A World Bank Group Flagship Report

WORLD DEVELOPMENT REPORT 2021

DATA FOR BETTER LIVES

WORLD BANK GROUP

A World Bank Group Flagship Report

WORLD DEVELOPMENT REPORT 2021

DATA FOR BETTER LIVES

WORLD BANK GROUP

This booklet contains the overview, as well as a list of contents and other front matter, from the *World Development Report 2021: Data for Better Lives*. doi: 10.1596/978-1-4648-1600-0. A PDF of the final book, once published, will be available at https://openknowledge.worldbank.org/ and http://documents.worldbank.org/, and print copies can be ordered at www.amazon.com. Please use the final version of the book for citation, reproduction, and adaptation purposes.

Cover design: Dania Kibbi, Base Three Studio

Cover illustration: Mijke Coebergh

Interior design: George Kokkinidis, Design Language, Brooklyn, New York, with updates and additional figure design by Dania Kibbi, Base Three Studio

Contents

Foreword

Data governance is the subject of intense debate in advanced economies and increasingly among large emerging markets. And yet many complex policy questions remain unanswered. In response, *World Development Report 2021: Data for Better Lives* surveys the emerging landscape and provides policy makers with a framework for thinking through the issues, opportunities, and trade-offs. One thing is clear: the perspective of lower-income countries has so far been largely absent from these global debates and urgently needs to be heard.

Data are a double-edged sword. On the one hand, they offer tremendous potential to create value by improving programs and policies, driving economies, and empowering citizens. On the other hand, data accumulation can lead to a concentration of economic and political power, raising the possibility that data may be misused in ways that harm citizens. Data are a resource that can be used and reused repeatedly to create more and more value, but there is a problem—the more data are reused, the higher is the risk of abuse.

It is hard to imagine a more dramatic example of these opportunities and tensions than the COVID-19 pandemic. Countries around the world have moved swiftly to repurpose mobile phone records to monitor the spread of the virus. But at the same time they have struggled to balance this benefit against privacy concerns and the risk of misuse.

Beyond pandemic times, the statistical capacity to produce and effectively use core economic and social data is limited. Many poor countries are unable to accurately track public finances, report on external debt, or monitor their development goals. Without such data, the ability to hold governments accountable and track progress withers.

Data governance arrangements to facilitate greater use of data while safeguarding against misuse remain in their infancy. The legal and regulatory frameworks for data are inadequate in lower-income countries, which all too often have gaps in critical safeguards as well as shortages of data-sharing measures. There, the data systems and infrastructure that enable interoperability and allow data to flow to more users are incomplete; less than 20 percent of low- and middle-income countries have modern data infrastructure such as colocation data centers and direct access to cloud computing facilities. Even where nascent data systems and governance frameworks exist, a lack of institutions with the requisite administrative capacity, decision-making autonomy, and financial resources holds back their effective implementation and enforcement.

To address these concerns, *World Development Report 2021* calls for a new social contract for data—one that enables the use and reuse of data to create economic and social *value,* promotes *equitable* opportunities to benefit from data, and fosters citizens' *trust* that they will not be harmed by misuse of the data they provide. However, in seeking such a social contract, lower-income countries are too often disadvantaged because they lack the infrastructure and skills to capture data and turn them into value; the scale and agency to participate equitably in global data markets and their governance; and the institutional and regulatory frameworks to create trust in data systems.

Forging a new social contract for data is a pressing domestic policy priority that will require strengthening national data systems and engaging all stakeholders at the national level. Because of the global scale of data, some of the most challenging aspects of the social contract also call for closer international cooperation to harmonize regulations and coordinate policies—bilaterally,

regionally, and globally. Critical areas for international engagement include reform of international taxation rights for data-driven businesses, World Trade Organization arrangements for trade in data-enabled services, regional collaboration on the development of data infrastructure, international harmonization of technical standards to support interoperability, and bilateral collaboration on law enforcement and antitrust regulation.

The World Bank stands ready to support its client countries on this important and challenging agenda. The findings of this *World Development Report* will shape support for client countries by identifying where public and private sector investments are the most critical, defining a rich program for policy reform and technical assistance, and highlighting areas in which global initiatives can help to convene and facilitate cross-border cooperation.

Realizing the full value of data will depend on a substantial commitment and effort, and it will be difficult. But the cost of failure is a world of missed opportunities and greater inequities.

David R. Malpass
President
The World Bank Group

Acknowledgments

This year's *World Development Report* (WDR) was prepared by a team led by directors Robert Cull, Vivien Foster, and Dean Jolliffe. Malarvizhi Veerappan served as report manager and as a member of the Report's leadership. The core team was composed of Jaffar Al-Rikabi, Adele Moukheibir Barzelay, Miriam Bruhn, Rong Chen, Niccolò Comini, Craig Hammer, Talip Kilic, Jan Loeprick, Daniel G. Mahler, Michael Minges, Martin Molinuevo, Nancy Morrison, David Newhouse, Sara Nyman, Vincent Francis Ricciardi III, David Satola, Dorothe Singer, Philip Wollburg, and Bilal Zia, together with research analysts Kenneth Zaul Moreno Sermeno, Nikkie Pacheco, and Elizabeth Salazar. Selome Missael Paulos provided the team with administrative support.

Davida Louise Connon, Hai-Anh H. Dang, Cem Dener, Samuel Paul Fraiberger, Lucas Kitzmüller, Aspasea Mckenna, and David Medine were members of the extended team. Rory Macmillan and Zia Mehrabi provided input as expert consultants. The Report was sponsored by the World Bank's Development Economics Vice Presidency and was prepared in close partnership with the Infrastructure Vice Presidency. Overall guidance was provided by Carmen M. Reinhart, Chief Economist, and Aart Kraay, Deputy Chief Economist. The team is especially grateful for the support and guidance provided by Makhtar Diop, who was at that time Vice President, Infrastructure, and Haishan Fu, Director, Data Group. The team is also appreciative of the comments and suggestions from Mari E. Pangestu, Managing Director, Development Policy and Partnerships. In the early months of the Report's preparation, Pinelopi Goldberg, then Chief Economist, provided guidance as well.

The team received suggestions from an advisory panel of high-level government officials, representatives of national statistical agencies and competition authorities, and private sector leaders: Ola Awad (President, Palestinian Central Bureau of Statistics), Dr. Long Chen (Director, Luohan Academy, representing Alibaba Group), Kersti Kaljulaid (President, Republic of Estonia), Francis W. Kariuki (Director-General, Competition Authority of Kenya), Zia Khan (Senior Vice President, Innovation, The Rockefeller Foundation), Ming Maa (President, Grab), Joyce Murray (Minister, Digital Government, Canada), Nandan Nilekani (Cofounder, Infosys, and architect of India's Data Empowerment architecture), Regina Opondo (Community Director, Kounkuey Design Initiative), Haleema Saeed (Director General, International Relations, Palestinian Central Bureau of Statistics), Kecuk Suhariyanto (Chief Statistician, BPS Statistics Indonesia), David Tudor (Group General Counsel, Prosus Group), Rodger Voorhies (President, Global Growth and Opportunity Division, Bill & Melinda Gates Foundation), Sheila Warren (Head, Blockchain and Data Policy, World Economic Forum), and Biratu Yigezu (Director, CSA Ethiopia). The team also received guidance from a technical board of leading academics: Emmanuelle Auriol (Research Faculty, University of Toulouse), Marshall Burke (Associate Professor, Stanford University), Luciano Floridi (Faculty, University of Oxford), Jan Kramer (Faculty and Chair, Internet and Telecommunications Business, University of Passau), Jacques Kremer (President, WorldWide Advisors, Inc.), Bruno Liebhaberg (Director General, Centre on Regulation in Europe), Fernando Perini (Senior Program Specialist, International Development Research Centre), John Pullinger (President, International Association for Official Statistics), Anna Scherbina (Associate Professor, Brandeis University), and Tommaso Valetti (Faculty, Imperial College London). The team valued their advice and found it very useful.

Finally, Shaida Badiee, Oliver Chinganya, Johannes Jutting, Stephen MacFeely, Angela Me, and Claire Melamed provided specialized reviews of chapter 9. However, the views expressed in the Report do not necessarily reflect those of these advisers and reviewers.

The team would also like to thank the following Bank staff, who were members of the "BrainsTrust," for their guidance during preparation of the Report: Luis Alberto Andres, João Pedro Azevedo, Andrea Barone, Kathleen Beegle, Tania Begazo, Jerome Bezzina, Calogero Carletto, Andrew L. Dabalen, Vyjayanti Desai, Marianne Fay, Michael Ferrantino, Marelize Gorgens, Mary Hallward-Driemeier, Kimberly D. Johns, Tim Kelly, Saki Kumagai, Daniel Lederman, William F. Maloney, Fredesvinda F. Montes Herraiz, James Neumann, Umar Serajuddin, Sharada Srinivasan, Davide Strusani, and Paolo Verme. Background and related research, along with dissemination, are being generously supported by the KDI School of Public Policy and Management (KDIS) Partnership Trust Fund, World Bank's Knowledge for Change Program (KCP, a multidonor trust fund), and the United States Agency for International Development.

The team drew on the analysis, research, and literature reviews of researchers and specialists from across the world. In addition, the team would like to thank the following for feedback and helpful discussions: Susan Ariel Aaronson, Ali Abbas, Karina Acosta, Sonia Ahmand, Sabina Alkire, Noam Angrist, Belinda Archibong, Kaushik Basu, Vitoria Beard, Theresa Beltramo, Willem Buiter, Tsenguunjav Byambasuren, Anupam Chander, Anindya Chatterjee, Nancy Chau, Katie Clancy, Julie Crowley, Shanta Devarajan, Gary Fields, Avi Goldfarb, Zihan Hu, Yingyi Jin, Ravi Kanbur, Carlos Lopez, Bentley MacLeod, Chelsea Madryga, Jose Antonio Ocampo, Kevin O'Neil, Ariel Ortiz-Bobea, Brian Park, Michael Pisa, Maria Poli, Cristian Pop-Eleches, Christian Reimsbach Kounatze, Elettra Ronchi, Benjamin David Roseth, Phet Sayo, Harman Sing Dhodi, Jenna Slotin, Michael Stanger, Joseph Stiglitz, Aberash Tariku, Jim Tebrake, Jeni Tennison, Anh Phuock Thien Nguyen, Kathryn Toure, Miguel Urquiola, Eric Verhoogen, Stefaan G. Verhulst, Sonali Vyas, and Yue Wang.

The following World Bank colleagues provided insightful comments, feedback, collaboration, and support: Rabah Arezki, Audrey Ariss, Angela Armstrong, Aziz Atamanov, Natalia Baal, Prospere R. Backiny-Yetna, Kathleen Beegle, Abdoullahi Beidou, Moussa Blimpo, Hana Brixi, Niklas Buehren, Greta L. Bull, Maurizio Bussolo, Cesar Calderon, Yu Cao, Greg Chen, Louise Cord, Paul Andres Corral Rodas, Jose Cuesta, Conrad Daly, Richard Damania, Olivia D'Aoust, Francesca Daverio, Zelalem Yilma Debebe, Asli Demirgüç-Kunt, Francesca de Nicola, Tami Ann Dokken, Mark Dutz, Alison Evans, Pablo Fajnzylber, Deon P. Filmer, Juliane Fries, Isis Gaddis, Emilia Galiano, Roberta V. Gatti, Tina George, Alejandro Medina Giopp, Chorching Goh, Lesly Goh, Markus Goldstein, Tania Gomez, Aparajita Goyal, Caren Grown, Boutheina Guermazi, Afef Haddad, Daniel Halim, Stephane Hallegatte, Nagaraja Harshadeep, Kristen Himelein, Johannes G. Hoogeveen, Chyi-Yun Huang, Elisabeth Huybens, Roumeen Islam, Ergys Islamaj, Ann-Sofie Jespersen, Anna Kalashyan, Woubet Kassa, Faruk Khan, Young Eun Kim, Soo Min Ko, Florence Kondylis, Ayhan Kose, Holly Krambeck, Megumi Kubota, Christoph Lakner, Somik Lall, Adriana Legovini, Norman Loayza, Nancy Lozano Gracia, Morgan Sofia Lucey, Cathrine Machingauta, Andrew D. Mason, Aaditya Mattoo, Anna Metz, Frederic Meunier, Sveta Milusheva, Miriam Muller, Mamta Murthi, James Neumann, Rochelle Glenene O'Hagan, Sandie Okoro, Madalina Papahagi, Utz Johann Pape, Elizaveta Perova, Tobias Pfutze, Martin Raiser, Laura Ralston, Martin Rama, Pooja Rao, Vijayendra Rao, Sheila Reszapi, Mark Roberts, Denis Robitaille, Carlos Rodriguez-Castelan, Dan Oliver Rogger, Clelia Rontoyanni, Rosemary Rop, Karla Viviana Saavedra Betancourt, Arisha Salman, Tara Sarma, Marc Tobias Schiffbauer, Ethel Sennhauser, Gabriel Sensenbrenner, Fatima Shah, Parmesh Shah, Nurlina Shaharuddin, Siddharth Sharma, Brian William Stacy, Clara Stinshoff, Prachi Tadsare, Daria Taglioni, Jomo Tariku, Jeff Thindwa, Hans Timmer, Margaret Triyana, Tea Trumbic, Vladimir Tsirkunov, Andrew Whitby, Mark Williams, Quentin Wodon, Keong Min Yoon, and Albert G. Zeufack.

