Alex, and Francois van Schalkwyk. 2016. "Open Data Intermediaries in the Agricultural Sector in Ghana." Washington, DC: World Wide Web Foundation.  
<http://webfoundation.org/docs/2016/12/WF-RP-Open-Data-Intermediaries-in-Agriculture-Ghana-Update.pdf>.
- Akerman, Anders, Ingvil Gaarder, and Magne Mogstad. 2015. "The Skill Complementarity of Broadband Internet." *The Quarterly Journal of Economics* 130 (4): 1781–1824.  
<https://doi.org/10.1093/qje/qjv028>.
- Aral, Sinan, Erik Brynjolfsson, and Lynn Wu. 2012. "Three-Way Complementarities: Performance Pay, Human Resource Analytics, and Information Technology." *Management Science*, March. <https://doi.org/10.1287/mnsc.1110.1460>.
- Argenton, Cédric, and Jens Prüfer. 2012. "Search Engine Competition with Network Externalities." *Journal of Competition Law and Economics* 8 (1): 73–105.
- Arraiz, I., M. Bruhn, and R. Stucchi. (2015). "Psychometrics as a Tool to Improve Screening and Access to Credit." IDB Working Paper Series 625. Washington, DC: Inter-American Development Bank. [https://www.eflglobal.com/wp-content/uploads/2015/10/Psychometrics\\_EFL1.pdf](https://www.eflglobal.com/wp-content/uploads/2015/10/Psychometrics_EFL1.pdf)
- Bartelsman, Eric J. 2013. "ICT, Reallocation and Productivity." 486. European Economy - Economic Papers 2008 - 2015. Directorate General Economic and Financial Affairs (DG ECFIN), European Commission. <https://ideas.repec.org/p/euf/ecopap/0486.html>.
- Bauer, Matthias, Martina F. Ferracane, and Erik van der Marel. 2016. "Tracing the Economic Impact of Regulations on the Free Flow of Data and Data Localization," May. <https://www.cigionline.org/publications/tracing-economic-impact-regulations-free-flow-data-and-data-localization>.
- Bedi, Tara, Aline Coudouel Aline, and Kenneth Simler. 2007. *More Than a Pretty Picture: Using Poverty Maps to Design Better Policies and Interventions*. Washington, DC: World Bank.
- Beer, Sebastian, and Jan Loeprick. 2015. "Profit Shifting: Drivers of Transfer (Mis)Pricing and the Potential of Countermeasures." *International Tax and Public Finance* 22 (3): 426–51.  
<https://doi.org/10.1007/s10797-014-9323-2>.
- Bengtsson, Linus, Xin Lu, Anna Thorson, Richard Garfield, and Johan von Schreeb. 2011. "Improved Response to Disasters and Outbreaks by Tracking Population Movements with Mobile Phone Network Data: A Post-Earthquake Geospatial Study in Haiti." *PLOS Medicine* 8 (8): e1001083. <https://doi.org/10.1371/journal.pmed.1001083>.
- Bhadury, Soumya, Sanjib Pohit, and Robert C. M. Beyer. 2018. "A New Approach to Nowcasting Indian Gross Value Added." 115. NCAER Working Papers. National Council of Applied Economic Research. <https://ideas.repec.org/p/nca/ncaerw/115.html>.



- Bieron, Brian, and Usman Ahmed. 2015. "E15 Initiative | Services, International Rulemaking, and the Digitisation of Global Commerce." *E15 Initiative* (blog). April 2015. <http://e15initiative.org/publications/services-international-rulemaking-and-the-digitization-of-global-commerce/>.
- Bjorkegren, D. and D. Grissen. (2015). "Behavior Revealed in Mobile Phone Usage Predicts Loan Repayment." [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2611775](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2611775)
- Blumenstock, Joshua, Gabriel Cadamuro, and Robert On. 2015. "Predicting Poverty and Wealth from Mobile Phone Metadata." *Science* 350 (6264): 1073–76. <https://doi.org/10.1126/science.aac4420>.
- Blumenstock, Joshua, Nathan Eagle, and Marcel Fafchamps, 2016. "Airtime transfers and mobile communications: Evidence in the aftermath of natural disasters." *Journal of Development Economics* 120: 157-181.