Nancy Morrison provided developmental guidance in drafting the Report, which was edited by Sabra Ledent and proofread by Gwenda Larsen and Catherine Farley. Robert Zimmermann verified the Report's extensive citations. Dania Kibbi, Base Three Studio, was the principal graphic designer. Special thanks go to Stephen Pazdan, who coordinated and oversaw

production of the Report; Mark McClure, who provided assistance; and the World Bank's Formal Publishing Program. The team would also like to thank Mary C. Fisk, who facilitated multiple translations of the overview and main messages; Patricia Katayama, who oversaw the overall publication process; and Deb Barker and Yaneisy Martinez, who managed the printing and electronic conversions of the Report and its many ancillary products. Naureen Aziz, Mark Felsenthal, Chisako Fukuda, David Mariano, and Mikael Reventar developed the communications and engagement strategy. Monique Pelloux Patron provided the team with resource management support. The team would also like to thank Maria Alyanak, Marcelo Buitron, Gabriela Calderon Motta, Maria del Camino Hurtado, and Alejandra Ramon for their help with coordination.

Beyond Words studio designed and developed the WDR 2021 website (http://wdr2021.worldbank.org) and data stories. The team comprised Kate Ashton, Rebecca Conroy, Jamie Gilman, Chris Hankins, Ceri Jones, Becca Muir, Richard Pullinger, Duncan Swain, and Lucy Swan. The illustrations for the website were created by Elisabetta Calabritto. Maarten Lambrechts and Adrianus Willem Tulp developed the website's data stories. Emmanuel Cosmas Maliganya, Vidhya Nagarajan, Balaji Natarajan, Somu Sinthamani, Anushka Thewarapperuma, and Roula Yazigi contributed to website development and dissemination.

This Report draws on background papers prepared by Meaza Abraham, Rodrigo Barajas, Noelia Carreras, Anupam Chander, Sandeep Chandy, Yuan Fang, Martina Francesca Ferracane, Angelina Fisher (New York University), Dayoung Park, Sharada Srinivasan, Clara Stinshoff, Thomas Streinz (New York University), Erik van der Marel, Dereje Wolde, Isabel Yu, and the WDR team. Federico Cardenas Chacon, Nicolas Conserva, Lillyana Sophia Daza Jaller, Paris Gkartzonikasm, New Doe Kaledzi, Olga Kuzmina, Aliaksandra Tyhrytskaya, and Yasmin Zand provided support for the data collection and research required for the Global Data Regulation Survey prepared for this Report.

Contributors to the special-feature spotlights in the Report are Diana Jimena Arango, Anna-Maria Bogdanova, Martin Brocklehurst, Libby Hepburn, Stephen MacFeely, Angela Me, David Rogers, Sven Schade, Lea Shanley, Vladimir Tsirkunov, and Mariana Varese. The team would also like to thank the World Bank colleagues who helped organize and facilitate consultations and advised on translations: Jaffar Al-Rikabi, Adele Moukheibir Barzelay, Rong Chen, Candyce Da Cruz Rocha, Nicole El Hajj, Vivien Foster, and Jevgenijs Steinbuks.

The WDR team consulted policy makers, international organizations, civil society organizations, private sector companies, development partners, academics, research institutions, as well as the offices of the World Bank's Executive Directors and Bank colleagues. This year's *World Development Report* faced the additional challenge of the COVID-19 pandemic, which made it impossible to conduct the usual face-to-face consultation meetings. Nonetheless, the team was able to turn this into an advantage because virtual consultation events enabled much wider participation by hundreds of stakeholders. Due to travel restrictions, the team conducted consultations using a digital format and partner networks, thereby allowing more people with diverse backgrounds to participate. Some of the events focused on regional and targeted stakeholder consultations in multiple languages, allowing hard-to-reach organizations to join the discussion. Interactions with core contributing units to the *World Development Report* as part of ongoing debates in certain areas such as taxes leveraged wider consultations taking place on these related issues. Team members also drew heavily on their own experiences and interactions with the many data professionals, statisticians, data scientists, civil society organizations, and policy makers working in often difficult conditions to use data to achieve development outcomes.

The consultations were launched in October 2019 with a high-level roundtable—a side event of the World Bank's annual meetings—and continued until October 2020. The team also conducted a series of bilateral consultations from April through June 2020 with the national statistical offices of Canada, Ethiopia, India, Indonesia, Mexico, the United Kingdom, and West Bank and Gaza. The team thanks the organizers of the many events that brought together a wide variety of stakeholders, including government officials, civil society organizations, academic institutions, the private sector, and international organizations. The organizers were Germany's Federal Ministry for Economic Cooperation and Development

(BMZ) and Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ); Canada's International Development Research Centre (IDRC), which organized consultations across Africa, Asia, and Latin America and the Caribbean in French, Spanish, and English; the Global Partnership for Sustainable Development Data (GPSDD); the GW Elliott School of International Affairs; Columbia and Cornell Universities; the Mastercard Advocacy Center of Excellence; A. T. Kearney's Global Business Policy Council; the Fletcher School of Law and Diplomacy's Institute for Business in the Global Context at Tufts University; Committee for the Coordination of Statistical Activities (CCSA); United Nations World Data Forum (UNWDF); Organisation for Economic Co-operation and Development's Development Assistance Committee (DAC); Global Data Barometer; and the Open Gov Hub.

The team organized and held a dedicated consultation event for civil society members and several bilateral consultations with technology firms, platform-based businesses, internet exchange points, payment industries, and cybersecurity firms in the private sector. They included Alibaba, Amazon, CADE (Brazilian competition authority), De Novo, Facebook, Google, LinkedIn, Lori Systems (Kenya), Mastercard, Power2SME (India), and UA-IX (Ukranian Internet Exchange Point). The team also reached out for guidance on specific topics from experts based in institutions that included Aelex Legal (law firm, Abuja, Nigeria), Atlantic Council (think tank), DataPrivacy.Com.Br (São Paulo, Brazil), Georgetown University, Hamu and Company (law firm, Lagos, Nigeria), Hogan Lovells (privacy and cybersecurity law firm), Hunton Andrews Kurth (global privacy and cybersecurity law firm), Interswitch (Lagos, Nigeria), New York University, Organisation for Economic Co-operation and Development, Oxford Internet Institute, Rockefeller Foundation, the United Nations Conference on Trade and Development (UNCTAD), the University of Southern California, and WilmerHale (data and cybersecurity law firm). Detailed information about all consultations and contributors can be found at https://www.worldbank.org/en/publication/wdr2021/consultations.

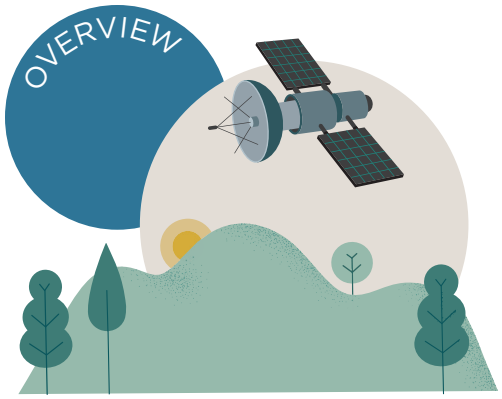
The team is grateful as well to the many World Bank colleagues who provided written comments during the formal Bank-wide review process. Those comments proved to be invaluable at a crucial stage in the Report's production.

The team apologizes to any individuals or organizations inadvertently omitted from this list and expresses its gratitude to all who contributed to this Report, including those whose names may not appear here. The team members would also like to thank their families for their support throughout the preparation of this Report.

OVERVIEW

You can have data without information, but you cannot have information without data.

—Daniel Keys Moran, computer programmer and science fiction author



Data, which are growing at an unprecedented rate, are becoming an integral part of the daily lives of most people everywhere. But how does that matter for the more than 700 million people living in extreme poverty? Is the explosion in the new types and uses of data improving their lives? Or will poor people and poor countries be left behind, creating a widening gap between those who reap the benefits of this new data-driven world and those who do not?

The innovations resulting from the creative new uses of data could prove to be one of the most life-changing events of this era for everyone. Like many general-purpose technologies such as the steam engine and electricity, the transformations emerging from the data revolution could touch all aspects of societies and economies. But such sweeping changes are not automatic. The productivity value of the steam engine and electricity was realized decades after they were first introduced. The delay occurred not because people did not recognize the importance of these innovations—sooner or later everyone did—but because the new manufacturing systems needed for these innovations to realize their economic potential could not take shape overnight. Just as electricity itself did not result in economic development, data alone will not improve well-being. Data can improve social and economic outcomes, but only if they are used systematically in ways that create information that generates insights that improve lives.

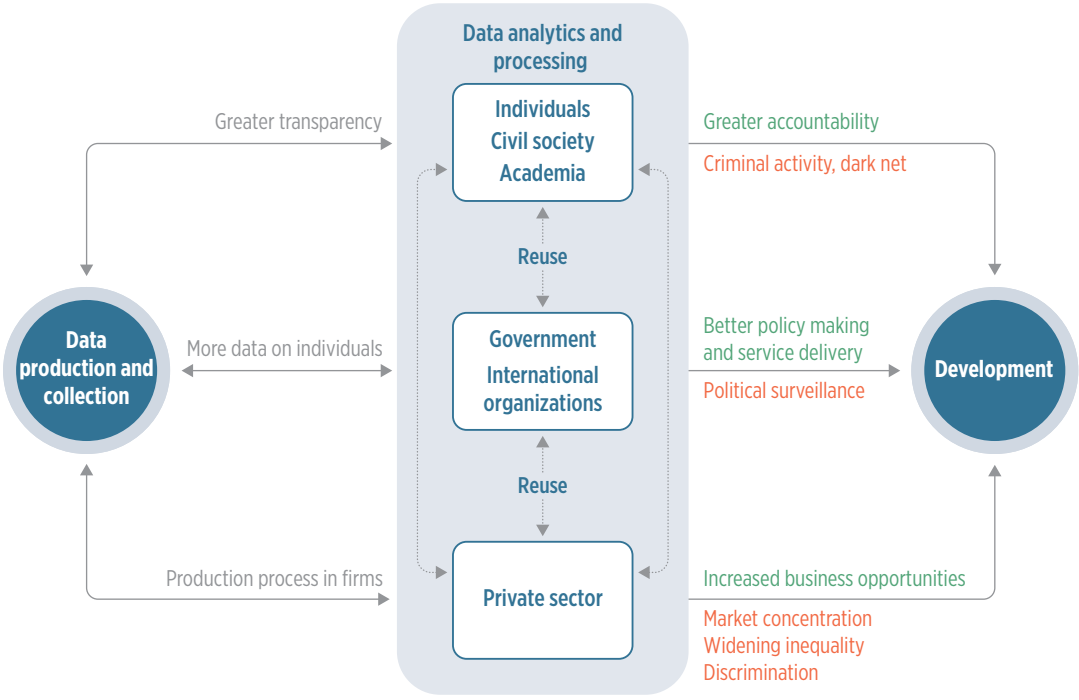
This Report aims to answer two fundamental questions. First, how can data better advance development objectives? Second, what kind of data governance arrangements are needed to support the generation and use of data in a safe, ethical, and secure way while also delivering value equitably?