- Blumenstock, Joshua. 2018. "Don't Forget People in the Use of Big Data for Development." *Nature* 561 (7722): 170–72. <https://doi.org/10.1038/d41586-018-06215-5>.
- Bok, Brandyn, Daniele Caratelli, Domenico Giannone, Argia M. Sbordon, and Andrea Tambalotti. 2018. "Macroeconomic Nowcasting and Forecasting with Big Data." *Annual Review of Economics* 10 (1): 615–43. <https://doi.org/10.1146/annurev-economics-080217-053214>.
- Bloom, Nicholas, Raffaella Sadun, and John Van Reenen. 2012. "Americans Do IT Better: US Multinationals and the Productivity Miracle." *American Economic Review* 102 (1): 167–201. <https://doi.org/10.1257/aer.102.1.167>.
- BRICS Competition Innovation Law & Policy Centre. 2019. "Digital Era Competition BRICS Report." <http://bricscompetition.org/materials/news/digital-era-competition-brics-report/>.
- Bruhn, Miriam, Gabriel Lara Ibarra, and David McKenzie. 2014. "The Minimal Impact of a Large-Scale Financial Education Program in Mexico City." *Journal of Development Economics* 108 (May): 184–89. <https://doi.org/10.1016/j.jdeveco.2014.02.009>.
- Bruhn, Miriam, Luciana de Souza Leão, Arianna Legovini, Rogelio Marchetti, and Bilal Zia. 2016. "The Impact of High School Financial Education: Evidence from a Large-Scale Evaluation in Brazil." *American Economic Journal: Applied Economics* 8 (4): 256–95. <https://doi.org/10.1257/app.20150149>.
- Brynjolfsson, Erik, and Lorin M. Hitt. 2000. "Beyond Computation: Information Technology, Organizational Transformation and Business Performance." *Journal of Economic Perspectives* 14 (4): 23–48. <https://doi.org/10.1257/jep.14.4.23>.
- Bugamelli, Matteo, and Patrizio Pagano. 2004. "Barriers to Investment in ICT." *Applied Economics* 36 (20): 2275–86. <https://doi.org/10.1080/0003684042000270031>.

- Burke, Marshall and David A. Lobell. 2017. "Satellite-based assessment of yield variation and its determinants in smallholder African systems." *Proceedings of the National Academy of Sciences (USA)* 114(9): 2189-2194.
- Cameron, Grant James, Hai-Anh Dang, Mustafa Dinc, James Stephen Foster, and Michael M Lokshin. 2019. "Measuring the Statistical Capacity of Nations." *Policy Research Working Paper; No. WPS 8693*.  
<http://documents.worldbank.org/curated/en/304431546956224461/Measuring-the-Statistical-Capacity-of-Nations>.
- Cisco. 2019. *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*.  
<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>.
- Corral, Paul, Joao Pedro Azevedo, Jonathan Karver, and Reena Badiani-Magnusson. 2019. "Going Municipal: Targeting Deprivation with New Evidence in Croatia." *Let's Talk Development*, World Bank, Washington, DC.  
<https://blogs.worldbank.org/developmenttalk/going-municipal-targeting-deprivation-new-evidence-croatia>
- Corrado, Carol, Jonathan Haskel, and Cecilia Jona-Lasinio. 2017. "Knowledge Spillovers, ICT and Productivity Growth." *Oxford Bulletin of Economics and Statistics* 79 (4): 592–618.  
<https://doi.org/10.1111/obes.12171>.
- Coyle, Diane, Stephannie Diepeveen, Julia Wdowin, Jeni Tennison, and Lawrence Kay. 2020. "The Value of Data - Policy Implications." Cambridge, United Kingdom: Bennett Institute and Open Data Institute. <https://www.bennettinstitute.cam.ac.uk/publications/value-data-policy-implications/>.
- Cremer, Charmes, Jacques, Yves-Alexandre de Montjoye, and Heike Schweitzer. 2019. "Competition Policy for The Digital Era." Website. Brussels: European Commission.  
<https://op.europa.eu:443/en/publication-detail/-/publication/21dc175c-7b76-11e9-9f05-01aa75ed71a1/language-en>.