One important message of this Report is that simply gathering more data is not the answer. Significant data shortfalls, particularly in poor countries, do exist, but the aim of this Report is to shift the focus toward *using data more effectively to improve development outcomes*, particularly for poor people in poor countries.

Advancing development objectives through data

Part I of this Report develops a conceptual framework that links data to development through three institutional pathways (figure O.1). The middle pathway is the use of data by governments and international organizations to support evidence-based policy making and improved service delivery. The top pathway is the use of data by civil society to monitor the effects of government policies and by individuals to enable them to monitor and access public and commercial services. The bottom pathway is the use of data by private firms in the production process—use that fuels their own growth as well as wider economic growth. One implication of the conceptual framework is that data alone cannot solve development problems: people (in society, governments, and firms) are the central actors transforming data into useful information that can improve livelihoods and lives.[1] Alongside capital, land, and labor, data are also an input to the development objectives that emerge along all three pathways. But, unlike capital, land, and labor, using data once does not diminish its value. Data that were initially collected with one intention can be reused for a completely different purpose (chapter 1).

Figure O.1 How data can support development: A theory of change



Source: WDR 2021 team.

Note: Positive impacts are shown in green; negative impacts are shown in red.

Disseminating, exchanging, and sharing data to enhance data reuse and repurposing

Because the potential of data to serve a productive use is essentially limitless, enabling the reuse and repurposing of data is critical if data are to lead to better lives. It is thus a central aspect of the conceptual framework. Figure O.1 uses two-way arrows to depict these flows. The two-way arrow between the private sector and government/international organizations indicates the reuse and repurposing of data originally collected for commercial purposes for public policy, and vice versa. Similarly, the two-way arrow between individuals/civil society/academia and government/international organizations indicates data being exchanged and reused by those parties. The final two-way arrows reflect the use of private sector data and data-driven applications by individuals/civil society/academia and the use of data and analysis generated by individuals/civil society/academia by firms. In practice, however, those holding data may be unwilling to exchange data. They may have concerns about data protection and security or the need to capture returns on investments in collecting data. Or they may hope to gain market power from accumulating data to capture economies of scale or obtain any other kind of political or competitive advantage from hoarding them.

The phrase "sharing and reuse" is shorthand used in this Report for all the types of transactions and exchanges of data that permit reuse, from government open data initiatives for sharing data to market-based transactions for data involving private firms. In theory, defining clear economic property rights over data should enable data to be traded widely on markets. But in practice, the extent of the data trade (beyond the market for advertising) has been limited by competing claims on ownership, tensions between the wide dissemination of data and incentives to accumulate more data for private commercial gain, and difficulties in assessing the quality and accuracy of data.

Each of the three pathways illustrated in figure O.1 shows how data can improve lives, but those same pathways create openings for data to be used in ways that harm people. Through the government pathway, data can be abused for political ends, such as politically motivated surveillance or discrimination along lines of ethnicity, religion, race, gender, disability

status, or sexual orientation. In the pathway running through individuals, there is the potential for cybercriminals to inflict considerable harm by stealing and manipulating sensitive information. The "dark net" is a vast parallel network of hidden websites that provides an underground digital platform for a wide array of criminal activities, facilitating illegal trade in drugs, counterfeit currency, stolen goods, credit card numbers, forged papers, firearms, and human organs. Similarly, through the private sector pathway, examples of harmful use include, among other things, the exploitation of information about consumer preferences and behavior to engage in aggressive or manipulative marketing techniques based on microtargeting of persuasive messages or to apply algorithms that facilitate collusion among market players.[2]

Unlocking data for the public good and safeguarding against misuses: Some COVID-19 examples

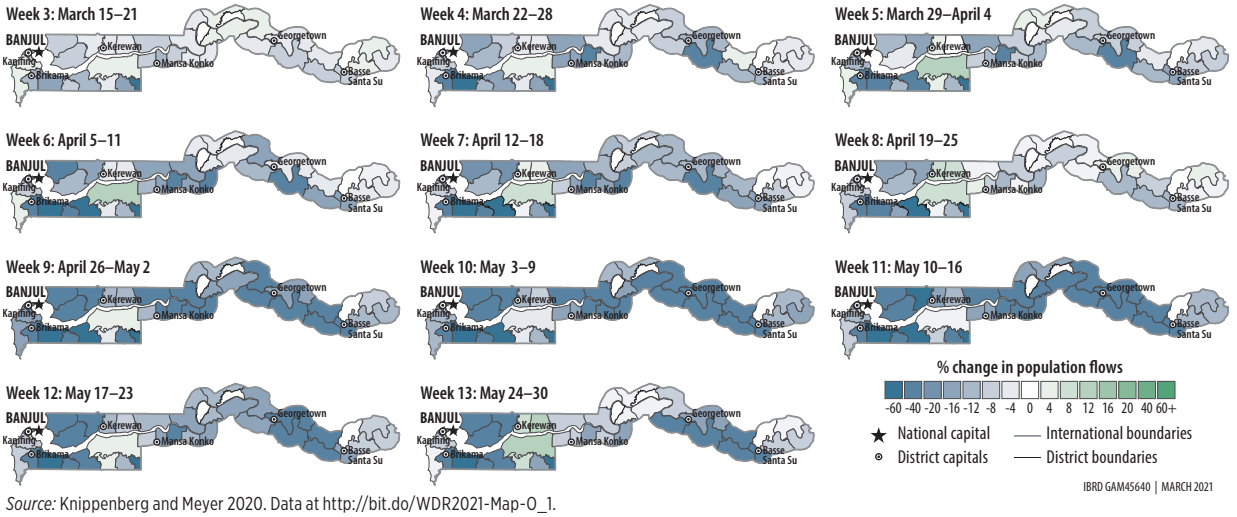
Many countries have used data to control the COVID-19 pandemic. This use includes tracking people's locations to better understand mobility patterns during lockdowns or to aid in disease contact tracing. Using call detail records (CDRs) from March through May 2020 aggregated to mask individual-level data, policy makers in The Gambia were able to review maps showing the movement of people across administrative boundaries (map O.1). These maps helped them understand the extent to which lockdowns were succeeding in reducing movement and allowed them to identify the factors linked to lockdown compliance and noncompliance and plan accordingly. Meanwhile, the government of Israel approved emergency regulations in March 2020 to allow the individual-level data collected from cellphones to be used to track people and then, through contact tracing, to curtail the spread of COVID-19.

CDRs were not created to aid public policy making or to allow the government to track the movements of individuals, but they are an example of data being reused and repurposed (flowing in the vertical channels in figure O.1). In Israel, these data were being collected before the pandemic, but they could be accessed only for national security purposes.[3]

These early efforts at repurposing CDRs to track infected individuals seemed to have a positive effect. In The Gambia, the maps helped reveal that the lockdown disproportionally affected poorer districts, indicating a need for relief and recovery efforts to target these areas. In Israel, analysis of the cellular data suggested their use led to identification of more than one-third of all of the country's coronavirus cases in the early weeks of the pandemic (more than 5,500 of the 16,200 people who had contracted the disease), possibly contributing to Israel's exceptionally low initial rates of coronavirus infections and deaths.

This new use of CDR data to track large parts of the population of Israel sparked debate and pushback over concerns about the potential misuse of the data by government. In Israel, many lawmakers raised privacy concerns, and the Supreme Court eventually

Map O.1 Use of aggregated cellphone records to track mobility week by week during COVID-19 lockdowns in The Gambia, March–May 2020



Source: Knippenberg and Meyer 2020. Data at http://bit.do/WDR2021-Map-O_1.

Note: Blue shades indicate outflow of people; green shades indicate inflow of people. A nationwide lockdown was imposed on March 22, 2020. Data were gathered using call detail records.

halted the program. The Court ruled in late April 2020 that the government must legislate the use of cellphone tracking and that "a suitable alternative, compatible with the principles of privacy, must be found."[4]

Many of the themes of this Report are illustrated in this example. The sharing and reuse of private sector CDR data with public authorities created social *value* by supporting the control of COVID-19 infections, thereby saving lives. At the same time, this transfer of data raised fundamental concerns about *trust*, with citizens concerned that their CDR data could then be repurposed by government officials for other unintended and potentially harmful purposes beyond public health. Issues of *equity* were also at stake. Whereas in a high-income country like Israel smartphone penetration was 93 percent, in a low-income country like The Gambia smartphone penetration was only 75 percent. In each case, that minority of the population lacking a smartphone was unable to generate CDR data and would not necessarily benefit directly from the public health protection afforded by contact tracing.

These examples also illustrate a key conundrum. The potential benefits that people realize in the form of improved policies and service delivery may increase rapidly as more data, especially personal data, are shared and reused—but the risks of data being misused increase as well. These potential benefits depend on data being disseminated or exchanged between parties. But parties must *trust* the systems, regulations, and institutions that underlie the security of such exchanges to willingly engage in them.

How can people trust that their data will be protected and that they will share in the value that data can produce? The mounting nature of such concerns suggests the need for a new social contract around data—that is, an agreement among all participants in the process of creating, reusing, and sharing data that fosters trust that they will not be harmed from exchanging data and that part of the value created by data will accrue equitably (figure O.2). The idea that societies engage in these sort of agreements, or social contracts, has existed for centuries, often linked to the writing of philosophers such as Thomas Hobbes, John Locke, and Jean-Jacques Rousseau.

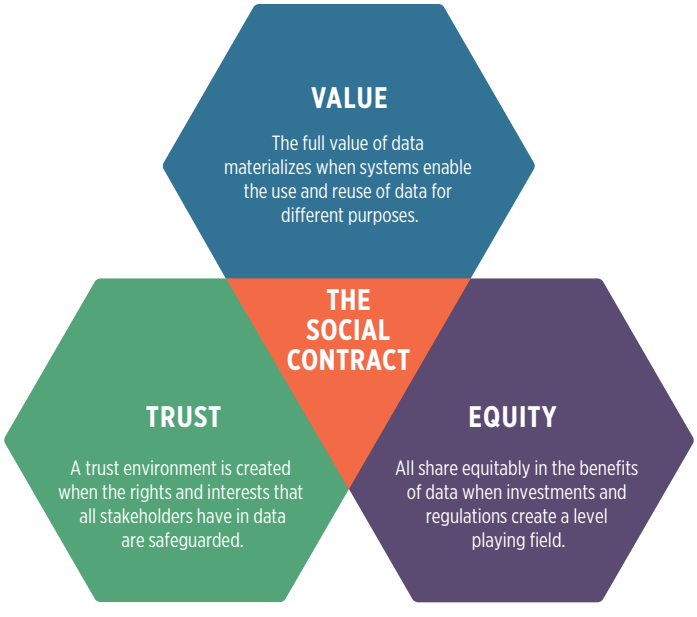
Legal systems, and governance more generally, can be viewed as instruments for establishing, facilitating, and enforcing social contracts. Persuading parties to abide by the rules of a social contract is not an easy task and will hinge on ensuring that the benefits from using data are shared in an equitable way—that is, everyone has something to gain. In this process, lower-income countries are too often disadvantaged, lacking, as they often do, the infrastructure and skills to capture data and turn them into value; the institutional and regulatory frameworks to create trust in data systems; and the scale and agency to participate equitably in global data markets and their governance.

With data reshaping our lives, our societies, and the world more generally, social contracts for data are needed both nationally and internationally, especially because of the cross-border nature of data transactions and flows. Spotlight 8.1 extends this idea of a social contract to the international realm, calling for a global consensus to ensure that data are safeguarded as a global public good and as a resource to achieve equitable and sustainable development.

The untapped potential of data; the evolving legal, regulatory, and governance frameworks for data generation, use, and reuse; the importance of country context (history, culture, governance, and political economy) in shaping appropriate frameworks; the role of technical capabilities for making the most of data safely; and the need for trust and more equitable sharing of the value of data—all these are the themes at the core of this *World Development Report*.

Part I of the Report begins by describing in more detail the potential development impact of data collected for public purposes—*public intent data* (chapter 2); data collected by the private sector as part of routine business processes—*private intent data* (chapter 3); and the synergies that arise from the joint use of different types of data (chapter 4). This distinction between public intent and private intent data is used

Figure O.2 A social contract for data founded on value, trust, and equity



Source: WDR 2021 team.

regardless of who collected the data or the methods used to gather the data (such as customer surveys, accounting records, or digital transactions).

Public intent data can improve service delivery, targeting, accountability, and empowerment

Public intent data hold great potential for designing, executing, and evaluating public programs and policy (chapter 2). Because public intent data are a prerequisite for many government functions, government agencies are the primary producers of these data by means of censuses, administrative data collection, and more. Citizens, civil society organizations, nongovernmental organizations, academic institutions, and international organizations contribute critically to the production of public intent data using surveys, crowdsourcing platforms, and other means.

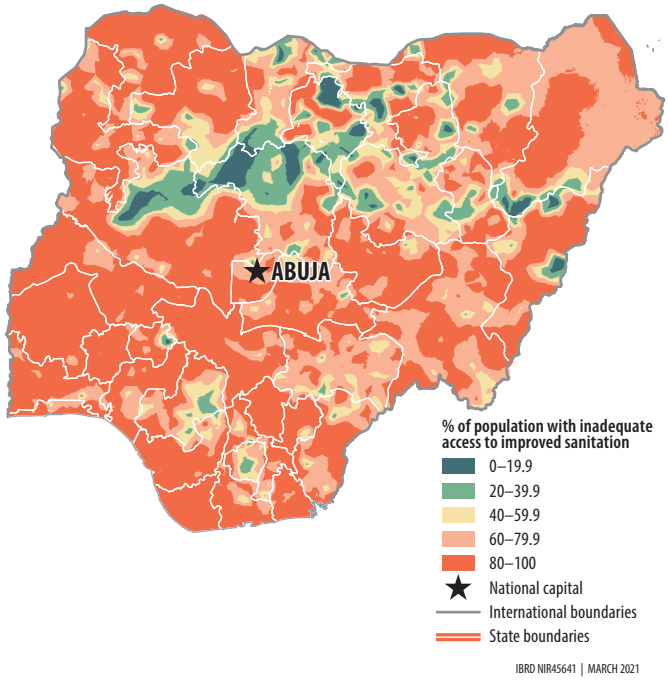
These kinds of data can lead to better lives through three main pathways: first, by improving policy making and service delivery; second, by prioritizing scarce resources and targeting them to reach marginalized populations and areas; and third, by holding government accountable and empowering individuals to make better choices through more information and knowledge.

An example from Nigeria illustrates the power of public intent data to improve and target service delivery. The 2015 National Water Supply and Sanitation Survey commissioned by Nigeria's government gathered data from households, water points, water schemes, and public facilities, including schools and health facilities. These data revealed that 130 million Nigerians (or more than two-thirds of the population at that time) did not meet the standard for sanitation set out by the Millennium Development Goals and that inadequate access to clean water was especially an issue for poor households and in certain geographical areas (map O.2).[5] In response to the findings from the report based on these data, President Muhammadu Buhari declared a state of emergency in the sector and launched the National Action Plan for the Revitalization of Nigeria's Water, Sanitation and Hygiene (WASH) Sector.[6]

The higher the quality of the data (in terms of features such as timeliness, accuracy, and resolution), the greater is their potential to generate value for development. Yet a variety of factors prevent countries—particularly low-income ones—from realizing greater value from data for the public good. These impediments include lack of resources, technical capacity, data governance, and demand for data-informed decision-making. The World Bank's Statistical Performance Indicators, released as part of this Report, identify gaps in the availability, quality, and usability of public intent data across 166 countries, focusing on features related to the timeliness, granularity, interoperability, and accessibility of those data.[7]

Unleashing the full potential of public intent data requires high-level prioritization of data in the policy process. Governments would then prioritize the production of high-quality data and the open and transparent use of data for decision-making. Transparency and reliability of official statistics can help build trust in government actions. A lack of transparency, such as not revealing a country's debt burden, can have harmful economic consequences and damage the public's trust in government (see spotlight 1.2). Fulfilling the potential of data requires long-term, stable financing of data; investments in statistical and technical capacity; and laws conducive to safe data production and reuse. Other areas that must be addressed include low levels of data literacy affecting the demand for data, policy makers' lack of incentives for and interest in using data, low trust in the quality of public intent data, and lack of infrastructure for accessing and using the data. These investments and initiatives rely on one another, and so failure to succeed in one area jeopardizes the overall value that data can bring to

Map O.2 Highly refined data pinpointed areas of Nigeria that needed better sanitation



Source: World Bank 2017. Data at http://bit.do/WDR2021-Map-O_2.

Note: Geographic hotspots of inadequate access to improved sanitation are shown from the least severe (■) to the most severe (■) in terms of the percentage of the population in that area that meets an international benchmark for sanitation.

development. Effective use of data can generate more demand for data, thereby justifying investments to produce more, and higher-quality, data.

Private intent data can fuel growth and boost development

Data collected and curated by the private sector for commercial purposes also hold great potential to spur development (chapter 3). Innovations in the use and application of data by businesses are creating tremendous economic value by enhancing data-driven decision-making and reducing transaction costs. A 2011 study of 179 large firms in the United States indicated that firms adopting data-driven decision-making increased their productivity by 5–6 percent relative to what would be expected in view of their other investments and use of information technology.[8]

Although data are in many ways an input to the production process of firms, much of the recent explosion of new data has come about as a by-product of economic activity, such as digitization of firm operations, mobile phone usage by individuals, digital transactions, and social media interactions. These data are collected at high frequency and can provide detailed information on individuals, businesses, economic outcomes, and phenomena. They not only enhance the economic efficiency of the firms themselves, but also offer potential to be repurposed for public policy needs such as COVID-19 tracking. For example, financial services providers are increasingly adopting alternative credit scoring techniques to solve the long-standing issue of lack of data on potential borrowers (or more specifically, asymmetric information) in banking. These techniques take advantage of users' digital footprints to assess creditworthiness for those who otherwise lack documentation. Two prominent examples of this approach are Lenddo, which operates in the Philippines, and Cignifi, which operates in Africa, Asia, and Latin America.

But these trends also come with new risks that must be addressed to ensure that the data-driven economy raises social welfare. Concerns are growing about excessive data collection, insufficient governance of data held by private firms, and inadequate protection of personal data. Many of these concerns revolve around the misuse of personal data. Such misuses include the failure of firms to properly protect the financial information of clients—exposing them to theft of funds or identity—or firms' engagement in unauthorized use of, or failure to protect, individuals' confidential health or location data.

Many of the processes through which firms create value with their data are driven by algorithms and machine learning. In these models, algorithms determine, among other things, what information, products, or services individuals are exposed to and at what price; what insurance packages they are offered; whether their loan applications are approved; what jobs they qualify for; and what medical advice they receive.

All these types of activities have the potential to significantly improve economic efficiency. For example, by consuming more data types and extracting relevant information from seemingly unrelated patterns, machine learning could generate credit scores for more individuals with greater precision. However, if the data fed into the machine learning embed discriminatory assumptions, machine learning will amplify that discrimination, not only producing harmful results, but also magnifying them.[9] This point brings to mind the decades-old data science adage "garbage in–garbage out," meaning that a data processing system such as machine learning is no better than the data it is given to process.[10] But there is a deeper concern: the output from machine learning is typically opaque and changes frequently as new data enter the system. Almost by design, it creates a rule that is not transparent, and so identifying discriminatory elements of the algorithm can be technically very challenging.

Often, data-driven markets exhibit positive network externalities, leading to increasing returns to scale and a propensity for a few large firms to dominate. The result can be the exclusion of smaller or more traditional firms to the detriment of local entrepreneurship, with possible risks for consumer welfare. These effects may be exacerbated in developing markets, where entrants find it harder to raise start-up capital and where there is limited human capital in data sciences. To counteract this, policy makers can address the underlying constraints to achieving scale, such as geoblocking (restricting access to internet content based on the user's geographical location) or lack of harmonization of data policies across countries. They can ensure that sector regulations and government support schemes provide a level playing field for all firms.

Combining and repurposing data can deepen their development impact

Combining and repurposing different types of data can enhance the impacts of data on development (chapter 4). Development problems are complex, spanning economic, cultural, environmental, demographic, and many other factors. Policy design based on data covering only one factor will be incomplete,

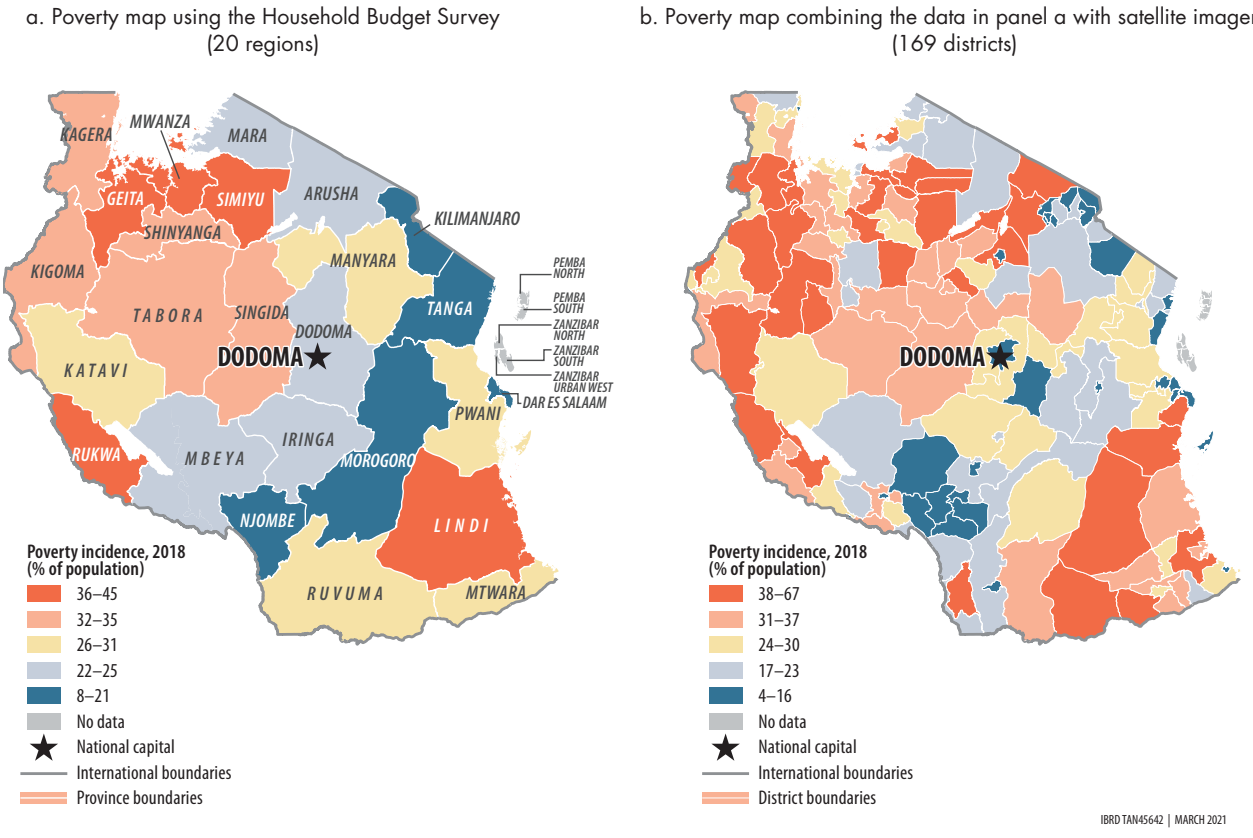
and sometimes ill-advised. Combining different types of data can fill data gaps and offer new perspectives on development problems.