- Croatian Bureau of Statistics (2016). *Small Area Estimates of Income Poverty in Croatia: Methodological Report*.  
[http://www.dzs.hr/ENG/DBHomepages/Personal%20Consumption%20and%20Poverty%20Indicators/Methodology\\_SILC\\_WB.pdf](http://www.dzs.hr/ENG/DBHomepages/Personal%20Consumption%20and%20Poverty%20Indicators/Methodology_SILC_WB.pdf)
- CSIS (Center for Strategic & International Studies). 2019. "Data Governance Principles for the Global Digital Economy." June 4, 2019. <https://www.csis.org/analysis/data-governance-principles-global-digital-economy>.

- Deaton, Angus. 2008. "Income, Health and Wellbeing Around the World: Evidence from the Gallup World Poll." *The Journal of Economic Perspectives : A Journal of the American Economic Association* 22 (2): 53–72. <https://doi.org/10.1257/jep.22.2.53>.
- Development Data Partnership. 2020. "The Development Data Partnership." Accessed February 2, 2020. <https://datapartnership.org/#>.
- Dhyne, Emmanuel, Joep Konings, Joep Konings, and Stijn Vanormelingen. 2018. "IT and Productivity: A Firm Level Analysis." 346. Working Paper Research. National Bank of Belgium. <https://ideas.repec.org/p/nbb/reswpp/201810-346.html>.
- Digital Civil Society Lab. 2020. "A Framework for Data Trusts." *Stanford PACS* (blog). 2020. <https://pacscenter.stanford.edu/research/digital-civil-society-lab/a-framework-for-data-trusts/>.
- Doss, Cheryl, Caitlin Kieran, and Talip Kilic. 2019. "Measuring Ownership, Control, and use of Assets." *Feminist Economics* 26(1): 1–25.
- Duggan, Victor, Sjamsu Rahardja, and Gonzalo Varela. 2013. *Service Sector Reform and Manufacturing Productivity: Evidence from Indonesia*. Policy Research Working Papers. Washington, DC: The World Bank. <https://doi.org/10.1596/1813-9450-6349>.
- Engstrom, Ryan, Jonathan Samuel Hersh, and David Locke Newhouse. 2017. "Poverty from Space : Using High-Resolution Satellite Imagery for Estimating Economic Well-Being." World Bank Policy Research Paper 8284. The World Bank.
- Entorf, Horst, and Jia Hou. 2018. "Financial Education for the Disadvantaged? A Review." SSRN Scholarly Paper ID 3167709. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3167709>.
- Ericsson. 2019. "Ericsson Mobility Report November 2019." <https://www.ericsson.com/en/mobility-report/reports/november-2019>
- European Union. 2016. *Regulation 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data and Repealing Directive 95/46/EC*. <https://eur-lex.europa.eu/eli/reg/2016/679/oj>.
- Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences." *The Quarterly Journal of Economics* 133 (4): 1645–92. <https://doi.org/10.1093/qje/qjy013>.
- Ferracane, Martina Francesca, and Erik Van Der Marel. 2019. "Do Data Policy Restrictions Inhibit Trade in Services?" DTE Working Paper No. 2. Brussels: European Center for International Political Economy. <http://cadmus.eui.eu//handle/1814/62325>.

- Frisancho, Verónica. 2018. "The Impact of School-Based Financial Education on High School Students and Their Teachers: Experimental Evidence from Peru." Inter-American Development Bank. <https://doi.org/10.18235/0001056>.
- Gal, Peter, Giuseppe Nicoletti, Theodore Renault, Stéphane Sorbe, and Christina Timiliotis. 2019. "Digitalisation and Productivity: In Search of the Holy Grail – Firm-Level Empirical Evidence from EU Countries," February. <https://doi.org/10.1787/5080f4b6-en>.
- Gallup Inc. 2020. "Gallup World Poll." What the Whole World Is Thinking. Gallup.Com. January 24, 2020. <https://www.gallup.com/services/170945/world-poll.aspx>.
- García Zaballos, Antonio, and Enrique Iglesias. 2017. "Data Centers and Broadband for Sustainable Economic and Social Development: Evidence from Latin America and the Caribbean." Inter-American Development Bank. <https://doi.org/10.18235/0000692>.
- GovLab. 2020a. "Open Data 500." Accessed February 3, 2020. <https://www.opendata500.com/us/>.
- GovLab. 2020b. "Data Collaboratives." Accessed February 2, 2020. <http://datacollaboratives.org/>.
- GSMA. 2018. "The Mobile Gender Gap Report." [https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/04/GSMA The Mobile Gender Gap Report 2018 32pp WEbV7.pdf](https://www.gsma.com/mobilefordevelopment/wp-content/uploads/2018/04/GSMA-The-Mobile-Gender-Gap-Report-2018-32pp-WEbV7.pdf)
- Gust, Christopher, and Jaime Marquez. 2004. "International Comparisons of Productivity Growth: The Role of Information Technology and Regulatory Practices." *Labour Economics*, Labour market consequences of new information technologies, 11 (1): 33–58. [https://doi.org/10.1016/S0927-5371\(03\)00055-1](https://doi.org/10.1016/S0927-5371(03)00055-1).
- Hansen, Matthew C., Alexander Krylov, Alexandra Tyukavina, Peter V. Potapov, Svetlana Turubanova, Bryan Zutta, Suspense Ifo, Belinda Margono, Fred Stolle, and Rebecca Moore. 2016. "Humid Tropical Forest Disturbance Alerts Using Landsat Data." *Environmental Research Letters* 11 (3). <https://doi.org/10.1088/1748-9326/11/3/034008>.
- Hillman, Jonathan E. 2018. "The Global Battle for Digital Trade." CSIS (Center for Strategic & International Studies). April 13, 2018. <https://www.csis.org/blogs/future-digital-trade-policy-and-role-us-and-uk/global-battle-digital-trade>.
- HM Treasury. 2019. "Unlocking Digital Competition, Report of the Digital Competition Expert Panel." London: HM Treasury. <https://www.gov.uk/government/publications/unlocking-digital-competition-report-of-the-digital-competition-expert-panel>.
- Humdata. (2020) "Welcome - Humanitarian Data Exchange." Accessed February 2, 2020. <https://data.humdata.org/>.

- Hoogeveen, Johannes, and Utz Pape. 2020. *"Data Collection in Fragile States: Innovations from Africa and Beyond."* Palgrave MacMillan.
- Hughes-Cromwick, Ellen, and Julia Coronado. 2019. "The Value of US Government Data to US Business Decisions." *Journal of Economic Perspectives* 33 (1): 131–46.  
<https://doi.org/10.1257/jep.33.1.131>.
- Indian Ministry of Commerce and Industry. 2019. "Draft National E-Commerce Policy: India's Data for India's Development." 2019.  
[https://dipp.gov.in/sites/default/files/DraftNational\\_e-commerce\\_Policy\\_23February2019.pdf](https://dipp.gov.in/sites/default/files/DraftNational_e-commerce_Policy_23February2019.pdf).
- Internet Society. 2019. "Policy Brief: Internet Shutdowns." *Internet Society* (blog). December 18, 2019. <https://www.internetsociety.org/policybriefs/internet-shutdowns>.
- Järv, Olle, Rein Ahas, Erki Saluveer, Ben Derudder, and Frank Witlox. 2012. "Mobile Phones in a Traffic Flow: A Geographical Perspective to Evening Rush Hour Traffic Analysis Using Call Detail Records." Edited by Renaud Lambiotte. *PLoS ONE* 7 (11): e49171.  
<https://doi.org/10.1371/journal.pone.0049171>.
- Kelly, Tim. 2016. "How the WDR16 Policy Framework Is Applied in the Union of Comoros." January 13, 2016. <http://blogs.worldbank.org/digital-development/how-wdr16-policy-framework-applied-union-comoros>.
- Kelly, Timothy John Charles, Rokuhei Fordyce Fukui, Michael Minges, Phillippa Biggs, Felicia Vacarelu, Miguel Luengo-Oroz, Mila Romanoff, et al. 2018. "Information and Communication for Development 2018: Data-Driven Development." 128301. Washington, DC: World Bank.  