As one example, public intent household surveys, which gather extensive data on living standards, consumption, income, and expenditures, are the basis for estimating national poverty rates in most countries. Because the survey instrument is so extensive and time-consuming to administer, the samples tend to be relatively small. Estimates of poverty are usually statistically valid for a nation and at some slightly finer level of geographic stratification, but rarely are such household surveys designed to provide the refined profiles of poverty that would allow policies to mitigate poverty to target the village level or lower. Meanwhile, for decades high-resolution poverty maps have been produced by estimating a model of poverty from survey data and then mapping this model onto census data, allowing an estimate of poverty for every household in the census data. A problem with this approach is that census data are available only once a decade (and in many poorer countries even less frequently).

Modifications of this approach have replaced population census data with CDR data or various types of remote sensing data (typically from satellites, but also from drones). This repurposing of CDR or satellite data can provide greater resolution and timelier maps of poverty. For example, using only household survey data the government of Tanzania was able to profile the level of poverty across only 20 regions of the country's mainland. Once the household survey data were combined with satellite imagery data, it became possible to estimate poverty for each of the country's 169 districts (map O.3). Combining the two data sources increased the resolution of the poverty picture by eightfold with essentially no loss of precision. Other examples of this innovative analysis are occurring in some of the world's most data-deficient environments such as Afghanistan and Rwanda, offering solutions to pressing data gaps.[11]

Examples of other ways of repurposing data include using online media and user-generated content to map water/flood events in real time for water management and food security and combining

Map O.3 Combining satellite imagery with household survey data increases the resolution of the poverty map of Tanzania



Source: World Bank 2019. Data at http://bit.do/WDR2021-Map-O_3.

satellite imagery data from public and private sources to monitor crop yields and forecast malnutrition.[12] Similarly, many examples in this Report highlight the potential for repurposing data to improve programs, policies, and outcomes in areas such as monitoring public health (including the spread of disease), managing crisis response and resource allocation, ensuring road safety in transport and transit, and monitoring illegal fishing and deforestation.

Novel ways to create and use data enable civil society to hold governments accountable for policies and to better monitor corruption.[13] For example, utilizing crowdsourced data and web scraping (extracting data from websites), social media discussion boards are emerging as ways in which local leaders can act against corrupt officials and receive real-time feedback on the impact of anticorruption policies. The "I paid a bribe" online initiative launched in 2011 by the Janaagraha Centre for Citizenship and Democracy in India has developed into one of the largest crowdsourced anticorruption platforms in the world. This tool collects citizens' reports of corrupt behavior and merges them with geospatial data to highlight problem areas. In doing so, it empowers individuals, civil society, and governments to fight corrupt behavior.

To encourage more efforts to repurpose and combine data sources, this Report describes ways in which donors, governments, and companies could invest in the people, partnerships, and research needed to leverage these new data sources for public benefit. Low-income countries should emphasize policy initiatives and investments in building the data skills of analysts and decision-makers; expanding tertiary education to encompass data science and analytics; promoting partnerships with universities and private companies in higher-income countries; strengthening the data literacy of senior government leadership; creating institutional environments that encourage the use of sophisticated data and evidence in policy making; and revamping national statistical offices to perform nontraditional roles with private intent data.

Aligning data governance with the social contract

A well-designed data governance framework allows countries to capture the full economic and social value of both public intent and private intent data and leverages synergies between them. This involves creating trust in the integrity of the data system, while ensuring that the benefits of data are equitably shared. Such a framework is the tangible expression of a country's social contract around data.

Part II of this Report describes these building blocks of data governance, which can deliver the potential benefits of data while safeguarding against harmful outcomes (figure O.3). These building blocks include data infrastructure policies (chapter 5); policies, laws, and regulations around data (chapter 6); related economic policies (chapter 7); and data governance institutions (chapter 8).

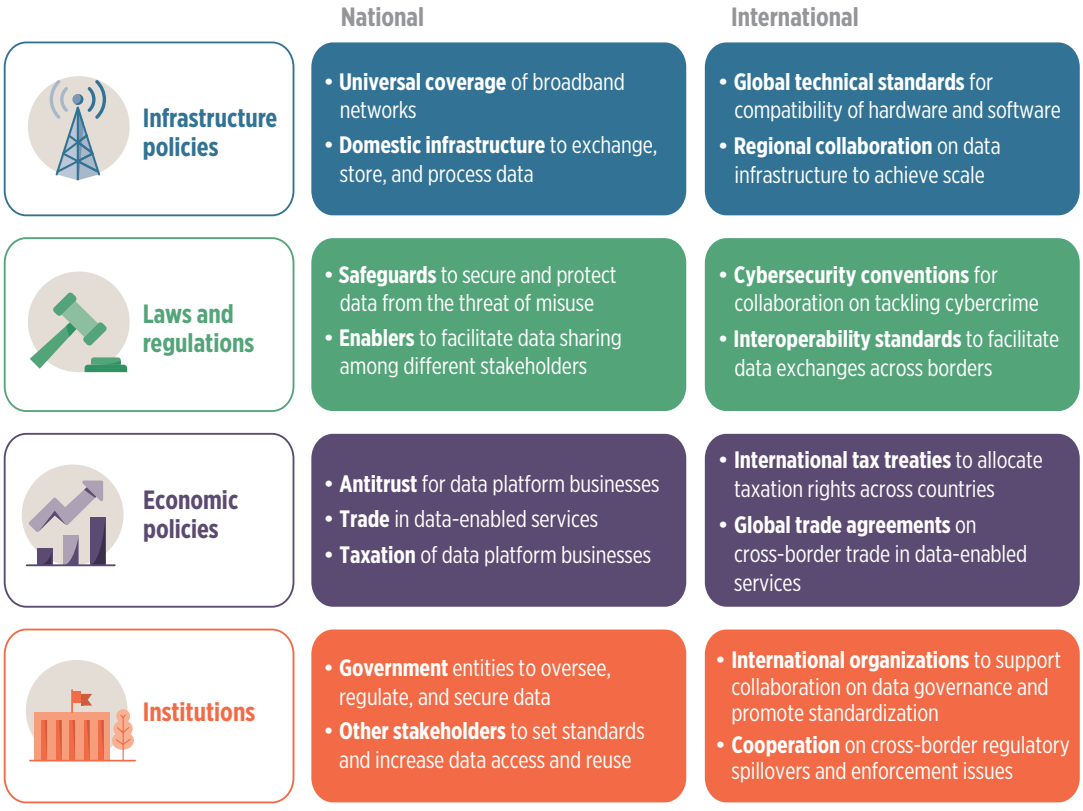
Although much of data governance is domestic in focus, an efficient and equitable resolution of many data governance challenges is possible only with international collaboration. Bilateral efforts are needed to manage cross-border spillovers of antitrust decisions and to join forces to combat cybercrime. Multilateral cooperation is essential to address global free-rider problems (such as data protectionism or tax evasion in data-enabled services) and to reduce transaction costs through harmonization of legal and technical standards for data protection and interoperability. At the same time, regional collaboration can help amplify the voice of low- and middle-income countries in global data governance negotiations and help realize scale economies in the development of data infrastructure.

Improving data infrastructure helps ensure equitable access for poor people in poor countries

The digital character of modern data calls for digital infrastructure—a prerequisite for collecting, exchanging, storing, processing, and distributing data (chapter 5). Yet the availability of such infrastructure is marked by inequity both within and between countries. Because the social and economic value of data infrastructure rises steeply as more and more citizens are connected, universal service policies have long existed to promote service rollout. In recognition of the transformative opportunities that broadband connectivity presents for both individuals and nations, the United Nations Broadband Commission has committed the international community to reaching 75 percent broadband-internet user penetration by 2025.[14]

That said, efforts to move toward universal access face fundamental challenges. First, because of the continual technological innovation in mobile technology service, coverage is a moving target. Whereas in 2018, 92 percent of the world's population lived within range of a 3G signal (offering speeds of 40 megabytes per second), that share dropped to 80 percent for 4G technology (providing faster speeds of 400 megabytes per second, which are needed for more sophisticated smartphone applications that can promote development). The recent commercial launch of 5G technology (reaching speeds of 1,000

Figure O.3 Data governance layers at the national and international levels



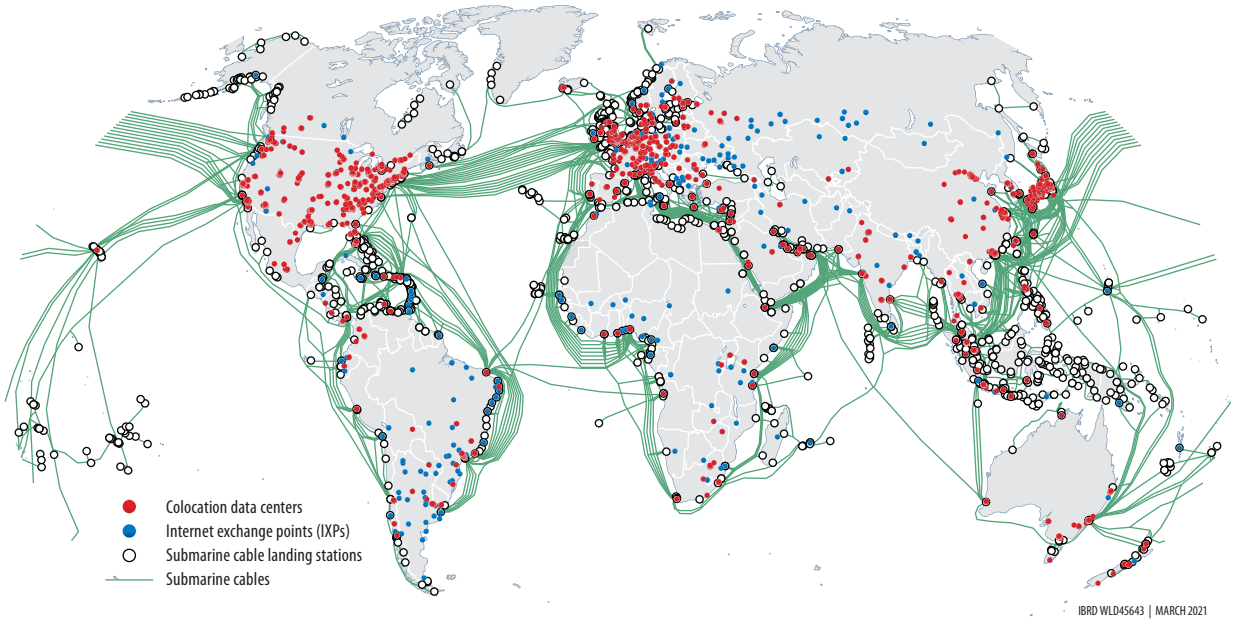
Source: WDR 2021 team.

megabytes per second) in a handful of leading-edge markets risks leaving the low-income countries even further behind. Policy makers can hasten technological upgrades by creating a supportive environment for private sector investment in the underpinning fiber-optic networks, while introducing more effective management of critical spectrum resources. Sharing infrastructure can also greatly reduce the cost of upgrades. Yet a careful balance must be struck between promoting competition in broadband provision wherever possible and encouraging cooperation between service providers in market segments where demand is too limited to support more than one infrastructure network.

The second challenge is that a substantial majority of the 40 percent of the world's population who do not use data services live within range of a broadband signal. Of people living in low- and middle-income countries who do not access the internet, more than two-thirds stated in a survey that they do not know what the internet is or how to use it, indicating that digital literacy is a major issue.[15] Affordability is also a factor in low- and middle-income countries, where the cost of an entry-level smartphone represents about 80 percent of monthly income of the bottom 20 percent of households.[16] Relatively high taxes and duties further contribute to this expense.[17] As costs come down in response to innovation, competitive pressures, and sound government policy, uptake in use of the internet will likely increase. Yet even among those who do use the internet, consumption of data services stands at just 0.2 gigabytes per capita per month, a fraction of what this Report estimates may be needed to perform basic social and economic functions online.