<http://documents.worldbank.org/curated/en/987471542742554246/Information-and-Communication-for-Development-2018-Data-Driven-Development>.
- Kende, Michael, and Bastiaan Quast. 2017. "Local Content Hosting: Speed, Visits, & Cost of Access." Internet Society (blog). April 11, 2017.  
<https://www.internetsociety.org/resources/doc/2017/a-case-study-in-local-content-hosting-speed-visits-and-cost-of-access/>.
- Kilic, Talip, Umar Serajuddin, Hiroki Uematsu, and Nobuo Yoshida. 2017. "Costing Household Surveys for Monitoring Progress toward Ending Extreme Poverty and Boosting Shared Prosperity." WPS7951. The World Bank.  
<http://documents.worldbank.org/curated/en/260501485264312208/Costing-household-surveys-for-monitoring-progress-toward-ending-extreme-poverty-and-boosting-shared-prosperity>.
- Kraemer, Moritz U. G., Nick Golding, Donal Bisanzio, Samir Bhatt, David M. Pigott, Sarah E. Ray, Oliver J. Brady, John S. Brownstein, Nuno R. Faria, Derek A. T. Cummings, Oliver G. Pybus, David L. Smith, Andrew J. Tatem, Simon I. Hay, and Robert C. Reiner. 2019. "Utilizing

- general human movement models to predict the spread of emerging infectious diseases in resource poor settings.” *Scientific Reports* 9 (11): 1-11 (2019).
- Lai, Shengjie, Elisabeth zu Erbach-Schoenberg, Carla Pezzulo, Nick W. Ruktanonchai, Alessandro Sorichetta, Jessica Steele, Tracey Li, Claire A. Dooley, and Andrew J. Tatem. 2019. “Exploring the Use of Mobile Phone Data for National Migration Statistics.” *Palgrave Communications* 5 (1): 1–10. <https://doi.org/10.1057/s41599-019-0242-9>.
- Lendle, Andreas, Marcelo Olarreaga, Simon Schropp, and Pierre-Louis Vezina. 2012. “There Goes Gravity : How EBay Reduces Trade Costs.” WPS6253. Washington, D.C: The World Bank. <http://documents.worldbank.org/curated/en/260421468147866905/There-goes-gravity-how-eBay-reduces-trade-costs>.
- Lobell, David B., George Azzari, Marshall Burke, Sydney Gourlay, Zhenong Jin, Talip Kilic, and Siobhan Murray. 2020a. “Eyes in the Sky, Boots on the Ground: Assessing Satellite- and Ground-Based Approaches to Crop Yield Measurement and Analysis.” *American Journal of Agricultural Economics* 102 (1): 202–19. <https://doi.org/10.1093/ajae/aaz051>.
- Lobell, David B., Stefania Di Tommaso, Calum You, Ismael Yacoubou Djima, Marshall Burke, and Talip Kilic. 2020b. “Sight for Sorghums: Comparisons of Satellite- and Ground-Based Sorghum Yield Estimates in Mali.” *Remote Sensing* 12 (1): 100. <https://doi.org/10.3390/rs12010100>.
- Lu, Xin, Linus Bengtsson, and Petter Holme, 2012. “Predictability of population displacement after the 2010 Haiti earthquake.” *Proceedings of the National Academy of Sciences* 109 (29): 11576–81.
- Lührmann, Melanie, Marta Serra-Garcia, and Joachim Winter. 2018. “The Impact of Financial Education on Adolescents’ Intertemporal Choices.” *American Economic Journal: Economic Policy* 10 (3): 309–32. <https://doi.org/10.1257/pol.20170012>.
- Manyika, James, Michael Chui, Diana Farrell, Steve Van Kuiken, Peter Groves, and Elizabeth Almasi Doshi. 2013. “Open Data: Unlocking Innovation and Performance with Liquid Information.” McKinsey Global Institute. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/open-data-unlocking-innovation-and-performance-with-liquid-information>.
- Manyika, James, Susan Lund, Jacques Bughin, Jonathan Woetzel, Kalin Stamenov, and Dhruv Dhingra. n.d. “Digital Globalization: The New Era of Global Flows.” McKinsey Global Institute. Accessed January 22, 2020. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/digital-globalization-the-new-era-of-global-flows>.
- Many Possibilities. 2019. “Africa Telecoms Infrastructure in 2019.” Many Possibilities (blog). 2019. <https://manypossibilities.net/2020/01/africa-telecoms-infrastructure-in-2019/>.

- Marsan, Carolyn Duffy. 2014. "Experts Say Economics and Politics Hamper Efficient Routing of Internet Data." IETF Journal, November. <https://www.ietfjournal.org/experts-say-economics-and-politics-hamper-efficient-routing-of-internet-data/>.
- McCarthy, Nancy, Talip Kilic, Alejandro de la Fuente, and Joshua M. Brubaker. 2018. "Shelter from the Storm? Household-Level Impacts of, and Responses to, the 2015 Floods in Malawi." *Economics of Disasters and Climate Change* 2 (3): 237–58. <https://doi.org/10.1007/s41885-018-0030-9>.
- McCaffrey, Mike, and Annabel Schiff. 2017. "Finclusion to Fintech: Fintech Product Development for Low-Income Markets." SSRN Scholarly Paper ID 3034175. Rochester, NY: Social Science Research Network. <https://papers.ssrn.com/abstract=3034175>.
- Milinovich, Gabriel J., Gail M. Williams, Archie C. A. Clements, and Wenbiao Hu. 2014. "Internet-Based Surveillance Systems for Monitoring Emerging Infectious Diseases." *The Lancet Infectious Diseases* 14 (2): 160–68.
- Mohnen, Pierre, Michael Polder, and George van Leeuwen. 2018. "ICT, R&D and Organizational Innovation: Exploring Complementarities in Investment and Production." Working Paper 25044. National Bureau of Economic Research. <https://doi.org/10.3386/w25044>.
- Núñez Reyes, Georgina, Júlia De Furquim, and Marcelo Pereira Dolabella. 2018. "Políticas de competencia para una economía digital: el marco regulatorio e institucional y el contexto internacional," June. <https://repositorio.cepal.org/handle/11362/43630>.
- OECD (Organization for Economic Co-operation and Development). 2013. "OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data - OECD." 2013. <https://www.oecd.org/internet/ieconomy/oecdguidelinesontheprivacyandtransborderflowsofpersonaldata.htm>.
- OECD. 2016. "Big Data: Bringing Competition Policy to the Digital Era." Background paper DAF/COMP(2016)14. Paris: Organization for Economic Cooperation and Development. [https://one.oecd.org/document/DAF/COMP\(2016\)14/en/pdf](https://one.oecd.org/document/DAF/COMP(2016)14/en/pdf).
- OECD. 2018a. "Rethinking Antitrust Tools for Multi-Sided Platforms 2018." Paris: Organisation for Economic Cooperation and Development. <https://www.oecd.org/competition/rethinking-antitrust-tools-for-multi-sided-platforms.htm>.
- OECD. 2018b. "Tax Challenges Arising from Digitalisation – Interim Report 2018 - Inclusive Framework on BEPS - En - OECD." OECD/G20 Base Erosion and Profit Shifting Project. Paris: Organisation for Economic Cooperation and Development. <http://www.oecd.org/ctp/tax-challenges-arising-from-digitalisation-interim-report-9789264293083-en.htm>.

- OECD. 2019. "Enhancing Access to and Sharing of Data - Reconciling Risks and Benefits for Data Re-Use Across Societies." Organisation for Economic Cooperation and Development. <https://www.oecd.org/sti/enhancing-access-to-and-sharing-of-data-276aaca8-en.htm>.
- OMF (Open Mobility Foundation). 2020. "Open Mobility Foundation | OMF." Accessed February 2, 2020. <https://www.openmobilityfoundation.org/>.
- Osgood-Zimmerman, Aaron, et al. 2018. "Mapping child growth failure in Africa between 2000 and 2015." *Nature* 555 (41): 41-47.
- Pathways for Prosperity Commission. 2019. "Digital Diplomacy: Technology Governance for Developing Countries." Oxford, UK: Pathways for Prosperity Commission. <https://pathwayscommission.bsg.ox.ac.uk/digital-diplomacy>.