A third challenge in expanding connectivity is its potential impact on global warming. The climate impacts of increased connectivity present a set of complicated trade-offs. In 2018 the electricity needed to support data infrastructure was equal to approximately 1 percent of global consumption—a significant draw with environmental consequences. But because of reliance on renewable energy–supported data infrastructure and increasing energy efficiencies, greenhouse gas emissions linked to data infrastructure are disproportionately lower than for other sectors. Furthermore, access to data infrastructure can have significant positive climatic effects as illustrated by the massive reduction in travel and increase in videoconferencing during COVID-19 (spotlight 5.2).

Map O.4 Data infrastructure is not yet widespread across all parts of the world



Sources: PeeringDB, Interconnection Database, https://www.peeringdb.com/; PCH Packet Clearing House, Packet Clearing House Report on Internet Exchange Point Locations (database), accessed December 14, 2020, https://www.pch.net/ixp/summary; TeleGeography, Submarine Cables (database), https://www.submarinecablemap.com/. Data at http://bit.do/WDR2021-Map-O_4.

Full participation in the data-driven economy entails not only connecting individual citizens but also developing adequate data infrastructure at the national level. For the most part, low- and middle-income countries lack domestic facilities to allow their own locally generated data to be exchanged (via internet exchange points, IXPs), stored (at colocation data centers), and processed (on cloud platforms)—see map O.4. Instead, many continue to depend on overseas facilities, requiring them to transfer large volumes of data in and out of the country—for which they pay a substantial penalty in terms of slower speed and higher prices.

Policy makers can do much to improve access to data infrastructure progressively. This process begins by encouraging the creation of domestic IXPs and then fostering a suitable investment climate for colocation data centers. In these centers, popular internet content can be stored locally, and access to overseas cloud infrastructure can be facilitated through the provision of on-ramps. Such facilities can be shared at the regional level, where suitable fiber-optic connectivity exists between countries and there is adequate regulatory harmonization. Because of the extremely high standards of reliability required for data infrastructure, as well as concerns about the carbon footprint of data, the ideal private sector investment climate should provide for reliable, clean, low-cost electricity, natural cooling, and negligible disaster risk—conditions that are not always readily met in low- and middle-income countries.

Data laws and regulations can help create an environment of trust

Trust in data transactions can be supported through a robust legal and regulatory framework encompassing both *safeguards* and *enablers* (chapter 6). The establishment of such a framework remains a work in progress across all country income groups (figure O.4).

Safeguards promote trust in data transactions by avoiding or limiting harm arising from the misuse of data. A fundamental prerequisite for trust in data systems is cybersecurity. Achieving adequate cybersecurity calls for creating a legal framework that obliges data controllers and processers to adopt technical systems to secure data.[18] To date, only a small minority of low- and middle-income countries have adopted adequate legal frameworks for cybersecurity. Kenya's new Data Protection Act stands out as a good example of comprehensive cybersecurity provisions.

Creation of an adequate legal framework for data protection is also critical. Such a framework should clearly differentiate between personal data (data that identify the individual) and nonpersonal data

(data that do not contain any personally identifiable information). Among middle-income countries, Mauritius is notable as having relatively well-developed safeguards for personal data. Indeed, it has distinguished itself as one of the first Sub-Saharan African countries to ratify the Council of Europe's Convention 108+ for the Protection of Individuals with Regard to the Processing of Personal Data.[19]

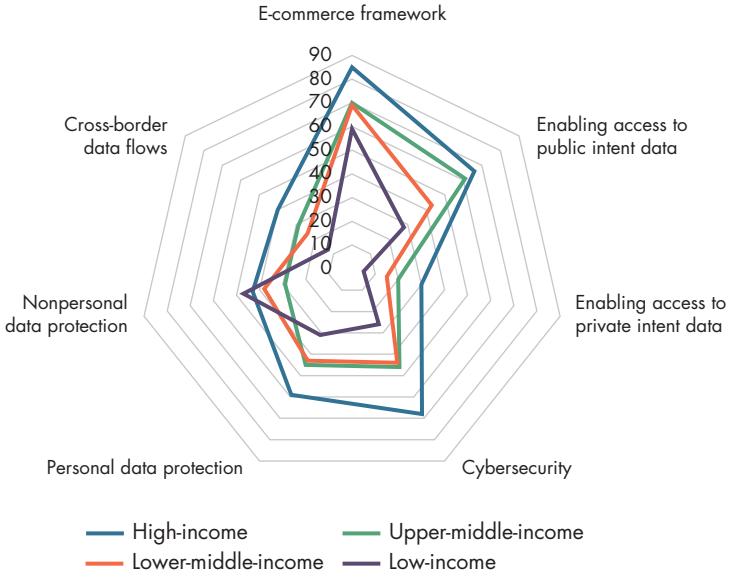
The protection of personal data is grounded in international human rights law, which requires that the interests of the data subject be adequately safeguarded *before* enabling any kind of data transaction.[20] This protection is usually achieved by compelling the subjects of data to provide some form of explicit consent for use of the data. But is such consent meaningful? Evidence indicates that it would take the average person 76 days a year to thoroughly read the numerous disclosure documents soliciting his or her consent to each website and application visited![21] This finding suggests the need to strengthen the legal obligations for data service providers to act in the best interest of the customers whose data are being used.

Because of the less sensitive nature of nonpersonal data, they can for the most part be adequately protected through intellectual property rights, allowing some balancing of interests between data protection and data reuse. However, this Report finds that most low-income countries surveyed do not have intellectual property rights in place for private intent data.

Complicating matters further, the distinction between personal data and nonpersonal data is becoming increasingly blurred. This blurring arises from the widespread mixing and processing of different data sources using sophisticated algorithms that may render nonpersonal data (such as from mobile phones) personally identifiable, or at least make it possible to identify specific social groups.

Enablers facilitate access to and reuse of data within and among stakeholder groups to ensure that the full social and economic value of data can be captured. The nature and extent of provisions to support data sharing differ markedly across public intent and private intent data. Significant efforts have been made around the world to safely disclose public intent data through open data policies (encouraging proactive publication of government data), together with access to information legislation (giving citizens a legally enforceable right to compel disclosure). For real impact, however, open data policies must be supported by a consistent protocol for classifying sensitive data, combined with interoperable technical standards, machine readable formats, and open licensing to facilitate subsequent reuse.

Figure O.4 The legal and regulatory framework for data governance remains a work in progress across all country income groupings



Source: WDR 2021 team, based on Global Data Regulation Survey conducted exclusively for this Report (https://microdata.worldbank.org/index.php/catalog/3866). Data at http://bit.do/WDR2021-Fig-O_4.

Note: The figure depicts the percentage of good practice laws and regulatory measures in place for countries covered by the survey in each country income group.

Governments have much less influence when it comes to disclosure of private intent data. Sharing of such data may serve as a remedy for the concentration of market power, such as in the Arab Republic of Egypt, where a merger between two major ride-hailing applications was made conditional on their sharing driver and rider information with smaller competitors. In other contexts, private intent data may also be critical for addressing important public policy challenges, such as the use of mobile phone records for contact tracing to control the spread of COVID-19. And yet relatively little attention has been paid so far to the possibility of incentivizing the exchange of private intent data through measures such as open licensing, data portability, and various types of data partnerships. Some countries—notably France—have nonetheless enacted legislation mandating the sharing of private sector data deemed to be in the public interest.[22]

Aligning data regulation with economic policy objectives can support the creation of value

Data play a central role in rapidly expanding platform-based business models. For example, search engines collect data on users' site visits, which they can sell to marketing companies so they can target

advertisements more precisely. These platform-based business models are becoming increasingly important in low- and middle-income countries. The design of legal and regulatory frameworks for data has a real impact on the buoyancy of data-driven businesses and requires difficult policy balances. Providing access to essential sources of market data, for example, may be critical for promoting competition among platform businesses, but it also may affect incentives for investment and innovation in data-driven businesses. Again, regulations designed to protect personal data may restrict cross-border data flows and materially affect a country's competitive edge in the burgeoning trade of data-enabled services (chapter 7).

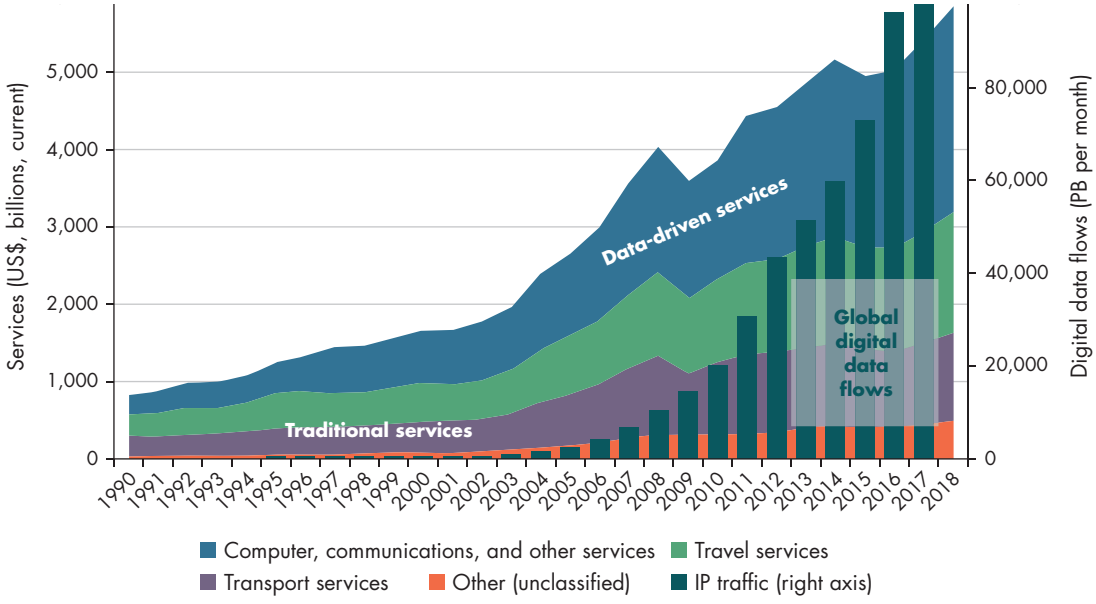
Competition and antitrust policy. Competition policy plays a critical role in ensuring that the value created by platform-based business models is equitably shared by producers and consumers. The presence of economies of scale in data collection externalities that increase the value of networks as more participants join platforms may lead to rapid accumulation of market power. Addressing such market dominance calls for two complementary strategies.

First, in countries that have sufficient capacity to enforce antitrust regulation, ex post antitrust enforcement should be applied—albeit with any adaptations that may be needed to address the challenges posed by data-driven businesses. For example, the standard test of market dominance—overpricing by a market leader—may not be meaningful in sectors where platforms routinely provide consumer services for free. However, even though several landmark antitrust cases involving platform businesses have emerged in middle-income countries, such as Egypt, India, and Mexico, this Report finds that not a single low-income country has completed such a case, despite the presence of the same globally dominant firms in these markets.

Second, in parallel with antitrust efforts, ex ante regulatory measures to make essential data accessible to rival firms and new entrants also merit serious consideration, as does empowering consumers to switch among competing providers by mandating full portability of their personal data. Care should always be taken to verify that access to data is critical for competition and does not unduly affect incentives for innovation in data-driven businesses.

Trade policy. Platform-based businesses also open up new avenues for international trade, entailing substantial cross-border flows of data (figure O.5). A country's regulatory framework for personal data protection has a material impact on participation in such trade,

Figure O.5 Since 1990, the global trade in data-driven services has grown exponentially and now constitutes half of trade in services



Source: WDR 2021 team calculations, based on World Bank, WITS (World Integrated Trade Solution) database, http://wits.worldbank.org/WITS/. Data at http://bit.do/WDR2021-Fig-O_5.