- Pratihast, Arun Kumar, Ben DeVries, Valerio Avitabile, Sytze De Bruin, Lammert Kooistra, Mesfin Tekle, and Martin Herold. 2014. "Combining Satellite Data and Community-Based Observations for Forest Monitoring." *Forests* 5 (10): 2464–89. <https://doi.org/10.3390/f5102464>.
- Radner, Roy, and Joseph Stiglitz. 1984. "A Nonconcavity in the Value of Information." In *Bayesian Models in Economic Theory*, edited by Marcel Boyer and Richard E. Kihlstrom, 33–52. Amsterdam: Elsevier.
- Romer, Paul M. 1990. "Endogenous Technological Change." *Journal of Political Economy* 98 (5): S71–102.
- Rowntree, Benjamin Seebohm. 2000. *Poverty: A Study of Town Life*. Centennial ed., (Original publication, London, 1901). Bristol: Policy Press.
- Sarwar, Mahrukh. 2017. "How Pakistan Turned Around Its Vaccination Program Using Technology." Information Technology University Punjab. *MIT Technology Review Pakistan* (blog). January 12, 2017. <http://www.technologyreview.pk/pakistan-turned-around-vaccination-program-using-technology/>.
- Serajuddin, Umar, Hiroki Uematsu, Christina Wieser, Nobuo Yoshida, and Andrew L. Dabalen. 2015. "Data Deprivation: Another Deprivation to End." WPS7252. The World Bank. <http://documents.worldbank.org/curated/en/700611468172787967/Data-deprivation-another-deprivation-to-end>.
- Singh, Poonam Khetrpal, and Mark Landry. 2019. "Harnessing the Potential of Digital Health in the WHO South-East Asia Region: Sustaining What Works, Accelerating Scale-up and Innovating Frontier Technologies." *WHO South-East Asia Journal of Public Health* 8 (2): 67. <https://doi.org/10.4103/2224-3151.264848>.
- Statistics South Africa. 2019. "General Household Survey, 2018." <http://www.statssa.gov.za/?p=12180>.



- Steele, Jessica E., Pål Roe Sundsøy, Carla Pezzulo, Victor A. Alegana, Tomas J. Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M. Iqbal, Khandakar N. Hadiuzzaman, Xin Lu, Erik Wetter, Andrew J. Tatem, and Linus Bengtsson. 2017. "Mapping Poverty Using Mobile Phone and Satellite Data." *Journal of The Royal Society Interface* 14 (127): 20160690. <https://doi.org/10.1098/rsif.2016.0690>.
- Stigler Center. 2019. "Stigler Center Committee on Digital Platforms, Final Report." Chicago: The University of Chicago Booth School of Business. <https://research.chicagobooth.edu/stigler/events/single-events/antitrust-competition-conference/digital-platforms-committee>.
- Sustainable Development Solutions Network (SDSN). 2015. "Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development." United Nations Sustainable Development Solutions Network, <http://unsdsn.org/wp-content/uploads/2015/04/Datafor-Development-Full-Report.pdf>.
- Tefficient. 2019. "Updated: Usage Up, But Monetisation Falter. 5G a Chance to Level Up." December 1, 2019. <https://tefficient.com/usage-up-but-monetisation-falters/>.
- Tiecke, Tobias G., Xianming Liu, Amy Zhang, Andreas Gros, Nan Li, Gregory Yetman, Talip Kilic, Siobhan Murray, Brian Blankespoor, Espen B. Prydz, Hai-Anh H. Dang. 2017. "Mapping the World Population One Building at a Time." *ArXiv:1712.05839*. <http://arxiv.org/abs/1712.05839>.
- United Nations. 2018. "The Sustainable Development Goals Report 2018 | Multimedia Library - United Nations Department of Economic and Social Affairs." New York, NY: United Nations. <https://www.un.org/development/desa/publications/the-sustainable-development-goals-report-2018.html>.
- UNCTAD (United Nations Conference on Trade and Development). 2019. "Digital Economy Report 2019 - Value Creating and Capture: Implications for Developing Economies." September 4, 2019. <https://unctad.org/en/pages/PublicationWebflyer.aspx?publicationid=2466>.
- USAID (U.S. Agency for International Development). 2019. "Artificial Intelligence in Global Health: Defining a Collective Path Forward." Center for Innovation and Impact, Innovating for Impact Series. Washington, DC: U.S. Agency for International Development. <https://www.usaid.gov/cii/ai-in-global-health>.
- Wardrop, Nicola. A., Warren C. Jochem, Tom J. Bird, Harrie R. Chamberlain, Dylan Clarke, Dan Kerr, Linus Bengtsson, Sabrina Juran, Vince Seaman, and Andrew J. Tatem. 2018. "Spatially Disaggregated Population Estimates in the Absence of National Population and Housing Census Data." *Proceedings of the National Academy of Sciences* 115 (14): 3529–37. <https://doi.org/10.1073/pnas.1715305115>.

- Wesolowski, Amy, et al., 2015. "Impact of human mobility on the emergence of dengue epidemics in Pakistan." *Proceedings of the National Academy of Sciences (USA)* 112: 11887-11892.
- World Bank. 2015. "Access to Electricity (% of Population) | Data." *World Development Indicators*. <https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS>.
- World Bank. 2016. "World Development Report 2016: Digital Dividends." Washington, DC: World Bank. <https://www.worldbank.org/en/publication/wdr2016>.
- World Bank. 2017. "World Development Report 2017: Governance and the Law." 112303. Washington, DC: The World Bank. <http://documents.worldbank.org/curated/en/774441485783404216/Main-report>.
- World Bank. 2018a. "*The State of Social Safety Nets 2018*." Washington, DC: World Bank.
- World Bank. 2018b. "*Information and Communications for Development 2018: Data-Driven Development*." Washington, DC: World Bank. <https://openknowledge.worldbank.org/handle/10986/30437>.
- World Bank. 2018c. "Data for Development: An Evaluation of World Bank Support for Data and Statistical Capacity." Washington, DC: Independent Evaluation Group, World Bank. <http://ieg.worldbankgroup.org/evaluations/data-for-development>.
- World Bank. 2019a. "World Development Report 2019: The Changing Nature of Work." Text/HTML. Washington, DC: World Bank Group. <https://www.worldbank.org/en/publication/wdr2019>.
- World Bank. 2019b. IDA19 Second Replenishment Meeting: Special Theme - Governance and Institutions. IDA19. Washington, DC: World Bank Group. <http://documents.worldbank.org/curated/en/696731563778743629/IDA19-Second-Replenishment-Meeting-Special-Theme-Governance-and-Institutions>
- World Bank. 2020a. "*Doing Business 2020: Comparing Business Regulation in 190 Economies*." Washington, DC: World Bank. <https://www.worldbank.org/en/publication/wdr2020>.
- World Bank. 2020b. "Data Innovation Project Success Stories from Around the World." January 22, 2020. <https://blogs.worldbank.org/opendata/data-innovation-project-success-stories-around-world>.
- The World Bank. 2020c. "Data on Statistical Capacity." Accessed February 3, 2020. <http://datatopics.worldbank.org/statisticalcapacity/>.
- World Economic Forum. 2013. "Technology Pioneers 2014." Geneva: World Economic Forum. <https://www.weforum.org/reports/technology-pioneers-2014/>.
- World Resource Institute. 2019. "Palm Oil Industry to Jointly Develop Radar Monitoring Technology to Detect Deforestation." World Resources Institute (blog). October 31, 2019.

<https://www.wri.org/news/2019/10/release-palm-oil-industry-jointly-develop-radar-monitoring-technology-detect>.

World Wide Web Foundation. 2020. "4thEdition | Open Data Barometer." Accessed February 3, 2020. [https://opendatabarometer.org/4thedition/?\\_year=2016&indicator=ODB](https://opendatabarometer.org/4thedition/?_year=2016&indicator=ODB).

Yang, Shihao, Samuel C. Kou, Fred Lu, John S. Brownstein, Nicholas Brooke, and Mauricio Santillana. 2017. "Advances in Using Internet Searches to Track Dengue." *PLoS Computational Biology* 13 (7).