Note: IP = Internet Protocol; PB = petabytes.

creating some tension between trust and value creation. Countries have adopted a variety of approaches for dealing with this. Some, notably the federal jurisdiction in the United States, permit open data flows based on private sector standards, with limited government involvement. Others, such as China, Nigeria, the Russian Federation, and Vietnam, apply more stringent regulatory requirements, requiring copies of certain personal data to be stored domestically (data localization) and state authorization for many international exchanges. In between are countries (including members of the European Union and others such as Argentina and South Africa) that make cross-border transfers of personal data conditional on whether the partner trading country offers an adequate data protection regime. This Report finds that a combination of well-defined domestic personal data protection measures with relative ease of cross-border movements appears to offer the most favorable environment for international trade in data-enabled services.[23]

Tax policy. Even though data-driven transactions are creating more economic activity, the governments of low- and middle-income countries are struggling to share equitably in this value by mobilizing the associated tax revenues. For indirect taxes (such as value added taxes), the revenue rights are clearly allocated to the country in which the final sales are made. However, the administrative capacity to capture this revenue is typically lacking. Estimates for East Asian countries suggest that losses to fiscal revenues could amount to as much as 1 percent of the gross domestic product (GDP) by 2030.[24] As for direct taxes (such as corporate taxes), agreed-on international rules are lacking for allocating rights to tax businesses that operate in markets without any physical presence. In the absence of such a consensus, an increasing number of countries have been resorting to the application of ad hoc digital service taxes as a compensatory measure.

Sound institutions and governance can improve the development impact of data

If institutions do not function well, policies and laws and regulations are unlikely to be implemented or enforced effectively, and infrastructure will not deliver on its potential. An effective institutional framework for data governance must fulfill several critical functions, such as setting policy objectives, developing supporting rules and standards, enforcing compliance with such regulations, and continually improving governance through learning and evaluation (chapter 8). For example, Uruguay's creation of a lead agency close to the Office of the President and acting with a whole-of-government perspective has been critical in driving the country's successful e-government reforms since 2007.

Although there is no one-size-fits-all approach for governments seeking to create robust institutional arrangements for data governance, certain institutional design characteristics are of universal importance. Institutions should be formally mandated, sufficiently resourced, and have the technical capacity needed to effectively undertake their functions in a coordinated manner across the whole of government. The main institutional actors within this framework often include data governance entities, data protection authorities, and cybersecurity agencies, as well as new types of institutions such as data trusts—accountability-oriented data intermediaries allowing individuals to pool their legal rights over data and assign them to trustees with explicit fiduciary duties. Institutional independence and functional autonomy may be critical in some cases to shield data governance institutions from undue political or commercial influence. Behavioral and cultural norms and political economy constraints often stymie reform efforts, creating implementation gaps, especially in low- and middle-income countries. Change management, collaborative leadership, and a culture of performance and incentives can help institutions overcome barriers to implementation and coordination and effectively perform their roles and responsibilities.

To maximize buy-in from all participants in the data governance ecosystem, including society more broadly, data management must be socially inclusive and perceived as legitimate. Legitimacy is enhanced when governments manage and use data in a transparent manner and are subject to meaningful systems of accountability. Nongovernmental actors and emerging mechanisms such as data intermediaries can play an important role in the ecosystem by helping governments and end users responsibly share and use data to better harness their development value, while safeguarding against the risks of misuse or abuse. Engaging with stakeholders, across society and internationally, in a collaborative and transparent manner will foster trust and legitimacy and strengthen the social contract around data use. For example, the Association of Southeast Asian Nations (ASEAN) has adopted a regional Framework on Digital Data Governance, which helps coordinate members' data governance arrangements with a view toward interoperability.

Moving toward an integrated national data system

A well-functioning data governance framework ensures that infrastructure, laws, economic policies, and institutions work together to support the use of data in a way that aligns with each society's values, while protecting individuals' rights over use of their data. This framework defines the rules, and associated compliance mechanisms, for how data can be safely shared, used, and reused by all stakeholders.

Part III of this Report concludes with an aspirational vision of an integrated national data system (INDS) that can deliver on the promise of producing high-quality data and then making data open in a way that they are both protected and accessible to be shared and reused by all stakeholders (chapter 9). The aspirational INDS works seamlessly with the governance structure. If the governance framework can be viewed as creating and enforcing the "rules of the road," the INDS can be seen as the "network of highways" that connect all users, ensuring safe passage of data to and from destinations.

The INDS is built on an intentional, whole-of-government, multistakeholder approach to data governance. It explicitly builds data production, protection, exchange, and use into planning and decision-making across government entities and actively *integrates* the various stakeholders from civil society, the public sector, and the private sector into the data life cycle and into the governance structures of the system.[25]

A well-functioning system requires people to produce, process, and manage high-quality data; people to populate the institutions that safeguard and protect the data against misuse; and people to draft, oversee, and implement data strategies, policies, and regulations. The system also needs people to hold the public and private sectors accountable and people capable of using data from the production process of private firms to improve policies in the public sector. All this requires robust data literacy so that a wide cross section of people benefit from an INDS.

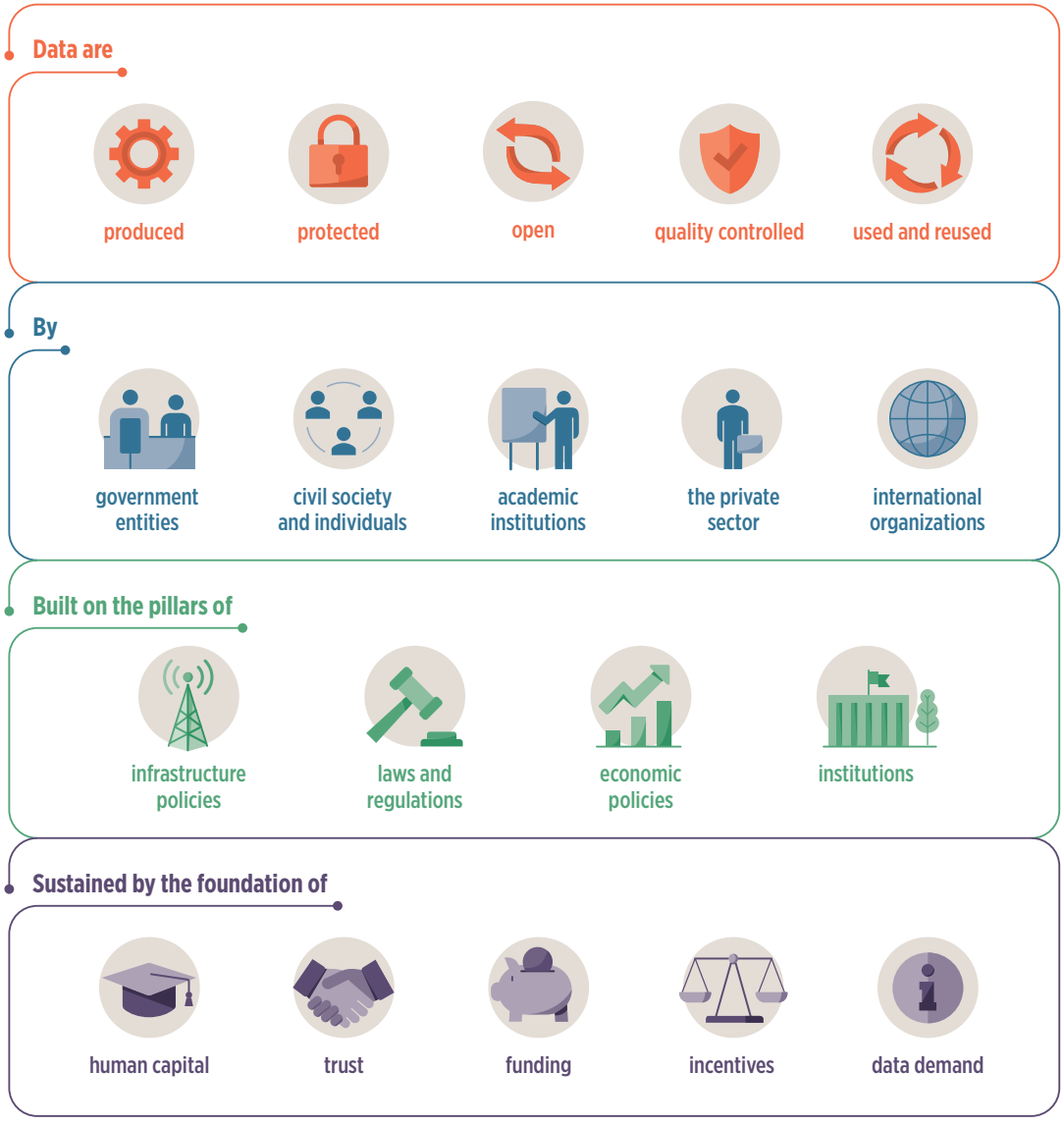
For a sound INDS, institutions and actors must also have the right incentives to produce, protect, and share data, and funding must be sufficient to implement the infrastructure and institutions needed for the system to function well. Finally, a culture of data use helps foster a high-quality supply of data and stimulate the demand for data-informed decision-making without which the national data system is not sustainable.

When government agencies, civil society, academia, and the private sector securely take part in a national data system, the potential uses of data expand and so does the potential impact on development. In fact, the more integrated the system and the more participants involved, the higher is the potential return. If two participants safely exchange data, data can flow in two directions. If three participants exchange data, data can flow in six directions, and with four participants, in 12 directions. As data are reused and repurposed, these connections will increase rapidly. Higher degrees of integration require close coordination and shared governance between participants, but such integration is otherwise compatible with a decentralized data architecture. The system is designed to ensure that data flow freely and safely—not remain in one place. A distributed system may be best placed to achieve this safe flow of data.

Even though most countries are far away from the aspirational goal of developing a well-functioning data system, setting sights on this target can provide countries with guidance on the next steps in developing such a system. How countries move toward this vision of an INDS will depend on their current capacity and the parameters of the social contract for data. There is no singular blueprint for how to build an INDS. Instead, this Report proposes a maturity model to help assess progress. Countries in the initial stages are likely to benefit the most from establishing the fundamentals for an integrated national data system. This includes developing policies and strategies aimed at better data governance, strengthening the technical capacity for data production and use of government agencies and the national statistical office, and promoting data literacy through education and training. With the fundamentals in place, governments can work on initiating and systemizing data flows across and between the participants in the national data system. This requires policies and standards that ensure the consistency and interoperability of data and institutions and infrastructure to enable the secure exchange of data that mitigates privacy risks. At advanced levels of data maturity, the goal is to optimize the system through shared data governance and collaboration between the various stakeholders from government, international organizations, civil society, and the private sector.

The structure of this system will differ from country to country, reflecting local norms for the safe reuse and sharing of data. Nonetheless, certain common attributes are needed to realize the

Figure O.6 What happens in an integrated national data system?



Source: WDR 2021 team.

development gains from reusing and sharing data. A well-functioning data system defines and establishes the authority and responsibility for data production, flow, and use in a nation. This system would build on the infrastructure, policies, laws and regulations, and institutions discussed here; integrate the many sources of data; and connect all the stakeholders (figure O.6).

For many countries, a system in which high-quality data flow and are used safely among various participants remains a distant vision. A low-income country suffering from high levels of poverty, fragility, and poor governance may struggle to produce even the most fundamental data, let alone set up a whole-of-government, multistakeholder approach to data governance. Yet keeping this vision in sight matters for all countries, even those struggling the most with data, because it can serve as a guide in making decisions on how to develop their data systems (box O.1).

Coalescing around a common understanding of a new social contract for data—one built on trust to produce value from data that are equitably distributed—and finding the right blueprint for building an

Box O.1 Toward an integrated national data system: Country examples

Important steps in the right direction. Many countries have adopted important initiatives that embody aspects of what is envisioned in an integrated national data system. South Africa's Department of Planning, Monitoring and Evaluation has developed a system that includes the data produced by citizens who monitor the performance of government programs. In Chile, civil society participation is mandated by the 2011 Law on Associations and Citizen Participation in Public Management, and the national statistical office has put in place a civil society council. The inclusion of multiple stakeholders in a national data system encourages sustainability and helps ensure that all participants have an opportunity to access and benefit from it. The Nepal Data Literacy Program, established in 2019, comprises a 100-hour modular, customizable pedagogy to support both technical skills building and efforts to enhance a culture of data use among Nepalis. The program is now partnering with Kathmandu University School of Management (KUSOM) to incorporate data literacy toolkits into the university programs and develop a data-driven course that will be free to other institutions and thousands of students.

A fully realized vision. In Estonia, the government has set up a national data system to safely manage citizens' personal data for use by government agencies and participating businesses. X-Road is an open-source data exchange layer solution that allows linked public and private databases to automatically share information, ensuring confidentiality, integrity, and interoperability between data exchange parties. It combines a technical solution (enabling technical architecture and a series of protocols) with a governance solution (the once-only principle enshrined in national law that obliges public sector agencies to refrain from duplicating data requests). Under this system, citizens have to supply government agencies and participating businesses with their information only once. It is then automatically transmitted to other participating entities. X-Road's cryptography protocols also enhance transparency because they log entries into the system and give individuals detailed insights into who is sharing their data and for what purposes.

The X-Road arrangement both builds on and enhances Estonia's social contract on data by providing trust, equity, and value. Its transparency engenders trust. Its national scope, available to all, promotes equity. Its ease and comprehensiveness provide value. To work well, this digital data system depends on some "analogue" components. Cooperation is fostered between government and the private sector and between components of infrastructure. Change management is built into the entire system, from its foundations in national law (and the social contract) to its design, uptake, and upkeep. A culture of trust and sharing (data sharing) is encouraged.

integrated national data system—one that unleashes the value of data to improve lives through creative, innovative applications by a widening array of users—are highly aspirational goals. Achieving these goals will require significant changes in how data are produced, managed, protected, shared, and used. Making these changes will be difficult and will depend on substantial commitment and effort, but the cost of failing to change is a world faced with greater inequities and many missed opportunities.

Notes

1. Chapter 1 reviews definitions of data and describes how the term *data* is used in this Report.
2. Amnesty International (2019); Zuboff (2019).
3. Scheer and Cohen (2020).
4. Scheer and Cohen (2020).
5. World Bank (2017).
6. FMWR (2018).
7. World Bank, Statistical Performance Indicators (database), http://documents.worldbank.org/curated/en/815721616086786412/Measuring-the-Statistical-Performance-of-Countries-An-Overview-of-Updates-to-the-World-Bank-Statistical-Capacity-Index.
8. Brynjolfsson, Hitt, and Kim (2011).
9. For an extensive discussion of this problem and many other concerns about machine learning, see O'Neil (2017).
10. For an early reference to "garbage in–garbage out" in the statistical literature, see Parzen (1964).
11. Aiken et al. (2020).
12. Burke and Lobell (2017); Osgood-Zimmerman et al. (2018).
13. An early illustration of how data can be used to improve accountability for public expenditure can be found in a study of the use of education budgets in Uganda by Reinikka and Svensson (2001).
14. See Broadband Commission for Sustainable Development, International Telecommunication Union, "Target 3:

Connectivity" (accessed October 31, 2020), https://broadbandcommission.org/Pages/targets/Target-3.aspx.

15. Chen (forthcoming). Analysis is based on Access Survey 2017–18 data collected by Research ICT Africa in 22 low- and middle-income countries across Africa, Asia, and Latin America.
16. GSMA (2019).
17. World Bank analysis of World Trade Organization ad valorem duties for "Telephones for cellular networks 'mobile telephones' or for other wireless networks" (Harmonized System code 851712).
18. ITU et al. (2018).
19. COE (2018).
20. Safeguards for personal data are grounded in a human rights framework based on international law. These safeguards have their origin in the establishment of the "rule of law" with the expression of individual rights in the Enlightenment and were codified in international law after World War II. They were further refined in the context of analog data in the 1970s and 1980s with the Fair Information Practices, the Council of Europe's 1981 Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data (Convention 108), and the first guidelines issued by the Organisation for Economic Co-operation and Development (OECD). The OECD guidelines and Convention 108 were updated in the digital context after launch of the World Wide Web in 1995 and continue to evolve.
21. Madrigal (2012).
22. OECD (2019).
23. Ferracane and van der Marel (forthcoming).
24. Al-Rikabi and Loeprick (forthcoming).
25. An integrated national data system does not imply that all data are integrated in a national database. Instead, various participants are integrated in a system in which data are safely flowing and used. This is akin to a national statistical system in the sense that an ensemble of participants jointly collects, protects, processes, and disseminates official statistics. But unlike in the national statistical system, the scope of an integrated national data system goes well beyond official statistics; it requires an intentional approach to governing the participants and their roles.

References

Aiken, Emily L., Guadalupe Bedoya, Aidan Coville, and Joshua E. Blumenstock. 2020. "Targeting Development Aid with Machine Learning and Mobile Phone Data: Evidence from an Anti-Poverty Intervention in Afghanistan." In COMPASS '20: Proceedings of the 3rd ACM SIGCAS Conference on Computing and Sustainable Societies, 310–11. New York: Association for Computing Machinery.

Al-Rikabi, Jaffar, and Jan Loeprick. Forthcoming. "Simulating Potential Tax Revenues from Data-Driven Platform Businesses in East Asia." WDR 2021 background paper, World Bank, Washington, DC.

Amnesty International. 2019. "Surveillance Giants: How the Business Model of Google and Facebook Threatens Human Rights." Report POL 30/1404/2019, Amnesty International, London. https://www.amnesty.org/en/documents/document/?indexNumber=pol30%2f1404%2f2019&language=en.

Brynjolfsson, Erik, Lorin M. Hitt, and Heekyung Hellen Kim. 2011. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" SSRN Scholarly Paper ID 1819486, Social Science Research Network, Rochester, NY. doi.org/10.2139/ssrn.1819486.

Burke, Marshall, and David Lobell. 2017. "Satellite-Based Assessment of Yield Variation and Its Determinants in Smallholder African Systems." *PNAS, Proceedings of the National Academy of Sciences* 114 (9): 2189–94. doi.org/10.1073/pnas.1616919114.

Chen, Rong. Forthcoming. "A Demand-Side View of Mobile Internet Adoption in the Global South." WDR 2021 background paper, World Bank, Washington, DC.

COE (Council of Europe). 2018. "Convention 108+: Convention for the Protection of Individuals with Regard to the Processing of Personal Data." COE, Strasbourg. https://rm.coe.int/convention-108-convention-for-the-protection-of-individuals-with-regar/16808b36f1.

Ferracane, Martina Francesca, and Erik van der Marel. Forthcoming. "Regulations on Personal Data: Differing Data Realms and Digital Services Trade." WDR 2021 background paper, World Bank, Washington, DC.

FMWR (Federal Ministry of Water Resources, Nigeria). 2018. "National Action Plan for Revitalization of the WASH Sector." FMWR, Abuja.

GSMA (GSM Association). 2019. "Connected Society: State of Mobile Internet Connectivity 2019." GSMA, London. https://www.gsma.com/mobilefordevelopment/resources/the-state-of-mobile-internet-connectivity-report-2019/.

ITU (International Telecommunication Union), World Bank, ComSec (Commonwealth Secretariat), CTO (Commonwealth Telecommunications Organisation), and NATO (Cooperative Cyber Defence Centre of Excellence, North Atlantic Treaty Organization). 2018. "Guide to Developing a National Cybersecurity Strategy: Strategic Engagement in Cybersecurity." ITU, Geneva. https://www.itu.int/pub/D-STR-CYB_GUIDE.01-2018.

Knippenberg, Erwin, and Moritz Meyer. 2020. "The Hidden Potential of Mobile Phone Data: Insights on COVID-19 in The Gambia." *Data Blog* (blog). September 10, 2020. https://blogs.worldbank.org/opendata/hidden-potential-mobile-phone-data-insights-covid-19-gambia.

Madrigal, Alexis C. 2012. "Reading the Privacy Policies You Encounter in a Year Would Take 76 Work Days." *Atlantic*, March 1, 2012. https://www.theatlantic.com/technology/archive/2012/03/reading-the-privacy-policies-you-encounter-in-a-year-would-take-76-work-days/253851/.

OECD (Organisation for Economic Co-operation and Development). 2019. *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-Use across Societies*. Paris: OECD. https://www.oecd-ilibrary.org/science-and-technology/enhancing-access-to-and-sharing-of-data_276aaca8-en.

O'Neil, Cathy. 2017. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York: Broadway Books.

Osgood-Zimmerman, Aaron, Anoushka I. Millear, Rebecca W. Stubbs, Chloe Shields, Brandon V. Pickering, Lucas Earl, Nicholas Graetz, et al. 2018. "Mapping Child Growth Failure in Africa between 2000 and 2015." *Nature* 555 (7694): 41–47. doi.org/10.1038/nature25760.

Parzen, Emanuel. 1964. "Review of Smoothing, Forecasting and Prediction of Discrete Time Series, by Robert Goodell Brown." *Journal of the American Statistical Association* 59 (307): 973–74. doi.org/10.2307/2283122.

Reinikka, Ritva, and Jakob Svensson. 2001 "Explaining Leakage of Public Funds." Policy Research Working Paper 2709, World Bank, Washington, DC.

Scheer, Steven, and Tova Cohen. 2020. "Israel Extends Coronavirus Cell Phone Surveillance by Three Weeks." *Emerging Markets* (blog), May 5, 2020. https://www.reuters.com/article/us-health-coronavirus-israel-surveillanc/israel-extends-coronavirus-cell-phone-surveillance-by-three-weeks-idUSKBN22H11I.

World Bank. 2017. *A Wake Up Call: Nigeria Water Supply, Sanitation, and Hygiene Poverty Diagnostic*. WASH Poverty Diagnostic Series. Washington, DC: World Bank. https://openknowledge.worldbank.org/handle/10986/27703.

World Bank. 2019. "Tanzania, Mainland Poverty Assessment 2019: Executive Summary." World Bank, Washington, DC. https://www.worldbank.org/en/country/tanzania/publication/tanzanias-path-to-poverty-reduction-and-pro-poor-growth.

Zuboff, Shoshana. 2019. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*. New York: PublicAffairs. https://www.hbs.edu/faculty/Pages/item.aspx?num=56791.

Contents of the *World Development Report 2021: Data for Better Lives*

ECO-AUDIT

Environmental Benefits Statement

The World Bank Group is committed to reducing its environmental footprint. In support of this commitment, we leverage electronic publishing options and print-on-demand technology, which is located in regional hubs worldwide. Together, these initiatives enable print runs to be lowered and shipping distances decreased, resulting in reduced paper consumption, chemical use, greenhouse gas emissions, and waste.

We follow the recommended standards for paper use set by the Green Press Initiative. The majority of our books are printed on Forest Stewardship Council (FSC)–certified paper, with nearly all containing 50–100 percent recycled content. The recycled fiber in our book paper is either unbleached or bleached using totally chlorine-free (TCF), processed chlorine–free (PCF), or enhanced elemental chlorine–free (EECF) processes.

More information about the Bank's environmental philosophy can be found at http://www.worldbank.org/corporateresponsibility.

green press INITIATIVE

Today's unprecedented growth of data and their ubiquity in our lives are signs that the data revolution is transforming the world. And yet much of the value of data remains untapped. Data collected for one purpose have the potential to generate economic and social value in applications far beyond those originally anticipated. But many barriers stand in the way, ranging from misaligned incentives and incompatible data systems to a fundamental lack of trust. *World Development Report 2021: Data for Better Lives* explores the tremendous potential of the changing data landscape to improve the lives of poor people, while also acknowledging its potential to open back doors that can harm individuals, businesses, and societies. To address this tension between the helpful and harmful potential of data, this Report calls for a new social contract that enables the use and reuse of data to create economic and social value, ensures equitable access to that value, and fosters trust that data will not be misused in harmful ways.

This Report begins by assessing how better use and reuse of data can enhance the design of public policies, programs, and service delivery, as well as improve market efficiency and job creation through private sector growth. Because better data governance is key to realizing this value, the Report then looks at how infrastructure policy, data regulation, economic policies, and institutional capabilities enable the sharing of data for their economic and social benefits, while safeguarding against harmful outcomes. The Report concludes by pulling together the pieces and offering an aspirational vision of an integrated national data system that would deliver on the promise of producing high-quality data and making them accessible in a way that promotes their safe use and reuse. By examining these opportunities and challenges, the Report shows how data can benefit the lives of all people, but particularly poor people in low- and middle-income countries.

WORLD BANK GROUP

SKU 33424